Noname manuscript No.
(will be inserted by the editor)

Second-order smoothing approximation to l_1 exact penalty function for nonlinear constrained optimization problems

Nguyen Thanh Binh

Received: date / Accepted: date

Abstract In this paper, a second-order smoothing approximation to the l_1 exact penalty function for nonlinear constrained optimization problems is presented. Error estimations are obtained among the optimal objective function values of the smoothed penalty problem, of the nonsmooth penalty problem and of the original optimization problem. Based on the smoothed penalty problem, an algorithm that has better convergence is presented. Numerical examples illustrate that this algorithm is efficient in solving nonlinear constrained optimization problems.

Keywords Exact penalty function · Smoothing method · Constrained optimization · Approximate optimal solution

Mathematics Subject Classification (2000) 90C30 · 65K05

1 Introduction

Consider the nonlinear inequality constrained optimization problem:

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

where $f, g_i : R^n \to R,\ i \in I = \{1, 2, \ldots, m\}$ are twice continuously differentiable functions and $X_0 = \{x \in R^n \mid g_i(x) \le 0,\ \ i = 1, 2, \ldots, m\}$ is the feasible set to (P).

To solve (P), many exact penalty function methods have been introduced in literatures, see, [1,3,4,5,7,13,24]. In 1967, Zangwill [24] first the classical l_1 exact penalty function as follows:

$$F(x, \rho) = f(x) + \rho \sum_{i=1}^{m} \max\{g_i(x), 0\}, \tag{1}$$

This work is supported by National Natural Science Foundation of China (Grant No. 11371242)

N.T. Binh(✉)
Department of Mathematics, Shanghai University, Shanghai, 200444, China
E-mail: bingbongyb@gmail.com

where $\rho > 0$ is a penalty parameter.

Note that penalty function (1) is not a smooth function. The obvious difficulty with the exact penalty function is that it is non-differentiable, which prevents the use of efficient minimization algorithms and may cause some numerical instability problems in its implementation. In order to avoid the drawback related to the non-differentiability, the smoothing methods of the exact penalty functions attracts much attention, see, [2,8,9,10,11,12,17,18,19,20,21,23,25]. Chen et al. [2] introduced a smooth function to approximate the classical l_1 penalty function by integrating the sigmoid function $1/(1+e^{-\alpha t})$. Lian [8] and Wu et al. [18] proposed a smoothing approximation to l_1 exact penalty function for inequality constrained optimization. Pinar et al. [12] also proposed a smoothing approximation to l_1 exact penalty function and an ε-optimal minimum can be obtained by solving the smoothed penalty problem. Xu et al. [20] discussed a second-order differentiability smoothing to the classical l_1 exact penalty function for constrained optimization problems. Meng et al. [10] introduced a smoothing of the square-root exact penalty function for inequality constrained optimization. However, little attention has been paid to smoothing the exact penalty function in terms of second-order differentiability. So, here we present a second-order smoothing approximation to the l_1 exact penalty function (1), and based on the smoothed penalty function obtained thereafter an algorithm for solving nonlinear constrained optimization problems is given in this paper.

The rest of this paper is organized as follows. In Section 2, we introduce a s-moothing function for the classical l_1 exact penalty function and some fundamental properties of the smoothing function. In Section 3, the algorithm based on the s-moothed penalty function is proposed and its global convergence is presented, with some numerical examples given. Finally, conclusions are given in Section 4.

2 Second-order smoothing penalty function

Let $q(t) = \max\{t,0\}$. Then, the penalty function (1) is turned into

$$G(x,\rho) = f(x) + \rho \sum_{i=1}^{m} q\left(g_i(x)\right), \tag{2}$$

where $\rho > 0$. The corresponding penalty optimization problem to $G(x,\rho)$ is defined as

$$(P_\rho) \quad \min\ G(x,\rho), \qquad \text{s.t. } x \in R^n.$$

In order to $q(t)$, we define function $q_\varepsilon(t) : R^1 \to R^1$ as

$$q_\varepsilon(t) = \begin{cases} 0 & \text{if } t < 0, \\ \dfrac{t^3}{9\varepsilon^2} & \text{if } 0 \le t < \varepsilon, \\ t + \dfrac{2}{3}\varepsilon e^{-\frac{t}{\varepsilon}+1} - \dfrac{14\varepsilon}{9} & \text{if } t \ge \varepsilon, \end{cases}$$

where $\varepsilon > 0$ is a smoothing parameter.

