A Smoothing Projected Newton-Type Algorithm for Semi-Infinite Programming

Liqun Qi,[*]   Chen Ling,[†]   Xiaojiao Tong[‡]   and   Guanglu Zhou[§]

August 10, 2006

Abstract

This paper presents a smoothing projected Newton-type method for solving the semi-infinite programming (SIP) problem. We first reformulate the KKT system of the SIP problem into a system of constrained nonsmooth equations. Then we solve this system by a smoothing projected Newton-type algorithm. At each iteration only a system of linear equations needs to be solved. The feasibility is ensured via the aggregated constraint under some conditions. Global and local superlinear convergence of this method is established under some standard assumptions. Preliminary numerical results are reported.

Key Words: Semi-infinite programming, KKT system, constrained equations, smoothing method, convergence.

1 Introduction

We consider the following semi-infinite programming (SIP) problem:

$$\min\{f(x): \ x \in X\}, \tag{1.1}$$

where $X = \{x \in R^n: \ g(x,v) \leq 0, \ \forall \ v \in V\}$, $f: R^n \to R$ and $g: R^n \times R^m \to R$ are twice continuously differentiable functions. In this paper, we assume that V is a nonempty compact box with

$$V = \{v \in R^m: \ a \leq v \leq b\},$$

---

[*]Department of Mathematics, City University of Hong Kong, Kowloon Tong, Kowloon, Hong Kong. E-mail: *maqilq@polyu.edu.hk*. His work is supported by the Hong Kong Research Grant Council.

[†]School of Mathematics and Statistics, Zhejiang University of Finance and Economics, Hangzhou, 310018, China. E-mail: *linghz@hzcnc.com*.

[‡]Institute of Mathematics, Changsha University of Science and Technology, Changsha, China. This work was done during her visit to The Hong Kong Polytechnic University. e-mail: *tongxj@cscu.edu.cn*.

[§]Department of Mathematics and Statistics, Curtin University of Technology, WA 6102, Australia. E-mail: *g.zhou@exchange.curtin.edu.au*. His work is supported by Australian Research Council.

where $a \in R^m$, $b \in R^m$, and $a < b$. Here the inequality $a < b$ means that $a_i < b_i$ for all $i = 1, 2, ..., m$.

The SIP problem arises from various applications such as approximation theory, optimal control, eigenvalue computation, mechanical stress of materials, and statistical design. The main difficulty for solving the SIP problem is that it has infinite constraints. Many methods have been proposed for the SIP problem. We refer readers to [8, 9, 15, 16, 23] for details.

Let

$$V(x) = \{v \in V : g(x, v) = 0\}.$$

It is well-known [22] that if x is a local minimizer of the SIP problem (1.1), and if the extended Mangasarian-Fromovitz constraint qualification (EMFCQ) holds at x, i.e., there exists a vector $d \in R^n$ such that

$$\nabla_x g(x, v)^T d < 0$$

for all $v \in V(x)$, then there are p positive numbers u_i such that

$$\nabla f(x) + \sum_{i=1}^{p} u_i \nabla_x g(x, v^i) = 0,$$

where $v^i \in V(x)$ for $i = 1, \cdots, p$ and $p \leq n$. Hence, the KKT system of the SIP problem (1.1) is as follows:

$$\begin{aligned}
&\nabla f(x) + \sum_{i=1}^{p} u_i \nabla_x g(x, v^i) = 0, \\
&g(x, v) \leq 0, \quad \forall \, v \in V, \\
&u_i > 0, \quad g(x, v^i) = 0, \ i = 1, \cdots, p.
\end{aligned} \tag{1.2}$$

In the KKT system (1.2), x is called a **stationary point** of the SIP problem, and $u \equiv (u_1, \cdots, u_p) \in R^p$ and v^i for $i = 1, \cdots, p$ are called its **Lagrange multiplier** and **attainers**, respectively.

The KKT system (1.2) can be analyzed further. By the definition of $V(x)$ and the second constrained condition of (1.2), $v^i \in V(x)$ $(i = 1, \cdots, p)$ imply that v^i $(i = 1, \cdots, p)$ are global minimizers of the following minimization problem:

$$\begin{aligned}
&\min \quad -g(x, v) \\
&s.t. \quad v \in V.
\end{aligned} \tag{1.3}$$

The KKT system of (1.3) can be written as

$$(v' - v)^T (-\nabla_v g(x, v)) \geq 0, \quad \forall \, v' \in V,$$

and it can be reformulated as a system of nonsmooth equations as follows (see [2, 5] for details):

$$\phi(x, v) = 0. \tag{1.4}$$

Here, $\phi(x, v)$ is defined as

$$\phi(x, v) := v - P(a, b, v + \nabla_v g(x, v)), \tag{1.5}$$

where the function P is the mid-function defined for all $j = 1, \cdots, m$, as

$$(P(a, b, w))_j = \begin{cases} a_j, & \text{if } w_j < a_j, \\ w_j, & \text{if } a_j \leq w_j \leq b_j, \\ b_j, & \text{if } b_j < w_j. \end{cases}$$

Thus, if a constraint qualification for (1.3) holds, then the KKT system of the SIP problem (1.1) can be reformulated as follows:

$$\begin{aligned} &\nabla f(x) + \sum_{i=1}^{p} u_i \nabla_x g(x, v^i) = 0, \\ &g(x, v) \leq 0, \quad \forall\, v \in V, \\ &u_i > 0, \quad g(x, v^i) = 0, \quad (i = 1, \cdots, p) \\ &\phi(x, v^i) = 0 \quad (i = 1, \cdots, p), \end{aligned} \tag{1.6}$$

because the system (1.4) is a first order necessary condition for v^i, $i = 1, 2, \cdots, p$ to be local solutions of (1.3).

Based on (1.6), a semismooth Newton method and a smoothing Newton method were presented in [21] and [10] respectively. The advantage of these two methods proposed in [21, 10] is that in every iteration only a system of linear equations needs to be solved. Moreover, these methods enjoy global and locally superlinear convergence. However, these two methods cannot ensure the feasibility of (1.1), because some previous information was replaced by (1.4), and the second constrained condition in (1.6) was removed. Quite recently, another iterative method for solving the KKT system of (1.1) was proposed in [29], in which the feasibility issue was considered. However, the method in [29] does not have locally superlinear convergence property.

Let

$$G(x) = \int_V [g(x, v)]_+ dv, \tag{1.7}$$

where $[x]_+ = \max\{0, x\}$. The function $G(x)$ was proposed in [25]. Then (1.6) is equivalent to

$$\begin{aligned} &\nabla f(x) + \sum_{i=1}^{p} u_i \nabla_x g(x, v^i) = 0, \\ &G(x) \leq 0, \\ &u_i > 0, \quad g(x, v^i) = 0, \quad (i = 1, \cdots, p) \\ &\phi(x, v^i) = 0 \quad (i = 1, \cdots, p). \end{aligned} \tag{1.8}$$

It is not difficult to show that $G(x)$ is nonsmooth but semismooth [18].

In this paper, we present a new method for solving the SIP problem by using a smoothing projected Newton-type algorithm. At each iteration only a system of linear equations needs to

3

be solved. The feasibility is ensured via the aggregated constraint $G(x) \leq 0$. Global and local superlinear convergence of this method is established under some standard assumptions. Some drawbacks of existing methods are overcome.

The rest of this paper is organized as follows. In Section 2, we study the properties of the smoothings of the functions $G(x)$ and $\phi(x, v)$ in (1.8). In Section 3, a smoothing projected Newton-type algorithm is presented to solve (1.8). This smoothing algorithm is an extension of the method for solving unconstrained nonsmooth equations presented in [19]. It can also be viewed as a smoothing version of the method proposed in [24]. In Section 4 we establish the global and local superlinear convergence of the new method. In Section 5, we give our numerical results, which show that our new method performs well. Some comments are made in the last section.

Some words about the notation. For a smooth (continuously differentiable) function $\Phi : R^n \to R^m$, we denote the Jacobian of Φ at $x \in R^n$ by $\Phi'(x)$, which is an $m \times n$ matrix. We denote the transposed Jacobian as $\nabla \Phi(x)$. For a function $f : R^n \times R^m \to R$ we denote $\nabla_x f(x, y)$ the gradient of f at (x, y) with respect to x and $\nabla^2_{xx} f(x, y)$, $\nabla^2_{xy} f(x, y)$ and $\nabla^2_{yy} f(x, y)$ denote, respectively, the $n \times n$, $n \times m$ and $m \times m$ Hessian matrices of f at (x, y). For a nonsmooth function $G(x)$, $\partial G(x)$ means the generalized Jacobian in the sense of Clarke [3]. If $M \in R^{t \times t}$, $M = (m_{ij})$, is any given matrix and $I, J \subseteq \{1, 2, \cdots, t\}$ are two subsets, then M_{IJ} stands for the $|I| \times |J|$ submatrix with elements m_{ij}, $i \in I$, $j \in J$. If $I = \{1, 2, \cdots, t\}$ or $J = \{1, 2, \cdots, t\}$, then M_{IJ} is simplified as $M_{\cdot J}$ or $M_{I \cdot}$, respectively. $\| \cdot \|$ denotes the Euclidean norm. If δ is a small quantity, $O(\delta)$ and $o(\delta)$ mean the same order and higher order small quantity respectively. $\Pi_W(\cdot)$ represents the orthogonal projection on a set W.

2   Some Preliminaries

In this section, we review some results on semismoothness and give some properties on smoothing approximation functions of the integral function $G(x)$ defined in (1.7) and $\phi(x, v)$ defined in (1.5). These results and properties will be used later.

Semismoothness was originally introduced by Mifflin [13] for functionals. In [20], Qi and Sun extended the definition of semismooth functions to $H : R^n \to R^n$. H is said to be semismooth at $x \in R^n$, if

$$\lim_{\substack{Q \in \partial H(x + th') \\ h' \to h, t \downarrow 0}} \{Qh'\}$$

exists for any $h \in R^n$. Semismoothness can also be defined equivalently as follows [12]:

Definition 2.1 *Let $H : R^n \to R^n$ be a locally Lipschitz function. We say that H is semismooth at x if*

(i) H is directionally differentiable at x; and

4

(ii) for any $h \to 0$ and $Q \in \partial H(x+h)$,

$$H(x+h) - H(x) - Qh = o(||h||).$$

H is said to be strongly semismooth at x if H is semismooth at x and for any $Q \in \partial H(x+h)$ and $h \to 0$,

$$H(x+h) - H(x) - Qh = O(||h||^2).$$

Here, $o(||h||)$ stands for a vector function of h, satisfying

$$\lim_{h \to 0} \frac{o(||h||)}{||h||} = 0,$$

while $O(||h||^2)$ stands for a vector function of h, satisfying

$$||O(||h||^2)|| \le M||h||^2$$

for all h satisfying $||h|| \le \delta$, and some $M > 0$ and $\delta > 0$. A function H is said to be a (strongly) semismooth function if it is (strongly) semismooth everywhere on R^n.

In [17], Qi defined the B-subdifferential of a locally Lipschitz function $H : R^n \to R^n$ at a point $x \in R^n$:

$$\partial_B H(x) = \left\{ Q \in R^{n \times n} \mid Q = \lim_{x^k \to x} H'(x^k), H \text{ is differentiable at } x^k \text{ for all } k \right\}.$$

It is seen that

$$\partial H(x) = \text{conv } \partial_B H(x).$$

The concept of the B-subdifferential will be used in the design of our algorithm. A locally Lipschitz function H is said to be BD-regular at $x \in R^n$ if all $Q \in \partial_B H(x)$ are nonsingular [17].

