Journal of Global Optimization **30:** 169–194, 2004.

A Smoothing Newton Method for Semi-Infinite Programming

DONG-HUI LI[1], LIQUN QI[2], JUDY TAM[2] and SOON-YI WU[3]
[1]*Department of Applied Mathematics, Hunan University Changsha, China;*
[2]*Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong;*
[3]*Institute of Applied Mathematics, National Cheng-Kung University, Tainan 700, Taiwan*

(Received 17 December 2002; accepted in revised form 6 January 2003)

Abstract. This paper is concerned with numerical methods for solving a semi-infinite programming problem. We reformulate the equations and nonlinear complementarity conditions of the first order optimality condition of the problem into a system of semismooth equations. By using a perturbed Fischer–Burmeister function, we develop a smoothing Newton method for solving this system of semismooth equations. An advantage of the proposed method is that at each iteration, only a system of linear equations is solved. We prove that under standard assumptions, the iterate sequence generated by the smoothing Newton method converges superlinearly/quadratically.

Key words: Semi-infinite programming, KKT condition, Semismooth equations, Smoothing Newton method

1. Introduction

Semi-infinite programming (SIP) is an exciting part of mathematical programming. It has strong practical backgrounds in approximation theory, optimal control and numerous engineering problems, etc. [17, 29]. For example, the enveloped constrained filter design in information technology requires that the response lies in a pre-prescribed envelope at all times [34]. This leads to an SIP problem.

Let $f : R^n \to R$, $g : R^{n+m} \to R$ be twice continuously differentiable, and $V \subset R^m$ be compact. Consider the following SIP problem

$$\min\{f(x), \quad x \in X\}, \tag{1}$$

where $X = \{x \in \Re^n : g(x, v) \leqslant 0, \ \forall v \in V\}$. If V is finite, the problem (1) is a finite optimization problem and is usually called a nonlinear programming (NLP) problem in the literature. One of the important methods for solving an SIP problem is the discretization method [29, 33]. In a discretization method, the infinite set V is approximated by a sequence of finite subsets $\{V_k\}$ such that V_k becomes denser and denser in V as k goes to the infinity. Then the SIP problem is approximated by a sequence of NLP problems

$$\min\{f(x),\quad x \in X_k\},\tag{2}$$

where $X_k = \{x \in \Re^n : g(x, v) \leqslant 0,\ \forall v \in V_k\}$, such that the solution x^k of (2), as hoped, converges to a solution of (1). At each iteration of a discretization method, an NLP problem is solved. Therefore, discretization methods especially suit for solving problems with a solution at which $g(x^*, \cdot)$ is (almost) constant on V or on a part of V. The almost constant property is a feature of some kinds of Chebyshev approximation problems.

However, in a general discretization method, the subset $V_k \subset V$ must be sufficiently dense in V when k is sufficiently large. This makes the algorithm computationally very expensive. The time needed to verify feasibility with respect to (2) and to solve this problem increases dramatically as the cardinality of V_k grows. To reduce the computational cost of discretization methods, a so-called reduction technique [**?**] was introduced which results in reduction based methods. Let the set V be specified by

$$V = \{c_j(v) \leqslant 0,\quad j = 1, \dots, q\},\tag{3}$$

where $c_j : R^m \to R$, $j = 1, \dots, q$ are twice continuously differentiable. The process of a typical reduction based method for solving an SIP problem with V specified by (3) is as follows. At iteration k, compute all local minimizers of the problem

$$\min\{-g(x^k, v),\quad v \in V\}.\tag{4}$$

Denote by S_k the set of all minimizers of (4). Solve the problem

$$\min\{f(x),\quad g(x, v) \leqslant 0,\ \forall v \in S_k\}\tag{5}$$

to get the next iterate x_{k+1}. Problems (4) and (5) are called outer and inner problems, respectively. Under some regular conditions [29], it has been proved that the set S_k is finite and hence the problem (5) reduces to a nonlinear programming problem. The regular conditions include:

(i) For every k and any $v \in V$, $\nabla c_j(v)$, $j \in I(v) \triangleq \{j \mid c_j(v) = 0\}$ is linearly independent.

(ii) At every stationary point of (4), the strong second order sufficiency conditions and the strictly complementarity condition hold.

Reduction based and discretization SQP type methods and trust region type methods have been studied by some authors [4, 13, 32]. Under certain conditions, these methods possess global convergence property. However, finding all local minimizers of (2) is very difficult and very expensive in computation. Discussion on the difficulties and numerical labor in finding all local minimizers of (2) can be found in [28].

In this paper, from a different point of view, we develop a new kind of iterative methods for finding a KKT point of an SIP problem. Let

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

It is well-known [31] that if x is a local minimum of the SIP problem (1), and if the extended Mangasarian–Fromovitz constraint qualification (EMFCQ) holds at x, i.e., there is a vector $h \in \Re^n$ such that

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

for all $v \in V(x)$. Then there are p positive numbers u_i and p vectors $v^i \in V(x)$ such that

$$\nabla f(x) + \sum_{i=1}^{p} u_i \nabla_x g(x, v^i) = 0, \tag{6}$$

with $p \leqslant n$. If the EMFCQ does not hold, an example given in [27] shows that the optimality condition (6) may not hold. We may also explicitly write out the conditions $x \in X$ and $v^i \in V(x)$ as

$$g(x, v) \leqslant 0, \ \forall v \in V, \tag{7}$$

and for $i = 1, \dots, p$,

$$u_i > 0, \ g(x, v^i) = 0. \tag{8}$$

Equations and inequalities (6), (7) and (8) are called the KKT system of the SIP problem (1). In a KKT system, x is called a **stationary point** of the SIP problem, and $u \in \Re^p$ and v^i for $i = 1, \dots, p$ are called its **Lagrange multiplier** and **attainers** respectively.

The KKT systems (6), (7) and (8) look like the KKT system of nonlinear programming problem

$$\min\{f(x), \ \ g(x, v^i) \leqslant 0, \ \ i = 1, \dots, p\}. \tag{9}$$

However, due to the restriction $v^i \in V(x)$ and the fact that p depends on x, the solution of (6), (7) and (8) is much more complicated than that of a general nonlinear programming problem. To develop numerical methods, we then analyze the restriction $v^i \in V(x)$, $i = 1, \dots, p$.

By the definition of $V(x)$, the condition $v^i \in V(x)$, $i = 1, \dots, p$ means that v^i, $i = 1, \dots, p$ are global minimizers of the NLP problem

$$\min\{-g(x, v) : c(v) \leqslant 0\}. \tag{10}$$

It is well-known that if a constraint qualification (CQ) for the NLP problem (10) holds, then there are p **auxiliary Lagrange multipliers** $w^i \in \Re^q$ for $i = 1, \ldots, p$ such that for $i = 1, \ldots, p$,

$$\begin{aligned} -\nabla_v g(x, v^i) + \sum_{j=1}^{q} w_j^i \nabla c_j(v^i) &= 0, \\ w_j^i \geqslant 0, \;\; c_j(v^i) &\leqslant 0, \\ w_j^i c_j(v^i) &= 0, \;\; \text{for } j = 1, \ldots, q. \end{aligned} \tag{11}$$

Well-known CQ's for NLP include the linear independence CQ (LICQ) [15], the Slater CQ (SLCQ) [15], the Mangasarian–Fromovitz CQ (MFCQ) [15], the constant rank CQ (CRCQ) [11], etc. [26].

We call an $x \in \Re^n$ with $u \in \Re^p$, $v^i \in \Re^m$ and $w^i \in \Re^q$, for $i = 1, \ldots, p$, $p \leqslant n$, satisfying (6), (8) and (11) a **substationary point** of the SIP problem.

System (11) is a first order necessary condition for v^i, $i = 1, \ldots, p$ to be local solutions of (10). If some second order sufficiency conditions hold for (10) at v^i for $i = 1, \ldots, p$, then v^i, $i = 1, \ldots, p$ are local solutions of (10). Note that system (11) is a system consisting of finitely many equations and inequalities. Thus, the conditions (6) and (8) together with $v^i \in V(x)$, $i = 1, \ldots, p$ are transformed into the following system

$$\begin{cases} \nabla f(x) + \sum_{i=1}^{p} u_i \nabla_x g(x, v^i) = 0, \\ u_i > 0, \;\; g(x, v^i) = 0, \;\; \text{for } i = 1, \ldots, p, \\ -\nabla_v g(x, v^i) + \sum_{j=1}^{q} w_j^i \nabla c_j(v^i) = 0, \;\; \text{for } i = 1, \ldots, p, \\ w_j^i \geqslant 0, \;\; c_j(v^i) \leqslant 0, \;\; \text{for } i = 1, \ldots, p; \;\; j = 1, \ldots, q, \\ w_j^i c_j(v^i) = 0, \;\; \text{for } i = 1, \ldots, p; \;\; j = 1, \ldots, q. \end{cases} \tag{12}$$

It is then desirable to develop numerical methods on the basis of (12). However we realize that in order to prove the nonsingularity conditions required by our algorithm, we need to modify the above system accordingly. Since $u_i > 0$ for $i = 1, 2, \cdots, p$, we may multiply the third equation in (12) by u_i and then further replace $u_i w_j^i$ by w_j^i for $i = 1, 2, \cdots, p$; $j = 1, 2, \cdots, q$. Thus system (12) is equivalent to the following:

$$\begin{cases} \nabla f(x) + \sum_{i=1}^{p} u_i \nabla_x g(x, v^i) \;=\; 0, \\ u_i > 0, \;\; g(x, v^i) \;=\; 0, \;\; \text{for } i = 1, \ldots, p, \\ -u_i \nabla_v g(x, v^i) + \sum_{j=1}^{q} w_j^i \nabla c_j(v^i) \;=\; 0, \;\; \text{for } i = 1, \ldots, p, \\ w_j^i \geqslant 0, \;\; c_j(v^i) \;\leqslant\; 0, \;\; \text{for } i = 1, \ldots, p; \;\; j = 1, \ldots, q, \\ w_j^i c_j(v^i) \;=\; 0, \;\; \text{for } i = 1, \ldots, p; \;\; j = 1, \ldots, q. \end{cases} \tag{13}$$

The KKT system and (13) are not equivalent. We should not forget the feasibility constraint (7). But (7) involves an infinite number of inequalities for x. As mentioned before, in most applications, the following assumption holds [18, 29, 32].

