

# Smooth exact penalty functions: a general approach

Maksim V. Dolgopolik<sup>1</sup>

Received: 11 January 2015 / Accepted: 25 March 2015 / Published online: 10 April 2015  
© Springer-Verlag Berlin Heidelberg 2015

**Abstract** In the article, we present a new perspective on the method of smooth exact penalty functions that is becoming more and more popular tool for solving constrained optimization problems. In particular, our approach to smooth exact penalty functions allows one to apply previously unused tools (namely, parametric optimization) to the study of these functions. We give a new simple proof of local exactness of smooth penalty functions that significantly generalizes all similar results existing in the literature. We also provide new necessary and sufficient conditions for a smooth penalty function to be globally exact.

**Keywords** Penalty function · Smooth exact penalty function · Perturbed optimization problem · Calmness

## 1 Introduction

The method of exact penalty functions [1–3] is a powerful tool for solving various constrained optimization problems. However, as it is well-known, exact penalty functions are usually nonsmooth, even in the case when the original problem is smooth. This obstacle makes it impossible to apply (without any transformation of the problem) well-developed and extensively studied methods of smooth unconstrained optimization to minimization of an exact penalty function.

Huyer and Neumaier in [4] proposed a new approach to exact penalization that allows one to overcome nonsmoothness of exact penalty functions. Namely, let the

---

✉ Maksim V. Dolgopolik  
maxim.dolgopolik@gmail.com

<sup>1</sup> Saint Petersburg State University, 35 Universitetskii ave., Petergof, Saint Petersburg 198504, Russia

original constrained optimization problem has the form

$$\min f(x) \text{ subject to } F(x) = 0, \quad x \in [\underline{x}, \bar{x}], \tag{1}$$

where  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  and  $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$  are smooth function,  $\underline{x}, \bar{x} \in \mathbb{R}^n$  are given vectors, and

$$[\underline{x}, \bar{x}] = \{x = (x_1, \dots, x_n) \in \mathbb{R}^n \mid \underline{x}_i \leq x_i \leq \bar{x}_i \quad \forall i \in \{1, \dots, n\}\}.$$

The new approach relies on the introduction of additional variable  $\varepsilon \geq 0$  in the following way. Choose  $w \in \mathbb{R}^m$ , and note that the problem (1) is equivalent to the optimization problem

$$\min_{x, \varepsilon} f(x) \text{ subject to } F(x) = \varepsilon w, \quad \varepsilon = 0, \quad x \in [\underline{x}, \bar{x}]. \tag{2}$$

Then one defines the new “smooth” penalty function for the augmented problem (2) as follows

$$F_\lambda(x, \varepsilon) = \begin{cases} f(x), & \text{if } \varepsilon = \Delta(x, \varepsilon) = 0, \\ f(x) + \frac{1}{2\varepsilon} \frac{\Delta(x, \varepsilon)}{1 - q\Delta(x, \varepsilon)} + \lambda\beta(\varepsilon), & \text{if } \varepsilon > 0, \Delta(x, \varepsilon) < q^{-1}, \\ +\infty, & \text{otherwise.} \end{cases} \tag{3}$$

where  $\lambda \geq 0$  is the penalty parameter,  $\Delta(x, \varepsilon) = \|F(x) - \varepsilon w\|^2$  is the constraint violation measure,  $\beta: [0, \bar{\varepsilon}] \rightarrow [0, +\infty)$  with  $\beta(0) = 0$  is the penalty term,  $q > 0$  and  $\bar{\varepsilon} > 0$  are some prespecified thresholds. Finally, one replaces the augmented problem (2) with the penalized problem

$$\min_{x, \varepsilon} F_\lambda(x, \varepsilon) \text{ subject to } (x, \varepsilon) \in [\underline{x}, \bar{x}] \times [0, \bar{\varepsilon}]. \tag{4}$$

Observe that the penalty function  $F_\lambda(x, \varepsilon)$  is smooth for any  $\varepsilon \in (0, \bar{\varepsilon})$  and  $x$  such that  $0 < \Delta(x, \varepsilon) < q^{-1}$  provided the function  $\beta$  is smooth on  $(0, \bar{\varepsilon})$ . Furthermore, it was proved in [4] that under standard assumption (namely, constraint qualification) the penalty function  $F_\lambda(x, \varepsilon)$  is locally exact. In other words,  $(x^*, \varepsilon^*)$  is a local minimum of the problem (4) if and only if  $\varepsilon^* = 0$  and  $x^*$  is a local minimum of the original problem (1). Consequently, one can apply methods of smooth unconstrained minimization to the penalized problem (4) in order to find a solution of the initial constrained optimization problem (1).

Later on, the approach of smooth exact penalty functions was generalized [5, 6] and successfully applied to various constrained optimization problems [7, 8], including some optimal control problems [9–11]. However, it should be noted that the existing proofs of the exactness of the smooth penalty function (3) and its various generalizations are quite complicated, and overburdened by technical details that overshadow understanding of the technique of smooth exact penalty functions. Also, the question

of when the problem (1) is actually equivalent to the problem (4) in terms of optimal solutions [in this case the penalty function  $F_\lambda(x, \varepsilon)$  is called exact] has not been discussed in the literature.

The aim of this article is to present a new perspective on the method of smooth exact penalty functions. This perspective allows one to apply previously unused tools (namely, parametric optimization) to the study and construction of smooth exact penalty functions. It also helped us to essentially simplify the proof of exactness of these functions. Another aim of this articles is to provide necessary and sufficient conditions for the smooth penalty function to be (globally) exact.

