J Glob Optim (2014) 58:51–73
DOI 10.1007/s10898-013-0111-9

A new class of exact penalty functions and penalty algorithms

Changyu Wang · Cheng Ma · Jinchuan Zhou

Received: 22 November 2011 / Accepted: 17 September 2013 / Published online: 29 September 2013

Abstract For nonlinear programming problems, we propose a new class of smooth exact penalty functions, which includes both barrier-type and exterior-type penalty functions as special cases. We develop necessary and sufficient conditions for exact penalty property and inverse proposition of exact penalization, respectively. Furthermore, we establish the equivalent relationship between these penalty functions and classical simple exact penalty functions in the sense of exactness property. In addition, a feasible penalty function algorithm is proposed. The convergence analysis of the algorithm is presented, including the global convergence property and finite termination property. Finally, numerical results are reported.

Keywords Nonlinear programming · Smooth and exact penalty function · Constraint qualifications · Penalty function algorithms

Mathematics Subject Classification 90C26 · 90C30

1 Introduction

The penalty function algorithm is an important method for solving constrained optimization problems, which, by augumenting a penalty term, transforms the original problem into a

This research was supported by the National Natural Science Foundation of China (10971118, 11271226, 11271233, 11101248, 10901096).

C. Wang
Institute of Operations Research, Qufu Normal University, Qufu, China
e-mail: wcy0537@126.com

C. Ma (✉)
College of Business, Qingdao University, Qingdao, China
e-mail: mc_0812@163.com

J. Zhou
Department of Mathematics, School of Science, Shandong University of Technology, Zibo, China
e-mail: jinchuanzhou@163.com

single unconstrained (or simple constrained) problem or into a sequence of unconstrained problems. For classical penalty function algorithms, we need to make the penalty parameter infinitely large in a limiting sense to recover an optimal solution of the original problem, which incurs numerical instability in implementation. To avoid this difficulty, we require the penalty function to satisfy the following property: recovering an exact solution of the original problem for reasonable finite values of the penalty parameter. The penalty function possessing this property is said to be exact. However, it should be identified that some exact penalty functions have the disadvantage that they either need Jacobian [7,9,14,15,20] or are no longer smooth (l_1 or l_∞ penalty functions etc. [1,3,4,6,12,16,18,19,21,22]).

In this paper, we are mainly concerned with the following nonlinear programming problem

$$(P)\quad \begin{aligned} &\min\ f(x)\\ &\text{s.t. } F(x)=0,\quad x\in[u,v],\end{aligned}$$

where $[u,v]=\{x\in\mathbb{R}^n|u\le x\le v\}$, $u\in(\{-\infty\}\cup\mathbb{R})^n$, $v\in(\{+\infty\}\cup\mathbb{R})^n$ and $\operatorname{int}[u,v]\neq\emptyset$. The functions $f:D\to\mathbb{R}$ and $F:D\to\mathbb{R}^m$ are continuously differentiable on an open set D satisfying $[u,v]\subset D$. We always assume that the feasible region is nonempty and the function f is bounded below on D, because f can be replaced by e^f otherwise.

Let $\omega\in\mathbb{R}^m$ be fixed. The problem (P) can be written equivalently as

$$\begin{aligned} \min\ & f(x)\\ \text{s.t.}\ & F(x)=\varepsilon\omega,\\ & x\in[u,v],\quad \varepsilon=0.\end{aligned}$$

For this problem, Huyer and Neumaier [13] introduce the following exact penalty function

$$f_\sigma(x,\varepsilon)=\begin{cases} f(x), & \text{if } \varepsilon=\Delta(x,\varepsilon)=0;\\ f(x)+\frac{1}{2\varepsilon}\frac{\Delta(x,\varepsilon)}{1-q\Delta(x,\varepsilon)}+\sigma\beta(\varepsilon), & \text{if } \varepsilon>0,\ \Delta(x,\varepsilon)<q^{-1};\\ +\infty, & \text{otherwise,}\end{cases}\tag{1.1}$$

where $q>0$ is a given positive constant, $\sigma>0$ is a penalty parameter, $\beta:[0,\overline{\varepsilon}]\to[0,\infty)$ is continuous on $[0,\overline{\varepsilon}]$ with $\overline{\varepsilon}>0$ and continuously differentiable on $(0,\overline{\varepsilon}]$ with $\beta(0)=0$. The term $\Delta(x,\varepsilon)$ measures the violation of the constraints, i.e.,

$$\Delta(x,\varepsilon)=\|F(x)-\varepsilon\omega\|^2=\sum_{j=1}^m\left(F_j(x)-\varepsilon\omega_j\right)^2.$$

The corresponding penalty problem to (1.1) is

$$(P_\sigma)\quad \begin{aligned} &\min\ f_\sigma(x,\varepsilon)\\ &\text{s.t. } (x,\varepsilon)\in[u,v]\times[0,\overline{\varepsilon}].\end{aligned}$$

Note that f_σ is continuously differentiable on $D_q=\{(x,\varepsilon)\in D\times(0,\overline{\varepsilon})|\Delta(x,\varepsilon)<q^{-1}\}$. The exact penalty property of f_σ is discussed thoroughly in [13], e.g., for smooth case, the exact penalty property is established if D_F (see (3.1)) is bounded and each $x\in D_F$ satisfies Mangasarian-Fromovitz constraint qualification [13, Theorem 2.1]; for nonsmooth case of (P), sufficient conditions are established by the regular technique and other strong conditions [13, Theorem 5.3].

The above penalty function f_σ is simple and exact. Here, we call a penalty function "simple", if it involves objective and constraint functions of the original problem (P) and does not involve gradients or the Jacobian. The main reason of f_σ to have significant differences with the classical simple exact penalty functions is that f_σ has good smoothness property, which is not enjoyed by the latter [2,10]. In this paper, we restrict our focus primarily on the case of (P) being smooth. The main results are listed as follows.

(i) The term $\frac{\Delta(x,\varepsilon)}{1-q\Delta(x,\varepsilon)}$ in (1.1) plays a role as a barrier term. However, to compute an interior point is not an easy thing in practical applications. Therefore, we extend this term to a class of convex functions. This allows us to present a unified framework for some barrier-type and exterior-type penalty functions, in which the latter are smooth on $[u, v] \times (0, \bar{\varepsilon})$ and have a larger smooth area than the former. The extended penalty function class (2.1) in Sect. 2 provides several alternatives in the design of penalty function algorithms.

(ii) Based on (2.1) in Sect. 2, the corresponding penalty problems for (P) are proposed. We firstly introduce the extended Mangasarian-Fromovitz constraint qualification, which generalizes the classical Mangasarian-Fromovitz constraint qualification from a single point to an infinite sequence. The sufficient condition for exact penalty property of $\widetilde{f}_\sigma$ is established, provided that ∇f is bounded on D_F and the extended Mangasarian-Fromovitz constraint qualification holds (see Theorem 3.1). This finding generalizes [13, Theorem 2.1]. Furthermore, necessary conditions for exact penalty property are obtained (see Theorem 3.2).

(iii) Following by a series of technical lemmas, we show that these necessary conditions are also necessary and sufficient conditions for inverse proposition of exact penalization (see Theorem 4.1). This result accurately characterizes the equivalence between $\widetilde{f}_\sigma$ and the classical simple exact penalty functions in the sense of exactness property. It demonstrates that $\widetilde{f}_\sigma$ possesses exactness property as the classical simple penalty functions as well as the smoothness property, which is not shared by the classical simple exact penalty functions.

(iv) It should be mentioned that, even if the original programming problem has optimal solutions, the penalty functions may fail to locate an optimal solution. In this case, it is impossible to obtain optimal solutions via penalty function algorithms, because in some penalty function algorithms, we need to find the optimal solution of penalty subproblem at each iteration. It also may occur even for the smooth penalty function with lower bound. For example, although the penalty function proposed in [8] is smooth and bounded from below, there exists an example to indicate that, the penalty function may fail to locate optimal solutions even though the original problem has an optimal solution. This causes the penalty function algorithm proposed in [8] and its revised version inapplicable. In addition, some papers discuss the convergence of iterative sequences generated by penalty function algorithms to FJ points or KKT points; see [5,17,20]. In order to avoid the cases mentioned above, we present a class of revised penalty function algorithms. The main feature of this class of algorithms is that if the optimal solution of the penalty subproblem does not exist, then we resort to finding a δ-optimal solution of the penalty subproblem. Note that the δ-optimal solution always exists, because penalty functions are bounded from below. This ensures the revised penalty function algorithms are always feasible. In addition, utilizing the exact penalty property and structural features of the class of penalty functions, we show that, under certain conditions, the proposed algorithm has finite termination property, i.e., the algorithm terminates at the optimal solution of the original problem after finitely many iterations; otherwise, by a perturbation theorem, the global convergence property is given, that is, every accumulation point of the sequence generated by the algorithm is an optimal solution of the original problem.

The organization of this paper is as follows. Section 2 extends the penalty function proposed by Huyer and Neumaier (1.1) to a class of penalty functions (2.1), establishes corresponding penalty problems and introduces the extended Mangasarian-Fromovitz constraint qualification. The necessary and sufficient conditions for exact penalty property are developed in Sect. 3. In Sect. 4, necessary and sufficient conditions for the inverse propositions of exact penalization are established. Additionally, the equivalence relationship between the new class of penalty functions and the classical simple exact penalty function in the sense of exactness property is established. Section 5 is devoted to the revised penalty function algorithm, where the finite termination property and global convergence property of the proposed algorithm are discussed, respectively. Numerical results are reported in Sect. 6.

2 A class of exact penalty functions

In this section, we extend the term $\frac{\Delta(x,\varepsilon)}{1-q\Delta(x,\varepsilon)}$ in (1.1) to a class of convex functions. This allows us to present a unified framework for some barrier-type and exterior-type penalty functions. Given $a \in (0, +\infty]$, let a function $\phi : [0, a) \to [0, +\infty)$ satisfy

(i_1) ϕ is convex and continuously differentiable on $[0, a)$ with $\phi(0) = 0$.

(i_2) $\phi'(t) > 0$ for all $t \in [0, a)$.