ASSUMPTION 1. *For any fixed x, the number of local minima of (10) is finite.*

This number depends upon x, and is unknown in general. Under Assumption 1, we may solve the finite system (13) by first finding its solution x, and check if (7) holds for these finitely many minima of (10) at x. If (7) holds at these points, then x is a substationary point of the SIP problem. In fact in some cases, it automatically holds. For example, if $g(x, \cdot)$ is a concave function, c is convex and $p \geqslant 1$, then a solution of the finite system (13) is a substationary point of the SIP problem automatically.

Note that under Assumption 1, omitting (7) may omit some solutions x with $p = 0$. But those x are solutions of

$$\nabla f(x) = 0, \tag{14}$$

which is in general not a difficult problem. Once we have found some solutions of (14), we then need to check if (7) holds for these solutions.

In this paper, we reformulate the system (13) into a system of semismooth equations. Then, we propose a smoothing Newton method to solve the system of semismooth equations. Under mild conditions, we prove that the proposed method is globally and superlinearly/quadratically convergent. The advantage of the proposed method is that, unlike discretization methods and reduction based methods in which, at each iteration, two nonlinear optimization problems have to be solved, the proposed method only solves a system of linear equations at each iteration.

The proposed method is an extension of smoothing Newton methods for solving semismooth equations arising from nonlinear programming problems, nonlinear complementarity problems and variational inequality problems etc.. We refer to [1, 24], a survey paper [23] and references therein for details on smoothing methods.

The organization of the paper is as follows. In the next section, we reformulate the system (13) into a system of semismooth equations. After introducing the concept of a smoothing function, we propose a smoothing Newton method and discuss some properties of the proposed method in Section 3. In Section 4, we establish global and superlinear/quadratic convergence of the proposed method. Numerical results are reported in Section 5. Finally, in Section 6, we discuss a special case of the method when V is a closed interval in R.

2. A Semismooth Equation Reformulation

In this section, we reformulate the system (13) into a system of semismooth equations. We first briefly review NCP and semismooth functions.

A function $\phi : \Re^2 \to \Re$ is called an NCP function [21] if $\phi(a, b) = 0$ if and only if $a \geqslant 0$, $b \geqslant 0$ and $ab = 0$. Two well-known NCP functions are the minimum function

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

and the Fischer–Burmeister function [6, 21]

$$\phi_{FB}(a, b) = \sqrt{a^2 + b^2} - a - b. \tag{15}$$

A locally Lipschitz function $F : \Re^n \to \Re^p$ is called semismooth [16, 19, 25] at $x \in \Re^n$ if F is directionally differentiable at x and for all $V \in \partial F(x + d)$ and $d \to 0$,

$$F'(x; d) = Vd + o(\|d\|),$$

where ∂F is the generalized Jacobian of F in the sense of Clarke [2]. F is called strongly semismooth [6, 21, 22, 25] at x if F is semismooth at x and for all $V \in \partial F(x + d)$ and $d \to 0$,

$$F'(x; d) = Vd + O(\|d\|^2).$$

Both the minimum function and the Fischer–Burmeister function are not smooth (continuously differentiable), but they are strongly semismooth.

By the use of the Fischer–Burmeister function ϕ_{FB} defined by (15), we may reformulate (13) as a system of semismooth equations:

$$\begin{cases} \nabla f(x) + \sum_{i=1}^{p} u_i \nabla_x g(x, v^i) = 0, \\ \phi_{FB}(u_i, -g(x, v^i)) = 0, \text{ for } i = 1, \ldots, p, \\ -u_i \nabla_v g(x, v^i) + \sum_{j=1}^{q} w_j^i \nabla c_j(v^i) = 0, \text{ for } i = 1, \ldots, p, \\ \phi_{FB}(w_j^i, -c_j(v^i)) = 0, \text{ for } i = 1, \ldots, p;\ j = 1, \ldots, q. \end{cases} \tag{16}$$

It is obvious that if ϕ_{FB} in (16) is replaced by any other NCP function, the equivalence between (13) and (16) remains true.

Nonlinear equation (16) transforms the system (13) into a semismooth equation of dimension $n + (1 + m + q)p$. It is then desired to solve it by some smoothing Newton methods. This is the goal of the next two sections. However, we understand that in appearance,(16) is not “totally" equivalent to (13). It allows the case that

$$u_i = 0, \qquad g(x, v^i) \leqslant 0.$$

If there is an $n + (1 + m + q)p$ dimensional vector satisfying (16), we may then drop the part indexed by i where $u_i = 0$. In this case, we get a solution of (13) which obviously satisfies (16). Hence, (13) and (16) are equivalent.

We also note that the parameter p depends upon the problem. One possibility is to use $p_k = r(x^k)$ at the $(k + 1)$th iteration to find x^{k+1}, where $r(x)$ is the rank of $\{\nabla_x g(x, v) : \ v \in V\}$. In the latter part of this paper, we will study the simple case which is in fact rather common in applications, i.e., p is known. For instance, if Assumption 1 holds and for any fixed x, $g(x, \cdot)$ is a concave function. Then $p = 0$ or 1. Since the solution for $p = 0$ can be checked by solving (14), as discussed before, under that additional assumption, a method for solving the case

$p = 1$, combining a solution of (14) satisfying (7), will then solve the problem. Even when p is unknown but small, say $p = 2$ or 3, we may still try $p = 1$ first. If it fails, we may then try $p = 2$, and so on. For the case when p is unknown and not small, it will not be in the scope of the present paper.

3. Smoothing Functions and A Smoothing Newton Method

In this section, we first introduce the concept of a smoothing function and then construct a smoothing function for the semismooth function H defined by (16). Let $\epsilon \neq 0$ be a parameter. We call H^ϵ a smoothing function of a semismooth function H if it is continuously differentiable everywhere and there is a constant $\mu > 0$ independent of ϵ such that

$$\|H^\epsilon(z) - H(z)\| \leqslant \mu\epsilon, \qquad \forall z.$$

The basic idea of a smoothing method for solving the semismooth equation $H(z) = 0$ is to generate a sequence $\{z^{\epsilon_k}\}$ that are (approximate) solutions of the smooth equation $H^{\epsilon_k}(z) = 0$, such that there exists at least one accumulation point of $\{z^{\epsilon_k}\}$ that, as hoped, is a solution of the semismooth equation $H(z) = 0$.

Smoothing Newton-like methods have received much attention in solving semismooth equations arising from the nonlinear complementarity problem, the variational inequality problem and the KKT system of the nonlinear programming problem in recent years [1, 7, 12, 14, 20, 23, 24, 36]. We want to develop a smoothing Newton method for solving smoothing Equations (16) arising from SIP. We first construct a smoothing function for the semismooth function H defined by (16).

Let $\phi_{FB}^\epsilon : R^2 \to R$ be the perturbed Fischer-Burmeister function defined by

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

It is obvious that for any $\epsilon > 0$, ϕ_{FB}^ϵ is differentiable everywhere and for each $\epsilon \geqslant 0$, we have

$$|\phi_{FB}^\epsilon(a, b) - \phi_{FB}(a, b)| \leqslant \epsilon, \qquad \forall (a, b) \in R^2. \tag{17}$$

In particular, $\phi_{FB}^0(a, b) = \phi_{FB}(a, b)$ for all $(a, b) \in R^2$. By direct computation, we get for every $\epsilon > 0$,

$$\nabla\phi_{FB}^\epsilon(a, b) = -\Big(1 - \frac{a}{\sqrt{a^2 + b^2 + \epsilon^2}},\ 1 - \frac{b}{\sqrt{a^2 + b^2 + \epsilon^2}}\Big). \tag{18}$$

$$\partial_B\phi_{FB}(a, b) = \begin{cases} -\Big(1 - \frac{a}{\sqrt{a^2 + b^2}},\ 1 - \frac{b}{\sqrt{a^2 + b^2}}\Big), & \text{if} a^2 + b^2 \neq 0, \\ \{-(\alpha,\ \beta) \mid (\alpha - 1)^2 + (\beta - 1)^2 = 1\}, & \text{if} a = b = 0, \end{cases} \tag{19}$$

where $\partial_B \phi_{FB}(z)$ stands for the B-differential of ϕ_{FB} at z.

$$\partial \phi_{FB}(a, b) = \begin{cases} -\Big(1 - \frac{a}{\sqrt{a^2+b^2}},\ 1 - \frac{b}{\sqrt{a^2+b^2}}\Big), & \text{if} a^2 + b^2 \neq 0, \\ \{-(\alpha,\ \beta) \mid (\alpha-1)^2 + (1-\beta)^2 \leqslant 1\}, \text{if} a = b = 0. \end{cases} \tag{20}$$

From equations (18) and (20), it is clear that for every $\epsilon > 0$

$$\nabla \phi_{FB}^{\epsilon}(0, 0) \in \partial \phi_{FB}(0, 0). \tag{21}$$

The following lemma further reveals the relation between ϕ_{FB}^{ϵ} and $\partial \phi_{FB}$.