Remark 1 Obviously, $q_\varepsilon(t)$ has the following attractive properties:

(i) For any $\varepsilon > 0,\ q_\varepsilon(t)$ is twice continuously differentiable on R, where

$$q'_\varepsilon(t) = \begin{cases} 0 & \text{if} \quad t < 0, \\ \dfrac{t^2}{3\varepsilon^2} & \text{if} \quad 0 \le t < \varepsilon, \\ 1 - \dfrac{2}{3}e^{-\frac{t}{\varepsilon}+1} & \text{if} \quad t \ge \varepsilon, \end{cases}$$

and

$$q''_\varepsilon(t) = \begin{cases} 0 & \text{if} \quad t < 0, \\ \dfrac{2t}{3\varepsilon^2} & \text{if} \quad 0 \le t < \varepsilon, \\ \dfrac{2}{3\varepsilon}e^{-\frac{t}{\varepsilon}+1} & \text{if} \quad t \ge \varepsilon. \end{cases}$$

(ii) $\lim\limits_{\varepsilon\to 0} q_\varepsilon(t) = q(t)$.

(iii) $q_\varepsilon(t)$ is convex and monotonically increasing in t for any given $\varepsilon > 0$. Property (iii) follow from (i) immediately.

Suppose that f and $g_i (i = 1,2,\ldots,m)$ are second-order continuously differentiable. Consider the penalty function for (P) given by

$$G_\varepsilon(x,\rho) = f(x) + \rho \sum_{i=1}^{m} q_\varepsilon\left(g_i(x)\right). \tag{3}$$

Clearly, $G_\varepsilon(x,\rho)$ is second-order continuously differentiable on R^n. Applying (3), the following penalty problem for (P) is obtained

$$(SP_{\rho,\varepsilon}) \quad \min\ G_\varepsilon(x,\rho), \qquad \text{s.t.}\ \ x \in R^n.$$

Now, the relationship between (P_ρ) and $(SP_{\rho,\varepsilon})$ is studied.

Lemma 1 *For any given $x \in R^n$, $\varepsilon > 0$ and $\rho > 0$, we have*

$$0 \le G(x,\rho) - G_\varepsilon(x,\rho) \le \frac{14m\rho\varepsilon}{9}. \tag{4}$$

Proof For $x \in R^n$ and $i \in I$, by the definition of $q_\varepsilon(t)$, we have

$$q\left(g_i(x)\right) - q_\varepsilon\left(g_i(x)\right) = \begin{cases} 0 & \text{if} \quad g_i(x) < 0, \\ g_i(x) - \dfrac{g_i(x)^3}{9\varepsilon^2} & \text{if} \quad 0 \le g_i(x) < \varepsilon, \\ \dfrac{14\varepsilon}{9} - \dfrac{2}{3}\varepsilon e^{-\frac{g_i(x)}{\varepsilon}+1} & \text{if} \quad g_i(x) \ge \varepsilon. \end{cases}$$

That is,

$$0 \le q\left(g_i(x)\right) - q_\varepsilon\left(g_i(x)\right) \le \frac{14\varepsilon}{9}, \qquad i = 1,2,\ldots,m.$$

Thus,

$$0 \le \sum_{i=1}^{m} q\left(g_i(x)\right) - \sum_{i=1}^{m} q_\varepsilon\left(g_i(x)\right) \le \frac{14m\varepsilon}{9},$$

which implies

$$0 \le \rho \sum_{i=1}^{m} q\left(g_i(x)\right) - \rho \sum_{i=1}^{m} q_\varepsilon\left(g_i(x)\right) \le \frac{14m\rho\varepsilon}{9}.$$

Therefore,

$$0 \le \left\{ f(x) + \rho \sum_{i=1}^{m} q\left(g_i(x)\right) \right\} - \left\{ f(x) + \rho \sum_{i=1}^{m} q_\varepsilon\left(g_i(x)\right) \right\} \le \frac{14m\rho\varepsilon}{9},$$

that is,

$$0 \le G(x,\rho) - G_\varepsilon(x,\rho) \le \frac{14m\rho\varepsilon}{9}.$$

The proof completes. □

A direct result of Lemma 1 is given as follows.

Corollary 1 *Let $\{\varepsilon_j\} \to 0$ be a sequence of positive numbers and assume x^j is a solution to $(SP_{\rho,\varepsilon})$ for some given $\rho > 0$. Let x' be an accumulation point of the sequence $\{x^j\}$. Then x' is an optimal solution to (P_ρ).*

Definition 1 For $\varepsilon > 0$, a point $x_\varepsilon \in R^n$ is called ε-feasible solution to (P) if $g_i(x_\varepsilon) \le \varepsilon,\ \forall i \in I$.

Definition 2 For $\varepsilon > 0$, a point $x_\varepsilon \in X_0$ is called ε-approximate optimal solution to (P) if

$$|f^* - f(x_\varepsilon)| \le \varepsilon,$$

where f^* is the optimal objective value of (P).

