RESEARCH ARTICLE
doi: 10.2306/scienceasia1513-1874.2018.44.268

ScienceAsia 44 (2018): **268-276**

A new exact penalty function method for nonlinear programming problems

Cheng MA

School of Business, Qingdao University, Qingdao, China

e-mail: mc_0812@163.com

Received 17 July 2017
Accepted 28 May 2018

ABSTRACT: In this paper, we present a new exact and smooth penalty function for nonlinear programming problems by adding only one variable no matter how many constraints. Through the smooth and exact penalty function, we can transform the nonlinear programming problems into unconstrained optimization models. We demonstrate that under some general conditions, when the penalty parameter $\sigma > 0$ is sufficiently large, the minimizer of this penalty function is the minimizer of the primal problem, which can be obtained after finite iterations. Alternatively, under some mild assumptions, sufficient conditions are derived for the local exactness property. The numerical results demonstrate that the new penalty function is reasonable and is an effective approach for solving a class of nonlinear programming problem with equality and inequality constraints.

MSC2010: 90C15

INTRODUCTION

In this paper, we consider the following constrained minimization problem:

$$\begin{cases} \min & f(x) \\ \text{s.t.} & F_j(x) = 0 \quad \forall j \in E, \\ & g_\ell(x) \leqslant 0 \quad \forall \ell \in I, \end{cases} \tag{P}$$

where $f : \mathbb{R}^n \to \mathbb{R}, F_j : \mathbb{R}^n \to \mathbb{R}$ and $g_\ell : \mathbb{R}^n \to \mathbb{R}$ ($j \in E, \ell \in I$) are continuously differentiable functions, E, I denote the index for equality and inequality constrained functions respectively.

There are various approaches for solving the constrained minimization such as SQP-trust region method[1], filter method[2], interior point method[3], penalty function method[4–6] etc. The interested reader can refer to reference and references therein.

The SQP-trust region method and the filter method have to solve a series infinite quadratic programming and there exists Maratos effect. The traditional penalty function method is a popular method. For example, we present the following penalty functions:

$$f_\sigma(x) = f(x) + \sigma\Big[\sum_{j\in E}|F_j(x)| + \sum_{\ell\in I}\max(0, g_\ell(x))\Big], \tag{1}$$

$$f_\sigma(x) = f(x) + \sigma\Big[\sum_{j\in E}F_j^2(x) + \sum_{\ell\in I}(\max(0, g_\ell(x)))^2\Big], \tag{2}$$

$$f_\sigma(x) = f(x) + \lambda(x)^T F(x) + \frac{\sigma}{2}F(x)^T F(x), \tag{3}$$

where $\sigma > 0$ is a penalty parameter, $F(x) = [F_j(x) : j \in E]$, $\lambda(x) = (\nabla F(x))^+ \nabla f(x)$, and $(\nabla F(x))^+$ denotes the generalized inverse matrix of $\nabla F(x)$. However, there are also some disadvantages. On the one hand, the penalty function (1) is usually continuous only in a neighborhood of the optimal solution, and may fail to be continuous in the whole region. On the other hand, the penalty function is exact and smooth, then it is not simple, and if the penalty function is simple and smooth, then it is not exact. For the above three penalty functions, it is well-known that (1) is a nonsmooth simple exact penalty function; (2) is a smooth simple penalty function, but it is not exact; (3) is a smooth exact penalty function, but it is not simple. Here, the word simple means that the penalty function only includes the functions of the primal problem rather than involves the derivative information of the primal problem.

Recently, a new exact penalty function[7,8] is given for the equality constrained minimization problem (Q), where a new approach is proposed by adding one variable to equality constrained minimization problem (Q) as follows:

$$\min_{x\in S} f(x), \tag{Q}$$

where

$$S=\{x\in[u,v]:F_j(x)=0\,\forall\, j\in E\},$$

$[u,v]$ is a box in $\mathbb{R}^n$ with nonempty interior given by $[u,v]=\{x\in R^n:u\leqslant x\leqslant v\}$ and $(\{-\infty\}\cup\mathbb{R})^n\ni u<v\in(\{+\infty\}\cup\mathbb{R})^n, f:D\to\mathbb{R}$ and $F_j:D\to\mathbb{R}$ $(j\in E)$ are continuously differentiable in an open set D containing $[u,v]$. Then fix $w_j\in\mathbb{R}$ for each $j\in E$ and consider the following equivalent problem:

$$\min_{(x,\varepsilon)\in S_0} f(x), \tag{Q}$$

where $S_{\bar\varepsilon}=\{(x,\varepsilon)\in[u,v]\times[0,\bar\varepsilon]:F_j(x)=\varepsilon w_j\ \forall\, j\in E\}$.

Let[7]

$$f_\sigma(x,\varepsilon)=\begin{cases} f(x), & \text{Case 1},\\ f(x)+\frac{1}{2\varepsilon}\frac{\Delta(x,\varepsilon)}{1-q\Delta(x,\varepsilon)}+\sigma\beta(\varepsilon), & \text{Case 2},\\ +\infty, & \text{otherwise.}\end{cases} \tag{4}$$

where, Case 1: $\varepsilon=0, x\in S$ and Case 2: $0<\varepsilon\leqslant\bar\varepsilon, \Delta(x,\varepsilon)<q^{-1}$, in addition, $\bar\varepsilon>0$ and $q>0$ are fixed and $\beta:[0,\bar\varepsilon]\to[0,+\infty)$ is continuous and continuously differentiable on $(0,\bar\varepsilon]$ with $\beta(0)=0$.

The corresponding penalty problem (Q_σ) is

$$\min_{(x,\varepsilon)\in[u,v]\times[0,\bar\varepsilon]} f_\sigma(x,\varepsilon), \tag{Q_σ}$$

where the constrained violation measure is

$$\Delta(x,\varepsilon)=\sum_{j\in E}(F_j(x)-\varepsilon w_j)^2.$$

$f_\sigma(x,\varepsilon)$ is an exact penalty function of the primal problem (Q), it is a continuously differentiable function on $\{(x,\varepsilon)\in[u,v]\times[0,\bar\varepsilon]:\varepsilon=0,x\in S\}$ or $\{(x,\varepsilon)\in[u,v]\times(0,\bar\varepsilon]:0<\varepsilon\leqslant\bar\varepsilon,\Delta(x,\varepsilon)<q^{-1}\}$ respectively, but it is not a continuously differentiable function on $\{(x,\varepsilon)\in[u,v]\times[0,\bar\varepsilon]:\varepsilon=0,x\in S$ or $0<\varepsilon\leqslant\bar\varepsilon,\Delta(x,\varepsilon)<q^{-1}\}$[7]. Therefore, it is worth noting that the penalty function (4) is not continuously differentiable in the whole region under the mathematical analysis sense. The main failure lies in the pair $(x^*,0)$ case, where $x^*\in S$, more concretely, for the case that $(x,0)\to(x^*,0)$ where $x\notin S$, the limits of $f_\sigma(x,\varepsilon)$ and $\nabla_{(x,\varepsilon)}f_\sigma(x,\varepsilon)$ as $(x,\varepsilon)\to(x^*,0)$, may not exist. However, the purpose of introducing the new variable ε is just to make some conveniences in theoretical analysis and practical computing in order to achieve the optimal solution. It is meaningless to set $\varepsilon=0$ at the beginning of the algorithm. Therefore, from the practical algorithmic point of view, this failure is trivial.

