Journal of Computational Mathematics, Vol. 24, No. 6, 2006, 719–732.

ON KARUSH-KUHN-TUCKER POINTS FOR A SMOOTHING METHOD IN SEMI-INFINITE OPTIMIZATION *1)

Oliver Stein
(*Department of Mathematics-C, RWTH Aachen University, Germany*)

Abstract

We study the smoothing method for the solution of generalized semi-infinite optimization problems from (O. Stein, G. Still: Solving semi-infinite optimization problems with interior point techniques, SIAM J. Control Optim., 42(2003), pp. 769–788). It is shown that Karush-Kuhn-Tucker points of the smoothed problems do not necessarily converge to a Karush-Kuhn-Tucker point of the original problem, as could be expected from results in (F. Facchinei, H. Jiang, L. Qi: A smoothing method for mathematical programs with equilibrium constraints, Math. Program., 85(1999), pp. 107–134). Instead, they might merely converge to a Fritz John point. We give, however, different additional assumptions which guarantee convergence to Karush-Kuhn-Tucker points.

Mathematics subject classification: 90C34, 90C25, 49M37, 65K05.
Key words: Generalized semi-infinite optimization, Stackelberg game, Constraint qualification, Smoothing, NCP function.

1. Introduction

This article studies a numerical solution method for so-called generalized semi-infinite optimization problems. These problems have the form

$$GSIP: \qquad \text{minimize} \quad f(x) \quad \text{subject to} \quad x \in M$$

with

$$M \;=\; \{\; x \in \mathbb{R}^n |\; g_i(x,y) \le 0 \quad \text{for all } y \in Y(x),\; i \in I \;\}$$

and

$$Y(x) \;=\; \{\; y \in \mathbb{R}^m |\; v_\ell(x,y) \le 0,\; \ell \in L \;\}.$$

All defining functions $f, g_i\,,\; i \in I = \{1,...,p\},\; v_\ell\,,\; \ell \in L = \{1,...,s\}$, are assumed to be real-valued and d times continuously differentiable on their respective domains with $d \ge 2$. The inclusion of equality constraints in the definitions of M and $Y(x)$ as well as of $i-$dependent index sets $Y(x)$ is straightforward and will not be considered here for the ease of presentation.

As opposed to a standard semi-infinite optimization problem *SIP*, the possibly infinite index set $Y(x)$ of inequality constraints is x-dependent in a *GSIP*. For surveys about *standard* semi-infinite optimization we refer to [6, 8, 17, 18], whereas the state of the art in *generalized* semi-infinite optimization is covered in [26, 27, 28] and in the monography [24]. The latter also contains a wide range of applications and the historical background of generalized semi-infinite programming.

* Received November 19, 2005.
1) This work was supported by a Heisenberg grant of the *Deutsche Forschungsgemeinschaft*.

A numerical solution method for a subclass of these problems was presented in [26]. It bases on a smoothing method which is also known from [2] for mathematical programs with complementarity constraints and which is essentially an interior point approach for a degenerate part of the problem. Section 2 explains the main features of this method.

In [26] we have shown that under weak assumptions global solutions of the smoothed problems converge to a global solution of *GSIP*, and that stationary points in the sense of Fritz John converge to a Fritz John point of *GSIP*. From the results in [2] it could be expected that without further assumptions even Karush-Kuhn-Tucker points of the smoothed problems converge to a Karush-Kuhn-Tucker point of *GSIP*.

The aim of the present article is to show that in the setting of *GSIP* this is actually *not* the case. We give, however, different additional assumptions which guarantee the convergence to a Karush-Kuhn-Tucker point. These are the contents of Sections 3 and 4.

2. Preliminaries

This section reviews the main ideas of the smoothing method from [26].

2.1. The Reduction Ansatz for convex lower level problems

The n-parametric so-called lower level problems of *GSIP* are given by

$$Q^i(x): \qquad \underset{y\in\mathbb{R}^m}{\text{maximize}} \quad g_i(x,y) \quad \text{subject to} \quad y \in Y(x)$$

with $i \in I$. Note that the upper level decision variable x is a parameter of the lower problem, and that the upper level index variable y is the decision variable of the lower level. For each parameter value x we can study the optimal value and the optimal points of the optimization problem $Q^i(x)$. More precisely, associated with $Q^i(x)$ are its optimal value function

$$\varphi_i(x) \;=\; \begin{cases} \sup\limits_{y\in Y(x)} g_i(x,y), & \text{if} \quad Y(x)\neq\emptyset \\ -\infty, & \text{else}, \end{cases}$$

and, in case of solvability, its solution set mapping

$$Y^i_\star(x) \;=\; \{y \in Y(x)|\; g_i(x,y) = \varphi_i(x)\}.$$

It is easily seen that M and the set $\{x \in \mathbb{R}^n|\; \varphi_i(x) \le 0,\; i \in I\}$ coincide.

Assumption 2.1. *For all $x \in \mathbb{R}^n$ the lower level problems $Q^i(x)$, $i \in I$, are convex, that is, the functions $-g_i(x,\cdot)$, $v_\ell(x,\cdot)$, $\ell \in L$, are convex on $\mathbb{R}^m$.*

Assumption 2.2. *For all $x \in \mathbb{R}^n$ the sets $Y(x)$ are bounded and satisfy the Slater condition, that is, there exists some $y^\star$ such that $v_\ell(x,y^\star) < 0$ for all $\ell \in L$.*

Under Assumptions 2.1 and 2.2 the sets $Y^i_\star(x)$ are nonempty and locally bounded around each $\bar{x} \in \mathbb{R}^n$ ([10]), so that the optimal value functions $\varphi_i(x) = \max_{y\in Y(x)} g_i(x,y)$, $i \in I$, are well-defined and continuous on $\mathbb{R}^n$ ([10]). In particular the feasible set M is closed.

For the derivation of stationarity conditions we concentrate on the nontrivial case of a point $\bar{x}$ from the boundary ∂M of M. Let $I_0(\bar{x}) = \{i \in I|\; \varphi_i(\bar{x}) = 0\}$ denote the set of active indices at $\bar{x}$. Since for $i \in I_0(\bar{x})$ the problem $Q^i(\bar{x})$ has vanishing optimal value, the set of its solution points can be described as

$$Y^i_\star(\bar{x}) \;=\; Y^i_0(\bar{x}) \;:=\; \{y \in Y(\bar{x}) |\; g_i(\bar{x}, y) = 0\},$$

and we have $i \in I_0(\bar{x})$ if and only if $Y^i_0(\bar{x}) \neq \emptyset$.