Many functions satisfy the conditions (i_1), (i_2), for example

$$\begin{array}{ll} \phi_1(t) = \frac{t}{(1-qt)^{\alpha}} & (a = q^{-1}, \alpha \geq 1); \\ \phi_2(t) = \tan(t) & (a = \frac{\pi}{2}); \\ \phi_3(t) = -\log(1 - t^{\alpha}) & (a = 1, \alpha \geq 1); \\ \phi_4(t) = t & (a = +\infty); \\ \phi_5(t) = e^t - 1 & (a = +\infty); \\ \phi_6(t) = \frac{1}{2}(\sqrt{t^2 + 4} + t) - 1 & (a = +\infty). \end{array}$$

Utilizing ϕ, a penalty function is given by

$$\widetilde{f}_{\sigma}(x, \varepsilon) = \begin{cases} f(x), & \text{if } \varepsilon = \Delta(x, \varepsilon) = 0, \\ f(x) + \frac{1}{2\varepsilon}\phi(\Delta(x, \varepsilon)) + \sigma\beta(\varepsilon), & \text{if } \varepsilon > 0, \Delta(x, \varepsilon) < a, \\ +\infty, & \text{otherwise.} \end{cases} \tag{2.1}$$

It is easy to see that $\widetilde{f}_{\sigma}$ is continuously differentiable on $D_a = \{(x, \varepsilon) \in D \times (0, \overline{\varepsilon}) | \Delta(x, \varepsilon) < a\}$. The barrier-type penalty functions correspond to the case of $a < +\infty$, and the exterior-type ones correspond to the case of $a = +\infty$.

The corresponding penalty problem is defined as

$$(\widetilde{P}_{\sigma}) \quad \min \ \widetilde{f}_{\sigma}(x, \varepsilon) \quad \text{s.t. } (x, \varepsilon) \in [u, v] \times [0, \overline{\varepsilon}].$$

Let's firstly recall Mangasarian-Fromovitz constraint qualification. Consider

$$\begin{cases} F(z) = 0, \\ g(z) \leq 0, \end{cases} \tag{2.2}$$

where $g : D \to \mathbb{R}^l$ is continuously differentiable, and F is defined as before. Denote by $\nabla F \in \mathbb{R}^{n \times m}$ the transpose of Jacobian of F.

Definition 2.1 For constraint system (2.2), we say that Mangasarian-Fromovitz constraint qualification (hereafter, MFCQ for short) holds at $z^* \in \mathbb{R}^n$, if

(1) $g(z^*) \leq 0$ and $\text{rank}(\nabla F(z^*)) = m$;
(2) there exists $p \in \mathbb{R}^n$ such that $\nabla F(z^*)^T p = 0$ and $\nabla g_j(z^*)^T p < 0$ for all $j \in J(z^*)$,

where $J(z^*) = \{j \in J | g_j(z^*) = 0\}$ and $J = \{1, 2, \dots, l\}$.

For an infinite sequence $K \subset \{1, 2, \dots\}$ and $\{z^k\}_{k\in K} \subset \mathbb{R}^n$, denote

$$J^+(K) = \{j \in J | \limsup_{k\in K, k\to\infty} g_j(z^k) \geq 0\} \quad \text{and} \quad J^-(K) = \{j \in J | \limsup_{k\in K, k\to\infty} g_j(z^k) < 0\}.$$

Definition 2.2 For constraint system (2.2), we say that an extended Mangasarian-Fromovitz constraint qualification (hereafter, EMFCQ for short) holds for $\{z^k\}_{k\in K}$, if there exist a matrix ∇F^* and an infinite subset $K_0 \subset K$ such that

(1) $\lim_{k\in K_0, k\to\infty} \nabla F(z^k) = \nabla F^*$ and $\text{rank}(\nabla F^*) = m$;
(2) there exists $p \in \mathbb{R}^n$ such that $(\nabla F^*)^T p = 0$ and $\limsup_{k\in K_0, k\to\infty} \nabla g_j(z^k)^T p < 0$ for all $j \in J^+(K_0)$.

Obviously, if MFCQ holds at z^*, then EMFCQ holds for the constant sequence $\{z^k\}_{k\in K}$ satisfying $z^k = z^*$ for all $k \in K$. Furthermore, from Definitions 2.1 and 2.2, we readily get the following result.

Proposition 2.1 *Let $\{z^k\}_{k\in K} \subset \mathbb{R}^n$. If the standard MFCQ holds for an accumulation point z^* of the sequence $\{z^k\}_{k\in K}$ for (2.2), then EMFCQ holds for $\{z^k\}_{k\in K}$.*

The following example shows that EMFCQ is suitable for an unbounded and infeasible sequence.

Example 2.1 Consider the following constraint system

$$\begin{cases} F(z) &= z_1 + z_2 + z_3^2 = 0, \\ g_1(z) &= z_1 - z_2 \leq 0, \\ g_2(z) &= z_1 - z_2 - z_3^2 \leq 0. \end{cases}$$

It is easy to see that EMFCQ holds for the unbounded sequence $\{z^k\} = \{(k+\frac{1}{k}, k, 1+\frac{1}{k})^T\}$. Now let us discuss EMFCQ in more detail. Clearly, from Definition 2.1, Definition 2.2 and Example 2.1, we generalize MFCQ from a single point z^* with $g(z^*) \leq 0$ to an infinite sequence $\{z^k\}_{k\in K}$. This sequence may be unbounded and infeasible, i.e., not necessarily satisfy $g(z^k) \leq 0$. Therefore, EMFCQ is suitable to weaken some assumptions on exactness property in the existing literature. In fact, it, to a great extent, mainly can be used to remove the level-bounded assumption. For instance, as stated in [13, Theorem 2.1], exactness property of the penalty function $f_\sigma(z, \varepsilon)$ is established, provided that the level set D_F is bounded and every point $z \in D_F$ satisfies MFCQ. These conditions ensure that the sequence $\{z^k\}_{k\in K}$ is bounded and every accumulation point satisfies MFCQ, where (z^k, ε_k) is a local optimal solution of the penalty problem $(\tilde{P}_\sigma)$. According to Proposition 2.1, EMFCQ holds for $\{z^k\}_{k\in K}$. Therefore, we can show that $\tilde{f}_\sigma(z, \varepsilon)$ is also exact (see Theorem 3.1 below) by requiring the validity of EMFCQ for $\{z^k\}_{k\in K}$ and the boundedness of ∇f over D_F. Therefore, [13, Theorem 2.1] is a special case of our result (see Corollary 3.1 and Remark 3.1 below).

Similarly, the level-bounded assumption in the convergence analysis of penalty function methods can be weakened by EMFCQ. For example, in [8], a class of penalty function methods (see [8, Algorithm 1]) was proposed for the nonlinear programming problem with inequality constraints. Notice that, in general, the iteration point z^k generated by Algorithm

1 is infeasible. By the assumptions that the level set Ω_ε is bounded and all global optimal solutions of the original problem satisfy MFCQ (see [8, Assumptions ($H1$) and ($H2$)]), an infinite sequence $\{z^k\}$ generated by Algorithm 1 is proven to be feasible as k sufficiently large. Similar to the above discussion, it immediately follows from Assumptions ($H1$), ($H2$) and [8, Lemma 3] that $\{z^k\}$ is bounded and every accumulation point satisfies MFCQ. Thus, we recover [8, Theorem 1] by replacing these assumptions by EMFCQ holds for $\{z^k\}_{k\in K}$ and ∇f is bounded on the level set. This indicates that [8, Theorem 1] is also a special case in our work.

3 Exact penalty property

This section deals mainly with the "exact" property of the penalty functions (2.1). In particular, [13, Theorem 2.1] is generalized by using EMFCQ, instead of standard MFCQ. Define the level set

$$D_F = \left\{x \in [u, v] \middle| \|F(x)\| \le \sqrt{a} + \bar{\varepsilon}\|\omega\|\right\}, \tag{3.1}$$

where $a \in (0, +\infty]$. If $a = +\infty$, then D_F reduces to $[u, v]$. Before proceeding, we need the following assumptions.

Assumption (A)
(A_1) ∇f is bounded on D_F;
(A_2) For any infinite subsequence $\{\sigma_k\}_{k\in K} \to +\infty$, EMFCQ holds for $\{x^k\}_{k\in K}$, where (x^k, ε_k) is a local optimal solution of $(\widetilde{P}_{\sigma_k})$ with finite value.

Lemma 3.1 *Suppose that there exists $\beta_1 > 0$ such that $\beta'(\varepsilon) \ge \beta_1$ for all $\varepsilon \in (0, \bar{\varepsilon}]$. If (x, ε) is a KKT point of $(\widetilde{P}_\sigma)$ with $\varepsilon > 0$, then*

$$2\beta_1\phi'(0)\sigma\varepsilon^2\frac{1}{\phi'(\Delta)^2} \le \|F(x)\|^2,$$

where Δ denotes $\Delta(x, \varepsilon)$ for simplification.

Proof If (x, ε) is a KKT point of $(\widetilde{P}_\sigma)$ with $\varepsilon > 0$, then by the construction of $\widetilde{f}_\sigma$, there exist $\lambda,\ \eta \in \mathbb{R}^n_+,\ \lambda_{n+1} \ge 0$, and $\eta_{n+1} \ge 0$ such that

$$\begin{aligned}
&\nabla f(x) + \frac{1}{\varepsilon}\phi'(\Delta)\nabla F(x)\big(F(x) - \varepsilon\omega\big) = \lambda - \eta,\\
&\inf(\lambda_i, x_i - u_i) = \inf(\eta_i, v_i - x_i) = 0, \qquad i = 1, 2, \ldots, n,\\
&-\frac{1}{2\varepsilon^2}\phi(\Delta) - \frac{1}{\varepsilon}\phi'(\Delta)\big(F(x) - \varepsilon\omega\big)^T\omega + \sigma\beta'(\varepsilon) = \lambda_{n+1} - \eta_{n+1}, \qquad (3.2)\\
&\lambda_{n+1} = \inf(\eta_{n+1}, \bar{\varepsilon} - \varepsilon) = 0. \qquad (3.3)
\end{aligned}$$

It follows from (3.2) and (3.3) that

$$-\frac{1}{2\varepsilon^2}\phi(\Delta) - \frac{1}{\varepsilon}\phi'(\Delta)\big(F(x) - \varepsilon\omega\big)^T\omega + \sigma\beta'(\varepsilon) \le 0,$$

from which and the fact that $\phi'(\Delta) > 0$ we have

$$-\frac{1}{\phi'(\Delta)}\phi(\Delta) - 2\varepsilon\big(F(x) - \varepsilon\omega\big)^T\omega + 2\varepsilon^2\sigma\beta'(\varepsilon)\frac{1}{\phi'(\Delta)} \le 0. \tag{3.4}$$