Lemma 2.1 *[17] Suppose that $H : R^n \to R^n$ is locally Lipschitz continuous and H is BD-regular at $x \in R^n$. Then there exist a neighborhood $N(x)$ of x and a constant C such that for any $y \in N(x)$ and $Q \in \partial_B H(y)$, Q is nonsingular and $||Q^{-1}|| \le C$.*

Lemma 2.2 *[14] Suppose that $H : R^n \to R^n$ is locally Lipschitz continuous and H is BD-regular at a solution x^* of $H(x) = 0$. If H is semismooth at x^*, then there exist a neighborhood $N(x^*)$ of x^* and a constant C such that for any $x \in N(x^*)$,*

$$||H(x)|| \ge C||x - x^*||.$$

Define $\bar{G} : R \times R^n \to R$ by

$$\bar{G}(t,x) = \int_V \bar{g}(t,x,v)dv,$$

where $\bar{g} : R \times R^n \times R^m \to R$ is defined by

$$\bar{g}(t, x, v) = \frac{\sqrt{(g(x, v))^2 + 4t^2} + g(x, v)}{2}. \tag{2.1}$$

The function $\bar{g}$ is the Chen-Harker-Kanzow-Smale smoothing function of $[g(x, v)]_+$. Other smoothing functions of $[g(x, v)]_+$ can be found in [19]. It is obvious that for any $t \neq 0$, $\bar{G}(t, x)$ is smooth with respect to variable x and

$$\nabla_x \bar{G}(t, x) = \int_V \nabla_x \bar{g}(t, x, v) dv. \tag{2.2}$$

We now study the semismoothness of $\bar{G}(t, x)$. To this end, We first consider a general case. Let $f(x, v) : R^n \times V \to R$ be continuous with respect to $v \in V$ for each fixed $x \in R^n$, and be locally Lipschitz with respect to x uniformly in $v \in V$, i.e., there exist a neighborhood N of 0 and a positive constant C such that

$$\|f(x + h, v) - f(x, v)\| \leq C \|h\|, \quad \forall\, h \in N, v \in V.$$

Let

$$\Gamma(x) = \int_V f(x, v) dv.$$

Obviously, $\Gamma : R^n \to R$ is locally Lipschitz continuous.

Proposition 2.1 *Suppose that $\partial_x f(x, v)$, viewed as a joint mapping of x and v, is upper semicontinuous, i.e., for every neighborhood N of $\partial_x f(x, v)$, there exists $\delta > 0$ such that*

$$\partial_x f(x', v') \subset N, \text{ for all } x' \in N_1(x, \delta), v' \in N_2(v, \delta),$$

where

$$N_1(x, \delta) = \{x' : \|x' - x\| \leq \delta\}$$

and

$$N_2(v, \delta) = \{v' : \|v' - v\| \leq \delta\} \cap V.$$

Then Γ is semismooth at $\bar{x}$ if $f(\cdot, v)$ is semismooth at $\bar{x}$ for every $v \in V$.

Proof. It follows from Proposition 1 in [18] that Γ is directionally differentiable at $\bar{x}$. On the other hand, by Theorem 2.7.2 in [3], we obtain

$$\partial \Gamma(x) \subset \int_V \partial_x f(x, v) dv. \tag{2.3}$$

This means that for any $Q \in \partial \Gamma(x)$, there exists a measurable mapping $v \to Q_v$ from V to R^n with $Q_v \in \partial_x f(x, v)$ a.e. such that for every $h \in R^n$,

$$Qh = \int_V Q_v h\, dv.$$

6

Take any $h \in R^n$ and $Q \in \partial\Gamma(\bar{x} + h)$. We have

$$\Gamma(\bar{x} + h) - \Gamma(\bar{x}) - Q\ h = \int_V (f(\bar{x} + h, v) - f(\bar{x}, v) - Q_v h)\, dv,$$

where $Q_v \in \partial_x f(\bar{x} + h, v)$, which implies

$$|\Gamma(\bar{x} + h) - \Gamma(\bar{x}) - Q\ h| \leq \int_V |f(\bar{x} + h, v) - f(\bar{x}, v) - Q_v h|\, dv. \tag{2.4}$$

To prove Γ is semismooth, it suffices to show that

$$\lim_{h \to 0} \frac{|\Gamma(\bar{x} + h) - \Gamma(\bar{x}) - Q\ h|}{\|h\|} = 0. \tag{2.5}$$

Since $f(\cdot, v)$ is semismooth at $\bar{x}$ for every fixed $v \in V$, we have

$$\lim_{h \to 0} \frac{|f(\bar{x} + h, v) - f(\bar{x}, v) - Q_v h|}{\|h\|} = 0, \text{ for all } Q_v \in \partial_x f(\bar{x} + h, v). \tag{2.6}$$

If there exist a neighborhood N of 0 and $C > 0$ such that

$$\frac{|f(\bar{x} + h, v) - f(\bar{x}, v) - Q_v h|}{\|h\|} \leq C, \tag{2.7}$$

for all $h \in N$, $Q_v \in \partial_x f(\bar{x} + h, v)$ and $v \in V$, then by the dominated convergence theorem, (2.5) follows from (2.6).

Now we prove (2.7). Since f is locally Lipschitz continuous at $\bar{x}$ uniformly in $v \in V$, there exist a neighborhood N of 0 and $C(\bar{x}) > 0$ such that

$$\frac{|f(\bar{x} + h, v) - f(\bar{x}, v)|}{\|h\|} \leq C(\bar{x}), \ \forall\ h \in N,\ v \in V. \tag{2.8}$$

On the other hand, the upper semicontinuity $\partial_x f(x, v)$ implies that for any $v \in V$ and neighborhood $N(v)$ of $\partial_x f(\bar{x}, v)$, there exist $\delta_v > 0$ such that

$$\partial_x f(\bar{x} + h, v') \subset N(v), \text{ for all } h \in N_1(0, \delta_v), v' \in N_2(v, \delta_v).$$

Obviously,

$$V \subset \cup_{v \in V} N_2(v, \delta_v).$$

By the compactness of V, there exist a finite number of neighborhoods, say $N_2(v_j, \delta_{v_j})$, $j = 1, 2, \cdots, m$ such that

$$V \subset \cup_{j=1}^m N_2(v_j, \delta_{v_j}).$$

Let $\bar{\delta} = \min\{\delta_{v_1}, \cdots, \delta_{v_m}\}$. Then we have

$$\cup_{v' \in V} \partial_x f(\bar{x} + h, v') \subset \cup_{j=1}^m N(v_j), \text{ for all } h \in N_1(0, \bar{\delta}).$$

7

It is well known that every $\partial_x f(\bar{x}, v_j)$ is compact, $j = 1, 2, \cdots, m$. Consequently, $\cup_{j=1}^m \partial_x f(\bar{x}, v_j)$ is compact and $\cup_{j=1}^m N(v_j)$ can be taken a bounded set. Hence, $\cup_{v' \in V} \partial_x f(\bar{x} + h, v')$ is bounded, which together with (2.8) implies (2.7) holds. We obtain the desired result and complete the proof of the theorem. ∎

Now we give some properties of the function $\bar{G}$.

Proposition 2.2 *The function $\bar{G}$ has the following properties:*

(i) It is twice continuously differentiable for any $t \neq 0$.

(ii) There exists a constant $C > 0$ such that for any $x \in R^n$

$$\left| \bar{G}(t, x) - G(x) \right| \leq C|t|.$$

(iii) The function $\bar{G}$ is semismooth.

Proof. It is obvious that (i) holds. The proof of (ii) is similar to that of Theorem 2.1 [11], so we omit it. Now we prove that (iii) holds.

By (i), we only need to show that (iii) holds on $\bar{z} = (0, \bar{x})$. Since the composition of semismooth functions is a semismooth function [6], $\bar{g}(t, x, v)$ is semismooth with respect to (t, x) for any fixed $v \in V$. To prove the semismoothness of $\bar{G}(t, x)$, by Proposition 2.1, we only need to show that $\partial_{(t,x)} \bar{g}(t, x, v)$ is upper semicontinuous with respect to (t, x, v) and $\bar{g}(t, x, v)$ is locally Lipschitz with respect to (t, x) uniformly in $v \in V$. By direct computation, we have

$$\partial_{(t,x)} \bar{g}(t, x, v) = \begin{cases} \left\{ \frac{1}{2} \left(\frac{4t}{\sqrt{g^2 + 4t^2}}, \left(1 + \frac{g}{\sqrt{g^2 + 4t^2}} \right) (\nabla_x g)^T \right) \right\}, & \text{if } t \neq 0 \\ \left\{ \left(0, (\nabla_x g)^T \right) \right\}, & \text{if } t = 0, \; g(x, v) > 0 \\ \left\{ (0, 0) \right\}, & \text{if } t = 0, \; g(x, v) < 0 \\ \left\{ \left(\lambda, \mu (\nabla_x g)^T \right) \mid \lambda^2 + (2\mu - 1)^2 \leq 1 \right\}, & \text{if } t = 0, \; g(x, v) = 0, \end{cases} \quad (2.9)$$

where g and $\nabla_x g$ express the corresponding values of functions at point (x, v). From (2.9), it is easy to verify $\partial_{(t,x)} \bar{g}(t, x, v)$ is upper semicontinuous with respect to (t, x, v) on $R \times R^n \times V$.

Now we verify that $\bar{g}(t, x, v)$ is locally Lipschitz with respect to (t, x) uniformly in $v \in V$. Let $z = (t, x)$. For z close to $\bar{z}$. There are two cases: (a) $t \neq 0$ and (b) $t = 0$.

Case (a). By the Mean-Value theorem, there exists a point $\tilde{z}$ in the open segment connecting z and $\bar{z}$ such that

$$\bar{g}(z, v) - \bar{g}(\bar{z}, v) = \nabla_z \bar{g}(\tilde{z}, v)^T (z - \bar{z}).$$

By (2.9), it is easy to know that there exists $C > 0$ such that

$$|\bar{g}(z, v) - \bar{g}(\bar{z}, v)| \leq C \left\| z - \bar{z} \right\|, \; \forall \, v \in V, \quad (2.10)$$

since g is continuously differentiable and V is compact.

8

Case (b). We have

$$
\begin{aligned}
|\bar{g}(z, v) - \bar{g}(\bar{z}, v)| &= |[g(x, v)]_+ - [g(\bar{x}, v)]_+| \\
&\leq 2|g(x, v) - g(\bar{x}, v)| \\
&\leq \|\nabla_x \bar{g}(\tilde{x}, v)\| \, \|x - \bar{x}\| \\
&= \|\nabla_x \bar{g}(\tilde{x}, v)\| \, \|z - \bar{z}\|,
\end{aligned}
$$

where $\tilde{x}$ is in the open segment connecting x and $\bar{x}$, the first inequality comes from the fact that $|[a]_+ - [b]_+| \leq 2|a - b|$. By this and the condition that g is continuously differentiable and V is compact, there exists $C > 0$ such that (2.10) holds. The proof is complete. ∎

Define $\varphi : R^4 \to R$ by

$$
\varphi(t, c, d, w) = \frac{c + \sqrt{(c - w)^2 + 4t^2}}{2} + \frac{d - \sqrt{(d - w)^2 + 4t^2}}{2},
$$

which is the Chen-Harker-Kanzow-Smale smoothing function for $mid(c, d, w)$. For $a, b, v \in R^m$, we define $\bar{\phi} : R \times R^m \times R^n \to R^m$ by

$$
(\bar{\phi}(t, x, v))_i = v_i - \varphi(t, a_i, b_i, v_i + (\nabla_v g(x, v))_i), \tag{2.11}
$$

where $i = 1, \cdots, m$. It is clear that $\bar{\phi}$ is smooth for $t \neq 0$.