The paper is organised as follows. In Sect. 2 we describe a new approach to smooth exact penalty functions. Some general results that draw a connection between exactness of the new penalty functions and some properties of perturbed optimization problems are presented in Sect. 3. In Sect. 4 we give a new simple proof of local exactness of a new penalty function that significantly generalizes all results on local exactness of smooth penalty functions existing in the literature. We also provide new conditions for the smooth penalty function to be globally exact.

## 2 How to construct a smooth exact penalty function?

Two main approaches are usually used for the study of exact penalty functions: direct approach, that is based on the use of error bound estimates and metric regularity, and indirect one, that relies on the consideration of a perturbed optimization problem. In the indirect approach [12–15], a perturbation of the initial problem is introduced as a tool for the study of a penalty function that is already defined. However, one can introduce a perturbation in order to *construct* a penalty function.

Namely, define the perturbed objective function for the problem (1)

$$g(x, \mu) = \begin{cases} f(x), & \text{if } \mu = \Delta(x, \mu) = 0, \\ f(x) + \frac{1}{\mu} \frac{\Delta(x, \mu)}{1 - q\Delta(x, \mu)}, & \text{if } \mu > 0, \Delta(x, \mu) < q^{-1}, \\ +\infty, & \text{otherwise,} \end{cases}$$

where  $\mu \geq 0$  is a perturbation parameter. Consider the perturbed optimization problem

$$\min_x g(x, \mu) \quad \text{subject to } x \in [\underline{x}, \bar{x}]. \quad (5)$$

It is clear that the problem above with  $\mu = 0$  is equivalent to the initial problem (1). Moreover, a penalization of the constraint  $F(x) = 0$  is achieved via the introduced perturbation. As a second step, note that the perturbed problem is equivalent to the problem

$$\min_{x, \mu} g(x, \mu) \quad \text{subject to } x \in [\underline{x}, \bar{x}], \quad \mu = 0.$$

Finally, introduce a penalty function for the problem above that penalizes only the constraint on the perturbation parameter, i.e.  $\mu = 0$ . The penalty function has the form

$$F_\lambda(x, \mu) = g(x, \mu) + \lambda\beta(\mu). \quad (6)$$

and it is smooth for any  $\mu \in (0, \bar{\varepsilon})$  and  $x \in \mathbb{R}^n$  such that  $0 < \Delta(x, \mu) < q^{-1}$ . Thus, the fact that the nonlinear constraints are taken into account via the perturbation (not penalization), and the fact that the penalty term is constructed only for a simple one dimensional constraint on the perturbation parameter allows one to avoid nonsmoothness that usually arises due to the restrictive requirements on a penalty term.

In the following section, we develop the approach discussed above in the general case, and show that exactness of the penalty function  $F_\lambda(x, \mu)$  of the form (6) is directly connected with some properties of the perturbed problem (5).

### 3 Exact penalty function for a perturbed optimization problem

Let  $X$  be a topological space,  $f: X \rightarrow \mathbb{R} \cup \{+\infty\}$  be a given function, and  $M, A \subset X$  be nonempty sets such that  $M \cap A \neq \emptyset$ . Hereafter, we study the following optimization problem:

$$\min f(x) \quad \text{subject to} \quad x \in M, \quad x \in A. \quad (\mathcal{P})$$

Denote by  $\Omega = M \cap A$  the set of feasible solutions of the problem ( $\mathcal{P}$ ). Denote also  $\mathbb{R}_+ = [0, +\infty)$  and  $\text{dom } f = \{x \in X \mid f(x) < +\infty\}$ . We suppose that the function  $f$  is bounded from below on  $\Omega$ .

Introduce a metric space of perturbation parameters  $(P, d)$ , and a perturbed objective function  $g: X \times P \rightarrow \mathbb{R} \cup \{+\infty\}$  such that there exists  $\mu_0 \in P$  for which the following conditions are satisfied:

1.  $g(x, \mu_0) = f(x)$  for any  $x \in \Omega$  (**consistency condition**);
2.  $\arg \min_{x \in A} g(x, \mu_0) = \arg \min_{x \in \Omega} f(x)$  (**exact penalization condition**).

With the use of the exact penalization condition one gets that the problem ( $\mathcal{P}$ ) is equivalent (in terms of optimal solutions) to the following optimization problem:

$$\min_{x, \mu} g(x, \mu) \quad \text{subject to} \quad x \in A, \quad \mu = \mu_0. \quad (7)$$

Furthermore, the consistency condition guarantees that if  $(x_0, \mu_0)$  with  $x_0 \in \Omega$  is a local minimum of the problem above, then  $x_0$  is a local minimum of the problem ( $\mathcal{P}$ ).

We apply the exact exact penalization technique to treat the problem (7). Namely, choose a function  $\beta: \mathbb{R}_+ \rightarrow \mathbb{R}_+ \cup \{+\infty\}$  such that  $\beta(t) = 0$  iff  $t = 0$ . For any  $\lambda \geq 0$  define the *penalty function*

$$F_\lambda(x, \mu) = g(x, \mu) + \lambda\beta(d(\mu, \mu_0))$$

and consider the following penalized problem

$$\min_{x, \mu} F_\lambda(x, \mu) \quad \text{subject to} \quad x \in A. \quad (8)$$

Observe that the function  $\lambda \rightarrow F_\lambda(x, \mu)$  is non-decreasing.

Our aim is to study a relation between (local or global) minima of the initial problem  $(\mathcal{P})$  and minima of the problem (8) in the context of the theory of exact penalty functions. Recall the definition of exact penalty function.