Theorem 1 *Let x^* be an optimal solution of problem (P_ρ) and x' be an optimal solution to $(SP_{\rho,\varepsilon})$ for some $\rho > 0$ and $\varepsilon > 0$. Then,*

$$0 \le G(x^*,\rho) - G_\varepsilon(x',\rho) \le \frac{14m\rho\varepsilon}{9}. \tag{5}$$

Proof From Lemma 1, for $\rho > 0$, we have that

$$0 \le G(x^*,\rho) - G_\varepsilon(x^*,\rho) \le \frac{14m\rho\varepsilon}{9},$$
$$0 \le G(x',\rho) - G_\varepsilon(x',\rho) \le \frac{14m\rho\varepsilon}{9}.$$

Under the assumption that x^* is an optimal solution to (P_ρ) and x' is an optimal solution to $(SP_{\rho,\varepsilon})$, we get

$$G(x^*,\rho) \le G(x',\rho),$$
$$G_\varepsilon(x',\rho) \le G_\varepsilon(x^*,\rho).$$

Therefore, we obtain that

$$\begin{aligned} 0 &\le G(x^*,\rho) - G_\varepsilon(x^*,\rho) \le G(x^*,\rho) - G_\varepsilon(x',\rho) \\ &\le G(x',\rho) - G_\varepsilon(x',\rho) \le \frac{14m\rho\varepsilon}{9}. \end{aligned}$$

That is,

$$0 \le G(x^*,\rho) - G_\varepsilon(x',\rho) \le \frac{14m\rho\varepsilon}{9}.$$

This completes the proof. □

Lemma 2 **([19])** *Suppose that x^* is an optimal solution to (P_ρ). If x^* is feasible to (P), then it is an optimal solution to (P).*

Theorem 2 *Suppose that x^* satisfies the conditions in Lemma 2 and x' be an optimal solution to $(SP_{\rho,\varepsilon})$ for some $\rho > 0$ and $\varepsilon > 0$. If x' is ε-feasible to (P). Then,*

$$0 \le f(x^*) - f(x') \le \frac{5m\rho\varepsilon}{3}, \tag{6}$$

that is, x' is an approximate optimal solution to (P).

Proof Since x' is ε-feasible to (P), it follows that

$$\sum_{i=1}^{m} q_\varepsilon\left(g_i(x')\right) \le \frac{m\varepsilon}{9}.$$

As x^* is a feasible solution to (P), we have

$$\sum_{i=1}^{m} q\left(g_i(x^*)\right) = 0.$$

By Theorem 1, we get

$$0 \le \left\{ f(x^*) + \rho \sum_{i=1}^{m} q\left(g_i(x^*)\right) \right\} - \left\{ f(x') + \rho \sum_{i=1}^{m} q_\varepsilon\left(g_i(x')\right) \right\} \le \frac{14m\rho\varepsilon}{9}.$$

Thus,

$$\rho \sum_{i=1}^{m} q_\varepsilon\left(g_i(x')\right) \le f(x^*) - f(x') \le \rho \sum_{i=1}^{m} q_\varepsilon\left(g_i(x')\right) + \frac{14m\rho\varepsilon}{9}.$$

That is,

$$0 \le f(x^*) - f(x') \le \frac{5m\rho\varepsilon}{3}.$$

By Lemma 2, x^* is actually an optimal solution to (P). Thus x' is an approximate optimal solution to (P). This completes the proof. □

Theorem 1 show that an approximate solution to $(SP_{\rho,\varepsilon})$ is also an approximate solution to (P_ρ) when the error ε is sufficiently small. By Theorem 2, an optimal solution to $(SP_{\rho,\varepsilon})$ is an approximate optimal solution to (P) if it is ε-feasible to (P).

Definition 3 For $x^* \in R^n$, we call $y^* \in R^m$ a Lagrange multiplier vector corresponding to x^* if and only if x^* and y^* satisfy that

$$\nabla f(x^*) + \sum_{i=1}^{m} y_i^* \nabla g_i(x^*) = 0, \tag{7}$$

$$y_i^* g_i(x^*) = 0, \;\; g_i(x^*) \le 0, \;\; y_i^* \ge 0, \;\; i \in I. \tag{8}$$

Theorem 3 *Let f and $g_i,\;\; i \in I$ in (P) are convex. Let x^* be an optimal solution of (P) and $y^* \in R^m$ a Lagrange multiplier vector corresponding to x^*. Then for some $\varepsilon > 0$,*

$$G(x^*, \rho) - G_\varepsilon(x, \rho) \le \frac{14m\rho\varepsilon}{9}, \qquad \forall x \in R^n, \tag{9}$$

provided that $\rho \ge y_i^*,\; i = 1, 2, \ldots, m$.