Next, we give a local description of M by finitely many *smooth* constraints for the case when certain regularity assumptions hold in the lower level problems. Let v be the column vector of the functions v_ℓ, $\ell \in L = \{1, ..., s\}$, let $\mathrm{diag}(\gamma)$ stand for the $(s, s)-$diagonal matrix with diagonal vector $\gamma \in \mathbb{R}^s$, and let $i \in I_0(\bar{x})$. Since each $\bar{y} \in Y^i_0(\bar{x})$ is a solution of $Q^i(\bar{x})$, and since the Slater condition (Assumption 2.2) holds in $Y(\bar{x})$, with the lower level Lagrange function

$$\mathcal{L}_i(x, y, \gamma) \;=\; g_i(x, y) - \gamma^\top v(x, y)$$

the Karush-Kuhn-Tucker theorem states that the following system of equalities and inequalities has a solution $\bar{\gamma}$:

$$\begin{aligned} \nabla_y \mathcal{L}_i(\bar{x}, \bar{y}, \gamma) &= 0 && (2.1)\\ -\mathrm{diag}(\gamma)\, v(\bar{x}, \bar{y}) &= 0 && (2.2)\\ -\gamma &\le 0 && (2.3)\\ v(\bar{x}, \bar{y}) &\le 0. && (2.4) \end{aligned}$$

The *Linear Independence Constraint Qualification (LICQ)* is said to hold at $\bar{y} \in Y(\bar{x})$ if the family of vectors $\nabla_y v_\ell(\bar{x}, \bar{y})$, $\ell \in L_0(\bar{x}, \bar{y})$, is linearly independent. Here $L_0(\bar{x}, \bar{y}) = \{\ell \in L|\; v_\ell(\bar{x}, \bar{y}) = 0\}$ denotes the set of lower level active indices, and $\nabla_y v_\ell$ stands for the *column* vector of partial derivatives of v_ℓ with respect to y. For brevity we will denote its transpose $\nabla_y^\top v_\ell$ (i.e., the Jacobian of v with respect to y) by $D_y v_\ell$. LICQ at $\bar{y}$ implies Assumption 2.2 as well as uniqueness of the multiplier $\bar{\gamma}$.

The point $\bar{y}$ is said to satisfy the *Strict Complementary Slackness (SCS)* condition, if $\bar{\gamma}_\ell > 0$, $\ell \in L_0(\bar{x}, \bar{y})$. Under LICQ the tangent space to $Y(\bar{x})$ at $\bar{y}$ can be described as $T_{\bar{y}} Y(\bar{x}) = \{\eta \in \mathbb{R}^m |\; D_y v_\ell(\bar{x}, \bar{y})\, \eta = 0,\; \ell \in L_0(\bar{x}, \bar{y})\}$. The point $\bar{y}$ is said to satisfy the *Second Order Sufficiency Condition (SOSC)* if the matrix $D^2_y \mathcal{L}_i(\bar{x}, \bar{y}, \bar{\gamma})|_{T_{\bar{y}} Y(\bar{x})}$ possesses only negative eigenvalues. Here, $D^2_y \mathcal{L}_i = D_y \nabla_y \mathcal{L}_i$ denotes the Hessian matrix of $\mathcal{L}_i$ with respect to y, and we have $D^2_y \mathcal{L}_i(\bar{x}, \bar{y}, \bar{\gamma})|_{T_{\bar{y}} Y(\bar{x})} = V^\top D^2_y \mathcal{L}_i(\bar{x}, \bar{y}, \bar{\gamma}) V$ for any matrix V of $m-$vectors which form a basis of the tangent space $T_{\bar{y}} Y(\bar{x})$. The Jacobian of (2.1), (2.2) with respect to (y, γ),

$$\begin{aligned} A^i(x, y, \gamma) \;&:=\; D_{(y,\gamma)} \begin{pmatrix} \nabla_y \mathcal{L}_i(x, y, \gamma) \\ -\mathrm{diag}(\gamma) v(x, y) \end{pmatrix} \\ &=\; \begin{pmatrix} D^2_y \mathcal{L}_i(x, y, \gamma) & -\nabla_y v(x, y) \\ -\mathrm{diag}(\gamma) D_y v(x, y) & -\mathrm{diag}(v(x, y)) \end{pmatrix}, \qquad (2.5) \end{aligned}$$

will play an important role throughout this paper.

Definition 2.1. *Let $\bar{x} \in \partial M$ and $i \in I_0(\bar{x})$. A point $\bar{y} \in Y^i_0(\bar{x})$ is called* nondegenerate global maximizer *of $Q^i(\bar{x})$ if LICQ holds at $\bar{y}$ and if SCS and SOSC are valid with the vector $\bar{\gamma}$ satisfying (2.1) – (2.4).*

Lemma 2.1 (cf., e.g., [24]) *Let Assumptions 2.1 and 2.2 be satisfied and let $\bar{x} \in \mathbb{R}^n$ as well as $i \in I$. Then a point $\bar{y}^i$ is a nondegenerate global maximizer of $Q^i(\bar{x})$ with corresponding multiplier vector $\bar{\gamma}^i$ if and only if (2.1) – (2.4) hold and if the Jacobian $A^i(x, y, \gamma)$ is nonsingular at $(\bar{x}, \bar{y}^i, \bar{\gamma}^i)$.*

Weak sufficient conditions for the nonsingularity of the matrices $A^i(\bar{x}, \bar{y}^i, \bar{\gamma}^i)$, $i \in I$, are given in [24].

Assumption 2.3 (Reduction Ansatz, cf. [7, 9, 29]) *For each $i \in I_0(\bar{x})$ all global maximizers of $Q^i(\bar{x})$ are nondegenerate.*

Under Assumptions 2.1 and 2.3 the global maximizers of $Q^i(\bar{x})$ are uniquely determined, that is, the sets $Y_0^i(\bar{x}) = \{\bar{y}^i\}$, $i \in I_0(\bar{x})$, are singletons.

An application of the implicit function theorem ([3]) shows that for each $\bar{y}^i$ with corresponding multiplier vector $\bar{\gamma}^i$ there are $C^{d-1}-$functions y^i and γ^i, defined on a neighborhood U^i of $\bar{x}$, such that $(y^i(\bar{x}), \gamma^i(\bar{x})) = (\bar{y}^i, \bar{\gamma}^i)$ and such that $y^i(x)$ is the locally unique local maximizer of $Q^i(x)$ with multiplier $\gamma^i(x)$. Hence, we may introduce the locally defined optimal value functions

$$\varphi_i : \; U^i \to \mathbb{R}, \quad x \mapsto g_i(x, y^i(x)), \; i \in I_0(\bar{x}).$$

Lemma 2.2 (cf., e.g., [12]) *The functions φ_i are of differentiability class C^d, and their gradients satisfy*

$$\nabla \varphi_i(\bar{x}) \;\; = \;\; \nabla_x \mathcal{L}_i(\bar{x}, \bar{y}^i, \bar{\gamma}^i).$$

Theorem 2.1 (Reduction Lemma, cf. [9, 23]) *Let Assumption 2.1 hold, let Assumption 2.3 be satisfied at $\bar{x}$, and put $U := \bigcap_{i \in I_0(\bar{x})} U^i$. Then the sets M and*

$$M_{\bar{x}} \;=\; \{\, x \in U | \; \varphi_i(x) \le 0, \; i \in I_0(\bar{x}) \,\}$$

coincide locally around $\bar{x}$.