From Theorem 3 in [19], Lemma 2.3 and Theorem 3.3 in [7], it is easy to prove the following results for $\bar{\phi}$.

Proposition 2.3 *The function $\bar{\phi}$ defined in (2.11) has the following properties:*
(i) It is twice continuously differentiable for $t \neq 0$.
(ii) It is semismooth. Furthermore, if g is twice Lipschitz continuously differentiable, it is strongly semismooth.
(iii) There exists a constant $C > 0$ such that for any $(x, v) \in R^{n+m}$ and $t \in R$,

$$
\left\| \bar{\phi}(t, x, v) - \phi(x, v) \right\| \leq C|t|.
$$

3 A Smoothing Projected Newton-Type Algorithm

Let

$$
\mathbf{v} = (v^1; v^2; ...; v^p).
$$

Define

$$
F(x, u, \mathbf{v}) = \nabla f(x) + \sum_{i=1}^{p} u_i \nabla_x g(x, v^i), \tag{3.1}
$$

$$
\mathbf{g}(x, \mathbf{v}) = \begin{pmatrix} g(x, v^1) \\ \vdots \\ g(x, v^p) \end{pmatrix}, \quad \hat{\phi}(x, \mathbf{v}) = \begin{pmatrix} \phi(x, v^1) \\ \vdots \\ \phi(x, v^p) \end{pmatrix}, \quad \tilde{\phi}(t, x, \mathbf{v}) = \begin{pmatrix} \bar{\phi}(t, x, v^1) \\ \vdots \\ \bar{\phi}(t, x, v^p) \end{pmatrix}.
$$

9

By introducing a slack variable $y \in R$ and relaxing $u_i > 0$ as $u_i \geq 0$, (1.8) can be written as the following system of nonsmooth equations with bounded constraints:

$$H(z) = 0,$$
$$u \geq 0, \quad y \geq 0,$$

(3.2)

where $z = (x, u, \mathbf{v}, y) \in R^n \times R^p \times R^{mp} \times R$, and

$$H(z) = \begin{pmatrix} F(x, u, \mathbf{v}) \\ \mathbf{g}(x, \mathbf{v}) \\ G(x) + y \\ \hat{\phi}(x, \mathbf{v}) \end{pmatrix}.$$

Clearly, if $z = (x, u, \mathbf{v}, y) \in R^n \times R^p \times R^{mp} \times R$ is a solution of (3.2), then we may get a solution of (1.8) by dropping those v^i in $\mathbf{v}$ satisfying $u_i = 0$.

Denote $w = (t, z) = (t, x, u, \mathbf{v}, y) \in R \times R^n \times R^p \times R^{mp} \times R$. Motivated by the smoothing method in [19] for a system of unconstrained nonsmooth equations and the method in [24] for a system of constrained nonsmooth equations, in this section we present a smoothing projected Newton-type method for solving (3.2). We define the following system of constrained equations:

$$\Phi(t, z) = 0$$
$$u \geq 0, y \geq 0,$$

(3.3)

where

$$\Phi(t, z) = \begin{pmatrix} t \\ \bar{H}(t, z) \end{pmatrix}, \quad \bar{H}(t, z) = \begin{pmatrix} F(x, u, \mathbf{v}) \\ \mathbf{g}(x, \mathbf{v}) \\ \bar{G}(t, x) + y \\ \tilde{\phi}(t, x, \mathbf{v}) \end{pmatrix}.$$

It is obvious that if (t, z) is a solution of (3.3) then z is a solution to (3.2). By Propositions 2.2 and 2.3, we have the following proposition.

Proposition 3.1 Φ *is smooth at* (t, z) *with* $t \neq 0$ *and semismooth at* $(0, z)$.

Let

$$W = \{w = (t, x, u, \mathbf{v}, y) : \ u \geq 0, y \geq 0\},$$

and

$$Z = \{(x, u, \mathbf{v}, y) \in R^n \times R^p \times R^{mp} \times R : \ u \geq 0, y \geq 0\}.$$

Define a merit function of (3.3) by

$$\Psi(w) = \frac{1}{2}\|\Phi(w)\|^2.$$

10

Then solving (3.3) is equivalent to finding a global solution of the following minimization problem:

$$\min \quad \Psi(w)$$
$$\text{s.t.} \quad u \geq 0, y \geq 0. \tag{3.4}$$

And w is a stationary point of (3.4) if it satisfies

$$\|\bar{d}_G(1)\| = 0. \tag{3.5}$$

Here,

$$\bar{d}_G(1) = \Pi_W(w - \gamma \nabla \Psi(w)) - w = \begin{pmatrix} -\gamma \nabla_t \Psi(w) \\ \Pi_Z(z - \gamma \nabla_z \Psi(w)) - z \end{pmatrix}, \tag{3.6}$$

where $\gamma > 0$ is a constant, $\Pi_W(\cdot)$ is an orthogonal projection operator onto W.

Let $\alpha \in (0,1)$ be a constant. For a sequence $\{w^k\}_{k=0}^{\infty}$, we define

$$\beta_0 = \beta(w^0) = \alpha \min\{1, \|\bar{d}_G^0(1)\|^2\},$$

and

$$\beta_k = \beta(w^k) := \begin{cases} \beta_{k-1}, & \text{if } \alpha \min\{1, \|\bar{d}_G^k(1)\|^2\} > \beta_{k-1} \\ \alpha \min\{1, \|\bar{d}_G^k(1)\|^2\}, & \text{otherwise.} \end{cases} \tag{3.7}$$

Now we state our smoothing projected Newton-type algorithm for solving (3.4).

Algorithm 3.1

Step 0. (Initialization)

Choose constants $\eta, \rho, \sigma \in (0,1)$, $p_1 > 0$, $p_2 > 2$ and $\alpha > 0, \bar{t} > 0$ with $\alpha \bar{t} < 1$. Let $\bar{w} = (\bar{t}, 0, 0, 0, 0)$, $t_0 = \bar{t}$ and $w^0 = (t_0, x^0, u^0, \mathbf{v}^0, y^0)$ with $u_i^0 \geq 0$ $(i = 1, \cdots, p)$; $y^0 \geq 0$. Set $k := 0$.

Step 1. (Stoping Test)

Let

$$\gamma_k = \min\left\{1, \frac{t_k}{|t_k + \nabla_t \bar{H}(w^k) \ \bar{H}(w^k)|}, \frac{\eta \|\Phi(w^k)\|}{\|\nabla \Psi(w^k)\|}, \frac{\eta \Psi(w^k)}{\|\nabla \Psi(w^k)\|^2}\right\}, \tag{3.8}$$

where $\nabla_t \bar{H}(w^k)$ is the first row of $\nabla \bar{H}(w^k)$. Compute $\bar{d}_G^k(1)$ by (3.6). If $\|\bar{d}_G^k(1)\| = 0$, stop. Otherwise, compute β_k by (3.7).

Step 2. (Compute Search Direction)

Compute d_G^k by

$$d_G^k = -\gamma_k \nabla \Psi(w^k) + \beta_k \bar{w}. \tag{3.9}$$

Compute d_N^k by solving the following linear system:

$$\Phi(w^k) + \Phi'(w^k)d_N^k = \beta_k \bar{w}. \tag{3.10}$$

11

If (3.10) has no solution or

$$-\nabla\Psi(w^k)^T d_N^k < p_1 \left\| d_N^k \right\|^{p_2},$$

then let $d_N^k := d_G^k$.

Step 3. (Line Search)

Let m_k be the smallest nonnegative integer m satisfying

$$\Psi(w^k + \bar{d}^k((\rho)^m)) \leq \Psi(w^k) + \sigma\nabla\Psi(w^k)^T \tilde{d}_G^k((\rho)^m), \tag{3.11}$$

where for any $\lambda \in [0,1]$,

$$\bar{d}^k(\lambda) = \tau^*(\lambda)\tilde{d}_G^k(\lambda) + (1 - \tau^*(\lambda))\tilde{d}_N^k(\lambda). \tag{3.12}$$

Here

$$\tilde{d}_G^k(\lambda) := \Pi_W(w^k + \lambda d_G^k) - w^k, \quad \tilde{d}_N^k(\lambda) := \Pi_W(w^k + \lambda d_N^k) - w^k, \tag{3.13}$$

$\tau^*(\lambda)$ is a solution of the following minimization problem:

$$\min_{\tau \in [0,1]} \frac{1}{2}\|\Phi(w^k) + \Phi'(w^k)[\tau\tilde{d}_G^k(\lambda) + (1 - \tau)\tilde{d}_N^k(\lambda)]\|^2.$$

Let $\lambda_k = (\rho)^{m_k}$ and $w^{k+1} = w^k + \bar{d}^k(\lambda_k)$.

Step 4. Set $k := k + 1$ and go to Step 1.

Remark. (a) Algorithm 3.1 is a smoothing version of the algorithm proposed in [24]. In [24], it is required that the merit function Ψ must be smooth. In this paper, we do not need this requirement.

(b) Similar to Lemma 3.1 [24], we have the following result about $\tau^*(\lambda)$.

$$\tau^*(\lambda) = \max\{0, \min\{1, \tau(\lambda)\}\}, \tag{3.14}$$

where $\tau(\lambda)$ is defined as

$$\tau(\lambda) = \begin{cases} 0, & \text{if } \Phi'(w^k)[\tilde{d}_G^k(\lambda) - \tilde{d}_N^k(\lambda)] = 0, \\ -\dfrac{[\Phi(w^k) + \Phi'(w^k)\tilde{d}_N^k(\lambda)]^T\Phi'(w^k)[\tilde{d}_G^k(\lambda) - \tilde{d}_N^k(\lambda)]}{\|\Phi'(w^k)[\tilde{d}_G^k(\lambda) - \tilde{d}_N^k(\lambda)]\|^2}, & \text{otherwise.} \end{cases}$$

The following projection properties are used in our analysis (see [1]).

Lemma 3.1 *The projection operator $\Pi_W(\cdot)$ with any convex set $W \subset R^n$ satisfies*
(i) for any $w \in W$,

$$[\Pi_W(w') - w']^T[\Pi_W(w') - w] \leq 0 \quad \text{for all } w' \in R^n;$$

(ii)

$$\|\Pi_W(w') - \Pi_W(w'')\| \leq \|w' - w''\| \quad \text{for all } w', w'' \in R^n;$$

(iii) Given $w, d \in R^n$, the function ζ defined by

$$\zeta(\lambda) = \|\Pi_W(w + \lambda d) - w\|/\lambda, \quad \lambda > 0$$

is non-increasing.

From the definition of β_k, the following proposition is obvious.