Motivated by this, in this paper, we propose a new exact penalty function for the nonlinear programming problems with equality and inequality constraints. The main feature of our penalty function is that we only need to add a variable ε for constraints. The merit function is considered as a function of x and ε simultaneously which has good smoothness, exactness properties, even without involving gradient and Jacobian matrices. It remains bounded below whenever $f(x)$ is bounded below, which is not shared by l_1 and quadratic penalty functions. With the appropriate assumptions, for sufficiently large σ, we verify that the minimizer (x^*,ε^*) of the penalty problem satisfies $\varepsilon^*=0$ if and only if x^* solves the original problem (P). This property demonstrates we can obtain the fact that x^* is optimal solution of original problem as long as $\varepsilon^*=0$ for the pair (x^*,ε^*). Furthermore, we present the result that, if a local optimal solution of the penalty problem satisfies the extended Mangasarian-Fromovitz constraint qualification, then the minimizer has the expression of $(x^*,0)$. As well known, the ill-conditioning introduced by a large penalty parameter may be detrimental. Therefore, for the new exact penalty function, we only require the penalty parameter to be increased by adding a relatively small constant in order to keep the penalty parameter as small as possible to avoid ill-conditioning, which is illustrated in numerical test section. For the penalty function algorithm, the global convergence property can be obtained.

AN EXACT AND SMOOTH PENALTY FUNCTION FOR EQUALITY AND INEQUALITY CONSTRAINED MINIMIZATION PROBLEM

We reformulate the feasible region as a set S as follows:

$$S=\{x:F_j(x)=0,\forall\, j\in E, g_\ell(x)\leqslant 0,\forall \ell\in I\}. \tag{5}$$

We introduce a new variable ε into the constraint function and define $S_\varepsilon=\{(x,\varepsilon):F_j(x)=\varepsilon^\gamma w_j,\forall\, j\in E, g_\ell(x)\leqslant\varepsilon^\gamma w_\ell,\forall\ell\in I\}$, where $w_j,w_\ell\in(0,1)$ for each $j\in E,\ell\in I$, and γ is a positive number. In particular, when $\varepsilon=0$, $S_\varepsilon=S$. We make some assumptions for (P):

(1) There exists a global minimizer for (P), this implies that $f(x)$ is bounded below on S;
(2) If $x^*\in L(\mathrm{P})$, then $L_{x^*}=\{x\in L(\mathrm{P}):f(x)=f(x^*)\}$ is a compact set, where $L(\mathrm{P})$ is the set of local minimization of (P).

The penalty function $f_\sigma(x,\varepsilon)$ and penalty problem (P_σ) can be formulated as follows

$$\min_{(x,\varepsilon)\in\mathbb{R}^n\times(-\bar{\varepsilon},\bar{\varepsilon})} f_\sigma(x,\varepsilon), \qquad (P_\sigma)$$

where

$$f_\sigma(x,\varepsilon)= \begin{cases} f(x) & \text{Case 1,}\\ f(x)-\varepsilon^\alpha\ln(1-\frac{\Delta(x,\varepsilon)}{\varepsilon^{2\delta}})+\sigma\varepsilon^\beta & \text{Case 2,}\\ +\infty & \text{otherwise} \end{cases} \qquad (6)$$

where Case 1: $\varepsilon=0, x\in S$ and Case 2: $\varepsilon\neq 0, 0<1-2\varepsilon^{-2\delta}\Delta(x,\varepsilon)<1$, $\alpha,\beta,\delta,\gamma$ are positive even numbers and $\beta>1$, in particular, $\gamma>\delta$ throughout this paper, $\sigma>0$ is a penalty parameter. Denote the summation of constraint violation as follows

$$\begin{aligned}\Delta(x,\varepsilon) &= \sum_{j\in E}(F_j(x)-\varepsilon^\gamma w_j)^2+\sum_{\ell\in I}(\max(0,g_\ell(x)-\varepsilon^\gamma w_\ell))^2\\ &= \sum_{j\in E}(F_j(x)-\varepsilon^\gamma w_j)^2+\sum_{\ell\in I^+(x,\varepsilon)}(g_\ell(x)-\varepsilon^\gamma w_\ell)^2.\end{aligned}$$

where $I^+(x,\varepsilon)=\{\ell\in I|g_\ell(x)\geqslant\varepsilon^\gamma w_\ell\}$.

For $\varepsilon>0$, $0<1-\varepsilon^{-2\delta}\Delta(x,\varepsilon)<1$, we have

$$\begin{aligned} f_\sigma(x,\varepsilon) &= f(x)-\varepsilon^\alpha\ln(1-\varepsilon^{-2\delta}\Delta(x,\varepsilon))+\sigma\varepsilon^\beta\geqslant f(x).\end{aligned}$$

provided $F_j(x)=0$ $(j\in E)$ and $g_\ell(x)\leqslant 0$ $(\ell\in I)$. Therefore, $f_\sigma(x,\varepsilon)$ is bounded below on $\mathbb{R}^n\times[-\bar{\varepsilon},\bar{\varepsilon}]$ whenever $f(x)$ is bounded below on the set D' consisting of $x\in\mathbb{R}^n$ satisfying

$$\|F(x)\|\leqslant\frac{\sqrt{2}}{2}\bar{\varepsilon}^\delta+\bar{\varepsilon}^\gamma\|w\|,$$
$$\|g(x)\|\leqslant\frac{\sqrt{2}}{2}\bar{\varepsilon}^\delta+\bar{\varepsilon}^\gamma\|w\|.$$

This is a reasonable condition since when f is bounded below on the feasible set, $\bar{\varepsilon}$ is small enough. To illustrate the theory developed, we consider the following simple nonlinear optimization problem:

$$\begin{aligned} \min\quad & x_1^3x_2^3\\ \text{s.t.}\quad & x_1^2+x_2^2=4;\\ & x_1\leqslant 2, x_2\leqslant 2.\end{aligned}$$

There are two global minimizers $x_1^*=-x_2^*=(\sqrt{2},-\sqrt{2})$ with $f(x_1^*)=f(x_2^*)=-8$. If we use the traditional penalty function, we have the following conclusions:

- l_1 penalty function:

  $$f_\sigma(x)=x_1^3x_2^3+\sigma(|\|x\|^2-4|+(x_1-2)^++(x_2-2)^+)$$

  is unbounded below, where $x^+=\max\{0,x\}$. Because when $x=(-m,m)^T$, $f_\sigma(x)\to-\infty$ as $m\to+\infty$.
- Quadratic penalty function:

  $$f_\sigma(x)=x_1^3x_2^3+\sigma((\|x\|^2-4)^2+((x_1-2)^+)^2+((x_2-2)^+)^2)$$

  is unbounded below, because $f_\sigma(x)\to-\infty$ for $x=(-m,m)^T$, as $m\to+\infty$.

