Solving semi-infinite programs by smoothing projected gradient method

Mengwei Xu[*], Soon-Yi Wu[†], and Jane J. Ye[‡]

October 16, 2013

Abstract. In this paper, we study a semi-infinite programming (SIP) problem with a convex set constraint. Using the value function of the lower level problem, we reformulate SIP problem as a nonsmooth optimization problem. Using the theory of nonsmooth Lagrange multiplier rules and Danskin's theorem, we present constraint qualifications and necessary optimality conditions. We propose a new numerical method for solving the problem. The novelty of our numerical method is to use the integral entropy function to approximate the value function and then solve SIP by the smoothing projected gradient method. Moreover we study the relationships between the approximating problems and the original SIP problem. We derive error bounds between the integral entropy function and the value function, and between locally optimal solutions of the smoothing problem and those for the original problem. Using certain second order sufficient conditions, we derive some estimates for locally optimal solutions of problem (VP). Numerical experiments show that the algorithm is efficient for solving SIP.

Key Words. Nonlinear semi-infinite programming problem, value function, integral entropy function, smoothing projected gradient algorithm, locally optimal solution.

2010 Mathematics Subject Classification. 65K05, 90C34.

[*]School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China. E-mail: xumengw@hotmail.com.

[†]Department of Mathematics, National Cheng Kung University/National Center for Theoretical Sciences, Tainan, Taiwan. E-mail: soonyi@mail.ncku.edu.tw.

[‡]Corresponding Author. Department of Mathematics and Statistics, University of Victoria, Victoria, B.C., Canada V8W 3R4. E-mail: janeye@uvic.ca. The research of this author was partially supported by NSERC.

1 Introduction.

The semi-infinite programming problem is an optimization problem with finite dimensional decision variable $x \in \mathbb{R}^n$ and infinite number of constraints. In this paper we consider the following (standard) nonlinear semi-infinite programming (SIP) problem:

$$\begin{array}{llll} \text{(SIP)} & \min\limits_{x} & f(x) & \\ & \text{s.t.} & g(x,y) \leq 0, & \forall y \in Y, \\ & & x \in X, & \end{array}$$

where $f : \mathbb{R}^n \to \mathbb{R}, g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, X is a subset of $\mathbb{R}^n$ and Y is a nonempty compact subsets of $\mathbb{R}^m$. Unless otherwise specified, in this paper we assume that f, g are continuously differentiable and X is a closed and convex set. In a more general case where Y is not fixed but x-dependent, the problem is called the generalized semi-infinite programming problem (see, e.g., [44]). In applications, additional semi-infinite constraints may be present. Whereas the techniques and results presented in this paper can be extended to tackle these more general problems (as in Examples 4.1 and 4.2), for the sake of simplicity we develop the main ideas only for SIP with one semi-infinite constraint.

Standard SIP problems have numerous applications and have been studied systematically since the 1960s. We refer to [16] for introduction to SIP problem, [31] for numerical methods of SIP, [12] for linear SIP and [27] for algorithmic aspects. We also refer to [15, 28] for excellent reviews with hundreds of references on SIP. The monograph [35] contains a detailed study of generalized SIP and the papers [13, 25, 36] survey the most recent development on the subject.

The difficulty of solving a SIP problem lies in that there are infinite many constraints. The earliest and the most common approach to solve SIP involving using the method of solving finite mathematical programming problems in one way or the other. The most common approach for solving SIP problems is the discretization method by which one chooses a finite grid of Y, solves a finite mathematical programs with Y replaced by the finite grid and updates the grid (see e.g. [27, 38, 40]). The discretization method is easy to implement but it is computationally costly and the cost per iteration increases dramatically as the cardinality of the finite grid of Y grows. Another approach, the reduction-based method, is to reduce the problem locally to a finite-dimensional nonlinear programming problem (see e.g. [2, 3, 15, 18, 39]). The reduction-based method, however, usually requires very strong assumptions including the conditions leading to the finiteness of the set of active constraints. Beyond discretization and reduction-based methods, one of the most important method is the so-called exchange method [1, 9, 14, 17, 20, 21, 25, 27, 32, 41, 42]. Zhang et.al [45] also propose a new exchange method for convex SIP problem which introduced a new dropping rule that only keeps those active constraints with positive Lagrange multipliers. Moreover, the algorithm does not require to solve a maximization problem over the index set at each iteration, it only needs to find some points such that a certain computation–easy criterion is satisfied. The Kurash-Kuch-Tucker (KKT) system of SIP can be reformulated to a system of semi-smooth equations. The authors [23, 26, 29, 30, 43] then apply the smoothing, semi-smooth Newton method, or Newton-type method to solve the KKT system. However, the methods presented in [23, 30] do not ensure the feasibility of the original SIP problem, the method of [43] does not have locally superlinear convergence property, and accumulation points of the sequences generated in [26, 29] are not necessarily stationary points of the SIP problem. Recently, [22] proposed a new smoothing Newton-type method for solving the SIP problem which have overcome these drawbacks.

More recent approach for dealing with SIP involves the following equivalent formulation.

$$
\begin{array}{lll}
\text{(VP)} \qquad & \min\limits_{x\in X} & f(x) \\
& \text{s.t.} & V(x) \leq 0,
\end{array}
$$

where $V(x) := \max\limits_{y\in Y} g(x,y)$ is the value function of the lower level problem. For the case where the lower level problem is concave (i.e., the lower level objective function $g(x,y)$ is concave in y and Y is a convex set), the Karush-Kuhn-Tucker (KKT) condition for the lower level problem is also sufficient for optimality. In this case, SIP can be reformulated and solved as a mathematical program with complementarity constraints (MPCC) [35, 37]. For the general case where the lower level problem may not be concave, Floudas and Stein [11] developed an adaptive convexification algorithm for solving SIP. Their idea is to adaptively construct concave relaxations of the lower level problem by using the αBB method ([10]), replace the relaxed lower level problems equivalently by the KKT conditions, and solve the resulting MPCCs. This approximation produces feasible iterates for the original SIP. Recently Shiu and Wu [34] combine the idea of the relaxed cutting plane method with convexification to solve SIP.

In this paper, we introduce a new method for SIP by solving the equivalent formulation (VP). Our method is based on the smoothing projected gradient algorithm that we recently introduced to solve a simple bilevel program in [24]. Problem (VP) is a nonsmooth problem since the value function $V(x)$ is generally nonsmooth even when the function $g(x,y)$ is smooth. By Danskin's theorem (see Proposition 2.1 below), the value function is Lipschitz continuous and its Clarke generalized gradients can be computed. By using the nonsmooth KKT conditions, we can then define stationary conditions for problem

(VP) under suitable constraint qualifications. To design a numerical algorithm, we use the following smooth approximation of the value function $V(x)$ introduced in [24]:

$$\gamma_\rho(x) := \rho^{-1} \ln \left(\int_Y \exp[\rho g(x,y)]dy \right)$$

where $\rho > 0$. We then use this smoothing function to replace the value function, solve a smoothing penalty problem, update the smoothing parameter ρ and the penalty parameter. Under suitable conditions, the iteration sequence converges to a stationary point of (VP). To give a quantitative measure on how good the approximation of the smoothing function is, we study the error bounds between the integral entropy function and the value function. We derive some relationships among the stationary points, the locally optimal solutions for the smoothing problems and the original problems. What is more, we derive some estimates for locally optimal solutions of problem (VP) under some second order sufficient conditions.

The rest of the paper is organized as follows. In Section 2, we study the optimality conditions and introduce some constraint qualifications. In Section 3, we study the smoothing projected gradient method and derive some properties of the integral entropy function including the error bounds between the integral entropy function and the value function. Moreover we give some error bounds between locally optimal solutions and current iterate of the algorithm. In this section we also give estimates for locally optimal solutions of the SIP problem under certain second order sufficient conditions. In Section 4, we report some results for our numerical experiments.

We adopt the following standard notation in this paper. For any two vectors a and b in $\mathbb{R}^n$, we denote by $a^T b$ their inner product. Given a function $G : \mathbb{R}^n \to \mathbb{R}^m$, we denote its Jacobian by $\nabla G(z) \in \mathbb{R}^{m\times n}$ and, if $m = 1$, the gradient $\nabla G(z) \in \mathbb{R}^n$ is considered as a column vector. For a set $\Omega \subseteq \mathbb{R}^n$, we denote by int Ω and co Ω the interior and the convex hull respectively. Let $N(x)$ be a neighbourhood of point x. In addition, we denote by $\mathbf{N}$ the set of nonnegative integers, $\exp(z)$ the exponential function and $|\cdot|$ the Lebesgue measure of a set.

2 Optimality conditions

The optimality conditions for SIP when $X = \mathbb{R}^n$ have been studied since the 1960s. A complete survey on the optimality conditions for linear, convex and smooth SIP problems can be found in [16, 25, 33]. In this section, we develop the constraint qualifications and necessary optimality conditions for (SIP) with X not necessarily equal to the whole space $\mathbb{R}^n$.

We first recall some background materials on nonsmooth analysis. For a convex set $C \subseteq \mathbb{R}^m$ and a point $z \in C$, the normal cone of C at z is given by

$$\mathcal{N}_C(z) := \{\zeta \in \mathbb{R}^m : \zeta^T(z' - z) \leq 0, \ \forall z' \in C\}$$

and the tangent cone of C at z is given by

$$\mathcal{T}_C(z) := \{d \in \mathbb{R}^m : (z^\nu - z)/\tau_\nu \to d \text{ for some } z^\nu \in C, \ z^\nu \to z, \ \tau_\nu \searrow 0\},$$

respectively.

Let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be Lipschitz continuous near $\bar{x}$. The Clarke generalized directional derivative of φ at $\bar{x}$ in direction d is defined by

$$\varphi^\circ(\bar{x}; d) := \limsup_{x \to \bar{x}, \ t \searrow 0} \frac{\varphi(x + td) - \varphi(x)}{t}.$$

The Clarke generalized gradient of φ at $\bar{x}$ is a convex and compact subset of $\mathbb{R}^n$, which is defined by

$$\partial\varphi(\bar{x}) := \{\xi \in \mathbb{R}^n : \xi^T d \leq \varphi^\circ(\bar{x}; d), \quad \forall d \in \mathbb{R}^n\}.$$

Note that, when φ is convex, the Clarke generalized gradient coincides with the subgradient in the sense of convex analysis, i.e.,

$$\partial\varphi(\bar{x}) = \{\xi \in \mathbb{R}^n : \xi^T(x - \bar{x}) \leq \varphi(x) - \varphi(\bar{x}), \quad \forall x \in \mathbb{R}^n\}$$

and, when φ is continuously differentiable at $\bar{x}$, we have $\partial\varphi(\bar{x}) = \{\nabla\varphi(\bar{x})\}$.