Proposition 3.2 *$\{\beta_k\}$ defined in (3.7) has the following properties:*
(i) $\{\beta_k\}$ is non-increasing sequence;
(ii) For all k, β_k satisfies

$$\beta_k \leq \alpha \min\{1, \|\bar{d}_G^k(1)\|^2\}.$$

Proposition 3.3 *Suppose that $w^k = (t^k, z^k) \in W$ with $t^k > 0$ is not a stationary point of (3.4). Then for any $\lambda \in (0, 1]$, it holds that*

$$\nabla\Psi(w^k)^T \tilde{d}_G^k(\lambda) \leq -\frac{\lambda}{\gamma_k}(1 - \alpha\bar{t})\|\bar{d}_G^k(1)\|^2 < 0. \tag{3.15}$$

Proof. In this proof, for simplicity, we drop the superscript k. For any $w = (t, z) \in W$ with $t > 0$, suppose that w is not a stationary point of (3.4). Then

$$\nabla\Psi(w) = \nabla\Phi(w)\Phi(w) = \begin{pmatrix} t + \nabla_t\bar{H}(w)\,\bar{H}(w) \\ \nabla_z\bar{H}(w)\,\bar{H}(w) \end{pmatrix} \equiv \begin{pmatrix} \nabla_t\Psi(w) \\ \nabla_z\Psi(w) \end{pmatrix},$$

where $\nabla_t\bar{H}(w)$ is the first row of $\nabla\bar{H}(w)$ and $\nabla_z\bar{H}(w)$ is the submatrix of $\nabla\bar{H}(w)$ obtained by just removing the first row of $\nabla\bar{H}(w)$. Obviously, $\tilde{d}_G(\lambda)$ can be written as

$$\tilde{d}_G(\lambda) \equiv \begin{pmatrix} (\tilde{d}_G(\lambda))_t \\ (\tilde{d}_G(\lambda))_z \end{pmatrix} = \begin{pmatrix} -\lambda\gamma(t + \nabla_t\bar{H}(w)\bar{H}(w)) + \lambda\beta(w)\bar{t} \\ \Pi_Z(z - \lambda\gamma\nabla_z\Psi(w)) - z \end{pmatrix}.$$

Then we have

$$\begin{aligned}
&(t + \nabla_t\bar{H}(w)\bar{H}(w))^T[-\lambda\gamma(t + \nabla_t\bar{H}(w)\bar{H}(w)) + \lambda\beta(w)\bar{t}\,] \\
&= -\lambda\gamma\left\|t + \nabla_t\bar{H}(w)\bar{H}(w)\right\|^2 + \lambda(t + \nabla_t\bar{H}(w)\bar{H}(w))^T\beta(w)\bar{t} \\
&\leq -\frac{\lambda}{\gamma}\|-\gamma\nabla_t\Psi(w)\|^2 + \frac{\lambda}{\gamma}\|-\gamma\nabla_t\Psi(w)\|\beta(w)\bar{t} \\
&\leq -\frac{\lambda}{\gamma}\|-\gamma\nabla_t\Psi(w)\|^2 + \frac{\lambda}{\gamma}\|-\gamma\nabla_t\Psi(w)\|(\alpha\bar{t}\,)\|\bar{d}_G(1)\| \\
&\leq -\frac{\lambda}{\gamma}\|-\gamma\nabla_t\Psi(w)\|^2 + \alpha\bar{t}\frac{\lambda}{\gamma}\|\bar{d}_G(1)\|^2,
\end{aligned} \tag{3.16}$$

13

where the second inequality comes from Proposition (3.2) (ii) and the fact that $\beta(w) \leq \alpha\|\bar{d}_G(1)\|$, the last inequality is due to $\| - \gamma\nabla_t\Psi(w)\| \leq \|\bar{d}_G(1)\|$ (see (3.6)). Thus,

$$
\begin{aligned}
&\nabla_z\Psi(w)^T[\Pi_Z(z - \lambda\gamma\nabla_z\Psi(w)) - z]\\
&= -\frac{1}{\lambda\gamma}[z - \lambda\gamma\nabla_z\Psi(w) - z]^T[\Pi_Z(z - \lambda\gamma\nabla_z\Psi(w)) - z]\\
&= \frac{1}{\lambda\gamma}[\Pi_Z(z - \lambda\gamma\nabla_z\Psi(w)) - (z - \lambda\gamma\nabla_z\Psi(w))]^T[\Pi_Z(z - \lambda\gamma\nabla_z\Psi(w)) - z]\\
&\quad - \frac{1}{\lambda\gamma}\|\Pi_Z(z - \lambda\gamma\nabla_z\Psi(w)) - z\|^2\\
&\leq -\frac{1}{\lambda\gamma}\|\Pi_Z(z - \lambda\gamma\nabla_z\Psi(w)) - z\|^2\\
&\leq -\frac{\lambda}{\gamma}\|\Pi_Z(z - \gamma\nabla_z\Psi(w)) - z\|^2, &(3.17)
\end{aligned}
$$

where the first and second inequalities come from Lemma 3.1 (i) and (iii), respectively. It follows from (3.16) and (3.17) that

$$
\begin{aligned}
&\nabla\Psi(w)^T\tilde{d}_G(\lambda) &(3.18)\\
&= (t + \nabla_t\bar{H}(w)\bar{H}(w))^T[-\lambda\gamma(t + \nabla_t\bar{H}(w)\bar{H}(w)) + \lambda\beta(w)\bar{t}] + \nabla_z\Psi(w)^T[\Pi_Z(z - \lambda\gamma\nabla_z\Psi(w)) - z]\\
&\leq -\frac{\lambda}{\gamma}\left[\| - \gamma\nabla_t\Psi(w)\|^2 + \|\Pi_z(z - \gamma\nabla_z\Psi(w)) - z\|^2\right] + \alpha\bar{t}\frac{\lambda}{\gamma}\|\bar{d}_G(1)\|^2\\
&= -\frac{\lambda}{\gamma}(1 - \alpha\bar{t})\|\bar{d}_G(1)\|^2 < 0. &(3.19)
\end{aligned}
$$

The proof is complete. ∎

Now we have the following conclusion which shows that Algorithm 3.1 is well-defined.

Theorem 3.1 *Suppose that $w^k = (t^k, z^k) \in W$ with $t^k > 0$ is not a stationary point of (3.4). Then there exists a constant $\lambda' \in (0, 1]$ such that for any $\lambda \in (0, \lambda']$, $\bar{d}^k(\lambda)$ is a descent direction of $\Psi(w^k)$ at w^k and*

$$
\Psi(w^k + \bar{d}^k(\lambda)) \leq \Psi(w^k) + \sigma\nabla\Psi(w^k)^T\tilde{d}_G^k(\lambda). \tag{3.20}
$$

Proof. By using Proposition 3.3, the conclusion can be proved in a similar way to the proof of Theorem 3.1 in [24] so we omit it. ∎

4 Convergence Analysis

In this section we analyze the global and local convergence of Algorithm 3.1 in the previous section. The following proposition is a key result which shows that Algorithm 3.1 can keep $t^k > 0$ at each iteration.

Proposition 4.1 *For each k, $k = 0, 1, \cdots$, $w^k = (t^k, z^k)$ satisfies*

$$t^k \geq \beta_k \bar{t}. \tag{4.1}$$

Furthermore, if w^k is not a stationary point of (3.4), then

$$t^k > 0. \tag{4.2}$$

Proof. We prove this proposition by induction. From the choices of t^0 and β_0 in Algorithm 3.1, it is obvious that (4.1) holds. Suppose that for any integer l, $w^l = (t^l, z^l)$ satisfies (4.1). Now we prove that $w^{l+1} = (t^{l+1}, z^{l+1})$ satisfies (4.1) as well. We denote

$$\bar{d}^l(\lambda_l) = \tau^*(\lambda_l)\tilde{d}^l_G(\lambda_l) + (1 - \tau^*(\lambda_l))\tilde{d}^l_N(\lambda_l) = \begin{pmatrix} (\bar{d}^l(\lambda_l))_t \\ (\bar{d}^l(\lambda_l))_z \end{pmatrix},$$

where λ_l is the accepted step-length at l-th iteration. It follows from Algorithm 3.1 that

$$
\begin{aligned}
(\bar{d}^l(\lambda_l))_t &= \tau^*(\lambda_l)\lambda_l[-\gamma_l(t^l + \nabla_t \bar{H}(w)\bar{H}(w)) + \beta(w^l)\bar{t}\,] + (1 - \tau^*(\lambda_l))\lambda_l[-t^l + \beta(w^l)\bar{t}\,] \\
&= -\lambda_l\gamma_l\tau^*(\lambda_l)(t^l + \nabla_t\bar{H}(w)\bar{H}(w)) - (1 - \tau^*(\lambda_l))\lambda_l t^l + \lambda_l\beta(w^l)\bar{t} \\
&\geq -\lambda_l\tau^*(\lambda_l)t^l - (1 - \tau^*(\lambda_l))\lambda_l t^l + \lambda_l\beta(w^l)\bar{t} \\
&= -\lambda_l t^l + \lambda_l\beta(w^l)\bar{t},
\end{aligned}
$$

where the inequality comes from the definition of γ_l (see (3.8)). Then we have

$$
\begin{aligned}
t^{l+1} - \beta(w^{l+1})\bar{t} &= t^l + (\bar{d}^l(\lambda_l))_t - \beta(w^{l+1})\bar{t} \\
&\geq (1 - \lambda_l)t^l + \lambda_l\beta(w^l)\bar{t} - \beta(w^{l+1})\bar{t} \\
&\geq (1 - \lambda_l)t^l + \lambda_l\beta(w^l)\bar{t} - \beta(w^l)\bar{t} \\
&= (1 - \lambda_l)t^l - (1 - \lambda_l)\beta(w^l)\bar{t} \geq 0, \tag{4.3}
\end{aligned}
$$

where the second inequality is due to the monotonicity property of $\beta(w^l)$ in Proposition 3.2, and the last inequality comes from that $t^l \geq \beta(w^l)\bar{t}$. Therefore, (4.1) holds for any nonnegative integer k. Furthermore, from (4.1) and that w^k is not a stationary point of (3.4), (4.2) holds. We complete the proof. ∎

Theorem 4.1 *Let $\{w^k\} \subset W$ be a sequence generated by Algorithm 3.1. Then any accumulation point of $\{w^k\}$ is a stationary point of (3.4).*

Proof. Proposition 4.1 shows that if our algorithm does not stop at a stationary point of (3.4), then $t^k > 0$ for any k. This means that Φ and Ψ are continuously differentiable at w^k. Hence, by using a similar way to the proof of Theorem 4.1 [24], we can prove the theorem holds. Here we omit the detailed proof. ∎

15

The following proposition shows that the stationary point of (3.4) is a feasible point of the original SIP problem under some conditions. Let

$$J_1(x,v) = \{i \mid v_i + (\nabla_v g(x,v))_i < a_i\}, \quad J_2(x,v) = \{i \mid a_i < v_i + (\nabla_v g(x,v))_i < b_i\}$$

$$J_3(x,v) = \{i \mid b_i < v_i + (\nabla_v g(x,v))_i\},$$

and

$$J_4(x,v) = \{i \mid v_i + (\nabla_v g(x,v))_i = a_i \text{ or } v_i + (\nabla_v g(x,v))_i = b_i\}.$$

Proposition 4.2 *Let $\{w^k\}$ be a sequence generated by by Algorithm 3.1 and $w^* = \lim_{k \in K}\{w^k\}$ with $t^* = 0$ be a stationary point of (3.4), where $K \subset \{1, 2, \cdots\}$. Suppose that there exists a sub-vector v^{j_0*} of $\mathbf{v}^*$ such that w^* satisfies the following conditions:*

(i) $g(x^*, v^{j_0*}) = 0$;

(ii) $\nabla^2_{xv}g(x^*, v^{j_0*}) = 0$;

(iii) The function $g(x^*, \cdot)$ is concave at v^{j_0*};

(iv) The set $J_4(x^*, v^{j_0*})$ is empty and $\left(\nabla^2_{v^{j_0}}g(x^*, v^{j_0*})\right)_{J_2^{j_0*}J_2^{j_0*}}$ is nonsingular, where $J_2^{j_0*} = J_2(x^*, v^{j_0*})$.