Rearranging (3.4) yields

$$-\frac{1}{\phi'(\Delta)}\phi(\Delta)+\Delta+\varepsilon^2\|\omega\|^2+2\varepsilon^2\sigma\beta'(\varepsilon)\frac{1}{\phi'(\Delta)}\le\|F(x)\|^2. \tag{3.5}$$

The convexity of ϕ, the condition (i_2) and the fact $\phi(0)=0$ ensure that

$$-\Delta\le\big(\phi(0)-\phi(\Delta)\big)\frac{1}{\phi'(\Delta)}=-\phi(\Delta)\frac{1}{\phi'(\Delta)}. \tag{3.6}$$

Combining (3.5) and (3.6) yields

$$2\varepsilon^2\sigma\beta'(\varepsilon)\frac{1}{\phi'(\Delta)}\le\|F(x)\|^2.$$

In consideration of $\beta'(\varepsilon)\ge\beta_1$ and the monotonicity of ϕ' (since ϕ is convex), the above inequality implies that

$$2\beta_1\phi'(0)\sigma\varepsilon^2\frac{1}{\phi'(\Delta)^2}\le\|F(x)\|^2.$$

□

Lemma 3.2 *Suppose that there exists $\beta_1>0$ such that $\beta'(\varepsilon)\ge\beta_1$ for all $\varepsilon\in(0,\bar{\varepsilon}]$. If (x^*,ε^*) is a local optimal solution of $(\widetilde{P}_\sigma)$ with finite optimal value, then x^* is a local optimal solution of (P) if and only if $\varepsilon^*=0$.*

Proof Since constraint functions of $(\widetilde{P}_\sigma)$ are all linear, then (x^*,ε^*) is also a KKT point of $(\widetilde{P}_\sigma)$. If x^* is a local optimal solution of (P), then $F(x^*)=0$. According to Lemma 3.1, we must have $\varepsilon^*=0$. Conversely, if $\varepsilon^*=0$, taking account of the finiteness of $\widetilde{f}_\sigma(x^*,0)$ and the construction of $\widetilde{f}_\sigma$, we have $F(x^*)=0$, and hence x^* is a local optimal solution of (P). □

Lemma 3.3 *Suppose that there exists $\beta_1>0$ such that $\beta'(\varepsilon)\ge\beta_1$ for all $\varepsilon\in(0,\bar{\varepsilon}]$. If (x^k,ε_k) is a KKT point of $(\widetilde{P}_{\sigma_k})$ with $\varepsilon_k>0$, then for $\sigma_k\to+\infty(k\to\infty)$, we have*

$$\lim_{k\to\infty}\frac{1}{\varepsilon_k}\phi'(\Delta_k)\sqrt{\Delta_k}=+\infty,$$

where $\Delta_k:=\Delta(x^k,\varepsilon_k)$.

Proof Lemma 3.1 implies that

$$\lim_{k\to\infty}\frac{1}{\varepsilon_k}\phi'(\Delta_k)\|F(x^k)\|=+\infty.$$

Note that

$$\frac{1}{\varepsilon_k}\phi'(\Delta_k)\|F(x^k)\|\le\phi'(\Delta_k)\left(\frac{1}{\varepsilon_k}\sqrt{\Delta_k}+\|\omega\|\right).$$

Therefore,

$$\lim_{k\to\infty}\phi'(\Delta_k)\left(\frac{1}{\varepsilon_k}\sqrt{\Delta_k}+\|\omega\|\right)=+\infty,$$

which, together with the monotonicity of ϕ', yields

$$\lim_{k\to\infty}\frac{1}{\varepsilon_k}\phi'(\Delta_k)\sqrt{\Delta_k}=+\infty.$$

□

Theorem 3.1 *Suppose that Assumptions* (A_1) *and* (A_2) *hold, and that there exists* $\beta_1 > 0$ *such that* $\beta'(\varepsilon) \geq \beta_1$ *for all* $\varepsilon \in (0, \overline{\varepsilon}]$. *When* $\sigma > 0$ *is sufficiently large, if* (x^*, ε^*) *is a local optimal solution of* $(\widetilde{P}_\sigma)$ *with finite optimal value, then* $\varepsilon^* = 0$. *Furthermore,* x^* *is a local optimal solution of* (P).

Proof It suffices to show that $\varepsilon^* = 0$. Suppose on the contrary that there exist $\sigma_k \to +\infty$ as $k \to \infty$ and a sequence of local optimal solutions (x^k, ε_k) of $(\widetilde{P}_{\sigma_k})$ with $\widetilde{f}_\sigma(x^k, \varepsilon_k)$ finite and $\varepsilon_k > 0$. Since all constraint functions are linear, then (x^k, ε^k) is a KKT point of $(\widetilde{P}_{\sigma_k})$, i.e., there exist $\lambda^k, \eta^k \in \mathbb{R}^n_+$ such that

$$\nabla f(x^k) + \frac{1}{\varepsilon_k}\phi'(\Delta_k)\nabla F(x^k)(F(x^k) - \varepsilon_k\omega) = \lambda^k - \eta^k, \tag{3.7}$$

$$\inf(\lambda_i^k, x_i^k - u_i) = \inf(\eta_i^k, v_i - x_i^k) = 0, \quad i = 1, 2, \ldots, n. \tag{3.8}$$

Since the index set $\{1, \ldots, n\}$ is finite, then there exists an infinite subset $K \subset \{1, 2, \ldots\}$ such that for $k \in K$,

$$x_i^k = u_i, \qquad i \in I_1, \tag{3.9}$$

$$x_i^k = v_i, \qquad i \in I_2, \tag{3.10}$$

$$u_i < x_i^k < v_i, \qquad i \in I_3, \tag{3.11}$$

where

$$I_1 \cup I_2 \cup I_3 = \{1, 2, \ldots, n\} \text{ and } I_i \cap I_j = \emptyset \text{ if } i \neq j.$$

Invoking Assumption (A_2), there exist an infinite subset $K_0 \subset K$ and $p \in \mathbb{R}^n$ such that

$$\lim_{k \in K_0, k \to \infty} \nabla F(x^k) = \nabla F^*, \ \ \text{rank}(\nabla F^*) = m, \ \ (\nabla F^*)^T p = 0, \tag{3.12}$$

$$p_i \begin{cases} > 0, & i \in I_1, \\ < 0, & i \in I_2. \end{cases} \tag{3.13}$$

Putting (3.7)–(3.11) together yields

$$\frac{\partial f(x^k)}{\partial x_i} + \frac{1}{\varepsilon_k}\phi'(\triangle_k)\Big(\nabla F(x^k)(F(x^k) - \varepsilon_k\omega)\Big)_i \begin{cases} \geq 0, & i \in I_1, \\ \leq 0, & i \in I_2, \\ = 0, & i \in I_3, \end{cases} \tag{3.14}$$

where $\partial f(x)/\partial x_i$ denotes the partial derivative of f with respect to x_i. Let

$$h_k = \frac{1}{\varepsilon_k}\phi'(\Delta_k)(F(x^k) - \varepsilon_k\omega) \text{ and } q_k = \frac{h_k}{\|h_k\|}.$$

Lemma 3.3 implies

$$\lim_{k \in K_0, k \to \infty} \|h_k\| = +\infty. \tag{3.15}$$

Since $\|q_k\| = 1$ is bounded, we can assume, without loss of generality, that

$$\lim_{k \in K_0, k \to \infty} q_k = \tilde{q} \neq 0. \tag{3.16}$$

It then follows from (3.14) that

$$\frac{\partial f(x^k)}{\partial x_i}\frac{1}{\|h_k\|} + \Big(\nabla F(x^k) q_k\Big)_i \begin{cases} \geq 0, & i \in I_1, \\ \leq 0, & i \in I_2, \\ = 0, & i \in I_3. \end{cases}$$

From Assumption (A_1), (3.12), (3.15), and (3.16), taking limits in the above formula yields

$$\Big((\nabla F^*)\tilde{q}\Big)_i \begin{cases} \geq 0, & i \in I_1, \\ \leq 0, & i \in I_2, \\ = 0, & i \in I_3. \end{cases} \tag{3.17}$$

Combining this with (3.12) implies

$$0 = p^T(\nabla F^*)\tilde{q} = \sum_{i\in I_1} p_i\Big((\nabla F^*)\tilde{q}\Big)_i + \sum_{i\in I_2} p_i\Big((\nabla F^*)\tilde{q}\Big)_i,$$

which, together with (3.13) and (3.17) yields $(\nabla F^*)\tilde{q} = 0$, and hence $\tilde{q} = 0$ since ∇F^* has full column rank by (3.12). This leads to a contradiction to $\tilde{q} \neq 0$ in (3.16). Therefore, the desired result readily follows from Lemma 3.2. □

As a corollary of Theorem 3.1, we have

Corollary 3.1 *Suppose that*

(1) *the set D_F is bounded,*
(2) *the standard MFCQ holds at each point of D_F,*
(3) *there exists $\beta_1 > 0$ such that $\beta'(\varepsilon) \geq \beta_1$ for all $\varepsilon \in (0, \bar{\varepsilon}]$.*

Then, whenever $\sigma > 0$ is large enough, every local optimal solution (x^*, ε^*) of $(\widetilde{P}_\sigma)$ with finite value satisfies $\varepsilon^* = 0$. Furthermore, x^* is a local optimal solution of (P).

Proof The validity of Assumption (A_1) comes from the condition (1), and Assumption (A_2) is due to the conditions (1), (2), and Proposition 2.1. □

Remark 3.1 In particular, Corollary 3.1 recovers [13, Theorem 2.1] by taking $\phi(t) = \frac{t}{1-qt}$.

Inspired by [13], a necessary condition for exact penalty property is given below. Toward this end, consider a class of functions $\Phi : [0, a) \to [0, +\infty)$, where $a > 0$, satisfying

(j_1) Φ is continuous and increasing on $[0, a)$ with $\Phi(0) = 0$,
(j_2) there exists $a' \in (0, a)$ such that $\Phi(t) \geq t$ for all $t \in [0, a']$.

