J Glob Optim (2010) 47:133–159
DOI 10.1007/s10898-009-9462-7

A new smoothing Newton-type algorithm for semi-infinite programming

Chen Ling · Qin Ni · Liqun Qi · Soon-Yi Wu

Received: 12 July 2007 / Accepted: 23 July 2009 / Published online: 1 August 2009

Abstract We consider a semismooth reformulation of the KKT system arising from the semi-infinite programming (SIP) problem. Based upon this reformulation, we present a new smoothing Newton-type method for the solution of SIP problem. The main properties of this method are: (a) it is globally convergent at least to a stationary point of the SIP problem, (b) it is locally superlinearly convergent under a certain regularity condition, (c) the feasibility is ensured via the aggregated constraint, and (d) it has to solve just one linear system of equations at each iteration. Preliminary numerical results are reported.

Keywords Semi-infinite programming (SIP) problem · KKT system · Nonsmooth equations · Smoothing method · Convergence

Chen Ling work was supported by the National Natural Science Foundation of China (10871168), the Zhejiang Provincial National Science Foundation of China (Y606168) and a Hong Kong Polytechnic University Postdoctoral Fellowship. Qin Ni work was supported by the National Natural Science Foundation of China (grant 10471062). Liqun Qi work was supported by the Hong Kong Research Grant Council (Grant PolyU 102307) and a Chair Professor Fund of the Hong Kong Polytechnic University.

C. Ling
School of Mathematics and Statistics, Zhejiang University of Finance and Economics,
310018 Hangzhou, China
e-mail: linghz@hzcnc.com

Q. Ni
Department of Mathematics, Nanjing University of Aeronautics and Astronautics, 210016 Nanjing, China
e-mail: niqfs@nuaa.edu.cn

C. Ling · L. Qi (✉)
Department of Applied Mathematics, The Hong Kong Polytechnic University,
Hung Hom, Kowloon, Hong Kong
e-mail: maqilq@polyu.edu.hk

S.-Y. Wu
Department of Mathematics, National Cheng Kung University/National Center for Theoretical Sciences,
Tainan, Taiwan
e-mail: soonyi@mail.ncku.edu.tw

1 Introduction

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

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

where $X = \{x \in R^n : g(x, v) \leq 0, \ \forall\, v \in V\}$, V is a nonempty compact subset of R^m, defined by $V = \{v \in R^m : c(v) \leq 0\}$, $f : R^n \to R$ and $g : R^n \times R^m \to R$ and $c : R^m \to R^q$ are twice continuously differentiable functions.

The SIP problem arises from various applications such as approximation theory, optimal control, eigenvalue computation, mechanical stress of materials, pollution control and statistical design. Therefore, the solution methods for SIP problems are very important. Since the main difficulty for solving the SIP problem is that it has infinite constraints, the main effort of existing methods is to reduce the infinite set V to a finite one. Many methods have been proposed for the SIP problem, which can be roughly divided into three types: (1) discretization methods, (2) exchange methods, and (3) local reduction methods. We refer readers to [2,4–10,13,15,20,27,31–33,36,37] for details.

Let

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

It is well-known [29] 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 $h \in R^n$ such that

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

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 $i \in \mathcal{P} := \{1, 2, \ldots, p\}$ and $p \leq n$. Hence, the KKT system of the SIP problem (1.1) is as follows:

$$\begin{cases} \nabla f(x) + \sum\limits_{i=1}^{p} u_i \nabla_x g(x, v^i) = 0, \\ g(x, v) \leq 0, \quad \forall\, v \in V, \\ u_i > 0, \quad g(x, v^i) = 0, \ i \in \mathcal{P}. \end{cases} \tag{1.2}$$

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

Moreover, by the definition of $V(x)$ and the second constrained condition of (1.2), $v^i \in V(x)$ $(i \in \mathcal{P})$ imply that v^i $(i \in \mathcal{P})$ are global minimizers of the following minimization problem:

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

The KKT system of (1.3) can be written as

$$\begin{cases} -\nabla_v g(x, v^i) + \sum_{j=1}^{q} w_j^i \nabla c_j(v^i) = 0, \\ w_j^i \geq 0, \ c_j(v^i) \leq 0, \\ w_j^i \, c_j(v^i) = 0, \ i \in \mathcal{P}, \ j \in \mathcal{Q}, \end{cases} \tag{1.4}$$

where $w^i \equiv (w_1^i, \ldots w_q^i) \in R^q$, $(i \in \mathcal{P})$ and $\mathcal{Q} := \{1, \ldots, q\}$. Thus, the system (1.2) and $v^i \in V(x)$ $(i \in \mathcal{P})$ are transformed into the following system:

$$\begin{cases} \nabla f(x) + \sum_{i=1}^{p} u_i \nabla_x g(x, v^i) = 0, \\ g(x, v) \leq 0, \quad \forall\, v \in V, \\ u_i > 0, \quad g(x, v^i) = 0, \ i \in \mathcal{P}, \\ -\nabla_v g(x, v^i) + \sum_{j=1}^{q} w_j^i \nabla c_j(v^i) = 0, \\ w_j^i \geq 0, \ c_j(v^i) \leq 0, \\ w_j^i \, c_j(v^i) = 0, \ i \in \mathcal{P}, \ j \in \mathcal{Q}. \end{cases} \tag{1.5}$$

It is then desirable to develop numerical methods on the basis of (1.5). However, we realize that in order to possess the conditions for the CD-regularity required by our algorithm, we need to modify the above system accordingly. The definition of the CD-regularity and the conditions for CD-regularity will be presented in Sects. 2 and 5, respectively. Since $u_i > 0$ for $i \in \mathcal{P}$, we may multiply the fourth equation in (1.5) by u_i and then further replace $u_i w_j^i$ by w_j^i for $i \in \mathcal{P}$; $j \in \mathcal{Q}$. Thus, in the case that $u_i > 0$ for $i \in \mathcal{P}$, the system (1.5) is equivalent to the following:

$$\begin{cases} \nabla f(x) + \sum_{i=1}^{p} u_i \nabla_x g(x, v^i) = 0, \\ g(x, v) \leq 0, \quad \forall\, v \in V, \\ u_i > 0, \quad g(x, v^i) = 0, i \in \mathcal{P}, \\ -u_i \nabla_v g(x, v^i) + \sum_{j=1}^{q} w_j^i \nabla c_j(v^i) = 0, \\ w_j^i \geq 0, \ c_j(v^i) \leq 0, \\ w_j^i \, c_j(v^i) = 0, \ i \in \mathcal{P}, \ j \in \mathcal{Q}. \end{cases} \tag{1.6}$$

Based on (1.6) except the feasibility constraints, a semismooth Newton method and a smoothing Newton method were presented in [26] and [14], respectively. The advantage of these two methods proposed in [14,26] 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). Recently, another iterative method for solving the KKT system of (1.1) was proposed in [39], in which the feasibility issue was considered. However, the method in [39] does not have locally superlinear convergence property. Quite recently, based on the constrained equations reformulation of the KKT system of the SIP problem with box parameter set V, two smoothing projected Newton-type algorithms for SIP problem were presented in [18,22]. However, the accumulation points of the sequences generated by those algorithms are not necessarily stationary points of the SIP problem.

In this paper, we present a new method for solving the SIP problem by using a smoothing Newton-type algorithm to solve (1.6). At each iteration only a system of linear equations needs to be solved. The feasibility is ensured via the aggregated constraint. Global and locally superlinear convergence of this method is established under some mild assumptions. Some drawbacks of existing methods are overcome.

The rest of this paper is organized as follows. In Sect. 2, we reformulate the system (1.6) into a system of semismooth equations by using an NCP function ϕ and an integral function G. In Sect. 3, we study the properties of the smoothing functions $\bar{G}(\cdot,\cdot)$ and $\bar{\phi}(\cdot,\cdot,\cdot)$ of $G(\cdot)$ and $\phi(\cdot,\cdot)$, respectively. In Sect. 4, a smoothing Newton-type algorithm is presented to solve (1.6). This smoothing algorithm is a modified version of the methods presented in [12,24]. In Sect. 5 we establish the global and locally superlinear convergence of the new method. In Sect. 6, we give our numerical results, which show that our new method performs well, whenever the evaluation of the integral function is not very expensive. Specially, for the SIP problem with higher dimension decision variable, the presented algorithm is hopeful. Some comments are made in the last section.

Some words about the notation. For a smooth (continuously differentiable) function $F: R^n \to R^m$, we denote the Jacobian of F at $x \in R^n$ by $DF(x)$, which is an $m \times n$ matrix. We denote the transposed Jacobian as $\nabla F(x)$. For a smooth function $g: R^n \times R^m \to R$, we denote by $\nabla_x g(x, y)$ the gradient of g at (x, y) with respect to x and by $\nabla^2_{xx} g(x, y)$, $\nabla^2_{xy} g(x, y) = D_y \nabla_x g(x, y)$ and $\nabla^2_{yy} g(x, y)$, the $n \times n$, $n \times m$ and $m \times m$ matrices of second order partial derivatives of g at (x, y), respectively. For a nonsmooth function $G(x)$, $\partial G(x)$ means the generalized Jacobian in the sense of Clarke [1]. If $\mathcal{T}$ is a finite set, we let $|\mathcal{T}|$ denote its cardinality, that is the number of elements of $\mathcal{T}$. For an $m \times n$ matrix M, a subset $\mathcal{I}$ of $\{1, 2, \ldots, m\}$ and a subset $\mathcal{J}$ of $\{1, 2, \ldots, n\}$, we use the notation $M_{\mathcal{I}\mathcal{J}}$ for the $|\mathcal{I}| \times |\mathcal{J}|$ sub-matrix obtained by deleting all rows $i \notin \mathcal{I}$ and all columns $j \notin \mathcal{J}$, and use the notation $M_{\mathcal{I}\cdot}$ ($M_{\cdot\mathcal{J}}$) for $|\mathcal{I}| \times n$ ($m \times |\mathcal{J}|$) the sub-matrix obtained by deleting all rows $i \notin \mathcal{I}$ (all columns $j \notin \mathcal{J}$). $\|\cdot\|$ 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.

2 A semismooth equation reformulation

In this section, we reformulate the system (1.6) into a system of semismooth equations. We first briefly review some concepts and results on semismoothness and NCP functions.

Let $H: R^n \to R^n$ be locally Lipschitzian continuous. In [21], Qi defined the B-subdifferential of a locally Lipschitz function H at a point $x \in R^n$:

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

Then the Clarke generalized Jacobian [1] of H at x is defined by

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

A locally Lipschitz function H is said to be CD-regular at $x \in R^n$ if all $Q \in \partial H(x)$ are nonsingular.

Semismoothness was originally introduced by Mifflin [16] for functionals. In [25], 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 [25]:

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

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

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

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

A function H is said to be a semismooth function if it is semismooth everywhere on R^n.