Theorem 2.1 shows that under the Reduction Ansatz the original problem *GSIP* is locally equivalent to the reduced problem $\min f|_{M_{\bar{x}}}$. Hence, local optimality conditions from finite optimization may be applied to yield results for the semi-infinite case.

For example, in view of Lemma 2.2 the *Mangasarian-Fromovitz Constraint Qualification (MFCQ)* is said to hold at $\bar{x}$ in $M_{\bar{x}}$ if there exists some vector $d \in \mathbb{R}^m$ such that

$$0 \;>\; D\varphi_i(\bar{x})\, d \;=\; D_x \mathcal{L}_i(\bar{x}, \bar{y}^i, \bar{\gamma}^i)\, d, \quad i \in I_0(\bar{x}).$$

Moreover, we obtain Fritz John type and Karush-Kuhn-Tucker type first order necessary optimality conditions ([11, 16]):

Theorem 2.2.

(i) Let Assumption 2.1 hold and let Assumption 2.3 be satisfied at a local minimizer $\bar{x}$ of GSIP. Then there exist multipliers $\kappa \ge 0$, $\lambda_i \ge 0$, $i \in I_0(\bar{x})$, not all equal to zero, such that

$$\kappa \nabla f(\bar{x}) \;+ \sum_{i \in I_0(\bar{x})} \lambda_i \nabla_x \mathcal{L}_i(\bar{x}, \bar{y}^i, \bar{\gamma}^i) \;=\; 0. \tag{2.6}$$

(ii) If, in addition to the assumptions of part a), MFCQ holds at $\bar{x}$, then in (2.6) one can choose $\kappa = 1$.

2.2. The numerical approach

Our numerical method replaces *GSIP* by a sequence of finite nonlinear programming problems which are numerically tractable and whose solutions or stationary points converge to a solution or a stationary point of *GSIP*, respectively. Unlike other numerical methods for semi-infinite programming, this approach does not discretize the index set $Y(x)$, but it takes advantage of the fact that the solution set of a regular convex lower level problem is characterized by its first order optimality condition.

Let us first recall that a function $\psi : \mathbb{R}^2 \to \mathbb{R}$ with

$$\psi(a, b) \;=\; 0 \quad \text{if and only if} \quad a \geq 0,\; b \geq 0,\; ab \;=\; 0$$

is called *NCP function.* Examples are the min-function, or natural residual function,

$$\psi^{NR}(a, b) \;=\; \tfrac{1}{2}\left(a + b - \sqrt{(a-b)^2}\right),$$

and the Fischer-Burmeister function ([4])

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

For numerical purposes one can regularize these nondifferentiable NCP functions. The so-called Chen-Harker-Kanzow-Smale function ([1, 14, 22]) is given by

$$\psi_\tau^{NR}(a, b) \;=\; \tfrac{1}{2}\left(a + b - \sqrt{(a-b)^2 + 4\tau^2}\right),$$

whereas the so-called smoothed Fischer-Burmeister function is

$$\psi_\tau^{FB}(a, b) \;=\; a + b - \sqrt{a^2 + b^2 + 2\tau^2}.$$

Obviously ψ_τ^{NR} and ψ_τ^{FB} are continuously differentiable for all $\tau \neq 0$, and for $\tau = 0$ they coincide with ψ^{NR} and ψ^{FB}, respectively. Moreover, both functions share the following important properties:

Lemma 2.3 ([14, 26]) *Let $\tau \neq 0$ and let ψ_τ denote one of the functions ψ_τ^{NR} and ψ_τ^{FB}. Then the following assertions hold:*

(i) We have $\psi_\tau(a, b) \;=\; 0$ if and only if $a > 0,\; b > 0,\; ab \;=\; \tau^2$.

(ii) At (a, b) with $\psi_\tau(a, b) = 0$ we have $D\psi_\tau(a, b) = (a + b)^{-1}(b, a)$. In particular, at a zero of ψ_τ the gradient of ψ_τ does not explicitly depend on τ.

In the sequel we mainly need the results of Lemma 2.3, so that we will not distinguish between ψ_τ^{NR} and ψ_τ^{FB} but simply write ψ_τ.

Our numerical approach bases on the observation from [25] that *GSIP* and the Stackelberg game ([21])

$$SG: \quad \min_{x, y^1, \ldots, y^p} \; f(x) \quad \text{s.t.} \quad g_i(x, y^i) \leq 0, \;\; \text{and} \;\; y^i \;\; \text{solves} \;\; Q^i(x),\; i \in I,$$

are equivalent problems whenever the index set $Y(x)$ is nonempty for all $x \in \mathbb{R}^n$. The latter is the case under Assumption 2.2.

Under Assumption 2.1 the restrictions "y^i solves $Q^i(x)$" in SG can be equivalently replaced by their first order optimality conditions: for each $i \in I$ there is a solution γ^i of (2.1) – (2.4). The latter statement is true under Assumption 2.2, since Slater's condition guarantees the existence of Karush-Kuhn-Tucker multipliers. By this reformulation, SG is equivalent to the following mathematical program with complementarity constraints, where $\mathrm{diag}(\gamma)$ denotes the $(s,s)-$diagonal matrix with diagonal vector $\gamma \in \mathbb{R}^s$:

$$MPCC: \quad \min_{x,y^1,\gamma^1\dots,y^p,\gamma^p} f(x) \quad \text{s.t.} \quad \begin{array}{rcl} g_i(x,y^i) & \le & 0 \\ \nabla_y \mathcal{L}_i(x,y^i,\gamma^i) & = & 0 \\ -\mathrm{diag}(\gamma^i)\, v(x,y^i) & = & 0 \\ -\gamma^i & \le & 0 \\ v(x,y^i) & \le & 0,\ i \in I. \end{array}$$

Unfortunately, numerical standard software cannot be expected to solve this problem since due to the appearance of complementarity conditions MFCQ is violated at all points of the feasible set of $MPCC$ ([20]). In [13, 19] it is shown that MFCQ is a necessary condition for the stability of smooth nonlinear programs under data perturbations and thus for the stability of numerical methods in the presence of round-off errors.