LEMMA 2. *Let $\{(a_k, b_k)\}$ be any nonzero sequence converging to some point (a, b) and $\{\epsilon_k\}$ be sequences of positive numbers, which converge to zero. The relation*

$$\lim_{k\to\infty} dist(\nabla \phi_{FB}^{\epsilon_k}(a_k, b_k),\ \partial \phi_{FB}(a_k, b_k)) = 0 \tag{22}$$

holds if and only if

$$\lim_{k\to\infty} \frac{\epsilon_k}{\sqrt{a_k^2 + b_k^2}} = 0. \tag{23}$$

Proof. By an elementary deduction, we have for any k

$$\begin{aligned} &\text{dist}(\nabla \phi_{FB}^{\epsilon_k}(a_k, b_k),\ \partial \phi_{FB}(a_k, b_k)) = \|\nabla \phi_{FB}^{\epsilon_k}(a_k, b_k) - \nabla \phi_{FB}(a_k, b_k))\| \\ &= \Big\|\Big(\frac{a_k}{\sqrt{a_k^2+b_k^2+\epsilon_k^2}} - \frac{a_k}{\sqrt{a_k^2+b_k^2}},\ \frac{b_k}{\sqrt{a_k^2+b_k^2+\epsilon_k^2}} - \frac{b_k}{\sqrt{a_k^2+b_k^2}}\Big)\Big\| \\ &= \Big\|\Big(\frac{a_k(\sqrt{a_k^2+b_k^2+\epsilon_k^2} - \sqrt{a_k^2+b_k^2})}{\sqrt{a_k^2+b_k^2+\epsilon_k^2}\sqrt{a_k^2+b_k^2}},\ \frac{b_k(\sqrt{a_k^2+b_k^2+\epsilon_k^2} - \sqrt{a_k^2+b_k^2})}{\sqrt{a_k^2+b_k^2+\epsilon_k^2}\sqrt{a_k^2+b_k^2}}\Big)\Big\| \\ &= \frac{\epsilon_k^2}{\sqrt{a_k^2+b_k^2+\epsilon_k^2}\Big(\sqrt{a_k^2+b_k^2+\epsilon_k^2} + \sqrt{a_k^2+b_k^2}\Big)} \\ &= \frac{\bar{\epsilon}_k^2}{\sqrt{1+\bar{\epsilon}_k^2}\Big(1+\sqrt{1+\bar{\epsilon}_k^2}\Big)}, \end{aligned} \tag{24}$$

where $\bar{\epsilon}_k = \epsilon_k/\sqrt{a_k^2 + b_k^2}$. Equality (24) implies that (22) holds if and only if $\bar{\epsilon}_k \to 0$ as desired. □

For simplicity, we denote $z = (x, u, v^1, \dots, v^p, w^1, \dots, w^p) \in R^{n+(1+m+q)p}$ and define some functions on $R^{n+(1+m+q)p}$ as follows.

$$
\begin{aligned}
L(z) &= \nabla f(x) + \sum_{i=1}^{p} u_i \nabla_x g(x, v^i),\\
\Psi^\epsilon(z) &= (\psi_1^\epsilon(z), \dots, \psi_p^\epsilon(z))^T,\\
\psi_i^\epsilon(z) &= \phi_{FB}^\epsilon(u_i, -g(x, v^i)), \quad i = 1, 2, \dots, p,\\
l(z) &= (l_1(z), \dots, l_p(z))^T,\\
l_i(z) &= -u_i \nabla_v g(x, v^i) + \sum_{j=1}^{q} w_j^i \nabla c_j(v^i), \quad i = 1, \dots, p,\\
\Phi^\epsilon(z) &= (\Phi_1^\epsilon(z)^T, \dots, \Phi_p^\epsilon(z)^T)^T,\\
\Phi_i^\epsilon(z) &= (\phi_{i1}^\epsilon(z), \dots, \phi_{iq}^\epsilon(z))^T, \quad i = 1, 2, \dots, p,\\
\phi_{ij}^\epsilon(z) &= \phi_{FB}^\epsilon(w_j^i, -c_j(v^i)), \quad i = 1, 2, \dots, p;\ j = 1, 2, \dots, q.
\end{aligned}
$$

Let

$$
H^\epsilon(z) = \begin{pmatrix} L(z) \\ \Psi^\epsilon(z) \\ l(z) \\ \Phi^\epsilon(z) \end{pmatrix} \tag{25}
$$

and $H = H^0$. It is clear that for each $\epsilon > 0$, H^ϵ is continuously differentiable and H is semismooth. Moreover, H is strongly semismooth if f, g and c are twice Lipschitz continuously differentiable. The system (13) is then reformulated into the following semismooth equation.

$$
H(z) = 0. \tag{26}
$$

In addition, it follows from (17) that

$$
\|H^\epsilon(z) - H(z)\| \leqslant \mu\epsilon, \quad \forall z, \tag{27}
$$

where $\mu = n + (1 + m + q)p$.

We turn to computing the Jacobian $\nabla H^\epsilon(z)$. By direct computation, we have

$$
\begin{aligned}
\nabla_x L(z) &= \nabla^2 f(x) + \sum_{i=1}^{p} u_i \nabla_{xx}^2 g(x, v^i) \in R^{n\times n},\\
\nabla_u L(z) &= (\nabla_x g(x, v^i)^T)_{i=1}^p \triangleq\ nabla_x g(x, v)^T \in R^{p\times n},\\
\nabla_v L(z) &= \begin{pmatrix} u_1 \nabla_{xv}^2 g(x, v^1) \\ \vdots \\ u_p \nabla_{xv}^2 g(x, v^p) \end{pmatrix} \triangleq \Lambda_2(z) \nabla_{vx}^2 g(x, v)^T \in R^{mp\times n},\\
\nabla_w L(z) &= 0 \in R^{qp\times n},
\end{aligned}
$$

where $\nabla_x g(x, v) = (\nabla_x g(x, v^1), \ldots, \nabla_x g(x, v^p)) \in R^{p\times n}$, $\Lambda_2(z) = \mathrm{diag}(u_i I_m)_{1\leqslant i\leqslant p} \in R^{mp\times mp}$ and $\nabla^2_{xv} g(x, v) = \left(\nabla^2_{xv} g(x, v^1), \ldots, \nabla^2_{xv} g(x, v^p)\right) \in R^{n\times mp}$. Denote for $i = 1, 2, \ldots, p$; $j = 1, 2, \ldots, q$

$$a_i^\epsilon(z) = 1 + \frac{g(x, v^i)}{\sqrt{u_i^2 + g(x, v^i)^2 + \epsilon^2}}, \quad b_i^\epsilon(z) = 1 - \frac{u_i}{\sqrt{u_i^2 + g(x, v^i)^2 + \epsilon^2}}$$

$$\beta_{ij}^\epsilon(z) = 1 + \frac{c_j(v^i)}{\sqrt{(w_j^i)^2 + c_j(v^i)^2 + \epsilon^2}}, \quad \gamma_{ij}^\epsilon(z) = 1 - \frac{w_j^i}{\sqrt{(w_j^i)^2 + c_j(v^i)^2 + \epsilon^2}}.$$

It is obvious that for every $\epsilon > 0$, inequalities

$$0 < a_i^\epsilon(z) < 2, \quad 0 < b_i^\epsilon(z) < 2, \quad (a_i^\epsilon(z) - 1)^2 + (b_i^\epsilon(z) - 1)^2 < 1$$

and

$$0 < \beta_{ij}^\epsilon(z) < 2, \quad 0 < \gamma_{ij}^\epsilon(z) < 2, \quad (\beta_{ij}^\epsilon(z) - 1)^2 + (\gamma_{ij}^\epsilon(z) - 1)^2 < 1$$

hold for any z, every $i = 1, \ldots, p$ and $j = 1, \ldots, q$. We also have

$$\begin{aligned}
&\nabla_x \Psi^\epsilon(z) = \nabla_x g(x, v)\mathrm{diag}(a_i^\epsilon(z)) \stackrel{\triangle}{=} \nabla_x g(x, v)\Lambda_1^\epsilon(z) \in R^{n\times p},\\
&\nabla_u \Psi^\epsilon(z) = -\mathrm{diag}(b_i^\epsilon(z)),\\
&\nabla_v \Psi^\epsilon(z) = \mathrm{diag}(\nabla_v g(x, v^i))\Lambda_1^\epsilon(z) \in R^{mp\times p},\\
&\nabla_w \Psi^\epsilon(z) = 0 \in R^{qp\times p},
\end{aligned}$$

where $\Lambda_1^\epsilon(z) = \mathrm{diag}(a_i^\epsilon(z))$. Moreover, we have

$$\begin{aligned}
\nabla_x l_i(z) &= -u_i \nabla^2_{vx} g(x, v^i) \in R^{p\times m},\\
\nabla_u l_i(z) &= -\mathrm{diag}(\nabla_v g(x, v^i)^T) \in R^{p\times mp},\\
\nabla_v l_i(z) &= \left(0, \ldots, 0, -u_i \nabla^2_{vv} g(x, v^i) + \sum_{j=1}^q w_j^i \nabla^2 c_j(v^i), 0, \ldots, 0\right)^T\\
&\stackrel{\triangle}{=} (0, \ldots, 0, M_{ii}(z), 0, \ldots, 0\Big)^T \in R^{mp\times m}\\
\nabla_w l_i(z) &= \left(0, \ldots, 0, \nabla c(v^i), 0, \ldots, 0\right)^T \in R^{qp\times m},
\end{aligned}$$

where $M_{ii}(z) = -u_i \nabla^2_{vv} g(x, v^i) + \sum_{j=1}^q w_j^i \nabla^2 c_j(v^i) \in R^{p\times p}$ and $\nabla c(v^i) = (\nabla c_1(v^i), \ldots, \nabla c_q(v^i)) \in R^{p\times q}$. The above equations imply

$$\begin{aligned}
&\nabla_x l(z) = -(u_1 \nabla^2_{vx} g(x, v^1), \ldots, u_p \nabla^2_{vx} g(x, v^p)) = -\nabla^2_{xv} g(x, v)\Lambda_2(z),\\
&\nabla_u l(z) = -\Lambda_1^\epsilon(z)\mathrm{diag}(\nabla_v g(x, v^i)^T),\\
&\nabla_v l(z) = \mathrm{diag}(M_{ii}(z)) \stackrel{\triangle}{=} M(z),\\
&\nabla_w l(z) = \mathrm{diag}\left(\nabla c(v^i)^T\right),
\end{aligned}$$

where for vectors or matrices $p^1, p^2, \dots, p^n$, the matrix diag $((p^i)^T)$ is defined by

$$\operatorname{diag}((p^i)^T) = \begin{pmatrix} (p^1)^T & & & \\ & (p^2)^T & & \\ & & \ddots & \\ & & & (p^n)^T \end{pmatrix}.$$

We also have

$$\nabla_x \Phi_j^\epsilon(v, w) = 0 \in R^{n\times q}, \quad \nabla_u \Phi_j^\epsilon(v, w) = 0 \in R^{m\times q}.$$

$$\nabla_v \phi_{ij}^\epsilon(z) = (0, \dots, 0, \beta_{i,j}^\epsilon(v, w)\nabla c_j(v^i)^T, 0, \dots, 0)^T \in R^{mp\times q},$$
$$\nabla_w \phi_{ij}^\epsilon(z) = (0, \dots, 0, -\gamma_{ij}^\epsilon(z), 0, \dots, 0)^T \in R^{qp\times q}$$

$$\nabla_v \Phi_i^\epsilon(v, w) = \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \\ \beta_{i1}^\epsilon \nabla c_1(v^i) & \dots & \beta_{iq}^\epsilon \nabla c_q(v^i) \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \nabla c(v^i) \\ 0 \\ \vdots \\ 0 \end{pmatrix} B_i^\epsilon(z) \in R^{mp\times q},$$

where $B_i^\epsilon(z) = \operatorname{diag}(\beta_{i1}^\epsilon(z), \dots, \beta_{iq}^\epsilon(z))$.