Then x^* is a feasible point of the original SIP problem (1.1).

Proof. Since w^* is a stationary point of (3.4) and the projected operator only works on the variables u and y, it is not difficult to see that

$$\lim_{k \in K} \nabla_{(t,x,\mathbf{v})}\Psi(w^k) = 0. \tag{4.4}$$

For any w^k with $t^k > 0$,

$$\nabla\Psi(w^k) = \nabla\Phi(w^k)\Phi(w^k), \tag{4.5}$$

where

$$\nabla\Phi(w^k) = \begin{bmatrix} 1 & 0_{1\times n} & 0 & \cdots & 0 & \nabla_t\bar{G}^k & \nabla_t\bar{\phi}_1^k & \cdots & \nabla_t\bar{\phi}_p^k \\ 0_{n\times 1} & \nabla_x F^k & \nabla_x g_1^k & \cdots & \nabla_x g_p^k & \nabla_x\bar{G}^k & \nabla_x\bar{\phi}_1^k & \cdots & \nabla_x\bar{\phi}_p^k \\ 0 & \nabla_x^T g_1^k & 0 & \cdots & 0 & 0 & 0_{1\times m} & \cdots & 0_{1\times m} \\ 0 & \nabla_x^T g_2^k & 0 & \cdots & 0 & 0 & 0_{1\times m} & \cdots & 0_{1\times m} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \nabla_x^T g_p^k & 0 & \cdots & 0 & 0 & 0_{1\times m} & \cdots & 0_{1\times m} \\ 0_{m\times 1} & u_1^k\nabla_{v^1}(\nabla_x g_1^k) & \nabla_{v^1}g_1^k & \cdots & 0 & 0_{p\times 1} & \nabla_{v^1}\bar{\phi}_1^k & \cdots & 0_{m\times m} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0_{m\times 1} & u_p^k\nabla_{v^p}(\nabla_x g_p^k) & 0 & \cdots & \nabla_{v^p}g_p^k & 0_{p\times 1} & 0_{m\times m} & \cdots & \nabla_{v^p}\bar{\phi}_p^k \\ 0 & 0_{1\times n} & 0 & \cdots & 0 & 1 & 0_{1\times m} & \cdots & 0_{1\times m} \end{bmatrix}$$

16

and $F^k = F(x^k, u^k, \mathbf{v}^k)$, $\bar{G}^k = \bar{G}(t^k, x^k)$, $\bar{\phi}_j^k = \bar{\phi}(t^k, x^k, v^{jk})$ and $g_j^k = g(x^k, v^{jk})$ for any $j = 1, 2, \cdots, p$. Consequently, from (4.4) and (4.5), we have

$$
\begin{aligned}
\lim_{k \in K} \nabla_{v^{j_0}} \Psi(w^k) &= \lim_{k \in K} \left(u_{j_0}^k \nabla_{v^{j_0}} (\nabla_x g_{j_0}^k), \nabla_{v^{j_0}} g_{j_0}^k, \nabla_{v^{j_0}} \bar{\phi}_{j_0}^k \right) \begin{pmatrix} F^k \\ g_{j_0}^k \\ \bar{\phi}_{j_0}^k \end{pmatrix} \\
&= \left(u_{j_0}^* \nabla_{v^{j_0}} (\nabla_x g_{j_0}^*), \nabla_{v^{j_0}} g_{j_0}^*, Q_{j_0}^* \right) \begin{pmatrix} F^* \\ g_{j_0}^* \\ \bar{\phi}_{j_0}^* \end{pmatrix} \\
&= Q_{j_0}^* \bar{\phi}_{j_0}^* = 0,
\end{aligned}
\tag{4.6}
$$

where the third equality comes from the assumed conditions (i) and (ii). Here $F^* = F(x^*, u^*, \mathbf{v}^*)$, $g_{j_0}^* = g(x^*, v^{j_0*})$, $\bar{\phi}_{j_0}^* = \bar{\phi}(t^*, x^*, v^{j_0*})$ and $Q_{j_0}^{*T} \in \partial_{v^{j_0}} \bar{\phi}_{j_0}^*$. On the other hand, it is easy to know that

$$
Q_{j_0}^{*T} = \begin{pmatrix} I_1 & 0 & 0 \\ -U_{21}^* & -U_{22}^* & -U_{23}^* \\ 0 & 0 & I_3 \end{pmatrix},
$$

where U_{2j}^*, $j = 1, \cdots, 3$, are the sub-matrices of the $\nabla_v^2 g(x^*, v^{j_0*})$, which are determined by the rows with the index $k \in J_2(x^*, v^{j_0*})$ and the columns with the index $l \in J_j(x^*, v^{j_0*})$; I_1 and I_3 are the identity matrices with $|J_1^{j_0*}|$ and $|J_3^{j_0*}|$ order, respectively. Moreover, by the assumed conditions (iv), it is obvious that $Q_{j_0}^*$ is nonsingular and hence $\bar{\phi}_{j_0}^* = 0$, that is

$$
\bar{\phi}(0, x^*, v^{j_0*}) = 0.
\tag{4.7}
$$

This implies

$$
(v - v^{j_0*})^T (-\nabla_v g(x^*, v^{j_0*})) \geq 0, \quad \forall\, v \in V,
$$

which shows that v^{j_0*} is a KKT point of the minimization problem (1.3) with $x = x^*$. Consequently, by the standard sufficient optimality theorem, we have that for any $v \in V$,

$$
g(x^*, v) \leq g(x^*, v^{j_0*}) = 0,
$$

because $g(x^*, \cdot)$ is concave at v^{j_0*}. This shows that x^* is a feasible point of (1.1). The proof is complete. ∎

For the general case that $g(x, \cdot)$ is not concave, we have the following result.

Proposition 4.3 *Let $\{w^k\}$ be a sequence generated by Algorithm 3.1 and $w^* = \lim_{k \in K} \{w^k\}$ with $t^* = 0$ be a stationary point of (3.4), where $K \subset \{1, 2, \cdots\}$. Suppose that w^* satisfies the following conditions:*

(i) $0 < \int_V \dfrac{1}{|g(x^*, v)|} dv < \infty$;

(ii) $\lim_{k \in K} \dfrac{t^k}{t^k + \nabla_t \bar{H}(w^k) \bar{H}(w^k)} \geq 1$;

(iii) For every $j = 1, 2, \cdots, p$, the set $J_4^j(x^*, v^*)$ is empty;

(iv) For every $j = 1, 2, \cdots, p$, $\bar{\phi}(0, x^*, v^{j*}) = 0$.

Then x^* is a feasible point of the original SIP problem (1.1).

Proof. It is obvious that $G(x^*) \geq 0$. Now we prove that $G(x^*) = 0$. From the definition of the stationary point, we have

$$
\begin{cases}
y^* > 0 & \Rightarrow \quad \dfrac{\partial}{\partial y}\Psi(w^*) = 0, \\[2ex]
y^* = 0 & \Rightarrow \quad \dfrac{\partial}{\partial y}\Psi(w^*) \geq 0.
\end{cases}
\tag{4.8}
$$

This implies

$$
y^* = 0,
\tag{4.9}
$$

because

$$
\frac{\partial \Psi(w^*)}{\partial y} = \bar{G}(t^*, x^*) + y^* > 0
$$

whenever $y^* > 0$, which contradicts to (4.8). Moreover, for any w^k with $t^k > 0$,

$$
\nabla_t \bar{\phi}(t^k, x^k, v^{jk})
$$

$$
= 2t^k \left(\frac{1}{\sqrt{\left(b_1 - v_1^{jk} - (\nabla_v g_k^j)_1\right)^2 + 4(t^k)^2}} - \frac{1}{\sqrt{\left(a_1 - v_1^{jk} - (\nabla_v g_k^j)_1\right)^2 + 4(t^k)^2}}, \right.
$$

$$
\left. \cdots, \frac{1}{\sqrt{\left(b_m - v_m^{jk} - (\nabla_v g_k^j)_m\right)^2 + 4(t^k)^2}} - \frac{1}{\sqrt{\left(a_m - v_m^{jk} - (\nabla_v g_k^j)_m\right)^2 + 4(t^k)^2}} \right)^T
\tag{4.10}
$$

and

$$
\nabla_t \bar{G}(t^k, x^k) = 2t^k \int_V \frac{1}{\sqrt{(g(x^k, v))^2 + 4(t^k)^2}} dv,
\tag{4.11}
$$

where $g_k^j = g(x^k, v^{jk})$ for any $j = 1, 2, \cdots, p$. Consequently, we obtain

$$
\lim_{k \in K} \frac{t^k}{t^k + \nabla_t \bar{H}(w^k) \bar{H}(w^k)} = \frac{t^k}{t^k + \sum_{j=1}^{p} (\nabla_t \bar{\phi}_j^k)^T \bar{\phi}_j^k + \nabla_t \bar{G}^k \left(\bar{G}^k + y^k \right)}
$$

$$
= \frac{1}{1 + 2G(x^*) \int_V \dfrac{1}{|g(x^*, v)|} dv},
$$

where $\bar{\phi}_j^k = \bar{\phi}(t^k, x^k, v^{jk})$, $\bar{G}^k = \bar{G}(t^k, x^k)$, and the last equality comes from (4.9), (4.10), (4.11) and the assumed conditions (iii) and (iv). Therefore, by the assumed conditions (i) and (ii), we have

$$
G(x^*) = 0,
$$

18

which shows that x^* is a feasible point of (1.1). We complete the proof. ∎

In addition, it is worth mentioning that Algorithm 3.1 has been mainly concerned with finding a stationary point of (3.4), which is not necessarily a solution of (3.2). If the accumulation point of the sequence generated by Algorithm 3.1 is not a solution of (3.2), we may use the Lagrangian globalization (LG) method presented in [27] to deal with it. In other words, we solve the system of equations (3.2) via two phases. Phase-I is to solve (3.4) by Algorithm 3.1 in which the global optimal solutions with the zero objective function value correspond to the solutions of the original constrained equations (3.2). If the point obtained in this phase, say w^*, is a stationary point of (3.4), but not a solution of (3.2), then we go to Phase-II in which we use a projected approach to find a better feasible point $\hat{w}$ such that $\|\Phi(\hat{w})\| < \|\Phi(\bar{w})\|$, where $\bar{w}$ is the point obtained in Phase I. For some details, see [27].