Given an NCP function ψ and $a, b \in \mathbb{R}^s$ we define the vectorization

$$\Psi(a,b) \;=\; (\psi(a_1\,,\,b_1),\dots,\psi(a_s\,,\,b_s))^\top,$$

so that $MPCC$ can be equivalently rewritten as the nonsmooth problem

$$P: \quad \min_{x,y^1,\gamma^1\dots,y^p,\gamma^p} f(x) \quad \text{s.t.} \quad \begin{array}{rcl} g_i(x,y^i) & \le & 0 \\ \nabla_y \mathcal{L}_i(x,y^i,\gamma^i) & = & 0 \\ \Psi(\gamma^i,-v(x,y^i)) & = & 0,\ i \in I. \end{array}$$

To smooth P we take an interior point approach for the lower level problems $Q^i(x)$, $i \in I$. In fact, for each $i \in I$ we replace the Karush-Kuhn-Tucker system (2.1) – (2.4) for (y^i,γ^i) by the perturbed system

$$\nabla_y \mathcal{L}_i(x,y^i,\gamma^i) \;=\; 0 \tag{2.7}$$

$$-\mathrm{diag}(\gamma^i)\, v(x,y^i) \;=\; \tau^2\, e_s \tag{2.8}$$

$$-\gamma^i \;\le\; 0 \tag{2.9}$$

$$v(x,y^i) \;\le\; 0 \tag{2.10}$$

depending on $\tau \in \mathbb{R}$ (and on x). Here we set $e_s = (1,\dots,1)^\top \in \mathbb{R}^s$. Note that for $\tau \neq 0$, under (2.8) the nonstrict inequalities (2.9), (2.10) are equivalent to their strict analogs. In view of Lemma 2.3(i), with one of the regularized NCP functions Ψ_τ in vector form, the problem P is thus embedded into the parameterized family of optimization problems

$$P_\tau: \quad \min_{x,y^1,\gamma^1\dots,y^p,\gamma^p} f(x) \quad \text{s.t.} \quad g_i(x,y^i) \;\le\; 0$$

$$\nabla_y \mathcal{L}_i(x,y^i,\gamma^i) \;=\; 0 \tag{2.11}$$

$$\Psi_\tau(\gamma^i,-v(x,y^i)) \;=\; 0,\ i \in I \tag{2.12}$$

with $P_0 = P$.

Note that the Jacobian of (2.7), (2.8) with respect to (y^i, γ^i) does not depend on τ and is given by $A^i(x, y^i, \gamma^i)$ from (2.5). Furthermore, by Lemma 2.3(ii) the Jacobian of (2.11), (2.12) with respect to (y^i, γ^i) at a point (x, y^i, γ^i) satisfying (2.11), (2.12) with $\tau \neq 0$ is

$$D_{(y^i,\gamma^i)} \begin{pmatrix} \nabla_y \mathcal{L}_i(x, y^i, \gamma^i) \\ \Psi_\tau(\gamma^i, -v(x, y^i)) \end{pmatrix} = \begin{pmatrix} Id & 0 \\ 0 & \operatorname{diag}(\gamma^i - v(x, y^i))^{-1} \end{pmatrix} \cdot A^i(x, y^i, \gamma^i) \tag{2.13}$$

where Id denotes the $(m, m)-$identity matrix, and where $\operatorname{diag}(\gamma^i - v(x, y^i))$ is nonsingular since all its diagonal entries are positive in view of (2.12) and Lemma 2.3(i). The Jacobian in (2.13) does not explicitly depend on τ, and it is nonsingular if and only if $A^i(x, y^i, \gamma^i)$ is.

The Jacobian of (2.11), (2.12) with respect to the complete variable vector $(x, y^1, \gamma^1, ..., y^p, \gamma^p)$ at a feasible point of P_τ thus does not explicitly depend on τ neither and can be written as

$$\begin{aligned} D\begin{pmatrix} \nabla_y \mathcal{L}_i(x, y^i, \gamma^i) \\ \Psi_\tau(\gamma^i, -v(x, y^i)) \end{pmatrix} &= \begin{pmatrix} Id & 0 \\ 0 & \operatorname{diag}(\gamma^i - v(x, y^i))^{-1} \end{pmatrix} \cdot \\ &\quad \left(\begin{pmatrix} D_x \nabla_y \mathcal{L}_i(x, y^i, \gamma^i) \\ -\operatorname{diag}(\gamma^i) D_x v(x, y^i) \end{pmatrix}, 0, ..., 0, A^i(x, y^i, \gamma^i), 0, ..., 0 \right). \end{aligned} \tag{2.14}$$

The following proposition shows that for $\tau \neq 0$ problem P_τ is numerically tractable in the sense that the inherent singularity in the equality constraints of problem P is removed.

Proposition 2.1 ([24]) *Let $\tau \neq 0$ and let $(x, y^1, \gamma^1, ..., y^p, \gamma^p)$ be a feasible point of P_τ such that for each $i \in I$ the matrix $A^i(x, y^i, \gamma^i)$ is nonsingular. Then the gradients of the equality constraints of P_τ are linearly independent at $(x, y^1, \gamma^1, ..., y^p, \gamma^p)$.*

Proof. The assertion immediately follows from (2.14) and the block structure of the matrix

$$D \begin{pmatrix} \nabla_y \mathcal{L}_1(x, y^1, \gamma^1) \\ \Psi_\tau(\gamma^1, -v(x, y^1)) \\ \vdots \\ \nabla_y \mathcal{L}_p(x, y^p, \gamma^p) \\ \Psi_\tau(\gamma^p, -v(x, y^p)) \end{pmatrix}.$$

The ideas presented so far lead to the following continuation method ([26]):

Numerical method

Step 1: Choose a sequence $\{\tau_\nu\}$ of nonzero reals with $\lim_{\nu\to\infty} \tau_\nu = 0$, a starting point $x^0 \in \mathbb{R}^n$, and a termination criterion.

Step 2: Compute a starting point $(x^{0,0}, y^{1,0,0}, \gamma^{1,0,0}, ..., y^{p,0,0}, \gamma^{p,0,0})$ for the solution of P_{τ_0} and set $\nu = 0$.

Step 3: Starting from $(x^{\nu,0}, y^{1,\nu,0}, ..., \gamma^{p,\nu,0})$, find a solution $(x^{\nu,\star}, y^{1,\nu,\star}, ..., \gamma^{p,\nu,\star})$ of P_{τ_ν} .

Step 4: If the termination criterion is violated at $(x^{\nu,\star}, y^{1,\nu,\star}, ..., \gamma^{p,\nu,\star})$, set $(x^{\nu+1,0}, y^{1,\nu+1,0}, ..., \gamma^{p,\nu+1,0}) = (x^{\nu,\star}, y^{1,\nu,\star}, ..., \gamma^{p,\nu,\star})$, $\nu := \nu + 1$, and go to Step 3.