Lemma 2.1 [21] *Suppose that $H : R^n \to R^n$ is locally Lipschitz continuous and H is CD-regular at $x \in R^n$. Then there exist a neighborhood $\mathcal{N}(x)$ of x and a constant C such that for any $y \in \mathcal{N}(x)$ and $Q \in \partial H(y)$, Q is nonsingular and $||Q^{-1}|| \le C$.*

Lemma 2.2 [19] *Suppose that $H : R^n \to R^n$ is locally Lipschitz continuous and H is CD-regular at a solution x^* of $H(x) = 0$. If H is semismooth at x^*, then there exist a neighborhood $\mathcal{N}(x^*)$ of x^* and a constant C such that for any $x \in \mathcal{N}(x^*)$,*

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

A function $\phi : R^2 \to R$ is called an NCP function if $\phi(a, b) = 0$ if and only if $a \ge 0$, $b \ge 0$ and $ab = 0$. Two well-known NCP functions are the minimum function

$$\phi_{\min}(a, b) := \min\{a, b\}$$

and the Fischer–Burmeister function

$$\phi_{FB}(a, b) = \sqrt{a^2 + b^2} - a - b. \tag{2.1}$$

Both the minimum function and the Fischer–Burmeister function are not smooth, but they are semismooth. Here and throughout this paper, we use the Fischer–Burmeister function.

Let

$$G(x) = \int_V [g(x, v)]_+ dv, \tag{2.2}$$

where $[x]_+ = \max\{0, x\}$. The function $G(x)$ was proposed to be used on SIP in [34]. It is not difficult to show that $G(x) \ge 0$ and G is nonsmooth but semismooth [23]. By the use of the functions ϕ and G defined by (2.1) and (2.2), respectively, (1.6) is reformulated as the following system:

$$\begin{cases} \nabla f(x) + \sum_{i=1}^{p} u_i \nabla_x g(x, v^i) = 0, \\ G(x) + s = 0, \\ \phi(u_i, -g(x, v^i)) = 0, \\ -u_i \nabla_v g(x, v^i) + \sum_{j=1}^{q} w_j^i \nabla c_j(v^i) = 0, \\ \phi(w_j^i, -c_j(v^i)) = 0,\ i \in \mathcal{P},\ j \in \mathcal{Q}, \end{cases} \tag{2.3}$$

which can be written as the following system of semismooth equations:

$$H(s, z) := \begin{pmatrix} G(x) + s \\ P(z) \end{pmatrix} = 0, \tag{2.4}$$

where

$$P(z) = \begin{pmatrix} \nabla f(x) + \sum_{i=1}^{p} u_i \nabla_x g(x, v^i) \\ \phi(u_1, -g(x, v^1)) \\ \vdots \\ \phi(u_p, -g(x, v^p)) \\ -u_1 \nabla_v g(x, v^1) + \sum_{j=1}^{q} w_j^1 \nabla c_j(v^1) \\ \vdots \\ -u_p \nabla_v g(x, v^p) + \sum_{j=1}^{q} w_j^p \nabla c_j(v^p) \\ \phi(w_1^1, -c_1(v^1)) \\ \vdots \\ \phi(w_q^1, -c_q(v^1)) \\ \vdots \\ \phi(w_1^p, -c_1(v^p)) \\ \vdots \\ \phi(w_q^p, -c_q(v^p)) \end{pmatrix},$$

$(s, z) = (s, x, u, \mathbf{v}, \mathbf{w}) \in R^{1+n+p(m+q+1)}$, $\mathbf{v} = (v^1, \ldots, v^p) \in R^{pm}$ and $\mathbf{w} = (w^1, \ldots, w^p) \in R^{pq}$. Note that $s \in R$ is an artificial variable which ensures the numbers of the variables in the system equal to the numbers of the equations. At the same time, we can write out the expression of the element U in $\partial H(s, z)$ and see that the introduction of artificial variable s can reduce the possible degeneration generated by the function $G(x)$.

It is similar to that stated in [14], that if there is an $1 + n + (m + q + 1)p$ dimensional vector satisfying (2.4) and $s = 0$, we may then drop the part indexed by i where $u_i = 0$. In this case, we get a solution of (1.6) which obviously satisfies (2.4). Hence, in this sense, (1.6) is equivalent to (2.4). In Sect. 4, we will present an algorithm for solving the system of nonsmooth equations (2.4).

3 Smoothing functions

The nonsmoothness of G and ϕ in (2.4) results in the difficulty of the implementation of the algorithm for solving (2.4). To overcome this drawback, in this section, we introduce the smoothing functions for G and ϕ, and recall some properties related to these smoothing functions. Let $t \in R$ be a parameter.

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

$$\bar{g}(t, x, v) = \frac{\sqrt{(g(x, v))^2 + 4t^2} + g(x, v)}{2}. \tag{3.5}$$

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 [24]. It is obvious that for any $t \neq 0$, $\bar{G}(t, x)$ is smooth with respect to variable x and

$$\nabla_x \bar{G}(t, x) = \int\limits_V \nabla_x \bar{g}(t, x, v) dv. \tag{3.6}$$

Define $\bar{\phi} : R^3 \to R$ by

$$\bar{\phi}(t, a, b) = \sqrt{a^2 + b^2 + t^2} - a - b.$$

Let $w \in R$ and $h : R^m \to R$ be continuously differentiable. Denote $\tilde{\phi} : R \times R \times R^m \to R$ as follow

$$\tilde{\phi}(t, w, v) = \bar{\phi}(t, w, h(v)). \tag{3.7}$$

For the functions $\bar{G}$ and $\tilde{\phi}$, we have the following propositions.

Proposition 3.1 [22] *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 with respect to (t, x).*

Proposition 3.2 [22] *The function $\tilde{\phi}$ defined in (3.7) 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 $(w, v) \in R \times R^n$*

$$\left\| \tilde{\phi}(t, w, v) - \phi(w, h(v)) \right\| \leq C|t|.$$

(iii) *The function $\tilde{\phi}$ is semismooth with respect to (t, w, v).*

Denote $y = (t, s, z) = (t, s, x, u, \mathbf{v}, \mathbf{w}) \in R^{2+n+p(m+q+1)}$. We define the following system of equations:

$$\Phi(y) = 0, \tag{3.8}$$

where

$$\Phi(y) = \begin{pmatrix} t \\ \bar{G}(t, x) + s \\ \bar{P}(t, z) \end{pmatrix}$$

and

$$\bar{P}(t,z) = \begin{pmatrix} \nabla f(x) + \sum_{i=1}^{p} u_i \nabla_x g(x, v^i) \\ \bar{\phi}(t, u_1, -g(x, v^1)) \\ \vdots \\ \bar{\phi}(t, u_p, -g(x, v^p)) \\ -u_1 \nabla_v g(x, v^1) + \sum_{j=1}^{q} w_j^1 \nabla c_j(v^1) \\ \vdots \\ -u_p \nabla_v g(x, v^p) + \sum_{j=1}^{q} w_j^p \nabla c_j(v^p) \\ \bar{\phi}(t, w_1^1, -c_1(v^1)) \\ \vdots \\ \bar{\phi}(t, w_q^1, -c_q(v^1)) \\ \vdots \\ \bar{\phi}(t, w_1^p, -c_1(v^p)) \\ \vdots \\ \bar{\phi}(t, w_q^p, -c_q(v^p)) \end{pmatrix}.$$

It follows from Propositions 3.1 and 3.2 that the following result holds.

Theorem 3.1 Φ *in* (3.8) *is semismooth.*

It is obvious that if $y = (t, s, z)$ with $s \geq 0$ is a solution of (3.8) then (s, z) is a solution to (2.4), and hence $z = (x, u, \mathbf{v}, \mathbf{w})$ is a solution of (1.6).

4 A smoothing Newton-type algorithm

In this section, motivated by the two methods in [12,24], we present a new smoothing Newton-type method for solving the KKT system of (1.1).

Define a merit function of (3.8) by

$$\theta(y) = \frac{1}{2}\|\Phi(y)\|^2.$$

Note that $\theta(\cdot)$ is smooth at any $y \in R^{2+n+p(m+q+1)}$ with $t > 0$.

Let $\gamma \in (0, 1)$ be a constant and choose $(\bar{t}, \bar{s}) \in R^2_{++}$ such that $\gamma\sqrt{\bar{t}^2 + \bar{s}^2} < 1$. Let $\bar{y} = (\bar{t}, \bar{s}, 0) \in R^2 \times R^{n+p(m+q+1)}$. For two sequences $\{y^k\}_{k=0}^{\infty} \subset R^2_{++} \times R^{n+p(m+q+1)}$ and $\{\alpha_k\}_{k=0}^{\infty} \subset R_{++}$, we define

$$\beta_0 = \beta(y^0) := \gamma \min\{1, \|\alpha_0 \nabla\theta(y^0)\|^2\}$$

and

$$\beta_k = \beta(y^k) := \begin{cases} \beta_{k-1}, & \text{if } \gamma \min\{1, \|\alpha_k \nabla\theta(y^k)\|^2\} > \beta_{k-1} \\ \gamma \min\{1, \|\alpha_k \nabla\theta(y^k)\|^2\}, & \text{otherwise.} \end{cases} \tag{4.1}$$

Then, because for any $y^0 \in R^2_{++} \times R^{n+p(m+q+1)}$ and $\alpha_0 > 0$, $\beta_0 \le \gamma < 1$, it follows that $(\bar{t}, \bar{s}) \ge \beta_0(\bar{t}, \bar{s})$.

Now, we state the smoothing Newton-type algorithm for solving (3.8).

Algorithm 4.1

Step 0. (Initialization)
Choose constants $\rho, \sigma, r \in (0, 1)$. Let $t^0 = \bar{t}$, $s^0 = \bar{s}$, $z^0 \in R^{n+p(m+q+1)}$ be an arbitrary point and $y^0 = (t^0, s^0, z^0)$. Set $k := 0$.

Step 1. (Stopping Test)
If $\|\nabla\theta(y^k)\| = 0$, stop. Otherwise, let

$$\alpha_k = \min\left\{1, \frac{s^k}{\bar{G}(t^k, x^k) + s^k}, \frac{t_k}{|\nabla_t \theta(y^k)|}\right\}, \tag{4.2}$$

and compute β_k by (4.1).

Step 2. (Compute Search Direction)
Compute $(d_N^k)_t$ and $(d_N^k)_s$ by

$$(d_N^k)_t = \beta_k \bar{t} - t^k, \quad (d_N^k)_s = \beta_k \bar{s} - s^k. \tag{4.3}$$

And compute $(d_N^k)_z$ by solving the following linear system

$$\nabla_z \bar{P}(t^k, z^k)^T (d_N^k)_z = -\left[\bar{P}(t^k, z^k) + \nabla_t \bar{P}(t^k, z^k)^T (d_N^k)_t\right]. \tag{4.4}$$

Let $d_N^k = ((d_N^k)_t, (d_N^k)_s, (d_N^k)_z)$.
Let d_G^k be computed by

$$d_G^k = -\alpha_k \nabla\theta(y^k) + \beta_k \bar{y}. \tag{4.5}$$

Step 3. (Computation of New Iterate)
If the solution d_N^k of (4.3)–(4.4) exists and

$$\theta(y^k + d_N^k) \le \sigma\theta(y^k), \tag{4.6}$$

then (*** fast step ***)

set $y^{k+1} := y^k + d_N^k$;

else (*** safe step ***)

let m_k be the smallest nonnegative integer m satisfying

$$\theta(y^k + r^m d_G^k) \le \theta(y^k) - \sigma\alpha_k\left(1 - \gamma\sqrt{\bar{t}^2 + \bar{s}^2}\right) r^m \|\nabla\theta(y^k)\|^2, \tag{4.7}$$

and set $y^{k+1} = y^k + r^{m_k} d_G^k$.

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

Remark It is remarked that $\bar{G}(t, x)$ in (3.8) and its derivative are not evaluated exactly. The functions `quad` or `dblquad` with the absolute error tolerance 10^{-6} in MATLAB are used to compute $\bar{G}(t, x)$ and its derivative. Numerical results show that this choice is proper.