$$\nabla_w \Phi_i^\epsilon(v, w) = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ -\gamma_{i1}^\epsilon(z) & 0 & \cdots & 0 \\ 0 & -\gamma_{i2}^\epsilon(z) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -\gamma_{iq}^\epsilon(z) \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \operatorname{diag}(-\gamma_{ij}^\epsilon(z))_{j=1}^q \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in R^{qp\times q}.$$

Hence

$$\nabla_x \Phi^\epsilon(z) = 0,$$
$$\nabla_u \Phi^\epsilon(z) = 0,$$
$$\nabla_v \Phi^\epsilon(z) = \operatorname{diag}\left(\nabla c(v^i) B_i^\epsilon(z)\right) \triangleq \operatorname{diag}(C_i(z)),$$
$$\nabla_w \Phi^\epsilon(z) = -\operatorname{diag}\left(\gamma_{ij}^\epsilon(z)\right).$$

From the above argument, we get the expression of $\nabla H^{\epsilon}(z)$ as follows.

$$\nabla H^{\epsilon}(z) = \begin{pmatrix} \nabla_x L(z) & \nabla_x \Psi^{\epsilon}(z) & \nabla_x l(z) & \nabla_x \Phi^{\epsilon}(z) \\ \nabla_u L(z) & \nabla_u \Psi^{\epsilon}(z) & \nabla_u l(z) & \nabla_u \Phi^{\epsilon}(z) \\ \nabla_v L(z) & \nabla_v \Psi^{\epsilon}(z) & \nabla_v l(z) & \nabla_v \Phi^{\epsilon}(z) \\ \nabla_w L(z) & \nabla_w \Psi^{\epsilon}(z) & \nabla_w l(z) & \nabla_w \Phi^{\epsilon}(z) \end{pmatrix}$$

$$= \begin{pmatrix} \nabla_x L(z) & \nabla_x g(x, v)\Lambda_1^{\epsilon}(z) & -\nabla^2_{xv} g(x, v)\Lambda_2(z) & \mathbf{0} \\ \nabla_x g(x, v)^T & -\mathrm{diag}(b_i^{\epsilon}(z)) & -\mathrm{diag}(\nabla_v g(x, v^i)^T) & \mathbf{0} \\ \Lambda_2(z)\nabla^2_{xv} g(x, v)^T & \mathrm{diag}(\nabla_v g(x, v^i))\Lambda_1^{\epsilon}(z) & M(z) & \mathrm{diag}(C_i(z)) \\ \mathbf{0} & \mathbf{0} & \mathrm{diag}\Big(\nabla c(v^i)^T\Big) & -\mathrm{diag}(\gamma_{ij}^{\epsilon}(z)) \end{pmatrix}. \tag{28}$$

We are going to state the steps of a smoothing Newton method. For simplicity, we use H^k, ∇H^k etc. to denote H^{ϵ_k} and ∇H^{ϵ_k} etc. Let

$$\xi_i^k = \sqrt{(u_i^2 + g(x, v^i)^2)_{z=z_k}}, \quad \xi_{ij}^k = \sqrt{((w_j^i)^2 + c_j(v^i)^2)_{z=z_k}}$$

$$J_k^1 = \{i \mid \xi_i^k \neq 0,\ i = 1, \dots, p\},$$

$$J_k^2 = \{(i, j) \mid \xi_{ij}^k \neq 0,\ i = 1, \dots, n, j = 1, \dots, q\}$$

and

$$J_k = J_k^1 \cup J_k^2 \qquad \xi^k = \min\{(\xi_i^k, i \in J_k^1),\ (\xi_{ij}^k, i, j \in J_k^2)\}.$$

Here and below, we use z_k instead of z^k to denote the kth iterate vector.

To make a smoothing Newton method be globally and superlinearly convergent, it is important to update the parameter ϵ_k in a proper way. Intuitively, a reasonable choice of ϵ_k is to let it satisfy

$$\lim_{k \to \infty} \mathrm{dist}(\nabla H^k(z_k),\ \partial_C H(z_k)) = 0. \tag{29}$$

Here, for a vector-valued function $\gamma = (\gamma_1, \dots, \gamma_m)^T$, the notation $\partial_C \gamma(z)$ is specified by

$$\partial_C \gamma(x) = \partial\gamma_1(x) \times \partial\gamma_2(x) \times \cdots \times \partial\gamma_m(x).$$

A general rule of how to let ϵ_k satisfy (29) has been discussed in [1]. In our algorithm, by means of Lemma 2, we specify the choice of ϵ_k such that (29) holds.

ALGORITHM 3 (A Smoothing Newton Method for SIP).

Initial *Given constants* $\alpha \in (0, 1/2)$, $\rho, \eta \in (0, 1)$ *and* $\sigma \in (0, (1-\alpha)/2)$. *Choose an initial point* $z_0 \in R^{n+(1+m+q)p}$ *and initial parameter* $\epsilon_0 \in (0, \frac{\alpha}{2\mu}\|H(z_0)\|)$.

Step 1 *Solve linear equation*

$$H(z_k) + \nabla H^k(z_k)p = 0. \tag{30}$$

Let p^k be a solution of (30).

Step 2 *Find the smallest nonnegative integer m such that the following inequality holds.*

$$\|H^k(z_k + \rho^m p^k)\|^2 - \|H^k(z_k)\|^2 \leqslant -2\sigma\rho^m \|H(z_k)\|^2. \tag{31}$$

Let m_k be the smallest nonnegative integer satisfying (31) and $\alpha_k = \rho^{m_k}$.

Step 3 *Let the next iterate be $z_{k+1} = z_k + \alpha_k p^k$.*

Step 4 *Stop if $H(z_{k+1}) = 0$.*

Step 5 *If $\|H(z_{k+1})\| \geqslant \alpha\|H(z_k)\| + \mu\alpha^{-1}\epsilon_k$, let $\epsilon_{k+1} = \epsilon_k$. Otherwise, choose a positive ϵ_{k+1} that satisfies*

$$\epsilon_{k+1} \leqslant \min\{\frac{1}{2}\epsilon_k,\ \xi^{k+1}\epsilon_k,\ \frac{\alpha}{2\mu}\|H(z_{k+1})\|\}. \tag{32}$$

Let $k := k + 1$ and go to Step 1.

Denote

$$K = \{0\} \cup \{k \mid \|H(z_k)\| < \alpha\|H(z_{k-1})\| + \mu\alpha^{-1}\epsilon_{k-1}\} \tag{33}$$

The following lemma summarizes some properties of the algorithm.

LEMMA 1 *Let $\{z_k\}$ and $\{\epsilon_k\}$ be generated by Algorithm 3. Then the following statements hold.*

(i) The positive sequence $\{\epsilon_k\}$ is nonincreasing and satisfies

$$\mu\epsilon_k \leqslant \alpha\|H(z_k)\|, \qquad \forall k = 1, 2, \ldots . \tag{34}$$

(ii) For every $k \in K$,

$$\epsilon_k \leqslant \frac{1}{2}\epsilon_{k-1}, \quad \beta_k \triangleq M\epsilon_k/\xi^k \leqslant \epsilon_{k-1},$$

and

$$dist(\nabla H^k(z_k),\ \partial_C H(z_k)) \leqslant \mu\beta_k \leqslant \mu\epsilon_{k-1}, \tag{35}$$

where $\mu = n + (1 + m + q)p$ and $M \geqslant \sup\{1, \|\nabla g(x, v)\|, \|\nabla c(v)\|\}$.

Proof. Inequality (34) is straightforward. We only prove (35).

Recall by (25) that

$$H^{\epsilon}(z) = \begin{pmatrix} L(z) \\ \Psi^{\epsilon}(z) \\ l(z) \\ \Phi^{\epsilon}(z) \end{pmatrix}, \qquad H(z) = H^{0}(z) = \begin{pmatrix} L(z) \\ \Psi^{0}(z) \\ l(z) \\ \Phi^{0}(z) \end{pmatrix}.$$

We verify (35) by showing that inequality

$$\operatorname{dist}\,(\nabla H_i^{\epsilon}(z_k), \partial H_i^{0}(z_k)) \leqslant \frac{M\epsilon_k}{2\xi^k} \leqslant \beta_k \tag{36}$$

holds for every i. For simplicity, we omit subscripts k in the proof.

For each $i = 1, 2, \ldots, p$, we have

$$\psi_i^0(z) = \phi_{FB}(u_i, -g(x, v^i)) = \sqrt{u_i^2 + g(x, v^i)^2} - (u_i - g(x, v^i)).$$

By a direct computation, we get

$$\partial_x \psi_i^0(z) = \begin{cases} \Big(1 + \dfrac{g(x, v^i)}{\sqrt{u_i^2 + g(x, v^i)^2}}\Big)\nabla_x g(x, v^i), & \text{if } u_i^2 + g(x, v^i)^2 \neq 0, \\ \{\rho \nabla_x g(x, v^i) \mid 0 \leqslant \rho \leqslant 2\}, & \text{if } u_i^2 + g(x, v^i)^2 = 0. \end{cases}$$

It is clear that if $u_i^2 + g(x, v^i)^2 = 0$, then $\nabla_x \psi_i^{\epsilon}(z) \in \partial_x \psi_i^0(z)$, which implies

$$\operatorname{dist}\,(\nabla \psi_i^{\epsilon}(z_k), \partial \psi_i^0(z_k)) = 0.$$

