J Glob Optim (2008) 41:245–266
DOI 10.1007/s10898-007-9228-z

The semismooth approach for semi-infinite programming under the Reduction Ansatz

Oliver Stein · Aysun Tezel

Received: 21 December 2006 / Accepted: 20 August 2007 / Published online: 26 September 2007

Abstract We study convergence of a semismooth Newton method for generalized semi-infinite programming problems with convex lower level problems where, using NCP functions, the upper and lower level Karush-Kuhn-Tucker conditions of the optimization problem are reformulated as a semismooth system of equations. Nonsmoothness is caused by a possible violation of strict complementarity slackness. We show that the standard regularity condition for convergence of the semismooth Newton method is satisfied under natural assumptions for semi-infinite programs. In fact, under the Reduction Ansatz in the lower level and strong stability in the reduced upper level problem this regularity condition is satisfied. In particular, we do not have to assume strict complementary slackness in the upper level. Numerical examples from, among others, design centering and robust optimization illustrate the performance of the method.

Keywords Generalized semi-infinite optimization · Semismooth Newton method · NCP function · CD-regularity · Reduction Ansatz

AMS subject classifications 2000 90C34 · 90C33 · 90C46 · 49M05 · 49M15

1 Introduction

This article studies a numerical solution method for generalized semi-infinite optimization problems, that is, problems of the type

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

O. Stein (✉)
School of Economics and Business Engineering, University of Karlsruhe, Karlsruhe, Germany
e-mail: stein@wior.uni-karlsruhe.de

A. Tezel
Department of Mathematics, Middle East Technical University, Ankara, Turkey
e-mail: atezel@metu.edu.tr

with

$$M = \{ x \in \mathbb{R}^n \mid g(x, y) \leq 0 \text{ for all } y \in Y(x) \}$$

and

$$Y(x) = \{ y \in \mathbb{R}^m \mid v_j(x, y) \leq 0, \ j \in Q \}.$$

All defining functions $f,\ g,\ v_j,\ j \in Q = \{1, \ldots, q\}$, are assumed to be real-valued and at least twice continuously differentiable on their respective domains. Moreover, we assume that the set-valued mapping $Y : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is locally bounded, that is, for each $\bar{x} \in \mathbb{R}^n$ there exists a neighborhood U of $\bar{x}$ such that $\bigcup_{x \in U} Y(x)$ is bounded in $\mathbb{R}^m$, and we require $Y(x) \neq \emptyset$ for all $x \in \mathbb{R}^n$.

Note that the possibly infinite index set $Y(x)$ of the semi-infinite inequality constraint is allowed to vary with x in a *GSIP*. As opposed to this, in a standard semi-infinite optimization problem (*SIP*) the index set is fixed, that is, we have $Y(x) \equiv Y$, and if Y is described by functional constraints, then the vector function v does not depend on x. For surveys about theory and methods for standard semi-infinite optimization we refer to [4,7,19], whereas introductions to generalized semi-infinite programming are given in [22] and in the recent tutorial paper [6].

If Y is even a finite set, we arrive at a usual nonlinear programming problem. Many solution methods for nonlinear programming problems base on solving their Karush-Kuhn-Tucker (KKT) system, that is, a necessary first order optimality condition. It is well-known that the complementarity conditions in the KKT system need special attention in any numerical approach. One possibility for their treatment is a reformulation by so-called NCP functions (see, e.g., [16] and the references therein), which reduces the problem to the solution of a certain system of equations which is either nonsmooth or smooth but degenerate. For special NCP functions these equations can be solved by so-called semismooth Newton methods where, in analogy to the standard Newton method, their convergence depends on a regularity condition in the solution point. It is important to note that the nonsmoothness of the system of equations stems from a possible lack of strict complementary slackness at a solution.

Such KKT methods have also been suggested for standard semi-infinite programming problems, where the KKT conditions take a somewhat more complicated form [7]. In particular, they involve an upper and a lower level problem. In the article [18] it was recently suggested to use NCP functions also for a nonsmooth reformulation of the KKT conditions in standard semi-infinite programming, and a regularity condition to guarantee convergence of a certain semismooth Newton method was proposed.

It turns out, however, that strict complementarity is a part of the regularity condition from [18], in the upper as well as in the lower level problem. A numerical method which searches a point with these regularity conditions would not need to use NCP functions but, in fact, already the standard Newton method would converge under these assumptions.

The aim of the present article is threefold: we point out an important pitfall in the solution of KKT systems for semi-infinite programs, we present a regularity condition which does not assume strict complementarity in the upper level problem, thus justifying the NCP function approach for semi-infinite programs, and at the same time we transfer this approach from standard to generalized semi-infinite programming. Note that this method will merely search for KKT points of the optimization problem, whereas global optimality plays a crucial role in the treatment of the so-called lower level problem (see Sect. 3.1).