**Definition 1** Let  $x^* \in \text{dom } f$  be a local minimum of the problem  $(\mathcal{P})$ . The penalty function  $F_\lambda$  is called *exact* at the point  $x^*$  [or, to be more precise, at the point  $(x^*, \mu_0)$ ] if there exists  $\lambda \geq 0$  such that  $(x^*, \mu_0)$  is local minimum of the problem (8). The greatest lower bound of all such  $\lambda$  is denoted by  $\lambda(x^*)$ .

**Definition 2** The penalty function  $F_\lambda$  is said to be *exact* if there exists  $\lambda \geq 0$  such that the function  $F_\lambda$  attains the global minimum on the set  $A \times P$ , and if  $(x^*, \mu^*) \in A \times P$  is an optimal solution of the problem (8), then  $\mu^* = \mu_0$ . The greatest lower bound of all such  $\lambda \geq 0$  is denoted by  $\lambda^*(g, \beta)$ .

- Remark 1* (i) Note that if  $(x^*, \mu^*) \in A \times P$  is an optimal solution of the problem (8) with  $\mu^* = \mu_0$ , then  $x^*$  is an optimal solution of the problem  $(\mathcal{P})$  by virtue of the exact penalization condition on the function  $g(x, \mu)$  and the fact that  $F_\lambda(x, \mu_0) = g(x, \mu_0)$ . Thus, the penalty function  $F_\lambda$  is exact if and only if there exists  $\lambda \geq 0$  such that the problem (8) is equivalent to the problem  $(\mathcal{P})$  in terms of optimal solutions.
- (ii) It is easy to verify that if the penalty function  $F_\lambda$  is exact, then for any  $\lambda > \lambda^*(g, \beta)$  the function  $F_\lambda$  attains the global minimum on the set  $A \times P$ , and if  $(x^*, \mu^*) \in A \times P$  is an optimal solution of the problem (8), then  $\mu^* = \mu_0$ .

It transpires that exactness of the penalty function  $F_\lambda$  is closely related to some properties of the perturbed optimization problem

$$\min_x g(x, \mu) \quad \text{subject to } x \in A \quad (\mathcal{P}_\mu).$$

Denote by  $h(\mu) = \inf_{x \in A} g(x, \mu)$  the *optimal value function* of this problem.

Recall that the problem  $(\mathcal{P}_{\bar{\mu}})$ ,  $\bar{\mu} \in P$ , is said to be  $\beta$ -calm at a point  $x^* \in A$  if there exists  $\lambda \geq 0, r > 0$  and a neighbourhood  $U$  of  $x^*$  such that

$$g(x, \mu) - g(x^*, \bar{\mu}) \geq -\lambda\beta(d(\mu, \bar{\mu})) \quad \forall x \in U \cap A \quad \forall \mu \in B(\bar{\mu}, r),$$

where  $B(\bar{\mu}, r) = \{\mu \in P \mid d(\mu, \bar{\mu}) \leq r\}$ .

*Remark 2* If  $\beta(t) \equiv t$ , then the concept of  $\beta$ -calmness coincides with the well-known concept of calmness of a perturbed optimization problem [12–15].

The following propositions describes a connection between exactness of the penalty function  $F_\lambda$  and calmness of the perturbed problem  $(\mathcal{P}_\mu)$  (cf. analogous result for classical penalty functions in [14]).

**Proposition 1** Let  $x^* \in \text{dom } f$  be a local minimum of the problem  $(\mathcal{P})$ . Then the penalty function  $F_\lambda$  is exact at the point  $x^*$  if and only if the problem  $(\mathcal{P}_{\mu_0})$  is  $\beta$ -calm at  $x^*$ .

*Proof* The validity of the proposition follows from the fact that the inequality

$$g(x, \mu) - g(x^*, \mu_0) \geq -\lambda\beta(d(\mu, \mu_0)) \quad \forall x \in U \cap A \quad \forall \mu \in B(\bar{\mu}, r),$$

holds true for some  $\lambda \geq 0, r \geq 0$  and a neighbourhood  $U$  of  $x^*$  iff for any  $x \in U \cap A$  and  $\mu \in B(\mu_0, r)$  one has

$$F_\lambda(x, \mu) = g(x, \mu) + \beta(d(\mu, \mu_0)) \geq g(x, \mu_0) = F_\lambda(x, \mu_0),$$

i.e. iff  $(x^*, \mu_0)$  is a local minimum of  $F_\lambda$  on the set  $A$ . □

We need an auxiliary definition. The optimal value function  $h$  is called  $\beta$ -calm from below at a point  $\bar{\mu} \in P$  if

$$\liminf_{\mu \rightarrow \mu_0} \frac{h(\mu) - h(\bar{\mu})}{\beta(d(\mu, \bar{\mu}))} > -\infty$$

(cf. the definition of calmness from below in [16]).

**Theorem 1** *Let  $\beta$  be strictly increasing. For the penalty function  $F_\lambda$  to be exact it is necessary and sufficient that the following assumptions hold true:*

1. *there exists an optimal solution of the problem  $(P)$ ;*
2. *the optimal value function  $h(\mu)$  is  $\beta$ -calm from below at  $\mu_0$ ;*
3. *there exists  $\lambda_0 \geq 0$  such that  $F_{\lambda_0}$  is bounded from below on the set  $A \times P$ .*

*Proof* Necessity. Since the penalty function  $F_\lambda$  is exact, then it attains the global minimum on the set  $A \times P$  for some  $\lambda \geq 0$ . Therefore, in particular,  $F_\lambda$  is bounded from below on  $A \times P$ .