Proof By the convexity of f and $g_i,\;\; i = 1, 2, \ldots, m$, we have

$$\begin{aligned} f(x) &\ge f(x^*) + \nabla f(x^*)^T (x - x^*), \;\; \forall x \in R^n, \\ g_i(x) &\ge g_i(x^*) + \nabla g_i(x^*)^T (x - x^*), \;\; \forall x \in R^n. \end{aligned} \tag{10}$$

Since x^* is an optimal solution of (P) and y^* is a Lagrange multiplier vector corresponding to x^*, by (7), (8), (9) and (10), we have

$$\begin{aligned} f(x) &\ge f(x^*) + \nabla f(x^*)^T (x - x^*) \\ &= f(x^*) - \sum_{i=1}^{m} y_i^* \nabla g_i(x^*)^T (x - x^*) \\ &\ge f(x^*) - \sum_{i=1}^{m} y_i^* \left(g_i(x) - g_i(x^*) \right) \\ &= f(x^*) - \sum_{i=1}^{m} y_i^* g_i(x). \end{aligned}$$

Since $g_i(x) \le g_i^+(x)$ ($g_i^+(x) = \max\{0, g_i(x)\},\; i \in I$), we have

$$\begin{aligned} G(x, \rho) = f(x) + \rho \sum_{i=1}^{m} g_i^+(x) &\ge f(x^*) - \sum_{i=1}^{m} y_i^* g_i(x) + \rho \sum_{i=1}^{m} g_i^+(x) \\ &\ge f(x^*) + \sum_{i=1}^{m} (\rho - y_i^*) g_i^+(x). \end{aligned}$$

Thus, for $\rho \ge y_i^*,\;\; i = 1, 2, \ldots, m$, we get $G(x, \rho) \ge f(x^*)$. Since x^* is feasible, then $f(x^*) = G(x^*, \rho)$ and by Lemma 1, we have

$$\begin{aligned} G(x^*, \rho) - G_\varepsilon(x, \rho) &= G(x^*, \rho) - G(x, \rho) + G(x, \rho) - G_\varepsilon(x, \rho) \\ &= f(x^*) - G(x, \rho) + G(x, \rho) - G_\varepsilon(x, \rho) \\ &\le G(x, \rho) - G_\varepsilon(x, \rho) \le \frac{14m\rho\varepsilon}{9}. \end{aligned}$$

This completes the proof. □

By Theorem 3, when the parameter ρ is sufficiently large, an approximate optimal solution to $(SP_{\rho,\varepsilon})$ is an approximate optimal solution to (P), where (P) is a convex problem. Therefore, we may obtain an approximate optimal solution to (P) by computing an approximate optimal solution to $(SP_{\rho,\varepsilon})$.

3 Algorithm and numerical examples

In this section, using the smoothed penalty function $G_\varepsilon(x,\rho)$, we propose an algorithm to solve nonlinear constrained optimization problems, defined as Algorithm I

Algorithm I

Step 1: Choose $x^0,\ \varepsilon > 0,\ \varepsilon_0 > 0,\ \rho_0 > 0,\ 0 < \eta < 1$ and $N > 1$, let $j = 0$ and go to Step 2.

Step 2: Use x^j as the starting point to solve

$$(SP_{\rho_j,\varepsilon_j}) \qquad \min_{x\in R^n} G_{\varepsilon_j}(x,\rho_j) = f(x) + \rho_j \sum_{i=1}^{m} q_{\varepsilon_j}\left(g_i(x)\right).$$

Let x^{j+1} be the optimal solution obtained (x^{j+1} is obtained by a quasi -Newton method).

Step 3: If x^{j+1} is ε-feasible to (P), then stop and we have obtained an approximate solution x^{j+1} of (P). Otherwise, let $\rho_{j+1} = N\rho_j,\ \ \varepsilon_{j+1} = \eta\varepsilon_j$ and $j = j+1$, then go to Step 2.

Remark 2 In this Algorithm I, as $N > 1$ and $0 < \eta < 1$, the sequence $\{\varepsilon_j\} \to 0$ $(j \to +\infty)$ and the sequence $\{\rho_j\} \to +\infty$ $(j \to +\infty)$.

In practice, it is difficult to compute $x^{j+1} \in arg \min\limits_{x\in R^n} G_{\varepsilon_j}(x,\rho_j)$. We generally look for the local minimizer or stationary point of $G(x,\rho_j,\varepsilon_j)$ by computing x^{j+1} such that $\nabla G_{\varepsilon_j}(x,\rho_j) = 0$.

