Optimality Conditions for Semi-Infinite and Generalized Semi-Infinite Programs via Lower Order Exact Penalty Functions

Xiaoqi Yang · **Zhangyou Chen** · **Jinchuan Zhou**

Abstract In this paper, we will study optimality conditions of semi-infinite programs and generalized semi-infinite programs by employing lower order exact penalty functions and the condition that the directional derivative of the objective function at the candidate point along any feasible direction for the linearized constraint set is non-negative. We consider three types of penalty functions for semi-infinite program and investigate the relationship among the exactness of these penalty functions. We employ lower order integral exact penalty functions and the second-order generalized derivative of the constraint function to establish optimality conditions for semi-infinite programs. We adopt the exact penalty function technique in terms of a classical augmented Lagrangian function for the lower level problems of generalized semi-infinite programs to transform them into standard semi-infinite programs and then apply our results for semi-infinite programs to derive the optimality condition for generalized semi-infinite programs. We will give various examples to illustrate our results and assumptions.

This paper is dedicated to Professor Elijah (Lucien) Polak for his 85th birthday.

_____

Xiaoqi Yang(✉)
Corresponding author. Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China
e-mail: mayangxq@polyu.edu.hk

Zhangyou Chen
College of Mathematics and Statistics, South-West Jiao Tong University, China
e-mail: zhangyouchen@home.swjtu.edu.cn

Jinchuan Zhou
Department of Mathematics, School of Science, Shandong University of Technology, Zibo, China
e-mail: jinchuanzhou@126.com

1 Introduction

Semi-infinite programming (SIP, in short) has been long under intensive study and has wide applications in Chebyshev approximation, engineering design, and optimal control etc.; see, e.g., [1–5], while generalized semi-infinite programming (GSIP, in short) has applications in reverse Chebyshev approximation, maneuverability of a robot and time optimal control etc.; see, e.g., [6–8].

The study of first-order necessary optimality conditions is important as they play a key role in the development of optimization theory and the design of optimal algorithms. Various types of Fritz-John (FJ, in short) and Karush-Kuhn-Tucker (KKT, in short) optimality conditions for (SIP) have been extensively investigated without or with a constraint qualification, see [9–12] and [2, 10, 13–16], respectively. Similar necessary optimality conditions for (GSIP) have also been studied without or with a constraint qualification in [6] and [15, 17–20], respectively. See Section 2 for a brief review on first-order optimality conditions of (SIP) and (GSIP).

It is known that another approach to study optimality conditions is by virtue of exact penalty functions. For nonlinear programs, l_1 penalty function theory and algorithms have been widely studied. Clarke [21] and Burke [22] used l_1 exact penalty functions to derive KKT necessary optimality conditions. On the other hand, Yang and Meng [23] and Meng and Yang [24] used $l_p(p \in [0, 1])$ exact penalty functions to derive KKT necessary optimality conditions together with some nonpositivity condition on the second-order directional derivative of the constraints. Lower order $l_p(p \in [0, 1])$ penalty functions have been investigated in recent years in a framework of a generalized augmented Lagrangian with a generalized augmenting function [25, 26], which is considered as a generalization of classical augmented Lagrangian theory with a convex quadratic augmenting function in [27] and a more general convex augmenting function in [28].

In this paper, we will study optimality conditions of (SIP) and (GSIP) by employing lower order exact penalty functions and the condition that the directional derivative of the objective function at the candidate point along any feasible direction for the linearized constraint set is non-negative. We will consider three types of penalty functions for (SIP) and investigate the relationship among the exactness of these penalty functions. We employ lower order integral exact penalty functions and the second-order generalized derivative of the constraint function to establish optimality conditions for (SIP). Our results extend the ones in [23] to (SIP) by using different types of l_p exact penalty functions. More specifically, we will propose three types of pth-order ($p \in]0, 1]$) penalty functions, generalizations of the l_∞ and l_p penalty functions in nonlinear programming problems, respectively. Under the assumption of exactness of the penalty functions, we derive first-order optimality conditions for (SIP). We will illustrate that the assumptions in our results are different from the extended Abadie CQ.

We will also investigate the optimality condition for (GSIP). We will adopt the exact penalty function technique in terms of a classical augmented Lagrangian function for the lower level problems of (GSIP) to transform (GSIP) into (SIP) and then apply our previously obtained results for (SIP) to derive the optimality condition for (GSIP).

The paper is organized as follows. In Section 2, we present a literature review for the optimality conditions of (SIP) and (GSIP). In Section 3, we give definitions of different types of penalty functions and study their relationship. In Section 4, we are devoted to the first-order optimality

condition for (SIP) under the exact penalization assumption. In Section 5, as an application of the results in Section 4, we study the first-order optimality condition of (GSIP) through a double penalization.

2 Literature Review

Consider the following semi-infinite program, denoted as (SIP):

$$\min \quad f(x) \quad \text{s.t.} \quad g(x, t) \le 0, t \in \Omega,$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \times \Omega \to \mathbb{R}$ are continuously differentiable real-valued functions, and Ω is a nonempty and compact set of parameters in $\mathbb{R}^m$. Let

$$X := \{x \in \mathbb{R}^n : g(x, t) \le 0, t \in \Omega\}$$

be the feasible set and, for $x \in X$, let

$$\Omega_x := \{t \in \Omega : g(x, t) = 0\}.$$

Let x^* be a locally optimal solution of (SIP).

By using a max-reformulation, SIP can be written as a nonlinear program with a nonsmooth constraint: $\min f(x)$ s.t. $\max_{t \in \Omega} g(x, t) \le 0$. Correspondingly, Fritz John [9] (see also Pschenichnyi [10]) derived the following FJ necessary optimality condition

$$0 \in \text{conv}\{\nabla f(x^*), \nabla_x g(x^*, t) \, (t \in \Omega_{x^*})\}. \tag{1}$$

Hettich and Jongen [11] and Borwein [12] derived the FJ necessary optimality condition (1) using a reduction technique.

Using a constraint qualification that every d with $\nabla_x g(x^*, t)^\top d \le 0$ for $t \in \Omega_{x^*}$ is a feasible direction and a Farkas lemma, Hettich and Kortanek [2] showed that the following necessary optimality condition holds

$$0 \in \nabla f(x^*) + \text{cl cone}\{\nabla_x g(x^*, t) \, (t \in \Omega_{x^*})\}. \tag{2}$$

Li, et al [13] introduced an extended Abadie CQ for a convex SIP for establishing the necessary optimality condition (2).

Furthermore, for a convex SIP with a Slater condition, Pschenichnyi [10] derived the following KKT condition

$$0 \in \nabla f(x^*) + \text{cone}\{\nabla_x g(x^*, t) \, (t \in \Omega_{x^*})\}. \tag{3}$$

For a nondifferentiable convex SIP, Lopez and Vercher [14] established (3) using the property of Farkas-Minkowski, implied by a Slater condition. For a differentiable SIP satisfying an extended Abadie constraint qualification (see (4)), it was shown in Stein [15] that (3) holds. For an SIP with a lower semi-continuous objective function and a locally Lipschitz constraint function, Zheng and

Yang [16] derived a KKT necessary condition (3) by using a generalized constraint qualification, which is an extension of the one in [2]. Under the condition that cone$\{\nabla_x g(x^*, t) \ (t \in \Omega_{x^*})\}$ is closed, (2) becomes (3).

Let $p > 0$. Pietrzykowski [29] introduced the following l_p integral penalty function

$$f(x) + \rho \int_\Omega g_+^p(x, t)\mathrm{d}\mu(t)$$

and established the convergence of the solution sequence of the penalty problems to an optimal solution of (SIP). Conn and Gould [30] established the exactness for a modified l_1 integral penalty function with positive feasible parametric areas as a denominator.

For $x \in X$, let
$$T_X(x) := \{d : \exists d^k \to d, \tau^k \downarrow 0, \text{ s.t. } x + \tau^k d^k \in X\}$$

be the contingent cone of the feasible set X at x and

$$D(x) := \{0 \neq d \in \mathbb{R}^n : \langle \nabla_x g(x, t), d \rangle \leq 0, \forall t \in \Omega_x\}$$

be the set of feasible directions for the linearized constraint set, see [36]. Clearly

$$T_X(x) \subset D(x) \cup \{0\}.$$

The extended Abadie CQ for SIP [15] is said to hold at $x \in X$ iff

$$T_X(x) = D(x) \cup \{0\}. \tag{4}$$

It is easy to show that

$$\langle \nabla f(x^*), d \rangle \geq 0, \ \forall d \in T_X(x^*). \tag{5}$$

In this paper, we are devoted to the study of the following optimality condition of (SIP)

$$\langle \nabla f(x^*), d \rangle \geq 0, \ \forall d \in D(x^*), \tag{6}$$

via its lower order l_p exact penalty function, where $T_X(x)$ in (5) is replaced by $D(x^*)$. (6) is stronger than (5) in general. If the extended Abadie CQ (4) holds, then (6) and (5) are equivalent. On the other hand, (6) is also nothing new, but another formulation of (2). Indeed, by Farkas Lemma [2], (6) is equivalent to (2). Thus (6) is weaker than the KKT optimality condition (3). It is known that some more constraint qualification is needed to establish the KKT optimality condition (3) from the condition (6), see [2].

Clearly (6) implies that

$$\max_{t \in \Omega_{x^*}} \{\langle \nabla f(x^*), d \rangle, \langle \nabla_x g(x^*, t), d \rangle\} \geq 0, \tag{7}$$

for all $d \in \mathbb{R}^n$. However, the reverse does not always hold as shown by an example. Consider the problem $\min x_2$ s.t. $tx_1 + x_2^2 \leq 0, t \in [-1, 1]$. The FJ optimality condition (1) holds at the unique feasible point $x^* = (0, 0)$, where $\Omega_{x^*} = [-1, 1], \nabla f(x^*) = (0, 1)^T, \nabla_x g(x^*, t) = (t, 0)^T, t \in [-1, 1]$,

4

while (6) does not hold with $D(0, 0) = \{0\} \times (\mathbb{R} \setminus \{0\})$. It is also true that (7) is equivalent to the FJ optimality condition (1) due to the compactness of the set $\{\nabla_x g(x^*, t) \ (t \in \Omega_{x^*})\}$. This follows from the fact that, for a convex set C, $z \in \mathrm{cl}\,(C)$ iff $\langle z, d \rangle \leq \sigma_C^*(d)$ for every $d \in \mathbb{R}^n$, where $\sigma_C^*(d) = \sup_{x \in C} \langle x, d \rangle$ is the support function of C, see Section 13 of [31]. In such a case, letting $C := \mathrm{co}\,\{\nabla f(x^*), \nabla_x g(x^*, t)(t \in \Omega_{x^*})\}$ and $z = 0$, we have

$$\sigma_C^*(d) = \max\{\langle \nabla f(x^*), d \rangle, \langle \nabla_x g(x^*, t), d \rangle(t \in \Omega_{x^*})\}, \quad \forall d \in \mathbb{R}^n$$

which gives the desired result.

Consider also the following generalized semi-infinite program, denoted as (GSIP),

$$\min \quad f(x) \quad \text{s.t.} \quad g(x, t) \leq 0, t \in \Omega \cap \Omega(x),$$

where Ω is a compact subset of $\mathbb{R}^m$, $\Omega(x) := \{t \in \mathbb{R}^m : v_i(x, t) \leq 0, i = 1, \cdots, l\}$ and the functions $f \colon \mathbb{R}^n \to \mathbb{R}$, $g \colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, and $v_i \colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}(i = 1, \cdots, l)$ are continuously differentiable. Both (SIP) and (GSIP) feature a 'bi-level' structure, while for (GSIP) this feature is often explicitly employed in its study. Let the Lagrange function of the lower level problem be

$$L(x, t, \lambda, \mu) := \lambda g(x, t) - \sum_{i=1}^{l} \mu_i v_i(x, t),$$

where $\lambda, \mu_i \geq 0, i = 1, \cdots, l$.