The following theorem plays a key role in the development of optimality conditions.

Proposition 2.1 (Danskin's Theorem) ([6, Page 99] or [8]) *Let $Y \subseteq \mathbb{R}^m$ be a compact set and $g(x, y)$ be a function defined on $\mathbb{R}^n \times \mathbb{R}^m$ that is continuously differentiable at $\bar{x}$. Then the value function*

$$V(x) := \max\{g(x, y) : y \in Y\}$$

is Lipschitz continuous near $\bar{x}$ and the Clarke generalized gradient of V at $\bar{x}$ is

$$\partial V(\bar{x}) = co\{\nabla_x g(\bar{x}, y) : y \in S(\bar{x})\}, \tag{2.1}$$

where $S(\bar{x})$ denotes the set of all maximizers of $g(\bar{x}, y)$ over $y \in Y$.

By the Carathéodory's Theorem, a closed convex set in $\mathbb{R}^n$ can be represented as a convex combination of not more than $n + 1$ points. Hence we have

$$\partial V(\bar{x}) = \left\{ \sum_{k=1}^{n+1} \lambda_k \nabla_x g(\bar{x}, \bar{y}_k) : \lambda_k \geq 0, \sum_{k=1}^{n+1} \lambda_k = 1, \bar{y}_k \in S(\bar{x}) \right\}. \tag{2.2}$$

By Danskin's theorem, the value function $V(x)$ is Lipschitz continuous at each point x and the Clarke generalized gradient can be calculated as in (2.2). This fact allows us to apply the nonsmooth KKT condition of Clarke [5, Proposition 6.4.4] to problem (VP) and obtain the nonsmooth KKT condition for (VP) under the calmness condition as defined below.

Definition 2.1 (Calmness) [5, Definition 6.4.1] Let $\bar{x}$ be a locally optimal solution of (VP). We say that (VP) is calm at $\bar{x}$ if $\bar{x}$ is also a locally optimal solution of the exact penalty problem

$$\begin{aligned} \min \quad & f(x) + \lambda \max\{V(x), 0\} \\ \text{s.t.} \quad & x \in X \end{aligned}$$

for some $\lambda > 0$.

Theorem 2.1 Let $\bar{x} \in X$ be a local minimizer of (SIP). Suppose that (VP) is calm at $\bar{x}$. Then there exists $\mu \geq 0$ such that

$$0 \in \nabla f(\bar{x}) + \mu \partial V(\bar{x}) + \mathcal{N}_X(\bar{x}), \tag{2.3}$$

$$\mu V(\bar{x}) = 0. \tag{2.4}$$

The necessary conditions in Theorem 2.1 involve the value function $V(x)$. However, using the formula (2.2), we can derive the conditions in terms of the original problem data of (SIP). Let $\bar{x}$ be a feasible solution of (SIP). Denote the active set of problem (SIP) at $\bar{x}$ by $Y_{act}(\bar{x}) = \{y \in Y : g(\bar{x}, y) = 0\}$. Note that when $V(\bar{x}) = 0$, $Y_{act}(\bar{x}) = S(\bar{x})$, otherwise if $V(\bar{x}) < 0$, then $Y_{act}(\bar{x}) = \emptyset$.

Theorem 2.2 Let $\bar{x} \in X$ be a local minimizer of (SIP). Suppose that (VP) is calm at $\bar{x}$. Then either $Y_{act}(\bar{x}) = \emptyset$ and $0 \in \nabla f(\bar{x}) + \mathcal{N}_X(\bar{x})$ or there exists $\bar{y}_k \in Y_{act}(\bar{x})$, for $1 \leq k \leq n+1$ and $\lambda_k \geq 0$ for $1 \leq k \leq n+1$ such that

$$0 \in \nabla f(\bar{x}) + \sum_{k=1}^{n+1} \lambda_k \nabla_x g(\bar{x}, \bar{y}_k) + \mathcal{N}_X(\bar{x}), \tag{2.5}$$

$$\lambda_k g(\bar{x}, \bar{y}_k) = 0, \ 1 \leq k \leq n+1. \tag{2.6}$$

Proof. By Theorem 2.1, (2.3) and (2.4) hold for some $\mu \geq 0$. Since $Y_{act}(\bar{x}) = \emptyset$ if and only if $V(\bar{x}) < 0$. When $Y_{act}(\bar{x}) = \emptyset$, (2.4) implies that $\mu = 0$. Hence $0 \in \nabla f(\bar{x}) + \mathcal{N}_X(\bar{x})$ in this case. When $V(\bar{x}) = 0$, $Y_{act}(\bar{x}) = S(\bar{x})$. (2.5) and (2.6) follow from formula (2.2). ■

The calmness condition may seem to be hard to verify. In fact this is not necessarily true since certain constraint qualifications lead to the calmness condition. Here we list two of them.

Definition 2.2 (EMFCQ) *A feasible point $\bar{x}$ is said to satisfy the extended Mangasarian-Fromovitz constraint qualification* (EMFCQ) *for problem* (SIP) *if there exists a direction $d \in \text{int } \mathcal{T}_X(\bar{x})$ such that*

$$\nabla_x g(\bar{x}, y)^T d < 0, \text{ for all } y \in Y_{act}(\bar{x}).$$

Definition 2.3 (NNAMCQ) *Let $\bar{x}$ be a feasible point of problem* (SIP)*. We say that the no nonzero abnormal multiplier constraint qualification* (NNAMCQ) *holds at $\bar{x}$ if either $V(\bar{x}) < 0$ or $V(\bar{x}) = 0$ but*

$$0 \notin \partial V(\bar{x}) + \mathcal{N}_X(\bar{x}); \tag{2.7}$$

equivalently if either $g(\bar{x}, y) < 0$ for all $y \in Y$ or $Y_{act}(\bar{x}) \neq \emptyset$ and there exists $(\lambda_1, \cdots, \lambda_{n+1}) \in \mathbb{R}^{n+1}$, $\lambda_k \geq 0$ for $1 \leq k \leq n+1$ not all zero and $\bar{y}_k \in Y_{act}(\bar{x})$ such that

$$0 \notin \sum_{k=1}^{n+1} \lambda_k \nabla_x g(\bar{x}, \bar{y}_k) + \mathcal{N}_X(\bar{x}).$$

Proposition 2.2 *Let $\bar{x}$ be a feasible point to* (SIP)*. Then the following implication always holds*

$$\text{EMFCQ} \Longrightarrow \text{NNAMCQ},$$

and the reverse implication holds provided by $\text{int } \mathcal{T}_{\mathrm{X}}(\bar{\mathrm{x}}) \neq \emptyset$. *If $\bar{x}$ is a locally optimal solution to* (SIP) *and* NNAMCQ *holds at $\bar{x}$ then* (VP) *is calm at $\bar{x}$.*

Proof. Since when $V(\bar{x}) = 0$, $Y_{act}(\bar{x}) = S(\bar{x})$, otherwise if $V(\bar{x}) < 0$ then $Y_{act}(\bar{x}) = \emptyset$. Thus by formula (2.2), it is easy to see that EMFCQ is equivalent to saying that either $V(\bar{x}) < 0$ or $V(\bar{x}) = 0$ but there exists a direction $d \in \text{int } \mathcal{T}_X(\bar{x})$ such that

$$v^T d < 0, \text{ for all } v \in \partial V(\bar{x})$$

which is the nonsmooth MFCQ for problem (VP). Consequently the proof of the first assertion follows from [19]. The second assertion follows from [5, Corollary 5 to Theorem 6.5.2]. ■

Based on Theorems 2.1 and 2.2, we define the stationary point for (SIP). Note that (VP) and (SIP) are exactly equivalent and the KKT conditions by Theorems 2.1 and 2.2 are also equivalent due to the formula (2.2).

Definition 2.4 (Stationary point) *We call a feasible point $\bar{x}$ a stationary point of problem* (VP) *if there exists $\mu \geq 0$ such that* (2.3) *and* (2.4) *hold. Equivalently we call a feasible point $\bar{x}$ a stationary point of problem* (SIP) *if either $Y_{act}(\bar{x}) = \emptyset$ and $0 \in \nabla f(\bar{x}) + \mathcal{N}_X(\bar{x})$ or there exist $\bar{y}_k \in Y_{act}(\bar{x})$, for $1 \leq k \leq n+1$ and $(\lambda_1, \cdots, \lambda_{n+1}) \in R^{n+1}$, $\lambda_k \geq 0$ for $1 \leq k \leq n+1$ such that* (2.5) *and* (2.6) *hold.*

The following corollary follows by Proposition 2.2 and Theorems 2.1 and 2.2.

Corollary 2.1 *Let $\bar{x} \in X$ be a local minimizer of* (SIP)*. Suppose that either* NNAMCQ *or* EMFCQ *holds at $\bar{x}$. Then $\bar{x}$ is a stationary point of* (SIP)*.*

In order to accommodate infeasible accumulation points in the numerical algorithm, we now extend EMFCQ and NNAMCQ to allow infeasible points.

Definition 2.5 (Extended EMFCQ) *A point $\bar{x} \in X$ is said to satisfy the extended Mangasarian Fromovitz constraint qualification* (Extended EMFCQ) *for problem* (SIP) *if either $g(\bar{x}, y) < 0$ for all $y \in Y$ or there exists a direction $d \in \operatorname{int} \mathcal{T}_X(\bar{x})$ such that*

$$\nabla_x g(\bar{x}, y)^T d < 0, \text{ for all } y \in S(\bar{x}).$$

Definition 2.6 (ENNAMCQ) *Let $\bar{x} \in X$. We say that the extended no nonzero abnormal multiplier constraint qualification* (ENNAMCQ) *holds at $\bar{x}$ for problem* (VP) *if either $V(\bar{x}) < 0$ or $V(\bar{x}) \geq 0$ but*

$$0 \notin \partial V(\bar{x}) + \mathcal{N}_X(\bar{x}).$$

Note that the extended EMFCQ and the extended NNAMCQ reduce to the EMFCQ and NNAMCQ respectively if x is a feasible point of problem (SIP).