Such a class of functions includes several important functions as special cases, for example

$$\begin{array}{lll} \Phi_1(t) = t^\alpha, & t \in [0, +\infty), & \alpha \in (0, 1], \\ \Phi_2(t) = -\log(1 - t^\alpha), & t \in [0, 1), & \alpha \in (0, 1], \\ \Phi_3(t) = e^t - 1, & t \in [0, +\infty). & \end{array}$$

Assumption (B) Let x^* be a feasible point of (P). There exist $\gamma > 0$ and a neighborhood $N(x^*)$ such that

$$f(x) - f(x^*) + \gamma\Phi(\|F(x)\|) \geq 0, \quad \forall x \in N(x^*) \cap [u, v], \tag{3.18}$$

where Φ satisfies the conditions (j_1) and (j_2).

The function Φ plays a key role in (3.18), which is illustrated by the following example.

Example 3.1

$$\begin{aligned} \min \quad & f(x) = x \\ \text{s.t.} \quad & F(x) = x^2 = 0, \\ & x \in (-\infty, +\infty). \end{aligned}$$

Clearly, the optimal solution is $x^* = 0$. If $\Phi(t) = t$, i.e., $\Phi(\|F(x)\|) = x^2$, then (3.18) is false for all $\gamma > 0$, while if $\Phi(t) = \sqrt{t}$, i.e., $\Phi(\|F(x)\|) = |x|$, then (3.18) is true by taking $\gamma = 1$.

Theorem 3.2 *Suppose that $\beta(\varepsilon) = \Phi(\varepsilon)$ for $\varepsilon > 0$ sufficiently small. If $(x^*, 0)$ is a local optimal solution of $(\widetilde{P}_{\sigma_0})$ with finite value for some $\sigma_0 > 0$, then Assumption (B) holds at x^* for Φ with $\gamma \geq \sigma_0 + \phi'(0)(1 + \|\omega\|)^2$.*

Proof Let $\gamma \geq \sigma_0 + \phi'(0)(1 + \|\omega\|)^2$. Suppose on the contrary that there exists a sequence $x^k \in [u, v]$ converging to x^* such that

$$f(x^k) - f(x^*) \leq f(x^k) - f(x^*) + \gamma \Phi(\|F(x^k)\|) < 0. \tag{3.19}$$

Since $(x^*, 0)$ is a local optimal solution of $(\widetilde{P}_{\sigma_0})$ and the corresponding optimal value is finite, it follows from the construction of $\widetilde{f}_\sigma$ that x^* is feasible (cf.(2.1)) and thus x^* is a local optimal solution of (P). Therefore, x_k is infeasible by (3.19). This, together with the continuity of F, implies that

$$\|F(x^k)\| > 0 \quad \text{and} \quad \lim_{k\to\infty} \|F(x^k)\| = \|F(x^*)\| = 0.$$

Let $\varepsilon_k = \|F(x^k)\|$, i.e., $\varepsilon_k > 0$ and $\varepsilon_k \to 0$ as $k \to \infty$. Notice that

$$\Delta_k = \|F(x^k) - \varepsilon_k \omega\|^2 \leq \varepsilon_k^2 (1 + \|\omega\|)^2. \tag{3.20}$$

Thus,

$$\lim_{k\to\infty} \Delta_k \leq \lim_{k\to\infty} \varepsilon_k^2 (1 + \|\omega\|)^2 = 0.$$

As a result,

$$\phi'(\Delta_k) \leq 2\phi'(0), \tag{3.21}$$

whenever k is sufficiently large, since ϕ' is continuous and $\phi'(0) > 0$. Since $(x^*, 0)$ is a local optimal solution of $(\widetilde{P}_{\sigma_0})$, then for k large enough we have

$$\begin{aligned}
0 &\leq f(x^k) - f(x^*) + \frac{1}{2\varepsilon_k}\phi(\Delta_k) + \sigma_0 \beta(\varepsilon_k) \\
&= f(x^k) - f(x^*) + \tfrac{1}{2\varepsilon_k}\phi(\Delta_k) + \sigma_0 \Phi(\varepsilon_k) && \textit{by Assumption } \beta(\varepsilon) = \Phi(\varepsilon) \\
&< \tfrac{1}{2\varepsilon_k}\phi(\Delta_k) + \sigma_0 \Phi(\varepsilon_k) - \gamma \Phi(\|F(x^k)\|) && \textit{by } (3.19) \\
&\leq \tfrac{1}{2\varepsilon_k}\phi'(\Delta_k)\Delta_k + \sigma_0 \Phi(\varepsilon_k) - \gamma \Phi(\varepsilon_k), && \textit{by the convexity of } \phi \textit{ and } \phi(0) = 0 \\
&\leq \varepsilon_k \phi'(0)(1 + \|\omega\|)^2 + (\sigma_0 - \gamma)\Phi(\varepsilon_k) && \textit{by } (3.20) \textit{ and } (3.21) \\
&\leq \Phi(\varepsilon_k)\Big[\phi'(0)(1 + \|\omega\|)^2 + \sigma_0 - \gamma\Big], && \textit{by condition } j_2 \\
&\leq 0, && \textit{since } \gamma \geq \sigma_0 + \phi'(0)(1 + \|\omega\|)^2
\end{aligned}$$

which leads to a contradiction. □

4 Inverse propositions for exact penalization

In this section, we shall show that, if x^* is a local optimal solution of (P), then Assumption (B) introduced in Sect. 3 is a necessary and sufficient condition for $(x^*, 0)$ to be a local optimal solution of $(\widetilde{P}_\sigma)$. Based on Theorem 3.2, we further establish the equivalence between this new class of exact penalty functions and the classical simple and exact penalty functions in the sense of "exact penalty property". Our conclusions clarify that this class of penalty functions (2.1) possesses not only exactness property as the classical simple penalty function, but also the smoothness property, which is not shared by the classical simple exact penalty function, however.

Theorem 4.1 *Let x^* be a local optimal solution of (P). The following two statements hold:*

(1) *If Assumption (B) holds at x^* for Φ, and $\beta(\varepsilon) \geq \Phi(\sqrt{\varepsilon})$ for $\varepsilon > 0$ sufficiently small, then $(x^*, 0)$ is a local optimal solution of $(\widetilde{P}_\sigma)$ for all $\sigma \geq \gamma$;*
(2) *Let $\beta(\varepsilon) = \Phi(\varepsilon)$ for $\varepsilon > 0$ sufficiently small. If there exists $\sigma_0 > 0$ such that $(x^*, 0)$ is a local optimal solution of $(\widetilde{P}_\sigma)$ for all $\sigma \geq \sigma_0$, then Assumption (B) holds at x^* for Φ with $\gamma \geq \sigma_0 + \phi'(0)(1 + \|\omega\|)^2$.*

Proof We only need to show the validity of part (1), since part (2) follows from Theorem 3.2. Suppose on the contrary that for some $\sigma \geq \gamma$, there exist $(x^k, \varepsilon_k) \in [u, v] \times (0, \overline{\varepsilon}]$, $x^k \to x^*$, and $\varepsilon_k \to 0$ as $k \to \infty$ such that $f(x^*) = \widetilde{f}_\sigma(x^*, 0) > \widetilde{f}_\sigma(x^k, \varepsilon_k)$,, i.e.,

$$\begin{aligned} 0 &> f(x^k) - f(x^*) + \frac{1}{2\varepsilon_k}\phi(\Delta_k) + \sigma\beta(\varepsilon_k) \\ &\geq f(x^k) - f(x^*) + \frac{1}{2\varepsilon_k}\phi'(0)\Delta_k + \sigma\beta(\varepsilon_k), \end{aligned} \tag{4.1}$$

where we have used the gradient inequality of convex functions, i.e., $\phi(\Delta_k) \geq \phi(0) + \phi'(0)\Delta_k = \phi'(0)\Delta_k$. Taking limits on both sides of (4.1) yields $\lim_{k\to\infty} \Delta_k/\varepsilon_k = 0$, which means that

$$\sqrt{\Delta_k} \leq \frac{1}{2}\sqrt{\varepsilon_k} \quad \text{and} \quad \sqrt{\varepsilon_k} \leq \frac{1}{2\|\omega\| + 1}$$

whenever k is sufficiently large. So,

$$\|F(x^k)\| \leq \sqrt{\Delta_k} + \varepsilon_k\|\omega\| \leq \left(\frac{1}{2} + \sqrt{\varepsilon_k}\|\omega\|\right)\sqrt{\varepsilon_k} \leq \sqrt{\varepsilon_k}. \tag{4.2}$$

Since $\varepsilon_k \to 0$, then by hypothesis, for k large enough,

$$\beta(\varepsilon_k) \geq \Phi(\sqrt{\varepsilon_k}). \tag{4.3}$$

Therefore, we conclude that for k sufficiently large,

$$\begin{aligned} 0 &> f(x^k) - f(x^*) + \frac{1}{2\varepsilon_k}\phi'(0)\Delta_k + \sigma\beta(\varepsilon_k) \\ &\geq f(x^k) - f(x^*) + \sigma\beta(\varepsilon_k) && \textit{by the nonnegativity of } \phi' \\ &\geq \sigma\beta(\varepsilon_k) - \gamma\Phi(\|F(x^k)\|) && \textit{by Assumption}(B) \\ &\geq \sigma\Phi(\sqrt{\varepsilon_k}) - \gamma\Phi(\|F(x^k)\|) && \textit{by } (4.3) \\ &\geq \Phi(\sqrt{\varepsilon_k})(\sigma - \gamma) && \textit{by the monotonicity of } \Phi \textit{ and } (4.4) \\ &\geq 0, && \textit{since } \sigma \geq \gamma \end{aligned}$$

which leads to a contradiction. □

The concept of regular zero is introduced in [13, Definition 4.1]. To be consistent with their notation, let us define $H : \mathbb{R}^n \to \mathbb{R}^p$ with $p = m + |I^*| \le n$ and

$$H(x) = \begin{pmatrix} F(x) \\ x_{I^*} - x^*_{I^*} \end{pmatrix},$$

where $|I^*|$ signifies the number of elements in $I^* = I_1^* \cup I_2^*$, I_1^* and I_2^* are, respectively, the active index sets of $u \le x^*$ and $x^* \le v$, i.e.,

$$I_1^* = \{i | u_i = x_i^*\} \text{ and } I_2^* = \{i | v_i = x_i^*\}.$$

According to Theorem 4.1, we obtain

Corollary 4.1 *Let x^* be a local optimal solution of (P) and $\beta(\varepsilon) \ge \sqrt{\varepsilon}$ for all ε sufficiently small. If x^* is a regular zero of H, then there exists $\gamma > 0$ such that $(x^*, 0)$ is a local optimal solution of $(\widetilde{P}_\sigma)$ for all $\sigma \ge \gamma$.*

Proof Since x^* is a regular zero of H, it then follows from [13, Lemmas 5.1 and 5.2] that Assumption (B) holds at x^* for $\Phi(t) = t$. Combining this and Theorem 4.1 yields the desired result. □

Remark 4.1 For the smooth case, Corollary 4.1 reduces to [13, Theorem 5.3] by taking $\phi(t) = \frac{t}{1-qt}$.