For this new penalty function, choosing $w_1=w_2=w_3=0.5$, we have

$$f_\sigma(x,\varepsilon) \begin{cases} x_1^3x_2^3 & \text{Case 1,}\\ x_1^3x_2^3-\varepsilon^\alpha\ln(1-\frac{\Delta(x,\varepsilon)}{\varepsilon^{2\delta}})+\sigma\varepsilon^\beta & \text{Case 2,}\\ +\infty & \text{otherwise,}\end{cases}$$

where Case 1: $\varepsilon=\Delta(x,\varepsilon)=0$ and Case 2: $\varepsilon\neq 0$, $0<1-2\varepsilon^{-2\delta}\Delta(x,\varepsilon)<1$ and

$$\begin{aligned}\Delta(x,\varepsilon)=(x_1^2+x_2^2-4-0.5\varepsilon^\gamma)^2 &+(\max(0,x_1-2-0.5\varepsilon^\gamma))^2\\ &+(\max(0,x_2-2-0.5\varepsilon^\gamma))^2.\end{aligned}$$

Since $f_\sigma(x,\varepsilon)=+\infty$ if $\|x\|\geqslant\sqrt{4+0.5\varepsilon^\gamma+\frac{\sqrt{2}}{2}\varepsilon^\delta}$ or $|x_1|\geqslant 2+\frac{\sqrt{2}}{2}\varepsilon^\delta+0.5\varepsilon^\gamma$ or $|x_2|\geqslant 2+\frac{\sqrt{2}}{2}\varepsilon^\delta+0.5\varepsilon^\gamma$, the bounded below of this new penalty function below can be verified easily.

In what follows, we shall show that, under some mild conditions, $f_\sigma(x,\varepsilon)$ is continuously differentiable with continuous limits on the part of the boundary with finite values.

Proposition 1 *Let $x\to x^*\in S, 0\neq\varepsilon\to\varepsilon^*=0$. Suppose that*

$$\begin{cases} 2\delta-\alpha>0,\\ \alpha-\delta-1>0,\\ \beta>1,\end{cases} \qquad (7)$$

then

$$\lim_{\substack{\varepsilon\to\varepsilon^*=0\\ x\to x^*\in S}} f_\sigma(x,\varepsilon)=f_\sigma(x^*,0)=f(x^*),$$

$$\lim_{\substack{\varepsilon\to\varepsilon^*=0\\ x\to x^*\in S}} \nabla_x f_\sigma(x,\varepsilon)=\nabla f(x^*),$$

$$\lim_{\substack{\varepsilon\to\varepsilon^*=0\\ x\to x^*\in S}} \frac{\partial f_\sigma(x,\varepsilon)}{\partial\varepsilon}=0.$$

Proof: From the fact that $\varepsilon \neq 0, 0 < 1 - 2\frac{\Delta(x,\varepsilon)}{\varepsilon^{2\delta}} < 1$, we have $\Delta(x,\varepsilon) = O(\varepsilon^{2\delta})$ and

$$\lim_{\substack{\varepsilon\to\varepsilon^*=0\\ x\to x^*\in S}} (1-\varepsilon^{-2\delta}\Delta(x,\varepsilon)) = c^* \in [\frac{1}{2}, 1],$$

when $\varepsilon \to \varepsilon^* = 0$, $x \to x^* \in S$. Thus, $\sum_{j\in E}(F_j(x) - \varepsilon^\gamma w_j) = O(\varepsilon^\delta)$ and $\sum_{\ell\in I^+(x,\varepsilon)}(g_\ell(x) - \varepsilon^\gamma w_\ell) = O(\varepsilon^\delta)$. From (7), we know $2\delta > \alpha$ and $\beta > 1$. This yields

$$\begin{aligned}
&\lim_{\substack{\varepsilon\to\varepsilon^*=0\\ x\to x^*\in S}} f_\sigma(x,\varepsilon)\\
&= \lim_{\substack{\varepsilon\to\varepsilon^*=0\\ x\to x^*\in S}} f(x) - \varepsilon^\alpha \ln\Big(1-\varepsilon^{-2\delta}\Delta(x,\varepsilon)\Big) + \sigma\varepsilon^\beta\\
&= f(x^*)
\end{aligned}$$

Notice that $f_\sigma(x,\varepsilon)$ is continuously differentiable in the set D. The gradient of $f_\sigma(x,\varepsilon)$ at (x,ε) is

$$\nabla_{(x,\varepsilon)} f_\sigma(x,\varepsilon) = \Big(\nabla_x f_\sigma(x,\varepsilon), \frac{\partial f_\sigma(x,\varepsilon)}{\partial \varepsilon}\Big)^T,$$

where

$$\nabla_x f_\sigma(x,\varepsilon) = \nabla f(x) + \varepsilon^{\alpha-2\delta}\frac{\partial_x \Delta(x,\varepsilon)}{1-\varepsilon^{-2\delta}\Delta(x,\varepsilon)} \quad (8)$$

and

$$\begin{aligned}
\frac{\partial f_\sigma(x,\varepsilon)}{\partial\varepsilon} &= -\alpha\varepsilon^{\alpha-1}\ln(1-\varepsilon^{-2\delta}\Delta(x,\varepsilon))\\
&\quad - \varepsilon^\alpha \frac{2\delta\varepsilon^{-2\delta-1}\Delta(x,\varepsilon) + \varepsilon^{-2\delta}\frac{\partial\Delta(x,\varepsilon)}{\partial\varepsilon}}{1-\varepsilon^{-2\delta}\Delta(x,\varepsilon)}\\
&\quad + \sigma\beta\varepsilon^{\beta-1}\\
&= -\alpha\varepsilon^{\alpha-1}\ln(1-\varepsilon^{-2\delta}\Delta(x,\varepsilon))\\
&\quad + \frac{\varepsilon^{\alpha-2\delta-1}}{1-\varepsilon^{-2\delta}\Delta(x,\varepsilon)}(-2\delta\Delta(x,\varepsilon)\\
&\quad - \varepsilon\frac{\partial\Delta(x,\varepsilon)}{\partial\varepsilon}) + \sigma\beta\varepsilon^{\beta-1}\\
&= -\alpha\varepsilon^{\alpha-1}\ln(1-\varepsilon^{-2\delta}\Delta(x,\varepsilon))\\
&\quad + \frac{\varepsilon^{\alpha-2\delta-1}}{1-\varepsilon^{-2\delta}\Delta(x,\varepsilon)} \times \Big(2(\gamma-\delta)\Delta(x,\varepsilon)\\
&\quad - 2\gamma(\sum_{j\in E}(F_j(x) - \varepsilon_k^\gamma w_j)F_j(x)\\
&\quad + \sum_{\ell\in I^+(x,\varepsilon)}(g_\ell(x) - \varepsilon^\gamma w_\ell)g_\ell(x)\Big)\\
&\quad + \sigma\beta\varepsilon^{\beta-1}. \qquad (9)
\end{aligned}$$

Combing (7), (8) and (9), we have

$$\lim_{\substack{\varepsilon\to\varepsilon^*=0\\ x\to x^*\in S}} \nabla_x f_\sigma(x,\varepsilon) = \nabla f(x^*)$$

and $\lim_{\substack{\varepsilon\to\varepsilon^*=0\\ x\to x^*\in S}} \frac{\partial f_\sigma(x,\varepsilon)}{\partial\varepsilon} = 0$. This yields the desired conclusion. □

ALGORITHM

Algorithm 1

Step 1 Choose $\tilde{\varepsilon}, \bar{\varepsilon} > 0$, $\eta > 0$ arbitrarily small, $\sigma_0 > 0$, $\rho > 0$ and $(x_0, \varepsilon_0) \in \mathbb{R}^n \times (-\bar{\varepsilon}, \bar{\varepsilon})$, $\varepsilon_0 \neq 0$, set $k := 0$.