Without any constraint qualification, Jongen, et al [6] derived a FJ necessary optimality condition with a FJ Lagrange function, that is, the constraint function $g(x, t)$ in (1) is replaced by the Lagrange function $L(x, t, \lambda, \mu)$. Based on a local representability of (GSIP) as an ordinary SIP and a lower level LICQ, Weber [17] described a necessary optimality condition of type (2). Rückmann and Shapiro [18] obtained a FJ condition with a KKT Lagrange function of the lower level problem by assuming a lower level MFCQ. Stein [19] obtained four kinds of FJ conditions, including a strengthened FJ condition for any lower level KKT Lagrange multiplier by assuming the directionally differentiability of the lower level optimal value function. Stein [15] and Ye and Wu [20] studied extended ACQ conditions for a GSIP.

On the other hand, a differentiable penalty function is used by Levitin [32] to eliminate the lower level constraints and (GSIP) can thus be approximated by a sequence of SIP's. Further developed by Polak and Royset [33] and Royset et al. [34], an l_∞ penalty function and an augmented Lagrangian function are adopted for (GSIP) with min-max form and also algorithms are proposed.

3 Max-Type and Integral-Type Penalty Functions

Recall that $X = \{x \in \mathbb{R}^n : g(x, t) \leq 0, t \in \Omega\}$ is the feasible set and, for $x \in X$,

$$\Omega_x = \{t \in \Omega : g(x, t) = 0\},$$

is the index set of active constraints of SIP.

Let $\mathbb{R}_+ := \{x \in \mathbb{R} : x \geq 0\}$ and $\mathbb{R}_{++} := \{x \in \mathbb{R} : x > 0\}$. A penalty function $F \colon \mathbb{R}^n \to \mathbb{R}$ is of the following form

$$F(x) := f(x) + \rho P(x),$$

where $\rho > 0$ is a penalty parameter and $P \colon \mathbb{R}^n \to \mathbb{R}_+$ is a penalty term satisfying

$$P(x) = 0 \Leftrightarrow x \in X.$$

Let x^* be a locally optimal solution of SIP. F is said to be exact at x^* iff there exists $\bar{\rho} > 0$ such that x^* is a locally optimal solution of F for any $\rho \geq \bar{\rho}$.

Let $p \in]0, 1]$ and $[a]_+ := \max\{a, 0\}$. We consider two types of penalty functions, i.e., max-type penalty and integral-type penalty functions. A pth-order max-type penalty function for SIP is defined as,

$$F^p_{max}(x) := f(x) + \rho \max_{t \in \Omega} g^p_+(x, t), \tag{8}$$

where $g^p_+(x, t) = ([g(x, t)]_+)^p$.

Let μ be a non-negative regular Borel measure defined on Ω with the support of μ being equal to Ω, that is $\mathrm{supp}(\mu) = \Omega$, where the support of μ is defined as the set of the points $t \in \Omega$ such that any open neighborhood V of t has a positive measure:

$$\mathrm{supp}(\mu) := \{t \in \Omega : \mu(V) > 0, \text{ for any open neighbourhood } V \text{ of } t\}.$$

Two pth-order integral-type penalty functions for SIP are defined, respectively, by

$$F^p_{int}(x) := f(x) + \rho \int_\Omega g^p_+(x, t) \, d\mu(t), \tag{9}$$

$$\bar{F}^p_{int}(x) := f(x) + \rho \left(\int_\Omega g_+(x, t) \, d\mu(t) \right)^p. \tag{10}$$

With the assumption $\mathrm{supp}(\mu) = \Omega$, $\int_\Omega g^p_+(x, t) \, d\mu(t)$ and $(\int_\Omega g_+(x, t) \, dt)^p$ are penalty terms, i.e.,

$$\int_\Omega g^p_+(x, t) \, d\mu(t) = 0 \iff g(x, t) \leq 0, \quad \forall t \in \Omega, \tag{11}$$

$$\left(\int_\Omega g_+(x, t) \, d\mu(t) \right)^p = 0 \iff g(x, t) \leq 0, \quad \forall t \in \Omega.$$

We show only the necessity of (11) since the sufficiency is trivial. Suppose that there exists $t_0 \in \Omega$ such that $g(x, t_0) > 0$. The continuity of g ensures the existence of neighborhood V of t_0 and $\alpha > 0$ such that $g(x, t) > \alpha$, for all $t \in V$. Thus,

$$\int_\Omega g^p_+(x, t) \, d\mu(t) \geq \int_V g^p_+(x, t) \, d\mu(t) \geq \alpha^p \mu(V) > 0,$$

where the last step comes from the fact that $\Omega = \mathrm{supp}(\mu)$. This leads to a contradiction.

Clearly, from the definition, if F^p_{max} (resp. F^p_{int} and $\bar{F}^p_{int}$) is exact at x^*, then so is $F^{\tilde{p}}_{max}$ (resp. $F^{\tilde{p}}_{int}$ and $\bar{F}^{\tilde{p}}_{int}$) for all $\tilde{p} \in]0, p[$. This is due to the fact that the local optimality of x^* guarantees that $g(x, t) < 1$ for all x in a neighborhood of x^* and thus $g^p_+(x, t) \leq g^{\tilde{p}}_+(x, t)$ if $\tilde{p} \in]0, p[$.

6

If the set Ω is finite, say $\Omega = \{t_1, \cdots, t_s\}$, then the feasible set X is defined by a finite number of inequalities and hence (SIP) becomes a standard nonlinear programming problem. Correspondingly, taking a regular Borel measure $\mu = \sum_{i=1}^{s} \delta_{t_i}$, where δ_{t_i} is a Dirac measure defined on Ω, the pth-order penalty functions (8), (9) and (10) take the form of

$$F_{max}^p(x) = f(x) + \rho \max_{1 \le i \le s} g_+^p(x, t_i), \tag{12}$$

$$F_{int}^p(x) = f(x) + \rho \sum_{i=1}^{s} g_+^p(x, t_i), \tag{13}$$

$$\bar{F}_{int}^p(x) = f(x) + \rho \left(\sum_{i=1}^{s} g_+(x, t_i) \right)^p, \tag{14}$$

respectively, where (13) and (14) are the pth-order penalty functions for a nonlinear programming problem investigated in Rubinov and Yang [25] for the existence of zero duality gap and exact penalty results and used in Meng and Yang [24] and Yang and Meng [23] for the study of optimality conditions. It is easy to see that the exactness of all three penalty functions (12), (13) and (14) are equivalent, see, e.g., Lemma 4.1 of [26].

Let μ be the Lebesgue measure. Note that Ω is a compact set of $\mathbb{R}^m$. If $\Omega = \mathrm{cl}\,\mathrm{int}\,\Omega$, then $\mathrm{supp}(\mu) = \Omega$. Here the support 'supp' is understood as the support of the Lebesgue measure with respect to Ω:

$$\mathrm{supp}(\mu) = \left\{ t \in \Omega : \int_{B(t,\delta) \cap \Omega} \mathrm{d}\mu(t) > 0 \text{ for all } \delta > 0 \right\}.$$

Especially, if Ω is a convex set with an interior point, then $\Omega = \mathrm{cl}\,\mathrm{int}\Omega$, see, e.g. Theorem 2.33 of [28] and so $\mathrm{supp}(\mu) = \Omega$.

Throughout the paper we will make the assumptions that $\Omega = \mathrm{supp}(\mu)$ whenever the integral-type penalty function is dealt with and that in all examples, for simplicity, μ is assumed to be the Lebesgue measure and $\mathrm{d}\mu(t)$ is simplified as $\mathrm{d}t$. For the details of real analysis, we refer the reader to the reference [35].

Next, we explore the relationships among the exact penalty functions F_{max}^p, $\bar{F}_{int}^p$ and F_{int}^p. First, let us introduce the following set, for $x \in \mathbb{R}^n$,

$$A_x := \arg\max_{t \in \Omega} g_+(x, t) = \{t \in \Omega : g_+(x, t) = \max_{t \in \Omega} g_+(x, t)\}.$$

Proposition 3.1. *Let $p \in\,]0, 1]$ and $\mathrm{supp}(\mu) = \Omega$. We have*

(i) *if F_{int}^p is an exact penalty function at x^*, so is $\bar{F}_{int}^p$;*

(ii) *if $\bar{F}_{int}^p$ is an exact penalty function at x^*, so is F_{max}^p;*

(iii) *if F_{max}^p is an exact penalty function at x^* and*

$$\liminf_{x \to x^*} \mu(A_x) > 0, \tag{15}$$

so is $\bar{F}_{int}^p$, where the definition of 'liminf' is referred to the one in [28].

7

Proof: (i) When $p = 1$, (i) is trivial. Let $p \in \,]0, 1[$. If F^p_{int} is an exact penalty function at x^*, there are a neighborhood $B(x^*)$ of x^* and $\bar{\rho} > 0$ such that, for any $x \in B(x^*)$ and $\rho \geq \bar{\rho}$,

$$f(x) + \rho \int_\Omega g_+(x, t)^p \, d\mu(t) \geq f(x^*), \quad \forall x \in B(x^*).$$

Following a special case of Hölder's inequality, see [35, Theorem 5.1], we have, for any $x \in B(x^*)$ and $\rho \geq \bar{\rho}\mu(\Omega)^{1-p}$,

$$f(x) + \rho \left(\int_\Omega g_+(x, t) \, d\mu(t) \right)^p \geq f(x) + \bar{\rho}\mu(\Omega)^{1-p}\mu(\Omega)^{p-1} \int_\Omega g_+(x, t)^p \, d\mu(t) \geq f(x^*).$$

(ii) It is easy to see that $g^p_+(x, t) \leq (\max_{t \in \Omega} g_+(x, t))^p = \max_{t \in \Omega} g^p_+(x, t), \forall x \in \mathbb{R}^n$. So,

$$\left(\int_\Omega g_+(x, t) \, d\mu(t) \right)^p \leq \left(\int_\Omega \max_{t \in \Omega} g_+(x, t) \, d\mu(t) \right)^p = [\mu(\Omega)]^p \max_{t \in \Omega} g^p_+(x, t),$$

where $\mu(\Omega) > 0$, since $\text{supp}(\mu) = \Omega$. By assumption, $\bar{F}^p_{int}$ is an exact penalty function at x^*, that is, there exists $\bar{\rho} > 0$ such that for every $\rho \geq \bar{\rho}$ there exists $\delta_\rho > 0$ satisfying

$$\bar{F}^p_{int}(x) = f(x) + \rho \left(\int_\Omega g_+(x, t) \, d\mu(t) \right)^p \geq f(x^*), \quad \forall x \in B(x^*, \delta_\rho). \tag{16}$$

Let $\hat{\rho} = [\mu(\Omega)]^p \bar{\rho}$. For each $\rho \geq \hat{\rho}$, we have from (16)

$$F^p_{max}(x) = f(x) + \rho \max_{t \in \Omega} g^p_+(x, t) \geq f(x) + \frac{\rho}{[\mu(\Omega)]^p} \left(\int_\Omega g_+(x, t) \, d\mu(t) \right)^p \geq f(x^*),$$

for any $x \in B(x^*, \frac{\delta_\rho}{[\mu(\Omega)]^p})$. This means that F^p_{max} is an exact penalty function at x^*.