3 Smoothing projected gradient algorithm

In this section, we use the smoothing projected gradient algorithm proposed in [24] to solve (SIP). We study the relationships between the smoothing problem and the original problem. Moreover we introduce some second order conditions under which the object value of the current iteration is very close to the locally optimal solution of the problem.

Definition 3.1 *Assume that, for a given $\rho > 0$, $g_\rho : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function. We say that $\{g_\rho : \rho > 0\}$ is a family of smoothing functions of g_0 if $\lim_{z \to x, \ \rho \uparrow \infty} g_\rho(z) = g_0(x)$ for any fixed $x \in \mathbb{R}^n$.*

Definition 3.2 [4] *We say that a family of smoothing functions $\{g_\rho : \rho > 0\}$ satisfies the gradient consistency property if for any $x \in \mathbb{R}^n$, $\limsup_{z \to x, \rho \uparrow \infty} \nabla g_\rho(z)$ is nonempty and $\limsup_{z \to x, \rho \uparrow \infty} \nabla g_\rho(z) \subseteq \partial g_0(x)$, where*

$$\limsup_{z \to x, \rho \uparrow \infty} \nabla g_\rho(z) := \left\{ \lim_{k \to \infty} \nabla g_{\rho_k}(z_k) : z_k \to x, \rho_k \uparrow \infty \right\}$$

is the set of all limit points.

For a given positive integer ρ, we define the integral entropy function as

$$\begin{aligned}\gamma_\rho(x) &:= \rho^{-1}\ln\left(\int_Y \exp[\rho g(x,y)]dy\right)\\ &= V(x)+\rho^{-1}\ln\left(\int_Y \exp[\rho(g(x,y)-V(x))]dy\right).\end{aligned}$$

In [24], the value function is defined to be the minimum value instead of the maximum value of the lower level objective function subject to $y \in Y$. If we set $f(x,y) := -g(x,y)$ in [24], we have $-V(x) = \min_{y\in Y} -g(x,y)$. Since $\partial(-V(x)) = -\partial V(x)$, we obtain the following results by applying [24, Theorem 5.1 and 5.5].

Lemma 3.1 $\{\gamma_\rho(x) : \rho > 0\}$ *is a family of smoothing function of the value function* $V(x)$ *which satisfies the gradient consistency property.*

The next proposition discusses the monotonicity of the entropy function γ_ρ.

Proposition 3.1 *For any given* $x \in X$*, if* $|S(x)| \geq 1$*, then the integral entropy function* $\gamma_\rho(x)$ *is a nonincreasing function of* ρ *while if* $|S(x)| < 1$*,* $\gamma_\rho(x)$ *is a nondecreasing function of* ρ *when* ρ *is sufficiently large.*

Proof. From the definition of $V(x)$, we have $g(x,y) - V(x) < 0$ for any $y \in Y \setminus S(x)$. It follows from the monotonicity of the exponential function that, for any positive numbers $\rho_1 > \rho_2$,

$$\int_{Y\setminus S(x)} \exp[\rho_2(g(x,y)-V(x))]dy > \int_{Y\setminus S(x)} \exp[\rho_1(g(x,y)-V(x))]dy. \tag{3.1}$$

By the definition of $\gamma_\rho(\cdot)$ and the fact that $g(x,y) - V(x) = 0$, for any $y \in S(x)$ we have that for any $x \in X$,

$$\begin{aligned}&\gamma_{\rho_1}(x) - \gamma_{\rho_2}(x)\\ &= \rho_1^{-1}\ln\left(\int_Y \exp[\rho_1(g(x,y)-V(x))]dy\right) - \rho_2^{-1}\ln\left(\int_Y \exp[\rho_2(g(x,y)-V(x))]dy\right)\\ &= \rho_1^{-1}\ln\left(|S(x)| + \int_{Y\setminus S(x)} \exp[\rho_1(g(x,y)-V(x))]dy\right)\\ &\quad -\rho_2^{-1}\ln\left(|S(x)| + \int_{Y\setminus S(x)} \exp[\rho_2(g(x,y)-V(x))]dy\right).\end{aligned}$$

Consider the case when $|S(x)| \geq 1$. Since $\ln z$ is monotonously increasing and non-negative valued when $z \geq 1$, by (3.1) we have $\gamma_{\rho_1}(x) - \gamma_{\rho_2}(x) \leq 0$ and, $\gamma_{\rho_1}(x) - \gamma_{\rho_2}(x) = 0$ only when $|S(x)| = 1$ and $|Y \setminus S(x)| = 0$.

We now consider the case when $|S(x)| < 1$. For any fixed $x \in X$ and $\rho > 0$, by the Mean Value Theorem one can find c which lies between $|S(x)|$ and $|S(x)| + \int_{Y\setminus S(x)} \exp[\rho(g(x,y) - V(x))]dy$ such that

$$\begin{aligned} &\ln\left(|S(x)| + \int_{Y\setminus S(x)} \exp[\rho(g(x,y) - V(x))]dy\right) \\ &= \ln|S(x)| + c^{-1}\int_{Y\setminus S(x)} \exp[\rho(g(x,y) - V(x))]dy. \end{aligned}$$

Thus, there exists c_1 lies between $|S(x)|$ and $|S(x)| + \int_{Y\setminus S(x)} \exp[\rho_1(g(x,y) - V(x))]dy$ and c_2 lies between $|S(x)|$ and $|S(x)| + \int_{Y\setminus S(x)} \exp[\rho_2(g(x,y) - V(x))]dy$ such that

$$\begin{aligned} &\gamma_{\rho_1}(x) - \gamma_{\rho_2}(x) \\ &= (\rho_1^{-1} - \rho_2^{-1})\ln|S(x)| + \rho_1^{-1}c_1^{-1}\int_{Y\setminus S(x)} \exp[\rho_1(g(x,y) - V(x))]dy \\ &\quad -\rho_2^{-1}c_2^{-1}\int_{Y\setminus S(x)} \exp[\rho_2(g(x,y) - V(x))]dy. \end{aligned}$$

Since $\ln|S(x)| < 0$ and $\int_{Y\setminus S(x)} \exp[\rho(g(x,y) - V(x))]dy$ tends to 0 at an exponential rate as $\rho \to \infty$. It follows that for $\rho_1 \geq \rho_2$ sufficiently large, $\gamma_{\rho_1}(x) - \gamma_{\rho_2}(x) \geq 0$.

■

It is interesting to estimate the bounds of the difference between the integral entropy function and the value function.

Proposition 3.2 *Let $x \in X$ and $\rho > 0$. If $|S(x)| \geq 1$, then we have*

$$0 \leq \gamma_\rho(x) - V(x) \leq \frac{1}{\rho}\ln|Y|. \tag{3.2}$$

If $|S(x)| < 1$, then there exists positive constants M_x and m_x such that

$$-\varepsilon_\rho(x) \leq \gamma_\rho(x) - V(x) \leq 0, \tag{3.3}$$

where

$$\varepsilon_\rho(x) := -\frac{1}{\rho}\ln(\rho^{-1}|Y|) - \ln\left(1 - \frac{M_x}{m_x}\frac{|Y|}{\rho}\right). \tag{3.4}$$

for ρ large enough. If $|S(x)| = 1$ and $|Y \setminus S(x)| = 0$, we have $\gamma_\rho(x) = V(x)$.

Proof. First we consider the case when $|S(x)| \geq 1$. Since $g(x,y) - V(x) = 0$ for $y \in S(x)$, we have

$$\begin{aligned} \gamma_\rho(x) &= V(x) + \rho^{-1}\ln\left(|S(x)| + \int_{Y\setminus S(x)} \exp[\rho(g(x,y) - V(x))]dy\right) \\ &\geq V(x). \end{aligned} \tag{3.5}$$

Since

$$\left(\int_Y \exp[\rho g(x,y)]dy\right)^{\frac{1}{\rho}} \leq |Y|^{\frac{1}{\rho}} \exp \max_{y\in Y} g(x,y) = |Y|^{\frac{1}{\rho}} \exp V(x),$$

by the monotonicity of the logarithmic function, we have

$$\gamma_\rho(x) \leq V(x) + \rho^{-1} \ln |Y|. \tag{3.6}$$

Therefore, (3.2) follows by (3.5) and (3.6).

Note that if $|S(x)| = 1$ and $|Y \setminus S(x)| = 0$, we have $\gamma_\rho(x) = V(x)$ by the definition of γ_ρ.

Finally we consider the case when $|S(x)| < 1$. Since $g(x,y) - V(x) < 0$ for any $y \in Y \setminus S(x)$, we have

$$\int_{Y\setminus S(x)} \exp[\rho(g(x,y) - V(x))]dy \to 0, \text{ as } \rho \to \infty,$$

which implies that when $|S(x)| < 1$, for a sufficiently large ρ,

$$|S(x)| + \int_{Y\setminus S(x)} \exp[\rho(g(x,y) - V(x))] \leq 1.$$

Therefore, for ρ large enough

$$\begin{aligned} \gamma_\rho(x) &= V(x) + \rho^{-1} \ln\left(|S(x)| + \int_{Y\setminus S(x)} \exp[\rho(g(x,y) - V(x))]dy\right) \\ &\leq V(x). \end{aligned} \tag{3.7}$$