Step 2 For the nonlinear programming problem (P) we construct the following penalty function

$$f_\sigma(x,\varepsilon) = \begin{cases} f(x) & \text{Case 1,}\\ f(x) - \varepsilon^\alpha \ln(1-\frac{\Delta(x,\varepsilon)}{\varepsilon^{2\delta}}) + \sigma\varepsilon^\beta & \text{Case 2,}\\ +\infty & \text{otherwise,}\end{cases}$$

where Case 1: $\varepsilon = \Delta(x,\varepsilon) = 0$ and Case 2: $\varepsilon \neq 0, 0 < 1 - 2\frac{\Delta(x,\varepsilon)}{\varepsilon^{2\delta}} < 1$, $\alpha, \beta, \delta, \gamma$ are positive even numbers, and $\beta \geqslant 2$,

$$\begin{aligned}
&\Delta(x,\varepsilon)\\
&= \sum_{j\in E}(F_j(x) - \varepsilon^\gamma w_j)^2\\
&\qquad + \sum_{\ell\in I}(\max(0, g_\ell(x) - \varepsilon^\gamma w_\ell))^2\\
&= \sum_{j\in E}(F_j(x) - \varepsilon^\gamma w_j)^2 + \sum_{\ell\in I^+(x,\varepsilon)}(g_\ell(x) - \varepsilon^\gamma w_\ell)^2
\end{aligned}$$

and $I^+(x,\varepsilon) = \{\ell \in I | g_\ell(x) \geqslant \varepsilon^\gamma w_\ell\}$. Use any unconstrained algorithm to solve

$$\min_{(x,\varepsilon)\in\mathbb{R}^n\times(-\bar{\varepsilon},\bar{\varepsilon})} f_\sigma(x,\varepsilon)$$

and denote the solution (x_k, ε_k) of (P_σ).

Step 3 If $|\varepsilon_k| \leqslant \tilde{\varepsilon}, \|\nabla_{(x,\varepsilon)} f_{\sigma_k}(x_k,\varepsilon_k)\| \leqslant \eta$, then stop. The point obtained x_k is an approximation solution of (P). Otherwise, choose $\sigma_{k+1} = \sigma_k + \rho$.

Step 4 Set $k := k+1$ and return to **Step 2**.

Lemma 1 *If* $(x_k, \varepsilon_k) \in L(P_{\sigma_k})$ *with finite* $f_{\sigma_k}(x_k, \varepsilon_k)$, $\varepsilon_k \neq 0$, *then* $(x_k, \varepsilon_k) \notin S_\varepsilon = \{(x,\varepsilon) \in \mathbb{R}^{n+1} : F_j(x) = \varepsilon^\gamma w_j, \forall j \in E, g_\ell(x) \leqslant \varepsilon^\gamma w_\ell, \forall \ell \in I\}$.

Proof: By $(x_k, \varepsilon_k) \in L(P_{\sigma_k})$ with finite $f_{\sigma_k}(x_k, \varepsilon_k)$, $\varepsilon_k \neq 0$, then $\frac{\partial f_{\sigma_k}}{\partial\varepsilon}(x_k,\varepsilon_k) = 0$, we have

$$\frac{\partial f_{\sigma_k}(x_k,\varepsilon_k)}{\partial\varepsilon}$$

$$= -\alpha\varepsilon_k^{\alpha-1}\ln\Big(1-\frac{\Delta(x_k,\varepsilon_k)}{\varepsilon_k^{2\delta}}\Big)+\frac{\varepsilon_k^{\alpha-2\delta-1}}{1-\varepsilon_k^{-2\delta}\Delta(x_k,\varepsilon_k)} \times\Big(2(\gamma-\delta)\Delta(x_k,\varepsilon_k) -2\gamma\Big(\sum_{j\in E}(F_j(x_k)-\varepsilon_k^{\gamma}w_j)F_j(x_k) +\sum_{\ell\in I^+(x_k,\varepsilon_k)}(g_\ell(x_k)-\varepsilon_k^{\gamma}w_\ell)g_\ell(x_k)\Big)\Big)+\beta\varepsilon_k^{\beta-1}\sigma_k = 0$$

If $(x_k,\varepsilon_k)\in S_{\varepsilon_k}$, then the left hand side of the above is equal to $\beta\sigma_k\varepsilon_k^{\beta-1}\neq 0$. This is a contradiction. Thus, $(x_k,\varepsilon_k)\notin S_{\varepsilon_k}$. □

Remark 1 In fact, the following result is true: if $\nabla_{(x,\varepsilon)}f_{\sigma_k}(x_k,\varepsilon_k)=0$ with finite $f_{\sigma_k}(x_k,\varepsilon_k)$ and $\varepsilon_k\neq 0$, then $(x_k,\varepsilon_k)\notin S_{\varepsilon_k}$.

Lemma 2 *If $(x_k,\varepsilon_k)\in L(P_{\sigma_k})$ with finite $f_{\sigma_k}(x_k,\varepsilon_k)$, $\varepsilon_k\neq 0$, $(x_k,\varepsilon_k)\to(x^*,\varepsilon^*)$, $\nabla F_j(x^*)$ for all $j\in E$, $\nabla g_\ell(x^*)$ for all $\ell\in I^+(x^*,\ \varepsilon^*)$ are linearly independent, and $2\delta-\alpha>0$, then $\varepsilon^*=0, x^*\in S$.*

Proof: We first show that $\varepsilon^*=0$. From $(x_k,\varepsilon_k)\in L(\mathrm{P}_{\sigma_k})$, one has

$$\nabla_x f_{\sigma_k}(x_k,\varepsilon_k)=0 \tag{10}$$

and

$$\frac{\partial f_{\sigma_k}(x_k,\varepsilon_k)}{\partial\varepsilon} = -\alpha\varepsilon_k^{\alpha-1}\ln\Big(1-\frac{\Delta(x_k,\varepsilon_k)}{\varepsilon_k^{2\delta}}\Big)+\frac{\varepsilon_k^{\alpha-2\delta-1}}{1-\varepsilon_k^{-2\delta}\Delta(x_k,\varepsilon_k)} \times\Big(2(\gamma-\delta)\Delta(x_k,\varepsilon_k) -2\gamma\Big(\sum_{j\in E}(F_j(x_k)-\varepsilon_k^{\gamma}w_j)F_j(x_k) +\sum_{\ell\in I^+(x_k,\varepsilon_k)}(g_\ell(x_k)-\varepsilon_k^{\gamma}w_\ell)g_\ell(x_k)\Big)\Big)+\beta\varepsilon_k^{\beta-1}\sigma_k = 0. \tag{11}$$

Rearranging (11), we have

$$-\alpha\varepsilon_k^{2\delta}\ln\Big(1-\frac{\Delta(x_k,\varepsilon_k)}{\varepsilon_k^{2\delta}}\Big)\Big(1-\frac{\Delta(x_k,\varepsilon_k)}{\varepsilon_k^{-2\delta}}\Big) +\Big(2(\gamma-\delta)\Delta(x_k,\varepsilon_k) -2\gamma\Big(\sum_{j\in E}(F_j(x_k)-\varepsilon_k^{\gamma}w_j)F_j(x_k) +\sum_{\ell\in I^+(x_k,\varepsilon_k)}(g_\ell(x_k)-\varepsilon_k^{\gamma}w_\ell)g_\ell(x_k)\Big)\Big) +\beta\varepsilon_k^{\beta-\alpha+2\delta}\sigma_k\Big(1-\frac{\Delta(x_k,\varepsilon_k)}{\varepsilon_k^{2\delta}}\Big) = 0. \tag{12}$$

Taking $\sigma_k\to+\infty$, the first term and the second term of (12) tend to finite. From the construction of the penalty function $f_\sigma(x,\varepsilon)$, one has

$$\lim_{k\to+\infty}\Big(1-\varepsilon_k^{-2\delta}\Delta(x_k,\varepsilon_k)\Big)\neq 0.$$

It holds that $\lim_{k\to+\infty}\varepsilon_k=\varepsilon^*=0$.