(iii) The condition (15) is equivalent to the existence of two constants $\varepsilon > 0$ and $\delta_1 > 0$ such that

$$\mu(A_x) \geq \varepsilon^{\frac{1}{p}}, \quad \forall x \in B(x^*, \delta_1).$$

Thus, for any $x \in B(x^*, \delta_1)$, we have

$$\left(\int_\Omega g_+(x, t) \, d\mu(t) \right)^p \geq \left(\int_{A_x} g_+(x, t) \, d\mu(t) \right)^p = [\mu(A_x)]^p \max_{t \in \Omega} g^p_+(x, t) \geq \varepsilon \max_{t \in \Omega} g^p_+(x, t). \tag{17}$$

Since F^p_{max} is an exact penalty function, there exists $\bar{\rho} > 0$ such that for every $\rho \geq \bar{\rho}$, $\delta_\rho > 0$ (with $\delta_\rho < \delta_1$) is found such that

$$F^p_{max}(x) = f(x) + \rho \max_{t \in \Omega} g^p_+(x, t) \geq f(x^*), \quad \forall x \in B(x^*, \delta_\rho). \tag{18}$$

Let $\hat{\rho} = \frac{\bar{\rho}}{\varepsilon}$. For any $\rho \geq \hat{\rho}$ (implying $\rho\varepsilon \geq \bar{\rho}$), it follows from (17) and (18) that

$$\bar{F}^p_{int}(x) = f(x) + \rho \left(\int_\Omega g_+(x, t) \, d\mu(t) \right)^p \geq f(x) + \rho\varepsilon \max_{t \in \Omega} g^p_+(x, t) \geq f(x^*),$$

for all $x \in B(x^*, \delta_{\rho\varepsilon})$. That is, $\bar{F}^p_{int}$ is an exact penalty function at x^*. $\qquad\square$

The extended MFCQ of (SIP) at x^* is said to hold (see, [37]) iff there exists $h \in \mathbb{R}^n$ such that

$$\nabla_x g(x^*, t)^T h < 0, \quad \forall t \in \Omega_{x^*}.$$

Proposition 3.2. *Assume that the extended MFCQ of (SIP) holds at x^* and f is locally Lipschitz continuous near x^*. Then F^1_{max} is an exact penalty function. If in addition, the condition (15) holds, then F^1_{int} is also exact.*

Proof: Let $C(\Omega)$ be the space of all continuous functionals on Ω, $G(x)(\cdot) := g(x, \cdot)$ and

$$K := C_-(\Omega) := \{y \in C(\Omega) : y(w) \le 0, \forall w \in \Omega\}.$$

Consider the following optimization problem

$$\min \quad f(x) \quad \text{s.t.} \quad G(x) \in K, t \in \Omega.$$

It follows from page 510 of [37] that the extended MFCQ is equivalent to the following Robinson's constraint qualification:

$$G(x^*) + DG(x^*)h \in \text{int}(K).$$

Again it follows from Proposition 3.111 of [37] that, if the Robinson's constraint qualification holds at x^* and f is locally Lipschitz continuous near x^*, then the penalty function

$$\theta(x) := f(x) + \rho\text{dist}(G(x), K)$$

is exact at x^*, where $\text{dist}(G(x), K) = \min\{\text{dist}(G(x), y) : y \in K\}$ and the distance between two functionals $\phi, \varphi \in C(\Omega)$ is defined as

$$\text{dist}(\phi, \varphi) := \max_{t \in \Omega} \|\phi(t) - \varphi(t)\|.$$

Now we have $\text{dist}(G(x), K) = \max_{t \in \Omega} g_+(x, t)$. Thus F^1_{max} is an exact penalty function.

If in addition, (15) holds, then, by Proposition 3.1 (iii), F^1_{int} is also exact.. $\qquad \square$

We now give various examples to illustrate the results of Propositions 3.1 and 3.2. First the exactness of penalty function $\bar{F}^p_{int}$ can indeed be strictly weaker than that of F^p_{int} as shown by the following example.

Example 3.1. *Consider the SIP problem*

$$\min \quad f(x) \quad \text{s.t.} \quad g(x, t) \le 0, t \in \Omega,$$

where

$$f(x) = \begin{cases} x^2, & \text{if } x \ge 0, \\ -x^{\frac{5}{3}}, & \text{otherwise,} \end{cases} \quad g(x, t) = 2tx - t^2 \text{ and } \Omega = [-1, 1].$$

Then, for $x \le 0$ sufficiently small,

$$\int_\Omega g_+(x, t)^{1/2} \, dt = \int_{2x}^0 (2tx - t^2)^{1/2} \, dt = \int_{-1}^1 x^2(1 - s^2)^{1/2} \, ds = \frac{\pi}{2}x^2,$$

where

$$t = (1 - s)x, \text{ and } \left(\int_\Omega g_+(x, t) \, dt\right)^{1/2} = \left(-\frac{4}{3}x^3\right)^{1/2}.$$

Thus it follows that $\bar{F}^{1/2}_{int}$ is exact at $x^* = 0$ and $F^{1/2}_{int}$ is not.

9

Clearly,

$$A_x = \begin{cases} \Omega_x, & \text{if } x \notin X, \\ \Omega, & \text{otherwise.} \end{cases}$$

Since $\mu(\Omega) > 0$ by the assumption $\text{supp}(\mu) = \Omega$, the condition (15) can be rewritten as

$$\liminf_{x \to x^*} \mu(A_x) = \liminf_{\substack{x \to x^* \\ x \notin X}} \mu(\Omega_x) > 0.$$

The following example shows that condition (15) in Proposition 3.1 is sufficient, not necessary, and that the extended MFCQ in Proposition 3.2 is also only sufficient for F^1_{max} to be an exact penalty function.

Example 3.2. *Consider the following linear SIP problem*

$$\min \quad x_1 \quad s.t. \quad tx_1 + t^3 x_2 \le 0, t \in [-1, 1].$$

It is easy to see that $X = \{(0,0)\}$, the point $x^* = (0,0)$ is the optimal solution and $\Omega_{x^*} = [-1, 1]$. It is obvious that $\liminf\limits_{x \to x^*} \mu(A_x) = 0$, i.e., condition (15) doesn't hold, and that the extended MFCQ does not hold. For the latter, we have $\nabla_x g(x^*, t)^T h = th_1 + t^3 h_2 = 0 \not< 0$ when $t = 0 \in \Omega_{x^*}$. Next, we check the exactness of integral-type penalty function

$$F^1_{int}(x_1, x_2) := x_1 + \rho \int_{-1}^{1} [tx_1 + t^3 x_2]_+ \, dt.$$

It suffices to consider the area $\{(x_1, x_2) : x_1 < 0\}$. For $x_1 < 0, 0 < -\frac{x_1}{x_2} \le 1$,

$$F^1_{int}(x_1, x_2) = x_1 + \rho \left(\int_{-\sqrt{-\frac{x_1}{x_2}}}^{0} + \int_{\sqrt{-\frac{x_1}{x_2}}}^{1} \right) tx_1 + t^3 x_2 \, dt = x_1 \left[1 - \frac{\rho}{2} \left(-1 - \frac{x_2}{2x_1} - \frac{x_1}{x_2} \right) \right].$$

Since $\frac{1}{2s} + s \ge \sqrt{2}$, for $s \in (0, 1]$, $F^1_{int}(x_1, x_2) \ge 0$ for ρ large enough. For $x_1 < 0, 1 < -\frac{x_1}{x_2}$, or $x_1 < 0, x_2 < 0$,

$$F^1_{int}(x_1, x_2) = x_1 + \rho \int_{-1}^{0} tx_1 + t^3 x_2 \, dt = x_1 \left[1 - \rho \left(\frac{1}{2} - \frac{1}{4} \left(-\frac{x_2}{x_1} \right) \right) \right].$$

In both above cases, for some $\rho > 0$, $F^1_{int} \ge 0$ holds. Thus F^1_{int} is exact at x^*, and so is F^1_{max}, by Proposition 3.1 (ii).

4 Optimality Conditions of SIP

In this section, we shall employ the exactness of F^p_{max}, F^p_{int} and $\bar{F}^p_{int}$ ($p \in \,]0, 1]$) to develop optimality conditions (6) of SIP. From Theorem 3.1, we know that exactness of F^p_{max} is weaker than that of $\bar{F}^p_{int}$ and the exactness of $\bar{F}^p_{int}$ is weaker than that of F^p_{int}. Let x^* be a locally optimal solution of SIP.

Two cases are treated separately according to the different value p. We consider first the case when $p = 1$.

(SIP) can be rewritten as

$$\min \quad f(x) \quad \text{s.t.} \quad \max_{t \in \Omega} g(x,t) \le 0. \tag{19}$$

The exactness of F^1_{max} is equivalent to saying that problem (19) has an l_1 exact penalty function in the usual sense, see Clarke [21]. By using the calculus of nonsmooth analysis, see e.g. Theorem 2.8.6 in [21], it is easy to prove that the exactness of F^1_{max} guarantees the KKT-type optimality condition (3), so does the exactness of F^1_{int} or that of $\bar{F}^1_{max}$.

Below is an example to show that the exactness of F^1_{max} is strictly weaker than that of F^1_{int}.

Example 4.1. *Consider the SIP problem with*

$$f(x) = -x_1 - x_2, g(x,t) = tx_1 + t^2 x_2, \Omega = [0, 1].$$

Then the feasible set is $X = \{x \in \mathbb{R}^2 : x_1 + x_2 \le 0, x_1 \le 0\}$ and thus $x^* = (0,0)$ is the optimal solution. We also have that

$$\max_{t \in [0,1]} g_+(x,t) = \begin{cases} x_1 + x_2, & \text{if } x_1 + x_2 \ge 0, x_1 + 2x_2 \ge 0, \\ 0, & \text{if } x_1 + x_2 \le 0, x_1 \le 0, \\ -\frac{x_1^2}{4x_2}, & \text{otherwise.} \end{cases}$$

Noting that $-x_1 - x_2 - \frac{x_1^2}{4x_2} = -\frac{(2x_2 + x_1)^2}{4x_2}$, it is easy to see that for $\rho \ge 1$, F^1_{max} is exact at x^*. x^* is a KKT point as well: by taking $\lambda = 1, t = 1, (-1, -1) + 1 \cdot (1, 1) = 0$. However, for $x \in \{x : x_1 \le 0, x_1 + x_2 \ge 0\}$,

$$\begin{aligned} F^1_{int}(x) &= -x_1 - x_2 + \rho \left(\frac{1}{2} x_1 + \frac{1}{3} x_2 - \frac{1}{6} \frac{x_1^3}{x_2^2} \right) \\ &= \frac{\rho}{6} \left(x_1 - \frac{x_1^3}{x_2^2} \right) + (\frac{\rho}{3} - 1)(x_1 + x_2) \\ &= \frac{(x_1 + x_2)}{6} \left((2\rho - 6) + \rho \frac{x_1}{x_2} - \rho \frac{x_1^2}{x_2^2} \right). \end{aligned}$$

So, by taking $x = (-s + s^2, s), s \downarrow 0$, we know that for ρ large enough $F^1_{int}(x) < 0$ and thus F^1_{int} is not exact at x^*.