In the rest of this section, we discuss some properties for Algorithm 4.1.

Lemma 4.1 *For any $\tilde{y} = (\tilde{t}, \tilde{s}, \tilde{z}) \in R^2_{++} \times R^{n+p(m+q+1)}$. Suppose that $\nabla\Phi(\tilde{y})$ is nonsingular, then there exist a closed neighborhood $\mathcal{N}(\tilde{y})$ of $\tilde{y}$ and a positive number $\tilde{\lambda} \in (0, 1]$*

such that for any $y = (t, s, z) \in \mathcal{N}(\tilde{y})$ *and all* $\lambda \in (0, \tilde{\lambda}]$ *we have* $(t, s) \in R^2_{++}$, $\nabla\Phi(y)$ *is invertible and*

$$\theta(y + \lambda d^y_G) \le \theta(y) - \lambda\sigma\alpha_y \left(1 - \gamma\sqrt{\bar{t}^2 + \bar{s}^2}\right) \|\nabla\theta(y)\|^2, \tag{4.8}$$

where

$$d^y_G = -\alpha_y \nabla\theta(y) + \beta(y)\bar{y} \tag{4.9}$$

and

$$\alpha_y = \min\left\{1, \frac{s}{\bar{G}(t,x) + s}, \frac{t}{|\nabla_t\theta(y)|}\right\}, \quad \beta(y) = \gamma \min\{1, \|\alpha_y \nabla\theta(y)\|^2\}.$$

Proof Since $\nabla\Phi(\tilde{y})$ is invertible and $(\tilde{t}, \tilde{s}) \in R^2_{++}$, there exists a closed neighborhood $\mathcal{N}(\tilde{y})$ of $\tilde{y}$ such that for any $y = (t, s, z) \in \mathcal{N}(\tilde{y})$ we have $(t, s) \in R^2_{++}$ and that $\nabla\Phi(y)$ is invertible. For any $y \in \mathcal{N}(\tilde{y})$ and $\lambda \in [0, 1]$, define

$$g_y(\lambda) = \theta(y + \lambda d^y_G) - \theta(y) - \lambda\nabla\theta(y)^T d^y_G,$$

then, it follows from the Mean Value Theorem that

$$g_y(\lambda) = \lambda \int_0^1 \left(D\theta(y + \tau\lambda d^y_G) - D\theta(y)\right) d^y_G d\tau.$$

Since $D\theta(\cdot)$ is uniformly continuous on $\mathcal{N}(\tilde{y})$ and $d^y_G \to d^{\tilde{y}}_G$ as $y \to \tilde{y}$, for all $y \in \mathcal{N}(\tilde{y})$

$$\lim_{\lambda\downarrow 0} g_y(\lambda)/\lambda = 0. \tag{4.10}$$

On the other hand, it is easy to see that $\beta(y) \le \gamma\alpha_y\|\nabla\theta(y)\|$ holds whether $\alpha_y\|\nabla\theta(y)\| \le 1$ or not. Therefore, by (4.9) and (4.10), we have that for all $\lambda \in [0, 1]$ and $y \in \mathcal{N}(\tilde{y})$,

$$\begin{aligned}
\theta(y + \lambda d^y_G) &= \theta(y) + \lambda\nabla\theta(y)^T d^y_G + g_y(\lambda) \\
&= \theta(y) - \lambda\alpha_y\|\nabla\theta(y)\|^2 + \lambda\nabla\theta(y)^T\beta(y)\bar{y} + g_y(\lambda) \\
&\le \theta(y) - \lambda\alpha_y\|\nabla\theta(y)\|^2 + \lambda\alpha_y\gamma\|\nabla\theta(y)\|^2\|\bar{y}\| + g_y(\lambda) \\
&= \theta(y) - \lambda\alpha_y\left(1 - \gamma\sqrt{\bar{t}^2 + \bar{s}^2}\right)\|\nabla\theta(y)\|^2 + o(\lambda).
\end{aligned} \tag{4.11}$$

Then from (4.11) we can find a positive number $\tilde{\lambda} \in (0, 1]$ such that for all $\lambda \in (0, \tilde{\lambda}]$ and $y \in \mathcal{N}(\tilde{y})$, (4.8) holds. □

We can get the following result directly from Lemma 4.1.

Proposition 4.1 *For any* $k \ge 0$, *if* $y^k \in R^2_{++} \times R^{n+p(m+q+1)}$ *and* $\nabla\Phi(y^k)$ *is nonsingular, then Algorithm 4.1 is well defined at the kth iteration.*

Proposition 4.2 *For each fixed* $k \ge 0$, *if* $(t^k, s^k) \in R^2_{++}$ *satisfies* $(t^k, s^k) \ge \beta_k(\bar{t}, \bar{s})$ *and* $\nabla\Phi(y^k)$ *is nonsingular, then we have*

$$(t^{k+1}, s^{k+1}) \ge \beta_{k+1}(\bar{t}, \bar{s}).$$

Proof By (4.3) and the fact that $\beta_{k+1} \leq \beta_k$, it is obvious that the conclusion is true when the fast step is taken. Now we prove that the conclusion also holds when the safe step is taken. It follows from (4.2) and (4.5) that

$$(d_G^k)_t \geq -t^k + \beta_k \bar{t}.$$

Consequently, we have that

$$\begin{aligned} t^{k+1} - \beta_{k+1}\bar{t} &= t^k + r^{m_k}(d_G^k)_t - \beta_{k+1}\bar{t} \\ &\geq (1 - r^{m_k})t^k + r^{m_k}\beta_k\bar{t} - \beta_k\bar{t} \\ &= (1 - r^{m_k})(t^k - \beta_k\bar{t}) \\ &\geq 0, \end{aligned}$$

where r^{m_k} is the acceptable step in Step 3 of Algorithm 4.1, the first inequality comes from the fact that $\beta_{k+1} \leq \beta_k$, the second inequality comes from the assumption that $t^k \geq \beta_k\bar{t}$. By an analogous way, we can prove that

$$s^{k+1} - \beta_{k+1}\bar{s} \geq 0.$$

Hence, we obtain the desired result and complete the proof. □

Theorem 4.1 *Suppose that for every $k \geq 0$, $\nabla\Phi(y^k)$ is nonsingular as long as $(t^k, s^k) \in R^2_{++}$ and $(t^k, s^k) \geq \beta_k(\bar{t}, \bar{s})$. Then an infinite sequence $\{y^k = (t^k, s^k, z^k)\}$ generated by Algorithm 4.1 satisfies that $(t^k, s^k) \in R^2_{++}$ and $(t^k, s^k) \geq \beta_k(\bar{t}, \bar{s})$.*

Proof First, since $y^0 = (\bar{t}, \bar{s}, z^0)$ satisfies $(\bar{t}, \bar{s}) \geq \beta_0(\bar{t}, \bar{s})$, we have from Propositions 4.1 and 4.2 that y^1 is well defined, $(t^1, s^1) \in R^2_{++}$ and $(t^1, s^1) \geq \beta_1(\bar{t}, \bar{s})$. Then, by repeatedly resorting to Propositions 4.1 and 4.2 we can prove that an infinite sequence $\{y^k\}$ is generated, $(t^k, s^k) \in R^2_{++}$ and $(t^k, s^k) \geq \beta_k(\bar{t}, \bar{s})$. The proof is complete. □

5 Convergence analysis

In this section, we prove the global and superlinear convergence of Algorithm 4.1. To this end, we first discuss the CD-regularity of Φ, which is a basic condition used frequently in convergence analysis.

Theorem 5.1 *Let $t \in R$. Then Φ is CD-regular at $y = (t, s, z)$ if $\bar{P}(t, \cdot)$ is CD-regular at z.*

Proof It is easy to see that $\bar{P}$ is regular. It then follows from Proposition 2.3.15 in [1] that

$$\partial_{(t,z)}\bar{P}(t, z) \subseteq \partial_t\bar{P}(t, z) \times \partial_z\bar{P}(t, z).$$

Consequently, by this, we can see that every element Q in $\partial\Phi(y)$ has the following form

$$Q = \begin{pmatrix} 1 & 0 & 0 \\ \zeta_1 & 1 & \zeta_2 \\ U_t & 0 & U_z \end{pmatrix}.$$

Here ζ_1 is the first component of ζ and ζ_2 is the sub-vector of ζ obtained by just removing the first component of ζ, where $\zeta \in \partial_{(t,z)}\bar{G}(t, x)$, $U_t \in \partial_t\bar{P}(t, z)$ and $U_z \in \partial_z\bar{P}(t, z)$. It is obvious that Q is nonsingular if U_z is nonsingular. We obtain the desired result and complete the proof. □

Let

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

$$l_i(z) = -u_i \nabla_v g(x, v^i) + \sum_{j=1}^{q} w_j^i \nabla c_j(v^i), \quad i \in \mathcal{P}.$$

And let

$$\nabla c(v^i) = \left(\nabla c_1(v^i), \ldots, \nabla c_q(v^i)\right), \quad i \in \mathcal{P}.$$

We make the following assumptions.

(A1) $D_x L(z)$ is positive semidefinite. Moreover, it is positive definite in the null space of $Span(\nabla_x g(x, v)^T)$. That is, $d^T D_x L(z) d > 0$ for all $d \in R^n \backslash \{0\}$ satisfying $\nabla_x g(x, v)^T d = 0$.
(A2) $D_{v^i} l_i(z)$ is positive semidefinite. Moreover, it is positive definite in the null space of $Span(\nabla c(v^i)^T)$. That is, $d^T D_{v^i} l_i(z) d > 0$ for all $d \in R^m \backslash \{0\}$ satisfying $\nabla c(v^i)^T d = 0$.

The following theorem comes from [14], which shows that Assumptions (A1) and (A2) are sufficient for $\nabla_z \bar{P}(t, z)$ to be nonsingular for every $t > 0$.

Theorem 5.2 *Let Assumptions* (A1) *and* (A2) *hold at* z. *Then* $\nabla_z \bar{P}(t, z)$ *is nonsingular for every* $t > 0$.

Remark By Theorems 5.1 and 5.2, we see that if Assumptions (A1) and (A2) hold at a considered point z, then $\nabla \Phi(y)$ is nonsingular for all $t > 0$.

Let $(0, \bar{z}) = (0, \bar{x}, \bar{u}, \bar{\mathbf{v}}, \bar{\mathbf{w}}) \in R^{1+n+p(1+m+q)}$ be a solution of $\bar{P}(t, z) = 0$. For $i \in \mathcal{P}$, let $I(\bar{v}^i) = \left\{j \in \mathcal{Q} \, : \, c_j(\bar{v}^i) = 0\right\}$. Before giving a sufficient condition for CD-regularity of $\bar{P}(0, \cdot)$ at $\bar{z}$, we also need the following assumptions.