If $u_i^2 + g(x, v^i)^2 \neq 0$, we get

$$\begin{aligned} &\operatorname{dist}\,(\nabla_x \psi_i^{\epsilon}(z), \partial_x \psi_i^0(z)) = \|\nabla_x \psi_i^{\epsilon}(z) - \nabla_x \psi_i^0(z)\| \\ &= \Big|\frac{1}{\sqrt{u_i^2 + g(x, v^i)^2 + \epsilon^2}} - \frac{1}{\sqrt{u_i^2 + g(x, v^i)^2}}\Big| |g(x, v^i)| \|\nabla_x g(x, v^i)\| \\ &= \frac{\epsilon^2 |g(x, v^i)| \|\nabla_x g(x, v^i)\|}{\sqrt{u_i^2 + g(x, v^i)^2 + \epsilon^2}\sqrt{u_i^2 + g(x, v^i)^2}\Big(\sqrt{u_i^2 + g(x, v^i)^2 + \epsilon^2} + \sqrt{u_i^2 + g(x, v^i)^2}\Big)} \\ &\leqslant \frac{\|\nabla_x g(x, v^i)\| \epsilon}{2\sqrt{u_i^2 + g(x, v^i)^2}} \leqslant \frac{M\epsilon}{2\xi} \leqslant \beta. \end{aligned}$$

Similarly, we have

$$\operatorname{dist}(\nabla_u \psi_i^{\epsilon}(z), \partial_u \psi_i^0(z)) = 0,$$

if $u_i^2 + g(x, v^i)^2 = 0$, and

$$\operatorname{dist}\,(\nabla_u \psi_i^{\epsilon}(z), \partial_u \psi_i^0(z)) \leqslant \frac{\epsilon}{2\sqrt{u_i^2 + g(x, v^i)^2}} \leqslant \frac{M\epsilon}{2\xi} \leqslant \beta,$$

if $u_i^2 + g(x, v^i)^2 \neq 0$; and

$$\text{dist}(\nabla_{v^i}\psi_i^\epsilon(z), \partial_{v^i}\psi_i^0(z)) = 0,$$

if $u_i^2 + g(x, v^i)^2 = 0$, and

$$\text{dist}\,(\nabla_{v^i}\psi_i^\epsilon(z), \partial_{v^i}\psi_i^0(z)) \leqslant \frac{\|\nabla_{v^i}g(x, v^i)\|\epsilon}{2\sqrt{u_i^2 + g(x, v^i)^2}} \leqslant \frac{M\epsilon}{2\xi} \leqslant \beta,$$

if $u_i^2 + g(x, v^i)^2 \neq 0$.

The above argument shows that for each $i = 1, 2, \ldots, p$, we always have

$$\text{dist}\,(\nabla\psi_i^\epsilon(z_k), \partial\psi_i^0(z_k)) \leqslant \frac{M\epsilon_k}{2\xi^k} \leqslant \beta_k.$$

Similarly, we can deduce that for each $i = 1, 2, \ldots, p$ and each $j = 1, 2, \ldots q$, it holds that

$$\text{dist}\,(\nabla\phi_{ij}^\epsilon(z_k), \partial\phi_{ij}^0(z_k)) \leqslant \frac{\|\nabla c(v_k)\|\epsilon_k}{2\xi^k} \leqslant \frac{M\epsilon}{2\xi} \leqslant \beta.$$

For any other i, we obviously have $\nabla H_i^\epsilon(z) = \nabla H_i(z)$, which implies

$$\text{dist}\,(\nabla H_i^\epsilon(z_k), \partial H_i^0(z_k)) = 0 \leqslant \beta_k.$$

Summarizing the above discussion, it is not difficult to get (35). □

4. Global and Superlinear Convergence of the Smoothing Newton Method

In this section, we prove the global and superlinear convergence of Algorithm 3. To this end, we need the following assumptions.

ASSUMPTION 4. (*i*) *The level set*

$$\Omega \stackrel{\triangle}{=} \{z \mid \|H(z)\| \leqslant \frac{1+2\alpha}{1-2\alpha}\|H(z_0)\|\} \tag{37}$$

is bounded.

(*ii*) *The functions f, g and c are twice continuously differentiable on Ω.*

(*iii*) *For every $\epsilon \in (0, \epsilon_0)$, the matrix $\nabla H^\epsilon(z)$ is nonsingular for any $z \in \Omega$.*

Assumption 5 (iii) will be justified under further assumptions later in this section. We will also show that the iterate sequence $\{z_k\}$ generated by Algorithm 3 remains in the level set Ω.

The following lemma shows that the proposed method is well defined.

LEMMA 2 *Algorithm 3 is well defined.*

Proof. It suffices to show that for every k, the line search step terminates finitely.

$$\begin{aligned} H^k(z_k)^T \nabla H^k(z_k)^T p_k &= -H^k(z_k)^T H(z_k) = -\|H(z_k)\|^2 \\ &\quad + (H(z_k) - H^k(z_k))^T H(z_k) \\ &\leqslant -\|H(z_k)\|^2 + \mu\epsilon_k \|H(z_k)\| \leqslant -(1-\alpha)\|H(z_k)\|^2 \\ &\leqslant -\sigma \|H(z_k)\|^2. \end{aligned}$$

This shows that inequalities (31) are satisfied for all m sufficiently large, and hence the line search in Step 3 terminates finitely. □

4.1. GLOBAL CONVERGENCE

LEMMA 3. *Let $\{a_k\}$ and $\{\bar\epsilon_k\}$ be positive sequences such that for some constant $\alpha \in (0, 1)$*

$$a_k \leqslant \alpha a_{k-1} + \bar\epsilon_{k-1}, \quad k = 1, 2, \ldots. \tag{38}$$

If $\sum_{k=0}^{\infty} \bar\epsilon_k < \infty$, then $\{a_k\}$ converges to zero. If in addition, $\bar\epsilon_{k+1} \leqslant \rho\bar\epsilon_k$ for all k with some constant $\rho \in (0, 1)$, then we have for all $k > 0$

$$a_k \leqslant \begin{cases} \left(a_0 + \dfrac{1}{|\rho - \alpha|}\bar\epsilon_0\right) r^k, & \text{if } \rho \neq \alpha, \\ (a_0 + k\bar\epsilon_0 \alpha^{-1})\alpha^k, & \text{if } \rho = \alpha, \end{cases} \tag{39}$$

where $r = \max\{\alpha, \rho\}$.

Proof. By the assumption that $\sum_{k=0}^{\infty} \bar\epsilon_k < \infty$, it is not difficult to show from (38) that $\{a_k\}$ satisfies the Cauchy condition and hence converges. Taking limits in both sides of (38) yields $\lim_{k\to\infty} a_k = 0$. Suppose further that $\bar\epsilon_{k+1} \leqslant \rho\bar\epsilon_k$. It follows from (38) that for each $k > 0$

$$\begin{aligned} a_k &\leqslant \alpha a_{k-1} + \bar\epsilon_{k-1} \\ &\leqslant \alpha^2 a_{k-2} + \alpha\bar\epsilon_{k-2} + \bar\epsilon_{k-1} \\ &\leqslant \alpha^k a_0 + \sum_{j=0}^{k-1} \alpha^{k-j-1}\bar\epsilon_j \\ &\leqslant \alpha^k a_0 + \sum_{j=0}^{k-1} \alpha^{k-j-1}\rho^j \bar\epsilon_0 \\ &= \begin{cases} \alpha^k a_0 + \alpha^{k-1}\dfrac{1-(\rho/\alpha)^k}{1-(\rho/\alpha)}\bar\epsilon_0, & \text{if } \rho \neq \alpha, \\ (a_0 + k\bar\epsilon_0\alpha^{-1})\alpha^k, & \text{if } \rho = \alpha. \end{cases} \end{aligned}$$

The last equality yields (39). □

LEMMA 4. *If the index set K defined by* (33) *is infinite, then the sequences* $\{\epsilon_k\}$, $\{\beta_k\}$ *and* $\{H(z_k)\}$ *converge to zero. Moreover, we have*

$$\lim_{k\to\infty, k\in K} dist(\nabla H^k(z_k), \partial_C H(z_k)) = 0. \tag{40}$$

Proof. That the sequence $\{\epsilon_k\}$ converges to zero follows from Lemma 3 (i) and (ii) immediately. Equality (40) then follows from (35). It remains to verify that $H(z_k)$ converges to zero. Let K consist of $k_0 = 0 < k_1 < k_2 < \cdots$. By Step 5, we have $\epsilon_k = \epsilon_{k_j}$ for each k satisfying $k_j \leqslant k < k_{j+1}$, which in turn, implies $H^k(z) = H^{k_j}(z)$ for any z and every k satisfying $k_j \leqslant k < k_{j+1}$. By Step 2, we have for every $k_j \leqslant k < k_{j+1}$

$$\|H^k(z_k)\| = \|H^{k_j}(z_k)\| \leqslant \|H^{k_j}(z_{k-1})\| \leqslant \cdots \leqslant \|H^{k_j}(z_{k_j})\|. \tag{41}$$

We then get

$$\begin{aligned}\|H(z_{k_{j+1}})\| &\leqslant \alpha\|H(z_{k_{j+1}-1})\| + \mu\alpha^{-1}\epsilon_{k_{j+1}-1}\\ &\leqslant \alpha\|H^{k_j}(z_{k_{j+1}-1})\| + \alpha\|H^{k_j}(z_{k_{j+1}-1}) - H(z_{k_j})\| + \mu\alpha^{-1}\epsilon_{k_j}\\ &\leqslant \alpha\|H^{k_j}(z_{k_j})\| + \mu(\alpha+\alpha^{-1})\epsilon_{k_j}\\ &\leqslant \alpha\|H(z_{k_j})\| + \mu(2\alpha+\alpha^{-1})\epsilon_{k_j}.\end{aligned} \tag{42}$$

Let $a_j = \|H(z_{k_j})\|$ and $\bar{\epsilon}_j = \mu(2\alpha+\alpha^{-1})\epsilon_{k_j}$. Then we have $\bar{\epsilon}_j \leqslant \frac{1}{2}\bar{\epsilon}_{j-1}$ for all $j \geqslant 1$. By using Lemma 6, we claim that the subsequence $\{\|H(z_{k_j})\|\}$ converges to zero.

For any k, let k_j be the largest index in K such that $k_j \leqslant k$. It then follows from (41) that

$$\|H(z_k)\| \leqslant \|H^{k_j}(z_k)\| + \mu\epsilon_{k_j} \leqslant \|H^{k_j}(z_{k_j})\| + \mu\epsilon_{k_j}. \tag{43}$$

This implies that the sequence $\{H(z_k)\}$ converges to zero. □

The next theorem establishes the global convergence of Algorithm 3.