Fix sufficiently large  $\lambda \geq 0$ , and a global minimizer  $(x^*, \mu^*)$  of  $F_\lambda$  on the set  $A \times P$ . Due to the exactness of  $F_\lambda$  one has  $\mu^* = \mu_0$ . Therefore, as it was mentioned above,  $x^*$  is an optimal solution of the problem  $(P)$ . Thus, the first assumption is valid as well.

Observe that

$$g(x^*, \mu_0) \geq h(\mu_0) := \inf_{x \in A} g(x, \mu_0) \geq \inf_{(x, \mu) \in A \times P} F_\lambda(x, \mu) = g(x^*, \mu_0).$$

Hence  $h(\mu_0) = g(x^*, \mu_0)$ . With the use of exactness of  $F_\lambda$  one gets that for all  $x \in A$  and  $\mu \in P$  the following holds

$$g(x, \mu) + \lambda\beta(d(\mu, \mu_0)) = F_\lambda(x, \mu) \geq F_\lambda(x^*, \mu_0) = g(x^*, \mu_0) = h(\mu_0)$$

or, equivalently,

$$g(x, \mu) - h(\mu_0) \geq -\lambda\beta(d(\mu, \mu_0)) \quad \forall x \in A \quad \forall \mu \in P.$$

Since the inequality above holds true for all  $x \in A$ , one obtains that

$$h(\mu) - h(\mu_0) \geq -\lambda\beta(d(\mu, \mu_0)) \quad \forall \mu \in P,$$

which implies the  $\beta$ -calmness from below of  $h$  at  $\mu_0$ .

Sufficiency. Let  $x^*$  be an optimal solution of the problem  $(\mathcal{P})$ . Then taking into account the exact penalization condition on the function  $g(x, \mu)$  one gets that  $h(\mu_0) = g(x^*, \mu_0)$ , i.e.  $x^*$  is a global minimum of the function  $x \rightarrow g(x, \mu_0)$ .

Since the optimal value function  $h$  is  $\beta$ -calm from below at  $\mu_0$ , then there exist  $\lambda_1 \geq 0$  and  $\delta > 0$  such that

$$h(\mu) - h(\mu_0) \geq -\lambda_1\beta(d(\mu, \mu_0)) \quad \forall \mu \in B(\mu_0, \delta).$$

Hence for any  $x \in A$  one has

$$g(x, \mu) - g(x^*, \mu_0) \geq h(\mu) - h(\mu_0) \geq -\lambda_1\beta(d(\mu, \mu_0)) \quad \forall \mu \in B(\mu_0, \delta)$$

or, equivalently, for all  $(x, \mu) \in A \times B(\mu_0, \delta)$  the following holds

$$F_{\lambda_1}(x, \mu) = g(x, \mu) + \lambda_1\beta(d(\mu, \mu_0)) \geq g(x^*, \mu_0) = F_{\lambda_1}(x^*, \mu_0).$$

On the other hand, if  $x \in A$  and  $\mu \notin B(\mu_0, \delta)$ , then for any  $\lambda \geq \lambda_2$ , where

$$\lambda_2 = \lambda_0 + \frac{g(x^*, \mu_0) - c}{\beta(\delta)}, \quad c = \inf_{(x, \mu) \in A \times P} F_{\lambda_0}(x, \mu) > -\infty,$$

one has

$$\begin{aligned} F_{\lambda}(x, \mu) &= F_{\lambda_0}(x, \mu) + (\lambda - \lambda_0)\beta(d(\mu, \mu_0)) \\ &\geq c + (\lambda - \lambda_0)\beta(\delta) \geq g(x^*, \mu_0) = F_{\lambda}(x^*, \mu_0). \end{aligned}$$

Thus, for any  $\lambda \geq \bar{\lambda} := \max\{\lambda_1, \lambda_2\}$  one has

$$F_{\lambda}(x, \mu) \geq F_{\lambda}(x^*, \mu_0) \quad \forall (x, \mu) \in A \times P$$

or, in other words, the penalty function  $F_{\lambda}$  attains the global minimum on  $A \times P$  at the point  $(x^*, \mu_0)$ . Let  $(\bar{x}, \bar{\mu}) \in A \times P$  be a different global minimizer of  $F_{\lambda}$  on  $A \times P$ . Let us show that  $\bar{\mu} = \mu_0$ , provided  $\lambda > \bar{\lambda}$ , then one concludes that the penalty function  $F_{\lambda}$  is exact.

Indeed, for any  $\lambda > \bar{\lambda}$ ,  $x \in A$  and  $\mu \neq \mu_0$  one has

$$F_{\lambda}(x^*, \mu_0) = F_{\bar{\lambda}}(x^*, \mu_0) \leq F_{\bar{\lambda}}(x, \mu) < F_{\lambda}(x, \mu),$$

since  $\beta(d(\mu, \mu_0)) > 0$ . Hence  $\bar{\mu} = \mu_0$  by virtue of the fact that  $(\bar{x}, \bar{\mu})$  is a global minimizer of  $F_{\lambda}$  on  $A \times P$ . □

Let us also point out a connection between calmness of the optimal value function  $h$  and calmness of the problem  $(\mathcal{P}_{\mu_0})$  at optimal solutions of the problem  $(\mathcal{P})$  in the case when the set  $A$  is compact.