We proceed to prove that $x^*\in S$. Together with (9), we can obtain

$$\varepsilon_k^{2\delta-\alpha}(1-\varepsilon_k^{-2\delta}\Delta(x_k,\varepsilon_k))\nabla f(x_k) +2\Big(\sum_{j\in E}(F_j(x_k)-\varepsilon_k^{\gamma}w_j)\nabla F_j(x_k) +\sum_{\ell\in I^+(x_k,\varepsilon_k)}(g_\ell(x_k)-\varepsilon_k^{\gamma}w_\ell)\nabla g_\ell(x_k)\Big) = 0.$$

Taking the limit in both sides, we have

$$\sum_{j\in E}(F_j(x^*)-(\varepsilon^*)^{\gamma}w_j)\nabla F_j(x^*) +\sum_{\ell\in I^+(x^*,\varepsilon^*)}(g_\ell(x^*)-(\varepsilon^*)^{\gamma}w_\ell)\nabla g_\ell(x^*)=0.$$

Since $\nabla F_j(x^*)$ $(j\in E)$, $\nabla g_\ell(x^*)$ $(\ell\in I^+(x^*,\varepsilon^*))$ are linearly independent, we have

$$F_j(x^*)-\varepsilon^{*^\gamma}w_j=0,\quad g_\ell(x^*)-\varepsilon^{*^\gamma}w_\ell=0,$$

for $\forall\, j\in E$ and $\ell\in I^+(x^*,\varepsilon^*)$. It implies that

$$\Delta(x^*,\varepsilon^*) =\sum_{j\in E}(F_j(x^*)-(\varepsilon^*)^{\gamma}w_j)^2 +\sum_{\ell\in I^+(x^*,\varepsilon^*)}(g_\ell(x^*)-(\varepsilon^*)^{\gamma}w_\ell)^2=0.$$

Therefore,

$$F_j(x^*)-(\varepsilon^*)^{\gamma}w_j=F_j(x^*)=0\ \forall j\in E;$$
$$g_\ell(x^*)-(\varepsilon^*)^{\gamma}w_\ell=g_\ell(x^*)=0\ \forall \ell\in I^+(x^*,\varepsilon^*).$$

We have $F_j(x^*)=0,\forall j\in E, g_\ell(x^*)\leqslant 0,\forall\ell\in I$, i.e., $x^*\in S$. The proof is completed. □

Based on above, we construct the following global convergence theorem.

Theorem 1 *Suppose that $(x_k,\varepsilon_k)\in L(P_{\sigma_k})$ generated by the Algorithm 1 with finite $f_{\sigma_k}(x_k,\varepsilon_k)$. For any accumulation point (x^*,ε^*), $\nabla F_j(x^*),\forall j\in E,\nabla g_\ell(x^*),\forall\ell\in I^+(x^*,\varepsilon^*)$ are linearly independent, then x^* is a local optimal solution of (P).*

Proof: From the conditions, we know there exists a subsequence $(x_k, \varepsilon_k)_{\aleph}$ such that $(x_k, \varepsilon_k) \to (x^*, \varepsilon^*)$, It then follows from Lemma 3 that $\varepsilon^* = 0$ and x^* is feasible point of the problem (P). Therefore, there exists a neighbor $o(x^*, 0)$ and consider an arbitrary point $(x, 0) \in o(x^*, 0) \cap (S \times \{0\})$, by the definition of (x_k, ε_k), one has

$$f(x^*) = f_\sigma(x^*, 0) \leqslant f_\sigma(x, 0) = f(x).$$

Therefore, x^* is a local optimal solution of (P). The proof is completed. □

Corollary 1 *Suppose that every local minimizer (x^*, ε^*) of the penalty problem (P_σ) with finite $f_\sigma(x^*, \varepsilon^*)$ and $\nabla F_j(x^*), \forall j \in E, \nabla g_\ell(x^*), \forall \ell \in I^+(x^*, \varepsilon^*)$ are linearly independent, then x^* is local minimizer of the primal problem (P) if and only if $\varepsilon^* = 0$.*

Proof: If x^* is local minimizer of the primal problem (P), then $F_j(x^*) = 0$ for all $j \in E$ and $g_\ell(x^*) \leqslant 0$ for all $\ell \in I$. Using a proof by contradiction, from Lemma 3, we have $\varepsilon^* = 0$. Alternatively, if $\varepsilon^* = 0$, in view of the construction of $f_\sigma(x, \varepsilon)$, x^* is feasible point of (P). From the hypothesis that $(x^*, 0)$ is optimal solution of (P_σ), x^* is local minimizer of the primal problem (P). □

Remark 2 Corollary 1 demonstrates another advantage of this penalty function is that ε can be regarded as an indicator variable of local (global) minimizer. In another words, under fairly general conditions, $\varepsilon^* = 0$ is equivalent to x^* is optimal solution of (P).

The next theorem explores that the finite termination property of the penalty function $f_\sigma(x, \varepsilon)$. Through this conclusion, the optimal solutions of primal problem (P) can be achieved within finite steps.

Theorem 2 *If $(x_k, \varepsilon_k) \in L(P_{\sigma_k})$ generated by Algorithm 1 with finite $f_{\sigma_k}(x_k, \varepsilon_k)$, $(x_k, \varepsilon_k) \to (x^*, \varepsilon^*)$ and $\nabla F_j(x^*), \forall j \in E, \nabla g_\ell(x^*), \forall \ell \in I^+(x^*, \varepsilon^*)$ are linearly independent, $\alpha, \beta, \gamma, \delta$ satisfy*

$$\alpha - \beta \geqslant 0 \tag{13}$$

then there exists $k_0 > 0$, when $k \geqslant k_0$, we have $\varepsilon_k = 0, x_k \in L(P)$.

Proof: We prove this theorem by contradiction. Assume the theorem is not true, then there exists a subsequence $\{(x_{n_k}, \varepsilon_{n_k})\}_{\aleph} \subseteq \{(x_k, \varepsilon_k)\}$ such that for any $k_0 > 0$, when $n_k \geqslant k_0, (x_{n_k}, \varepsilon_{n_k}) \in L(\mathrm{P}_{n_k})$ with finite $f_{\sigma_{n_k}}(x_{n_k}, \varepsilon_{n_k})$ and $\varepsilon_{n_k} \neq 0$ and the conditions of Theorem 2 hold for such subsequence. From the statement of Lemma 1, $(x_{n_k}, \varepsilon_{n_k}) \notin S_{\varepsilon_{n_k}}$ holds. From (8), we know that

$$\begin{aligned}
&\frac{\partial f_{\sigma_{n_k}}(x_{n_k}, \varepsilon_{n_k})}{\partial \varepsilon} \\
&= -\alpha \varepsilon_{n_k}^{\alpha-1} \ln\Big(1 - \varepsilon_{n_k}^{-2\delta} \Delta(x_{n_k}, \varepsilon_{n_k})\Big) \\
&\quad + \frac{\varepsilon_{n_k}^{\alpha-1}}{1 - \varepsilon_{n_k}^{-2\delta} \Delta(x_{n_k}, \varepsilon_{n_k})} \Big(2(\gamma - \delta) \frac{\Delta(x_{n_k}, \varepsilon_{n_k})}{\varepsilon_{n_k}^{2\delta}} \\
&\qquad - 2\gamma \varepsilon_{n_k}^{-2\delta} \Big(\sum_{j \in E} (F_j(x_{n_k}) - \varepsilon_{n_k}^{\gamma} w_j) F_j(x_{n_k}) \\
&\qquad\quad + \sum_{\ell \in I^+(x_{n_k}, \varepsilon_{n_k})} (g_\ell(x_{n_k}) - \varepsilon_{n_k}^{\gamma} w_\ell) g_\ell(x_{n_k}) \Big)\Big) \\
&\quad + \beta \varepsilon_{n_k}^{\beta-1} \sigma_{n_k} \\
&= 0
\end{aligned} \tag{14}$$