Corollary 4.2 *Let x^* be a strict local optimal solution of (P) and $\beta(\varepsilon) \ge \sqrt{\varepsilon}$ for $\varepsilon > 0$ sufficiently small. If MFCQ holds at x^*, then there exists $\gamma > 0$ such that $(x^*, 0)$ is a local optimal solution of $(\widetilde{P}_\sigma)$ for all $\sigma \ge \gamma$.*

Proof Since MFCQ holds at x^*, it then follows from [10, Theorem 4.4] that Assumption (B) holds at x^* for $\Phi(t) = t$. Therefore, the desired result follows from Theorem 4.1. □

Actually, these two sufficient conditions given in Corollaries 4.1 and 4.2 are independent. This is illustrated by the following examples.

Example 4.1

$$F(x) = x_1 - x_2 = 0,$$
$$x = (x_1, x_2) \in [0, +\infty) \times [0, \infty).$$

It is easy to see that MFCQ holds at $x^* = (0, 0)$, while x^* is not a regular zero of H.

Now let us show that the second-order sufficient conditions guarantee the validity of Assumption (B). Recall that, for (P), the second-order sufficient conditions are said to hold at x^* if

(a) x^* is a KKT point, i.e., there exist $\lambda^*, \eta^* \in \mathbb{R}^n_+$, and $\mu^* \in \mathbb{R}^m$ such that

$$\nabla f(x^*) - \lambda^* + \eta^* + \nabla F(x^*)\mu^* = 0, \tag{4.4}$$
$$\inf(\lambda_i^*, x_i^* - u_i) = \inf(\eta_i^*, v_i - x_i^*) = 0, \quad i = 1, 2, \ldots, n. \tag{4.5}$$

(b) the matrix $\nabla^2_{xx} L(x^*, \mu^*)$ is positive definite on the cone $\{d \ne 0 | \nabla F(x^*)d = 0, d_i = 0, \ \textit{as}\ \lambda_i^* > 0 \textit{ or } \eta_i^* > 0\}$, where $L(x, \mu) = f(x) + \mu^T F(x)$.

The following example shows that the second-order sufficient conditions are independent of MFCQ and regular zero of H.

Example 4.2

$$\begin{aligned} \min \quad & f(x) = x_1^2 + x_2^2 \\ \text{s.t.} \quad & F(x) = x_1 - x_2 = 0, \\ & x = (x_1, x_2) \in [0, +\infty) \times (-\infty, 0]. \end{aligned}$$

It is easy to check that at the point $x^* = (0, 0)$, the second-order sufficient conditions hold, while MFCQ and regular zero of H do not hold true.

Corollary 4.3 *If $\beta(\varepsilon) \geq \sqrt{\varepsilon}$ for all $\varepsilon > 0$ sufficiently small and the second-order sufficient conditions hold at x^*, then there exists $\gamma > 0$ such that $(x^*, 0)$ is a local strict optimal solution of $(\widetilde{P}_\sigma)$ for all $\sigma \geq \gamma$.*

Proof Since the second-order sufficient conditions hold at x^*, it then readily follows from [10, Theorem 4.6] that Assumption (B) is valid for $\Phi(t) = t$ with the inequality being strict. The desired result follows from Theorem 4.1. □

Now we propose a new sufficient condition for the validity of Assumption (B) from a different way.

Proposition 4.1 *Let x^* be a local optimal solution of (P). If x^* is a KKT point and*

$$\lim_{\substack{x \in [u,v] \to x^* \\ F(x) \neq 0}} \frac{1}{\Phi(\|F(x)\|)} \int_0^1 \Big(\nabla_x L(x^* + s(x - x^*), \mu^*) - \nabla_x L(x^*, \mu^*) \Big)^T (x - x^*) ds = 0, \tag{4.6}$$

where $\lambda^*, \eta^* \in \mathbb{R}^n_+$, $\mu^* \in \mathbb{R}^m$ are the corresponding Lagrangian multipliers, then Assumption (B) holds at x^* for Φ with $\gamma = 1 + \|\mu^*\|$.

Proof It is sufficient to show the existence of a neighborhood of x^*, say $N(x^*)$, such that

$$f(x) - f(x^*) + \gamma \Phi(\|F(x)\|) \geq 0, \quad x \in N(x^*) \cap [u, v]. \tag{4.7}$$

Since x^* is a local optimal solution of (P), then there exists a neighborhood of x^*, say $\widetilde{N}(x^*)$, such that (4.7) holds true for all $x \in \widetilde{N}(x^*) \cap [u, v]$ satisfying $F(x) = 0$. Now consider the case of $x \in \widetilde{N}(x^*) \cap [u, v]$ with $F(x) \neq 0$. Putting (4.4), (4.5), and (4.6) together yields

$$\begin{aligned} f(x) - f(x^*) &= \nabla f(x^*)^T (x - x^*) + \int_0^1 \Big(\nabla f(x^* + s(x - x^*)) - \nabla f(x^*) \Big)^T (x - x^*) ds \\ &= \sum_{i \in I_1^*} \lambda_i^* (x_i - x_i^*) - \sum_{i \in I_2^*} \eta_i^* (x_i - x_i^*) - \mu^{*T} \nabla F(x^*)^T (x - x^*) \\ &\quad + \int_0^1 \Big(\nabla f(x^* + s(x - x^*)) - \nabla f(x^*) \Big)^T (x - x^*) ds \\ &\geq -\mu^{*T} \nabla F(x^*)^T (x - x^*) + \int_0^1 \Big(\nabla f(x^* + s(x - x^*)) - \nabla f(x^*) \Big)^T (x - x^*) ds \end{aligned}$$

$$= -\mu^{*T}F(x) + \int_0^1 \mu^{*T}\Big(\nabla F(x^* + s(x-x^*)) - \nabla F(x^*)\Big)^T (x-x^*)ds$$

$$+ \int_0^1 \Big(\nabla f(x^* + s(x-x^*)) - \nabla f(x^*)\Big)^T (x-x^*)ds$$

$$= -\mu^{*T}F(x) + \int_0^1 \Big(\nabla_x L(x^* + s(x-x^*), \mu^*) - \nabla_x L(x^*, \mu^*)\Big)^T (x-x^*)ds$$

$$= -\mu^{*T}F(x) + o(\Phi(\|F(x)\|)). \tag{4.8}$$

Note that there exists a neighborhood $N(x^*) \subset \widetilde{N}(x^*)$ of x^* such that

$$\frac{1}{\Phi(\|F(x)\|)}\Big|o\Big(\Phi(\|F(x)\|)\Big)\Big| \le \frac{1}{2}, \tag{4.9}$$

whenever $x \in N(x^*) \cap [u, v]$ with $F(x) \neq 0$. Therefore,

$$\begin{aligned} f(x) - f(x^*) + \gamma\Phi(\|F(x)\|) &\ge \gamma\Phi(\|F(x)\|) - \mu^{*T}F(x) + o\big(\Phi(\|F(x)\|)\big) && \textit{by } (4.8)\\ &\ge \gamma\Phi(\|F(x)\|) - \|\mu^*\|\|F(x)\| + o\big(\Phi(\|F(x)\|)\big)\\ &\ge \Phi(\|F(x)\|)(\gamma - \|\mu^*\|) + o\big(\Phi(\|F(x)\|)\big) && \textit{by } \Phi(t) \ge t\\ &= \Phi(\|F(x)\|) + o\big(\Phi(\|F(x)\|)\big) && \textit{by } \gamma = 1 + \|\mu^*\|\\ &\ge \frac{1}{2}\Phi(\|F(x)\|) && \textit{by } (4.9)\\ &> 0. \end{aligned}$$

This yields the inequality as desired. □

It should be emphasized that the condition (4.6) is also independent of other conditions given in the previous discussion, which is illustrated by the following example.

Example 4.3

$$\begin{aligned} \min\ & f(x) = -|x|^{\frac{3}{2}}\\ \text{s.t.}\ & F(x) = x^2 = 0,\\ & x \in (-\infty, +\infty). \end{aligned}$$

By a simple calculation, we know that, at $x^* = 0$, the condition (4.6) holds when $\Phi(t) = \sqrt{t}$, while MFCQ, the second-order sufficient conditions, and regular zero of F do not hold at x^*. In addition, the condition (4.6) is also true for Example 4.2 when $\Phi(t) = t$.

Applying Theorem 4.1 and Proposition 4.1 yields the following corollary.

Corollary 4.4 *Let x^* be a local optimal solution of (P) and $\beta(\varepsilon) \ge \Phi(\sqrt{\varepsilon})$ for all $\varepsilon > 0$ sufficiently small. If x^* is a KKT point and the condition (4.6) holds, then $(x^*, 0)$ is a local optimal solution of $(\widetilde{P}_\sigma)$ with $\sigma \ge 1 + \|\mu^*\|$.*

We turn our attention to the relationship between this new class of exact penalty functions and the classical simple penalty functions in the sense of "exactness" property. Define

$$f_\gamma(x) = f(x) + \gamma\Phi(\|F(x)\|),$$

where Φ satisfies (j_1) and (j_2). Clearly, f_γ is a classical simple and exact penalty function. The associated penalty problem is

$$(P_\gamma) \quad \min \ f_\gamma(x) \quad \text{s.t.}\ x \in [u, v].$$

Note that (3.18) in Assumption (B) is equivalent to saying that x^* is a local optimal solution of (P_γ). Let x^* be a local optimal solution of (P). According to Theorem 4.1, the relationship between $(\widetilde{P}_\sigma)$ and (P_γ) is summarized as follows:

(1) Suppose that $\beta(\varepsilon) \geq \Phi(\sqrt{\varepsilon})$ for $\varepsilon > 0$ sufficiently small. If there exists $\gamma > 0$ such that x^* is a local optimal solution of the penalty problem (P_γ), then $(x^*, 0)$ is a local optimal solution of the penalty problem $(\tilde{P}_\sigma)$ when $\sigma > 0$ is sufficiently large.
(2) Suppose that $\beta(\varepsilon) = \Phi(\varepsilon)$ for $\varepsilon > 0$ sufficiently small. If there exists $\sigma > 0$ such that $(x^*, 0)$ is a local optimal solution of the penalty problem $(\tilde{P}_\sigma)$, then x^* is a local optimal solution of the penalty problem (P_γ) when $\gamma > 0$ is sufficiently large.