The article is organized as follows. In Sect. 2 we briefly review the semismooth Newton method and a basic convergence result. Section 3 recalls the semismooth system of equations

which corresponds to the KKT system of *SIP*, and derives such a system for *GSIP*. Here, Sect. 3.1 collects basic facts about the bilevel structure of semi-infinite programming, and Sect. 3.2 reviews the KKT approach from [18] for standard semi-infinite programs, pointing out the mentioned pitfall. Section 3.3 is devoted to the Reduction Ansatz, a basic regularity condition in semi-infinite programming. We need it not only to formulate our convergence result, but also to derive an appropriate semismooth system of equations for the generalized semi-infinite case in Sect. 3.4.

Section 4 contains our main result, a convergence condition for the application of the semismooth Newton method to *GSIP* where strict complementarity does not have to be assumed in the upper level problem. Section 5 illustrates the performance of the semismooth Newton method for several generalized and standard semi-infinite programs arising in applications, and Sect. 6 concludes this article with some final remarks. Some auxiliary results about NCP functions and block matrices are given in an appendix.

2 Preliminaries on the semismooth Newton approach

For a locally Lipschitzian function $F : \mathbb{R}^n \to \mathbb{R}^m$ let $\partial F(x)$ denote Clarke's generalized Jacobian at x [2]. F is called *semismooth* at $x \in \mathbb{R}^n$ if F is directionally differentiable at x and if for all $V \in \partial F(x + d)$ and $d \to 0$ we have

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

In some sense, semismoothness is equivalent to the uniform convergence of directional derivatives in all directions [17]. Semismoothness was originally introduced by Mifflin for functionals [14]. In [17], the definition of semismooth functions was extended to $F : \mathbb{R}^n \to \mathbb{R}^m$. Convex functions, piecewise linear functions and smooth functions are examples of semismooth functions.

Furthermore, F is called *strongly semismooth* [16] at x if F is semismooth at x and if for all $V \in \partial F(x + d)$ and $d \to 0$ we have

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

For other definitions and properties of semismoothness we refer to [17].

In analogy to the standard Newton method, the basic iteration of the semismooth Newton approach for solving the equation $F(z) = 0$ is [17]

$$z^{k+1} = z^k - (W^k)^{-1} F(z^k) \tag{1}$$

with $W^k \in \partial F(z^k)$.

To study convergence properties of the semismooth Newton method, the concept of CD regularity was introduced. Here, CD stands for the Clarke subdifferential [16]. In fact, let $F : \mathbb{R}^n \to \mathbb{R}^n$ be semismooth. Then F is called *CD-regular* at a point $\bar{x}$, if all matrices in $\partial F(\bar{x})$ are nonsingular [16].

Theorem 2.1 ([17]) *Suppose that $\bar{x}$ is a solution of $F(x) = 0$, and F is semismooth and CD-regular at $\bar{x}$. Then the iteration method (1) is well defined and $\{x^k\}$, the sequence generated by (1), converges to $\bar{x}$ q-superlinearly in a neighborhood of $\bar{x}$. If F is strongly semismooth at $\bar{x}$, then the convergence is q-quadratic.*

3 Semismooth optimality conditions for *SIP* and *GSIP*

3.1 The bilevel structure of semi-infinite programming

The theoretical and numerical treatment of *GSIP* is closely related to the bilevel structure of semi-infinite programming. In fact, under our assumptions it is easy to see that the semi-infinite constraint in *GSIP* is equivalent to

$$\varphi(x) \;=\; \max_{y \in Y(x)} g(x, y) \;\leq\; 0,$$

which means that the feasible set M of *GSIP* is the lower level set of some optimal value function:

$$M \;=\; \{\, x \in \mathbb{R}^n \,|\, \varphi(x) \leq 0 \,\}.$$

The function φ is the optimal value function of the so-called *lower level problem*

$$Q(x): \quad \max_{y \in \mathbb{R}^m} g(x, y) \qquad \text{subject to} \qquad v_j(x, y) \;\leq\; 0, \;\; j \in Q. \tag{2}$$

In contrast to the upper level problem which consists in minimizing f over M, in the lower level problem x plays the role of an n-dimensional parameter, and y is the decision variable.

The main computational problem in semi-infinite programming is that the lower level problem has to be solved to *global optimality*, even if only a stationary point of the upper level problem is sought. In fact, standard NLP solvers can only be expected to produce a *local* maximizer y_{loc} of $Q(\bar{x})$ which is not necessarily a global maximizer y_{glob}. Even if $g(\bar{x}, y_{\text{loc}}) \leq 0$ holds, $\bar{x}$ might be infeasible since $g(\bar{x}, y_{\text{loc}}) \leq 0 < \varphi(\bar{x}) = g(\bar{x}, y_{\text{glob}})$ cannot be ruled out in general.