From (14), we obtain

$$\begin{aligned}
&-\alpha \varepsilon_{n_k}^{\alpha-\beta} \ln\Big(1 - \frac{\Delta(x_{n_k}, \varepsilon_{n_k})}{\varepsilon_{n_k}^{2\delta}}\Big) + \frac{\varepsilon_{n_k}^{\alpha-\beta}}{1 - \varepsilon_{n_k}^{-2\delta} \Delta(x_{n_k}, \varepsilon_{n_k})} \\
&\quad \times \Big(2(\gamma - \delta) \Delta(x_{n_k}, \varepsilon_{n_k}) \\
&\qquad - 2\gamma \Big(\sum_{j \in E} (F_j(x_{n_k}) - \varepsilon_{n_k}^{\gamma} w_j) F_j(x_{n_k}) \\
&\quad + \sum_{\ell \in I^+(x_{n_k}, \varepsilon_{n_k})} (g_\ell(x_{n_k}) - \varepsilon_{n_k}^{\gamma} w_\ell) g_\ell(x_{n_k}) \Big)\Big) + \beta \sigma_{n_k} \\
&= 0.
\end{aligned} \tag{15}$$

From Lemma 2, we derive $\varepsilon_{n_k} \to \varepsilon^* = 0$, $x_{n_k} \to x^* \in S$. Combining with that $\varepsilon_{n_k} \neq 0, 0 < 1 - 2\varepsilon_{n_k}^{-\delta} \Delta(x_{n_k}, \varepsilon_{n_k}) < 1$, we have

$$\lim_{\substack{\varepsilon_{n_k} \to \varepsilon^* = 0 \\ x_{n_k} \to x^* \in S}} \Big(1 - \varepsilon_{n_k}^{-2\delta} \Delta(x_{n_k}, \varepsilon_{n_k})\Big) = c^* \in [\frac{1}{2}, 1].$$

Let $\alpha - \beta \geqslant 0$, then the first term and the second one of (15) tend to finite, and the third term tends to infinite, which is impossible. It implies that such subsequence cannot exist. Therefore, there exists $k_0 > 0$, when $k \geqslant k_0$, $\varepsilon_k = 0$, $(x_k, 0) \in L(\mathrm{P}_{\sigma_k})$. Thus, by $(x_k, 0) \in L(\mathrm{P}_{\sigma_k})$, there exists a neighbor $o(x_k, 0)$ at $(x_k, 0), \sigma_k > 0$, for all $(x, 0) \in o((x_k, 0), \sigma_k) \cap (S \times \{0\})$, it holds

$$f(x_k) = f_{\sigma_k}(x_k, 0) \leqslant f_{\sigma_k}(x, 0) = f(x).$$

Thus, $x_k \in L(\mathrm{P})$. The proof is completed. □

LOCAL EXACTNESS PROPERTY

In this section, we shall show that, under fairly general conditions and some additional hypothesis, $(x^*,0)$ is a local optimal solution of penalty problem (P_σ) if x^* is a local minimizer of the original problem (P) for sufficiently large penalty parameter σ.

We now consider the nonsmooth case. Assume $f(x)$ and $F_j(x)$ $(j \in E)$, and $g_\ell(x)$ $(\ell \in I)$ are nonsmooth functions. In order to regularize f and g, we embed $f(x), F_j(x), j \in E$ and $g_\ell(x), \ell \in I$ into the smoothing function $f(x,\varepsilon), F_j(x,\varepsilon), \forall j \in E$ and $g_\ell(x,\varepsilon), \forall \ell \in I$ by introducing the above variable ε. Therefore, the introduced additional variable ε play critical roles in solving the problem (P). The variable ε has active actions not only in perturbation for constraint system no matter how many constrained functions, but also in regularization of the nonsmooth case. After regularization, the regularized functions $f(x,\varepsilon), F_j(x,\varepsilon)$ and $g_\ell(x,\varepsilon)$ are continuously differentiable in (x,ε), when $\varepsilon \neq 0$ and satisfy

$$
\begin{aligned}
f(x) &= f(x,0) = \lim_{\varepsilon\to 0} f(x,\varepsilon)\\
F_j(x) &= F_j(x,0) = \lim_{\varepsilon\to 0} F_j(x,\varepsilon), \forall j \in E\\
g_\ell(x) &= g_\ell(x,0) = \lim_{\varepsilon\to 0} g_\ell(x,\varepsilon), \forall \ell \in I.
\end{aligned}
$$

We consider the following system

$$
\begin{cases}
\min\limits_{(x,\varepsilon)\in\mathbb{R}^{n+1}} & f(x,\varepsilon)\\
\text{s.t.} & F_j(x,\varepsilon) = 0, \quad \forall j \in E,\\
 & g_\ell(x,\varepsilon) \leqslant 0, \quad \forall \ell \in I.
\end{cases} \qquad (P_\varepsilon)
$$

Now we introduce the definition of error bound[9].

Definition 1 We denote $x \in \mathbb{R}^n$ satisfies the following system

$$
\begin{cases}
F(x) = 0,\\
g(x) \leqslant 0,
\end{cases}
$$

as a set S. This system is said to satisfy a local error bound at x^*, if there exist positive constants $k > 0$ and $\delta > 0$ such that

$$
\operatorname{dist}(x \mid S) \leqslant k(\|F(x)\| + \|g(x)^+\|)
$$

holds, for all $x \in x^* + \delta\mathbb{B}$, where $\mathbb{B}$ is the closed unit ball in $\mathbb{R}^n$.

In the following part, the conditions that the error bound for (P) exist are considered. We make some assumptions:

(A$_1$) $f(\cdot,0)$ is Lipschitz continuous with Lipschitz constant L.

(A$_2$) The Mangasarian-Fromovitz constraint qualification holds at $(x^*,0)$.

We know the assumption (A$_2$) guarantees[9] the error bound condition holds. Furthermore, combining with Corollaries 2.3.1 and 2.4.1[9] or Theorem 3.1[10], we obtain the following conclusion.

Lemma 3 *If (A$_1$) and (A$_2$) hold, there exist a neighborhood N_0 of x^*, and a constant $\tau > 0$ such that*

$$
f(x,0) \geqslant f(x^*,0) - \tau\Big(\sum_{j\in E}\|F_j(x,0)\| + \sum_{\ell\in I}\|g_\ell(x,0)^+\|\Big)
$$

holds.