(A3) For each $i \in \mathcal{P}$, $\bar{u}_i > 0$.
(A4) The vectors $\nabla_x g(\bar{x}, \bar{v}^i)$, $i \in \mathcal{P}$ are linearly independent.
(A5) For each $i \in \mathcal{P}$, the vectors $\nabla c_j(\bar{v}^i)$, $j \in I(\bar{v}^i)$ are linearly independent.
(A6) For all $(d^T, \xi_1^T, \ldots, \xi_p^T)^T \in S(\bar{x}, \bar{\mathbf{v}}) \backslash \{0\}$,

$$d^T D_x L(\bar{z}) d + 2 \sum_{i=1}^{p} \bar{u}_i d^T \nabla^2_{xv} g(\bar{x}, \bar{v}^i) \xi_i - \sum_{i=1}^{p} \xi_i^T D_{v^i} l_i(\bar{z}) \xi_i < 0,$$

where $S(\bar{x}, \bar{\mathbf{v}})$ be the set of all $(d^T, \xi_1^T, \ldots, \xi_p^T)^T \in \Re^n \times \Re^{mp}$ satisfying

$$\nabla_x g(\bar{x}, \bar{v}^i)^T d + \nabla_v g(\bar{x}, \bar{v}^i)^T \xi_i = 0, \quad \text{for } i = 1, 2, \ldots, p.$$

Now, we state and prove a theorem which shows that Assumptions (A3)–(A6) are sufficient for $\bar{P}(0, \cdot)$ to be CD-regular at the solution $\bar{z}$ of $\bar{P}(0, z) = 0$.

Theorem 5.3 *Suppose that* $(0, \bar{z})$ *is a solution of* $\bar{P}(t, z) = 0$ *and Assumptions* (A3)–(A6) *hold. Then* $\bar{P}(0, \cdot)$ *is CD-regular at* $\bar{z}$.

Proof It is readily to know that the elements of $\partial_z \bar{P}(0, \bar{z})$ are of the form

$$W = \begin{pmatrix} D_x L(\bar{z}) & \nabla_x g(\bar{x}, \bar{\mathbf{v}}) & F & 0 \\ \nabla_x g(\bar{x}, \bar{\mathbf{v}})^T & 0 & G & 0 \\ -F^T & -G^T & H & Q \\ 0 & 0 & U & V \end{pmatrix},$$

where

$$\begin{aligned} \nabla_x g(\bar{x}, \bar{\mathbf{v}}) &= \left(\nabla_x g(\bar{x}, \bar{v}^1), \ldots, \nabla_x g(\bar{x}, \bar{v}^p)\right), \quad F = \left(\bar{u}_1 \nabla^2_{xv} g(\bar{x}, \bar{v}^1), \ldots, \bar{u}_p \nabla^2_{xv} g(\bar{x}, \bar{v}^p)\right), \\ G &= \mathrm{diag}\left(D_v g(\bar{x}, \bar{v}^1), \ldots, D_v g(\bar{x}, \bar{v}^p)\right), \quad H = \mathrm{diag}\left(D_{v^1} l_1(\bar{z}), \ldots, D_{v^p} l_p(\bar{z})\right), \\ Q &= \mathrm{diag}\left(\nabla c(\bar{v}^1), \ldots, \nabla c(\bar{v}^p)\right) \end{aligned}$$

and

$$U = \mathrm{diag}\left(\Lambda^1 \nabla c(\bar{v}^1)^T, \ldots, \Lambda^p \nabla c(\bar{v}^p)^T\right), \quad V = \mathrm{diag}\left(\Gamma^1, \ldots, \Gamma^p\right)$$

with $\Lambda^i = \mathrm{diag}\left(a_{i1}, \ldots, a_{iq}\right)$ and $\Gamma^i = \mathrm{diag}\left(b_{i1}, \ldots, b_{iq}\right)$ satisfying

$$\begin{cases} (a_{ij}, b_{ij}) = (0, -1), & \text{if } j \in \mathcal{Q} \backslash I(\bar{v}^i), \\ (a_{ij}, b_{ij}) = (1, 0), & \text{if } j \in \{j \in I(\bar{v}^i) \ : \ \bar{w}^i_j > 0\}, \\ (a_{ij}, b_{ij}) \in \left\{(a, b) \in R^2 \ : \ (a-1)^2 + (b+1)^2 \le 1\right\}, & \text{if } j \in \{j \in I(\bar{v}^i) \ : \ \bar{w}^i_j = 0\}. \end{cases} \tag{5.12}$$

It is only necessary to prove that W is nonsingular under the given assumptions. By (5.12), it is clear that W is nonsingular if and only if

$$\bar{W} = \begin{pmatrix} \nabla_x L(\bar{z}) & \nabla_x g(\bar{x}, \bar{\mathbf{v}}) & F & 0 \\ \nabla_x g(\bar{x}, \bar{\mathbf{v}})^T & 0 & G & 0 \\ -F^T & -G^T & H & \bar{Q} \\ 0 & 0 & \bar{U} & \bar{V} \end{pmatrix}$$

is nonsingular, where

$$\bar{Q} = \mathrm{diag}\left(\nabla c(\bar{v}^1)_{\cdot I(\bar{v}^1)}, \ldots, \nabla c(\bar{v}^p)_{\cdot I(\bar{v}^p)}\right), \ \bar{V} = \mathrm{diag}\left(\Gamma^1_{I(\bar{v}^1) I(\bar{v}^1)}, \ldots, \Gamma^p_{I(\bar{v}^p) I(\bar{v}^p)}\right)$$

and

$$\bar{U} = \mathrm{diag}\left(\Lambda^1_{I(\bar{v}^1) I(\bar{v}^1)} (\nabla c(\bar{v}^1)_{\cdot I(\bar{v}^1)})^T, \ldots, \Lambda^p_{I(\bar{v}^p) I(\bar{v}^p)} (\nabla c(\bar{v}^p)_{\cdot I(\bar{v}^p)})^T\right).$$

Moreover, it is easy to see that for $i \in \mathcal{P}$ and $j \in \{j \in I(\bar{v}^i) \ : \ \bar{w}^i_j = 0\}$, $b_{ij} = -1$ provided $a_{ij} = 0$. In this case we delete the row and column which includes $b_{ij} = -1$, the obtained matrix has the same nonsingularity as $\bar{W}$. Without loss of generality, we assume that $a_{ij} > 0$ for $i \in \mathcal{P}$ and $j \in \{j \in I(\bar{v}^i) \ : \ \bar{w}^i_j = 0\}$. It is clear that $a_{ij}^{-1} b_{ij} \le 0$ for $i \in \mathcal{P}$ and $j \in I(\bar{v}^i)$.

Suppose that

$$\bar{W}\begin{pmatrix} d_1 \\ d_2 \\ \xi_1 \\ \vdots \\ \xi_p \\ \zeta_1 \\ \vdots \\ \zeta_p \end{pmatrix} = 0, \tag{5.13}$$

where $d_1 \in R^n$, $d_2 \in R^p$, $\xi_i \in R^m$ and $\zeta_i \in R^{|I(\bar{v}^i)|}$ $(i \in \mathcal{P})$. Then (5.13) implies

$$D_x L(\bar{z})d_1 + \nabla_x g(\bar{x}, \bar{\mathbf{v}})d_2 + \sum_{i=1}^{p} \bar{u}_i \nabla^2_{xv} g(\bar{x}, \bar{v}^i)\xi_i = 0, \tag{5.14}$$

$$\nabla_x g(\bar{x}, \bar{v}^i)^T d_1 + \nabla_v g(\bar{x}, \bar{v}^i)^T \xi_i = 0, \quad i \in \mathcal{P}, \tag{5.15}$$

$$-\bar{u}_i \nabla^2_{vx} g(\bar{x}, \bar{v}^i)d_1 - \nabla_v g(\bar{x}, \bar{v}^i)d_{2i} + D_{v^i} l_i(\bar{z})\xi_i + \nabla c(\bar{v}^i)_{.I(\bar{v}^i)}\zeta_i = 0, \quad i \in \mathcal{P}, \tag{5.16}$$

$$\nabla c_j(\bar{v}^i)^T \xi_i + a_{ij}^{-1} b_{ij}\zeta_{ij} = 0, \quad i \in \mathcal{P},\ j \in I(\bar{v}^i), \tag{5.17}$$

where d_{2i} and ζ_{ij} are the ith component of d_2 and the jth component of ζ_i, respectively. By (5.14), it follows that

$$d_1^T D_x L(\bar{z})d_1 + d_1^T \nabla_x g(\bar{x}, \bar{\mathbf{v}})d_2 + \sum_{i=1}^{p} \bar{u}_i d_1^T \nabla^2_{xv} g(\bar{x}, \bar{v}^i)\xi_i = 0. \tag{5.18}$$

By (5.15) and (5.16), we know that for every $i \in \mathcal{P}$,

$$d_{2i} \nabla_x g(\bar{x}, \bar{v}^i)^T d_1 - \bar{u}_i \xi_i^T \nabla^2_{vx} g(\bar{x}, \bar{v}^i)d_1 + \xi_i^T D_{v^i} l_i(\bar{z})\xi_i + \xi_i^T \nabla c(\bar{v}^i)_{.I(\bar{v}^i)}\zeta_i = 0,$$

which implies

$$d_1^T \nabla_x g(\bar{x}, \bar{\mathbf{v}})d_2 - \sum_{i=1}^{p} \bar{u}_i \xi_i^T \nabla^2_{vx} g(\bar{x}, \bar{v}^i)d_1 + \sum_{i=1}^{p} \xi_i^T D_{v^i} l_i(\bar{z})\xi_i + \sum_{i=1}^{p} \xi_i^T \nabla c(\bar{v}^i)_{.I(\bar{v}^i)}\zeta_i = 0. \tag{5.19}$$

By (5.17), we obtain that

$$\sum_{i=1}^{p} \xi_i^T \nabla c(\bar{v}^i)_{.I(\bar{v}^i)}\zeta_i + \sum_{i=1}^{p} \sum_{j \in I(\bar{v}^i)} a_{ij}^{-1} b_{ij}\zeta_{ij}^2 = 0. \tag{5.20}$$

By (5.18) and (5.19), it holds that

$$d_1^T D_x L(\bar{z})d_1 + 2\sum_{i=1}^{p} \bar{u}_i \xi_i^T \nabla^2_{vx} g(\bar{x}, \bar{v}^i)d_1 - \sum_{i=1}^{p} \xi_i^T D_{v^i} l_i(\bar{z})\xi_i - \sum_{i=1}^{p} \xi_i^T \nabla c(\bar{v}^i)_{.I(\bar{v}^i)}\zeta_i = 0,$$

which implies, together with (5.20), that

$$d_1^T D_x L(\bar{z})d_1 + 2\sum_{i=1}^{p} \bar{u}_i \xi_i^T \nabla^2_{vx} g(\bar{x}, \bar{v}^i)d_1 - \sum_{i=1}^{p} \xi_i^T D_{v^i} l_i(\bar{z})\xi_i = -\sum_{i=1}^{p} \sum_{j \in I(\bar{v}^i)} a_{ij}^{-1} b_{ij}\zeta_{ij}^2 \geq 0 \tag{5.21}$$

where the inequality comes from the fact that $a_{ij}^{-1}b_{ij} \leq 0$ for $i \in \mathcal{P}$ and $j \in I(\bar{v}^i)$. Since $(d^T, \xi_1^T, \ldots, \xi_p^T)^T$ satisfies (5.15), by (5.21) and (A6), we know that $d_1 = 0$ and $\xi_i = 0$ for $i \in \mathcal{P}$. Consequently, by (A3), it follows that $d_2 = 0$ from (5.14). Moreover, by (A5) and (5.16), it follows that $\zeta_i = 0$ for $i \in \mathcal{P}$. We obtain the desired result and complete the proof. □

Under the strict complementarity assumption for the lower level problem, a sufficient condition for CD-regularity is given in [30]. In the theorem above, we do not assume strict complementary slackness in the lower level problem, but the upper level strict complementarity condition is need. In order to obtain another condition for CD-regularity of $\bar{\Phi}(0, \cdot)$, in which the strict complementarity conditions in both the upper level and the lower level problems are not assumed, we further need the following assumptions.