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

$$\begin{aligned}
I^0(x) &= \{i \mid g_i(x) < 0,\ \ i \in I\},\\
I_\varepsilon^+(x) &= \{i \mid g_i(x) \geq \varepsilon,\ \ i \in I\},\\
I_\varepsilon^-(x) &= \{i \mid 0 \leq g_i(x) < \varepsilon,\ \ i \in I\}.
\end{aligned}$$

Then, the following result is obtained.

Theorem 4 *Assume that* $\lim\limits_{\|x\|\to+\infty} f(x) = +\infty$. *Let* $\{x^j\}$ *be the sequence generated by Algorithm I. Suppose that the sequence* $\{G_{\varepsilon_j}(x^j,\rho_j)\}$ *is bounded. Then* $\{x^j\}$ *is bounded and any limit point* x^* *of* $\{x^j\}$ *is feasible to (P), and satisfies*

$$\lambda\nabla f(x^*) + \sum_{i\in I} \mu_i \nabla g_i(x^*) = 0, \tag{11}$$

where $\lambda \geq 0$ *and* $\mu_i \geq 0,\ i = 1,2,\ldots,m$.

Proof First, we will prove that $\{x^j\}$ is bounded. Note that

$$G_{\varepsilon_j}(x^j,\rho_j) = f(x^j) + \rho_j \sum_{i=1}^{m} q_{\varepsilon_j}\left(g_i(x^j)\right), \qquad j = 0,1,2,\ldots, \tag{12}$$

and by the definition of $q_\varepsilon(t)$, we have

$$\sum_{i=1}^{m} q_{\varepsilon_j}\left(g_i(x^j)\right) \geq 0. \tag{13}$$

Suppose to the contrary that $\{x^j\}$ is unbounded. Without loss of generality, we assume that $\|x^j\| \to +\infty$ as $j \to +\infty$. Then, $\lim\limits_{j\to+\infty} f(x^j) = +\infty$ and from (12) and (13), we have

$$G_{\varepsilon_j}(x^j,\rho_j) \geq f(x^j) \to +\infty,\ \rho_j > 0,\ j = 0,1,2,\ldots,$$

which results in a contradiction since the sequence $\{G_{\varepsilon_j}(x^j,\rho_j)\}$ is bounded. Thus $\{x^j\}$ is bounded.

We show next that any limit point x^* of $\{x^j\}$ is feasible to (P). Without loss of generality, we assume that $\lim\limits_{j\to+\infty} x^j = x^*$. Suppose that x^* is not feasible to (P). Then there exits some $i \in I$ such that $g_i(x^*) \geq \alpha > 0$. Note that

$$\begin{aligned} G_{\varepsilon_j}(x^j,\rho_j) = f(x^j) &+ \rho_j \sum_{i \in I^+_{\varepsilon_j}(x^j)} \left(g_i(x^j) + \frac{2}{3}\varepsilon_j e^{-\frac{g_i(x^j)}{\varepsilon_j}+1} - \frac{14\varepsilon_j}{9} \right) \\ &+ \rho_j \sum_{i \in I^-_{\varepsilon_j}(x^j)} \frac{g_i(x^j)^3}{9\varepsilon_j^2}. \end{aligned} \tag{14}$$

If $j \to +\infty$, then for any sufficiently large j, the set $\{i \mid g_i(x^j) \geq \alpha\}$ is not empty. Because I is finite, then there exists an $i_0 \in I$ that satisfies $g_{i_0}(x^j) \geq \alpha$. If $j \to +\infty, \rho_j \to +\infty, \varepsilon_j \to 0$, it follows from (14) that $G_{\varepsilon_j}(x^j,\rho_j) \to +\infty$, which contradicts the assumption that $\{G_{\varepsilon_j}(x^j,\rho_j)\}$ is bounded. Therefore, x^* is feasible to (P).

Finally, we show that (11) holds. By Step 2 in Algorithm I, $\nabla G_{\varepsilon_j}(x^j,\rho_j) = 0$, that is

$$\begin{aligned} \nabla f(x^j) &+ \rho_j \sum_{i \in I^+_{\varepsilon_j}(x^j)} \left(1 - \frac{2}{3} e^{-\frac{g_i(x^j)}{\varepsilon_j}+1} \right) \nabla g_i(x^j) \\ &+ \rho_j \sum_{i \in I^-_{\varepsilon_j}(x^j)} \frac{1}{3\varepsilon_j^2} g_i(x^j)^2 \nabla g_i(x^j) = 0. \end{aligned} \tag{15}$$

For $j = 1,2,\ldots,$ let

$$\gamma_j = 1 + \sum_{i \in I^+_{\varepsilon_j}(x^j)} \rho_j \left(1 - \frac{2}{3} e^{-\frac{g_i(x^j)}{\varepsilon_j}+1} \right) + \sum_{i \in I^-_{\varepsilon_j}(x^j)} \frac{\rho_j}{3\varepsilon_j^2} g_i(x^j)^2. \tag{16}$$