THEOREM 5. *The sequence* $\{z_k\}$ *generated by Algorithm 3 is contained in* Ω. *Moreover, we have*

$$\lim_{k\to\infty} H(z_k) = 0. \tag{44}$$

Consequently, every accumulation point of $\{z_k\}$ *is a solution of* $H(z) = 0$.

Proof. For any $k > 0$, let k_j be the largest index in K such that $k_j \leqslant k$. Note that $\alpha < 1/2$. From Lemma 6 and (42), we have for any $k \geqslant k_j$,

$$\begin{aligned}\|H(z_{k_j})\| &\leqslant \Big(\|H(z_0)\| + \frac{1}{1/2-\alpha}\mu(2\alpha+\alpha^{-1})\epsilon_0\Big)(\frac{1}{2})^j\\ &\leqslant \Big(1 + \frac{\alpha}{(1/2-\alpha)2\mu}\mu(2\alpha+\alpha^{-1})\Big)(\frac{1}{2})^j\|H(z_0)\|\\ &= \Big(1 + \frac{1+2\alpha^2}{1-2\alpha}\Big)(\frac{1}{2})^j\|H(z_0)\|,\end{aligned}$$

where the second inequality follows from the choice of ϵ_0. This together with (43) implies

$$\begin{aligned}\|H(z_k)\| &\leqslant \Big(1+\frac{1+2\alpha^2}{1-2\alpha}\Big)(\frac{1}{2})^j\|H(z_0)\|+(\frac{1}{2})^j\mu\epsilon_0\\ &\leqslant \Big(1+\frac{1+2\alpha^2}{1-2\alpha}+\frac{\alpha}{2}\Big)(\frac{1}{2})^j\|H(z_0)\|\\ &\leqslant \Big(1+\frac{1+2\alpha}{1-2\alpha}\Big)(\frac{1}{2})^j\|H(z_0)\|.\end{aligned}$$

This shows that $\{z_k\}\subset\Omega$.

We turn to verifying (44). By Lemma 7, it suffices to show that the index set K defined by (33) is infinite. For the sake of contradiction, we assume that K is finite. By Algorithm 1, there is an index $\bar{k}$ such that $\epsilon_k=\epsilon_{\bar{k}}\overset{\triangle}{=}\bar{\epsilon}$ for all $k\geqslant\bar{k}$. Let $\bar{H}=H^{\bar{\epsilon}}$. It then follows from Step 5 of Algorithm 1 that $\|H(z_k)\|\geqslant\alpha\|H(z_{k-1})\|+\mu\bar{\epsilon}\geqslant\mu\bar{\epsilon}>0$ holds for all $k\geqslant\bar{k}$.

By the line search condition (31), we have

$$\alpha_k\|H(z_k)\|^2\leqslant\|\bar{H}(z_k)\|^2-\|\bar{H}(z_{k+1})\|^2.$$

Summarizing these inequalities yields that $\alpha_k\|H(z_k)\|^2\to 0$ as $k\to\infty$. However, the sequence $\{\|H(z_k)\|\}$ is bounded away from zero. We claim that the sequence $\{\alpha_k\}$ goes to zero as k goes to infinity. By the line search rule, when k is sufficiently large, inequality (31) does not hold for $\alpha'_k\overset{\triangle}{=}\alpha_k/\rho$. That is, the following inequality holds for all k sufficiently large.

$$\|\bar{H}(z_k+\alpha'_k p^k)\|^2-\|\bar{H}(z_k)\|^2>-2\sigma\alpha'_k\|H(z_k)\|^2. \tag{45}$$

Since $\{z_k\}$ is bounded, there is a subsequence $\{z_k\}_{k\in K_1}$ converging to some point $\bar{z}\in\Omega$. By the singularity assumption of $\nabla\bar{H}(\bar{z})$, the linear equation (30) implies that the sequence $\{p_k\}_{k\in K_1}$ is also bounded. Without loss of generality, we may assume that $\{p_k\}_{k\in K_1}$ converges to some vector $\bar{p}$. By taking limits in both sides of (30) as k goes to infinity with $k\in K_1$, we get

$$\nabla\bar{H}(\bar{z})\bar{p}=-H(\bar{z}). \tag{46}$$

Dividing by α'_k and then letting k go to infinity with $k\in K_1$ in both sides of (45), we get

$$\bar{H}(\bar{z})^T\nabla\bar{H}(\bar{z})\bar{p}\geqslant-2\sigma\|H(\bar{z})\|^2.$$

The last inequality together with (46) implies

$$\bar{H}(\bar{z})^T H(\bar{z})\leqslant 2\sigma\|H(\bar{z})\|^2.$$

Further analyzing the last inequality, we deduce

$$2\sigma\|H(\bar{z})\|^2 \geqslant \|H(\bar{z})\|^2 - H(\bar{z})^T(H(\bar{z}) - \bar{H}(\bar{z})) \geqslant \|H(\bar{z})\|^2 - \mu\bar{\epsilon}$$
$$\|H(\bar{z})\| \geqslant (1-\alpha)\|H(\bar{z})\|^2.$$

Since $\sigma \in (0, (1-\alpha)/2)$, the last inequality implies $H(\bar{z}) = 0$, which is a contradiction. The contradiction shows that K must be infinite. The proof is then complete. □

Theorem 5 shows that under appropriate conditions, every accumulation point of $\{z_k\}$ is a solution of $H(z) = 0$. If some additional assumptions are further assumed, the whole sequence $\{z_k\}$ converges to that solution. Moreover, the convergence rate is superlinear. These are the goals of the next subsection.

4.2. SUPERLINEAR CONVERGENCE

THEOREM 6. *Let $\{z_k\}$ be generated by Algorithm* 1. *If there is an accumulation point z^* of $\{z_k\}_{k\in K}$ at which every matrix of $\partial_C H(z^*)$ is nonsingular, then the whole sequence $\{z_k\}$ converges to z^*. Moreover, the convergence is superlinear. If in addition, the functions f, g and c are twice Lipschitz continuously differentiable, then the convergence rate is quadratic.*

Proof. Let $\{z_k\}_{k\in K_0} \subset \{z_k\}_{k\in K}$ converge to z^*. By Lemma 7 (ii) and the assumption that every matrix in $\partial_C H(z^*)$ is nonsingular, when k is sufficiently large, $\nabla H^k(z_k)$ is nonsingular and there is a constant $M > 0$ such that $\|\nabla H^k(z_k)^{-1}\| \leqslant M$ holds for all $k \in K_0$ sufficiently large. Let $V_k \in \partial_C H(z_k)$ satisfy $\mathrm{dist}(\nabla H^k(z_k), \partial_C H(z_k)) = \|\nabla H^k(z_k) - V_k\|$. We get for all k sufficiently large

$$\begin{aligned}\|z_k + p_k - z^*\| &= \|\nabla H^k(z_k)^{-1}(H(z_k) - \nabla H^k(z_k)(z_k - z^*))\| \\ &\leqslant M\Big(\|H(z_k) - H(z^*) - V_k(z_k - z^*)\| + \\ &\quad \|(V_k - \nabla H^k(z_k))(z_k - z^*)\|\Big) \\ &= o(\|z_k - z^*\|). \end{aligned} \tag{47}$$

Again, by the nonsingularity of elements in $\partial_C H(z^*)$, there is a constant $\kappa > 0$ such that $\|H(z_k)\| \geqslant \kappa\|z_k - z^*\|$. On the other hand, the Lipschitz continuity of H implies $\|H(z_k + p_k)\| = O(\|z_k + p_k - z^*\|) = o(\|z_k - z^*\|) = o(\|H(z_k)\|)$. By Step 5 and the definition of K, $\mu\epsilon_k \leqslant \frac{1}{2}\alpha\|H(z_k)\|$ for all $k \in K_0 \subset K$, which implies

$$\|H^k(z_k)\| \geqslant \|H(z_k)\| - \mu\epsilon_k \geqslant (1 - \frac{1}{2}\alpha)\|H(z_k)\|.$$

Therefore, we have

$$\begin{aligned}\|H^k(z_k+p_k)\|^2 - \|H^k(z_k)\|^2 &\leqslant \Big(\|H(z_k+p_k)\| + \mu\epsilon_k\Big)^2 - (1-\frac{1}{2}\alpha)^2\|H(z_k)\|^2 \\ &\leqslant \Big(o(\|H(z_k)\|) + \frac{1}{2}\alpha\|H(z_k)\|\Big)^2 - (1-\frac{1}{2}\alpha)^2 \\ &\quad \|H(z_k)\|^2 \\ &= -\Big((1-\frac{1}{2}\alpha)^2 - \frac{\alpha^2}{4}\Big)\|H(z_k)\|^2 + o(\|H(z_k)\|^2) \\ &= -(1-\alpha)\|H(z_k)\|^2 + o(\|H(z_k)\|^2).\end{aligned}$$

Since $\sigma \in (0,\ (1-\alpha)/2)$, the last equality implies that when $k \in K_0$ is sufficiently large, the unit step is always accepted. Therefore, by the fact that $\|H(z_k+p_k)\| = o(\|H(z_k)\|)$, we have $\|H(z_{k+1})\| = o(\|H(z_k)\|)$. Consequently, $\|H(z_{k+1})\| \leqslant \rho\|H(z_k)\| \leqslant \rho\|H(z_k)\| + \mu\alpha^{-1}\epsilon_k$. This shows that $k+1 \in K_0$. By means of the induction principle, we claim that the index set K_0 contains all except finitely many indices. Thus the sequence $\{z_k\}$ converges to z^*. The superlinear convergence follows from (47).

If f, g and c are twice Lipschitz continuously differentiable, then H is strongly semismooth. In a way similar to (47), we can deduce that there is a constant $\bar{m} > 0$ such that the inequality $\|z_k + p_k - z^*\| \leqslant \bar{M}\|z_k - z^*\|^2$ holds for all $k \in K_0$ sufficiently large. Therefore, the convergence rate is quadratic. □

4.3. REGULARITY

We conclude this section by giving a sufficient condition for $\nabla H^\epsilon(z)$ to be nonsingular. For $z = (x, u, v, w) \in R^{n+(1+m+q)p}$, let $Q = \{1, 2, \cdots, q\}$ and $P = \{1, 2, \cdots, p\}$. We make the following assumptions:

ASSUMPTIN 5 *(i) $\nabla_x L(z)$ is positive semidefinite. Moreover, it is positive definite in the null space of Span$(\nabla_x g(x, v)^T)$. That is, $d^T\nabla_x L(z)d > 0$ for all $d \in R^n$ satisfying $\nabla_x g(x, v)^T d = 0$.*
(ii) $\nabla_v l_i(z)$ is positive semidefinite. Moreover, it is positive definite in the null space of Span$(\nabla c(v^i)^T)$. That is $d^T M_{ii}(z)d > 0$ for all d satisfying $\nabla c(v^i)^T d = 0$.