Give a partition $\{Y_1, \cdots, Y_\rho\}$, $|Y_j| = \frac{|Y|}{\rho}$, $j \in \{1, \cdots, \rho\}$. For any $x \in X$, $\bar{y} \in S(x)$, we can find an index $\bar{j}$ such that $\bar{y} \in Y_{\bar{j}}$. Since $g(x,y)$ is continuously differentiable, by the Mean Value Theorem, for any $y \in Y_{\bar{j}}$, there exists y' which lies between y and $\bar{y}$ such that

$$\exp g(x,y) = \exp g(x,\bar{y}) + [\nabla_y \exp g(x,y')]^T (y - \bar{y}).$$

Let $m_x = \min_{y\in Y_{\bar{j}}} \exp g(x,y)$ and $M_x = \max_{y\in Y_{\bar{j}}} \|\nabla_y \exp g(x,y)\|$, thus for any $y \in Y_{\bar{j}}$,

$$\exp g(x,y) \geq \exp g(x,\bar{y}) - M_x |Y_{\bar{j}}|. \tag{3.8}$$

It follows from the previous proof and (3.8) that

$$\begin{aligned} &\left(\int_Y \exp[\rho(g(x,y) - V(x))]dy\right)^{\frac{1}{\rho}} \\ &\geq \left(\frac{|Y|}{\rho}\left(\min_{Y_1} \exp[\rho(g(x,y) - V(x))] + \cdots + \min_{Y_\rho} \exp[\rho(g(x,y) - V(x))]\right)\right)^{\frac{1}{\rho}} \end{aligned}$$

$$
\begin{aligned}
&\geq \left(\frac{|Y|}{\rho} \min_{Y_{\bar{j}}} \exp[\rho(g(x,y) - V(x))] \right)^{\frac{1}{\rho}} \\
&= \left(\frac{|Y|}{\rho} \left(\min_{Y_{\bar{j}}} \exp[g(x,y)] \exp[-V(x)] \right)^{\rho} \right)^{\frac{1}{\rho}} \\
&= \left(\frac{|Y|}{\rho} \left(\frac{\min_{Y_{\bar{j}}} \exp[g(x,y)]}{\exp[\max_{y \in Y} g(x,y)]} \right)^{\rho} \right)^{\frac{1}{\rho}} \\
&\geq \left(\frac{|Y|}{\rho} \left(\frac{\exp[g(x,\bar{y})] - M_x |Y_{\bar{j}}|}{\exp[g(x,\bar{y})]} \right)^{\rho} \right)^{\frac{1}{\rho}} \\
&= \left(\frac{|Y|}{\rho} \right)^{\frac{1}{\rho}} \left(1 - \frac{M_x}{\exp[g(x,\bar{y})]} \frac{|Y|}{\rho} \right) \\
&\geq \left(\frac{|Y|}{\rho} \right)^{\frac{1}{\rho}} \left(1 - \frac{M_x}{m_x} \frac{|Y|}{\rho} \right).
\end{aligned}
$$

By the monotonicity of the logarithmic function,

$$
\ln \left(\int_Y \exp[\rho(g(x,y) - V(x))] dy \right)^{\frac{1}{\rho}} \geq \frac{1}{\rho} \ln(\rho^{-1}|Y|) + \ln \left(1 - \frac{M_x}{m_x} \frac{|Y|}{\rho} \right).
$$

From the definition of $\gamma_\rho(x)$,

$$
\begin{aligned}
\gamma_\rho(x) - V(x) &= \rho^{-1} \ln \left(\int_Y \exp[\rho(g(x,y) - V(x))] dy \right) \\
&\geq \frac{1}{\rho} \ln(\rho^{-1}|Y|) + \ln \left(1 - \frac{M_x}{m_x} \frac{|Y|}{\rho} \right). \qquad (3.9)
\end{aligned}
$$

We complete the proof.

■

For a given positive integer ρ, we consider the following smoothing problem of (VP):

$$
\begin{array}{lll}
(\mathrm{VP})_\rho & \min\limits_{x \in X} & f(x) \\
& \text{s.t.} & \gamma_\rho(x) \leq 0.
\end{array}
$$

Denote the feasible regions for problem (VP) and $(\mathrm{VP})_\rho$ by $\mathcal{F}$ and $\mathcal{F}_\rho$ respectively. We now investigate the relationships among the stationary points, the optimal solutions for the smoothing problems and the original problem. First we give conditions under which an accumulation point of the stationary sequence of the smoothing problems is a stationary point of the original problem (VP).

Theorem 3.1 *Assume that x_ρ is a stationary point of problem $(\mathrm{VP})_\rho$, i.e., x_ρ satisfies*

$$
0 \in \nabla f(x_\rho) + \lambda_\rho \nabla \gamma_\rho(x_\rho) + \mathcal{N}_X(x_\rho), \qquad (3.10)
$$

$$
\lambda_\rho \gamma_\rho(x_\rho) = 0, \gamma_\rho(x_\rho) \leq 0, \qquad (3.11)
$$

where $\lambda_\rho \geq 0$. Let x^* be an accumulation point of the sequence $\{x_\rho\}$ as $\rho \uparrow \infty$. If either the sequence of the multipliers $\{\lambda_\rho\}$ is bounded or the NNAMCQ for problem (VP) holds at x^*. Then x^* is a stationary point of the problem (VP).

Proof. Without loss of generality, assume that $\lim\limits_{\rho\to\infty} x_\rho = x^*$. By the gradient consistency property of γ_ρ, there exists a subsequence $K_0 \subseteq \mathbf{N}$ such that

$$\lim_{\rho\to\infty,\, \rho\in K_0} \nabla\gamma_\rho(x_\rho) \in \partial V(x^*).$$

When the NNAMCQ holds at x^*, the sequence $\{\lambda_\rho\}$ must be bounded otherwise a contradiction arises. So we only need to prove the result for the case when the sequence $\{\lambda_\rho\}$ is bounded. In this case, there is a subsequence $\bar{K}_0 \subseteq K_0$ such that $\{\lambda_\rho\}$ is convergent. Let $\bar{\lambda} := \lim\limits_{\rho\to\infty,\, \rho\in\bar{K}_0} \lambda_\rho \geq 0$. By letting $\rho \to \infty$ with $\rho \in \bar{K}_0$ in (3.10),

$$0 \in \nabla f(x^*) + \bar{\lambda}\partial V(x^*) + \mathcal{N}_X(x^*). \tag{3.12}$$

Taking a limit in (3.11) for $\rho \in \bar{K}_0$, we get $V(x^*) \leq 0$ and $\bar{\lambda}V(x^*) = 0$ from Definition 3.1. From the above discussion, we know that x^* is a stationary point of (VP). ■

The relationship between the locally optimal solutions of the approximating problems $(\mathrm{VP})_\rho$ and the original problem (VP) can be derived as follows.

Proposition 3.3 *Let x^* be a locally optimal solution of problem* (VP) *and x_ρ is a locally optimal solution of problem* $(\mathrm{VP})_\rho$ *in a neighbourhood of x^* for some sufficiently large ρ. If the calmness condition holds at x^* for problem* (VP) *with the exact penalty λ^*, then we have*

$$0 \leq f(x^*) - f(x_\rho) \leq \lambda^*\varepsilon_\rho(x_\rho), \tag{3.13}$$

where $\varepsilon_\rho(x)$ is defined as in (3.4).

Proof. From the calmness condition, x^* is also a locally optimal solution of problem

$$\min_{x\in X} f(x) + \lambda^* \max\{V(x), 0\}.$$

Thus for sufficiently large $\rho > 0$ we have

$$f(x_\rho) + \lambda^* \max\{V(x_\rho), 0\} \geq f(x^*) + \lambda^* \max\{V(x^*), 0\} = f(x^*). \tag{3.14}$$

By Proposition 3.2, $\max\{V(x_\rho), 0\} \leq \varepsilon_\rho(x_\rho)$. If we also have that $x^* \in \mathcal{F}_\rho$, (3.13) follows by (3.14). ■

Theorem 3.1 and Proposition 3.3 suggest a method for finding stationary points and locally optimal solutions of (VP) by which one solves the approximating problem $(\mathrm{VP})_\rho$

and updates the penalty parameter λ_ρ. However the problem $(\mathrm{VP})_\rho$ is a constrained optimization problem which may be solved by using a penalty method. In this paper, we suggest to use the smoothing projected gradient algorithm [24, Algorithm 3.1]. The idea is to approximate the function $\max\{x, 0\}$ by $\frac{1}{2}(\sqrt{x^2+\rho^{-1}}+x)$ and the value function $V(x)$ by its smoothing function $\gamma_\rho(x)$ to obtain the following smooth penalty problem of $(\mathrm{VP})_\rho$:

$$(\mathrm{VP})_\rho^\lambda \qquad \begin{array}{ll} \min & G_\rho^\lambda(x) := f(x) + \dfrac{\lambda}{2}\left(\sqrt{\gamma_\rho^2(x)+\rho^{-1}}+\gamma_\rho(x)\right) \\ \text{s.t.} & x \in X. \end{array}$$

We now describe the smoothing projected algorithm [24, Algorithm 3.1] applied to our problem (SIP). Note that we denote by P_X the projection operator onto X, that is,

$$P_X[x] := \operatorname{argmin}\{\|z-x\| : z \in X\}.$$

Algorithm 3.1 *Let $\{\beta, \gamma, \sigma_1, \sigma_2\}$ be constants in $(0,1)$ with $\sigma_1 \le \sigma_2$, $\{\sigma, \sigma', \hat{\eta}\}$ be constants in $(1,\infty)$ and ϵ be a small positive number. Choose an initial point $x^0 \in X$, an initial smoothing parameter $\rho_0 > 0$, an initial penalty parameter $\lambda_0 > 0$ and set $k := 0, s := 0$.*

1. *Let $z_0^k := x^k$ and $z_{s+1}^k := P_X[z_s^k - \alpha_s \nabla G_{\rho_k}^{\lambda_k}(z_s^k)]$, where $\alpha_s := \beta^{l_s}$, $l_s \in \{0, 1, 2\cdots\}$ is the smallest number satisfying*

$$G_{\rho_k}^{\lambda_k}(z_{s+1}^k) - G_{\rho_k}^{\lambda_k}(z_s^k) \le \sigma_1 \nabla G_{\rho_k}^{\lambda_k}(z_s^k)^T \left(z_{s+1}^k - z_s^k\right) \tag{3.15}$$

   and $\beta^{l_s} \ge \gamma$, or $\bar{\alpha}_s := \beta^{l_s-1}$ such that $\bar{z}_{s+1}^k := P_X[z_s^k - \bar{\alpha}_s \nabla G_{\rho_k}^{\lambda^k}(z_s^k)]$ satisfies

$$G_{\rho_k}^{\lambda_k}(\bar{z}_{s+1}^k) - G_{\rho_k}^{\lambda_k}(z_s^k) > \sigma_2 \nabla G_{\rho_k}^{\lambda_k}(z_s^k)^T \left(\bar{z}_{s+1}^k - z_s^k\right). \tag{3.16}$$

   If

$$\frac{\|z_{s+1}^k - z_s^k\|}{\alpha_s} < \hat{\eta}\rho_k^{-1}, \tag{3.17}$$

   set $x^{k+1} := z_{s+1}^k$, $\rho_{k+1} := \sigma\rho_k$, $s := 0$, go to Step 2. Otherwise, set $s = s+1$, and go to Step 1.

2. *If*

$$\gamma_{\rho_k}(x^{k+1}) \le 0, \tag{3.18}$$

   go to Step 3. Otherwise, set $\lambda_{k+1} := \sigma'\lambda_k$, $k = k+1$, and go to Step 1.