Then $\gamma_j > 1$. From (15), we have

$$\frac{1}{\gamma_j}\nabla f(x^j) + \sum_{i\in I^+_{\varepsilon_j}(x^j)} \frac{\rho_j\left(1-\frac{2}{3}e^{-\frac{g_i(x^j)}{\varepsilon_j}+1}\right)}{\gamma_j}\nabla g_i(x^j)$$
$$+\sum_{i\in I^-_{\varepsilon_j}(x^j)} \frac{\rho_j\varepsilon_j^{-2}}{3\gamma_j} g_i(x^j)^2\nabla g_i(x^j) = 0. \tag{17}$$

Let

$$\lambda^j = \frac{1}{\gamma_j},$$
$$\mu_i^j = \frac{\rho_j\left(1-\frac{2}{3}e^{-\frac{g_i(x^j)}{\varepsilon_j}+1}\right)}{\gamma_j}, \qquad i\in I^+_{\varepsilon_j}(x^j),$$
$$\mu_i^j = \frac{\rho_j\varepsilon_j^{-2}}{3\gamma_j} g_i(x^j)^2, \qquad i\in I^-_{\varepsilon_j}(x^j),$$
$$\mu_i^j = 0, \qquad i\in I\setminus\left(I^+_{\varepsilon_j}(x^j)\cup I^-_{\varepsilon_j}(x^j)\right).$$

Then we have

$$\lambda^j + \sum_{i\in I}\mu_i^j = 1,\ \ \forall j, \tag{18}$$
$$\mu_i^j \geq 0,\ \ i\in I,\ \ \forall j.$$

When $j\to\infty$, we have that $\lambda^j\to\lambda\geq 0,\ \mu_i^j\to\mu_i\geq 0,\ \forall i\in I$. By (17) and (18), as $j\to+\infty$, we have

$$\lambda\nabla f(x^*) + \sum_{i\in I}\mu_i\nabla g_i(x^*) = 0,$$
$$\lambda + \sum_{i\in I}\mu_i = 1.$$

For $i\in I^0(x^*)$, as $j\to+\infty$, we get $\mu_i^j\to 0$. Therefore, $\mu_i = 0,\ \forall i\in I^0(x^*)$. So, (11) holds, and this completes the proof. □

Now, we will solve some nonlinear constrained optimization problems with Algorithm I on MATLAB. In each of the following examples, the MATLAB 7.12 subroutine *fmincon* is used to obtain the local minima of problem $(SP_{\rho_j,\varepsilon_j})$. The numerical results of each example are presented in the following tables. It is shown that Algorithm I yield some approximate solutions that have a better objective function value in comparison with some other algorithms.

Note: j is the number of iteration in the Algorithm I.

ρ_j is constrain penalty parameter at the j'th iteration.

Table 1 Numerical results of Algorithm I with $x^0=(0,0,0,0),\ \rho_0=10,\ N=4$

j	ρ_j	ε_j	$f(x^j)$	$g_1(x^j)$	$g_2(x^j)$	$g_3(x^j)$	x^j
1	10	0.02	-44.271512	0.009467	0.0015424	-1.866763	(0.170216,0.836027, 2.010594,-0.966369)
2	40	0.0002	-44.234025	0.000047	0.000077	-1.883044	(0.169563,0.835533, 2.008644,-0.964884)
3	160	0.000002	-44.233662	-0.000000	-0.000072	-1.888579	(0.168232,0.834156, 2.010050,-0.963345)

x^j is a solution at the j'th iteration in the Algorithm I.
$f(x^j)$ is an objective value at x^j.
$g_i(x^j)\ (i=1,\ldots,m)$ is a constrain value at x^j.

Example 1 Consider the example in [8],

$$
\begin{aligned}
\text{(P3.1)}\quad \min\quad & f(x)=x_1^2+x_2^2+2x_3^2+x_4^2-5x_1-5x_2-21x_3+7x_4\\
\text{s.t.}\quad & g_1(x)=2x_1^2+x_2^2+x_3^2+2x_1+x_2+x_4-5\le 0,\\
& g_2(x)=x_1^2+x_2^2+x_3^2+x_4^2+x_1-x_2+x_3-x_4-8\le 0,\\
& g_3(x)=x_1^2+2x_2^2+x_3^2+2x_4^2-x_1-x_4-10\le 0.
\end{aligned}
$$

Let $x^0=(0,0,0,0),\ \rho_0=10,\ N=4,\ \varepsilon_0=0.02,\ \eta=0.01$ and $\varepsilon=10^{-6}$. Numerical results of Algorithm I for solving (P3.1) are given in Table 1.