5 Penalty function methods

In this section, we present a revised penalty function algorithm via $\tilde{f}_\sigma$. The algorithm is always feasible, since $\tilde{f}_\sigma$ is bounded from below. It should be noted that the parameter ε in $\tilde{f}_\sigma$ plays a key role in the framework of the algorithm. It follows from the Assumption (A) and Theorem 3.1 that the proposed algorithm terminates at the optimal solutions of (P) after finitely many iterations (see Theorem 5.1). We further analyze the case in the absence of Assumption (A). In this case, we present a perturbation theorem about the algorithm (see Theorem 5.2). This allows us to obtain the global convergence property of the algorithm (see Corollary 5.1). Furthermore, necessary and sufficient conditions for zero duality gap are also derived between the original problem (P) and its dual problem (L) defined by the penalty function $\tilde{f}_\sigma$ (see Corollary 5.2).

Assumption (C) The function β satisfies $\inf_{\varepsilon \geq \varepsilon_0} \beta(\varepsilon) > 0$ for all $\varepsilon_0 > 0$.

Indeed, the function β given in the previous discussion satisfies this assumption.

Algorithm 5.1 Let $\alpha \in (0, 1)$ be sufficiently small number, $\delta_0 > 0$, $\sigma_0 \geq 1$, and $k := 0$;
Step 1. If $argmin(\widetilde{P}_{\sigma_k}) \neq \emptyset$, solve

$$(x^k, \varepsilon_k) \in argmin(\widetilde{P}_{\sigma_k})$$

and go to Step 2. Otherwise, solve

$$(x^k, \varepsilon_k) \in \delta_k - argmin(\widetilde{P}_{\sigma_k}),$$

i.e.,

$$\widetilde{f}_\sigma(x^k, \varepsilon_k) \leq \inf\{\widetilde{f}_\sigma(x, \varepsilon) | (x, \varepsilon) \in [u, v] \times [0, \bar{\varepsilon}]\} + \delta_k,$$

and go to Step 3.
Step 2. If $\varepsilon_k = 0$, stop; Otherwise, go to Step 4.
Step 3. If $\varepsilon_k = 0$ and $\delta_k \leq \alpha$, stop; Otherwise, go to Step 4.
Step 4. Let

$$\delta_{k+1} = \frac{1}{2}\delta_k \ \text{ and } \ \sigma_{k+1} = \begin{cases} \sigma_k, & \text{if } \varepsilon_k = 0, \\ \rho\sigma_k, & \text{otherwise;} \end{cases}$$

where $\rho > 1$ is a constant.
Step 5. Let $k := k + 1$ and go back to step 1.

This algorithm is always feasible, since the existence of δ-optimal solution is ensured by the lower-boundedness of $\widetilde{f}_\sigma$ over $D \times [0, \overline{\varepsilon}]$. In addition, due to the smoothness of $\widetilde{f}_{\sigma_k}(\cdot, \varepsilon)$ on D_F, many descent algorithms can be used to find (x^k, ε_k). Given $\eta \geq 0$, define a perturbation function of (P) as follows

$$\theta(\eta) = \inf\{f(x)|x \in [u, v], \|F(x)\| \leq \eta\}.$$

Clearly, θ is a nonincreasing function with $\theta(0)$ equal to the optimal value of (P). Since $f(x)$ is lower-bounded on $[u, v]$, then the limit of θ as η approaches to 0^+ exists and

$$\lim_{\eta \to 0^+} \theta(\eta) \leq \theta(0), \tag{5.1}$$

i.e., θ is upper semicontinuous at zero.

Theorem 5.1 *Under the assumptions of Theorem 3.1, let* $\{(x^k, \varepsilon_k)\}$ *be a sequence generated by Algorithm 5.1 and* $(x^k, \varepsilon_k) \in argmin(\widetilde{P}_{\sigma_k})$. *Then there exists* k_0 *such that when* $\sigma_k \geq \sigma_{k_0}, \varepsilon_k = 0$, *i.e.,* x^k *is an optimal solution of* (P).

Proof From Theorem 3.1, there exists k_0 such that when $\sigma_k \geq \sigma_{k_0}, \varepsilon_k = 0$. It readily follows from the finiteness of $\tilde{f}_{\sigma_k}(x^k, 0)$ and the definition of $\tilde{f}_\sigma$ that $F(x^k) = 0$, i.e., $\tilde{f}_{\sigma_k}(x^k, 0) = f(x^k)$. Note that $(x^k, 0) \in argmin(\widetilde{P}_{\sigma_k})$ yields that $x^k \in argmin(P)$. □

Lemma 5.1 *Let* β *satisfy Assumption* (C), *and* $\{(x^k, \varepsilon_k)\}$ *be an infinite sequence generated by Algorithm 5.1. Then*

(1) $\exists k_0 > 0$ *such that* $\varepsilon_k > 0$ *for all* $k \geq k_0$;
(2) $\lim_{k \to \infty} \varepsilon_k = 0$;
(3) $\lim_{k \to \infty} F(x^k) = 0$.

Proof Note that the condition $\delta_k \leq \alpha$ is always true as k large enough, since $\delta_{k+1} = \delta_k/2$ by Step 4. This, according to Step 2 and Step 3, means that the termination condition is $\varepsilon_k = 0$. In other words, the algorithm will generate an infinite sequence only if $\varepsilon_k > 0$ as k sufficiently large. By the iteration strategy on σ_k in Step 4, we have

$$\lim_{k \to \infty} \sigma_k = \infty. \tag{5.2}$$

From Step 1, we have

$$\begin{aligned} f(x^k) + \frac{1}{2\varepsilon_k}\phi(\Delta_k) + \sigma_k\beta(\varepsilon_k) &\leq \inf\{\widetilde{f}_{\sigma_k}(x, \varepsilon)|(x, \varepsilon) \in [u, v] \times [0, \overline{\varepsilon}]\} + \delta_k \\ &\leq f(\overline{x}) + \delta_k, \end{aligned} \tag{5.3}$$

where $\overline{x}$ is a feasible point of (P). Note that f is bounded from below on D and ϕ is nonnegative. It immediately follows from (5.2) and (5.3) that $\lim_{k \to \infty} \beta(\varepsilon_k) = 0$, which, together with Assumption (C), further implies that

$$\lim_{k \to \infty} \varepsilon_k = 0. \tag{5.4}$$

Similarly, according to the nonnegativity of β, we get from (5.3) and (5.4) that $\lim_{k \to \infty} \phi(\Delta_k) = 0$, and hence

$$\lim_{k \to \infty} \Delta_k = 0, \tag{5.5}$$

where we have used the gradient inequality of convex function ϕ, i.e., $\phi(\Delta_k) \geq \phi'(0)\Delta_k \geq 0$. Since $\|F(x^k)\| \leq \sqrt{\Delta_k} + \varepsilon_k\|\omega\|$, putting (5.4) and (5.5) together yields $\lim_{k\to\infty} \|F(x^k)\| = 0$ as claimed. □

The global convergence property of Algorithm 5.1 is given below.

Theorem 5.2 *Let β satisfy Assumption (C). The following statements hold:*

(1) *Assume Algorithm 5.1 stops after finitely many iterations, then x^k is an either optimal solution or α-optimal solution of (P).*
(2) *If an infinite sequence is generated, then*

$$\lim_{k\to\infty} f(x^k) = \lim_{\eta\to 0^+} \theta(\eta),$$

and

$$\lim_{k\to\infty} \inf\{\widetilde{f}_{\sigma_k}(x,\varepsilon) \mid (x,\varepsilon) \in [u,v] \times [0,\overline{\varepsilon}]\} = \lim_{\eta\to 0^+} \theta(\eta).$$

Proof (1). Suppose the algorithm stops at the k-th iteration, then $\varepsilon_k = 0$. Since $\widetilde{f}_{\sigma_k}(x^k, 0)$ is finite, then $F(x^k) = 0$, and hence $\widetilde{f}_{\sigma_k}(x^k, 0) = f(x^k)$ by the definition of $\widetilde{f}_\sigma$. Therefore, according to Step 1, we have $(x^k, 0) \in argmin(\widetilde{P}_{\sigma_k})$ or $(x^k, 0) \in \delta_k - argmin(\widetilde{P}_{\sigma_k})$. This together with $\widetilde{f}_{\sigma_k}(x^k, 0) = f(x^k)$ and $\delta_k \leq \alpha$ yields $x^k \in argmin(P)$ or $x^k \in \alpha - argmin(P)$.