Since, in the following, we aim to use the approach from [18] and replace the lower level problem by its KKT conditions, we must make sure that a solution of the KKT system is a global maximizer. We emphasize that otherwise one might compute *infeasible points* for the semi-infinite problem, which is a major pitfall of the approach at hand. In particular, the concept of *substationary points* from [18] may entail infeasibility.

A natural assumption under which a solution of the KKT conditions leads to a global maximizer is convexity of the lower level problem, that is, for each $x \in \mathbb{R}^n$ the function $g(x, \cdot)$ is concave, and the set $Y(x)$ is convex. We thus make the following assumption throughout the remainder of this article.

Assumption 3.1 For all $x \in \mathbb{R}^n$ the lower level problem $Q(x)$ is convex.

Note that in generalized semi-infinite optimization many relevant applications have convex lower level problems (see also Sect. 5). On the other hand, in standard semi-infinite optimization this situation is rather rare. However, techniques using lower level convexity, like the one discussed in the present paper, can be combined with adaptive convexification methods like the one from [3] to treat semi-infinite programs with *nonconvex* lower level problems.

3.2 Semismooth optimality conditions for standard *SIP*

Let us first consider the standard semi-infinite case. As mentioned in the introduction, a standard semi-infinite problem has the form

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

with

$$M = \{ x \in \mathbb{R}^n | \, g(x, y) \leq 0 \text{ for all } y \in Y \}$$

and

$$Y = \{ y \in \mathbb{R}^m | \, v_j(y) \leq 0, \ j \in Q \}.$$

To formulate first and second order optimality conditions, the following notation will be used. For a continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, we denote the gradient in row form by $Df(x) = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n} \right]$, and in column form by $\nabla f(x) = D^T f(x)$. For a continuously differentiable function $F : \mathbb{R}^n \to \mathbb{R}^r$ we denote the Jacobian of F at $x \in \mathbb{R}^n$ by $DF(x)$ whereas the transposed Jacobian is $\nabla F(x)$. For a function $g : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}$ we denote by $\nabla_x g(x, y)$ the gradient of g at (x, y) with respect to x and by $\nabla^2_{xx} g(x, y)$, $\nabla^2_{xy} g(x, y) = D_y \nabla_x g(x, y)$ and $\nabla^2_{yy} g(x, y)$, the respective $n \times n$, $n \times r$ and $r \times r$ matrices of second order partial derivatives of g at (x, y).
For $\bar{x} \in M$ let

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

denote the (possibly infinite) *set of active indices* of $\bar{x}$. Note that $Y_0(\bar{x})$ coincides with the set of global maximizers of $Q(\bar{x})$ in the case $\varphi(\bar{x}) = 0$.

The *Extended Mangasarian-Fromovitz Constraint Qualification (EMFCQ)* holds at $\bar{x}$, if there is a vector $d \in \mathbb{R}^n$ such that

$$D_x g(\bar{x}, y)d < 0 \quad \text{for all} \quad y \in Y_0(\bar{x}).$$

Theorem 3.2 ([9]) *Let $\bar{x}$ be a local minimizer of SIP at which EMFCQ is satisfied. Then there are a $p \in \{0, \ldots, n\}$, multipliers $\bar{\mu}_i \geq 0$ and active indices $\bar{y}^i \in Y_0(\bar{x})$, $i \in P = \{1, 2, \ldots, p\}$, such that*

$$\nabla f(\bar{x}) + \sum_{i=1}^{p} \bar{\mu}_i \nabla_x g(\bar{x}, \bar{y}^i) = 0. \tag{3}$$

Next we complement the upper level first order condition from Theorem 3.2 by a lower level first order condition. In fact, since the active indices $\bar{y}^i \in Y_0(\bar{x})$, $i \in P$, are global solutions of $Q(\bar{x})$, under some constraint qualification like Slater's condition in the lower level problem (2), there exist vectors of Lagrange multipliers $\bar{\gamma}^i \in \mathbb{R}^q$ such that

$$\begin{gathered} \nabla_y g(\bar{x}, \bar{y}^i) - \textstyle\sum_{j=1}^{q} \bar{\gamma}^i_j \nabla_y v_j(\bar{y}^i) = 0, \quad i \in P, \\ \bar{\gamma}^i_j \geq 0, \quad v_j(\bar{y}^i) \leq 0, \quad \bar{\gamma}^i_j v_j(\bar{y}^i) = 0, \quad i \in P, \ j \in Q. \end{gathered} \tag{4}$$

Recall that by Assumption 3.1 the KKT conditions of the lower level problem are sufficient for feasibility in the semi-infinite problem.