3. If

$$\|P_X[x^{k+1} - \nabla G_{\rho_k}^{\lambda_k}(x^{k+1})] - x^{k+1}\| = 0, \tag{3.19}$$

or $\|x^{k+1} - x^k\| \le \epsilon$, *terminate. Otherwise, set* $k = k+1$, *and go to Step 1.*

Theorem 3.2 [24, Theorems 3.1, 3.2 and 3.3] *Let* $\{x^k\}$ *be a sequence generated by Algorithm 3.1.*

(1) *If* x^* *is an accumulation point of* $\{x^k\}$ *and the sequence* $\{\lambda_k\}$ *is bounded, then* x^* *is a stationary point of* (VP).

(2) *If* $\lim_{k\to\infty} x^k = x^*$ *and the* ENNAMCQ *holds at* x^*, *then the sequence* $\{\lambda_k\}$ *is bounded and hence* x^* *is a stationary point of problem* (VP).

(3) *Assume that the* ENNAMCQ *holds for* (VP) *at any point* $x \in X$ *satisfying* $V(x) \ge 0$. *Then any accumulation point of* $\{x^k\}$ *is a stationary point of* (VP).

Suppose that Algorithm 3.1 terminates within finite iterations when the condition (3.19) holds at x^{k+1}. When certain second order sufficient condition holds, we can estimate the true optimal objective function value for a nearby locally optimal solution with the value $f(x^{k+1})$. To derive these results, we first state and prove the following lemma.

Lemma 3.2 *Let* $\{x^k\}$ *be a sequence generated by Algorithm 3.1. Assume that* $\bar{x}$ *is an accumulation point of the sequence. Let* x^* *be a locally optimal solution of problem* (VP) *in a neighbourhood of* $\bar{x}$. *If* $\{\lambda_k\}$ *is bounded, then for sufficiently large* k,

$$f(x^*) - f(x^{k+1}) \le c\rho_{k+1}^{-1}, \tag{3.20}$$

where $c := L\hat{\eta}\frac{\sigma}{\sigma-1}$ *with* $\sigma \in (1, +\infty)$ *as in Algorithm 3.1 and* $L := |\nabla f(x^{k+1})| + 1$.

Proof. Since $\{\lambda_k\}$ is bounded, we know that $\gamma_{\rho_k}(x^{k+1}) \le 0$ from condition (3.18). Since $\gamma_\rho(x)$ is a smoothing function of $V(x)$, it follows that $V(\bar{x}) \le 0$. Therefore by the optimality of x^*, $f(x^*) \le f(\bar{x})$.

When k is large enough, from the Taylor expansion,

$$f(\bar{x}) = f(x^{k+1}) + \nabla f(x^{k+1})^T(\bar{x} - x^{k+1}) + o(\|\bar{x} - x^{k+1}\|).$$

Thus,

$$\begin{aligned} |f(\bar{x}) - f(x^{k+1})| &= |\nabla f(x^{k+1})^T(\bar{x} - x^{k+1}) + o(\|\bar{x} - x^{k+1}\|)| \\ &\le |\nabla f(x^{k+1})|\|\bar{x} - x^{k+1}\| + \|\bar{x} - x^{k+1}\| \\ &\le (|\nabla f(x^{k+1})| + 1)\|\bar{x} - x^{k+1}\|. \end{aligned}$$

By condition (3.17) and the continuity of f on a bounded set,

$$\begin{aligned} |f(\bar{x}) - f(x^{k+1})| &\leq L\|\bar{x} - x^{k+1}\| \\ &\leq L\sum_{l\geq k}\|x^{l+1} - x^{l+2}\| < L\hat{\eta}\sum_{l\geq k}\rho_{l+1}^{-1} \\ &= L\hat{\eta}\rho_{k+1}^{-1}(1 + \sigma^{-1} + \sigma^{-2} + \cdots) = c\rho_{k+1}^{-1}. \end{aligned}$$

By the Taylor expansion we may take $L := |\nabla f(x^{k+1})| + 1$. Thus the proof is complete. ■

Theorem 3.3 *Assume that f is twice continuously differentiable. Let x^{k+1} be a point generated by Algorithm 3.1 which is terminated when condition* (3.19) *holds. Let $\mu_k := \mu_{\rho_k}^{\lambda_k}(x^{k+1})$ with $\mu_\rho^\lambda(x) := \frac{\lambda}{2}\left(1 + \frac{\gamma_\rho(x)}{\sqrt{\gamma_\rho^2(x)+\rho^{-1}}}\right)$. Suppose that for every nonzero vector $d \in \mathcal{T}_X(x^{k+1})$,*

$$d^T\nabla^2 L_{\rho_k}^{\lambda_k}(x^{k+1})d > 0, \tag{3.21}$$

where $L_\rho^\lambda(x) := f(x) + \mu_k\gamma_\rho(x)$ and there exists an $\varepsilon > 0$ such that $|\mu_k\gamma_{\rho_k}(x^{k+1})| \leq \varepsilon$. Assume that x^* is a locally optimal solution of problem (VP)*and $\{\lambda_k\}$ is bounded. When $|S(x^{k+1})| < 1$, we have*

$$-(\varepsilon + \mu_k \max\{0, \rho_k^{-1}\ln|Y|\}) \leq f(x^*) - f(x^{k+1}) \leq c\rho_{k+1}^{-1}, \tag{3.22}$$

where c is a constant denoted in Lemma 3.2. When $|S(x^{k+1})| \geq 1$, we have

$$-(\varepsilon + \mu_k\rho_k^{-1}\ln|Y|) \leq f(x^*) - f(x^{k+1}) \leq 0. \tag{3.23}$$

Proof. By condition (3.19) we have

$$0 \in \nabla G_{\rho_k}^{\lambda_k}(x^{k+1}) + \mathcal{N}_X(x^{k+1}) = \nabla f(x^{k+1}) + \mu_k\nabla\gamma_{\rho_k}(x^{k+1}) + \mathcal{N}_X(x^{k+1}).$$

Since the second order condition (3.21) holds, x^{k+1} is a locally optimal solution of the penalized problem of $(VP)_{\rho_k}$:

$$\min_{x\in X} f(x) + \mu_k\gamma_{\rho_k}(x). \tag{3.24}$$

Thus,

$$f(x^{k+1}) - \varepsilon \leq f(x^{k+1}) + \mu_k\gamma_{\rho_k}(x^{k+1}) \leq f(x^*) + \mu_k\gamma_{\rho_k}(x^*). \tag{3.25}$$

By Proposition 3.2, if $|S(x^*)| \geq 1$, $\gamma_{\rho_k}(x^*) \leq \rho_k^{-1}\ln|Y|$ while if $|S(x^*)| < 1$, $\gamma_{\rho_k}(x^*) \leq 0$.

When $|S(x^{k+1})| < 1$, (3.22) follows from the Lemma 3.2 and (3.25).

When $|S(x^{k+1})| \geq 1$, we have $V(x^{k+1}) \leq 0$ from (3.18) and Proposition 3.2, thus x^{k+1} is a feasible solution of (VP). Then $f(x^*) \leq f(x^{k+1})$, which guarantees (3.23). ■

Theorem 3.4 *Assume that f is twice continuously differentiable. Let x^{k+1} be a point generated by Algorithm 3.1 which is terminated when condition* (3.19) *holds. Furthermore suppose that the second order sufficient condition (SOSC) for problem* $(\mathrm{VP})_{\rho_k}^{\lambda_k}$ *holds: for every nonzero vector $d \in \mathcal{T}_X(x^{k+1})$,*

$$d^T \nabla^2 G_{\rho_k}^{\lambda_k}(x^{k+1}) d > 0. \tag{3.26}$$

For any neighbourhood of x^{k+1}, if there exists a locally optimal solution of problem (VP) *which is denoted by x^* and $\{\lambda_k\}$ is bounded, then*

$$-\lambda_k(2\rho_k^{-\frac{1}{2}} + \max\{0, \rho_k^{-1}\ln|Y|\}) \le f(x^*) - f(x^{k+1}) \le c\rho_{k+1}^{-1}, \tag{3.27}$$

where c is a constant denoted in Lemma 3.2. In particular, when $|S(x^{k+1})| \ge 1$, we have

$$-\lambda_k(2\rho_k^{-\frac{1}{2}} + \rho_k^{-1}\ln|Y|) \le f(x^*) - f(x^{k+1}) \le 0. \tag{3.28}$$

Proof. From the definition of $G_\rho^\lambda(\cdot)$ and $\gamma_{\rho_k}(x^{k+1}) \le 0$, we have

$$\begin{aligned} G_{\rho_k}^{\lambda_k}(x^{k+1}) &= f(x^{k+1}) + \frac{\lambda_k}{2}\left(\sqrt{\gamma_{\rho_k}^2(x^{k+1}) + \rho_k^{-1}} + \gamma_{\rho_k}(x^{k+1})\right) \\ &\ge f(x^{k+1}) + \frac{\lambda_k}{2}\left(|\gamma_{\rho_k}(x^{k+1})| + \gamma_{\rho_k}(x^{k+1})\right) \\ &= f(x^{k+1}). \end{aligned} \tag{3.29}$$

Since the conditions (3.19) and (3.26) hold, we know that x^{k+1} is a locally optimal solution of problem $(\mathrm{VP})_{\rho_k}^{\lambda_k}$, thus

$$G_{\rho_k}^{\lambda_k}(x^{k+1}) \le G_{\rho_k}^{\lambda_k}(x^*) = f(x^*) + \frac{\lambda_k}{2}\left(\sqrt{\gamma_{\rho_k}^2(x^*) + \rho_k^{-1}} + \gamma_{\rho_k}(x^*)\right). \tag{3.30}$$

From (3.29) - (3.30), we have

$$f(x^*) \ge f(x^{k+1}) - \frac{\lambda_k}{2}\left(\sqrt{\gamma_{\rho_k}^2(x^*) + \rho_k^{-1}} + \gamma_{\rho_k}(x^*)\right). \tag{3.31}$$

Consider the case where $\gamma_{\rho_k}(x^*) \le 0$,

$$\begin{aligned} \sqrt{\gamma_{\rho_k}^2(x^*) + \rho_k^{-1}} + \gamma_{\rho_k}(x^*) &\le \sqrt{(\gamma_{\rho_k}(x^*) - \rho_k^{-\frac{1}{2}})^2} + \gamma_{\rho_k}(x^*) \\ &= -\gamma_{\rho_k}(x^*) + \rho_k^{-\frac{1}{2}} + \gamma_{\rho_k}(x^*) \\ &= \rho_k^{-\frac{1}{2}}. \end{aligned} \tag{3.32}$$

If $\gamma_{\rho_k}(x^*) > 0$, then x^* is not a feasible solution of problem $(VP)_{\rho_k}$, which implies that $|S(x^*)| \ge 1$ by Proposition 3.2. From (3.2), we have $\gamma_{\rho_k}(x^*) \le \rho_k^{-1}\ln|Y|$, thus,

$$\begin{aligned} \sqrt{\gamma_{\rho_k}^2(x^*) + \rho_k^{-1}} + \gamma_{\rho_k}(x^*) &\le \sqrt{(\gamma_{\rho_k}(x^*) + \rho_k^{-\frac{1}{2}})^2} + \gamma_{\rho_k}(x^*) \\ &= 2\gamma_{\rho_k}(x^*) + \rho_k^{-\frac{1}{2}} \\ &\le 2\rho_k^{-1}\ln|Y| + \rho_k^{-\frac{1}{2}}. \end{aligned} \tag{3.33}$$

Thus we have

$$f(x^*) - f(x^{k+1}) \geq -\frac{\lambda_k}{2}(\rho_k^{-\frac{1}{2}} + 2\max\{0, \rho_k^{-1}\ln|Y|\}) \tag{3.34}$$

from (3.31)-(3.33).