**Theorem 2** *Let the set  $A$  be compact, and the function  $g(x, \mu)$  be lower semicontinuous (l.s.c.) on  $A \times B(\mu_0, r)$  for some  $r > 0$ . Then for the optimal value function  $h$  to be  $\beta$ -calm from below at  $\mu_0$  it is necessary and sufficient that the problem  $(\mathcal{P}_{\mu_0})$  is  $\beta$ -calm at every optimal solution of the problem  $(\mathcal{P})$ .*

*Proof* Necessity. Fix an optimal solution  $x^*$  of the problem  $(\mathcal{P})$ . From the definition of  $\beta$ -calmness it follows that there exist  $\lambda \geq 0$  and  $r > 0$  such that

$$h(\mu) - h(\mu_0) \geq -\lambda\beta(d(\mu, \mu_0)) \quad \forall \mu \in B(\mu_0, r).$$

Since  $x^*$  is an optimal solution of the problem  $(\mathcal{P})$ , then  $h(\mu_0) = g(x^*, \mu_0)$  due to the exact penalization condition on  $g(x, \mu)$ . Therefore

$$g(x, \mu) - g(x^*, \mu_0) \geq h(\mu) - h(\mu_0) \geq -\lambda\beta(d(\mu, \mu_0)) \quad \forall (x, \mu) \in A \times B(\mu_0, r).$$

Thus, the problem  $(\mathcal{P}_{\mu_0})$  is  $\beta$ -calm at  $x^*$ .

Sufficiency. Taking into account the facts that the set  $A$  is compact and the function  $g(x, \mu)$  is l.s.c. one gets that the function  $g(\cdot, \mu_0)$  attains the global minimum on the set  $A$ , and the set  $A^*$  of all global minima of  $g(\cdot, \mu_0)$  on  $A$  is compact. Furthermore, from the exact penalization condition it follows that  $A^*$  is also the set of all optimal solutions of the problem  $(\mathcal{P})$ . Hence the problem  $(\mathcal{P}_{\mu_0})$  is  $\beta$ -calm at every  $x^* \in A^*$ . Therefore for any  $x^* \in A^*$  there exist  $\lambda(x^*) \geq 0$ ,  $r(x^*) \geq 0$  and a neighbourhood  $U(x^*)$  of  $x^*$  such that

$$g(x, \mu) - g(x^*, \mu_0) \geq -\lambda(x^*)\beta(d(\mu, \mu_0)) \quad \forall (x, \mu) \in U(x^*) \times B(\mu_0, r(x^*)).$$

Applying the compactness of  $A^*$  one obtains that there exists  $x_1^*, x_2^*, \dots, x_n^* \in A^*$  such that  $A^* \subset \bigcup_{k=1}^n U(x_k^*)$ . Denote

$$U = \bigcup_{k=1}^n U(x_k^*), \quad \bar{\lambda} = \max_{k \in 1:n} \lambda(x_k^*), \quad \bar{r} = \min_{k \in 1:n} r(x_k^*).$$

Then for any  $x \in U$  one has

$$g(x, \mu) - h(\mu_0) \geq -\bar{\lambda}\beta(d(\mu, \mu_0)) \quad \forall \mu \in B(\mu_0, \bar{r}), \tag{9}$$

due to the fact that  $g(x^*, \mu_0) = h(\mu_0) = \inf_{x \in A} g(x, \mu_0)$  for any  $x^* \in A^*$ .

Set  $K = A \setminus U$ . Since  $A^* \subset U$ , then for any  $x \in K$  one has  $g(x, \mu_0) > h(\mu_0)$  (recall that  $A^*$  is the set of all global minima of  $g(\cdot, \mu_0)$  on  $A$ ). Consequently, applying the lower semicontinuity of the function  $g(x, \mu)$  one gets that for any  $x \in K$  there exist  $\delta(x) > 0$  and a neighbourhood  $V(x)$  of  $x$  such that

$$g(y, \mu) > h(\mu_0) \quad \forall (y, \mu) \in V(x) \times B(\mu_0, \delta(x)).$$

Observe that from the facts that  $U$  is open and  $A$  is compact it follows that  $K = A \setminus U$  is also compact. Hence and from the inequality above it is easy to show that there exists  $\delta > 0$  such that



$$g(x, \mu) > h(\mu_0) \quad \forall (x, \mu) \in K \times B(\mu_0, \delta).$$

Therefore taking into account (9) one gets that

$$g(x, \mu) - h(\mu_0) \geq -\bar{\lambda}\beta(d(\mu, \mu_0)) \quad \forall (x, \mu) \in A \times B(\mu_0, \min\{\delta, \bar{r}\}),$$

which yields

$$h(\mu) - h(\mu_0) \geq -\bar{\lambda}\beta(d(\mu, \mu_0)) \quad \forall \mu \in B(\mu_0, \min\{\delta, \bar{r}\}).$$

Thus, the optimal valued function  $h$  is  $\beta$ -calm from below at  $\mu_0$ . □

Combining Theorems 1 and 2, and Proposition 1 one obtains that the following result holds true.

**Corollary 1** *Let the set  $A$  be compact, the function  $g(x, \mu)$  be l.s.c. on  $A \times B(\mu_0, r)$  for some  $r > 0$ , and the function  $\beta$  be strictly increasing. Then the penalty function  $F_\lambda$  is exact if and only if it is exact at every optimal solution of the problem (P), and there exists  $\lambda_0 \geq 0$  such that  $F_{\lambda_0}(x, \mu)$  is bounded from below on  $A \times P$ .*

### 4 Smooth exact penalty functions

Let us apply the theory developed in the previous section to the study of smooth exact penalty functions. Let  $X$  and  $Y$  be metric spaces,  $A \subset X$  be a nonempty set, and  $\Phi: X \rightrightarrows Y$  be a given set-valued mapping with closed images. For any subset  $C \subset X$  and  $x_0 \in X$  denote by

$$d(x_0, C) = \inf_{x \in C} d(x_0, x)$$

the distance between  $C$  and  $x_0$ . For any  $y \in Y$  denote, as usual,  $\Phi^{-1}(y) = \{x \in X \mid y \in \Phi(x)\}$ .