Now we present an important theoretical result of the local exactness proof. Before proving this result, some more assumptions are first given as follows.

(H$_1$) δ, β, γ are positive even integers and satisfy $\delta \geqslant \beta$ and $\gamma \geqslant \beta$;

(H$_2$) For sufficiently small $0 < \varepsilon' \ll 1$,

$$
\|g_\ell(x,\varepsilon) - g_\ell(x,0)\| \leqslant K\varepsilon^\beta,
$$
$$
\|F_j(x,\varepsilon) - F_j(x,0)\| \leqslant K\varepsilon^\beta,
$$

for all $\ell \in I, j \in E$, and $\varepsilon \in [-\varepsilon', 0) \cup (0, \varepsilon']$

(H$_3$) $|f(x,\varepsilon) - f(x,0)| \leqslant K\varepsilon^\beta$, the domain of ε as **(H$_2$)**;

Based on the above hypothesis, we will present the main results in this section.

Theorem 3 *Suppose the assumptions **(H$_1$)**–**(H$_3$)** hold, for sufficiently large σ, there are a neighborhood $N \subseteq N_0$ of x^* and sufficiently small $0 < \varepsilon' \ll 1$ such that*

$$
f_\sigma(x,\varepsilon) > f_\sigma(x^*,0) = f(x^*)
$$

for all $(x,\varepsilon) \in N \times [-\varepsilon',0) \cup (0,\varepsilon']$. In particular, $(x^*,0)$ is a local minimizer of $f_\sigma(x,\varepsilon)$.

Proof: Let the neighborhood $N \subseteq N_0$ of x^* be sufficiently small such that

$$
\sup_{x\in N}\{f(x^*,0) - f(x,0)\} \leqslant 1,
$$

and assume that the penalty parameter

$$
\sigma \geqslant K + \tau(K+2)(|E| + |I|).
$$

We divide into two cases for further analysis.

Case 1. $\Delta(x,\varepsilon) \geqslant \varepsilon^{2\delta}$;

Case 2. $\Delta(x,\varepsilon) < \varepsilon^{2\delta}$, for $x \in N, \varepsilon \in [-\varepsilon', 0) \cup (0, \varepsilon']$.

Case 1. By the construction of penalty function, $f_\sigma(x,\varepsilon) = +\infty$. Therefore, $f_\sigma(x,\varepsilon) > f_\sigma(x^*,0)$.

Case 2. We have $\Delta(x,\varepsilon) < \varepsilon^{2\delta}$, i.e.,

$$\sum_{j\in E}(F_j(x)-\varepsilon^\gamma w_j)^2 + \sum_{\ell\in I^+(x,\varepsilon)}(g_\ell(x)-\varepsilon^\gamma w_\ell)^2 < \varepsilon^{2\delta},$$

this yields that

$$\begin{aligned}\|F_j(x,\varepsilon)\| &\leqslant \varepsilon^\gamma|w_j| + \|F_j(x,\varepsilon)-\varepsilon^\gamma w_j\|\\ &< \varepsilon^\gamma|w_j| + \varepsilon^\delta,\\ \|g_\ell(x,\varepsilon)\| &\leqslant \varepsilon^\gamma|w_\ell| + \|g_\ell(x,\varepsilon)-\varepsilon^\gamma w_\ell\|\\ &< \varepsilon^\gamma|w_\ell| + \varepsilon^\delta.\end{aligned}$$

Furthermore, together with Lemma 3 and assumptions $(\mathbf{H_1})$–$(\mathbf{H_3})$

$$\begin{aligned}&f(x^*,0)\\ &\leqslant f(x,0) + \tau\Big(\sum_{j\in E}\|F_j(x,0)\| + \sum_{\ell\in I^+(x,0)}\|g_\ell(x,0)\|\Big)\\ &\leqslant f(x,\varepsilon) + K\varepsilon^\beta + \tau\Big(\sum_{j\in E}\|F_j(x,\varepsilon)\| + \sum_{j\in E}K\varepsilon^\beta\\ &\quad + \sum_{\ell\in I^+(x,0)}\|g_\ell(x,\varepsilon)\| + \sum_{\ell\in I^+(x,0)}K\varepsilon^\beta\Big)\\ &< f(x,\varepsilon) + K\varepsilon^\beta + \tau\Big(\sum_{j\in E}\varepsilon^\gamma|w_j| + \varepsilon^\delta|E|\\ &\quad + K\varepsilon^\delta|E| + \sum_{\ell\in I}\varepsilon^\gamma|w_\ell| + \varepsilon^\delta|I| + K\varepsilon^\delta|I|\Big)\\ &\leqslant f(x,\varepsilon) + K\varepsilon^\beta + \tau(K+2)(|E|+|I|)\varepsilon^\beta\\ &\leqslant f(x,\varepsilon) + \sigma\varepsilon^\beta.\end{aligned}$$

where $|E|, |I|$ denote the dimension of equality constraint and inequality constraint respectively. The second inequality follows from $(\mathbf{H_2})$ and $(\mathbf{H_3})$. The fourth inequality follows immediately from the assumption $(\mathbf{H_1})$. Therefore, $f(x^*,0) < f(x,\varepsilon) + \sigma\varepsilon^\beta \leqslant f_\sigma(x,\varepsilon)$. This yields the inequality as desired. □

NUMERICAL EXAMPLES

To give some insight into the behavior of the algorithm presented in this paper. We use $\|\nabla_{(x,\varepsilon)} f_\sigma(x,\varepsilon)\| \leqslant 10^{-6}$ as stopping criteria. Tables Table 1–5 show the computational results for the corresponding problem with the following items: the penalty parameter σ_k, x_k, ε_k of the final iterate and $f(x_k)$ the function value of f at the final x_k, and the constraint violation measure $\Delta(x_k,\varepsilon_k)$. In this section, the parameters used in this algorithm are set as $\alpha = 5, \beta = 1.9, \gamma = 4$ and $\delta = 3$.

Example 1

$$\begin{aligned}\min\quad & 5x_1x_2x_3 - \tfrac{1}{2}x_1^2 + 10(x_1-1)^2 - 2x_2x_3\\ & -x_3 - \tfrac{3}{2}x_2^2 - x_3^2,\\ \text{s.t.}\quad & -x_1^2 - x_3^2 - x_1 - 2x_2 - x_3 + 2 = 0,\\ & x_1 + \tfrac{3}{4} \geqslant 0,\\ & (x_1-x_3)^2 + x_2^3 - 0.1x_1 + 0.05x_1^2 + 1.05 \geqslant 0.\end{aligned}$$

We choose $x_0 = (0,0,0), \varepsilon_0 = 20$ as initial point. $\rho = 5$. The optimal solution and optimal value are $x^* = (1,-1,1)$ and $f(x^*) = -7.0000$ of the above example.

Table 1 Numerical results of Example 1

σ_k	x_k	ε_k	$f(x_k)$	$\Delta(x_k,\varepsilon_k)$
10	(1.065, -1.128, 0.405)	0.071	-4.456	0
15	(1.000, -0.839, 0.888)	-0.000	-5.467	0
20	(1.150, -1.003, 0.835)	0.037	-6.585	2.5619e-009
25	(0.991, -0.994, 1.005)	0.011	-6.932	1.8309e-012
30	(1.012, -1.000, 0.986)	0.000	-6.991	0

Example 2

$$\begin{aligned}\min\quad & x_1^2 + x_1x_2 + 2x_2^2 - 6x_1 - 14x_2 - 12x_3\\ \text{s.t.}\quad & x_1 + x_2 + x_3 = 20;\\ & x_1 + 2x_2 \leqslant 30;\\ & x_1, x_2, x_3 \geqslant 0.\end{aligned}$$

Here, we choose $x_0 = (7,7,7), \varepsilon_0 = 2$ as initial point. $\rho = 5$. The optimal solution and optimal value are $x^* = (0, 0.5, 19.5)$ and $f(x^*) = -240.5$.