Compared with $p = 1$, the case of $p \in]0, 1[$ needs more effort. First, we briefly describe some notation that will be used in the sequel. Given a continuous function $h : \mathbb{R}^n \to \mathbb{R}$, the upper Dini-directional derivative of h at a point x in the direction $d \in \mathbb{R}^n$ is defined by

$$D_+ h(x; d) := \limsup_{\lambda \downarrow 0} \frac{h(x + \lambda d) - h(x)}{\lambda}.$$

A function $h : \mathbb{R}^n \to \mathbb{R}$ is called to be $C^{1,1}$ if it is continuously differentiable and its derivative is locally Lipschitz function on $\mathbb{R}^n$. The generalized upper and lower second-order directional derivatives of a $C^{1,1}$ function h at x in the direction $d \in \mathbb{R}^n$ are defined, respectively, by

$$h^{\circ\circ}(x; d) := \limsup_{y\to x, \lambda\downarrow 0} \frac{\langle \nabla h(y + \lambda d), d\rangle - \langle \nabla h(y), d\rangle}{\lambda},$$

$$h_{\circ\circ}(x; d) := \liminf_{y\to x, \lambda\downarrow 0} \frac{\langle \nabla h(y + \lambda d), d\rangle - \langle \nabla h(y), d\rangle}{\lambda},$$

see [38–40]. Recall $D(x) = \{0 \neq d \in R^n : \langle \nabla_x g(x,t), d\rangle \leq 0 \ \forall t \in \Omega_x\}$ and let

$$\Omega_x^{=}(d) := \{t \in \Omega_x : \langle \nabla_x g(x,t), d\rangle = 0\},$$
$$\Omega_x^{<}(d) := \{t \in \Omega_x : \langle \nabla_x g(x,t), d\rangle < 0\},$$

where $D(x)$ is called the cone of feasible directions for the linearized constraint set for (SIP), see [36].

Before stating our main result, we present two lemmas, which are needed in our proof, one of which is compiled from [23] for the estimates of the upper Dini-directional derivative $D_+h(x, t; \cdot)$ when h is a pth order penalty term.

Lemma 4.1. (Fatou's Lemma [35, Chapter 11, Corollary 5.7]) *Let* $(Z, \mathfrak{M}, \mu)$ *be a measure space and* $f_k : Z \to [0, +\infty[$ *be integrable for each k. Assume that*

$$\liminf_{k\to\infty} \int_Z |f_k(t)| \, d\mu(t)$$

exists. Then

$$\int_Z \liminf_{k\to\infty} f_k(t) \, d\mu(t) \leq \liminf_{k\to\infty} \int_Z f_k(t) \, d\mu(t).$$

Lemma 4.2. [23] *Let* $\bar{h}(x) = (\max\{h(x), 0\})^p$ *with* $p \in \,]0, 1[$ *and h be continuously differentiable at x.*

(i) *If* $h(x) < 0$*, then* $D_+\bar{h}(x; d) = 0$*;*

(ii) *If* $h(x) = 0$ *and* $\langle \nabla h(x), d\rangle < 0$*, then* $D_+\bar{h}(x; d) = 0$*;*

(iii) *If* $p = 0.5$*,* $h(x) = 0$ *and* $\langle \nabla h(x), d\rangle = 0$*, then* $D_+\bar{h}(x; d) \leq \sqrt{\max\{\frac{1}{2}h^{\circ\circ}(x; d), 0\}}$*;*

(iv) *If* $p \in \,]0.5, 1[$*,* $h(x) = 0$*,* $\langle \nabla h(x), d\rangle = 0$ *and* $h^{\circ\circ}(x; d)$ *is finite, then* $D_+\bar{h}(x; d) = 0$*;*

(v) *If* $p \in \,]0, 0.5[$*,* $h(x) = 0$*,* $\langle \nabla h(x), d\rangle = 0$ *and* $h^{\circ\circ}(x; d) < 0$*, then* $D_+\bar{h}(x; d) = 0$*.*

Now we establish a necessary optimality condition for SIP by virtue of the exact penalty function F_{int}^p.

Theorem 4.1. *Let $p \in]0, 1[$ and F^p_{int} be exact at x^*. Under any one of the three assumptions below,*
(i) $p = 0.5$ and $g^{\circ\circ}(x^*, t; d) \leq 0$ for all $t \in \Omega^=_{x^*}(d)$ and $d \in D(x^*)$,
(ii) $p \in]0.5, 1[$ and $g(\cdot, t)$ is $C^{1,1}$, for all $t \in \Omega^=_{x^*}(d)$, and
(iii) $p \in]0, 0.5[$ and $g^{\circ\circ}(x^*, t; d) < 0$, for all $t \in \Omega^=_{x^*}(d)$ and $d \in D(x^*)$,
we have

$$\langle \nabla f(x^*), d \rangle \geq 0, \forall d \in D(x^*). \tag{20}$$

Proof: For $p \in]0, 1[$ and $d \in D(x^*)$, we have

$$0 \leq D_+ F^p_{int}(x^*; d) = \langle \nabla f(x^*), d \rangle + \rho \limsup_{\lambda \downarrow 0} \int_\Omega \frac{g^p_+(x^* + \lambda d, t)}{\lambda} \, d\mu(t). \tag{21}$$

Firstly, we claim that for any $t \in \Omega$,

$$\frac{g^p_+(x^* + \lambda d, t)}{\lambda} \to 0 \quad \text{as } \lambda \downarrow 0. \tag{22}$$

Note that $\Omega = (\Omega \backslash \Omega_{x^*}) \cup \Omega^<_{x^*}(d) \cup \Omega^=_{x^*}(d)$. For $t \in \Omega \backslash \Omega_{x^*}$, $g(x^*, t) < 0$. Thus by (a) of Lemma 4.2,

$$D_+ g^p_+(x^*, t; d) = 0.$$

For $t \in \Omega^<_{x^*}(d)$, $g(x^*, t) = 0$ and $\langle \nabla g(x^*, t), d \rangle < 0$. Thus, by (b) of Lemma 4.2, we still have

$$D_+ g^p_+(x^*, t; d) = 0.$$

For the case $t \in \Omega^=_{x^*}(d)$, that is $g(x^*, t) = \langle \nabla g(x^*, t), d \rangle = 0$, we consider the following three subcases:
Subcase (i) $p = 0.5$. By (c) of Lemma 4.2 and the assumption, we have

$$0 \leq D_+ g^p_+(x^*, t; d) \leq \sqrt{\max\{\frac{1}{2} g^{\circ\circ}(x^*, t; d), 0\}} = 0.$$

Subcase (ii) $p \in]0.5, 1[$. By (d) of Lemma 4.2 and the assumption $g(\cdot, t)$ is $C^{1,1}$, we have $D_+ g^p_+(x^*, t; d) = 0$.
Subcase (iii) $p \in]0, 0.5[$. By (e) of Lemma 4.2 and the assumption $g^{\circ\circ}(x^*, t; d) < 0$, we have $D_+ g^p_+(x^*, t; d) = 0$.

In all, we have $D_+ g^p_+(x^*, t; d) = 0$ for any $t \in \Omega$. Noting the non-negativity of $\frac{g^p_+(x^* + \lambda d, t)}{\lambda}$, thus we obtain the claim (22).

Secondly, we claim that for any constant $\epsilon > 0$ there exists $\lambda^* > 0$ such that

$$\frac{g^p_+(x^* + \lambda d, t)}{\lambda} \leq \epsilon, \tag{23}$$

holds for any $(t, \lambda) \in \Omega \times]0, \lambda^*[$. Indeed, by the continuity of $\frac{g^p_+(x^* + \lambda d, t)}{\lambda}$ with respect to (t, λ), for any $t \in \Omega$, there exist a neighborhood V_t of t and $\lambda_t > 0$ such that $\frac{g^p_+(x^* + \lambda d, t')}{\lambda} \leq \epsilon$, for all

13

$(t', \lambda) \in V_t \times]0, \lambda_t[$. By the compactness of Ω, there are $t^1, \cdots, t^k$ belonging to Ω such that $\Omega \subset \bigcup_k V_{t^k}$. By letting $\lambda^* = \min_{1 \le i \le k} \lambda_{t^i}$, we obtain the claim (23).

By applying Fatou's Lemma to the sequence of functions $\left\{ \epsilon - \frac{g_+^p(x^* + \lambda d, t)}{\lambda} : \lambda \downarrow 0 \right\}$, we have

$$\liminf_{\lambda \downarrow 0} \int_\Omega \left(\epsilon - \frac{g_+^p(x^* + \lambda d, t)}{\lambda} \right) d\mu(t) \ge \mu(\Omega)\epsilon - \int_\Omega \limsup_{\lambda \downarrow 0} \frac{g_+^p(x^* + \lambda d, t)}{\lambda} d\mu(t)$$

$$= \mu(\Omega)\epsilon - \int_\Omega D_+ g_+^p(x^*, t; d) \, d\mu(t)$$

$$= \mu(\Omega)\epsilon.$$

So the last term in the right hand side equality (21) vanishes. Thus we have proved

$$\langle \nabla f(x^*), d \rangle \ge 0, \quad \forall d \in D(x^*).$$

$\square$

Remark 4.1. *It should be noted that when $p \in]0, 1[$, we need to use integral-type exact penalty functions, instead of max-type exact penalty functions, to deal with necessary optimality conditions of SIP. The reason why we do so can be shown by the following example.*

Example 4.2. *Consider the following linear SIP*

$$\min \quad x \quad s.t. \quad 2tx - t^2 \le 0, \quad t \in [-1, 1].$$

The optimal solution is $x^* = 0$, $\Omega_{x^*} = \{0\}$ and $D(x^*) = \{0 \ne d \in \mathbb{R}\}$. It is easy to see that $F_{max}^{\frac{1}{2}}$ is exact at x^*. We now show that $F_{int}^{\frac{1}{2}}$ is not exact at x^*. Indeed, for $x \in]-\frac{1}{2}, 0[$ sufficiently small, we have

$$F_{int}^{\frac{1}{2}}(x) = x + \rho \int_{-1}^1 (2tx - t^2)_+^{\frac{1}{2}} dt = x + \rho \int_{2x}^0 (2tx - t^2)^{\frac{1}{2}} dt = x + \frac{\rho\pi}{2} x^2 < 0 = F_{int}^{\frac{1}{2}}(0).$$

Since the constraint function is linear, then the second-order condition $g^{\circ\circ}(x^*, t; d) = 0$ is true for all t and d. But

$$\langle \nabla f(x^*), d \rangle < 0,$$

whenever $d \in \mathbb{R}_-$. So, for this example, we cannot develop the optimality conditions by only assuming the exactness of $F_{max}^{\frac{1}{2}}$ and the second-order conditions presented in Theorem 4.1.

Remark 4.2. *The following example modified from one in [23] shows that extended Abadie CQ is satisfied but the assumptions of Theorem 4.1 are not. On the other hand, the situation that the extended Abadie CQ is not satisfied but the assumptions in Theorem 4.1 are satisfied is illustrated in Example 4.4.*

Example 4.3. *Consider the following SIP problem*

$$\min \quad x \quad s.t. \quad t(x^2 - x) \le 0, t \in [1, 2].$$

14

Then $X = [0, 1]$, and $x^* = 0$ is the unique optimal solution. It is easy to show that

$$\Omega_{x^*} = [1, 2], \ T_X(x^*) = (\text{cone } \cup_{t \in \Omega_{x^*}} \nabla_x g(x^*, t))^\circ = [0, \infty[,$$

$$g^{\circ\circ}(x^*, t; d) = 2td^2 \ \text{and} \ D(x^*) =]0, \infty[.$$

Thus the extended Abadie CQ is satisfied. We have that

$$F_{int}^{\frac{1}{2}}(x) = x + \left(\rho \int_1^2 t^{\frac{1}{2}} \, dt\right)[(x^2 - x)]_+^{\frac{1}{2}} = \frac{(1 - \hat{\rho}^2)x^2 + \hat{\rho}^2 x}{x - \sqrt{x^2 - x}} \geq 0,$$

for all sufficiently small $x < 0$, where $\hat{\rho} = \rho \int_1^2 t^{\frac{1}{2}} \, dt$, and that

$$F_{int}^{\frac{1}{2}}(x) \geq 0,$$

for $x \geq 0$ and thus $F_{int}^{\frac{1}{2}}$ is exact, but the assumption (i) of Theorem 4.1 is not satisfied.