By (3) and (4), we arrive at the following equalities and inequalities:

$$\begin{gathered} \nabla f(\bar{x}) + \textstyle\sum_{i=1}^{p} \bar{\mu}_i \nabla_x g(\bar{x}, \bar{y}^i) = 0, \\ \bar{\mu}_i \geq 0, \quad g(\bar{x}, \bar{y}^i) = 0, \quad i \in P, \\ \nabla_y g(\bar{x}, \bar{y}^i) - \textstyle\sum_{j=1}^{q} \bar{\gamma}^i_j \nabla_y v_j(\bar{y}^i) = 0, \quad i \in P, \\ \bar{\gamma}^i_j \geq 0, v_j(\bar{y}^i) \leq 0, \ \bar{\gamma}^i_j v_j(\bar{y}^i) = 0, \quad i \in P, \ j \in Q. \end{gathered} \tag{5}$$

Recall that a function $\psi : \mathbb{R}^2 \to \mathbb{R}$ is called an *NCP-function* [13] if

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

Here, NCP stands for *nonlinear complementarity problem*. Important examples of NCP functions are the *Fischer-Burmeister* function

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

and the *min* function

$$\psi_{min}(a, b) = -\min\{a, b\}.$$

With any NCP function ψ we may reformulate (5) as a system of equations:

$$\begin{array}{c} \nabla f(\bar{x}) + \sum_{i=1}^{p} \bar{\mu}_i \nabla_x g(\bar{x}, \bar{y}^i) = 0, \\ \psi(\bar{\mu}_i, -g(\bar{x}, \bar{y}^i)) = 0, \quad i \in P, \\ \nabla_y g(\bar{x}, \bar{y}^i) - \sum_{j=1}^{q} \bar{\gamma}_j^i \nabla_y v_j(\bar{y}^i) = 0, \quad i \in P, \\ \psi(\bar{\gamma}_j^i, -v_j(\bar{y}^i)) = 0, \quad i \in P,\ j \in Q. \end{array} \tag{6}$$

As observed in [18], the system of Eq. 6 is not directly equivalent to (5). The system (6) allows the case that

$$\bar{\mu}_i = 0, \quad g(\bar{x}, \bar{y}^i) < 0.$$

However, if there is an $n + (m + q + 1)p$ dimensional vector, say $(\bar{x}, \bar{\mu}, \bar{y}, \bar{\gamma})$, satisfying (6), we may drop the variables indexed by i with $\bar{\mu}_i = 0$. Thus, we get a solution of (5). On the other hand, a solution of (5) obviously satisfies (6). In this sense, (5) and (6) are *equivalent*, and the solution of (5) amounts to finding a zero of the function $T : \mathbb{R}^N \to \mathbb{R}^N$

$$T(z) = \begin{pmatrix} \nabla f(x) + \sum_{i=1}^{p} \mu_i \nabla_x g(x, y^i) \\ \psi(\mu_1, -g(x, y^1)) \\ \vdots \\ \psi(\mu_p, -g(x, y^p)) \\ \nabla_y g(x, y^1) - \sum_{j=1}^{q} \gamma_j^1 \nabla_y v_j(y^1) \\ \psi(\gamma_1^1, -v_1(y^1)) \\ \vdots \\ \psi(\gamma_q^1, -v_q(y^1)) \\ \vdots \\ \nabla_y g(x, y^p) - \sum_{j=1}^{q} \gamma_j^p \nabla_y v_j(y^p) \\ \psi(\gamma_1^p, -v_1(y^p)) \\ \vdots \\ \psi(\gamma_q^p, -v_q(y^p)) \end{pmatrix} \tag{7}$$

where $N = n + (m + q + 1)p$ with $z = (x^T, \mu^T, y^T, \gamma^T)^T \in \mathbb{R}^{n+(m+q+1)p}$, $x \in \mathbb{R}^n$, $\mu \in \mathbb{R}^p$, $y \in \mathbb{R}^{mp}$ and $\gamma \in \mathbb{R}^{qp}$.

Recall that we assume twice continously differentiable functions g and v. If, in the definition of T, we use the special NCP functions ψ_{FB} or ψ_{min}, then a result from [16] guarantees that T is *strongly semismooth*. We may thus apply the semismooth Newton method from Sect. 2 to T and expect q-quadratic convergence under CD-regularity in the solution point.

As we will see in Sect. 4, CD-regularity is strongly related to the so-called Reduction Ansatz in the lower level problem, which we present next. In addition, the Reduction Ansatz will also allow us to construct the function T for *generalized* semi-infinite optimization problems.