We emphasize that this method only takes advantage of specially structured *lower* level problems to reduce the numerical computation of a solution for *GSIP* to the solution of a sequence of finite optimization problems. On the other hand, the latter nonlinear finite problems are *not* assumed to possess any special properties, so that a starting point x^0 has to be chosen in advance.

In step 2 we clearly choose $x^{0,0} = x^0$ as the starting point for iteration $\nu = 0$. In order to obtain the corresponding starting values $(y^{1,0,0}, ..., \gamma^{p,0,0})$ numerically, one might try to find a zero of

$$\begin{pmatrix} \nabla_y \mathcal{L}_i(x^0, y^i, \gamma^i) \\ \Psi_{\tau_0}(\gamma^i, -v(x^0, y^i)) \end{pmatrix}$$

for each $i \in I$. A better method is given in [24].

Step 3 is a "black box" which stands for any standard solution method for nonlinear finite optimization problems. To use our solution approach, a user thus only has to construct the auxiliary problems P_τ for several values of τ, and solve those with his favorite NLP software.

Termination criteria might be the relative error of optimal points or of optimal values, as well as the error in the first order optimality condition for *GSIP*, and combinations thereof. We point out that the availability of a first order optimality condition is crucial for the numerical performance of the method. In contrast to this, an analogous approach for the solution of general classes of *MPCC*s suffers from the drawback that good termination criteria are difficult to check numerically ([2], [15], [20]).

For implementation issues, numerical results and further details about this method we refer to [24, 26].

2.3. Convergence results

As shown in [26], the smoothing approach embeds *GSIP* into the family of problems

$$GSIP_\tau : \quad \min_x \ f(x) \quad \text{s.t.} \quad x \in M_{GSIP_\tau}$$

with $M_{GSIP_\tau} = \mathrm{pr}_x(M_{P_\tau})$, the orthogonal projection of $M(P_\tau)$ to $\mathbb{R}^n$. We denote the unfolded feasible set of M_{GSIP_τ} by

$$\mathcal{M}_{GSIP} \ = \ \{ \ (x, \tau) \in \mathbb{R}^n \times \mathbb{R} | \ x \in M_{GSIP_\tau} \ \}.$$

Now let Assumption 2.1 hold, let Assumption 2.3 be satisfied at some point $\bar{x} \in M_{GSIP}$, and let $\bar{y}^i$ denote the unique solution of $Q^i(\bar{x})$ with unique multiplier vector $\bar{\gamma}^i$ for $i \in I_0(\bar{x})$. Then for each $i \in I_0(\bar{x})$ the point

$(x, y^i, \gamma^i, \tau) = (\bar{x}, \bar{y}^i, \bar{\gamma}^i, 0)$ solves (2.7), (2.8), and the Jacobian with respect to (y^i, γ^i) of the latter system of equations, $A^i(\bar{x}, \bar{y}^i, \bar{\gamma}^i)$, is nonsingular. Thus, the implicit function theorem can be applied to obtain locally unique $C^{d-1}-$functions $(y^i(x,\tau), \gamma^i(x,\tau))$ with $(y^i(\bar{x},0), \gamma^i(\bar{x},0)) = (\bar{y}^i, \bar{\gamma}^i)$, $i \in I_0(\bar{x})$, such that the points $(x, y^i(x,\tau), \gamma^i(x,\tau), \tau)$ solve (2.7), (2.8) for all (x,τ) in some neighborhood $\mathcal{U}^i$ of $(\bar{x}, 0)$.

Due to Assumption 2.3 in particular the strict complementary slackness condition is satisfied, so that for all $i \in I_0(\bar{x})$ we have $v_\ell(\bar{x}, \bar{y}^i) < 0, \ \ell \in L_0^c(\bar{x}, \bar{y}^i)$, and $\bar{\gamma}^i_\ell > 0, \ \ell \in L_0(\bar{x}, \bar{y}^i)$. By continuity also $v_\ell(x, y^i(x,\tau)) < 0, \ \ell \in L_0^c(\bar{x}, \bar{y}^i)$, and $\gamma^i_\ell(x,\tau) > 0, \ \ell \in L_0(\bar{x}, \bar{y}^i)$, hold for all $(x,\tau) \in \mathcal{U}^i$ and for sufficiently small $\mathcal{U}^i$. In view of (2.8) and $\tau^2 > 0$ this implies $\gamma^i_\ell(x,\tau) > 0, \ \ell \in L_0^c(\bar{x}, \bar{y}^i)$, and $v_\ell(x, y^i(x,\tau)) < 0, \ \ell \in L_0(\bar{x}, \bar{y}^i)$, on $\mathcal{U}^i$. The implicit functions $y^i(x,\tau)$ and $\gamma^i(x,\tau)$ thus also satisfy (2.9), (2.10), even strictly. Because of Lemma 2.3(i) the latter means that the same implicit functions are the locally unique solutions of (2.11), (2.12).

Using the implicitly defined functions y^i, $i \in I_0(\bar{x})$, we introduce the following subset of the unfolding space $\mathbb{R}^n \times \mathbb{R}$ in the neighborhood $\mathcal{U} := \bigcap_{i \in I_0(\bar{x})} \mathcal{U}^i$ of $(\bar{x}, 0)$:

$$\mathcal{M}_{(\bar{x},0)} \;=\; \{\; (x,\tau) \in \mathcal{U} |\; g_i(x, y^i(x,\tau)) \le 0,\; i \in I_0(\bar{x}) \;\}.$$

Lemma 2.4 (Parametric Reduction Lemma, cf. [26])
Let Assumption 2.1 hold and let Assumption 2.3 be satisfied at some point $\bar{x} \in M_{GSIP}$. Then the sets $\mathcal{M}_{GSIP}$ and $\mathcal{M}_{(\bar{x},0)}$ coincide locally around $(\bar{x}, 0)$.

In particular, for fixed τ close to 0 locally around $\bar{x}$ the finite optimization problem

$$(GSIP_{\bar{x}})_\tau : \quad \min_x \; f(x) \quad \text{s.t.} \quad g_i(x, y^i(x,\tau)) \le 0,\; i \in I_0(\bar{x}),$$

is equivalent to $GSIP_\tau$. Choosing $\tau = 0$ we see that the Parametric Reduction Lemma implies Theorem 2.1, that is, $\bar{x}$ is a local solution of *GSIP* if and only if $\bar{x}$ solves

$$(GSIP_{\bar{x}})_0 : \quad \min_x \; f(x) \quad \text{s.t.} \quad g_i(x, y^i(x,0)) \le 0,\; i \in I_0(\bar{x}).$$

Lemma 2.4 is the main tool to prove the following convergence results.