Example 4.4. *Consider the following SIP problem*

$$\min \quad f(x) = x^3 \quad \text{s.t.} \quad g(x, t) = tx^6 \leq 0, \ t \in \Omega = [0, 1].$$

The optimal solution is $x^* = 0$. The extended Abadie CQ is invalid at x^*, since $\nabla_x g(x^*, t) = 0$ for all $t \in \Omega$ and

$$\{0\} = T_X(x^*) \neq D(x^*) = \mathbb{R} \setminus \{0\}.$$

Theorem 4.1 is applicable for the case of $p = \frac{1}{2}$. In fact, by a simple calculation, we have for $x \geq 0$,

$$F_{int}^{\frac{1}{2}}(x) = x^3 + \rho \int_0^1 \max\{tx^6, 0\}^{\frac{1}{2}} \, dt \geq 0;$$

and for $x < 0$,

$$F_{int}^{\frac{1}{2}}(x) = x^3 + \rho \int_0^1 \max\{tx^6, 0\}^{\frac{1}{2}} \, dt = (1 - \rho \int_0^1 t^{\frac{1}{2}} \, dt)x^3 = (1 - \frac{2}{3}\rho)x^3 \geq 0,$$

whenever $\rho \geq \frac{3}{2}$. So, $F_{int}^{\frac{1}{2}}$ is an exact penalty function at x^*. It is clear that

$$g^{\circ\circ}(x^*, t; d) = 0 \leq 0,$$

for all $t \in \Omega$ and $d \in \mathbb{R}$. Thus the assumption (i) of Theorem 4.1 is satisfied. Indeed, KKT condition (3) also holds for this example.

Next we employ the exactness of $\bar{F}_{int}(p \in]0, 1[)$ to develop the optimality condition (20) of SIP. We first give a proposition that is needed in the proof of the necessary optimality conditions (20) of SIP.

Proposition 4.1. *Let $g: \mathbb{R}_+ \to \mathbb{R}_+$ be a non-negative function, $f: \mathbb{R}_+ \to \mathbb{R}$ be a continuous and strictly increasing function and $\lambda_0 \in \mathbb{R}_+$. Then*

$$\limsup_{\lambda \to \lambda_0} f(g(\lambda)) \leq f(\limsup_{\lambda \to \lambda_0} g(\lambda)). \tag{24}$$

Proof: Let $\{\lambda^k\}$ be a sequence such that $\lim_{k\to\infty} \lambda_k = \lambda_0$ and

$$\limsup_{\lambda\to\lambda_0} f(g(\lambda)) = \lim_{k\to\infty} f(g(\lambda^k)).$$

Then $\limsup_{k\to\infty} g(\lambda^k) \le \limsup_{\lambda\to\lambda_0} g(\lambda)$. As the limit $\lim_{k\to\infty} f(g(\lambda^k))$ exists and f is strictly increasing, $\lim_{k\to\infty} g(\lambda^k)$ exists. If $\lim_{k\to\infty} g(\lambda^k) = \infty$, then $\limsup_{\lambda\to\lambda_0} g(\lambda) = \infty$. Thus (24) holds. Assume now that $\lim_{k\to\infty} g(\lambda^k) < \infty$. By the continuity and monotonicity of f,

$$\lim_{k\to\infty} f(g(\lambda^k)) = f(\lim_{k\to\infty} g(\lambda^k)) \le f(\limsup_{\lambda\to\lambda_0} g(\lambda)),$$

and the assertion (24) holds. $\qquad\square$

Theorem 4.2. *Let $p \in {]}0,1{[}$ and $\bar{F}_{int}^p$ be exact at x^*. Under any one of the three assumptions in Theorem 4.1, the necessary optimality condition* (20) *holds.*

Proof: Let $d \in D(x^*)$. We have that

$$
\begin{aligned}
0 &\le D_+\bar{F}_{int}^p(x^*, d) \\
&= \langle \nabla f(x^*), d \rangle + \rho \limsup_{\lambda\downarrow 0} \frac{(\int_\Omega g_+(x^* + \lambda d, t)\, \mathrm{d}\mu(t))^p}{\lambda} \\
&\le \langle \nabla f(x^*), d \rangle + \rho \left(\limsup_{\lambda\downarrow 0} \frac{\int_\Omega g_+(x^* + \lambda d, t)\, \mathrm{d}\mu(t)}{\lambda^{1/p}} \right)^p,
\end{aligned}
\tag{25}
$$

where the last inequality follows from Proposition 4.1. Note that

$$\limsup_{\lambda\downarrow 0} \frac{\int_\Omega g_+(x^* + \lambda d, t)\, \mathrm{d}\mu(t)}{\lambda^{1/p}} = \limsup_{\lambda\downarrow 0} \int_\Omega \left(\frac{g_+^p(x^* + \lambda d, t)}{\lambda} \right)^{1/p} \mathrm{d}\mu(t).$$

Following exactly the proof in Theorem 4.1, we have for any $t \in \Omega$,

$$\frac{g_+^p(x^* + \lambda d, t)}{\lambda} \to 0 \quad \text{as } \lambda \to 0.$$

and that for any constant $\epsilon > 0$ there exists $\lambda^* > 0$ such that for any $(t, \lambda) \in \Omega \times {]}0, \lambda^*{[}$,

$$\frac{g_+^p(x^* + \lambda d, t)}{\lambda} \le \epsilon.$$

By applying Fatou's Lemma to the sequence $\left\{ \epsilon^{1/p} - \left(\frac{g_+^p(x^* + \lambda d, t)}{\lambda} \right)^{1/p} \right\}$ and Proposition 4.1, we

16

have

$$\liminf_{\lambda\downarrow 0} \int_\Omega \left(\epsilon^{1/p} - \left(\frac{g^p_+(x^* + \lambda d, t)}{\lambda} \right)^{1/p} \right) \mathrm{d}\mu(t)$$

$$\geq \mu(\Omega)\epsilon^{1/p} - \int_\Omega \limsup_{\lambda\downarrow 0} \left(\frac{g^p_+(x^* + \lambda d, t)}{\lambda} \right)^{1/p} \mathrm{d}\mu(t)$$

$$\geq \mu(\Omega)\epsilon^{1/p} - \int_\Omega \left(\limsup_{\lambda\downarrow 0} \frac{g^p_+(x^* + \lambda d, t)}{\lambda} \right)^{1/p} \mathrm{d}\mu(t)$$

$$= \mu(\Omega)\epsilon^{1/p} - \int_\Omega \left(D_+ g^p_+(x^*, t; d) \right)^{1/p} \mathrm{d}\mu(t)$$

$$= \mu(\Omega)\epsilon^{1/p}.$$

Thus the second term of the last equation in (25) vanishes. This completes the proof. □

5   Optimality Conditions of GSIP

Assume that $\Omega \cap \Omega(x) \neq \emptyset$ for all $x \in \mathbb{R}^n$. As in Polak and Royset [33], we will associate (GSIP) with an SIP problem via augmented Lagrangian of the lower level problem and derive optimality conditions for (GSIP) by applying Theorems 4.1 to the resulted SIP.

Let val(GSIP) be the optimal value of the problem (GSIP), $X_{(\text{GSIP})}$ be the feasible set of (GSIP) and $\hat{x}$ be a locally optimal solution of (GSIP). Let $x \in \mathbb{R}^n$. The lower level problem associated with (GSIP) is

$$Q(x) \qquad \max_{t\in\Omega} \quad g(x, t) \quad \text{s.t.} \quad v_i(x, t) \leq 0, i = 1, \cdots, l.$$

Let val$Q(x)$ be the optimal value of the problem $Q(x)$. Clearly

$$x \in X_{(\text{GSIP})} \quad \text{iff} \quad \text{val}Q(x) \leq 0.$$

Let $\bar{f}(x, \mu, c) := f(x)$ and

$$\bar{g}(x, t, \mu, c) := g(x, t) - \frac{1}{2c} \sum_{i=1}^l \{([cv_i(x, t) + \mu_i]_+)^2 - \mu_i^2\}, (x, \mu, c) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}_{++}.$$

Consider the following SIP problem, denoted as (SIPg),

$$\min_{(x,\mu,c)\in\mathbb{R}^n\times\mathbb{R}^l\times\mathbb{R}_{++}} \quad \bar{f}(x, \mu, c) \quad \text{s.t.} \quad \bar{g}(x, t, \mu, c) \leq 0, t \in \Omega.$$

Let val(SIPg) be the optimal value of (SIPg) and $X_{(\text{SIP}g)}$ be the feasible set of (SIPg).

We will show the equivalence between (GSIP) and (SIPg) in Proposition 5.1. For a fixed $x \in \mathbb{R}^n$, it is clear that $\bar{g}(x, t, \mu, c)$ is the classical augmented Lagrangian of the lower level problem $Q(x)$, see [27]. There are other generalized augmented Lagrangians that may be used to

reformulate (GSIP) as an SIP, see [41, 42]. The reason that we choose to use the classical augmented Lagrangian is that it is a $C^{1,1}$ function, see [39], as long as $g(\cdot, t)$ and each $v_i(\cdot, t)$ is twice continuously differentiable. It is worth noting that the $C^{1,1}$ property plays a crucial role in the establishment of optimality conditions when using lower order exact penalty functions, see [23].

Furthermore, for function $\alpha(x) := [\max\{\beta(x), 0\}]^2$, where β is a C^2 function, we have

$$\nabla \alpha(x) = 2\max\{\beta(x), 0\}\nabla\beta(x), \tag{26}$$

$$\alpha_{\circ\circ}(x; d) = \begin{cases} 0, & \text{if } \beta(x) \leq 0, \\ 2[\beta(x)d^T\nabla^2\beta(x)d + \langle\nabla\beta(x), d\rangle^2], & \text{if } \beta(x) > 0, \end{cases} \tag{27}$$

see [39].

Before proceeding, we will investigate the relations between the optimal solutions of (GSIP) and (SIPg). Let $\bar{H}(x, \mu, c) := \max_{t\in\Omega} \bar{g}(x, t, \mu, c)$. It is easy to see that (x, μ, c) is feasible for (SIPg) if $\bar{H}(x, \mu, c) \leq 0$ and that

$$\bar{H}(x, \mu, c) \geq \text{val}Q(x), \ \forall(x, \mu, c) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}_{++}.$$

Thus we obtain the relation between the optimal values of (GSIP) and (SIPg) as follows

$$\text{val(SIP}g) \geq \text{val(GSIP)}.$$

Let $v(x, u) = \max_{t\in\Omega} \hat{g}(x, t, u)$, where

$$\hat{g}(x, t, u) := \begin{cases} g(x, t), & \text{if } v(x, t) \leq u, \\ -\infty, & \text{otherwise.} \end{cases}$$

Then $v(x, 0)$ is the optimal value of the lower level problem $Q(x)$.