3.3 The Reduction Ansatz

Consider a feasible point $\bar{x}$ of *GSIP* and its set of active indices

$$Y_0(\bar{x}) = \{y \in Y(\bar{x}) |\ g(\bar{x}, y) = 0\}.$$

Suppose that the following conditions (Q-I)–(Q-III) hold at some $\bar{y} \in Y_0(\bar{x})$ in $Q(\bar{x})$:

(Q-I) The *linear independence constraint qualification*:

$(LI)_{Q(\bar{x})}$ $\{\nabla_y v_j(\bar{x}, \bar{y}) | j \in Q_0(\bar{x}, \bar{y})\}$ is a linearly independent family, where $Q_0(\bar{x}, \bar{y}) = \{j \in Q \mid v_j(\bar{x}, \bar{y}) = 0\}$ is the set of lower level active indices at $\bar{y} \in Y(\bar{x})$.
Because of Q-I, we have the following lower level Karush-Kuhn-Tucker conditions: there exists a unique vector of Lagrange multipliers $\bar{\gamma} \in \mathbb{R}^q$ such that

$$(\mathrm{KKT})_{Q(\bar{x})} \quad \left\{ \begin{array}{c} \nabla_y g(\bar{x}, \bar{y}) - \sum_{j=1}^q \bar{\gamma}_j \nabla_y v_j(\bar{x}, \bar{y}) = 0 \\ v_j(\bar{x}, \bar{y}) \leq 0 \\ \bar{\gamma}_j \geq 0 \\ \bar{\gamma}_j v_j(\bar{x}, \bar{y}) = 0,\ j \in Q. \end{array} \right\} \tag{8}$$

(Q-II) *Strict complementarity*: for each $j \in Q$: $\bar{\gamma}_j > 0$, $v_j(\bar{x}, \bar{y}) = 0$ or $\bar{\gamma}_j = 0$, $v_j(\bar{x}, \bar{y}) < 0$.

(Q-III) The *second order sufficiency condition*:

$$(SOSC)_{Q(\bar{x})} \quad \left\{ \begin{array}{l} \eta^T \nabla_y^2 \mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma}) \eta < 0 \ \text{ for all } \eta \in G_{Q(\bar{x})} \backslash \{0\}, \text{ where} \\ G_{Q(\bar{x})} = \{\eta \in \mathbb{R}^m |\ D_y v_j(\bar{x}, \bar{y}) \eta = 0,\ j \in Q_0(\bar{x}, \bar{y})\} \end{array} \right\} \tag{9}$$

with $\mathcal{L}(\bar{x}, y, \gamma) = g(\bar{x}, y) - \sum_{j=1}^q \gamma_j v_j(\bar{x}, y)$, the *Lagrangian* associated with $Q(\bar{x})$.

Under the conditions (Q-I) to (Q-III) $\bar{y}$ is called a *nondegenerate* global maximizer of the lower level problem in the sense of Jongen/Jonker/Twilt [10]. The *Reduction Ansatz* is said to hold at $\bar{x} \in M$ if *all* elements $\bar{y} \in Y_0(\bar{x})$ are nondegenerate for the lower level problem.

Suppose that the Reduction Ansatz holds, then we can reduce *GSIP* locally to a smooth finite optimization problem $GSIP_{red}$, the so-called reduced *GSIP*, as given in the next theorem.

Theorem 3.3 ([8]) *Let the Reduction Ansatz be satisfied at a feasible point $\bar{x}$ of GSIP. Then,*

(a) The active index set is finite, $Y_0(\bar{x}) = \{\bar{y}^1, \bar{y}^2, \ldots, \bar{y}^p\}$, and there exist neighborhoods $U_{\bar{x}}$ of $\bar{x}$ and $V_{\bar{y}^i}$ of $\bar{y}^i$ and unique C^1–functions
$y^i : U_{\bar{x}} \to V_{\bar{y}^i}$, where $y^i(\bar{x}) = \bar{y}^i$,
$\gamma^i : U_{\bar{x}} \to \mathbb{R}^q$, where $\gamma^i(\bar{x}) = \bar{\gamma}^i$,
such that for every $x \in U_{\bar{x}}$ the value $y^i(x)$ is the unique local maximizer of $Q(x)$ in $V_{\bar{y}^i}$ with corresponding Lagrange multiplier vector $\gamma^i(x)$.

(b) The following finite reduction holds: $\bar{x}$ is a solution of GSIP, locally in a neighborhood $U_{\bar{x}}$ of $\bar{x}$, if and only if $\bar{x}$ is a local solution of the so-called reduced problem

$$GSIP_{red}: \min_{x \in U_{\bar{x}}} f(x) \quad s.t. \ \varphi_i(x) = g(x, y^i(x)) \leq 0, \quad for\ all\ i = 1, 2, \ldots, p.$$