Theorem 2.3 ([26]) *Let Assumption 2.1 hold, let $(\tau_\nu)_{\nu\in\mathbb{N}}$ be a sequence with* $\lim_{\nu\to\infty} \tau_\nu = 0$, *and let $(x^\nu, y^{1,\nu}, \gamma^{1,\nu} ..., y^{p,\nu}, \gamma^{p,\nu})_{\nu\in\mathbb{N}}$ be a sequence of global solutions of P_{τ_ν}, $\nu \in \mathbb{N}$. If $x^\star$ is an accumulation point of the sequence $(x^\nu)_{\nu\in\mathbb{N}}$ such that Assumption 2.3 holds at $x^\star$ and such that MFCQ holds at some solution of $GSIP_0(x^*)$, then $x^\star$ is a global solution of GSIP.*

Theorem 2.3 is primarily of theoretical interest as numerical standard software may not find global solution points of the problems P_{τ_ν}, $\nu \in \mathbb{N}$. One can at most expect a point which satisfies some stationarity condition. Consequently a numerical solution method for *GSIP* can also merely be expected to find stationary points in the sense of Theorem 2.2. For stationarity in the sense of Fritz John such a result is known:

Theorem 2.4 ([26]) *Let Assumption 2.1 hold, let $(\tau_\nu)_{\nu\in\mathbb{N}}$ be a sequence with* $\lim_{\nu\to\infty} \tau_\nu = 0$, *and let $(x^\nu, y^{1,\nu}, \gamma^{1,\nu} ..., y^{p,\nu}, \gamma^{p,\nu})$ be Fritz John points of P_{τ_ν}, $\nu \in \mathbb{N}$, with an accumulation point $(x^\star, y^{1,\star}, \gamma^{1,\star} ..., y^{p,\star}, \gamma^{p,\star})$. Let Assumption 2.3 hold at $x^\star$, and let the matrices $A^i(x^\star, y^{i,\star}, \gamma^{i,\star})$, $i \in I \setminus I_0(x^\star)$, be nonsingular. Then $x^\star$ is a Fritz John point of GSIP.*

Weak sufficient conditions for the existence of accumulation points in the assumption of Theorem 2.4 are given in [24].

Comparing these convergence results to those from [2] for $MPCCs$, it seems that it should be possible to prove a stronger result, namely that without further assumptions Karush-Kuhn-Tucker points of P_{τ_ν} converge to a Karush-Kuhn-Tucker point of *GSIP*. This is, however, not the case as we will see next.

3. Convergence of Karush-Kuhn-Tucker Points

In the following example Karush-Kuhn-Tucker points of the problems P_τ converge to a Fritz John point of *GSIP* which is *not* a Karush-Kuhn-Tucker point.

Example 3.1. *Let $n = m = 1$ and put $f(x) = x$, $g(x,y) = x^2 + y$, $v_1(y) = y$, $v_2(y) = -1 - y$. Then $Y = [-1, 0]$ is a compact set and we obtain $\varphi(x) = \max_{y \in Y} g(x,y) = x^2$. It follows that*

$$M_{GSIP} \;=\; \{\; x \in \mathbb{R} |\; \varphi(x) \leq 0 \;\} \;=\; \{0\}$$

so that the only feasible point is $\bar{x} = 0$. Its active index set is

$$Y_0(0) \;=\; \{\; y \in Y |\; g(0,y) = 0 \;\} \;=\; \{0\},$$

and $\bar{y} = 0$ is a nondegenerate global maximizer of the convex lower level problem

$$Q(0): \qquad \max\, y \quad s.t. \quad -1 \leq y \leq 0.$$

Thus the Reduction Ansatz (Assumption 2.3) holds at $\bar{x} = 0$, and the corresponding implicit function for the solution point is $y(x) \equiv 0$. This leads to $\varphi(x) = g(x, y(x)) = x^2$ and $M_{\bar{x}} = \{x \in \mathbb{R} |\; x^2 \leq 0\}$, so that $\bar{x} = 0$ is a Fritz John point of $f|_{M_{\bar{x}}}$, but not a Karush-Kuhn-Tucker point.

On the other hand, for $\tau \in \mathbb{R}$ the smoothed problem P_τ becomes

$$\min_{x,y,\gamma_1,\gamma_2} \; x \quad s.t. \quad \begin{aligned} x^2 + y &\leq 0 \\ 1 - \gamma_1 + \gamma_2 &= 0 \\ \psi_\tau(\gamma_1\,,\, -y) &= 0 \\ \psi_\tau(\gamma_2\,,\, 1 + y) &= 0 \end{aligned}$$

where ψ_τ denotes one of the smoothed NCP functions ψ_τ^{NR}, ψ_τ^{FB}. It is not hard to see that for a feasible point $(x, y, \gamma_1, \gamma_2)$ we necessarily have

$$y \;=\; -\tau^2 - \tfrac{1}{2} + \sqrt{\tau^4 + \tfrac{1}{4}}$$

which is negative for $\tau \neq 0$. P_τ is then equivalent to the problem

$$\min_x \; x \quad s.t. \quad x^2 \;\leq\; \tau^2 + \tfrac{1}{2} - \sqrt{\tau^4 + \tfrac{1}{4}}$$

with the solution $x(\tau) = -\sqrt{\tau^2 + \frac{1}{2} - \sqrt{\tau^4 + \frac{1}{4}}}$ which is obviously a Karush-Kuhn-Tucker point. However, the limit $\bar{x} = 0 = \lim_{\tau \to 0} x(\tau)$ is not a Karush-Kuhn-Tucker point of GSIP, but only a Fritz John point.

Example 3.1 shows that without further assumptions Karush-Kuhn-Tucker points of the smoothed problems do not necessarily converge to a Karush-Kuhn-Tucker point of *GSIP*. In the following we will give such further assumptions.

Lemma 3.1. *Let Assumption 2.1 hold, let Assumption 2.3 be satisfied at a point $\bar{x} \in M_{GSIP}$, and let $(x, y^1, \gamma^1, ..., y^p, \gamma^p)$ be a Karush-Kuhn-Tucker point of P_τ with (x, τ) sufficiently close to $(\bar{x}, 0)$. Moreover, let the matrices $A^i(\bar{x}, \bar{y}^i, \bar{\gamma}^i)$, $i \in I \setminus I_0(\bar{x})$, be nonsingular. Then x is a Karush-Kuhn-Tucker point of $(GSIP_{\bar{x}})_\tau$.*

Proof. The feasibility of $(x, y^1, \gamma^1, ..., y^p, \gamma^p)$ for P_τ and Lemma 2.3(i) imply particularly that (2.7) and (2.8) hold. Since the matrices $A^i(x, y^i, \gamma^i)$ are nonsingular for $i \in I_0(\bar{x})$ by the Reduction Ansatz and continuity, and also for $i \notin I_0(\bar{x})$ by the additional assumption of this lemma, for (x, τ) sufficiently close to $(\bar{x}, 0)$ the point y^i coincides with the unique solution $y^i(x, \tau)$ from the application of the implicit function theorem to (2.7), (2.8) for each $i \in I$. By Lemma 2.4 x is thus feasible for $(GSIP_{\bar{x}})_\tau$.