(2). If the algorithm does not stop after finitely many iterations, then we have $\lim_{k\to\infty} \sigma_k = +\infty$. By Lemma 5.1, there exists a k_0 such that $\varepsilon_k > 0$ for all $k \geq k_0$. Because $\widetilde{f}_{\sigma_k}(x^k, \varepsilon_k)$ is finite, then

$$\widetilde{f}_{\sigma_k}(x^k, \varepsilon_k) = f(x^k) + \frac{1}{2\varepsilon_k}\phi(\Delta_k) + \sigma_k\beta(\varepsilon_k).$$

The continuity of β and the fact $\beta(0) = 0$ guarantee the existence of $\overline{\varepsilon}_k > 0$ such that $\lim_{k\to\infty} \overline{\varepsilon}_k = 0$ and

$$\lim_{k\to\infty} \sigma_k\beta(\overline{\varepsilon}_k) = 0. \tag{5.6}$$

Indeed, this is ensured by choosing $\overline{\varepsilon}_k$ such that $\beta(\overline{\varepsilon}_k) \leq 1/\sigma_k^2$. Choose another sequence $\{\xi_k\}$ with $\xi_k > 0$ and $\lim_{k\to\infty} \xi_k = 0$. According to the definition of infimum in θ, there exist $y^k \in [u, v]$ such that $\|F(y^k)\| \leq \overline{\varepsilon}_k$, and

$$f(y^k) \leq \theta(\overline{\varepsilon}_k) + \xi_k. \tag{5.7}$$

Let $\overline{\Delta}_k = \|F(y^k) - \overline{\varepsilon}_k\omega\|^2$. Since $\|F(y^k)\| \leq \overline{\varepsilon}_k$, then

$$\overline{\Delta}_k \leq \overline{\varepsilon}_k^2(1 + \|\omega\|)^2, \tag{5.8}$$

from which and the fact $\lim_{k\to\infty} \overline{\varepsilon}_k = 0$, we get $\overline{\Delta}_k < a$ for all k sufficiently large. Thus, $f_{\sigma_k}(y^k, \overline{\varepsilon}_k)$ is finite by (2.1). Lemma 5.1 implies that, for any $\eta > 0$, we have $\|F(x^k)\| \leq \eta$ as k large enough. Hence, $\theta(\eta) \leq f(x^k)$ holds by the definition of θ. Therefore,

$$\begin{aligned}
\theta(\eta) &\le f(x^k)\\
&\le f(x^k) + \frac{1}{2\varepsilon_k}\phi(\Delta_k) + \sigma_k\beta(\varepsilon_k)\\
&= \widetilde{f}_{\sigma_k}(x^k, \varepsilon_k)\\
&\le \inf\{\widetilde{f}_{\sigma_k}(x, \varepsilon) | (x, \varepsilon) \in [u, v] \times [0, \overline{\varepsilon}]\} + \delta_k\\
&\le \widetilde{f}_{\sigma_k}(y^k, \overline{\varepsilon}_k) + \delta_k\\
&= f(y^k) + \frac{1}{2\overline{\varepsilon}_k}\phi(\overline{\Delta}_k) + \sigma_k\beta(\overline{\varepsilon}_k) + \delta_k\\
&\le \theta(\overline{\varepsilon}_k) + \xi_k + \frac{1}{2\overline{\varepsilon}_k}\phi'(\overline{\Delta}_k)\overline{\Delta}_k + \sigma_k\beta(\overline{\varepsilon}_k) + \delta_k\\
&\le \theta(\overline{\varepsilon}_k) + \xi_k + \frac{1}{2}\phi'(\overline{\Delta}_k)\overline{\varepsilon}_k(1 + \|\omega\|)^2 + \sigma_k\beta(\overline{\varepsilon}_k) + \delta_k,
\end{aligned}$$

where the fifth inequality comes from (5.7) and the convexity of ϕ as before, and the last inequality is due to (5.8). Using (5.6) and the arbitrariness of $\eta > 0$, we get the desired result by taking limits on both sides of above inequality. □

Corollary 5.1 *Let β satisfy Assumption (C), and $\{(x^k, \varepsilon_k)\}$ be an infinite sequence generated by Algorithm 5.2. The following statements hold.*

(1) $\lim_{k\to\infty} f(x^k) = \theta(0)$ *if and only if θ is lower semicontinuous at zero.*
(2) *If x^* is an accumulation point of $\{x^k\}$, then x^* is a global optimal solution of (P).*

Proof (1). This follows directly from Part 2) of Theorem 5.2.
(2). If x^* is an accumulation point, then $F(x^*) = 0$ by Lemma 5.1, i.e., x^* is feasible. According to Part 2) in Theorem 5.2, we get from (5.1) that

$$\theta(0) \ge \lim_{\eta\to 0^+} \theta(\eta) = \lim_{k\to\infty} f(x^k) = f(x^*) \ge \theta(0),$$

which implies $f(x^*) = \theta(0)$, i.e., x^* is a global optimal solution of (P). □

Using the penalty function $\widetilde{f}_\sigma$, define the associated dual optimization problem

$$(L) \quad \max l(\sigma) \quad \text{s.t. } \sigma \ge 0,$$

where $l(\sigma) = \inf\{\widetilde{f}_\sigma(x, \varepsilon) \mid (x, \varepsilon) \in [u, v] \times [0, \overline{\varepsilon}]\}$. It is easy to see that the weak dual theorem holds, i.e.,

$$\sup_{\sigma\ge 0} l(\sigma) \le \theta(0).$$

Note that l is a nondecreasing function. Therefore, invoking Part 2) in Theorem 5.2, we develop the following necessary and sufficient conditions for the validity of zero duality gap property, i.e.,

$$\sup_{\sigma\ge 0} l(\sigma) = \theta(0).$$

Corollary 5.2 *Let β satisfy Assumption (C). Then the zero duality gap property between (P) and (L) holds if and only if θ is lower semicontinuous at zero.*

6 Numerical results

To give some insight into the behavior of Algorithm 5.1, numerical tests are performed on four nonlinear programming problems with equality constraints obtained from [11]. The algorithm is implemented in Matlab 7.8.0 and runs on Intel Core 2 CPU 2.39 GHz with 1.99 GB memory. We use $\|\nabla_{(x,\varepsilon)} \widetilde{f}_\sigma(x,\varepsilon)\| \le 10^{-6}$ as stopping criteria. Tables 1, 2, 3 and 4 show the computational results for the corresponding problems with the following items:

$\phi_i(t)$ $(i = 1, 2, \ldots, 6)$	as defined in Section 2,
σ_k	the penalty parameter,
x^k, ε_k	the final iterate,
$\Delta(x^k, \varepsilon_k)$	the violation measure of the constraints,
$\widetilde{f}_{\sigma_k}(x^k, \varepsilon_k)$	the value of penalty function $\widetilde{f}_\sigma(x,\varepsilon)$ at the final (x^k, ε_k) with the penalty parameter σ_k.

We make numerical tests using different choices of ϕ_i, $i = 1, 4, 6$ defined in Sect. 2, where $\phi_1(t)$ is barrier-type penalty function and $\phi_4(t)$, $\phi_6(t)$ are exterior-type penalty functions. Here, $\beta(\varepsilon)$ in (2.1) is set as $\sqrt{\varepsilon}$ and $\rho = 2$ or $\rho = 4$. Tables 1, 2, 3 and 4 illustrate the practical behavior of the algorithm. The penalty subproblem can be (approximately) solved by unconstrained smooth minimization techniques since the proposed penalty function has

Table 1 Numerical results of Example 6.1

$\phi_i(t)$	σ_k	x^k	ε_k	$\Delta(x^k, \varepsilon_k)$	$\widetilde{f}_{\sigma_k}(x^k, \varepsilon_k)$
$\phi_1(t) = \frac{t}{1-0.1t}$	1	(1.0242, 1.0620, −0.0033, 0.0214)	0.0060	5.3129e−4	−0.9024
	2	(0.9955, 0.9884, 0.0445, 0.0494)	5.7578e−5	2.0030e−8	−0.9801
	4	(1.0000, 1.0000, −5.7647e−4,0.0032)	1.7402e−7	5.2608e−10	−0.9968
$\phi_4(t)$	1	(0.9994, 0.9984, −0.0252, −0.0157)	0.0014	3.7181e−6	−0.9600
	2	(1.0001, 1.0004, 0.0185, −0.0047)	1.3967e−4	2.7711e−7	−0.9755
	4	(0.9999, 0.9999, 0.0003, −0.0022)	2.6207e−7	9.9595e−10	−0.9960
$\phi_6(t)$	1	(1.0064, 1.0296, 0.0860, −0.0031)	0.0028	2.9153e−4	−0.9274
	2	(0.9953, 0.9878, 0.0439, −0.0528)	5.4413e−7	4.8769e−9	−0.9919
	4	(0.9998, 0.9993, 9.3599e−4, 0.0152)	8.6056e−9	1.3469e−11	−0.9990

Table 2 Numerical results of Example 6.2

$\phi_i(t)$	σ_k	x^k	ε_k	$\Delta(x^k, \varepsilon_k)$	$\widetilde{f}_{\sigma_k}(x^k, \varepsilon_k)$
$\phi_1(t) = \frac{t}{1-0.1t}$	1	(1.0026, 1.0019, 0.9979, 0.9936, 0.9971)	6.4899e−8	2.5015e−7	0.0100
	2	(0.9997, 0.9908, 1.0062, 1.0216, 1.0003)	5.6659e−8	2.3655e−11	0.0010
	4	(0.9972, 1.0071, 0.9962, 0.9852, 1.0028)	1.3359e−8	8.6253e−12	0.0010
$\phi_4(t)$	1	(0.9915, 0.9905, 1.0090, 1.0272, 1.0084)	6.4169e−5	2.1140e−7	0.0100
	2	(1.0034, 1.0016, 0.9978, 0.9941, 0.9966)	1.2454e−6	6.8508e−9	0.0050
	4	(0.9979, 1.0006, 1.0003, 0.9999, 1.0021)	1.0000e−6	6.5704e−11	0.0040
$\phi_6(t)$	1	(0.9835, 0.9875, 1.0135, 1.0397, 1.0163)	3.7318e−5	4.7781e−7	0.0100
	2	(1.0221, 1.0006, 0.9922, 0.9840, 0.9784)	1.5117e−6	1.2133e−8	0.0050
	4	(1.0083, 1.0038, 0.9947, 0.9856, 0.9917)	8.8059e−7	6.7950e−10	0.0040

Table 3 Numerical results of Example 6.3

$\phi_i(t)$	σ_k	x^k	ε_k	$\Delta(x^k, \varepsilon_k)$	$\widetilde{f}_{\sigma_k}(x^k, \varepsilon_k)$
$\phi_1(t) = \frac{t}{1-0.01t}$	10	(−0.1083, 0.0360, 0.5365, −0.4644,0.0360)	1.3508e−6	7.7016e−6	5.3317
	20	(−0.0935, 0.0311, 0.5130, −0.4506,0.0311)	1.0000e−6	6.0195e−11	5.3266
	40	(−0.0936, 0.0312, 0.5130,−0.4505,0.0312)	1.0000e−6	5.7181e−11	5.3266
$\phi_4(t)$	10	(−0.0701, 0.0551, 0.7457, −0.1994,0.2297)	0.0299	0.0475	3.5825
	20	(−0.0936, 0.0312, 0.5124, −0.4499,0.0312)	1.0515e−6	5.8772e−11	5.3266
	40	(−0.0944, 0.0315, 0.5126, −0.4495,0.0315)	1.1337e−6	9.8943e−11	5.3265
$\phi_6(t)$	10	(−0.0365, 0.0864, 1.0685, 0.1549,0.5093)	0.0588	0.2727	1.7232
	20	(−0.0997, 0.0332, 0.5246, −0.4583,0.0332)	1.0000e−6	5.7256e−8	5.3280
	40	(−0.0942, 0.0314, 0.5149, −0.4521,0.0314)	1.0626e−6	3.6097e−6	5.3265

good smoothness property. Here, in our numerical experiments, trust-region methods are employed for solving the penalty subproblems.