(A4′) The vectors $\nabla_x g(\bar{x}, \bar{v}^i)$, $i \in \mathcal{P}(\bar{z})$ are linearly independent, where $\mathcal{P}(\bar{z}) = \{i \in \mathcal{P} : g(\bar{x}, \bar{v}^i) = 0\}$.

(A7) For any $(d_1^T, d_2^T, \xi_1^T, \ldots, \xi_p^T)^T \in \Lambda(\bar{z})$,

$$d_1^T D_x L(\bar{z}) d_1 - 2 \sum_{i \in \mathcal{P}(\bar{z})} \xi_i^T \nabla_v g(\bar{x}, \bar{v}^i) d_{2i} + \sum_{i=1}^{p} \xi_i^T D_{v^i} l_i(\bar{z}) \xi_i > 0, \quad (5.22)$$

where $\Lambda(\bar{z})$ is the set of all $(d_1^T, d_2^T, \xi_1^T, \ldots, \xi_p^T)^T \in R^{n+|\mathcal{P}(\bar{z})|+mp}$ satisfying

$$D_x L(\bar{z}) d_1 + (\nabla_x g(\bar{x}, \bar{\mathbf{v}})_{\cdot \mathcal{P}(\bar{z})}) d_2 + \sum_{i=1}^{p} \bar{u}_i \nabla_{xv}^2 g(\bar{x}, \bar{v}^i) \xi_i = 0$$

and $(d_1^T, \xi_1^T, \ldots, \xi_p^T)^T \neq 0$.

Theorem 5.4 *Suppose that* $(0, \bar{z})$ *is a solution of* $\bar{P}(t, z) = 0$ *and Assumptions* (A4′), (A5) and (A7) *hold. Then* $\bar{P}(0, \cdot)$ *is CD-regular at* $\bar{z}$.

Proof Since we do not assume strict complementary slackness in the upper level problem, every element W of $\partial_z \bar{P}(0, \bar{z})$ is of the form

$$W = \begin{pmatrix} D_x L(\bar{z}) & \nabla_x g(\bar{x}, \bar{\mathbf{v}}) & F & 0 \\ \Lambda \nabla_x g(\bar{x}, \bar{\mathbf{v}})^T & \Gamma & \Lambda G & 0 \\ -F^T & -G^T & H & Q \\ 0 & 0 & U & V \end{pmatrix},$$

where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$ and $\Gamma = \mathrm{diag}(\gamma_1, \ldots, \gamma_p)$ satisfying

$$\begin{cases} (\lambda_i, \gamma_i) = (0, -1), & \text{if } i \in \mathcal{P} \backslash \mathcal{P}(\bar{z}), \\ (\lambda_i, \gamma_i) = (1, 0), & \text{if } i \in \{i \in \mathcal{P}(\bar{z}) : \bar{u}_i > 0\}, \\ (\lambda_i, \gamma_i) \in \left\{(a, b) \in R^2 : (a-1)^2 + (b+1)^2 \leq 1\right\}, & \text{if } i \in \{i \in \mathcal{P}(\bar{z}) : \bar{u}_i = 0\}, \end{cases} \quad (5.23)$$

and other notation are the same as those used in the proof of Theorem 5.3. By (5.23), we know that for $i \in \{i \in \mathcal{P}(\bar{z}) : \bar{u}_i = 0\}$, $\gamma_i = -1$ provided $\lambda_i = 0$. In this case we delete the row and column which includes γ_i, the obtained matrix has the same nonsingularity as W. The similar conclusion also holds for $i \in \mathcal{P}$ and $j \in \{j \in I(\bar{v}^i) : \bar{w}_j^i = 0\}$. Hence, without loss of generality, we assume that $\lambda_i > 0$ ($i \in \{i \in \mathcal{P}(\bar{z}) : \bar{u}_i = 0\}$) and $a_{ij} > 0$ ($i \in \mathcal{P}$, $j \in \{j \in I(\bar{v}^i) : \bar{w}^i_j = 0\}$). It is easy to see that W is nonsingular if and only if

$$\tilde{W} = \begin{pmatrix} D_x L(\bar{z}) & \nabla_x g(\bar{x}, \bar{\mathbf{v}})_{\cdot\mathcal{P}(\bar{z})} & F & 0 \\ (\nabla_x g(\bar{x}, \bar{\mathbf{v}})_{\cdot\mathcal{P}(\bar{z})})^T & \Lambda^{-1}_{\mathcal{P}(\bar{z})}\Gamma_{\mathcal{P}(\bar{z})} & G_{\mathcal{P}(\bar{z})\cdot} & 0 \\ -F^T & -(G_{\mathcal{P}(\bar{z})\cdot})^T & H & \bar{Q} \\ 0 & 0 & \bar{U} & \bar{V} \end{pmatrix}$$

is nonsingular, where $\Lambda_{\mathcal{P}(\bar{z})} = \Lambda_{\mathcal{P}(\bar{z})\mathcal{P}(\bar{z})}$, $\Gamma_{\mathcal{P}(\bar{z})} = \Gamma_{\mathcal{P}(\bar{z})\mathcal{P}(\bar{z})}$, and $\bar{Q}$, $\bar{U}$ and $\bar{V}$ are the same as those used in Theorem 5.3.

Suppose that

$$\tilde{W} \begin{pmatrix} d_1 \\ d_2 \\ \xi_1 \\ \vdots \\ \xi_p \\ \zeta_1 \\ \vdots \\ \zeta_p \end{pmatrix} = 0, \tag{5.24}$$

where $d_1 \in R^n$, $d_2 \in R^{|\mathcal{P}(\bar{z})|}$, $\xi_i \in R^m$ and $\zeta_i \in R^{|I(\bar{v}^i)|}$ ($i \in \mathcal{P}$). Then (5.24) implies

$$D_x L(\bar{z})d_1 + \nabla_x g(\bar{x}, \bar{\mathbf{v}})_{\cdot\mathcal{P}(\bar{z})} d_2 + \sum_{i=1}^{p} \bar{u}_i \nabla^2_{xv} g(\bar{x}, \bar{v}^i)\xi_i = 0, \tag{5.25}$$

$$\nabla_x g(\bar{x}, \bar{v}^i)^T d_1 + \lambda_i^{-1}\gamma_i d_{2i} + \nabla_v g(\bar{x}, \bar{v}^i)^T \xi_i = 0, \quad i \in \mathcal{P}(\bar{z}), \tag{5.26}$$

$$-\bar{u}_i \nabla^2_{vx} g(\bar{x}, \bar{v}^i)d_1 - \nabla_v g(\bar{x}, \bar{v}^i)d_{2i} + D_{v^i} l_i(\bar{z})\xi_i + \nabla c(\bar{v}^i)_{\cdot I(\bar{v}^i)}\zeta_i = 0, \quad i \in \mathcal{P}(\bar{z}), \tag{5.27}$$

$$-\bar{u}_i \nabla^2_{vx} g(\bar{x}, \bar{v}^i)d_1 + D_{v^i} l_i(\bar{z})\xi_i + \nabla c(\bar{v}^i)_{\cdot I(\bar{v}^i)}\zeta_i = 0, \quad i \in \mathcal{P}\backslash\mathcal{P}(\bar{z}), \tag{5.28}$$

$$\nabla c_j(\bar{v}^i)^T \xi_i + a_{ij}^{-1} b_{ij}\zeta_{ij} = 0, \quad i \in \mathcal{P},\ j \in I(\bar{v}^i). \tag{5.29}$$

By (5.25)–(5.29), we have

$$\begin{aligned} & d_1^T D_x L(\bar{z})d_1 - 2\sum_{i\in\mathcal{P}(\bar{z})} \xi_i^T \nabla_v g(\bar{x}, \bar{v}^i)d_{2i} + \sum_{i=1}^{p} \xi_i^T D_{v^i} l_i(\bar{z})\xi_i \\ & -\sum_{i=1}^{p}\sum_{j\in I(\bar{v}^i)} a_{ij}^{-1} b_{ij}\zeta_{ij}^2 - \sum_{i\in\mathcal{P}(\bar{z})} \lambda_i^{-1}\gamma_i d_{2i}^2 = 0, \end{aligned} \tag{5.30}$$

which, together with the fact that $a_{ij}^{-1} b_{ij} \leq 0$ for $i \in \mathcal{P}$, $j \in I(\bar{v}^i)$ and $\lambda_i^{-1}\gamma_i \leq 0$ for $i \in \mathcal{P}(\bar{z})$, implies

$$d_1^T D_x L(\bar{z})d_1 - 2\sum_{i\in\mathcal{P}(\bar{z})} \xi_i^T \nabla_v g(\bar{x}, \bar{v}^i)d_{2i} + \sum_{i=1}^{p} \xi_i^T D_{v^i} l_i(\bar{z})\xi_i \leq 0. \tag{5.31}$$

By (5.25), (5.31) and (A7), we know that $d_1 = 0$ and $\xi_i = 0$ for $i \in \mathcal{P}$. Consequently, by (5.25), we obtain

$$\nabla_x g(\bar{x}, \bar{\mathbf{v}})_{\cdot\mathcal{P}(\bar{z})} d_2 = 0,$$

which implies

$$\sum_{i\in\mathcal{P}(\bar z)} d_{2i}\nabla_x g(\bar x, \bar v^i) = 0.$$

By (A4′), $d_2 = 0$. Moreover, by (5.27) and (5.28),

$$\sum_{j\in I(\bar v^i)} \nabla c_j(\bar v^i)\zeta_{ij} = 0, \quad i \in \mathcal{P}.$$

By (A5), $\zeta_i = 0,\ i \in \mathcal{P}$. Therefore, $(d_1^T, d_2^T, \xi_1^T, \ldots, \xi_p^T, \zeta_1^T, \ldots, \zeta_p^T)^T = 0$, which shows that $\tilde W$ is nonsingular. We obtain the desired result and complete the proof. □

Remark (a) By Theorems 5.1, 5.3 and 5.4, we see that if Assumptions (A3)–(A6), or Assumptions (A4′), (A5) and (A7) hold at the solution $(0, 0, \bar z)$ of $\Phi(y) = 0$, then Φ is CD-regular at $(0, 0, \bar z)$. This shows that Algorithm 4.1 can actually handle the SIP problem with the violated strict complementarity in the lower level problem (or in both the upper level and the lower level problems). (b) For a solution $\bar y$ of $\Phi(y) = 0$, if $\bar u_i = 0$ and $g(\bar x, \bar v^i) = 0$ for some index i, then we get from (2.3) that

$$\sum_{j=1}^{q} \bar w_j^i \nabla c_j(\bar v^i) = 0 \quad \text{and} \quad \bar w_j^i \ge 0,\ c_j(\bar v^i) \le 0,\ \bar w_j^i\, c_j(\bar v^i) = 0 \quad \text{for } j \in \mathcal{Q}.$$

Hence, $\nabla c_j(\bar v^i),\ j \in I(\bar v^i)$ are linearly dependent, provided at least one $\bar w_j^i \ne 0$. It means that the possible violation of the strict complementarity in the upper level is shifted to the explicit violation of LICQ in the lower level. At same time, the optimality criteria for the lower level becomes a Fritz–John-type condition. In this paper, we focus attention on the case where $\bar u \ne 0$. Thus, we may obtain a solution of (1.6) by dropping the part indexed by i with $\bar u_i = 0$. (c) The Lagrange multiplier $\bar u_i$ corresponding to the objective function $g(x, \cdot)$ in the lower level is the Lagrange multiplier corresponding to the constraint $g(\cdot, v^i)$ in the upper level. Assumption (A7) involves the first- and the second-order terms not only of the lower, but also of the upper level. It is different from the second order optimality conditions for SIP problems given in [28]. As future work, we will work on how to find a better condition under which the considered $\bar P(0, \cdot)$ is CD-regular.