In view of

$$D\, g_i(x, y^i) \;=\; (D_x g_i(x, y^i),\; 0, ...0, D_y g_i(x, y^i), 0, ..., 0)$$

and (2.14) the Karush-Kuhn-Tucker condition for P_τ at $(x, y^1, \gamma^1 ..., y^p, \gamma^p)$ states that there exist multipliers $\lambda_i \geq 0$, $i \in I$, $\varrho^1, ..., \varrho^p \in \mathbb{R}^m$, and $\sigma^1, ..., \sigma^p \in \mathbb{R}^s$ with

$$\begin{aligned} 0 \;&=\; Df(x) + \sum_{i \in I} \lambda_i \, D_x g_i(x, y^i) \\ &\quad + \sum_{i \in I} ((\varrho^i)^\top, (\tilde{\sigma}^i)^\top) \begin{pmatrix} D_x \nabla_y \mathcal{L}_i(x, y^i, \gamma^i) \\ -\mathrm{diag}(\gamma^i) D_x v(x, y^i) \end{pmatrix} && (3.1) \\ 0 \;&=\; \lambda_i \,(\, D_y g_i(x, y^i), 0\,) + ((\varrho^i)^\top, (\tilde{\sigma}^i)^\top)\, A^i(x, y^i, \gamma^i),\; i \in I && (3.2) \\ 0 \;&=\; \lambda_i \cdot g_i(x, y^i),\; i \in I, && (3.3) \end{aligned}$$

where

$$\tilde{\sigma}^i \;=\; \mathrm{diag}(\gamma^i - v(x, y^i))^{-1} \sigma^i,\; i \in I.$$

Because of $y^i = y^i(x, \tau)$, the equations (3.3) become

$$0 \;=\; \lambda_i \cdot g_i(x, y^i(x, \tau)),\; i \in I. \tag{3.4}$$

Once again by the feasibility of $(x, y^1, \gamma^1, ..., y^p, \gamma^p)$ for P_τ and a simple continuity argument we obtain $g_i(x, y^i) = g_i(x, y^i(x, \tau)) < 0$ for $i \notin I_0(\bar{x})$ and (x, τ) sufficiently close to $(\bar{x}, 0)$. Hence, (3.4) implies

$$\lambda_i \;=\; 0, \quad i \notin I_0(\bar{x}). \tag{3.5}$$

Together with the nonsingularity of the matrices $A^i(x, y^i, \gamma^i),\; i \notin I_0(\bar{x})$, and with (3.2), this yields

$$((\varrho^i)^\top, (\tilde{\sigma}^i)^\top) \;=\; 0, \quad i \notin I_0(\bar{x}). \tag{3.6}$$

Next, (3.5), (3.6), and the nonsingularity of the matrices $A^i(x, y^i, \gamma^i)$, $i \in I_0(\bar{x})$, allow us to reduce (3.1), (3.2) to

$$\begin{aligned} 0 \;&=\; Df(x) + \sum_{i \in I_0(\bar{x})} \lambda_i \, D_x g_i(x, y^i) && (3.7) \\ &\quad - \sum_{i \in I_0(\bar{x})} \lambda_i \,(\, D_y g_i(x, y^i), 0\,)(A^i(x, y^i, \gamma^i))^{-1} \begin{pmatrix} D_x \nabla_y \mathcal{L}_i(x, y^i, \gamma^i) \\ -\mathrm{diag}(\gamma^i) D_x v(x, y^i) \end{pmatrix}. \end{aligned}$$

Plugging the implicit functions $y^i(x, \tau)$ and $\gamma^i(x, \tau)$ into (2.7), (2.8) and differentiating the resulting constant function with respect to x yields the equation

$$A^i(x, y^i, \gamma^i) \begin{pmatrix} D_x y^i(x, \tau) \\ D_x \gamma^i(x, \tau) \end{pmatrix} \;=\; -\begin{pmatrix} D_x \nabla_y \mathcal{L}_i(x, y^i, \gamma^i) \\ -\mathrm{diag}(\gamma^i) D_x v(x, y^i) \end{pmatrix}$$

for all (x,τ) in a neighborhood of $(\bar{x},0)$. Consequently (3.7) can be further reduced to

$$\begin{aligned} 0 &= Df(x) + \sum_{i\in I_0(\bar{x})} \lambda_i \left(D_x g_i(x,y^i) + D_y g_i(x,y^i) D_x y^i(x,\tau)\right) \\ &= Df(x) + \sum_{i\in I_0(\bar{x})} \lambda_i \, D_x[\, g_i(x,y^i(x,\tau))\,]. \end{aligned} \tag{3.8}$$

This proves that x satisfies the desired Karush-Kuhn-Tucker condition for $(GSIP_{\bar{x}})_\tau$.

We remark that Lemma 3.1 can be reversed: under the given nondegeneracy assumptions, for a Karush-Kuhn-Tucker point x of $(GSIP_{\bar{x}})_\tau$ with (x,τ) sufficiently close to $(\bar{x},0)$, the point $(x,y^1(x,\tau),\gamma^1(x,\tau),...,y^p(x,\tau),\gamma^p(x,\tau))$ is a Karush-Kuhn-Tucker point of P_τ. Its feasibility is obtained via Lemma 2.4, and the appropriate definitions of multipliers immediately follow from the proof of Lemma 3.1.

Theorem 3.1. *Let Assumption 2.1 hold, let $(\tau_\nu)_{\nu\in\mathbb{N}}$ be a sequence with $\lim_{\nu\to\infty}\tau_\nu=0$, and let $(x^\nu,y^{1,\nu},\gamma^{1,\nu}...,y^{p,\nu},\gamma^{p,\nu})$ be Karush-Kuhn-Tucker points of P_{τ_ν} , $\nu\in\mathbb{N}$, with multipliers $\lambda_i^\nu\geq 0$, $i\in I$, $\varrho^{1,\nu},...,\varrho^{p,\nu}\in\mathbb{R}^m$, $\sigma^{1,\nu},...,\sigma^{p,\nu}\in\mathbb{R}^s$, and with an accumulation point $(x^\star,y^{1,\star},\gamma^{1,\star}...,y^{p,\star},\gamma^{p,\star})$. Let Assumption 2.3 hold at $x^\star$, let the matrices $A^i(x^\star,y^{i,\star},\gamma^{i,\star})$, $i\in I\setminus I_0(x^\star)$, be nonsingular, and let the sequence $(\lambda_i^\nu)_\nu$ be bounded for each $i\in I$. Then $x^\star$ is a Karush-Kuhn-Tucker point of GSIP.*