Fix an element  $y_0 \in Y$ , and consider the following optimization problem:

$$\min f(x) \quad \text{subject to} \quad y_0 \in \Phi(x), \quad x \in A. \tag{10}$$

Note that the set  $\Omega$  of feasible solutions of this problem has the form  $\Omega = \Phi^{-1}(y_0) \cap A$ .

Following the general technique proposed above and the method of smooth exact penalty functions [6], define  $P = \mathbb{R}_+$ , fix a non-decreasing function  $\phi: \mathbb{R}_+ \cup \{+\infty\} \rightarrow \mathbb{R}_+ \cup \{+\infty\}$  such that  $\phi(t) = 0$  iff  $t = 0$ , and introduce the perturbed objective function

$$g(x, \mu) = \begin{cases} f(x), & \text{if } x \in \Omega, \mu = 0, \\ +\infty, & \text{if } x \notin \Omega, \mu = 0, \\ f(x) + \frac{1}{\mu}\phi(d(y_0, \Phi(x))^2), & \text{if } \mu > 0. \end{cases}$$

Clearly, the function  $g(x, \mu)$  satisfies the consistency condition and the exact penalization condition with  $\mu_0 = 0$ .

Introduce the penalty function

$$F_\lambda(x, \mu) = g(x, \mu) + \lambda\beta(\mu),$$

where  $\beta: \mathbb{R}_+ \rightarrow \mathbb{R}_+ \cup \{+\infty\}$  is a non-decreasing function such that  $\beta(\mu) = 0$  iff  $\mu = 0$ . Let us obtain sufficient conditions for  $F_\lambda(x, \mu)$  to be exact. In order to formulate these conditions, recall that a set valued mapping  $\Phi$  is said to be *metrically subregular* with respect to the set  $A$  with constant  $a > 0$  at a point  $(\bar{x}, \bar{y}) \in X \times Y$  with  $\bar{y} \in \Phi(\bar{x})$  and  $\bar{x} \in A$  if there exists a neighbourhood  $U$  of  $\bar{x}$  such that

$$d(\Phi(x), \bar{y}) \geq ad \left( x, \Phi^{-1}(\bar{y}) \cap A \right) \quad \forall x \in U \cap A.$$

Thus,  $\Phi$  is metrically subregular with respect to the set  $A$  iff the restriction of  $\Phi$  to  $A$  is metrically subregular in the usual sense. See [17, 18] and references therein for extensive study of metric subregularity.

**Theorem 3** *Let  $x^* \in \text{dom } f$  be a local minimum of the problem (10), the function  $f$  be Lipschitz continuous near  $x^*$ , and the set-valued mapping  $\Phi$  be metrically subregular with respect to the set  $A$  with constant  $a > 0$  at  $(x^*, y_0)$ . Suppose also that the following assumptions are satisfied:*

1. *there exists  $\phi_0 > 0$  and  $t_0 > 0$  such that  $\phi(t) \geq \phi_0 t$  for any  $t \in [0, t_0]$ ;*
2. *there exists  $\beta_0 > 0$  and  $\bar{\mu} > 0$  such that  $\beta(\mu) \geq \beta_0 \mu$  for any  $\mu \in [0, \bar{\mu}]$ .*

*Then the penalty function  $F_\lambda(x, \mu)$  is exact at the point  $x^*$ . Moreover, one has*

$$\lambda(x^*) \leq \frac{L^2}{4\phi_0\beta_0a^2}, \tag{11}$$

where  $L \geq 0$  is a Lipschitz constant of  $f$  near  $x^*$ .

*Proof* Since  $x^*$  is a local minimum of the problem (10), then there exists  $\rho > 0$  such that  $f(x) \geq f(x^*)$  for any  $x \in B(x^*, \rho) \cap \Omega$ . Suppose, for a moment, that there exists  $\delta \in (0, \rho)$  such that

$$f(x) - f(x^*) \geq -Ld(x, \Omega) \quad \forall x \in B(x^*, \delta) \setminus \Omega, \tag{12}$$

where  $L \geq 0$  is a Lipschitz constant of  $f$  near  $x^*$ . Then applying the metric subregularity of  $\Phi$  with respect to  $A$ , and the fact that the function  $\phi$  is non-decreasing, one gets that for any  $\mu > 0$  and  $x \in B(x^*, r) \cap A$ , where  $r = \min\{\delta, \sqrt{t_0}/a\}$ , the following inequalities hold true

$$\begin{aligned}
 g(x, \mu) - g(x^*, 0) &= f(x) - f(x^*) + \frac{1}{\mu} \phi \left( d(y_0, \Phi(x))^2 \right) \\
 &\geq -Ld(x, \Omega) + \frac{\phi_0 a^2}{\mu} d \left( x, \Phi^{-1}(y_0) \cap A \right)^2 \\
 &= -Ld(x, \Omega) + \frac{\phi_0 a^2}{\mu} d(x, \Omega)^2.
 \end{aligned}$$