Theorem 5.5 *Let $\{y^k\}$ be the sequences generated by Algorithm 4.1 and y^* be a limit point of a subsequence $\{y^k\}_{k\in K}$. Suppose that for every $k \ge 0$, $\nabla\Phi(y^k)$ is nonsingular as long as $(t^k, s^k) \in R^2_{++}$ and $(t^k, s^k) \ge \beta_k(\bar t, \bar s)$, $\bar P(t^*, \cdot)$ is CD-regular at z^* and $\{y^k\}_{k\in K}$ satisfies*

$$\liminf_{k\in K,\ k\to\infty} \frac{t^k}{|\nabla_t \theta(y^k)|} > 0 \tag{5.32}$$

and

$$\liminf_{k\in K,\ k\to\infty} \frac{s^k}{|s^k + \bar G(t^k, x^k)|} > 0. \tag{5.33}$$

Then y^* is a solution of $\Phi(y) = 0$.

Proof It follows from Proposition 4.2 that an infinite sequence $\{y^k\}$ is generated such that $(t^k, s^k) \ge \beta_k(\bar t, \bar s)$ for all $k \ge 0$. From the design of Algorithm 4.1, $\theta\left(y^{k+1}\right) < \theta\left(y^k\right)$ for all $k \ge 0$. Hence the two sequences $\left\{\theta(y^k)\right\}$ and $\{\beta_k\}$ are monotonically decreasing. Since $\theta(y^k)$, $\beta_k \ge 0$ $(k \ge 0)$, there exist θ^*, $\beta^* \ge 0$ such that $\theta(y^k) \to \theta^*$ and $\beta_k \to \beta^*$ as

$k \to \infty$. If $\theta^* = 0$, then from the continuity of $\theta(\cdot)$ and $\beta(\cdot)$ we have $\theta(y^*) = \theta^* = 0$ and obtain the desired result. Suppose that $\theta^* > 0$. This implies that we eventually take only safe steps in Algorithm 4.1 since otherwise we would have

$$\theta(y^{k+1}) \le \sigma\theta(y^k)$$

for infinitely many k which imply $\theta^* = 0$. Therefore, we can assume without loss of generality that all steps are safe steps.

Since $\bar{P}(t^*, \cdot)$ is CD-regular at z^*, by Theorem 5.1, Φ is CD-regular at y^*. Furthermore, by (5.32) and (5.33), it is easy to see that

$$\liminf_{k\in K,\ k\to\infty} \alpha_k > 0,$$

which implies that $\beta^* > 0$ and $(t^*, s^*) \ge \beta^*(\bar{t}, \bar{s})$, we see that $(t^*, s^*) \in R^2_{++}$. Then $\nabla\Phi(y^*)$ exists and is nonsingular from the CD-regularity of $\bar{P}(t^*, \cdot)$ at z^*. Hence, from Lemma 4.1 there exists a closed neighborhood $\mathcal{N}(y^*)$ of y^* and a positive number $\tilde{\lambda} \in (0, 1]$ such that for any $y = (t, s, z) \in \mathcal{N}(y^*)$ and all $\lambda \in (0, \tilde{\lambda}]$ we have $(t, s) \in R^2_{++}$, $\nabla\Phi(y)$ is invertible and (4.8) holds. Therefore, for a nonnegative integer l such that $r^l \in (0, \tilde{\lambda}]$, we have

$$\theta(y^k + r^l d_G^k) \le \theta(y^k) - \sigma\alpha_k \left(1 - \gamma\sqrt{\bar{t}^2 + \bar{s}^2}\right) r^l \|\nabla\theta(y^k)\|^2.$$

for all sufficiently large k. This contradicts the fact that the sequence $\{\theta(y^k)\}$ converges to $\theta^* > 0$ because $\alpha_k\|\nabla\theta(y^k)\|^2 \ge \alpha_k^2\|\nabla\theta(y^k)\|^2 > \beta^*/\gamma > 0$. So, we complete our proof. □

Remark Let $\{y^k\}$ be the sequences generated by Algorithm 4.1 and y^* be a limit point of a subsequence $\{y^k\}_{k\in K}$. Suppose that the assumed conditions in Theorem 5.5 hold. Then $G(x^*) = 0$ since $s^* \ge 0$ and $\Phi(y^*) = 0$ by Theorem 5.5, which implies that x^* is a feasible solution of (1.1). Consequently, from the CD-regularity of Φ at y^*, we have

$$\|x^k - x^*\| \le \|y^k - y^*\| = O(\|\Phi(y^k) - \Phi(y^*)\|) = O(\varepsilon)$$

provided $\|\Phi(y^k)\| \le \varepsilon$, which shows that x^k is an approximate feasible solution of (1.1).

In the following proposition, we present a condition under which (5.32) holds.

Proposition 5.1 *Let $\{y^k\}$ be a sequence generated by Algorithm 4.1 and $y^* = \lim_{k\in K}\{y^k\}$ for some subset $K \subset \{1, 2, \ldots\}$. Suppose that y^* satisfies*

$$\int_V \frac{1}{|g(x^*, v)|} dv < \infty. \tag{5.34}$$

Then we have

$$\lim_{k\in K} \frac{t^k}{|\nabla_t\theta(y^k)|} > 0.$$

Proof By direct computation, we have that for any $a, b \in R$, $x \in R^n$ and $t > 0$,

$$\begin{aligned} |\nabla_t\bar{\phi}(t, a, b)\bar{\phi}(t, a, b)| &= \frac{t}{\sqrt{a^2 + b^2 + t^2}}(\sqrt{a^2 + b^2 + t^2} - a - b) \\ &\le \frac{t}{\sqrt{a^2 + b^2 + t^2}}(\sqrt{a^2 + b^2 + t^2} + |a| + |b|) \\ &\le 3t \end{aligned} \tag{5.35}$$

and

$$\nabla_t \bar{G}(t, x) = 2t \int_V \frac{1}{\sqrt{(g(x, v))^2 + 4t^2}} dv. \tag{5.36}$$

Consequently,

$$\begin{aligned}
|\nabla_t \theta(y^k)| &\leq t^k + \sum_{i=1}^{p} |\nabla_t \bar{\phi}(t^k, u_1^k, -g(x^k, v^{1k}))\bar{\phi}(t^k, u_1^k, -g(x^k, v^{1k}))| \\
&\quad + \sum_{i=1}^{p} \sum_{j=1}^{q} |\nabla_t \bar{\phi}(t^k, w_j^{ik}, -c_j(v^{ik}))\bar{\phi}(t^k, w_j^{ik}, -c_j(v^{ik}))| \\
&\quad + \nabla_t \bar{G}(t^k, x^k)(\bar{G}(t^k, x^k) + s^k) \\
&\leq t^k + 3p(q+1)t^k + 2t^k(\bar{G}(t^k, x^k) + s^k) \\
&\quad \times \int_V \frac{1}{\sqrt{(g(x^k, v))^2 + 4(t^k)^2}} dv.
\end{aligned} \tag{5.37}$$

Hence, we have

$$\lim_{k \in K} \frac{t^k}{|\nabla_t \theta(y^k)|} \geq \frac{1}{1 + 3p(q+1) + 2(G(x^*) + s^* + 1) \int_V \frac{1}{|(g(x^*, v))|} dv} > 0.$$

We obtain the desired result and complete the proof. □

In the rest of this section, we investigate the local convergence rate of Algorithm 4.1. We make the following standard assumption:

(B1) Let $y^* = (t^*, s^*, z^*) = (0, 0, z^*)$ be an accumulation point of the sequence $\{y^k\}$ generated by Algorithm 4.1. Suppose $\lim_{k \in K} y^k = y^*$ for some subset $K \subset \{1, 2, \ldots\}$, y^* is a solution of the system of equations (3.8) and $\bar{P}(0, \cdot)$ is CD-regular at z^*.

We need the following proposition which has already been shown by Moré and Sorensen [17].

Proposition 5.2 *Assume that $w^* \in R^l$ is an isolated accumulation point of a sequence $\{w^k\} \subseteq R^l$ such that, for every subsequence $\{w^k\}_K$ converging to w^*; there is an infinite subset $\tilde{K} \subseteq K$ such that $\{\|w^{k+1} - w^k\|\}_{\tilde{K}} \to 0$. Then the whole sequence w^k converges to w^*.*

The original version of the following result is due to Facchinei and Soares [3]; here we cite a slight different version from Kanzow and Qi [12].

Proposition 5.3 *Let $G : R^l \to R^l$ be locally Lipschitz continuous, $w^* \in R^l$ with $G(w^*) = 0$ such that all elements in $\partial G(w^*)$ are nonsingular, and assume that there are two sequences $\{w^k\} \subseteq R^l$ and $\{d^k\} \subseteq R^l$ with $\{w^k\} \to w^*$ and $\|w^k + d^k - w^*\| = o(\|w^k - w^*\|)$. Then $\|G(w^k + d^k)\| = o(\|G(w^k)\|)$.*

Theorem 5.6 *Suppose that $\{y^k\}$ is a sequence generated by Algorithm 4.1 and y^* is a point satisfying* (B1). *Then the whole sequence $\{y^k\}$ converges to y^*, and*

$$\|y^{k+1} - y^*\| = o(\|y^k - y^*\|). \tag{5.38}$$

Proof First, since $\bar{P}(0,\cdot)$ is CD-regular at z^*, by Theorem 5.1 it follows that Φ is CD-regular at y^*, and hence y^* is a locally isolated solution of (3.8). Since $\{\theta(y^k)\}$ decreases monotonically and $y^k \to y^*$ as $k \in K,\ k \to \infty$, we have that $\{\theta(y^k)\} \to \theta(y^*) = 0$ on the whole sequence. Consequently, every accumulation point of $\{y^k\}$ is a solution of (3.8), and hence y^* is an isolated accumulation point of $\{y^k\}$. Now let $\{y^k\}_{\bar{K}}$ be a subsequence converging to y^*. Then

$$\{\|\Phi(y^k)\|\}_{\bar{K}} \to \|\Phi(y^*)\| = 0.$$

It is easy to see that $|(d_N^k)_t| = O(\|\Phi(y^k)\|)$ and $|(d_N^k)_s| = O(\|\Phi(y^k)\|)$ by (4.3) and (4.1). Consequently, since $\bar{P}(0,\cdot)$ is CD-regular at z^*, by (4.4) and Lemma 2.1, $\|(d_N^k)_z\| = O(\|\Phi(y^k)\|)$. Therefore, we have

$$\|d_N^k\| = O(\|\Phi(y^k)\|). \tag{5.39}$$

On the other hand, by (4.5), it is clear that

$$\|d_G^k\| = O(\|\Phi(y^k)\|). \tag{5.40}$$

Based on (5.39) and (5.40), it is not difficult to prove that there exists an infinite subsequence $\tilde{K}$ of $\bar{K}$ such that

$$\left\{\|y^{k+1} - y^k\|\right\}_{\tilde{K}} \to 0.$$

By Proposition 5.2, we know that the whole sequence $\{y^k\}$ converges to y^*.