Proof. For sufficiently large $\nu\in\mathbb{N}$ all assumptions of Lemma 3.1 are satisfied, so that x^ν is a Karush-Kuhn-Tucker point of $(GSIP_{x^\star})_{\tau_\nu}$ satisfying

$$0 \;=\; Df(x^\nu) + \sum_{i\in I_0(\bar{x})} \lambda_i^\nu \, D_x[\, g_i(x^\nu, y^i(x^\nu,\tau^\nu))\,]. \tag{3.9}$$

Since for all $i\in I$ the sequence of multipliers $(\lambda_i^\nu)_\nu$ is bounded, it converges without loss of generality to some $\lambda_i^\star\geq 0$. By the continuity of the appearing functions we can let ν tend to infinity in (3.9) and obtain

$$0 \;=\; Df(x^\star) + \sum_{i\in I_0(\bar{x})} \lambda_i^\star \, D_x[\, g_i(x^\star, y^i(x^\star,0))\,].$$

Due to Lemma 2.2 $x^\star$ is thus a Karush-Kuhn-Tucker point for *GSIP* in the sense of Theorem 2.2(ii).

4. Final Remarks

With the same techniques as in Section 3 it is easy to show that under the Reduction Ansatz at $\bar{x}\in M_{GSIP}$ the MFCQ holds at a feasible point $(x,y^1,\gamma^1,...,y^p,\gamma^p)$ of P_τ with (x,τ) sufficiently close to $(\bar{x},0)$ if and only if MFCQ holds at x in the feasible set of $(GSIP_{\bar{x}})_\tau$. Since MFCQ is stable under small perturbations, assuming MFCQ at $\bar{x}\in M_{GSIP}$ thus guarantees MFCQ at $(x,y^1,\gamma^1,...,y^p,\gamma^p)$ for (x,τ) sufficiently close to $(\bar{x},0)$. The stationary points of the smoothed problems as well as of the original problem can then only be Karush-Kuhn-Tucker points.

We emphasize that Assumption 2.1 on lower level convexity is satisfied in a number of real-life applications of *generalized* semi-infinite optimization ([24]), but rarely in *standard* semi-infinite optimization. Still some of the techniques discussed here can be used to solve standard semi-infinite problems with *nonconvex* lower levels. As an appropriate framework the so-called adaptive convexification algorithm is presented in [5].

Acknowledgments. The author wishes to thank the anonymous referee for her or his precise remarks which helped to significantly improve the exposition of results in this article.

References

[1] B. Chen, P.T. Harker, A non-interior-point continuation method for linear complementarity problems, *SIAM J. Matrix Anal. Appl.*, **14** (1993), 1168–1190.

[2] F. Facchinei, H. Jiang, L. Qi, A smoothing method for mathematical programs with equilibrium constraints, *Math. Program.*, **85** (1999), 107–134.

[3] A.V. Fiacco, G.P. McCormick, Nonlinear Programming: Sequential Unconstrained Minimization Techniques, Wiley, New York, 1968.

[4] A. Fischer, A special Newton-type optimization method, *Optimization*, **24** (1992), 269–284.

[5] C.A. Floudas, O. Stein, The adaptive convexification algorithm: a feasible point method for semi-infinite programming, Preprint No. 116, Department of Mathematics - C, RWTH Aachen University, 2006.

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

[7] R. Hettich, H. Th. Jongen, Semi-infinite programming: conditions of optimality and applications, in Optimization Techniques, J. Stoer, ed., Part 2, Lecture Notes in Control and Inform. Sci., Vol. 7, Springer, Berlin, 1978, 1–11.

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

[9] R. Hettich, G. Still, Second order optimality conditions for generalized semi-infinite programming problems, *Optimization*, **34** (1995), 195–211.

[10] W.W. Hogan, Directional derivatives for extremal value functions with applications to the completely convex case, *Oper. Res.*, **21** (1973), 188–209.

[11] F. John, Extremum problems with inequalities as subsidiary conditions, in Studies and Essays, R. Courant Anniversary Volume, Interscience, New York, 1948, 187–204.

[12] H. Th. Jongen, P. Jonker, F. Twilt, Nonlinear Optimization in Finite Dimensions, Kluwer, Dordrecht, 2000.

[13] H.Th. Jongen, G.W. Weber, Nonlinear optimization: characterization of structural stability, *J. Global Optim.*, **1** (1991), 47–64.

[14] C. Kanzow, Some noninterior continuation methods for linear complementarity problems, *SIAM J. Matrix Anal. Appl.*, **17** (1996), 851–868.

[15] Z. Luo, J. Pang, D. Ralph, Mathematical Programs with Equilibrium Constraints, Cambridge University Press, Cambridge, 1996.

[16] O.L. Mangasarian, S. Fromovitz, The Fritz John necessary optimality conditions in the presence of equality and inequality constraints, *Journal of Mathematical Analysis and Applications*, **17** (1967), 37–47.

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

[18] R. Reemtsen, J.-J. Rückmann, eds., Semi-Infinite Programming, Kluwer, Boston, 1998.

[19] S.M. Robinson, Stability theory for systems of inequalities, part II: differentiable nonlinear systems, *SIAM J. Numer. Anal.*, **13** (1976), 497–513.

[20] H. Scheel, S. Scholtes, Mathematical programs with complementarity constraints: stationarity, optimality, and sensitivity, *Math. Oper. Res.*, **25** (2000), 1–22.

[21] K. Shimizu, Y. Ishizuka, J.F. Bard, Nondifferentiable and Two-Level Mathematical Programming, Kluwer, Boston, 1997.

[22] S. Smale, Algorithms for solving equations, in Proceedings of the International Congress of Mathematicians, AMS, Providence, 1987, 172–195.

[23] O. Stein, On Parametric Semi-infinite Optimization, Thesis, Shaker, Aachen, 1997.

[24] O. Stein, Bi-level Strategies in Semi-infinite Programming, Kluwer, Boston, 2003.

[25] O. Stein, G. Still, On generalized semi-infinite optimization and bilevel optimization, *Eur. J. Oper. Res.*, **142** (2002), 444–462.

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

[27] G. Still, Generalized semi-infinite programming: numerical aspects, *Optimization*, **49** (2001), 223–242.

[28] G. Still, Solving generalized semi-infinite programs by reduction to simpler problems, *Optimization*, **53** (2004), 19–38.

[29] W. Wetterling, Definitheitsbedingungen für relative Extrema bei Optimierungs- und Approximationsaufgaben, *Numer. Math.*, **15** (1970), 122–136.