Table 2 Numerical results of Example 2

σ_k	x_k	ε_k	$f(x_k)$	$\Delta(x_k,\varepsilon_k)$
10	(-0.033, -0.021, 20.041)	0.28	-239.22	0.002
15	(-0.009, 0.498, 19.508)	0.16	-240.14	9.827e-005
20	(-0.006, 0.480, 19.527)	0.13	-240.20	3.704e-005

Example 3

$$\left\{\begin{aligned}\min\quad & x_1^3 + 2x_2^2x_3 + 2x_3,\\ \text{s.t.}\quad & x_1^2 + x_2 + x_3^2 = 4,\\ & x_1^2 - x_2 + 2x_3 \leqslant 2,\\ & x_1, x_2, x_3 \geqslant 0.\end{aligned}\right.$$

Here, we choose $x_0 = (-2,-2,1), \varepsilon_0 = 2$ and $x_0 = (-1,2,-1), \varepsilon_0 = 2$ as initial points, respectively. $\rho = 5$. The optimal solution and optimal value are $x^* = (0,4,0)$ and $f(x^*) = 0$ of the above example.

Table 3 Numerical result of Example 3

σ_k	x_k	ε_k	$f(x_k)$	$\Delta(x_k, \varepsilon_k)$
25	(0.000, 4.013, -0.627)	0.866	-3.846	0.396
30	(0.000, 3.992, -0.416)	0.739	-1.240	0.157
35	(0.008, 4.000, -0.002)	0.114	0.001	2.159e-006

Example 4

$$\begin{cases} \min & \cos x_1 \sin x_2 - \frac{x_1}{x_2^2+1} \\ \text{s.t.} & -1 \leqslant x_1 \leqslant 2, \\ & -1 \leqslant x_2 \leqslant 1. \end{cases}$$

Initial point is $x_0 = (4, 0), \varepsilon_0 = 2$. $\rho = 2$.The optimal solution and optimal value are $x^* = (2, 0.1058)$ and $f(x^*) = -2.02181$ of the above example.

Table 4 Numerical result of Example 4

σ_k	x_k	ε_k	$f(x_k)$	$\Delta(x_k, \varepsilon_k)$
2	(2.000, 0.109)	0.009	-2.021	6.055e-013
4	(2.000, 0.052)	0.013	-2.015	5.745e-012
6	(0.696, -0.889)	-0.000	-0.985	0
8	(1.997, 0.104)	-0.000	-2.019	0
10	(2.000, 0.107)	-0.000	-2.021	0

Example 5

$$\begin{cases} \min & x_1^2 + x_2^2 + (x_3 - 10)^2 + 4(x_4 - 5)^2 \\ & +(x_5 - 3)^2 + 2(x_6 - 1)^2 + 5x_7^2 + 7x_8^2 + 2x_9^2 \\ & +(x_{10} - 7)^2 + x_1 x_2 - 14x_1 - 16x_2 + 45 \\ \text{s.t.} & 3(x_1 - 2)^2 + 4(x_2 - 3)^2 + 2x_3^2 - 7x_4 \leqslant 120, \\ & 5x_1^2 + 4(x_3 - 6)^2 + 8x_2 - 2x_4 \leqslant 40, \\ & \frac{1}{2}(x_1 - 8)^2 + 2(x_2 - 4)^2 + 3x_5^2 - x_6 \leqslant 30, \\ & x_1^2 + 2(x_2 - 2)^2 - 2x_1 x_2 + 14x_5 - 6x_6 \leqslant 0, \\ & 4x_1 + 5x_2 - 3x_7 + 9x_8 \leqslant 105, \\ & 10x_1 - 8x_2 - 17x_7 + 2x_8 \leqslant 0, \\ & 12(x_9 - 8)^2 - 3x_1 + 6x_2 - 7x_{10} \leqslant 0, \\ & -8x_1 + 2x_2 + 5x_9 - 2x_{10} \leqslant 12, \\ & x_1, x_2, \cdots, x_{10} \geqslant 0. \end{cases}$$

The optimal solution and optimal value are $x^* = (1.8388, 3.3026, 7.3159, 5.1275, 0.9962, 1.4294, 0, 0, 6.0187, 8.7721)$ and $f(x^*) = 74.0196$ of the above example. We choose $x_0 = (1, 0, 0, 0, 0, 0, 0, 0, 0.6, 1.1)$, $\varepsilon_0 = 5.1$ as initial point. $\rho = 2$.

Table 5 Numerical result of Example 5

σ_k	x_k	ε_k	$f(x_k)$	$\Delta(x_k, \varepsilon_k)$
2	(3.72,5.91,9.69,5.01,3.00,1.00,0,0,1.77,7.17)	4.8	22.4	0.11
4	(2.99,5.61,8.86,5.03,2.69,1.07,0,0,3.10,7.37)	4.3	63.6	0.59
6	(1.92,3.18,7.26,5.11,1.16,1.76,0,0,6.00,9.08)	0.1	74.6	1.46e-006
8	(1.72,3.70,7.18,5.11,1.00,1.69,0,0,6.09,8.68)	0.0	74.7	7.84e-010

Acknowledgements: Cheng Ma's work is supported by the National Natural Science Foundation (11401331,11671220), China Postdoctoral Science Foundation (2016M592148), the Postdoctoral Science Foundation of Shandong Province (201603063), the Postdoctoral Science Foundation of Qingdao city (2016032) and the humanities and social sciences project of Shandong provincial university (J17RA107).

REFERENCES

1. Bazaraa MS, Sherali HD, Shetty CM (2006) *Nonlinear Programming: Theory and Algorithms* (3rd edition). John Wiley & Sons.
2. Wachter A, Biegler LT (2006) On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Math Program* **106**, 25–57.
3. Chen L, Goldfarb D (2006) Interior-point l_2-penalty methods for nonlinear programming with strong global convergence properties. *Math Program* **108**, 1–36.
4. Han SP, Mangasarian OL (1979) Exact penalty functions in nonlinear programming. *Math Program* **17**, 251–69.
5. Lian SJ, Duan YQ (2016) Smoothing of the lower-order exact penalty function for inequality constrained optimization. *J Inequal Appl* **185**, 1–12.
6. Lian SJ, Zhang LS (2012) A simple smooth exact penalty function for smooth optimization problem. *J Syst Sci Complex* **25(5)**, 521–8.
7. Huyer W, Neumair A (2003) A new exact penalty function. *SIAM J Control Optim* **13**, 1141–59.
8. Wang CY, Ma C, Zhou JC (2014) A New Class of Exact Penalty Functions and Penalty Algorithms. *J Global Optim* **58(1)**, 58–73.
9. Burke JV (1991) An exact penalization viewpoint of constrained optimization. *SIAM J Control Optim* **29**, 968–98.
10. Pang JS (1997) Error bound in Mathematical Programming. *Math Program* **79**, 299–332.