In the rest of this section, we analyze the local convergence of Algorithm 3.1. We make the following standard assumption:

(A1) Let $w^* = (t^*, z^*) = (0, z^*)$ be an accumulation point of the sequence $\{w^k\}$ generated by Algorithm 3.1. Suppose $\lim_{k \in K} w^k = w^*$ for some subset $K \subset \{1, 2, \cdots\}$, w^* is a solution of the system of equations (3.3) and Φ is BD-regular at w^*.

BD-regularity can be satisfied without special difficulty.

Let $G(x, \mathbf{v})$ be the set of $(d, \xi_1, \zeta_1, \cdots, \xi_p, \zeta_p) \in R^n \times R^{(|J_2(x,v^1)|+|J_4(x,v^1)|)} \times \cdots \times R^{(|J_2(x,v^p)|+|J_4(x,v^p)|)}$ satisfying that for any $j = 1, \cdots, p$,

$$d^T \nabla_x g(x, v^j) = 0$$

and

$$(\nabla_v^T (\nabla_x g(x, v^j)))_{\cdot J_2^j}^T d + (\nabla_{vv}^2 g(x, v^j))_{J_2^j J_2^j} \xi_j + (\nabla_{vv}^2 g(x, v^j))_{J_2^j J_4^j} \zeta_j = 0,$$

where $J_2^j = J_2(x, v^j)$ and $J_4^j = J_4(x, v^j)$. In order to give a sufficient condition of BD-regularity, we make the following assumptions.

(A2) For all $j = 1, \cdots, p$, $u_j > 0$.

(A3) The vectors $\left\{\nabla_x g(x, v^j), j = 1, \cdots, p\right\}$ are linearly independent.

(A4) For all $(d, \xi_1, \zeta_1, \cdots, \xi_p, \zeta_p) \in G(x, \mathbf{v}) \backslash \{0\}$,

$$d^T \nabla_x^T F(x, u, \mathbf{v}) d - \sum_{j=1}^p u^j \xi_j^T (\nabla_{vv}^2 g(x, v^j))_{J_2^j J_2^j} \xi_j - \sum_{j=1}^p u^j \zeta_j^T (\nabla_{vv}^2 g(x, v^j))_{J_4^j J_4^j} \zeta_j$$
$$- 2 \sum_{j=1}^p u^j \xi_j^T (\nabla_{vv}^2 g(x, v^j))_{J_4^j J_2^j}^T \zeta_j > 0.$$

Remark. Suppose that (i) the matrix $\nabla_x F(x, u, \mathbf{v})$ is positive definite; (ii) for every $j = 1, 2, \ldots, p$, the set J_4^j is empty; and (iii) the matrix $(\nabla_{vv}^2 g(x, v^j))_{J_2^j J_2^j}$ is negative definite whenever $J_2^j \neq \emptyset$. Then the assumption (A4) holds automatically.

19

Theorem 4.2 *Suppose that $w^* = (t^*, z^*) = (t^*, x^*, u^*, \mathbf{v}^*, y^*)$ is a solution of (3.3) and satisfies (A2)-(A4). Then Φ is BD-regular at w^*.*

Proof. For the sake of simplicity, we prove the conclusion for $p = 1$. The proof of the conclusion for the general case with $p > 1$ is similar. Without loss of generality, we assume

$$J_1^* := J_1(x^*, v^*) = \{1, 2, \cdots k_1\}, \qquad J_2^* := J_2(x^*, v^*) = \{k_1 + 1, \cdots k_2\},$$
$$J_3^* := J_3(x^*, v^*) = \{k_2 + 1, \cdots k_3\}, \quad J_4^* := J_4(x^*, v^*) = \{k_3 + 1, \cdots m\},$$

where $1 \leq k_1 \leq k_2 \leq k_3 \leq m$. It is obvious that $w^* = (t^*, z^*)$ to be a solution of (3.3) implies $t^* = 0$. Moreover, we have, by $\bar{\phi}(0, x^*, v^*) = 0$, that

$$v^* - \text{mid}(a, b, v^* + \nabla_v g(x^*, v^*)) = 0, \tag{4.12}$$

which combines with the definition of the mid function to yield

$$(\nabla_v g(x^*, v^*))_i = 0 \tag{4.13}$$

for $i \in J_2^* \cup J_4^*$. Consequently, by direct computation, we obtain that for any $Q \in \partial_B \Phi(w^*)$,

$$Q = \begin{bmatrix}
1 & 0_{1\times n} & 0 & 0_{1\times m_1} & 0_{1\times m_2} & 0_{1\times m_3} & 0_{1\times m_4} & 0 \\
0_{n\times 1} & A & B & u^*C_1 & u^*C_2 & u^*C_3 & u^*C_4 & 0_{n\times 1} \\
0 & B^T & 0 & D_1 & 0_{1\times m_2} & D_3 & 0_{1\times m_4} & 0 \\
F_1 & F_2 & 0 & 0_{1\times m_1} & 0_{1\times m_2} & 0_{1\times m_3} & 0_{1\times m_4} & 1 \\
H_1 & 0_{m_1\times n} & 0_{m_1\times 1} & I_{m_1\times m_1} & 0_{m_1\times m_2} & 0_{m_1\times m_3} & 0_{m_1\times m_4} & 0_{m_1\times 1} \\
H_2 & -C_2^T & 0_{m_2\times 1} & -U_{21} & -U_{22} & -U_{23} & -U_{24} & 0_{m_2\times 1} \\
H_3 & 0_{m_3\times n} & 0_{m_3\times 1} & 0_{m_3\times m_1} & 0_{m_3\times m_2} & I_{m_3\times m_3} & 0_{m_3\times m_4} & 0_{m_3\times 1} \\
H_4 & -VC_4^T & 0_{m_4\times 1} & -VU_{41} & -VU_{42} & -VU_{43} & \Lambda I_4 - VU_{44} & 0_{m_4\times 1}
\end{bmatrix}, \tag{4.14}$$

where $m_1 = k_1$, $m_2 = k_2 - k_1$, $m_3 = k_3 - k_2$, $m_4 = m - k_3$ and

$$A = \nabla_x^T F(x^*, u^*, v^*), \qquad B = \nabla_x g(x^*, v^*),$$

$$C_j = (\nabla_v^T (\nabla_x g(x^*, v^*)))_{\cdot J_j^*} \qquad \text{for any} \quad j = 1, \cdots, 4,$$

$$U_{ij} = (\nabla_{vv}^2 g(x^*, v^*))_{J_i^* J_j^*} \qquad \text{for any} \quad i = 2, 4; j = 1, \cdots, 4,$$

$F_1 \in \partial_t G(0, x^*)$, $F_2 \in \partial_x G(0, x^*)$; H_j, $j = 1, \cdots, 4$, are the sub-vectors of $H \in \partial_t \bar{\phi}(0, x^*, v^*)$ with the index $k \in J_j^*$, respectively; I_4 is the identity matrix with m_4 order; $V = (I_4 - \Lambda)$ and $\Lambda = \text{diag}(\lambda_1, \cdots, \lambda_{m_4})$ with $\lambda_i \in [0, 1]$ for $i = 1, \cdots, m_4$.

From the structure of Q, we assume that, without loss of generality, $\lambda_i \in (0, 1)$ for $i = 1, \cdots, m_4$. Moreover, it is easy to see that the non-singularity of Q is the same as the matrix

$$\tilde{Q} = \begin{bmatrix}
A & B & u^*C_2 & u^*C_4 \\
B^T & 0 & 0_{1\times m_2} & 0_{1\times m_4} \\
C_2^T & 0_{m_2\times 1} & U_{22} & U_{24} \\
VC_4^T & 0_{m_4\times 1} & VU_{42} & -\Lambda I_4 + VU_{44}
\end{bmatrix}.$$

20

Suppose that

$$\tilde{Q}\begin{pmatrix} d_1 \\ d_2 \\ \xi \\ \zeta \end{pmatrix} = 0, \qquad (4.15)$$

where $d_1 \in R^n$, $d_2 \in R$, $\xi \in R^{m_2}$ and $\zeta \in R^{m_4}$. Then (4.15) implies

$$Ad_1 + Bd_2 + u^* C_2 \xi + u^* C_4 \zeta = 0, \qquad (4.16)$$

$$B^T d_1 = 0, \qquad (4.17)$$

$$C_2^T d_1 + U_{22}\xi + U_{24}\zeta = 0 \qquad (4.18)$$

and

$$V C_4^T d_1 + V U_{42}\xi - (\Lambda I_4 - V U_{44})\zeta = 0. \qquad (4.19)$$

Multiplication (4.16) with d_1^T yields

$$d_1^T A d_1 + d_1^T B d_2 + u^* d_1^T C_2 \xi + u^* d_1^T C_4 \zeta = 0,$$

which, together with (4.17), (4.18) and (4.19), shows

$$d_1^T A d_1 - u^* \xi^T U_{22}\xi - u^* \zeta^T U_{24}^T \xi - u^* \zeta^T U_{44}\zeta + u^* \zeta^T (\Lambda V^{-1})\zeta - u^* \xi^T U_{42}^T \zeta = 0.$$

It is obvious that $U_{24}^T = U_{42}$. Hence, from $u^* > 0$ and the fact that the matrix $\Lambda V^{-1} = \text{diag}\,(\lambda_1/(1 - \lambda_1), \cdots, \lambda_{m_4}/(1 - \lambda_{m_4}))$ is positive definite, we have

$$\begin{aligned} d_1^T \nabla_x^T F(x^*, u^*, v^*) d_1 - u^* \xi^T (\nabla_{vv}^2 g(x^*, v^*))_{J_2^* J_2^*} \xi \\ - u^* \zeta^T (\nabla_{vv}^2 g(x^*, v^*))_{J_4^* J_4^*} \zeta - 2u^* \xi^T (\nabla_{vv}^2 g(x^*, v^*))_{J_4^* J_2^*}^T \zeta \le 0. \end{aligned} \qquad (4.20)$$

On the other hand, by (4.17) and (4.18), it follows that $(d_1, \xi, \zeta) \in G(x^*, v^*)$. Thus, by (4.20) and the assumption (A4), we have that $(d_1, \xi, \zeta) = 0$. Consequently, from (4.16) and (A2), we obtain that $d_2 = 0$. Hence $(d_1, d_2, \xi, \zeta) = 0$, which shows that Q is nonsingular. Therefore, Φ is BD-regular at z^*. This completes the proof. ∎

From the BD-regularity condition and semismoothness of function Φ, we have the following lemma by using Lemmas 2.1 and 2.2.