From Table 1, it is said that an approximate ε-feasible solution to (P3.1) is obtained at the 3'th iteration, that is $x^3=(0.168232,0.834156,2.010050,-0.963345)$ with corresponding objective function value $f(x^3)=-44.233662$. It is easy to check that the x^3 is feasible solution to (P3.1). The solution we obtained is slightly better than the solution obtained in the 4'th iteration by method in [8] (the objective function value $f(x^*)=-44.23040$) for this example. Further, with the same parameters $\rho_0,\ N,\ \varepsilon_0,\ \eta$ as above, we change the starting point to $x^0=(1,1,1,1)$ or $x^0=(6,6,6,6)$. New numerical results by Algorithm I are given in Table 2 and Table 3.

It is easy to see from Tables 2 and 3 that the convergence of Algorithm I is the same and the objective function values are almost the same. That is to say, the efficiency of Algorithm I does not completely depend on the starting point x^0. Then, we can choose any starting point for Algorithm I.

Example 2 Consider the example in [18],

$$
\begin{aligned}
\text{(P3.2)}\quad \min\quad & f(x)=-2x_1-6x_2+x_1^2-2x_1x_2+2x_2^2\\
\text{s.t.}\quad & x_1+x_2\le 2,\\
& -x_1+2x_2\le 2,\\
& x_1,\ x_2\ge 0.
\end{aligned}
$$

Table 2 Numerical results of Algorithm I with $x^0 = (1,1,1,1),\ \rho_0 = 10,\ N = 4$

j	ρ_j	ε_j	$f(x^j)$	$g_1(x^j)$	$g_2(x^j)$	$g_3(x^j)$	x^j
1	10	0.02	-44.271512	0.009467	0.0015424	-1.866763	(0.170216,0.836027, 2.010594,-0.966369)
2	40	0.0002	-44.234025	0.000047	0.000077	-1.883044	(0.169563,0.835533, 2.008644,-0.964884)
3	160	0.000002	-44.233355	-0.000113	-0.000079	-1.900244	(0.166329,0.831255, 2.012529,-0.960615)

Table 3 Numerical results of Algorithm I with $x^0 = (6,6,6,6),\ \rho_0 = 10,\ N = 4$

j	ρ_j	ε_j	$f(x^j)$	$g_1(x^j)$	$g_2(x^j)$	$g_3(x^j)$	x^j
1	10	0.02	-44.271512	0.009467	0.0015424	-1.866763	(0.170216,0.836027, 2.010594,-0.966369)
2	40	0.0002	-44.234025	0.000047	0.000077	-1.883044	(0.169563,0.835533, 2.008644,-0.964884)
3	160	0.000002	-44.232449	-0.000613	-0.000188	-1.856917	(0.159767,0.840231, 2.011450,-0.963346)

Let

$$g_1(x) = x_1 + x_2 - 2,\ \ g_2(x) = -x_1 + 2x_2 - 2,$$
$$g_3(x) = -x_1,\ \ g_4(x) = -x_2.$$

Thus problem (P3.2) is equivalent to the following problem:

$$\begin{aligned}
\text{(P3.2')}\quad \min\ & f(x) = -2x_1 - 6x_2 + x_1^2 - 2x_1x_2 + 2x_2^2 \\
\text{s.t. } & g_1(x) = x_1 + x_2 - 2 \le 0, \\
& g_2(x) = -x_1 + 2x_2 - 2 \le 0, \\
& g_3(x) = -x_1 \le 0, \\
& g_4(x) = -x_2 \le 0.
\end{aligned}$$

Let $x^0 = (0,0),\ \rho_0 = 8,\ N = 10,\ \varepsilon_0 = 0.01,\ \eta = 0.01$ and $\varepsilon = 10^{-6}$. Numerical results of Algorithm I for solving (P3.2') are given in Table 4.

By Table 4, an approximate optimal solution to (P3.2') is obtained at the 3'th iteration, that is $x^* = (0.800000, 1.200000)$ with corresponding objective function value $f(x^*) = -7.200000$. The solution we obtained is similar with the solution obtained in the 4'th iteration by method in [18] (the objective function value $f(x^*) = -7.2000$) for this example.

4 Conclusions

This paper has presented a second-order smoothing approximation to the l_1 exact penalty function and an algorithm based on this smoothed penalty problem. It is shown that the optimal solution to the $(SP_{\rho,\varepsilon})$ is an approximate optimal solution

Table 4 Numerical results of Algorithm I with $x^0 = (0,0)$, $\rho_0 = 8$, $N = 10$

j	ρ_j	ε_j	$f(x^j)$	$g_1(x^j)$	$g_2(x^j)$	x^j
1	8	0.01	-7.228666	0.010245	-0.397951	(0.806147,1.204098)
2	80	0.0001	-7.200091	0.000032	-0.399994	(0.800019,1.200013)
3	800	0.000001	-7.200000	0.000000	-0.400000	(0.800000,1.200000)

to the original optimization problem under some mild conditions. Numerical results show that the Algorithm I has a better convergence for a global approximate solution.