Thus, (3.27) follows from the Lemma 3.2. Furthermore, if $|S(x^{k+1})| \geq 1$, then x^{k+1} is a feasible solution of problem (VP) from Proposition 3.2. Since x^* is a locally optimal solution of problem (VP), we have $f(x^{k+1}) \geq f(x^*)$, which guarantees (3.28) together with (3.34). The proof is complete. ∎

4 Numerical results

In this section, we illustrate the computational behaviour and convergence results presented in Section 3. For the numerical illustration in this section we implement Algorithm 3.1 in MATLAB.

In Examples 4.1 and 4.2 there are two semi-infinite constraints. We now explain how to modify our results to the multiple semi-infinite constraints case:

$$\begin{array}{lll} (\mathrm{SIP})_p & \min\limits_{x} & f(x) \\ & \text{s.t.} & g_j(x,y) \leq 0, \quad \forall y \in Y, \quad j = 1, \cdots, p \\ & & x \in X, \end{array}$$

where $g_j : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ $(j = 1, \ldots, p)$ are continuously differentiable and the rest of the problem data follow the same assumptions as (SIP).

We first extend the extended EMFCQ to $(\mathrm{SIP})_p$. Let $S_j(\bar{x})$ denotes the set of all maximizers of $g_j(\bar{x}, y)$ over $y \in Y$ and denote the index set

$$I(\bar{x}) := \{j = 1, \cdots, p : g_j(\bar{x}, y) \geq 0 \quad \forall y \in S_j(\bar{x})\}.$$

We say that a point $\bar{x} \in X$ satisfies the extended EMFCQ for problem $(\mathrm{SIP})_p$ if either $g_j(\bar{x}, y) < 0$ for all $y \in Y$ and all $j = 1, \cdots, p$ or there exists a direction $d \in \mathrm{int}\, \mathcal{T}_X(\bar{x})$ such that $\nabla_x g_j(\bar{x}, y)^T d < 0,$ for all $y \in S_j(\bar{x})$ and all $j \in I(\bar{x})$. The following result provides a sufficient condition for the extended EMFCQ that is very easy to verify.

Proposition 4.1 *Let $\bar{x}$ be an interior point of X. If either $g_j(\bar{x}, y) < 0$ for all $y \in Y$ and all $j = 1, \cdots, p$ or there exists an index $i \in \{1, \cdots, n\}$ such that $\frac{\partial g_j(\bar{x},y)}{\partial x_i} > 0$ or $\frac{\partial g_j(\bar{x},y)}{\partial x_i} < 0$, for any $y \in S_j(\bar{x})$, $j \in I(\bar{x})$, then the extended EMFCQ holds at $\bar{x}$.*

Proof. Consider the case when there exists an index i_0 such that $\frac{\partial g_j(\bar{x},y)}{\partial x_{i_0}} > 0$, for any $y \in S_j(\bar{x})$, and any $j \in I(\bar{x})$. We select a direction d such that $d_i := 0$ if $i \neq i_0$ and

$d_{i_0} := -1$. Then we have

$$\nabla_x g_j(\bar{x}, y)^T d < 0, \text{ for all } y \in S_j(\bar{x}), \ j \in I(\bar{x}). \tag{4.1}$$

When there exists an index i_0 such that $\frac{\partial g_j(\bar{x},y)}{\partial x_{i_0}} < 0$, for any $y \in S_j(\bar{x})$ and any $j \in I(\bar{x})$, we select $d_i := 0$ if $i \neq i_0$ and $d_{i_0} := 1$. Thus (4.1) also holds. Therefore, the extended EMFCQ holds at $\bar{x}$. ■

For any $j = 1, \cdots, p$, denote the value functions by

$$V_j(x) := \max\{g_j(x, y) : y \in Y\}.$$

We approximate the value function $V_j(x)$ by the corresponding integral entropy function:

$$\gamma_\rho^j(x) := \rho^{-1} \ln \left(\int_Y \exp[\rho g_j(x, y)] dy \right), j = 1, \cdots, p.$$

We modify Algorithm 3.1 and apply it to $(\mathrm{SIP})_p$ in a straight forwarded manner. To verify that an accumulation point is a stationary point, we give the following proposition.

Proposition 4.2 *Let $\{x^{k+1}\}$ be a sequence generated by Algorithm 3.1. Suppose that for any $j = 1, \cdots, p$, the exact penalty parameter sequence $\{\lambda_j^k\}_k$ is bounded, and there exist a subset $K \subset \mathbf{N}$ such that $\lim_{k\to\infty, k\in K} x^k = x^*$ and a subset $J \subseteq \{1, \cdots, p\}$ such that $\lim_{k\to\infty, k\in K} \gamma_{\rho_k}^j(x^{k+1}) = 0$, $j \in J$. Let $v_j := \lim_{k\to\infty, k\in K} \nabla\gamma_{\rho_k}^j(x^{k+1}) \quad j \in J$. If*

$$\nabla f(x^*)^T d \geq 0$$

for all d in the linearization cone of the feasible region:

$$\mathcal{L}(x^*) := \{d \in \mathcal{T}_X(x^*) : v_j^T d \leq 0 \quad j \in J\},$$

then x^* is a stationary point of (VP).

Proof. Since for any $j = 1, \cdots, p$, $\{\lambda_j^k\}_k$ is bounded, by the rule of Algorithm 3.1 $\gamma_{\rho_k}^j(x^{k+1}) \leq 0$ for sufficiently large k. Since $\gamma_\rho^j(x)$ is a smoothing function of the value function $V_j(x)$ and the gradient consistency holds, we have $V_j(x^*) = \lim_{k\to\infty} \gamma_{\rho_k}^j(x^{k+1}) = 0$ and $v_j \in \partial V_j(x^*)$, $j = 1, \cdots, p$. By assumptions, $d = 0$ is an optimal solution to the following linearized problem:

$$\begin{aligned} \min_d \quad & \Phi(d) := \nabla f(x^*)^T d \\ \text{s.t.} \quad & v_j^T d \leq 0, \ j \in J, \\ & d \in \mathcal{T}_X(x^*). \end{aligned}$$

Since the objective function and the constraint functions are all linear in variable d, the KKT condition holds at the optimal solution. Hence there exist multipliers $\mu_j \geq 0$, $j \in J$ such that

$$0 \in \nabla f(x^*) + \sum_{j \in J} \mu_j v_j + \mathcal{N}_X(x^*).$$

Thus x^* is a stationary point of (VP). ■

Example 4.1 [11, Example 5.1] Finding the Chebyshev approximation of the function $\sin(\pi y)$ by a quadratic function on the interval $Y = [0, 1]$ amounts to solving the following SIP:

$$\begin{array}{ll} \min & f(x) := x_4 \\ \text{s.t.} & g_1(x,y) := \sin(\pi y) - x_3 y^2 - x_2 y - x_1 - x_4 \leq 0, \quad \forall\, y \in Y = [0,1], \\ & g_2(x,y) := -\sin(\pi y) + x_3 y^2 + x_2 y + x_1 - x_4 \leq 0, \quad \forall\, y \in Y = [0,1]. \end{array}$$

In our numerical experiment, we set the initial point $x^0 = (1, 5, -3, 3)$, the parameters $\beta = 0.9,\ \gamma = 0.5,\ \sigma_1 = \sigma_2 = 10^{-6},\ \rho_0 = \lambda_0 = 100,\ \hat{\eta} = 5 * 10^6,\ \sigma = \sigma' = 10, \epsilon = 10^{-12}$.

Since the stopping criteria $\|x^{k+1} - x^k\| \leq \epsilon$ holds, we terminate at the 10th iteration (185 CPU seconds) with $\lambda_k = 100$ and $\rho_k = 1.0 * 10^{12}$. We obtain an accumulation point $x^* \approx (-0.028, 4, -4, 0.028)$.

We select $d = (0, 0, 0, 1)^T$. Since $\nabla_x g_j(x,y)^T d = -1 < 0$ and for any $y \in Y$, $j = 1, 2$, by Proposition 4.1 the extend EMFCQ holds at every point $x \in \mathbb{R}^4$. Thus the sequence $\{\lambda_k^j\}$, $j = 1, 2$ must be bounded and any accumulation point must be a stationary point of (VP). We now verify this. Since $\lim_{k\to\infty} \gamma^1_{\rho_k}(x^{k+1}) \approx -7.053 * 10^{-6}$, $\lim_{k\to\infty} \gamma^2_{\rho_k}(x^{k+1}) \approx -7.169 * 10^{-6}$ (which we consider them to be zero), and

$$\begin{aligned} v_1 &= \lim_{k\to\infty} \nabla \gamma^1_{\rho_k}(x^{k+1}) \approx (-1, -0.5, -0.3618, -1), \\ v_2 &= \lim_{k\to\infty} \nabla \gamma^2_{\rho_k}(x^{k+1}) \approx (1, 0.4933, 0.3656, -0.9999), \end{aligned}$$

the linearization cone

$$\begin{aligned} \mathcal{L}(x^*) : \ &= \ \{d \in \mathbb{R}^4 : v_1^T d \leq 0, v_2^T d \leq 0\} \\ &\approx \ \{d \in \mathbb{R}^4 : d = \alpha_1(-0.0064, 0.0056, 0.01, 0) + \alpha_2(1.473, -2.967, 0, 0.01) \\ &\qquad + \alpha_3(-0.732, 1.4834, 0, 0) + \alpha_4(-0.742, 1.4834, 0, 0), \alpha_i \geq 0, i = 1, \cdots, 4\}. \end{aligned}$$

Since $\nabla f(x^*) \approx (0, 0, 0, 1)$, it follows that $\nabla f(x^*)^T d \geq 0$, for any $d \in \mathcal{L}(x^*)$.