(c) The functions φ_i from part (b) are of class C^2, and for all $x \in U_{\bar{x}}$ their gradients satisfy

$$D_x \varphi_i(x) = D_x \mathcal{L}(x, y^i(x), \gamma^i(x)). \tag{10}$$

Remark 3.4 For standard *SIP* the formula in Theorem 3.3(c) simplifies to

$$D_x \varphi_i(x) = D_x g(x, y^i(x)).$$

3.4 Semismooth optimality conditions for *GSIP*

If $\bar{x} \in M$ is a local minimizer of *GSIP* at which the Reduction Ansatz holds, then, by Theorem 3.3, $\bar{x}$ is also a local minimizer of the locally reduced problem $GSIP_{red}$, and necessary optimality conditions for $\bar{x}$ in the reduced problem are also necessary optimality conditions for $\bar{x}$ in the original problem. In particular, if $\bar{x} \in M$ is a local minimizer of *GSIP* at which the Reduction Ansatz and the Mangasarian-Fromovitz constraint qualification hold, then there exist a $p \in \{1, \ldots, n\}$, and multipliers $\bar{\mu}_i \geq 0$, $i \in P = \{1, \ldots, p\}$, such that

$$\nabla f(\bar{x}) + \sum_{i=1}^{p} \bar{\mu}_i \nabla \varphi_i(\bar{x}) = 0.$$

Note that all constraints in $GSIP_{red}$ are active at $\bar{x}$ by construction, that is, we have $\varphi_i(\bar{x}) = 0$ for all $i \in P$.

Using $\varphi_i(\bar{x}) = g(\bar{x}, \bar{y}^i)$, the fact that each $\bar{y}^i$ is a global maximizer of the lower level problem, as well as the evaluation of (10) at $\bar{x}$, we arrive at the system

$$\begin{aligned}
&\nabla f(\bar{x}) + \textstyle\sum_{i=1}^{p} \bar{\mu}_i \nabla_x \mathcal{L}(\bar{x}, \bar{y}^i, \bar{\gamma}^i) = 0,\\
&\bar{\mu}_i \geq 0, \quad g(\bar{x}, \bar{y}^i) = 0, \quad i \in P,\\
&\nabla_y g(\bar{x}, \bar{y}^i) - \textstyle\sum_{j=1}^{q} \bar{\gamma}_j^i \nabla_y v_j(\bar{x}, \bar{y}^i) = 0, \quad i \in P,\\
&\bar{\gamma}_j^i \geq 0, \ v_j(\bar{x}, \bar{y}^i) \leq 0, \quad \bar{\gamma}_j^i v_j(\bar{x}, \bar{y}^i) = 0, \quad i \in P, \ j \in Q.
\end{aligned} \tag{11}$$

With any NCP function ψ, along the same lines as in Sect. 3.2 the solution of (11) is seen to be equivalent to finding a zero of the function

$$T(z) = \begin{pmatrix} \nabla f(x) + \sum_{i=1}^{p} \mu_i \nabla_x \mathcal{L}(x, y^i, \gamma^i) \\ \psi(\mu_1, -g(x, y^1)) \\ \vdots \\ \psi(\mu_p, -g(x, y^p)) \\ \nabla_y g(x, y^1) - \sum_{j=1}^{q} \gamma_j^1 \nabla_y v_j(x, y^1) \\ \psi(\gamma_1^1, -v_1(x, y^1)) \\ \vdots \\ \psi(\gamma_q^1, -v_q(x, y^1)) \\ \vdots \\ \nabla_y g(x, y^p) - \sum_{j=1}^{q} \gamma_j^p \nabla_y v_j(x, y^p) \\ \psi(\gamma_1^p, -v_1(x, y^p)) \\ \vdots \\ \psi(\gamma_q^p, -v_q(x, y^p)) \end{pmatrix} \tag{12}$$

with $z = (x^T, \mu^T, y^T, \gamma^T)^T \in \mathbb{R}^N$, $N = n + (m + q + 1)p$, $x \in \mathbb{R}^n$, $\mu \in \mathbb{R}^p$, $y \in \mathbb{R}^{mp}$ and $\gamma \in \mathbb{R}^{qp}$. Again, T is strongly semismooth under our assumptions.

We emphasize that for a standard *SIP* the term $\nabla_x \mathcal{L}(x, y^i, \gamma^i)$ in T is replaced by $\nabla_x g(x, y^i)$, and $v_j(x, y^i)$ by $v_j(y^i)$, that is, the function T from (12) generalizes the function T from (7) from *SIP* to *GSIP*. Note that for standard *SIP* it is not necessary to assume the Reduction Ansatz to derive the function T. However, the situation for *GSIP* is not essentially more restrictive, since below we will anyway assume the Reduction Ansatz at the solution point for our convergence result, and the Reduction Ansatz at local minimizers of *GSIP* is a weak assumption [5].

4 Convergence of the semismooth Newton method

In this section we wish to apply the semismooth Newton approach from Sect. 2 to find a zero of the function T from (12). In particular, we want to use Theorem 2.1 and, thus, find a sufficient condition for CD-regularity of T in a solution point $\bar{z}$.

The part of this condition concerning the lower level problem will be the Reduction Ansatz, while in the upper level problem we will assume the so-called *Robinson condition* [20].