Lemma 4.1 *There exist positive constants κ and ϵ such that for every w^k satisfying $\|w^k - w^*\| \le \epsilon$,*
(i) $\Phi'(w^k)$ is nonsingular and satisfies

$$\|\Phi'(w^k)\| \le \kappa.$$

(ii)

$$\|\Phi(w^k)\| = \sqrt{2}\Psi(w^k)^{\frac{1}{2}} = O(\|w^k - w^*\|).$$

21

Lemma 4.2 *For all $k \in K$ sufficiently large,*
(i)
$$\beta(w^k) = O(\Psi(w^k)) = O(\|w^k - w^*\|^2);$$

(ii) and for any $\lambda \in (0,1]$
$$w^k + \lambda d_N^k = (1 - \lambda)w^k + \lambda w^* + \lambda o(\Psi(w^k)^{\frac{1}{2}}). \tag{4.21}$$

Proof. From the definition of $\beta(w^k)$, the choice of γ_k, the projection property and Lemma 4.1, for w^k sufficiently close to w^*,

$$\beta(w^k) \le \alpha \|\bar{d}_G^k(1)\|^2 \le \alpha \gamma_k^2 \|\nabla\Psi(w^k)\|^2 \le \alpha\eta\Psi(w^k) = \frac{\alpha\eta}{2}\|\Phi(w^k)\|^2 = O(\|w^k - w^*\|^2).$$

This shows (i) holds. It follows from (i) and Lemma 4.1 that

$$
\begin{aligned}
w^k + \lambda d_N^k &= w^k + \lambda \Phi'(w^k)^{-1}[-\Phi(w^k) + \beta(w^k)\bar{w}] \\
&= w^k - \lambda\Phi'(w^k)^{-1}[\Phi(w^k) - \Phi(w^*) - \Phi'(w^k)(w^k - w^*)] \\
&\quad -\lambda(w^k - w^*) + \lambda\Phi'(w^k)^{-1}\beta(w^k)\bar{w} \\
&= (1 - \lambda)w^k + \lambda w^* + \lambda o(\|w^k - w^*\|) + \lambda O(\Psi(w^k)) \\
&= (1 - \lambda)w^k + \lambda w^* + \lambda o(\Psi(w^k)^{\frac{1}{2}}),
\end{aligned}
$$

where the third equality is due to the semismoothness of Φ and (i). (ii) is proved. The proof is complete ∎

Lemma 4.3 *For $k \in K$ large enough,*
$$\tilde{d}_N^k(\lambda) = -\lambda(w^k - w^*) + \lambda o(\Psi(w^k)^{\frac{1}{2}}) \tag{4.22}$$

and
$$\nabla\Psi(w^k)^T \tilde{d}_N^k(\lambda) \le -\mu\lambda\Psi(w^k), \tag{4.23}$$

where μ is any constant in $(0, 2)$.

Proof. From Lemma 4.2 and the property of a projector, we obtain that

$$
\begin{aligned}
\tilde{d}_N^k(\lambda) &= \Pi_W(w^k + \lambda d_N^k) - w^k \\
&= \Pi_W[(1 - \lambda)w^k + \lambda w^* + \lambda o(\Psi(w^k)^{\frac{1}{2}})] - w^k \\
&= \Pi_W[(1 - \lambda)w^k + \lambda w^*] - w^k \\
&\quad + \left\{ \Pi_W[(1 - \lambda)w^k + \lambda w^* + \lambda o(\Psi(w^k)^{\frac{1}{2}})] - \Pi_W[(1 - \lambda)w^k + \lambda w^*] \right\} \\
&= -\lambda(w^k - w^*) + \lambda o(\Psi(w^k)^{\frac{1}{2}}),
\end{aligned}
$$

where the last equality comes from $(1 - \lambda)w^k + \lambda w^* \in W$ and the projection property (see Lemma 3.1 (ii)). It follows from (4.22) that

$$
\begin{aligned}
\nabla \Psi(w^k)^T \tilde{d}_N^k(\lambda) &= -\lambda \Phi(w^k)^T \Phi'(w^k)(w^k - w^*) + \lambda o(\Psi(w^k)) \\
&= -2\lambda \Psi(w^k) + \lambda \Phi(w^k)^T [\Phi(w^k) - \Phi(w^*) - \Phi'(w^k)(w^k - w^*)] + \lambda o(\Psi(w^k)) \\
&\leq -\mu \lambda \Psi(w^k),
\end{aligned}
$$

where the last inequality comes from the semismoothness of Φ and Lemma 4.1. We complete the proof. ■

Lemma 4.4 *We have that for $k \in K$ large enough,*
(i)

$$
\tau^*(\lambda)_k \leq o(1), \tag{4.24}
$$

where $\tau^*(\lambda)_k$ is defined as in (3.14).
(ii)

$$
\bar{d}^k(\lambda) = -\lambda(w^k - w^*) + \lambda o(\Psi(w^k)^{\frac{1}{2}}). \tag{4.25}
$$

(iii)

$$
\nabla \Psi(w^k)^T \bar{d}^k(\lambda) = -2\lambda \Psi(w^k) + \lambda o(\Psi(w^k)). \tag{4.26}
$$

Proof. By using Lemma 4.3, this lemma can be proved in a similar way to the proof of Theorem 3.2 [24]. We omit the detailed proof. ■

Now we prove that the convergence rate of Algorithm 3.1 is locally superlinear under the BD-regularity condition.

Theorem 4.3 *Suppose that $\{w^k\}$ is a sequence generalized by Algorithm 3.1 and w^* is a point satisfying (A1). Then the whole sequence $\{w^k\}$ converges to w^* superlinearly.*

Proof. From Lemma 4.4 we have that for sufficiently large $k \in K$,

$$
\|w^k + \bar{d}^k(1) - w^*\| = o(\Psi(w^k)^{\frac{1}{2}}) = o(\|\Phi(w^k)\|) = o(\|w^k - w^*\|), \tag{4.27}
$$

and

$$
\begin{aligned}
\Psi(w^k + \bar{d}^k(1)) &= \frac{1}{2} \|\Phi(w^k + \bar{d}^k(1))\|^2 \\
&= \frac{1}{2} \|\Phi(w^k + \bar{d}^k(1)) - \Phi(w^*)\|^2 \\
&= O(\|w^k + \bar{d}^k(1) - w^*\|^2) \\
&= o(\Psi(w^k)), \tag{4.28}
\end{aligned}
$$

23

where the last equality is due to (4.27). Thus,

$$
\begin{aligned}
-\nabla\Psi(w^k)^T \tilde{d}_G^k(1) &\leq \|\nabla\Psi(w^k)\|\|\tilde{d}_G^k(1)\| \\
&= \|\nabla\Psi(w^k)\|\|\Pi_W(w^k - \gamma_k\nabla\Psi(w^k) + \beta(w^k)\bar{w}) - w^k\| \\
&\leq \|\nabla\Psi(w^k)\|[\|\gamma_k\nabla\Psi(w^k)\| + O(\Psi(w^k))] \\
&\leq \eta\Psi(w^k) + o(\Psi(w^k)),
\end{aligned} \tag{4.29}
$$

where the second inequality is due to the property of $\beta(w^k)$ and the projection property, and the last inequality comes from the choice of γ_k. It follows (4.28) and (4.29) that

$$
\Psi(w^k) + \sigma\nabla\Psi(w^k)^T \tilde{d}_G^k(1) \geq (1 - \sigma\eta)\Psi(w^k) + o(\Psi(w^k)) \geq o(\Psi(w^k)) = \Psi(w^k + \bar{d}^k(1)), \quad (4.30)
$$

which implies

$$
w^{k+1} = w^k + \bar{d}^k(1),
$$

for k sufficiently large. Moreover, from (4.27) we conclude that w^k converges to w^* superlinearly. We complete the proof. ∎

5 Preliminary Numerical Results

In this section, we report our preliminary numerical test results. We implemented Algorithm 3.1 in MATLAB and the numerical experiments were done by using a Pentium III 733MHz computer with 256 MB of RAM. We compared Algorithm 3.1 with `fseminf` that is a solver for SIP based on an implementation of the discretization SQP method in MATLAB toolbox. We tested 12 problems which we call problems 1-12. Problems 1-3 and 7 are from [28]. Problem 4 comes from [26] with a revised region. Problem 5 is a problem modified from [29], and Problem 6 is from [4]. Problems 8-12 are some problems in which the dimension of the parameter v is 2.

Throughout the computational experiments, we use $\|\bar{d}_G^k(1)\| \leq 10^{-6}$ as the stopping criterion for Algorithm 3.1. The values of $\bar{G}(t, x)$ and $\nabla\bar{G}(t, x)$ were computed by using the function `quad` in MATLAB when V is an interval in R and the function `dblquad` when V is a box set in R^2. The parameters used in the algorithm are specified as follows:

$$
\eta = 0.9, \ \rho = 0.5, \ \sigma = 0.001, \ \alpha = 0.5, \ \bar{t} = 0.9, \ p_1 = 1.0e - 10, \ p_2 = 2.1.
$$

The starting point u^0 and y^0 for all problems are set $t^0 = \bar{t}, \ u^0 = 0.05\mathbf{e}, \ y^0 = 0.5$, where $\mathbf{e}$ is the vector of ones. For the solver `fseminf`, we use all the default values.

Problem 1.

$$
f(x) = 1.21\exp(x_1) + \exp(x_2), \quad g(x, v) = v - exp(x_1 + x_2),
$$

$$
V = [-10, 1], p = 1, (x_0, v_0) = (1, 1, 1).
$$

Problem 2.

$$f(x) = x_1^2 + x_2^2 + x_3^2, \quad g(x, v) = x_1 + x_2\exp(x_3 v) + \exp(2v) - 2\sin(4v),$$

$$V = [0, 1], p = 1, (x_0, v_0) = (1, 1, 1, 1).$$

Problem 3.

$$f(x) = \frac{1}{3}x_1^2 + \frac{1}{2}x_1 + x_2^2, \quad g(x, v) = (1 - x_1^2 v^2)^2 - x_1 v^2 - x_2^2 + x_2,$$

$$V = [-1, 1], p = 1, (x_0, v_0) = (-1, -1, 1).$$

Problem 4.

$$f(x) = x_1^2 + (x_2 - 3)^2, \quad g(x, v) = x_2 - 2 + x_1\sin(v/x_2 - 0.5),$$

$$V = [0, 10], p = 1, (x_0, v_0) = (1, -1, 1).$$

Problem 5.

$$f(x) = \frac{1}{2}x^T x, \quad g(x, v) = 3 + 4.5\sin(4.7\pi(v - 1.23)/8) - \sum_{i=1}^{n} x_i v^{i-1},$$

$$V = [0, 1], n = 10, p = 1, (x_0, v_0) = (0, 0, \cdots, 0, 1).$$

Problem 6.

$$f(x) = (x_1 - 2x_2 + 5x_2^2 - x_2^3 - 13)^2 + (x_1 - 14x_2 + x_2^2 + x_2^3 - 29)^2,$$

$$g(x, v) = x_1^2 + 2x_2 v^2 + \exp(x_1 + x_2) - \exp(v), V = [0, 1], p = 1, (x_0, v_0) = (1, -1, 1).$$

Problem 7.

$$f(x) = x_1^2 + x_2^2 + x_3^3,$$

$$g(x, v) = x_1(v_1 + v_2^2 + 1) + x_2(v_1 v_2 - v_2^2) + x_3(v_1 v_2 + v_2^2 + v_2) + 1,$$

$$V = [0, 1] \times [0, 1], p = 1, (x_0, v_0) = (1, 1, 1, 1, 0).$$

Problem 8.

$$f(x) = x_1^2 + x_2^2 + x_3^2, \quad g(x, v) = x_1 + x_2\exp(x_3 v_1) + \exp(2v_2) - 2\sin(4v_1),$$

$$V = [0, 1] \times [0, 1], p = 2, (x_0, v_0) = (-1, -1, -1, 0, 1, 1, 0).$$

Problem 9.