Hence by Proposition 4.2, x^* is a stationary point of problem (VP).

Example 4.2 [11, Example 5.2]Consider the SIP associated with the design centring problem:

$$\begin{aligned}
\min \quad & f(x) := -x_3 \\
\text{s.t.} \quad & g_1(x,y) := 0.3\sin(\pi(x_1 + x_3\cos y)) - (x_2 + x_3 \sin y) \le 0, \quad \forall\, y \in Y = [0,1], \\
& g_2(x,y) := (x_1 + x_3\cos y)^2 + 0.3(x_2 + x_3\sin y)^2 - 1 \le 0, \quad \forall\, y \in Y = [0,1].
\end{aligned}$$

In our numerical experiment, we set the initial point $x^0 = (0.5, 0.5, 0.5)$, the parameters $\beta = 0.9,\ \gamma = 0.5,\ \sigma_1 = 0.95,\ \sigma_2 = 0.98,\ \rho_0 = \lambda_0 = 100,\ \hat{\eta} = 5 * 10^5,\ \sigma = 20,\ \sigma' = 10,$ $\epsilon = 10^{-8}$.

Since the stopping criterion $\|x^{k+1} - x^k\| \le \epsilon$ holds, we terminate at the 26th iteration (55 CPU seconds) with $\lambda_k = 1000,\ \rho_k = 6.71 * 10^{35}$ and the objective function value $-x_3^{k+1} = -0.77686$. We obtain an accumulation point $x^* \approx (0, 0.962, 0.77686)$. Note that

$$\nabla_x g_1(x,y) = \begin{pmatrix} 0.3\pi\cos(\pi(x_1 + x_3\cos y)) \\ -1 \\ 0.3\cos(\pi(x_1 + x_3\cos y))\cos y - \sin y \end{pmatrix}$$

$$\nabla_x g_2(x,y) = \begin{pmatrix} 2(x_1 + x_3\cos y) \\ 0.6(x_2 + x_3\sin y) \\ 2x_1 + 0.6x_2\sin y + x_3(2\cos^2 y + 0.6\sin^2 y) \end{pmatrix}.$$

We select $d = (-1, 0, 0)^T$, then $\nabla_x g_j(x,y)^T d < 0$, for any $y \in Y$, $x \in N(x^*)$ and $j = 1, 2$, which implies the extended EMFCQ holds. Thus the sequence $\{\lambda_k^j\}$, $j = 1, 2$ must be bounded and hence the accumulation point must be a stationary point of (VP).

Since $\lim_{k\to\infty} \gamma_{\rho_k}^1(x^{k+1}) \approx -2.71 * 10^{-4} < 0$, $\lim_{k\to\infty} \gamma_{\rho_k}^2(x^{k+1}) \approx -2.22 * 10^{-16}$ (which we consider it to be zero), and

$$\begin{aligned}
v_1 &= \lim_{k\to\infty} \nabla\gamma_{\rho_k}^1(x^{k+1}) \approx (0.474, -1, 1.106), \\
v_2 &= \lim_{k\to\infty} \nabla\gamma_{\rho_k}^2(x^{k+1}) \approx (-0.706, 0.824, 1.55),
\end{aligned}$$

the linearization cone

$$\begin{aligned}
\mathcal{L}(x^*) : \ &= \ \{d \in \mathbb{R}^3 : v_2^T d \le 0\} \\
&\approx \ \{d \in \mathbb{R}^3 : d = \alpha_1(1.168, 1, 0) + \alpha_2(-2.2, 0, -1) \\
&\qquad + \alpha_3(1.417, 0, 0), \alpha_i \in R_+, i = 1, \cdots, 3\}.
\end{aligned}$$

Since $\nabla f(x^*) \approx (0, 0, -1)$, it follows that $\nabla f(x^*)^T d \ge 0$ for $d \in \mathcal{L}(x^*)$. Hence x^* is a stationary point of problem (VP).

In the above two examples, we almost terminated at the same point as in [11] and we did not need too much cpu time.

Example 4.3 [7, Example 6]Consider the SIP:

$$\begin{array}{ll} \min & f(x) := (x_1 - 2x_2 + 5x_2^2 - x_2^3 - 13)^2 + (x_1 - 14x_2 + x_2^2 + x_2^3 - 29)^2 \\ \text{s.t.} & g(x,y) := x_1^2 + 2x_2 y^2 + \exp(x_1 + x_2) - \exp(y) \leq 0, \quad \forall\, y \in Y = [0,1]. \end{array}$$

In our numerical experiment, we choose the initial point $x^0 = (1, 2)$ and the parameters $\beta = 0.9,\ \gamma = 0.5,\ \sigma_1 = \sigma_2 = 10^{-6},\ \rho_0 = \lambda_0 = 100,\ \hat{\eta} = 2 * 10^5,\ \sigma = \sigma' = 10,\ \epsilon = 10^{-12}$.

Since the stopping criterion $\|x^{k+1} - x^k\| \leq \epsilon$ holds, we terminate at the 8th iteration (79.6 CPU seconds) such that $f(x^{k+1}) = 97.159$, where $\lambda_k = 1000$ and $\rho_k = 1.0 * 10^{10}$. We obtain an accumulation point $x^* \approx (0.719951, -1.450488)$. Note that,

$$\nabla_x g(x,y) = \begin{pmatrix} 2x_1 + \exp(x_1 + x_2) \\ 2y^2 + \exp(x_1 + x_2) \end{pmatrix},$$

thus $\frac{\partial g(x,y)}{\partial x_2} > 0$, for any $y \in Y$, and for all x, which implies the extended EMFCQ holds at all x from Proposition 4.1. Thus the sequence $\{\lambda_k\}$ must be bounded and hence the accumulation point must be a stationary point of (VP).

Since $\lim\limits_{k\to\infty} \gamma_{\rho_k}(x^{k+1}) \approx -7.074 * 10^{-5}$ (which we consider it to be zero) and

$$v := \lim_{k\to\infty} \nabla\gamma_{\rho_k}(x^{k+1}) \approx (1.92, 0.48),$$

the linearization cone

$$\mathcal{L}(x^*) \quad \approx \quad \{d \in \mathbb{R}^2 : d = \alpha_1(-0.25, 1) + \alpha_2(-0.52, 0), \alpha_i \geq 0, i = 1, 2\}.$$

Since $\nabla f(x^*) \approx (-9.458, -2.37)$, it follows that $\nabla f(x^*)^T d \geq 0$ for any $d \in \mathcal{L}(x^*)$. Hence x^* is indeed a stationary point of problem (VP).

The last two examples satisfy the second order conditions introduced in Theorems 3.3 and 3.4. So we can give estimates for locally optimal solutions.

Example 4.4 [7, Example 2]

$$\begin{array}{ll} \min & f(x) := \dfrac{1}{3}x_1^2 + \dfrac{1}{2}x_1 + x_2^2 \\ \text{s.t.} & g(x,y) := (1 - x_1^2 y^2)^2 - x_1 y^2 - x_2^2 + x_2 \leq 0, \qquad \forall\, y \in Y = [0,1]. \end{array}$$

In our numerical experiment, we choose the initial point $x^0 = (0, 0)$, the parameters $\beta = 0.9,\ \gamma = 0.5,\ \sigma_1 = 10^{-6},\ \sigma_2 = 10^{-6},\ \rho_0 = 100,\ \lambda_0 = 100,\ \hat{\eta} = 5 * 10^5,\ \sigma = \sigma' = 10.$

We terminated after 9 iterations (4664 CPU seconds) when $\|\nabla G_{\rho_k}^{\lambda_k}(x^{k+1})\| = 8.9 * 10^{-6}$ (which we consider it to be zero) with $x^{k+1} = (-0.750008, -0.618064)$, $f(x^{k+1}) = 0.194540$

and $\gamma_{\rho_k}(x^{k+1}) = -6.719 * 10^{-5}$, $\mu_k := \mu_{\rho_k}^{\lambda_k}(x^{k+1}) = 0.5527$, where $\lambda_k = 1000$ and $\rho_k = 1.0 * 10^{11}$.

Note that,

$$\nabla_x g(x,y) = \begin{pmatrix} -4(1-x_1^2y^2)x_1 - y^2 \\ -2x_2+1 \end{pmatrix},$$

for all x, there exists a $d \in \mathbb{R}^2$ with $d_1 = 0$ such that $\nabla_x g(x,y)^T d > 0$, which implies the extended EMFCQ holds. Thus $\{\lambda_k\}$ is bounded.

Since

$$\nabla^2 L_{\rho_k}^{\lambda_k}(x^{k+1}) = \begin{pmatrix} 0.6667 & 0 \\ 0 & 0.8944 \end{pmatrix},$$

is a positive definite matrix, the second order condition is satisfied. Obviously $|S(x^{k+1})| \leq 1$ since $|Y| = 1$. Assume that x^* is a locally optimal solution of problem (VP) in a neighbourhood of x^{k+1}. We also note that x^{k+1} does not change after $k \geq 8$. By Theorem 3.3, we can estimate the value of $f(x^*)$.

Since $Y = [0,1]$, $\ln|Y| = 0$. We also have $\varepsilon = |\mu_k \gamma_{\rho_k}(x^{k+1})| = 3.714 * 10^{-5}$, $c\rho_{k+1}^{-1} = 1.006 * 10^{-6}$. Thus using Theorem 3.3 we conclude that if x^* is a locally optimal solution of problem (VP) in a neighbourhood of x^{k+1}, then we must have

$$0.19454 - 3.714 * 10^{-5} \leq f(x^*) \leq 0.19454 + 1.006 * 10^{-6}.$$

Example 4.5 [7, Example 14]Consider the SIP

$$\begin{aligned} \min \quad & f(x) := c^2 \exp(x_1) + \exp(x_2) \\ \text{s.t.} \quad & g(x,y) := y - \exp(x_1 + x_2), \quad \forall\, y \in Y = [0,1]. \end{aligned}$$

In [7], the author used a projected method to solve the first order formulation of the problem and get a solution $x^* = (-\ln|c|, \ln|c|)$, $y^* = 1$ such that $f(x^*) = 2|c|$.