Next we recall some concepts from Rockafellar [27]. Problem $Q(x)$ is said to satisfy the *quadratic growth condition* iff there is a $c \geq 0$ such that $\bar{g}(x, t, 0, c)$ is bounded above as a function of $t \in \Omega$. Problem $Q(x)$ is said to be *stable of degree* 2 iff there is a neighborhood U of the origin in $\mathbb{R}^l$ and a C^2 function $\pi_x : U \to \mathbb{R}$ such that

$$v(x, u) \leq \pi_x(u), \forall u \in U, \text{ and } v(x, 0) = \pi_x(0).$$

Lemma 5.1. (Rockafellar [27]) *Under the quadratic growth condition of $Q(x)$, we have*

$$\text{val } Q(x) = \min_{(\mu, c)\in\mathbb{R}^l\times\mathbb{R}_{++}} \bar{H}(x, \mu, c)$$

iff the problem $Q(x)$ is stable of degree 2.

Therefore we have

Proposition 5.1. *Assume that, for all $x \in \mathbb{R}^n$, $Q(x)$ satisfies the quadratic growth condition and is stable of degree 2. Then problems (GSIP) and (SIPg) have the same optimal value, i.e., val(GSIP) = val(SIPg), and furthermore,*
(i) if $\hat{x}$ is a locally optimal solution of (GSIP), then there exists $(\hat{\mu}, \hat{c}) \in \mathbb{R}^l\times\mathbb{R}_{++}$ such that $(\hat{x}, \hat{\mu}, \hat{c})$ is a locally optimal solution of (SIPg);
(ii) if $(\hat{x}, \hat{\mu}, \hat{c})$ is a locally optimal solution of (SIPg), then $\hat{x}$ is a locally optimal solution of (GSIP).

Proof: (i) Let $\hat{x}$ be a locally optimal solution of (GSIP). There is a neighborhood $W(\hat{x})$ of $\hat{x}$ such that $f(x) \geq f(\hat{x})$ for all $x \in W(\hat{x}) \cap X_{(\text{GSIP})}$. For any $(x, \mu, c) \in W(\hat{x}) \times \mathbb{R}^l \times \mathbb{R}_{++}$ satisfying $\bar{H}(x, \mu, c) \leq 0$, by Lemma 5.1, there exists $(\mu_x, c_x) \in \mathbb{R}^l \times \mathbb{R}_{++}$ such that

$$\bar{H}(x, \mu, c) \geq \bar{H}(x, \mu_x, c_x) = \text{val}Q(x).$$

Then $\text{val}Q(x) \leq 0$. That is $x \in X_{(\text{GSIP})}$. Especially, for $\hat{x}$ there is $(\hat{\mu}, \hat{c})$ satisfying

$$\bar{H}(\hat{x}, \hat{\mu}, \hat{c}) = \text{val}Q(\hat{x}) \leq 0.$$

Thus $(\hat{x}, \hat{\mu}, \hat{c}) \in X_{SIP_g}$. We obtain $\bar{f}(x, \mu, c) = f(x) \geq f(\hat{x}) = \bar{f}(\hat{x}, \hat{\mu}, \hat{c})$ and this completes the proof.

(ii) The proof proceeds similarly. For any $x \in W(\hat{x})$ satisfying $\text{val}Q(x) \leq 0$, there is (μ_x, c_x) such that $\bar{H}(x, \mu_x, c_x) = \text{val}Q(x) \leq 0$, that is (x, μ_x, c_x) is feasible for (SIPg). Then

$$f(x) = \bar{f}(x, \mu_x, c_x) \geq \bar{f}(\hat{x}, \hat{\mu}, \hat{c}) = f(\hat{x}).$$

Thus

$$\text{val(GSIP)} = \text{val(SIPg)}.$$

$\square$

Under the assumptions of Proposition 5.1, we transform (GSIP) into an equivalent SIP problem (SIPg). For $(x, \mu, c) \in X_{(\text{SIP}_g)}$, let $\Omega_{(x,\mu,c)} = \{t \in \Omega : \bar{g}(x, t, \mu, c) = 0\}$ be the index set of active constraints of val(SIPg) at (x, μ, c).

It follows from (26) that we have

$$\nabla_x \bar{g}(x, t, \mu, c) = \nabla_x g(x, t) - \nabla_x^T v(x, t)[cv(x, t) + \mu]_+,$$

$$\nabla_\mu \bar{g}(x, t, \mu, c) = -\frac{1}{c}([cv(x, t) + \mu]_+ - \mu),$$

$$\nabla_c \bar{g}(x, t, \mu, c) = \frac{1}{2c^2} \sum_{i=1}^{l} \{([cv_i(x, t) + \mu_i]_+)^2 - \mu_i^2 - 2c[cv_i(x, t) + \mu_i]_+ v_i(x, t)\}.$$

Let $(x, \mu, c) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}_{++}$ and $t \in \Omega$. Define the index sets

$$\hat{I}_{(x,\mu,c)}^+(t) := \{i \in \{1, \cdots, l\} : cv_i(x, t) + \mu_i > 0\},$$

$$\hat{I}_{(x,\mu,c)}^0(t) := \{i \in \{1, \cdots, l\} : cv_i(x, t) + \mu_i = 0\},$$

and the cone of feasible directions for the linearized constraint set for (SIPg) by

$$D(x, \mu, c) := \{0 \neq d = (d_1, d_2, d_3) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R} :$$
$$\langle \nabla_x \bar{g}(x, t, \mu, c), d_1 \rangle + \langle \nabla_\mu \bar{g}(x, t, \mu, c), d_2 \rangle + \nabla_c \bar{g}(x, t, \mu, c)d_3 \leq 0, \ \forall t \in \Omega_{(x,\mu,c)}\}.$$

Lemma 5.2. *Let $\hat{x}$ be a locally optimal solution of (GSIP) and the assumptions of Proposition 5.1 hold. Let $(\hat{\mu}, \hat{c})$ be the corresponding multiplier as in (1) of Proposition 5.1 such that $(\hat{x}, \hat{\mu}, \hat{c})$ is a locally optimal solution of (SIPg). Then the cone $D(\hat{x}, \hat{\mu}, \hat{c})$ is of the following form*

$$\{0 \neq d = (d_1, d_2, d_3) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R} : \langle \nabla_x g(\hat{x}, t) - \nabla_x^T v(\hat{x}, t)\hat{\mu}, d_1 \rangle \leq 0, \ \forall t \in \Omega_{(\hat{x},\hat{\mu},\hat{c})}\}. \tag{28}$$

Proof: Let $\tilde{t} \in \Omega(\hat{x}, \hat{\mu}, \hat{c})$. By the assumption that $(\hat{x}, \hat{\mu}, \hat{c})$ is a locally optimal solution of (SIPg), $\tilde{t}$ solves the problem $\max_{t \in \Omega} \bar{g}(\hat{x}, t, \hat{\mu}, \hat{c})$ for which the following hold

$$\min_{(\mu, c)} \max_{t \in \Omega} \bar{g}(\hat{x}, t, \mu, c) = \max_{t \in \Omega} \min_{(\mu, c)} \bar{g}(\hat{x}, t, \mu, c) = \mathrm{val}Q(\hat{x}) = \max_{t \in \Omega \cap \Omega(\hat{x})} g(\hat{x}, t).$$

Or equivalently, that $(\tilde{t}, \hat{\mu}, \hat{c})$ is a saddle point of the augmented Lagrangian $\bar{g}(\hat{x}, \cdot, \cdot, \cdot)$. Thus $\tilde{t}$ solves the lower level problem

$$\max\{g(\hat{x}, t) : t \in \Omega, v(\hat{x}, t) \leq 0\}$$

and augmenting Lagrange multiplier $\hat{\mu}$ is also a Lagrange multiplier of the lower level problem, see [27]. By the first order necessary optimality condition, we have

$$0 \in \nabla_t [g(\hat{x}, \hat{t}) - \hat{\mu}^T v(\hat{x}, \hat{t})] + N_\Omega(\hat{t}), \langle \hat{\mu}, v(\hat{x}, \hat{t}) \rangle = 0, v(\hat{x}, \hat{t}) \leq 0, \hat{\mu} \geq 0.$$

It is easy to see that $\hat{I}^+_{(\hat{x}, \hat{\mu}, \hat{c})}(\hat{t}) = \{i \in \{1, \cdots, l\} : \hat{\mu}_i > 0\}$ and

$$[\hat{c} v_i(\hat{x}, \hat{t}) + \hat{\mu}_i]_+ = \hat{\mu}_i, i = 1, \cdots, l.$$

Thus $\nabla_\mu \bar{g}(\hat{x}, \hat{t}, \hat{\mu}, \hat{c}) = 0, \nabla_c \bar{g}(\hat{x}, \hat{t}, \hat{\mu}, \hat{c}) = 0$. As $\tilde{t} \in \Omega_{(\hat{x}, \hat{\mu}, \hat{c})}$ is arbitrary, (28) holds. $\qquad \square$

Let $p = 1$ and let the assumptions of Proposition 5.1 hold and

$$G^1_{max}(x, \mu, c) := \bar{f}(x, \mu, c) + \max_{t \in \Omega} \bar{g}_+(x, t, \mu, c)$$

be exact at $(\hat{x}, \hat{\mu}, \hat{c})$ where the pair $(\hat{\mu}, \hat{c})$ is obtained from Proposition 5.2. Then it is easy to see from the case of (SIP), see Section 3, that the following KKT-type optimality condition holds:

$$0 \in \nabla f(\hat{x}) + \mathrm{cone}\{\nabla_x g(\hat{x}, t) - \nabla_x^T v(\hat{x}, t)\hat{\mu}, t \in \Omega_{(\hat{x}, \hat{\mu}, \hat{c})}\}.$$

For $p \in]0, 1[$, let the integral-type double penalty function of (SIPg) be defined by

$$G^p_{int}(x, \mu, c) := \bar{f}(x, \mu, c) + \rho \int_\Omega \bar{g}^p_+(x, t, \mu, c) \, d\mu(t).$$

By applying Theorem 4.1 and Lemma 5.2, we have.

Theorem 5.1. *Let the assumptions of Proposition 5.1 hold. Let $\hat{x}$ be a locally optimal solution of (GSIP) and G^p_{int} be exact at the point $(\hat{x}, \hat{\mu}, \hat{c})$ where the pair $(\hat{\mu}, \hat{c})$ is obtained from Proposition 5.1. Then, under one of the following assumptions,*
(i) $p \in]0.5, 1[$,
(ii) $p = 0.5$ and $\bar{g}^{\circ\circ}_{(x,\mu,c)}(\hat{x}, t, \hat{\mu}, \hat{c}; d) \leq 0$ for all $d \in D(\hat{x}, \hat{\mu}, \hat{c})$ and $t \in \Omega_{(\hat{x}, \hat{\mu}, \hat{c})}$ with

  $\langle \nabla_{(x,\mu,c)} \bar{g}(\hat{x}, t, \hat{\mu}, \hat{c}), d \rangle = 0$, and
(iii) $p \in]0, 0.5[$ and $\bar{g}^{\circ\circ}_{(x,\mu,c)}(\hat{x}, t, \hat{\mu}, \hat{c}; d) < 0$ for all $d \in D(\hat{x}, \hat{\mu}, \hat{c})$ and $t \in \Omega_{(\hat{x}, \hat{\mu}, \hat{c})}$ with

  $\langle \nabla_{(x,\mu,c)} \bar{g}(\hat{x}, t, \hat{\mu}, \hat{c}), d \rangle = 0$,

  we have

$$\langle \nabla f(\hat{x}), d_1 \rangle \geq 0, \tag{29}$$

for all $d_1 \in \mathbb{R}^n$ satisfying $\langle \nabla_x g(\hat{x}, t) - \nabla_x^T v(\hat{x}, t)\hat{\mu}, d_1 \rangle \leq 0, t \in \Omega_{(\hat{x}, \hat{\mu}, \hat{c})}$.