In fact, consider $\bar{x} \in M$ and the locally reduced problem $GSIP_{red}$ where, according to the definition of T, we neglect the fact that all constraints φ_i, $i \in P$, are active by their definition. Let $P_0(\bar{x}) = \{i \in P \mid \varphi_i(\bar{x}) = 0\}$ be the set of active indices at $\bar{x}$ for the upper level problem. Recall from Theorem 3.3(c) that the auxiliary functions φ_i, $i \in P$, in the reduced problem $GSIP_{red}$ are twice continuously differentiable, so that it makes sense to impose a second order regularity condition on $GSIP_{red}$.

The Robinson condition is said to hold at $\bar{x}$ if the following conditions (GSIP-I) and (GSIP-II) are satisfied:

(GSIP-I) The *linear independence constraint qualification*:

$(LI)_{GSIP}$ $\{\nabla_x \mathcal{L}(\bar{x}, \bar{y}^i, \bar{\gamma}^i) \mid i \in P_0(\bar{x})\}$ is a linearly independent family.

If $\bar{x}$ is a local minimizer, then there exists a unique vector $\bar{\mu} \in \mathbb{R}^p$ of Lagrange multipliers with

$$(\mathrm{KKT})_{GSIP} \quad \nabla f(\bar{x}) + \sum_{i=1}^{p} \bar{\mu}_i \nabla_x \mathcal{L}(\bar{x}, \bar{y}^i, \bar{\gamma}^i) = 0,$$
$$\bar{\mu}_i \geq 0,\ g(\bar{x}, \bar{y}^i) \leq 0,\ \bar{\mu}_i g(\bar{x}, \bar{y}^i) = 0,\ i \in P.$$

(GSIP-II) The *strong second order sufficiency condition*:

$$(SSOSC)_{GSIP} \quad \left\{ \begin{array}{l} \xi^T \nabla_x^2 L(\bar{x}, \bar{y}, \bar{\mu}, \bar{\gamma}) \xi > 0 \text{ for all } \xi \in G_{GSIP} \backslash \{0\} \text{ with} \\ G_{GSIP} = \{d \in \mathbb{R}^n \mid D_x \mathcal{L}(\bar{x}, \bar{y}^i, \bar{\gamma}^i) d = 0 \text{ for } i \in P_+(\bar{x})\}. \end{array} \right\}$$

with $P_+(\bar{x}) = \{i \in P_0(\bar{x}) \mid \bar{\mu}_i > 0\}$. Here, $\nabla_x^2 L(\bar{x}, \bar{y}, \bar{\mu}, \bar{\gamma})$ stands for the Hessian of the Lagrangian $f(x) + \sum_{i=1}^{p} \mu_i \varphi_i(x)$ of $GSIP_{red}$, that is, for $\nabla_x^2 f(\bar{x}) + \sum_{i=1}^{p} \bar{\mu}_i \nabla_x^2 \varphi_i(\bar{x})$, with

$$\nabla_x^2 \varphi_i(\bar{x}) = \nabla_x^2 \mathcal{L}(\bar{x}, \bar{y}^i, \bar{\gamma}^i) - \begin{pmatrix} \nabla_{yx}^2 \mathcal{L}_i \\ -D_x v_{Q_0^i}(\bar{x}, \bar{y}^i) \end{pmatrix}^T \begin{pmatrix} D_y^2 \mathcal{L}_i & -\nabla_y v_{Q_0^i}(\bar{x}, \bar{y}^i) \\ -D_y v_{Q_0^i}(\bar{x}, \bar{y}^i) & 0 \end{pmatrix}^{-1} \begin{pmatrix} \nabla_{yx}^2 \mathcal{L}_i \\ -D_x v_{Q_0^i}(\bar{x}, \bar{y}^i) \end{pmatrix}$$

where $D_x v_{Q_0^i}$ stands for the matrix with rows $D_x v_j,\ j \in Q_0^i = Q_0(\bar{x}, \bar{y}^i)$. $D_y v_{Q_0^i}$ is defined similarly.

It is well-known [13] that the Robinson condition at $\bar{x}$ just means that $\bar{x}$ is a strongly stable local minimizer. Note that the Robinson condition does *not* assume strict complementary slackness. In the standard *SIP* case, $\nabla_x \mathcal{L}(\bar{x}, \bar{y}^i, \bar{\gamma}^i)$ reduces to $\nabla_x g(\bar{x}, \bar{y}^i)$, and in the Hessian of the Lagrangian we obtain

$$\nabla_x^2 \varphi_i(\bar{x}) = \nabla_x^2 g(\bar{x}, \bar{y}^i) - \begin{pmatrix} \nabla_{yx}^2 g(\bar{x}, \bar{y}^i) \\ 0 \end{pmatrix}^T \begin{pmatrix} D_y^2 \mathcal{L}_i & -\nabla v_{Q_0^i}(\bar{y}^i) \\ -D v_{Q_0^i}(\bar{y}^i) & 0 \end{pmatrix}^{-1} \begin{pmatrix} \nabla_{yx}^2 g(\bar{x}, \bar{y}^i) \\ 0 \end{pmatrix}$$

Now consider a zero z of T from (12). Then the KKT conditions mentioned in (Q-I) and (GSIP-I) hold by definition of T. The remaining conditions in the Reduction Ansatz and the Robinson condition are algebraic conditions on the involved functions which can also be imposed independently of the fact that we deal with lower level global maximizers and upper level local minimizers. In this sense, we can make the following assumption:

Assumption 4.1 The Reduction Ansatz and the Robinson condition hold at z.