Note that the function  $h(t) = -Lt + \phi_0 a^2 t^2 / \mu$  attains the global minimum at the point  $\mu L / 2\phi_0 a^2$ , and

$$h \left( \frac{\mu L}{2\phi_0 a^2} \right) = -\frac{L^2}{4\phi_0 a^2} \mu.$$

Hence for any  $\mu > 0$  and  $x \in B(x^*, r) \cap A$  one has

$$g(x, \mu) - g(x^*, 0) \geq -\frac{L^2}{4\phi_0 a^2} \mu. \tag{13}$$

On the other hand, if  $x \in B(x^*, r) \cap A$  and  $\mu = 0$ , then either  $x \notin \Omega$  and  $g(x, \mu) = +\infty \geq g(x^*, 0)$  or  $x \in \Omega$  and  $g(x, \mu) = f(x) \geq f(x^*) = g(x^*, 0)$  (recall that  $r \leq \delta < \rho$ ). Therefore the inequality (13) is satisfied for any  $x \in B(x^*, r) \cap A$  and  $\mu \geq 0$ , which yields that

$$F_\lambda(x, \mu) = g(x, \mu) + \lambda\beta(\mu) \geq g(x, \mu) + \lambda\beta_0\mu \geq g(x^*, 0) = F_\lambda(x^*, 0)$$

for all  $(x, \mu) \in B(x^*, r) \times [0, \bar{\mu}]$  and for any  $\lambda \geq L^2 / 4\phi_0\beta_0 a^2$ . Thus, the penalty function  $F_\lambda(x, \mu)$  is exact at  $x^*$ , and (11) holds true.

It remains to show that the inequality (12) is valid for some  $\delta > 0$ . Indeed, fix  $x \in B(x^*, \rho/2) \setminus \Omega$ . By the definition of the distance between a point and a set there exists a sequence  $\{x_n\} \subset \Omega$  such that  $d(x, x_n) \rightarrow d(x, \Omega)$  as  $n \rightarrow \infty$ . Moreover, without loss of generality one can suppose that  $d(x, x_n) \leq \rho/2$  for any  $n \in \mathbb{N}$ , since  $d(x, x^*) \leq \rho/2$  and  $x^* \in \Omega$ . Consequently, one has

$$d(x_n, x^*) \leq d(x_n, x) + d(x, x^*) \leq \frac{\rho}{2} + \frac{\rho}{2} = \rho,$$

which implies  $f(x_n) \geq f(x^*)$ . Therefore applying the Lipschitz continuity of  $f$  near  $x^*$  one obtains that for any  $n \in \mathbb{N}$  the following inequalities holds true

$$f(x) - f(x^*) = f(x) - f(x_n) + f(x_n) - f(x^*) \geq f(x) - f(x_n) \geq -Ld(x, x_n).$$

Passing to the limit as  $n \rightarrow \infty$  one gets the desired result. □

Applying Corollary 1 and the theorem above one can easily obtain sufficient conditions for the penalty function  $F_\lambda$  to be exact.

**Theorem 4** *Let the set  $A$  be compact. Suppose that the following assumptions are satisfied:*

1.  $f$  is l.s.c. on  $A$ , and locally Lipschitz continuous near optimal solutions of the problem  $(\mathcal{P})$ ;
2.  $\Phi$  is metrically subregular with respect to the set  $A$  at  $(x^*, y_0)$  for any optimal solution  $x^*$  of the problem  $(\mathcal{P})$ ;
3. the mapping  $x \rightarrow d(y_0, \Phi(x))$  is continuous on  $A$ ;
4.  $\phi$  is l.s.c., and there exists  $\phi_0 > 0$  and  $t_0 > 0$  such that  $\phi(t) \geq \phi_0 t$  for any  $t \in [0, t_0]$ ;
5.  $\beta$  is strictly increasing and there exists  $\beta_0 > 0$  and  $\bar{\mu} > 0$  such that  $\beta(\mu) \geq \beta_0 \mu$  for any  $\mu \in [0, \bar{\mu}]$ .

Then the penalty function  $F_\lambda$  is exact.

*Proof* Applying Theorem 3, and taking into account the assumptions of the theorem one obtains that the penalty function  $F_\lambda$  is exact at every optimal solution of the problem  $(\mathcal{P})$ . Since  $f$  is l.s.c. on  $A$ , and the set  $A$  is compact, then  $f$  is bounded from below on this set. Hence the function  $g(x, \mu)$  is bounded from below on  $A \times \mathbb{R}_+$ , which implies that the penalty function  $F_\lambda$  is also bounded from below on  $A \times \mathbb{R}_+$  for any  $\lambda \geq 0$ .

Let us show that the function  $g(x, \mu)$  is l.s.c. on  $A \times \mathbb{R}_+$ . Then with the use of Corollary 1 one obtains the desired result.

For any  $\varepsilon > 0$  introduce the function

$$g_\varepsilon(x, \mu) = f(x) + \frac{1}{\mu + \varepsilon} \phi \left( d(y_0, \Phi(x))^2 \right) \quad (x, \mu) \in A \times \mathbb{R}_+.$$

Taking into account the fact that the function  $x \rightarrow d(y_0, \Phi(x))$  is continuous on  $A$ , and  $\phi$  is l.s.c., one gets that the function  $x \rightarrow \phi(d(y_0, \Phi(x))^2)$  is l.s.c. on  $A$  as well. Hence and from the lower semicontinuity of  $f$  it follows that the function  $g_\varepsilon(x, \mu)$  is l.s.c. on  $A \times \mathbb{R}_+$ . Note that

$$g(x, \mu) = \sup_{\varepsilon > 0} g_\varepsilon(x, \mu) \quad \forall (x, \mu) \in A \times \mathbb{R}_+.$$

Therefore the function  $g(x, \mu)$  is l.s.c. on  $A \times \mathbb{R}_+$  as the supremum of a family of l.s.c. functions. □

Theorem 3 can be modified to the case of more general functions  $g(x, \mu)$  and  $F_\lambda(x, \mu)$ . In particular, let

$$g(x, \mu) = \begin{cases} f(x), & \text{if } x \in \Omega, \mu = 0, \\ +\infty, & \text{if } x \notin \Omega, \mu = 0, \\ f(x) + \frac{1}{\mu^\alpha} \phi \left( d(y_0, \Phi(x))^2 \right), & \text{if } \mu > 0. \end{cases}$$

and

$$F_\lambda(x, \mu) = g(x, \mu) + \lambda \beta(\mu),$$

where  $\alpha > 0$  (cf. [8,9,11]). The following result holds true.