In [23], $c = 1.5$ is set and a semismooth Newton method is used to obtain $x^* = (-0.405, 0.405)$ such that $f(x^*) = 3$.

In our numerical experiment, we set $c = 1.5$ and choose the initial point $x^0 = (1, 0.5)$, the parameters $\beta = 0.9$, $\gamma = 0.5$, $\sigma_1 = \sigma_2 = 10^{-6}$, $\rho_0 = \lambda_0 = 100$, $\hat{\eta} = 2 * 10^5$, $\sigma = \sigma' = 10$.

We terminated after 9 iterations (4696 CPU seconds) when $\|\nabla G_{\rho_k}^{\lambda_k}(x^{k+1})\| = 1.68 * 10^{-5}$ (which we consider it to be zero) with $x^{k+1} = (-0.405477, 0.405439)$, $f(x^{k+1}) = 3$, $\gamma_{\rho_k}(x^{k+1}) = -1.26 * 10^{-5}$ and $\mu_{\rho_k}^{\lambda_k}(x^{k+1}) = 1.5$, where $\lambda_k = 100$ and $\rho_k = 1.0 * 10^{11}$.

Obviously, $|S(x^{k+1})| \leq 1$. Note that,

$$\nabla_x g(x,y) = \begin{pmatrix} -\exp(x_1+x_2) \\ -\exp(x_1+x_2) \end{pmatrix},$$

thus $\frac{\partial g(x,y)}{\partial x_1} < 0$, for any $y \in Y$ for all x, which implies the extend EMFCQ holds at each point x from Proposition 4.1. Consequently the penalty parameter sequence $\{\lambda_k\}$ is bounded.

Since

$$\nabla^2 G_{\rho_k}^{\lambda_k}(x^{k+1}) = \begin{pmatrix} 227152 & 227150 \\ 227150 & 227152 \end{pmatrix},$$

is a positive definite matrix, the SOSC in Theorem 3.4 is satisfied.

Since $Y = [0,1]$, $\ln|Y| = 0$. By calculation, we obtain $c\rho_{k+1}^{-1} = 6.936 * 10^{-7}$. Thus by Theorem 3.4, we conclude that if x^* is a locally optimal solution of problem (VP) in a neighbourhood of x^{k+1}, then we must have

$$3 - 6.32 * 10^{-4} \leq f(x^*) \leq 3 + 6.936 * 10^{-7}.$$

References

[1] E.J. Anderson and P. Nash, Linear Programming in Infinite-Dimension over Topological Vector Space: Theory and Applications, John Wiley & Sons, 1987.

[2] B. Bhattacharjee, P. Lemonidis, W.H. Green Jr. and P.I. Barton, *Global solution of semi-infinite programs*, Math. Program., **103**(2005), 283-307.

[3] M.J. Cánovas, A. Hantoute, M.A. López and J. Parra, *Stability of indices in the KKT conditions and metric regularity in convex semi-infinite optimization*, J. Optim. Theory Appl., **139**(2008), 485-500.

[4] X. Chen, R.S. Womersley and J.J. Ye, *Minimizing the condition number of a Gram matrix*, SIAM J. Optim., **21**(2011), 127-148.

[5] F.H. Clarke, Optimization and Nonsmooth Analysis, Wiley-Interscience, New York, 1983.

[6] F.H. Clarke, Yu.S. Ledyaev, R.J. Stern and P.R. Wolenski, Nonsmooth Analysis and Control Theory, Springer, New York, 1998.

[7] I.D. Coope and G.A. Watson, *A projected Lagrangian algorithm for semi-infinite programming*, Math. Program., **32**(1985), 337-356.

[8] J.M. Danskin, The Theory of Max-Min and its Applications to Weapons Allocation Problems, Springer, New York, 1967.

[9] S.C. Fang, C.J. Lin and S.-Y. Wu, *Solving quadratic semi-infinite programming problems by using relaxed cutting-plane scheme*, J. Comput. Appl. Math., **129**(2001), 89-104.

[10] C.A. Floudas, Deterministic Global Optimization, Theory, Methods and Applications, Kluwer, Dordrecht, 2000.

[11] C.A. Floudas and O. Stein, *The adaptive convexification algorithm: a feasible point method for semi-infinite programming*, SIAM J. Optim., **18**(2007), 1187-1208.

[12] M.A. Goberna and M. A. López, Linear Semi-infinite Optimization, Wiley, Chichester, 1998.

[13] F. Guerra Vśzquez, J.-J. Rückmann, O. Stein and G. Still, *Generalized semi-infinite programming: a tutorial*, J. Compu. Appl. Math., **217**(2008), 394-419.

[14] S.A. Gustafson and K.O. Kortanek, *Numerical treatment of a class of semi-infinite programming problems*, Naval Res. Logis. Quart., **20**(1973), 477-504.

[15] R.P. Hettich and H.T. Jongen, Semi-infinite Programming: Conditions of Optimality and Applications, in Optimization Techniques, ser. Lecture Notes in Control and Information Sciences, J. Stoer, Ed. Springer, Berlin, 1978.

[16] R.P. Hettich and K.O. Kortanek, *Semi-infinite programming: theory, methods, and applications*, SIAM Rev, **35**(1993), 380-429.

[17] H. Hu, *A one-phase algorithm for semi-infinite linear programming*, Math. Program., **46**(1990), 85-103.

[18] H.T. Jongen, P. Jonker and F. Twilt, *Critical sets in parametric optimization*, Math. Program., **34**(1986), 333-353.

[19] A. Jourani, *Constraint qualifications and Lagrange multipliers in nondifferentiable programming problems*, J. Optim. Theory Appl., **81**(1994), 533-548.

[20] K. Kortanek and H. No, *A central cutting plane algorithm for convex semi-infinite programming problems*, SIAM J. Optim., **3**(1993), 901-918.

[21] H.C. Lai and S.-Y. Wu, *On linear semi-infinite programming problems: an algorithm*, Numer. Func. Anal. Optim., **13**(1992), 287-304.

[22] C. Ling, Q. Ni, L. Qi and S.-Y. Wu, *A new smoothing Newton-type algorithm for semi-infinite programming*, J. Global Optim., **47**(2010), 133-159.

[23] D.H. Li, L. Qi, J. Tam and S.-Y. Wu, *A smoothing Newton method for semi-infinite programming*, J. Global Optim., **30**(2004), 169-194.

[24] G.H. Lin, M. Xu and J.J. Ye, *On solving simple bilevel programs with a nonconvex lower level program*, Math. Program., series A, DOI 10.1007/s10107-013-0633-4.

[25] M. López and G. Still, *Semi-infinite programming*, Eur. J. Opera. Res., **180**(2007), 491-518.

[26] Q. Ni, C. Ling, L. Qi and K.L. Teo, *A truncated projected Newton-type algorithm for large-scale semi-infinite programming*, SIAM J. Optim., **16**(2006), pp. 1137-1154.

[27] E. Polak, Optimization: Algorithms and Consistent Approximations, Springer, Berlin, 2006.

[28] E. Polak, *On the mathematical foundations of nondifferentiable optimization in engineering design*, SIAM Rev., **29**(1987), pp. 21-89.

[29] L. Qi, C. Ling, X.J. Tong and G. Zhou, *A smoothing projected Newton-type algorithm for semi-infinite programming*, Comput. Optim. Appl., **42**(2009), 1-30.

[30] L. Qi, S.-Y. Wu and G. Zhou, *Semismooth Newton methods for solving semi-infinite programming problems*, J. Global Optim., **27**(2003), 215-232.

[31] R. Reemtsen and S. Görner, Numerical methods for semi-infinite programming: a survey, in Semi-Infinite Programming, R. Reemtsen and J.-J. Rückmann, eds., Kluwer, Boston, 1998, 195-275.

[32] K. Roleff, A stable multiple exchange algorithm for linear SIP, R. Hettich, Ed. Springer, Berlin, 1979.

[33] A. Shapiro, *Semi-infinite programming, duality, discretization and optimality conditions*, Optim., **58**(2009), 133-161.

[34] T.-J Shiu and S.-Y. Wu, *Relaxed cutting plane method with convexification for solving nonlinear semi-infinite programming problems*, Comput. Optim. Appl., **53**(2012), 91-113.

[35] O. Stein, Bilevel-Strategies on Semi-infinite Programming, Kluwer, Boston, 2003.

[36] O. Stein, *How to solve a semi-infinite optimization problem*, Eur. J. Oper. Res., **223**(2012), 312-320.

[37] O. Stein and G. Still, *Solving semi-infinite optimization problems with interior point techniques*, SIAM J. Control Optim., **42**(2003), 769-788.

[38] G. Still, *Discretization in semi-infinite programming: the rate of convergence*, Math. Program., **91**(2001), 53-69.

[39] Y. Tanaka, M. Fukushima and T. Ibaraki, *A globally convergent SQP method for semi-infinite nonlinear optimization*, J. Comput. Appl. Math, **23**(1988), 141-153.

[40] K.L. Teo, X.Q. Yang and L.S. Jennings, *Computational discretization algorithms for functional inequality constrained optimization*, Ann. Oper. Res, **28**(2000), 215-234.

[41] S.-Y. Wu, and S.C. Fang, *Solving convex programs with infinitely many linear constraints by a relaxed cutting plane method*, Comput. Math. Appl., **38**(1999), 23-33.

[42] S.-Y. Wu, S.C. Fang and C.J. Lin, *Relaxed cutting plane method for solving linear semi-infinite programming problems*, J. Optim. Theory Appl., **99**(1998), 759-779.

[43] S.-Y. Wu, D.H. Li, L. Qi and G. Zhou, *An iterative method for solving KKT system of the semi-infinite programming*, Optim. Methods Soft., **20**(2005), 629-643.

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

[45] L.P. Zhang, S.-Y. Wu and M.A. López, *A new exchange method for convex semi-infinite programming*, SIAM J. Optim., **20**(2010), 2959-2977.