References

1. M.S. Bazaraa and J.J. Goode, Sufficient conditions for a globally exact penalty function without convexity, *Mathematical Programming Study,* **19**, (1982), 1-15.
2. C.H. Chen and O.L. Mangasarian, Smoothing methods for convex inequalities and linear complementarity problems, *Math. Program.,* **71**, (1995), 51-69.
3. G. Di Pillo and L. Grippo, An exact penalty function method with global conergence properties for nonlinear programming problems, *Math. Program.,* **36**, (1986), 1-18.
4. G. Di Pillo and L. Grippo, Exact penalty functions in contrained optimization, *SIAM J. Control. Optim.,* **27**, (1989), 1333-1360.
5. S.P. Han and O.L. Mangasrian, Exact penalty function in nonlinear programming, *Math. Program.,* **17**, (1979), 257-269.
6. X.X. Huang and X.Q. Yang, Convergence analysis of a class of nonlinear penalization methods for constrained optimization via first-order necessary optimality conditions, *J. Optim. Theory Appl.,* **116**, (2003), 311-332.
7. J.B. Lasserre, A globally convergent algorithm for exact penalty functions, *Eur. J. Oper. Res.,* **7**, (1981), 389-395.
8. S.J. Lian, Smoothing approximation to l_1 exact penalty function for inequality constrained optimization, *Appl. Math. Comput.,* **219**, (2012), 3113-3121.
9. B.Z. Liu, On smoothing exact penalty functions for nonlinear constrained optimization problems, *J. Appl. Math. Comput.,* **30**, (2009), 259-270.
10. Z.Q. Meng, C.Y. Dang and X.Q. Yang, On the smoothing of the square-root exact penalty function for inequality constrained optimization, *Comput. Optim. Appl.,* **35**, (2006), 375-398.
11. K.W. Meng, S.J. Li and X.Q. Yang, A robust SQP method based on a smoothing lower order penalty function, *Optimization,* **58**, (2009), 22-38.
12. M.C. Pinar and S.A. Zenios, On smoothing exact penalty function for convex constrained optimization, *SIAM J. Optim.,* **4**, (1994), 486-511.
13. E. Rosenberg, Exact penalty functions and stability in locally Lipschitz programming, *Math. Program.,* **30**, (1984), 340-356.
14. A.M. Rubinov, B.M. Glover and X.Q. Yang, Extended Lagrange and penalty function in continuous optimization, *Optimization,* **46**, (1999), 327-351.
15. A.M. Rubinov, X.Q. Yang and A.M. Bagirov, Penalty functions with a small penalty parameter, *Optim. Methods Softw.,* **17**, (2002), 931-964.
16. A.M. Rubinov and X.Q. Yang, Nonlinear Lagrangian and Penalty Functions in Optimization, *Kluwer Academic, Dordrecht,* (2003).
17. 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,* **53**, (2004), 57-68.
18. 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, *J. Ind. Manag. Optim.,* **1**, (2005), 533-547.
19. X.S. Xu, Z.Q. Meng, J.W. Sun and R. Shen, A penalty function method based on smoothing lower order penalty function, *J. Comput. Appl. Math.,* **235**, (2011), 4047-4058.
20. X.S. Xu, Z.Q. Meng, J.W. Sun, L.Q. Huang and R. Shen, A second-order smooth penalty function algorithm for constrained problems, *Comput. Optim. Appl.,* **55**, (2013), 155-172.

21. X.Q. Yang, Smoothing approximations to nonsmooth optimization problems, *J. Aust. Math. Soc. B,* **36**, (1994), 274-285.
22. X.Q. Yang and X.X. Huang, A nonlinear Lagrangian approach to constrained optimization problems, *SIAM J. Optim.,* **11**, (2001), 1119-1144.
23. X.Q. Yang, Z.Q. Meng, X.X. Huang and G.T.Y. Pong, Smoothing nonlinear penalty function for constrained optimization, *Numer. Funct. Anal. Optim.,* **24**, (2003), 357-364.
24. W.I. Zangwill, Nonlinear programming via penalty function, *Mgmt. Sci.,* **13**, (1967), 334-358.
25. S.A. Zenios, M.C. Pinar and R.S. Dembo, A smooth penalty function algorithm for network-structured problems, *Eur. J. Oper. Res.,* **64**, (1993), 258-277.