Note that Assumption 5 (i) and (ii) are stronger than the second order optimality conditions for the first and the second level programs, respectively. However, both conditions do not need the linear independence of active constraint gradients.

The following Theorem shows that Assumption 5 is sufficient for $\nabla H^\epsilon(z)$ to be nonsingular.

THEOREM 7. *Let Assumption 5 hold. Then $\nabla H^\epsilon(z)$ is nonsingular for every $\epsilon > 0$.*

Proof. From the expression of $\nabla H^\epsilon(z)$, it suffices to show that the following matrix $Q^\epsilon(z)$ is nonsingular.

$$Q^\epsilon(z) \triangleq \begin{pmatrix} \nabla_x L(z) & -\nabla_x g(x,v) & -\nabla^2_{xv} g(x,v)\Lambda_2(z) & \mathbf{0} \\ \nabla_x g(x,v)^T & \mathrm{diag}(a_i^\epsilon(z)^{-1} b_i^\epsilon(z)) & -\mathrm{diag}(\nabla_v g(x,v^i)^T) & \mathbf{0} \\ \Lambda_2(z)\nabla^2_{xv} g(x,v)^T & -\mathrm{diag}(\nabla_v g(x,v^i)) & M(z) & -\mathrm{diag}(\nabla c(v^i)) \\ \mathbf{0} & \mathbf{0} & -\mathrm{diag}\left(\nabla c(v^i)^T\right) & \mathrm{diag}(\beta_{ij}^\epsilon(z)^{-1}\gamma_{ij}^\epsilon(z)) \end{pmatrix}.$$

Let $d = (d_1, d_2, d_3, d_4)$ be a solution of $Q^\epsilon(z)d = 0$. This means

$$\begin{cases} \nabla_x L(z)d_1 - \nabla_x g(x,v)d_2 - \nabla^2_{xv} g(x,v)\Lambda_2(z)d_3 = 0, \\ \nabla_x g(x,v)^T d_1 + \mathrm{diag}(a_i^\epsilon(z)^{-1} b_i^\epsilon(z))d_2 - \mathrm{diag}(\nabla_v g(x,v^i)^T)d_3 = 0, \\ \Lambda_2(z)\nabla^2_{xv} g(x,v)^T d_1 - \mathrm{diag}(\nabla_v g(x,v^i))d_2 + M(z)d_3 - \mathrm{diag}(\nabla c(v^i))d_4 = 0, \\ -\mathrm{diag}\left(\nabla c(v^i)^T\right)d_3 + \mathrm{diag}(\beta_{ij}^\epsilon(z)^{-1}\gamma_{ij}^\epsilon(z))d_4 = 0. \end{cases} \tag{48}$$

It then follows that

$$d_1^T \nabla_x L(z)d_1 + d_2^T \mathrm{diag}(a_i^\epsilon(z)^{-1} b_i^\epsilon(z))d_2 + d_3^T \nabla_v l(z)d_3 + d_4^T \mathrm{diag}(\beta_{ij}^\epsilon(z)^{-1}\gamma_{ij}^\epsilon(z))d_4 = 0. \tag{49}$$

Since $\nabla_x L(z)$ and $\nabla_v l(z)$ are positive semidefinite, and $\mathrm{diag}(a_i^\epsilon(z)^{-1} b_i^\epsilon(z))$ and $\mathrm{diag}(\beta_{ij}^\epsilon(z)^{-1}\gamma_{ij}^\epsilon(z))$ are positive definite, we get $d_1^T \nabla_x L(z)d_1 = 0$, $d_3^T \nabla_v l(z)d_3 = 0$, $d_2 = 0$ and $d_4 = 0$. Therefore, we get from (48)

$$\begin{cases} \nabla_x g(x,v)^T d_1 - \mathrm{diag}(\nabla_v g(x,v^i)^T)d_3 = 0, \\ \mathrm{diag}\left(\nabla c(v^i)^T\right)d_3 = 0. \end{cases} \tag{50}$$

The last equation of (50) together with Assumption 5 (ii) implies $d_3 = 0$. It then follows from the first equation of (50) and Assumption 5 (i) that $d_1 = 0$. This shows that zero is the unique solution of $Q^\epsilon(z)d = 0$. Consequently, $Q^\epsilon(z)$ is nonsingular. The proof is complete. □

A semismooth function F is said to be CD-regular at z if every element in $\partial F(x)$ is nonsingular. If z is a solution of (26), Qi et al. [27] gave a sufficient condition for H to be CD-regular at z. We give the conditions and conclusion as follows. For details, see [27].

ASSUMPTION 6 *(i)* $u_i > 0 \quad \forall i = 1, 2, \ldots, p$.
(ii) The vectors $\nabla_x g(x, v^i)$, $i = 1, 2, \ldots, p$ *are linearly independent.*
(iii) For each $i = 1, 2, \ldots, p$, *the vectors* $\nabla c_j(v^i)$, $j \in I(v^i) \triangleq \{j : c_j(v^i) = 0\}$ *are linearly independent.*
(iv) $w_j^i - c_j(v^i) \neq 0$, $\forall i \in P$ *and* $j \in Q$.
(v) For all $(d, \xi_1, \ldots, \xi_p) \in G(x, v)\backslash\{0\}$,

$$d^T \nabla_x L(z)d + \sum_{i=1}^{p} \xi_i^T \nabla_v l(x, u_i, v^i, w^i)\xi_i > 0,$$

where $G(x, v)$ be the set of all $(d, \xi_1, \cdots, \xi_p) \in R^n \times R^{mp}$ satisfying

$$d^T \nabla_x g(x, v^i) - \xi_i^T \nabla_v g(x, v^i) = 0 \text{ for } i \in P,$$

and

$$\xi_i^T \nabla c_j(v^i) = 0 \text{ for } i \in P,\ j \in I(v^i).$$

The following theorem comes from [27].

THEOREM 8. *Suppose that $z(x, u, v, w)$ is a solution of (26) and satisfies the conditions of Assumption 6. Then H is CD-regular at z.*

5. Numerical Results

We realize that the parameter p depends upon the problem. In implementing the algorithm, we consider the simple case in which p is known. This is in fact rather common in applications. For instance, if Assumption 1 holds and for any fixed x, $g(x, \cdot)$ is a concave function. Then $p = 0$ or 1. Even when p is unknown but small, say $p = 2$ or 3, we may try $p = 1$ first. If it fails, we may then try $p = 2$, and so on. In the following, we shall first give three examples with $p = 1$, and one example with $p = 2$. (The case that p is unknown and not small will be studied in future research.)

To illustrate the computational behavior of the proposed algorithm in Section 3, it was then implemented in MATLAB (Version 6.0.0.88 Realse 12) and run on a COMPUCON PC (Pentium II/300 64 MB/4.3 GB) for the following examples from [3]. These examples were also being used in [27]. Throughout the computational experiments, the parameters used in the algorithm were $\rho = 0.5$, $\alpha = 0.05$, $\epsilon_0 = 10^{-8}$, and $\sigma = 10^{-4}$. We will terminate our iteration when $\|H(z^k)\| < 10^{-6}$. The numerical results are summarized in Table 1, where *Iter.* denotes the number of iterations, *Fun. Evalu.* the number of function evaluations, x^k and v^k the final iterate and $f(x^k)$ the function value of f at the final iterate x^k.

EXAMPLE 7 $f(x) = 2.25 \exp(x_1) + \exp(x_2)$, $g(x, v) = v - \exp(x_1 + x_2)$, $V = [0, 1]$.$p = 1$, $x^0 = (-1.5, 0)^T$, $v^0 = 0$.

EXAMPLE 8 $f(x) = x_1^2 + x_2^2 + x_3^2$, $g(x, v) = x_1 + x_2 \exp(x_3 v) + \exp(2v) - 2\sin(4v)$, $V = [0, 1]$. $p = 1$, $x^0 = (0, 0, -0.5)^T$, $v^0 = 0$.

EXAMPLE 9 $f(x) = (x_1 - 2x_2 + 5x_2^2 - x_2^3 - 13)^2 + (x_1 - 14x_2 + x_2^2 + x_2^3 - 29)^2$, $g(x, v) = x_1^2 + 2x_2 v^2 + \exp(x_1 + x_2) - \exp(v)$, $V = [0, 1]$. $p = 1$, $x^0 = (0, -2)^T$, $v^0 = 0$.

EXAMPLE 10 $f(x) = \frac{1}{3}x_1^2 + \frac{1}{2}x_1 + x_2^2$, $g(x, v) = (1 - x_1^2 v^2)^2 - x_1 v^2 - x_2^2 + x_2$, $V = [0, 1]$. $p = 2$, $x^0 = (0, 0)^T$, $v_1^0 = 0.5$, $v_2^0 = 1$.

Table 1. Results for Algorithm 3

Example	p	Iter.	Fun. Evalu.	x^k	$f(x^k)$
1	1	3	22	(-4.05e-01,4.05e-01)	3.00e+00
2	1	5	39	(-2.13e-01,-1.36e+00,1.85e+00)	5.33e+00
3	1	3	26	(7.20e-01,-1.45e+00)	9.72e+01
4	2	2	22	(-7.50e-01,-6.18e-01)	1.94e-01

6. A Special Case

This section considers a special case where V is an closed interval in R. Without loss of generality, let $V = [0, 1]$. SIP with this V has been studied by many authors (see [13] and references therein). It is obvious that Algorithm 3 can be applied to solve this problem directly. Taking into account the particular structure of V, we are going to develop a smoothing Newton method that is particularly useful for this problem.