Now we prove that (5.38) holds. Let

$$\Psi(t,z) = \begin{pmatrix} t \\ \bar{P}(t,z) \end{pmatrix}.$$

Then, from Lemma 2.1, for all (t,z) sufficiently close to (t^*, z^*),

$$\left\|\nabla\Psi(t,z)^{-1}\right\| = O(1).$$

Hence, from the special structure of $\nabla\Phi(y)$, Definition 2.1 and Lemma 2.1, for (t^k, z^k) sufficiently close to $(0, z^*)$, we have

$$\begin{aligned}
&\left\|\left(t^k, z^k\right) + \left((d_N^k)_t, (d_N^k)_z\right) - \left(0, z^*\right)\right\| \\
&= \left\|\left(t^k, z^k\right) + (\nabla\Psi(t^k, z^k)^T)^{-1}\left[-\Psi(t^k, z^k) + \beta_k\left(\bar{t}, 0\right)\right] - \left(0, z^*\right)\right\| \\
&= O\left(\left\|\Psi(t^k, z^k) - \Psi(0, z^*) - \nabla\Psi(t^k, z^k)^T\left(\left(t^k, z^k\right) - \left(0, z^*\right)\right)\right\|\right) + O(\beta_k \bar{t}) \\
&= o\left(\left\|\left(\left(t^k, z^k\right) - \left(0, z^*\right)\right)\right\|\right) + O\left(\theta(y^k)\right).
\end{aligned} \tag{5.41}$$

Noticing that Φ is locally Lipschitz continuous at $(0, 0, z^*)$, we know that for all y^k sufficiently close to y^*,

$$\theta(y^k) = \frac{1}{2}\|\Phi(y^k)\|^2 = O(\|y^k - y^*\|^2), \tag{5.42}$$

which implies, together with the second expression in (4.3), that

$$|s^k + (d_N^k)_s - s^*| = \beta_k \bar{s} = O(\|y^k - y^*\|^2). \tag{5.43}$$

Therefore, we know, by combining (5.41), (5.42) and (5.43), that for all y^k sufficiently close to y^*,

$$\|y^k + d_N^k - y^*\| = o(\|y^k - y^*\|). \tag{5.44}$$

Consequently, by Proposition 5.3, we have that eventually only fast steps will be taken in Step 3 of Algorithm 4.1 and

$$\|y^{k+1} - y^*\| = o(\|y^k - y^*\|).$$

We obtain the desired result and complete the proof. □

The next corollary follows from Theorems 5.5 and 5.6.

Corollary 5.1 *Let $\{y^k\}$ be the sequences generated by Algorithm 4.1 and y^* be a limit point of a subsequence $\{y^k\}_{k\in K}$. Suppose that the assumed conditions in Theorem 5.5 hold. Then the whole sequence $\{y^k\}$ converges to y^*, and*

$$\|y^{k+1} - y^*\| = o(\|y^k - y^*\|). \tag{5.45}$$

6 Preliminary numerical results

In this section, we report our preliminary numerical test results. We implemented Algorithm 4.1 in both MATLAB and FORTRAN 77. We tested 14 problems which we call Problems 1–14. Problems 1–3 are from [38]. Problem 4 comes from [35] with a revised region. Problem 5 is first presented in this paper, and Problem 6 is from [2]. Problems 7–12 are some examples in which the dimension of the parameter v is 2. While Problems 13–14 are two examples with more higher dimension decision variable, which come from [39] and [11], respectively.

Problem 1

$$f(x) = 1.21\exp(x_1) + \exp(x_2), \quad g(x, v) = v - \exp(x_1 + x_2),$$
$$V = [0, 1], \ \ p = 1, (x^0, v^0) = (2, -2, 1).$$

Problem 2

$$f(x) = x_1^2 + x_2^2 + x_3^2, \ \ g(x, v) = x_1 + x_2\exp(x_3 v) + \exp(2v) - 2\sin(4v),$$
$$V = [0, 1], \ \ p = 1, (x^0, v^0) = (-2, 0, 4, 1).$$

Problem 3

$$f(x) = \tfrac{1}{3}x_1^2 + \tfrac{1}{2}x_1 + x_2^2, \ \ g(x, v) = (1 - x_1^2 v^2)^2 - x_1 v^2 - x_2^2 + x_2,$$
$$V = [0, 1], \quad p = 1, (x^0, v^0) = (-4, -1, 1).$$

Problem 4

$$f(x) = x_1^2 + (x_2 - 3)^2, \ \ g(x, v) = x_2 - 2 + x_1\sin(v/(x_2 - 0.5)),$$
$$V = [0, 3], \ \ p = 1, (x^0, v^0) = (1, 6, 1).$$

Problem 5

$$f(x) = 2x_1^2 + 2x_1x_3 + 4x_2^2 + x_3^2,$$
$$g(x, v) = x_1 + x_1^2\sin(2v) + 3x_1x_2 + x_2^2\cos(3v) + x_3^2 - v,$$
$$V = [0, 3\pi], \ \ p = 1, (x^0, v^0) = (2, 3, 4, 1).$$

Problem 6

$$f(x) = (x_1 - 2x_2 + 5x_2^2 - x_2^3 - 13)^2 + (x_1 - 14x_2 + x_2^2 + x_2^3 - 29)^2,$$
$$g(x, v) = x_1^2 + 2x_2 v + \exp(x_1 + x_2) - \exp(v),$$
$$V = [0, 1],\ \ p = 1,\ (x^0, v^0) = (1, -1, 1).$$

Problem 7

$$f(x) = \tfrac{1}{3}x_1^2 + \tfrac{1}{2}x_1 + x_2^2,$$
$$g(x, v) = (1 - x_1^2 v_1^2)^2 - x_1 v_2^2 - x_2^2 + x_2,$$
$$V = [0, 2] \times [0, 1],\quad p = 2,\quad (x^0, v^0) = (-1, -1, 0, 0, 0, 1).$$

Problem 8

$$f(x) = (x_1 - 2)^2 + x_2^2,\quad g(x, v) = x_1^2\cos(v_1) + x_2\sin(v_2) - 4,$$
$$V = [0, \pi] \times [0, \pi],\quad p = 1,\quad (x^0, v^0) = (-1, -1, 1, 0).$$

Problem 9

$$f(x) = x_1^2 + x_2^2 + x_3^3,$$
$$g(x, v) = x_1(v_1 + v_2^2 + 1) + x_2(v_1 v_2 - v_2^2) + x_3(v_1 v_2 + v_2^2 + v_2) + 1,$$
$$V = [0, 1] \times [0, 1],\quad p = 1,\quad (x^0, v^0) = (1, 1, 1, 1, 1).$$

Problem 10

$$f(x) = x_1^2 + x_2^2 + x_3^2,$$
$$g(x, v) = x_1 + x_2\exp(x_3 v_1) - \exp(2x_1 v_2) + \sin(4v_1),$$
$$V = [0, 1] \times [0, 1],\quad p = 2,\quad (x^0, v^0) = (1, 1, 1, 1, 1, 0, 1).$$

Problem 11

$$f(x) = (x_1 - 3)^2 + x_2^2 - x_2,$$
$$g(x, v) = x_1^2 v_1\cos(v_1 v_2) + (x_2 - 1)v_1^2\sin(v_2 x_1 - \tfrac{13}{9}\pi) - 4v_2 + x_1,$$
$$V = [0, 2] \times [1, 2],\quad p = 1,\quad (x_0, v_0) = (1, 1, 0, 0).$$

Problem 12

$$f(x) = \tfrac{1}{2}(x_1^2 + x_2^2 + x_3^2 + x_4^2),$$
$$g(x, v) = \sin(v_1 v_2) - x_1 - x_2 v_1 - x_3 v_2 - x_4 v_1 v_2,$$
$$V = [0, 1] \times [0, 1],\quad p = 1,\quad (x^0, v^0) = (2, 2, 2, 2, 1, 0).$$

Problem 13

$$f(x) = \int_0^1 \left(\sum_{i=1}^{n} x_i t^{i-1} - \tan t \right)^2 dt,\quad g(x, v) = \tan v - \sum_{i=1}^{n} x_i v^{i-1},$$
$$V = [0, 1],\quad p = 1.$$

Problem 14

$$f(x) = \frac{1}{2}x^T x,\quad g(x, v) = 3 + 4.5\sin(4.7\pi(v - 1.23)/8) - \sum_{i=1}^{n} x_i v^{i-1},$$
$$V = [0, 1],\quad p = 1,\ (x^0, v^0) = (2, 2, \ldots, 2, 1).$$

Table 1 The last three iterates generated by Algorithm 4.1

Problem	k	$\|\Phi(y^k)\|$	Problem	k	$\|\Phi(y^k)\|$	Problem	k	$\|\Phi(y^k)\|$
1	5	1.3496e-2	2	7	2.9284e-3	3	8	8.6888e-3
	6	9.0575e-5		8	5.1470e-6		9	1.2333e-4
	7	4.0332e-9		9	2.4003e-11		10	2.0364e-8
4	17	3.6864e-6	5	7	1.8260e-2	6	18	9.7817e-3
	18	1.3323e-6		8	2.6251e-4		19	3.3488e-5
	19	4.7863e-7		9	7.7329e-8		20	5.1996e-10
7	14	2.5145e-4	8	6	3.4413e-3	9	8	2.6080e-2
	15	1.1943e-6		7	1.5087e-5		9	4.7719e-4
	16	1.9282e-8		8	3.7231e-9		10	1.2153e-7
10	7	4.6280e-3	11	6	4.7301e-3	12	8	8.2845e-4
	8	1.1212e-5		7	2.7131e-5		9	8.6175e-5
	9	1.5885e-10		8	1.0377e-7		10	8.8161e-7

We first implemented Algorithm 4.1 for Problems 1–12 in MATLAB and the numerical experiments were done by using a Pentium III 733MHz computer with 256 MB of RAM. We compared Algorithm 4.1 with `fseminf` that is a solver for SIP based on an implementation of the discretization SQP method in MATLAB toolbox. We use $\|\Phi(y^k)\| \leq 10^{-6}$ as the stopping criterion for Algorithm 4.1, in this case, the obtained final iteration x^k is an approximation of a feasible point of (1.1) under certain assumptions. 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 algorithm are specified as follows

$$\gamma = 0.5, \ \ \rho = 0.5, \ \ \sigma = 0.001, \ \ \bar{t} = \bar{s} = 0.5.$$

The starting points t^0, s^0 for all problems are set $t^0 = \bar{t}$, $s^0 = \bar{s}$. The starting points u^0, w^0 are equal to 1.0**e**, 1.0**e** for Problems 1–12, where **e** is the vector of ones. For the solver `fseminf`, we use all the default values.

In the test of Problems 1–12, the values of p are estimated by using the following adaptive strategy. First, we let $p = 1$ and use Algorithm 4.1 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 4.1 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 10 of 12 test problems and $p = 2$ for other two test problems by the above method.

The test results for Problems 1–12 are summarized in Tables 1 and 2. In Table 1, $\Phi(y^k)$ is the value of the function $\Phi(y)$ in (3.8) 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; $f(x^k)$ 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 (2.2) at the final iteration.