Remark 5.1. *If $(\hat{x}, \hat{\mu}, \hat{c})$ solves G_{int}^p, then so does $(\hat{x}, \hat{\mu}, c)$ for any $c \geq \hat{c}$. Since for any (x, t, μ) and $c_1 \leq c_2$, $\bar{g}(x, t, \mu, c_1) \geq \bar{g}(x, t, \mu, c_2)$, it follows that $G_{int}^p(x, \mu, c_1) \geq G_{int}^p(x, \mu, c_2)$.*

We now compute the generalized second-order directional derivative $\bar{g}_{(x,\mu,c)}^{\circ\circ}(\hat{x}, t, \hat{\mu}, \hat{c}; d)$ for $d \in D(\hat{x}, \hat{\mu}, \hat{c})$ and $t \in \Omega(\hat{x}, \hat{\mu}, \hat{c})$.

Lemma 5.3. *Let $d \in D(\hat{x}, \hat{\mu}, \hat{c})$ and $t \in \Omega(\hat{x}, \hat{\mu}, \hat{c})$. Then the following formula holds:*

$$
\bar{g}_{(x,\mu,c)}^{\circ\circ}(\hat{x}, t, \hat{\mu}, \hat{c}; d) = d_1^T [\nabla_{xx}^2 g(\hat{x}, t) - \sum_{i=1}^{l} \hat{\mu}_i \nabla_{xx}^2 v_i(\hat{x}, t)] d_1
$$

$$
- \sum_{i \in \hat{I}_{\hat{z}}^+(t)} (\sqrt{\hat{c}} d_1^T \nabla_x v_i(\hat{x}, t) + \frac{d_{2i}}{\sqrt{\hat{c}}})^2 + \sum_{i=1}^{l} \frac{d_{2i}^2}{\hat{c}}. \tag{30}
$$

Proof: By definition $\bar{g}(x, t, \mu, c)$, let

$$
\bar{g}(x, t, \mu, c) := k(x, t, \mu, c) + h(x, t, \mu, c), \quad (x, \mu, c) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}_{++}
$$

where

$$
k(x, t, \mu, c) := g(x, t) + \sum_{i=1}^{l} \frac{\mu_i^2}{2c}, \qquad h(x, t, \mu, c) = -\sum_{i=1}^{l} h_i(x, t, \mu, c),
$$

$$
h_i(x, t, \mu, c) := \frac{([cv_i(x, t) + \mu_i]_+)^2}{2c} \quad \text{for } i = 1, \cdots, l.
$$

For function k, k is a C^2 function and

$$
k_{(x,\mu,c)}^{\circ\circ}(\hat{x}, t, \hat{\mu}, \hat{c}; d) = d_1^T \nabla_{xx}^2 g(\hat{x}, t) d_1 + \frac{d_2^{\top} d_2}{\hat{c}} + \mu^T \mu \frac{d_3^2}{\hat{c}^3} - \sum_{i=1}^{l} 2\frac{\hat{\mu}_i d_{2i} d_3}{\hat{c}^2}.
$$

For function h, we claim that

$$
h^{\circ\circ}(\hat{x}, t, \hat{\mu}, \hat{c}; d) = \left(-\sum_{i=1}^{m} h_i \right)^{\circ\circ} (\hat{x}, t, \hat{\mu}, \hat{c}; d)
$$

$$
= -\sum_{i \in \hat{I}_{\hat{z}}^+(t)} (h_i)_{\circ\circ}(\hat{x}, t, \hat{\mu}, \hat{c}; d) = -\sum_{i \in \hat{I}_{\hat{z}}^+(t)} d^T \nabla^2 h_i(\hat{x}, t, \hat{\mu}, \hat{c}) d \tag{31}
$$

where the third equality follows from the fact that h_i is C^2 for $i \in \hat{I}^+(t)$. We now prove the second equality. On one hand,

$$
\left(-\sum_{i=1}^{m} h_i \right)^{\circ\circ} (\hat{x}, t, \hat{\mu}, \hat{c}; d) = -\left(\sum_{i=1}^{m} h_i \right)_{\circ\circ} (\hat{x}, t, \hat{\mu}, \hat{c}; d)
$$

$$
\leq -\sum_{i \in \hat{I}^+(t)} (h_i)_{\circ\circ}(\hat{x}, t, \hat{\mu}, \hat{c}; d) - \sum_{i \notin \hat{I}^+(t)} (h_i)_{\circ\circ}(\hat{x}, t, \hat{\mu}, \hat{c}; d)
$$

$$
= -\sum_{i \in \hat{I}^+(t)} (h_i)_{\circ\circ}(\hat{x}, t, \hat{\mu}, \hat{c}; d). \qquad \text{(by (27))}
$$

21

On the other hand, since the h_i's are C^2 for $i \notin \hat{I}^0(t)$, we have

$$\left(\sum_{i=1}^{m} h_i\right)_{\circ\circ}(\hat{x}, t, \hat{\mu}, \hat{c}; d) = \sum_{i \notin \hat{I}^0(t)} (h_i)_{\circ\circ}(\hat{x}, t, \hat{\mu}, \hat{c}; d) + \left(\sum_{i \in \hat{I}^0(t)} h_i\right)_{\circ\circ}(\hat{x}, t, \hat{\mu}, \hat{c}; d)$$

$$= \sum_{i \in \hat{I}^+(t)} (h_i)_{\circ\circ}(\hat{x}, t, \hat{\mu}, \hat{c}; d) + \left(\sum_{i \in \hat{I}^0(t)} h_i\right)_{\circ\circ}(\hat{x}, t, \hat{\mu}, \hat{c}; d), \tag{32}$$

where, for $i \in \hat{I}^+(t)$,

$$(h_i)_{\circ\circ}(\hat{x}, t, \hat{\mu}, \hat{c}; d) = \hat{\mu}_i d_1^T \nabla_x^2 v_i(\hat{x}, t) d_1 + \hat{c} \langle \nabla_x v_i(\hat{x}, t), d_1 \rangle^2 + \frac{d_{2i}^2}{\hat{c}} + \frac{\hat{\mu}_i^2 d_3^2}{\hat{c}^3}$$

$$+ 2d_{2i} \langle \nabla_x v_i(\hat{x}, t), d_1 \rangle - 2 \frac{\hat{\mu}_i d_{2i} d_3}{\hat{c}^2}.$$

Next, we claim that the second term of the right hand side of the equation (32) vanishes. By definition of second-order directional derivative, for any sequences $(x^\nu, \mu^\nu, c^\nu) \to (\hat{x}, \hat{\mu}, \hat{c})$ and $\lambda^\nu \downarrow 0$ as $\nu \to \infty$, the last term is no larger than

$$\liminf_{\nu \to \infty} \frac{\langle \nabla(\sum_{i \in \hat{I}^0(t)} h_i)((x^\nu, t, \mu^\nu, c^\nu) + \lambda^\nu d) - \nabla(\sum_{i \in \hat{I}^0(t)} h_i)(x^\nu, t, \mu^\nu, c^\nu), d \rangle}{\lambda^\nu},$$

This is further equal to zero by choosing any particular sequences $(x^\nu, \mu^\nu, c^\nu) \to (\hat{x}, \hat{\mu}, \hat{c})$ and $\lambda^\nu \downarrow 0$ such that for all $i \in \hat{I}^0(t)$,

$$\frac{c^\nu v_i(x^\nu, t) + \mu_i^\nu}{2c^\nu} < 0 \text{ and } \frac{(c^\nu + \lambda^\nu d_3) v_i(x^\nu + \lambda^\nu d_1, t) + (\mu_i^\nu + \lambda^\nu d_{2i})}{2(c^\nu + \lambda^\nu d_3)} < 0.$$

Indeed, the sequences can be chosen as follows. Let $(x^\nu, \mu^\nu, c^\nu) = (\hat{x}, \mu^\nu, \hat{c})$ with $\mu_i^\nu \uparrow \hat{\mu}_i$ for all $i \in \hat{I}^0(t)$ and $\mu_i^\nu = \hat{\mu}_i$ for the other i's. Then the first strict inequality is satisfied. This also means that the continuous function $F_i(x, \mu, c) := \frac{cv_i(x,t) + \mu_i}{2c}$ is strictly less than 0 at $(\hat{x}, \mu^\nu, \hat{c})$. Thus, for given $d = (d_1, d_2, d_3) \in D(\hat{x}, \hat{\mu}, \hat{c})$, there is $\lambda^\nu \downarrow 0$ such that $F_i(\hat{x} + \lambda^\nu d_1, \mu^\nu + \lambda^\nu d_2, c^\nu + \lambda^\nu d_3) < 0$. This is just the second strict inequality. Combining these two parts, the second term of the right hand side of the equation (32) vanishes. Thus formula (31) follows. The conclusion follows easily. □

We have the following corollary of Theorem 5.1.

Corollary 5.1. *Assume that the following conditions hold:*

(i) $G_{int}^{\frac{1}{2}}(x, \mu, c)$ is exact at $(\hat{x}, \hat{\mu}, \hat{c})$;

(ii) $g(\cdot, t)$ and $-v_i(\cdot, t)$ $(i = 1, \cdots, l)$ are concave for each $t \in \Omega$; and

(iii) $\hat{I}^+(t) = \{1, \cdots, l\}$ and $\langle \nabla_x v_i(\hat{x}, t), d_1 \rangle = 0$ for $d \in D(\hat{x}, \hat{\mu}, \hat{c}), t \in \Omega(\hat{x}, \hat{\mu}, \hat{c})$ and $i \in \hat{I}^+(t)$.

Then (29) holds.

Proof: Under our assumptions and the formula (30), we have that

$$\bar{g}^{\circ\circ}(\hat{x}, t, \hat{\mu}, \hat{c}; d) = d_1^T [\nabla_{xx}^2 g(\hat{x}, t) - \sum_{i=1}^{l} \hat{\mu}_i \nabla_{xx}^2 v_i(\hat{x}, t)] d_1.$$

By the assumed concavities, inequality $\bar{g}^{\circ\circ}(\hat{x}, t, \hat{\mu}, \hat{c}; d) \leq 0$ holds. So condition (ii) of Theorem 5.1 holds. Thus our assertion follows. □

Finally, we present below an example which verifies Theorem 5.1 (ii).

Example 5.1. *Consider the following GSIP problem*

$$\min_{x \in \mathbb{R}} \quad x^3 \quad s.t. \quad -t \leq 0, t \in \Omega \cap \Omega(x),$$

with $\Omega = [-1, 1], \Omega(x) = \{t \in \mathbb{R} : x^3 - t \leq 0\}$. Then $X = \{x : x \geq 0\}$ and the solution $\hat{x} = 0$ is unique. Let

$$\bar{g}(x, t, \mu, c) := -t - \frac{1}{2c}([c(x^3 - t) + \mu]_+^2 - \mu^2)$$

and the p-th order integral-type double penalty function ($p \in]0, 1[$)

$$G_{int}^p(x, \mu, c) := x^3 + \rho \int_{-1}^1 [\bar{g}(x, t, \mu, c)]_+^p \, dt.$$

The lower level problem $Q(x)$: $\min_{t \in [-1,1]} -t$ s.t. $x^3 - t \leq 0$, has a unique KKT multiplier. The perturbed problem of the lower level problem $Q(x)$ is

$$\max_{t \in [-1,1]} -t \text{ s.t. } x^3 - t \leq u.$$

Its value function is $v(x, u) = u - x^3$ for all points near $(x, u) = (0, 0)$, and is twice continuously differentiable. Thus the lower level problem is stable of degree 2 (in a neighborhood of the origin). Therefore, the original GSIP can be equivalently transformed into an SIP in a neighborhood of the origin.