We rewrite the box constraint $t \in [0, 1]$ in two inequality constraints $t \leqslant 1$ and $-t \leqslant 0$. Since these two constraint functions are linear, if $g(x, \cdot)$ is concave, the KKT points of (10) are global solutions without additional assumption on the inner problem. When $V = [0, 1]$, the KKT system of (10) is as follows.

$$\begin{cases} -g_t'(x, t) + w_1 - w_2 = 0, \\ \min\{w_1,\ 1 - t\} = 0, \\ \min\{w_2,\ -t\} = 0. \end{cases} \tag{51}$$

Consequently, the KKT system of (1) is written as

$$\begin{cases} \nabla f(x) + \sum_{i=1}^{p} u_i \nabla_x g(x, t_i) = 0, \\ u_i \geqslant 0, \quad g(x, t_i) \leqslant 0, \ \text{ for } i = 1, \ldots, p, \\ u_i g(x, t_i) = 0, \ \text{ for } i = 1, \ldots, p, \\ -g_t'(x, t_i) + w_1^i - w_2^i = 0, \ \text{ for } i = 1, \ldots, p, \\ \min\{w_1,\ 1 - t_i\} = 0, \ \text{ for } i = 1, \ldots, p, \\ \min\{w_2,\ -t_i\} = 0, \ \text{ for } i = 1, \ldots, p. \end{cases} \tag{52}$$

By using the mid function, system (51) can be rewritten as a compact form

$$\phi(x, t) \overset{\triangle}{=} \text{mid}\{t,\ -g_t'(x, t),\ t - 1\} = 0, \tag{53}$$

where $\text{mid}\{a, b, c\}$ denotes the median value of scalars a, b and c. Therefore, the KKT system (52) can be reformulated as a system of nonlinear equations.

$$\begin{cases} \nabla f(x) + \sum_{i=1}^{p} u_i \nabla_x g(x, t_i) = 0, \\ \phi_{FB}(u_i,\ -g(x, t_i)) = 0, \ \text{ for } i = 1, \ldots, p, \\ \phi(x, t_i) = 0, \ \text{ for } i = 1, \ldots, pn, \end{cases} \tag{54}$$

where ϕ_{FB} is the Fischer–Burmeister function. Compared with (52), system (54) has less equations.

The function ϕ is also strongly semismooth if g is twice continuously differentiable as a function of t, and hence Equation (54) is a system of semismooth equations. It is then not difficult to develop a smoothing Newton method in a way similar to previous section.

Let $\rho : R \to R$ be a continuous density function with a bounded absolute mean, that is,

$$\int_{-\infty}^{\infty} |s|\rho(s)ds < \infty.$$

Then the Gabriel-Moré smoothing function is defined by

$$\phi_{GM}^{\epsilon}(x, t) = \int_{-\infty}^{\infty} \text{mid}\{t,\ t-1,\ g_t'(x, t) - \epsilon^2 s\}\rho(s)ds. \tag{55}$$

The Gabriel-Moré smoothing function contains, as a special case, the Chen-Harker-Kanzow-Smale smoothing function, which corresponds to the density function

$$\rho(s) = \frac{1}{(s^2+1)^{3/2}}.$$

It has been shown that for every $\epsilon > 0$, $\phi_{GM}^{\epsilon}(x, \cdot)$ is continuously differentiable everywhere. Moreover, the function ϕ_{GM}^{ϵ} possesses similar properties as those of ϕ_{FB}^{ϵ}. There is no difficulty to propose a smoothing Newton method for solving (54) by using the Gabriel–Moré smoothing function or the Chen–Harker–Kanzow–Smale smoothing function. The related global and superlinear convergence of the method can be proved in a way similar to Section 4. We omit the details.

7. Acknowledgements

The work was partially supported by The Hong Kong Polytechnic University Postdoctoral Fellowship and the Research Grant Council of Hong Kong. This work was initiated when Liqun Qi was visiting National Cheng-Kung University. That trip was supported by the National Sciences Council of Taiwan.

References

1. Chen, X., Qi L. and Sun, D. (1998), Global and superlinear convergence of the smoothing Newton method and its application to general box constrained variational inequalities, *Mathematics of Computation* 67, 519–540.
2. Clarke, F.H. (1983), *Optimization and Nonsmooth Analysis*, Wiley, New York.
3. Coope, I.D. and Watson, G.A. (1985), A projected Lagrangian algorithm for semi-infinite programming, *Mathematical Programming* 32, 337–356.
4. Coope, I.D. and Price, C.J. (1998), Exact penalty function methods for nonlinear semi-infinite programming. In: Reemtsen, R. and Rückmann, J. (eds.), *Semi-Infinite Programming*, Kluwer Academic Publishers, Boston, pp. 137–157.

5. De Luca, T., Facchinei, F. and Kanzow, C. (1996), A semismooth equation approach to the solution of nonlinear complementarity problems, *Mathematical Programming* 75, 407–439.
6. Fischer, A. (1997), Solution of monotone complementarity problems with locally Lipschitzian functions, *Mathematical Programming* 76, 513–532.
7. Fukushima, M. and Qi, L. (1998), *Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*, Kluwer Academic Publishers, Boston, MA.
8. Gramlich, G., Hettich, R. and Sachs, E. (1995), Sqp-methods in semi-infinite programming, *SIAM Journal on Optimization* 5, 641–658.
9. Hettich, R.and Jongen, H. Th. (1978), Semi-infinite programming: conditions of optimality and applications, *Lecture Notes in Control and Information Sciences* 7, 1–11.
10. Hettich, R. and Kortanek, K. (1993), Semi-infinite programming: theory, methods and applications, *SIAM Review* 35, 380–429.
11. Janin, R. (1984), Direction derivative of the marginal function in nonlinear programming, *Mathematical Programming Study* 21, 127–138.
12. Kanzow, C. and Pieper, H. (1999), Jacobian smoothing methods for general complementarity problems, *SIAM Journal on Optimization* 9, 342–373.
13. C.T. Lawrence and Tits, A.L. (1998), Feasible sequential quadratic programming for finely discretized problems from SIP. In: Reemtsen, R. and Rückmann, J. (eds.), *Semi-Infinite Programming*, Kluwer Academic Publishers, Boston, pp. 159–193.
14. Li, D.H. and Fukushima, M. (2000), Smoothing Newton and quasi-Newton methods for mixed complementarity problems, *Computational Optimization and Applications* 17, 203–230.
15. Mangasarian, O.L. (1969), *Nonlinear Programming*, McGraw-Hill, New York.
16. Pang, J.-S. and Qi, L. (1993), Nonsmooth equations: Motivation and algorithms, *SIAM Journal on Optimization* 3, 443–465.
17. Polak, E. (1997), *Optimization: Algorithms and Consistent Approximation*, Springer, New York.
18. Polak, E. and Tits, A. (1982), A recursive quadratic programming algorithm for semi-infinite optimization problems, *Applied Mathematics and Optimization* 8, 325–349.
19. Qi, L. (1993), Convergence analysis of some algorithms for solving nonsmooth equations, *Mathematics of Operations Research* 18, 227–244.
20. Qi, L. and Chen, X. (1995), A globally convergent successive approximation method for several nonsmooth equations, *SIAM Journal on Control and Optimization* 33, 402–418.
21. Qi, L. (1999), Regular pseudo-smooth NCP and BVIP functions and globally and quadratically convergent generalized Newton methods for complementarity and variational inequality problems, *Mathematics of Operations Research* 24, 440–471.
22. Qi, L. and Jiang, H. (1997), Semismooth Karush-Kuhn-Tucker equations and convergence analysis of Newton methods and quasi-Newton methods for solving these equations, *Mathematics of Operations Research* 22, 301–325.
23. Qi, L. and Sun, D. (1999), A survey of some nonsmooth equations and smoothing Newton methods, in Eberhard, A., Glover, B., Hill, R. and Ralph, D. (eds.), *Progress in Optimization: Contributions from Australasia*, Kluwer Academic Publishers, Boston, 121–146.
24. Qi, L., Sun, D. and Zhou, G. (2000), A new look at smoothing Newton methods for nonlinear complementarity problems and box constrained variational inequalities, *Mathematical Programming* 87, 1–35.
25. Qi, L. and Sun, J. (1993), A nonsmooth version of Newton's method, *Mathematical Programming* 58, 353–367.
26. Qi, L. and Wei, Z.(2000), On the constant positive linear dependence condition and its applications to SQP methods, *SIAM Journal on Optimization* 10, 963–981 Corrigendum, *SIAM Journal on Optimization* 11 (2001) 1145–1146.
27. Qi, L., Wu, S.-Y. and Zhou, G., Semismooth Newton for solving semi-infinite programming problems, to appear in: *Journal of Global Optimization*.

28. Reemtsen, R. and Görner, S. (1998), Numerical methods for semi-infinite programming: a survey. In Reemtsen, R. and Gc̈kmann, J. (eds.), *Semi-Infinite Programming*, Kluwer Academic Publishers, Boston, MA, pp. 195-275.
29. Reemsten, R. and Rückmann, J. (1998), *Semi-Infinite Programming*, Kluwer Academic Publishers, Boston, MA.
30. Rückmann, J.J. and Shapiro, A. (2001), Second-order optimality conditions in generalized semi-infinite programming, *Set-Valued Analysis* 9, 169–186.
31. Shapiro, A. (1998), First and second order optimality conditions and perturbation analysis of semi-infinite programming problems. In: Reemsten, R. and Rückmann, J. (eds.), *Semi-Infinite Programming*, Kluwer Academic Publishers, Boston, MA, pp. 103–133.
32. Tanaka, Y., Fukushima, M. and Ibaraki, T. (1988), A globally convergent SQP method for semi-infinite nonlinear optimization, *Journal of Computational and Applied Mathematics* 23, 141–153.
33. Teo, K.L., Yang, X.Q. and Jennings, L.S. (2000), Computational discretization algorithms for functional inequality constrained optimization, *Annals of Operations Research* 98, 215–234.
34. Vo, B., Cantoni, A. and Teo, K.L. (1997), Envelope constrained filter with linear interpolator, *IEEE Transactions on Signal Processing* 45, 1405–1414.
35. Watson, G.A. (1981), Globally convergent methods for semi-infinite programming, *BIT* 21, 362–373.
36. Xiu, N. and Zhang, J. (2001), A smoothing Gauss-Newton method for the generalized HLCP, *Journal of Computational and Applied Mathematics* 129, 195–208.
37. Yamashita, N. and Fukushima, M. (1997), Modified Newton methods for solving semismooth reformulations of monotone complementarity problems, *Mathematical Programming* 76, 469–491.