**Theorem 5** Let  $x^* \in \text{dom } f$  be a local minimum of the problem (10), the function  $f$  be Lipschitz continuous near  $x^*$ , and the set-valued mapping  $\Phi$  be metrically subregular with respect to the set  $A$  at  $(x^*, y_0)$ . Suppose that the following assumptions are satisfied:

1.  $\phi(t) \geq \phi_0 t^\gamma$  for any  $t \in [0, t_0]$ , and for some  $\phi_0 > 0$ ,  $\gamma > 0$  and  $t_0 > 0$ ;
2.  $\beta(\mu) \geq \beta_0 \mu^\sigma$  for any  $\mu \in [0, \bar{\mu}]$ , and for some  $\beta_0 > 0$ ,  $\sigma > 0$  and  $\bar{\mu} > 0$ .

Suppose also that

$$\gamma > \frac{1}{2}, \quad \sigma \leq \frac{\alpha}{2\gamma - 1}.$$

Then the penalty function  $F_\lambda(x, \mu)$  is exact at the point  $x^*$ .

*Proof* Arguing in the same way as in the proof of Theorem 3 one can easily verify that there exists  $\Theta > 0$  such that

$$g(x, \mu) - g(x^*, 0) \geq -\Theta \mu^{\frac{\alpha}{2\gamma-1}} \quad \forall x \in B(x^*, r) \cap A \quad \forall \mu \geq 0,$$

where  $r > 0$  is sufficiently small. Then applying the assumption on the function  $\beta$  one can show that  $F_\lambda$  is exact at  $x^*$ , provided  $\sigma \leq \alpha/(2\gamma - 1)$ .  $\square$

**Acknowledgments** The author was supported by the Russian Foundation for Basic Research (Project No. 14-01-31521 mol\_a) and Saint Petersburg State University (Project No. 9.38.205.2014).

## References

1. Di Pillo, G., Grippo, L.: Exact penalty functions in constrained optimization. *SIAM. J. Control Optim.* **27**, 1333–1360 (1989)
2. Burke, J.V.: An exact penalization viewpoint on constrained optimization. *SIAM. J. Control Optim.* **29**, 968–998 (1991)
3. Demyanov, V.F.: Nonsmooth optimization. In: Di Pillo, G., Schoen, F. (eds.) *Nonlinear Optimization. Lecture Notes in Mathematics*, vol. 1989, pp. 55–164. Springer, Heidelberg (2010)
4. Huyer, W., Neumaier, A.: A new exact penalty function. *SIAM. J. Optim.* **13**, 1141–1158 (2003)
5. Bingzhuang, L., Wenling, Z.: A modified exact smooth penalty function for nonlinear constrained optimization. *J. Inequal. Appl.* **1**, 173 (2012)
6. Wang, C., Ma, C., Zhou, J.: A new class of exact penalty functions and penalty algorithms. *J. Glob. Optim.* **58**, 51–73 (2014)
7. Ma, C., Li, X., Cedric Yiu, K.-F., Zhang, L.-S.: New exact penalty function for solving constrained finite min-max problems. *Appl. Math. Mech. Engl. Ed.* **33**, 253–270 (2012)
8. Lin, Q., Loxton, R., Teo, K.L., Wu, Y.H., Yu, C.: A new exact penalty method for semi-infinite programming problems. *J. Comput. Appl. Math.* **261**, 271–286 (2014)
9. Li, B., Yu, C.J., Teo, K.L., Duan, G.R.: An exact penalty function method for continuous inequality constrained optimal control problem. *J. Optim. Theory. Appl.* **151**, 260–291 (2011)
10. Jiang, C., Lin, Q., Yu, C., Teo, K.L., Duan, G.-R.: An exact penalty method for free terminal time optimal control problem with continuous inequality constraints. *J. Optim. Theory. Appl.* **154**, 30–53 (2012)
11. Lin, Q., Loxton, R., Teo, K.L., Wu, Y.H.: Optimal feedback control for dynamic systems with state constraints: an exact penalty approach. *Optim. Lett.* **8**, 1535–1551 (2014)
12. Clarke, F.H.: A new approach to lagrange multipliers. *Math. Oper. Res.* **1**, 165–174 (1976)
13. Clarke, F.H.: *Optimization nonsmooth analysis*. Wiley, New York (1983)
14. Burke, J.V.: Calmness and exact penalization. *SIAM. J. Control Optim.* **29**, 493–497 (1991)

15. Uderzo, A.: Exact penalty functions and calmness for mathematical programming under nonlinear perturbations. *Nonlinear. Anal.* **73**, 1596–1609 (2010)
16. Penot, J.-P.: Calmness and stability properties of marginal and performance functions. *Numer. Funct. Anal. Optim.* **25**, 287–308 (2004)
17. Azé, D.: A unified theory for metric regularity of multifunctions. *J. Convex. Anal.* **13**, 225–252 (2006)
18. Kruger, A.Y.: Error bounds and metric subregularity. *Optim.* (2014). doi:[10.1080/02331934.2014.938074](https://doi.org/10.1080/02331934.2014.938074)