$$f(x) = x_1^2 + x_2^2 + x_3^2, \quad g(x, v) = x_1 + x_2\exp(x_3 v_1) - \exp(2x_1 v_2) + \sin(4v_1),$$

$$V = [0, 1] \times [0, 1], p = 2, (x_0, v_0) = (-0.2, -0.2, -0.2, 0, 1, 1, 0).$$

Problem 10.

$$f(x) = x_1^2/3 + x_1/2 + x_2^2, \quad g(x, v) = (1 - x_1^2 v_1^2)^2 - x_1 v_2^2 - x_2^2 + x_2.$$

$$V = [0, 2] \times [0, 2], p = 2, (x_0, v_0) = (-0.2, -0.2, 1, 0, 0, 1).$$

Problem 11.

$$f(x) = 0.5(x_1^2 + x_2^2 + x_3^2 + x_4^2), \quad g(x, v) = \sin(v_1 v_2) - x_1 - x_2 v_1 - x_3 v_2 - x_4 v_1 v_2,$$

$$V = [0, 1] \times [0, 1], p = 1, (x_0, v_0) = (-0.5, -0.5, -0.5, -0.5, 0, 1).$$

Problem 12.

$$f(x) = 0.5(x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 + x_6^2),$$

$$g(x, v) = \exp(v_1^2 + v_2^2) - (x_1 + x_2 v_1 + x_3 v_2 + x_4 v_1^2 + x_5 v_1 v_2 + x_6 v_2^2),$$

$$V = [0, 1] \times [0, 1], p = 1, (x_0, v_0) = (-2, -2 - 2, -2, -2, -2, 1, 1).$$

In all above test problems, the values of p are estimated by using the following adaptive strategy. First, we let $p = 1$ and use Algorithm 3.1 to solve a test problem. If this test problem can be solved within 30 iterations, then we let $p = 1$ be the number of attainers at the solution. Otherwise, we let $p = 2$ and use Algorithm 3.1 to solve this test problem again. If this test problem can be solved within 30 iterations, then we let $p = 2$ be the number of attainers. If this fails again, then we let $p = 3$ and then do the above procedure until we find a number p ($p \leq n$) which is the estimated number of attainers. It is interesting that we get $p = 1$ for 9 of 12 test problems and p $= 2$ for other three test problems by the above method.

The test results are summarized in Tables 1 and 2. In Table 1, $\bar{d}_G^k(1)$ is the value of the function $\bar{d}_G(1)$ defined in (3.6) at the k-th iteration. In Table 2, **n.it** represents the number of the total iterations; **cpu** is the total cost time in seconds for solving the SIP problem; $\Psi(w^k)$ denote the value of the merit function $\Psi(w)$ of (3.3) at the final iteration; $f(x^k)$ is the value of the objective function in the SIP problem at the final iteration; and $G(x^k)$ is the value of the function $G(x)$ of (1.7) at the final iteration.

The results reported in Tables 1 and 2 show that Algorithm 3.1 performs well for these test problems. From Table 1, we can see that Algorithm 3.1 indeed has superlinear convergence property. From Table 2, we can see that Algorithm 3.1 uses less CPU time than `fseminf` for 9 test problems and `fseminf` uses less CPU time than Algorithm 3.1 for other 3 test problems. Moreover, it appears from Table 2 that Algorithm 3.1 indeed can ensure the feasibility of the test problems.

The numerical tests reported in the paper are very preliminary. Further experience with testing and with actual applications will be necessary and we leave it as our future research topic. In addition, we notice that for problems 1-7, 11 and 12, when $p \geq 2$, these test problems cannot be

solved by Algorithm 3.1 within 30 iterations. For problems 8-10, when $p = 1$, these three test problems cannot be solved by Algorithm 3.1 within 30 iterations. This means that it is important to choose a suitable number p when we use Algorithm 3.1 to solve the SIP problem. When the size of the SIP problem and the number p are large, the above method to determine the number p may be expensive in computation. As future work, we will work on how to find a good way to determine a suitable number p in the KKT system of the SIP problem.

Problem	k	$\bar{d}_G^k(1)$	Problem	k	$\bar{d}_G^k(1)$
1	4	0.0053	2	7	0.0141
	5	2.5166e-5		8	4.9812e-4
	6	3.6233e-10		9	4.5572e-7
3	5	0.0036	4	7	2.0688e-6
	6	9.671e-5		8	1.2389e-6
	7	3.8475e-9		9	4.2285e-7
5	2	0.0027	6	3	6.3414e-4
	3	3.5370e-5		4	3.2046e-5
	4	1.3002e-10		5	1.2032e-8
7	5	0.0228	8	5	3.7704e-5
	6	7.3722e-4		6	1.0023e-6
	7	4.1271e-7		7	2.7485e-10
9	8	0.0075	10	7	9.0029e-4
	9	4.0859e-5		8	2.2744e-6
	10	7.8563e-8		9	1.6448e-9
11	6	0.0086	12	3	9.5606e-4
	7	0.0016		4	1.3685e-7
	8	7.8146e-7		5	2.4486e-15

Table 1: The last three iterates generated by Algorithm 3.1 for Problems 1-12.

Problem	Algorithm 3.1					fseminf			
	$n.it$	cpu	$\Psi(w^k)$	$f(x^k)$	$G(x^k)$	$n.it$	cpu	$f(x^k)$	$G(x^k)$
1	6	0.09	9.1606e-19	2.2	0	16	0.53	2.1989	4.2815e-7
2	9	0.17	7.4361e-11	5.3347	0	23	0.42	5.3307	1.1142e-5
3	7	0.13	2.3462e-14	0.1945	0	4	0.12	0.1945	3.6407e-12
4	9	0.31	7.1952e-9	1	4.7985e-6	10	0.66	1	9.0699e-9
5	4	0.30	4.0342e-19	0.0657	0	2	0.42	0.0656	1.7460e-7
6	5	0.11	2.0507e-11	97.1589	1.3039e-9	8	0.17	97.1589	7.4551e-14
7	7	4.81	1.0514e-13	1	0	7	4.91	1	0
8	7	14.22	1.2988e-16	27.4166	0	6	5.33	27.3065	6.1921e-8
9	10	2.72	1.3677e-13	0	0	3	2.98	3.1516e-5	0
10	9	1.48	2.6635e-13	0.382	5.8388e-7	15	2.66	0.382	0
11	8	2.28	1.3049e-11	0.0885	0	1	2.00	0.0885	1.8080e-21
12	5	0.73	3.7729e-29	4.5498	0	1	3.38	4.5498	0

Table 2: Test results for Algorithm 3.1 and `fseminf`

6   Final Remarks

In this paper we have presented a smoothing projected Newton-type algorithm for solving the KKT system of the SIP problem. First, we reformulate the infinite constraints of the SIP problem to a constraint by using an integral function. Then, the KKT system of the SIP problem is written as a system of constrained nonsmooth equations, and solved by a smoothing projected Newton-type method. Under some standard assumptions, we prove the global and local superlinear convergence properties of this method. Compared with the existing methods such as discretization methods, exchange methods and local reduction methods, our method only needs to solve a system of linear equations at each iteration. Compared with the methods proposed in [21, 10], our method can ensure the feasibility of (1.1). As future work, one problem is to find a good way to determine a suitable number p in the KKT system of the SIP problem. Another problem is to find conditions of quadratic convergence of our method.

References

[1] P.H. Calamai and J.J. Moré, *Projected gradient methods for linear constrained problems*, Math. Prog., 39 (1987), 93-116.

[2] C. Chen, L. Qi and D. Sun, *Global and superlinear convergence of the smoothing Newton method and its application to general box constrained variational inequalities*, Math. Comp., 67 (1998), 519-540.

[3] F.H. Clarke, *Optimization and Nonsmooth Analysis*, John Wiley and Sons, New York, 1983.

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

[5] F. Facchinei and J.S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems, I-II*, Springer-Verlag, New York, 2003.

[6] A. Fischer, *Solution of monotone complementarity problems with locally Lipschitzian functions,* Math. Prog., 76 (1997), 513-532.

[7] S.A. Gabriel and J.J. Moré, *Smoothing of mixed complementarity problems*, in: M.C. Ferris and J.S. Pang, eds., Complementarity and Variational Problems: State of the Art, SIAM, Philadelphia, 1997, 105-116.

[8] M.A. Goberna and M.A. López, *Semi-Infinite Programming: Recent Advances*, Kluwer Academic Publishers, 2001.

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

[10] 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 .

[11] C. Ling, L. Qi, G. Zhou and S.Y. Wu, *Global convergence of a robust smoothing SQP method for semi-infinite programming*, J. Optim. Theory Appl., to appear.

[12] F. Meng, D. Sun and G. Zhao, *Semismoothness of solutions to generalized equations and the Moreau-Yosida regularization*, Technical report, Department of Mathematics, National University of Singapore, Singapore, June 2004.

[13] R. Mifflin, *Semismooth and semiconvex functions in constrained optimization,* SIAM J. Control Optim., 15 (1977), 957-972.

[14] J.S. Pang and L. Qi, *Nonsmooth equations: Motivation and algorithms*, SIAM J. Optim., 3 (1993), 443-465.

[15] E. Polak, *On the mathematical foundation of nondifferentiable optimization in engineering design*, SIAM Review, 29 (1987), 21-89.

[16] E. Polak, *Optimization: Algorithms and Consistent Approximation*, Springer-Verlag, New York, 1997.

[17] L. Qi, *Convergence analysis of some algorithms for solving nonsmooth equations*, Math. Oper. Res., 18 (1993), 227-244.

[18] L. Qi, A. Shapiro and C. Ling, *Differentiability and semismoothness properties of integral functions and their applications*, Math. Prog., Ser. A, 102 (2005), 223-248.

[19] L. Qi, D. Sun and G. Zhou, *A new look at smoothing Newton methods for nonlinear complementarity problems and box constrained variational inequalities*, Math. Prog., 87 (2000), 1-35.

[20] L. Qi and J. Sun, *A nonsmooth version of Newton's method*, Math. Prog., 58 (1993), 353-367.

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

[22] A. Shapiro, *First and second order optimality conditions and perturbation analysis of semi-infinite programming problems*, in R. Reemtsen and J. Rükmann eds., Semi-Infinite Programming, Kluwer Academic Publishers, Boston, 1998, 103-133.

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

[24] D. Sun, R.S. Womersley and H. Qi, *A feasible semismooth asymptotically Newton method for mixed complementarity problems*, Math. Prog., 94 (2002), 167-187.

[25] K.L. Teo, V. Rehbock and L.S. Jennings, *A new computational algorithm for functional inequality constrained optimization problems*, Automatica, 29 (1993), 789-792.

[26] K.L. Teo, X.Q. Yang and L.S. Jennings, *Computational discretization algorithms for functional inequality constrained optimization*, Ann. Oper. Res., 98 (2000), 215-234.

[27] X.J. Tong, L. Qi and Y.F. Yang, *The Lagrangian globalization method for nonsmooth constrained equations*, Comput. Optim. Appl., 33 (2006), 89-109.

[28] G.A. Watson, *Numerical experiments with globally convergent methods for semi-infinite programming problems*, in: A.V. Fiacco and K.O. Kortanek, eds., Semi-Infinite Programming and Applications, Springer-Verlag, Berlin, 1981, 193-205.

[29] 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 Softw., 20 (2005), 629-643.