We will verify that penalty function G_{int}^1 is not exact but $G_{int}^{\frac{1}{2}}$ is exact at $(\hat{x}, \hat{\mu}, \hat{c}) = (0, 1, \hat{c})$ for some $\hat{c} > 0$. For this we only need to consider points near $(0, 1, \hat{c})$. It is also enough to consider the case when $x \leq 0$.

Let $x \leq 0$. The effective integral interval of y consists of the two parts:

$$A := \{t : x \leq 0, c(x^3 - t) + \mu \leq 0, \bar{g}(x, t, \mu, c) \geq 0\},$$

$$B := \{t : x \leq 0, c(x^3 - t) + \mu \geq 0, \bar{g}(x, t, \mu, c) \geq 0\}.$$

Then $A = \{t : x \leq 0, x^3 + \frac{\mu}{c} \leq t \leq \frac{\mu^2}{2c}\}$ and

$$B = \left\{ t : x \leq 0, t \leq x^3 + \frac{\mu}{c}, \right.$$
$$\left. t \in \left[x^3 - \frac{1 - \mu + \sqrt{(1-\mu)^2 - 2cx^3}}{c}, x^3 - \frac{1 - \mu - \sqrt{(1-\mu)^2 - 2cx^3}}{c} \right] \right\}.$$

By choosing $\hat{c}$ large enough, and since (x, μ, c) is in a neighborhood of $(0, 1, \hat{c})$, then $A = \emptyset$, and

$$B = \left\{ t \in \left[x^3 - \frac{1 - \mu + \sqrt{(1-\mu)^2 - 2cx^3}}{c}, x^3 - \frac{1 - \mu - \sqrt{(1-\mu)^2 - 2cx^3}}{c} \right] : x \leq 0 \right\}.$$

23

So,

$$G_{\text{int}}^p(x,\mu,c) = x^3 + \rho \int_{x^3-(1-\mu+\sqrt{(1-\mu)^2-2cx^3})/c}^{x^3-(1-\mu-\sqrt{(1-\mu)^2-2cx^3})/c} \left[-\frac{c}{2}t^2 - (1-\mu-cx^3)t - \frac{c}{2}x^6 - x^3\mu \right]^{\frac{1}{p}} \, \mathrm{d}t$$

$$= x^3 + \rho \left(\frac{c}{2}\right)^{\frac{1}{p}} \int_{-\sqrt{(1-\mu)^2-2cx^3}/c}^{\sqrt{(1-\mu)^2-2cx^3}/c} [((1-\mu)^2 - 2cx^3)/c^2 - t^2]^{\frac{1}{p}} \, \mathrm{d}t.$$

It is easy to calculate that

$$G_{\text{int}}^1(x,1,\hat{c}) = x^3 + \frac{\rho\hat{c}}{2} \int_{-\sqrt{-2\hat{c}x^3}/\hat{c}}^{\sqrt{-2\hat{c}x^3}/\hat{c}} -\frac{2x^3}{\hat{c}} - t^2 \, \mathrm{d}t = x^3 + \frac{4\rho}{3} \cdot \left(\frac{2}{\hat{c}}\right)^{\frac{1}{2}} \cdot (-x)^{\frac{9}{2}},$$

$$G_{\text{int}}^{\frac{1}{2}}(x,\mu,c) = x^3 + \rho \cdot \sqrt{\frac{c}{2}} \cdot \frac{(1-\mu)^2 - x^3}{c^2} \cdot \int_{-1}^{1} \sqrt{1-t^2} \, \mathrm{d}t$$

$$= x^3 + \rho \cdot \sqrt{\frac{c}{2}} \cdot \frac{(1-\mu)^2 - x^3}{c^2} \cdot \frac{\pi}{2}$$

$$\geq x^3 + \rho \cdot \sqrt{\frac{c}{2}} \cdot \frac{1}{c^2} \cdot \frac{\pi}{2} \cdot (-x^3).$$

Then $G_{\text{int}}^{\frac{1}{2}}(x,\mu,c) \geq 0$ near $(0,1,\hat{c})$. Thus G_{int}^1 is not exact and $G_{\text{int}}^{\frac{1}{2}}$ is exact.

We also have $D(\hat{x},\hat{\mu},\hat{c}) = \mathbb{R}^3$, and $\bar{g}^{\circ\circ}(\hat{x},t,\hat{\mu},\hat{c};d) \equiv 0$. Hence, all the conditions in Theorem 5.1 (ii) are satisfied and thus the optimality condition (29) for GSIP holds.

References

1. Krabs, W.: Optimization and Approximation. Translated from the German by Philip M. Anselone and Ronald B. Guenther, A Wiley-Interscience Publication, John Wiley & Sons Ltd., Chichester (1979)

2. Hettich, R. and Kortanek, K.: Semi-infinite programming: theory, methods, and applications. SIAM Review **35**, 380–429 (1993)

3. Polak, E.: Optimization: Algorithms and Consistent Approximations. Applied Mathematical Sciences **124**. Springer-Verlag, New York (1997)

4. Reemtsen, R. and Rückmann, J.-J. editors. Semi-Infinite Programming. Nonconvex Optimization and its Applications **25**. Kluwer Academic Publishers, Boston, MA (1998)

5. Goberna, M. Á. and López, M. A. editors. Semi-Infinite Programming: Recent Advances. Nonconvex Optimization and its Applications **57**. Kluwer Academic Publishers, Dordrecht (2001)

6. Jongen, H. T., Rückmann, J.-J. and Stein, O.: Generalized semi-infinite optimization: a first order optimality condition and examples. Math. Programming, **83**, 145–158 (1998)

7. Stein, O.: Bi-level Strategies for Semi-infinite Programming, Kluwer Academic Publishers, Boston (2003)

8. Still, G.: Generalized semi-infinite programming: theory and methods. Europ. J. Oper. Res. **119**, 301–313 (1999)

9. John, F.: Extremum problems with inequalities as subsidiary conditions. Studies and Essays, Courant Anniversary Volume, Friederichs, K.O., Neugebauer, O.E., Stocker, J.J., eds., Wiley, New York, 187-204 (1948)

10. Pschenichnyi, B. N.: Necessary Conditions for an Extremum. Marcel Dekker, New York (1971)

11. Hettich, R. and Jongen, H. T.: Semi-infinite programming: conditions of optimality and applications. In Optimization techniques (Proc. 8th IFIP Conf., Würzburg, 1977), Part 2, Lecture Notes in Control and Information Sci. **7**, 1–11. Springer, Berlin (1978)

12. Borwein, J. M.: Direct theorems in semi-infinite convex programming. Math. Programming **21**, 301–318 (1981)

13. Li, W., Nahak, C. and Singer, I.: Constraint qualifications for semi-infinite systems of convex inequalities. SIAM J. Optim. **11**, 31–52 (2000)

14. López, M. A. and Vercher, E.: Optimality conditions for nondifferentiable convex semi-infinite programming. Math. Programming **27**, 307–319 (1983)

15. Stein, O.: On constraint qualifications in nonsmooth optimization. J. Optim. Theory Appl. **121**, 647-671 (2004)

16. Zheng, X. Y. and Yang, X. Q.: Lagrange multipliers in nonsmooth semi-infinite optimization problems. Math. Oper. Res. **32**, 168–181 (2007)

17. Weber, G.-W.: Generalized semi-infinite optimization: on some foundations. Vychisl. Tekhnol. **4**, 41–61 (1999)

18. Rückmann, J. J. and Shapiro, A.: First-order optimality conditions in generalized semi-infinite programming. J. Optim. Theory Appl. **101**, 677-691 (1999)

19. Stein, O.: First-order optimality conditions for degenerate index sets in generalized semi-infinite optimization. Math. Oper. Res. **26**, 565-582 (2001)

20. Ye, J. J. and Wu, S. Y.: First order optimality conditions for generalized semi-infinite programming problems. J. Optim. Theory Appl., **137**, 419-434 (2008)

21. Clarke, F. H.: Optimization and Nonsmooth Analysis. Canadian Mathematical Society Series of Monographs and Advanced Texts. John Wiley & Sons Inc., New York (1983)

22. Burke, J. V.: An exact penalization viewpoint of constrained optimization. SIAM J. Cont. & Optim. **29**, 968–998 (1991)

23. Yang, X. Q. and Meng, Z. Q.: Lagrange multipliers and calmness conditions of order p. Math. Oper. Res. **32**, 95–101 (2007)

24. Meng, K. W. and Yang, X. Q.: Optimality conditions via exact penalty functions. SIAM J. Optim., **20**, 3208–3231 (2010)

25. Rubinov, A. and Yang, X.Q.:   Lagrange-type Functions in Constrained Nonconvex Optimization. Kluwer Academic Publishers, Dordrecht, Netherlands (2003)

26. Huang, X.X. and Yang, X.Q.: A unified augmented Lagrangian approach to duality and exact penalization. Math. Oper. Res. **28**, 533-552 (2003)

27. Rockafellar, R. T.:   Augmented Lagrange multiplier functions and duality in nonconvex programming. SIAM J. Control **12**, 268–285 (1974)

28. Rockafellar, R. T. and Wets, R. J.-B.:   Variational Analysis. Springer-Verlag, Berlin (1998)

29. Pietrzykowski, T.: A generalization of the potential method for conditional maxima on the Banach, reflexive spaces. Numer. Math. **18**, 367–372 (1971/72)

30. Conn, A. R. and Gould, N. I. M.: An exact penalty function for semi-infinite programming. Math. Programming, **37**, 19–40 (1987)

31. Rockafellar, R. T.: Convex Analysis. Princeton University Press (1970)

32. Levitin, E. S.: Reduction of generalized semi-infinite programing problems to semi-infinite or piece-wise smooth programming problems. Preprint, 8-2001, University of Trier (2001)

33. Polak, E. and Royset, J. O.:   On the use of augmented Lagrangians in the solution of generalized semi-infinite min-max problems. Comput. Optim. Appl. **31**, 173–192 (2005)

34. Royset, J. O., Polak, E. and Der Kiureghian, A.: Adaptive approximations and exact penalization for the solution of generalized semi-infinite min-max problems. SIAM J. Optim. **14**, 1–33 (2003)

35. Lang, S.:   Real and Functional Analysis, volume 142 of Graduate Texts in Mathematics. Springer-Verlag, New York, third edition (1993)

36. R. Fletcher.   Practical Methods of Optimization. A Wiley-Interscience Publication. John Wiley & Sons Ltd., Chichester, second edition (1987)

37. Bonnans, J. F. and Shapiro, A.: Perturbation Analysis of Optimization Problems. Springer Series in Operations Research. Springer-Verlag, New York (2000)

38. Cominetti, R. and Correa, R.: A generalized second-order derivative in nonsmooth optimization. SIAM J. Control Optim. **28**, 789–809 (1990)

39. Hiriart-Urruty, J.-B., Strodiot, J.-J. and Nguyen, V. H.: Generalized hessian matrix and second-order optimality conditions for problems with $C^{1,1}$ data. Appl. Math. & Optim. **11**, 43–56 (1984)

40. Yang, X. Q.: Generalized second-order derivatives and optimality conditions. Nonlinear Anal. **23**, 767–784, (1994)

41. Burachik, R. S. and Rubinov, A.: Abstract convexity and augmented Lagrangians. SIAM J. Optim. **18**, 413-436 (2007)

42. Nedic, A. and Ozdaglar, A.: Separation of nonconvex sets with general augmenting functions. Math. Oper. Res. **33**, 587-605 (2008)