The results reported in Tables 1 and 2 show that Algorithm 4.1 performs well for these test problems. From Table 1, we can see that Algorithm 4.1 indeed has superlinear convergence property. From Table 2, we can see that Algorithm 4.1 uses less CPU time than `fseminf` for 7 test problems and `fseminf` uses less CPU time than Algorithm 4.1 for other 5 test problems. Moreover, it appears from Table 2 that Algorithm 4.1 indeed can ensure the feasibility of the test problems.

We also implemented Algorithm 4.1 for Problems 13–14 in FORTRAN 77 by using a Pentium III 1133 MHz computer with 256 MB memory. The dimensions (N) of the two problems are chosen by 20, 40, 60, 80, 100, 200, 400, 1000 and 2000. All calculation within the driving programs, test problems and optimization code are carried out in double precision. In the test of the two problems, the termination condition is $\|\Phi(y^k)\| \leq 10^{-5}$, the starting points u^0, w^0 are set 0.5**e** and 0.5**e**, respectively, and other parameters are same to that in the test of Problems 1–12.

The test results for Problems 13 and 14 are given in Tables 3 and 4, respectively. Problem 13 is dense, i.e. its Hessian of Lagrangian function is not sparse. Here, Algorithm 4.1 is used for solving Problem 13, whose dimensions range from 20 to 200. Table 3 shows that Algorithm 4.1 performs well for solving some medium dense SIP problems. Table 4 shows that Algorithm 4.1 performs very well for solving Problem 14 with the different dimensions. Specially, the iteration number almost has no increase when $N \geq 200$.

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 1–6, 8–9 and 11–12, when $p \geq 2$, these test

Table 2 Test results for Algorithm 4.1 and `fseminf`

Problem	Algorithm 4.1				`fseminf`			
	n.it	cpu	$f(x^k)$	$G(x^k)$	n.it	cpu	$f(x^k)$	$G(x^k)$
1	7	0.05	2.2	0	7	0.17	2.1989	7.804e-8
2	9	0.17	5.3347	3.456e-13	30	0.50	5.3242	7.467e-5
3	10	0.13	0.1945	0	3	0.03	0.1945	0
4	19	0.16	1	6.357e-9	10	0.14	1	2.568e-3
5	9	0.33	0	0	7	0.06	0	0
6	20	0.28	97.1589	0	8	0.19	97.1589	2.010e-24
7	16	1.92	0.3820	2.054e-12	13	2.23	0.3820	1.221e-7
8	8	0.91	0	0	1	1.67	0	0
9	10	13.75	1	0	7	4.78	1	0
10	9	3.64	0	0	6	4.75	0	0
11	8	1.23	1.0191	0	5	2.78	1.0191	0
12	10	1.58	0.0885	0	2	1.88	0.0885	1.611e-10

Table 3 Test results of Problem 13

N	ITK	CPU	$\bar{G}(t^k, x^k)$	$\theta(y^k)$	$f(x^k)$
20	38	0.23	8.46e-8	7.53e-11	1.09
40	129	2.37	2.42e-9	8.71e-11	2.26
60	138	6.18	1.51e-7	6.33e-11	2.71
80	161	10.71	7.55e-9	4.65e-12	3.75
100	192	19.10	1.58e-8	4.41e-11	2.98
200	210	80.14	3.75e-9	1.80e-13	3.41

Table 4 Test results of Problem 14

N	ITK	CPU	$\bar{G}(t^k, x^k)$	$\theta(y^k)$	$f(x^k)$
60	6	0.03	2.03e-9	5.03e-9	0.02947
100	7	0.06	2.55e-12	1.96e-11	0.02941
200	9	0.18	1.62e-14	1.69e-13	0.02942
400	9	0.30	1.41e-10	7.15e-10	0.02942
1000	9	0.86	2.44e-9	7.05e-9	0.02941
2000	11	1.17	7.68e-11	1.13e-9	0.02941

problems can not be solved by Algorithm 4.1 within 30 iterations. For problems 7 and 10, when $p = 1$, the two test problems can not be solved by Algorithm 4.1 within 30 iterations. This means that it is important to choose a suitable number p when we use Algorithm 4.1 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. In addition, If V in (1.1) is a nonpolyhedral index set, then our method cannot be used directly. 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. From Tables 3 and 4, we see that our algorithm is hopeful for SIP problem with more higher dimension decision variable. It is hoped that an improved version of Algorithm 4.1 may also be capable of handling high dimensional index sets.

7 Final remarks

In this paper we have presented a smoothing 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 nonsmooth equations, and solved by a smoothing Newton-type method. Under certain 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 [14,26], our method can ensure the feasibility of (1.1). As future work, one problem is to find a way to determine a suitable number p in the KKT system of the SIP problem. Another problem is to find conditions which ensure the quadratic convergence of our method.

Acknowledments We would like to thank the anonymous referees for their helpful comments and suggestions which lead to a significantly improved version of this article.

References

1. Clarke, F.H.: Optimization and Nonsmooth Analysis. Wiley, New York (1983)
2. Coope, I.D., Watson, G.A.: A projected Lagrangian algorithm for semi-infinite programming. Math. Prog. **32**, 337–356 (1985)
3. Facchinei, F., Soares, J.: A new merit function for nonlinear complementarity problems and a related algorithm. SIAM J. Optim. **7**, 225–247 (1997)
4. Fang, S.C., Wu, S.Y.: An inexact approach to solving linear semi-infinite programming problems. Optimization **28**, 291–299 (1994)
5. Goberna, M.A., López, M.A.: Optimal value function in semi-infinite programming. J. Optim. Theory Appl. **59**, 261–279 (1988)

6. Goberna, M.A., López, M.A.: Semi-Infinite Programming: Recent Advances. Kluwer, Dordrecht (2001)
7. Gustafson, S.A.: On numerical analysis in semi-infinite programming. In: Hettich, R. (ed.) Semi-Infinite Programming, Lecture Notes in Control and Information Science, vol 15, pp. 51–65. Springer, New York (1979)
8. Gustafson, S.A.: A three-phase algorithm for semi-infinite programming. In: Fiacco, A.V., Kortanek, K.O. (eds.) Semi-Infinite Programming and Applications, pp. 138–157. Springer, Berlin (1983)
9. Hettich, R., Kortanek, K.O.: Semi-infinite programming: theory, methods, and applications. SIAM Review **35**, 380–429 (1993)
10. Hettich, R., Zencke, P.: Numerische Methoden der Approximation und Semi-infinite Optimierung. Teubner, Stuttgart (1982)
11. Ito, S., Lin, Y., Teo, K.L.: A dual parametrization method for convex semi-infinite programming. Ann. Oper. Res. **98**, 189–214 (2000)
12. Kanzow, C., Qi, H.D.: A QP-free constrained Newton-type method for variational inequality problems. Math. Prog. **85**, 81–106 (1999)
13. Lai, H.C., Wu, S.Y.: On linear semi-infinite programming problems: an algorithm. Numer. Funct. Anal. Optim. **13**, 287–304 (1992)
14. Li, D.H., Qi, L., Tam, J., Wu, S.Y.: A smoothing Newton method for semi-infinite programming. J. Global Optim. **30**, 169–194 (2004)
15. Lin, C.J., Fang, S.C., Wu, S.Y.: A dual affine scaling based algorithm for solving linear semi-infinite programming problems. In: Du, D.Z., Sun, J. (eds.) Advances in Optimization and Approximation, pp. 217–233. Kluwer, London (1994)
16. Mifflin, R.: Semismooth and semiconvex functions in constrained optimization. SIAM J. Control Optim. **15**, 957–972 (1977)
17. Moré, J.J., Sorensen, D.C.: Computing a trust region step. SIAM J. Sci. Stat. Comput. **4**, 553–572 (1983)
18. Ni, Q., Ling, C., Qi, L., Teo, K.L.: A truncated projected Newton-type algorithm for large scale semi-infinite programming. SIAM J. Optim. **16**, 1137–1154 (2006)
19. Pang, J.S., Qi, L.: Nonsmooth equations: motivation and algorithms. SIAM J. Optim. **3**, 443–465 (1993)
20. Polak, E., Tits, A.L.: A recursive quadratic programming algorithm for semi-infinite programming problems. Appl. Math. Optim. **8**, 325–349 (1982)
21. Qi, L.: Convergence analysis of some algorithms for solving nonsmooth equations. Math. Oper. Res. **18**, 227–244 (1993)
22. Qi, L., Ling, C., Tong, X.J., Zhou, G.: A smoothing projected Newton-type algorithm for semi-infinite programming. Comput. Optim. Appl. **42**, 1–30 (2009)
23. Qi, L., A., Shapiro, Ling, C.: Differentiability and semismoothness properties of integral functions and their applications. Math. Prog. **102**, 223–248 (2005)
24. Qi, L., Sun, D., Zhou, G.: A new look at smoothing Newton methods for nonlinear complementarity problems and box constrained variational inequalities. Math. Prog. **87**, 1–35 (2000)
25. Qi, L., Sun, J.: A nonsmooth version of Newton's method. Math. Prog. **58**, 353–367 (1993)
26. Qi, L., Wu, S.Y., Zhou, G.: Semismooth Newton methods for solving semi-infinite programming problems. J. Global Optim. **27**, 215–232 (2003)
27. Roleff, K: A stable multiple exchange algorithm for linear SIP. In: Hettich, R. (ed.) Semi-Infinite Programming, Lecture Notes in Control and Information Science, vol 15, pp. 83–96. Springer, New York (1979)
28. Rückmann, J.J., Shapiro, A.: Second order optimality conditions in generalized semi-infinite programming. Set-Valued Anal. **9**, 169–186 (2001)
29. Shapiro, A.: First and second order optimality conditions and perturbation analysis of semi-infinite programming problems. In: Reemtsen, R., Rükmann, J. (eds.) Semi-Infinite Programming, pp. 103–133. Kluwer, Boston (1998)
30. Stein, O., Tezel, A.: The semismooth approach for semi-infinite programming under the reduction Ansatz. J. Global Optim. **41**, 245–266 (2008)
31. Still, G.: Discretization in semi-infinite programming: the rate of convergence. Math. Prog. **91**, 53–69 (2001)
32. Sheu, R.L., Wu, S.Y., Fang, S.C.: A primal-dual infeasible-interior-point algorithm for linear semi-infinite programming. Comput. Math. Applic. **29**, 7–18 (1995)
33. Tanaka, Y., Fukushima, M., Ibaraki, T.: A globally convergent SQP method for semi-infinite nonlinear optimization. J. Comp. Appl. Math. **23**, 141–153 (1988)
34. Teo, K.L., Rehbock, V., Jennings, L.S.: A new computational algorithm for functional inequality constrained optimization problems. Automatica **29**, 789–792 (1993)
35. Teo, K.L., Yang, X.Q., Jennings, L.S.: Computational discretization algorithms for functional inequality constrained optimization. Ann. Oper. Res. **98**, 215–234 (2000)
36. Todd, M.J.: Interior-point algorithms for semi-infinite programming. Math. Prog. **65**, 217–245 (1994)

37. Watson, G.A.: A multiple exchange algorithm for multivariate Chebyshev approximation. SIAM J. Numer. Anal. **12**, 46–52 (1975)
38. Watson, G.A.: Numerical experiments with globally convergent methods for semi-infinite programming problems. In: Fiacco, A.V., Kortanek, K.O. (eds.) Semi-Infinite Programming and Applications, pp. 193–205. Springer, Berlin (1983)
39. Wu, S.Y., Li, D.H., Qi, L., Zhou, G.: An iterative method for solving KKT system of the semi-infinite programming. Optim. Methods Softw. **20**, 629–643 (2005)