Example 6.1 ([11])

$$\begin{aligned} \min \quad & f(x) = -x_1 \\ \text{s.t.} \quad & F_1(x) = x_2 - x_1^3 - x_3^2 = 0, \\ & F_2(x) = x_1^2 - x_2 - x_4^2 = 0. \end{aligned}$$

The point $\bar{x} = (1, 1, 0, 0)$ is the unique (global) minimizer with the optimal objective function value -1.0000.

Example 6.2 ([11])

$$\begin{aligned} \min f(x) &= (x_1 - x_2)^2 + (x_2 - x_3)^2 + (x_3 - x_4)^4 + (x_4 - x_5)^4 \\ \text{s.t. } F_1(x) &= x_1 + x_2^2 + x_3^3 - 3 = 0, \\ F_2(x) &= x_2 - x_3^2 + x_4 - 1 = 0. \end{aligned}$$

The point $\bar{x} = (1, 1, 1, 1, 1)$ is the (global) minimizer with the optimal objective function value 0.

Example 6.3 ([11])

$$\begin{aligned} \min f(x) &= (4x_1 - x_2)^2 + (x_2 + x_3 - 2)^2 + (x_4 - 1)^2 + (x_5 - 1)^2 \\ \text{s.t. } F_1(x) &= x_1 + 3x_2 = 0, \\ F_2(x) &= x_3 + x_4 - 2x_5 = 0, \\ F_3(x) &= x_2 - x_5 = 0. \end{aligned}$$

The point $\bar{x} = (-\frac{33}{349}, \frac{11}{349}, \frac{180}{349}, -\frac{158}{349}, \frac{11}{349})$ is a minimizer with the optimal objective function value $\frac{1859}{349}$.

Example 6.4 ([11])

$$\begin{aligned} \min f(x) &= \sum_{j=1}^{10} x_j \left(c_j + \ln \frac{x_j}{x_1 + \cdots + x_{10}} \right) \\ \text{s.t. } F_1(x) &= x_1 + 2x_2 + 2x_3 + x_6 + x_{10} - 2 = 0, \end{aligned}$$

Table 4 Numerical results of Example 6.4

$\phi_i(t)$	σ_k	x^k	ε_k	$\Delta(x^k, \varepsilon_k)$	$\widetilde{f}_{\sigma_k}(x^k, \varepsilon_k)$
$\phi_1(t) = \frac{t}{1-0.01t}$	1	(0.1191,0.9396,7.7552e−5, 0.0028, 0.4978,0.0015, 0, 0, 0, 0)	1.5441e−6	3.8468e−8	−30.5591
	4	(0.9749, 0.5096, 0.0002, 0.0054, 0.4945, 0.0053, 0, 0, 0, 0)	1.1971e−6	2.5382e−7	−29.1162
	16	(0.0998,0.9482,1.2600e−5,0.0014,0.4975,0.0035,0,7.4320e−5,0,0.0002)	9.7000e−6	1.2211e−7	−30.5731
$\phi_4(t)$	2	(0.1002,0.9487,2.3861e−4,0.0013,0.4985,0.0017,6.7980e−7,7.7881e−7,0,1.2701e−5)	1.0000e−5	2.5026e−12	−30.5803
	4	(0.1013, 0.9480, 0, 0.0011,0.4981,0.0025,5.370e−7, 0, 0, 0)	1.0000e−8	3.8461e−13	−30.5826
	8	(0.1083, 0.9438, 1.3200e−6, 2.6780e−4, 0.4978, 0.0040, 0, 0, 0, 0)	1.0000e−8	8.8444e−8	−30.5815
$\phi_6(t)$	1	(0.1003,0.9486,0.0003,0.0013,0.4985,0.0017,1.1820e−6,2.1168e−6,0,2.5201e−5)	1.0000e−5	2.1877e−7	−30.5873
	4	(0.1003,0.9486,0.0003,0.0013,0.4985,0.0018,4.3192e−6,1.5402e−6,0,1.4780e−5)	1.0000e−5	2.1907e−7	−30.5778
	16	(0.1059, 0.9458, 2.9000e−7, 0.0011, 0.4982, 0.0023, 0, 0, 0, 0)	1.0000e−8	2.7457e−13	−30.5813

$$F_2(x) = x_4 + 2x_5 + x_6 + x_7 - 1 = 0,$$
$$F_3(x) = x_3 + x_7 + x_8 + 2x_9 + x_{10} = 0,$$
$$x_i \geq 0, i = 1, 2, \ldots, 10.$$

where $c_1 = -6.089, c_2 = -17.164, c_3 = -34.054, c_4 = -5.914, c_5 = -24.721, c_6 = -14.986, c_7 = -24.100, c_8 = -10.708, c_9 = -26.662, c_{10} = -22.179$. The point $\bar{x} = (0.1083, 0.9438, 0, 0.0004, 0.4978, 0.0040, 0, 0, 0, 0)$ is a (but not unique) local minimizer with the optimal objective function value -30.581215.

As shown, for whatever choices of ϕ_i, with the penalty parameter gradually increasing, the values of indicator variable ε_k and constraint violation measure $\Delta(x^k, \varepsilon_k)$ tend to zero as desired. Additionally, it is not difficult to observe that the minimizers can be obtained without requirements of large penalty parameters for the choices of ϕ_i considered here. Numerical performances verify the correctness of the developed theory as desired. For example, as illustrated in Tables 1, 2, 3 and 4, the iterates (x^k, ε_k) are already quite close to the point $(\bar{x}, 0)$, where $\bar{x}$ is a minimizer of the original problem. As stated in Theorem 3.1, the optimal solution of the penalty problem must take the form of (x^k, ε_k) with $\varepsilon_k = 0$ for sufficiently large penalty parameters, which further means x^k is a local optimal solution of the original problem (P).

In summary, our numerical experiments on four examples of nonlinear programming problems with equality constraints confirm the efficiency of the proposed algorithm. As shown in Tables 1, 2, 3 and 4, the numerical outputs for the different choices for $\phi_i, i = 1, 4, 6$ seem to have no significant differences, which demonstrates the class of convex functions presenting a unified framework for some barrier-type and exterior-type functions are effective for solving nonlinear programming problems. Nevertheless, there exists a little difference in the algorithm implementation process for solving exterior-type penalty functions and barrier-type penalty functions; for barrier-type penalty functions, we must take the additional constraint $\Delta(x, \varepsilon) < a$ into account in contrast to exterior-type penalty functions, for instance, $a = q^{-1}$ for $\phi_1(t)$, $a = \frac{\pi}{2}$ for $\phi_2(t)$, and $a = 1$ for $\phi_3(t)$.

Acknowledgments The authors would like to thank referees for their constructive comments and suggestions, which significantly improved the presentation of the paper. The authors also thank Prof. Zhiyuan Tian in Qingdao University for his help.

References

1. Antczak, T.: Exact penalty functions method for mathematical programming problems involving invex functions. Eur. J. Oper. Res. **198**, 29–36 (2009)
2. Bazaraa, M., Sherali, H., Shetty, C.: Nonlinear Programming: Theory and Algorithms. Wiley, New York (2006)
3. Burke, J.: Calmness and exact penalization. SIAM J. Control Optim. **29**, 493–497 (1991)
4. Cominetti, R., Dussault, J.P.: Stable exponential-penalty algorithm with superlinear convergence. J. Optim. Theory Appl. **83**, 285–390 (1994)
5. Chen, L., Goldfarb, D.: Interior-point l_2 penalty methods for nonlinear programming with strong global convergence properties. Math. Program. **108**, 1–36 (2006)
6. Evans, J.P., Gould, F.J., Tolle, J.W.: Exact penalty functions in nonlinear programming. Math. Programm. **4**, 72–97 (1973)
7. Fletcher, R.: An exact penalty function for nonlinear programming with inequalities. Math. Program. **5**, 129–150 (1973)
8. Gonzaga, C.C., Castillo, R.A.: A nonlinear programming algorithm based on non-coercive penalty functions. Math. Program. **96**, 87–101 (2003)
9. Han, S.P., Mangasarian, O.L.: A dual differentiable exact penalty function. Math. Program. **25**, 293–306 (1983)

10. Han, S.P., Mangasarian, O.L.: Exact penalty functions in nonlinear programming. Math. Program. **17**, 251–269 (1979)
11. Hock, W., Schittkowski, K.: Test Examples for Nonlinear Programming Codes. Springer, New York (1981)
12. Hoheisel, T., Kanzowa, C., Outrata, J.: Exact penalty results for mathematical programs with vanishing constraints. Nonlinear Anal. **72**, 2514–2526 (2010)
13. Huyer, W., Neumaier, A.: A new exact penalty function. SIAM J. Optim. **13**, 1141–1158 (2003)
14. Li, W., Peng, J.: Exact penalty functions for constrained minimization problems via regularized gap function for variational inequalities. J. Global Optim. **37**, 85–94 (2007)
15. Lucidi, S.: New results on a continuously differentiable exact penalty function. SIAM J. Optim. **2**, 558–574 (1992)
16. Luo, Z.Q., Pang, J.S., Ralph, D.: Mathematical Programs with Equilibrium Constraints. Cambridge University Press, Cambridge (1996)
17. Meng, Z.Q., Dang, C.Y., Yang, X.Q.: On the smoothing of the square-root exact penalty function for inequality constrained optimization. Comput. Optim. Appl. **35**, 375–398 (2006)
18. Di Pillo, G: Exact penalty methods, algorithms for continuous optimization. In: Spedicato, E. (ed.). Kluwer, The Netherlands (1994)
19. Di Pillo, G., Grippo, L.: An exact penalty function method with global convergence properties for nonlinear programming problems. Math. Program. **36**, 1–18 (1986)
20. Di Pillo, G., Lucidi, S.: An augmented Lagrangian function with improved exactness properties. SIAM J. Optim. **12**, 376–406 (2001)
21. Tseng, P., Bertsekas, D.P.: On the convergence of the exponential multiplier method for convex programming. Math. Program. **60**, 1–19 (1993)
22. Zaslavski, A.J.: Existence of exact penalty for constrained optimization problems in Hilbert spaces. Nonlinear Anal. **67**, 238–248 (2007)