Theorem 4.2 *Suppose that $\bar{z} = (\bar{x}^T, \bar{\mu}^T, \bar{y}^T, \bar{\gamma}^T)^T$ is a zero of T from (12) with the choices $\psi = \psi_{FB}$ or $\psi = \psi_{min}$, and that Assumption 4.1 holds at $\bar{z}$. Then T is CD-regular at $\bar{z}$.*

Proof Let $\bar{z}$ be a zero of T at which Assumption 4.1 holds, and let ψ denote either one of the two NCP functions ψ_{FB} and ψ_{min}. We only consider the case of two active indices, that is, $p = 2$, the general case running along the same lines. We will distinguish two cases, depending on whether upper level strict complementary slackness holds or not.

Case 1 Strict complementarity holds in the upper level problem.

We have $\psi(\bar{\mu}_i, -g(\bar{x}, \bar{y}^i)) = 0$ and $(\bar{\mu}_i, -g(\bar{x}, \bar{y}^i)) \neq 0$, so that ψ is differentiable at $(\bar{\mu}_i, -g(\bar{x}, \bar{y}^i))$ with gradient

$$D\psi(\bar{\mu}_i, -g(\bar{x}, \bar{y}^i)) = \begin{cases} (-1, 0), & i \notin P_0(\bar{x}) \\ (0, -1), & i \in P_0(\bar{x}) \end{cases}$$

for $i = 1, 2$ (see (25) in Appendix A.1).

The Reduction Ansatz in the lower level problem implies $\psi(\bar{\gamma}_j^i, -v_j(\bar{x}, \bar{y}^i)) = 0$ and $(\bar{\gamma}_j^i, -v_j(\bar{x}, \bar{y}^i)) \neq 0$, so that ψ is also differentiable at $(\bar{\gamma}_j^i, -v_j(\bar{x}, \bar{y}^i))$ with gradient

$$D\psi(\bar{\gamma}_j^i, -v_j(\bar{x}, \bar{y}^i)) = \begin{cases} (-1, 0), & j \notin Q_0(\bar{x}, \bar{y}^i) \\ (0, -1), & j \in Q_0(\bar{x}, \bar{y}^i) \end{cases}$$

for $i = 1, 2$ and $j \in Q$. Together this means that T is differentiable at $\bar{z}$, and its Jacobian is the matrix

$$DT(\bar{z}) = \begin{pmatrix} \mathcal{B}_{11} & \mathcal{B}_{12} & \mathcal{B}_{13} \\ \mathcal{B}_{21} & \mathcal{B}_{22} & 0 \\ \mathcal{B}_{31} & 0 & \mathcal{B}_{33} \end{pmatrix} \tag{13}$$

with the following blocks:

$$\mathcal{B}_{11} = \begin{pmatrix} D_x^2 f(\bar{x}) + \sum_{i=1}^2 \bar{\mu}_i \nabla_x^2 \mathcal{L}_i & \nabla_x \mathcal{L}_1 & \nabla_x \mathcal{L}_2 \\ \lambda_1 D_x g(\bar{x}, \bar{y}^1) & \theta_1 & 0 \\ \lambda_2 D_x g(\bar{x}, \bar{y}^2) & 0 & \theta_2 \end{pmatrix}$$

with $\nabla_x^2 \mathcal{L}_i = \nabla_x^2 \mathcal{L}(\bar{x}, \bar{y}^i, \bar{\gamma}^i)$, etc.,

$$\lambda_i = \begin{cases} 0, & i \notin P_0(\bar{x}) \\ 1, & i \in P_0(\bar{x}) \end{cases}, \quad \theta_i = \begin{cases} -1, & i \notin P_0(\bar{x}) \\ 0, & i \in P_0(\bar{x}) \end{cases}, \quad i = 1, 2,$$

$$\mathcal{B}_{12} = \begin{pmatrix} \bar{\mu}_1 \nabla_{xy}^2 \mathcal{L}_1 & -\bar{\mu}_1 \nabla_x v(\bar{x}, \bar{y}^1) \\ \lambda_1 D_y g(\bar{x}, \bar{y}^1) & 0 \\ 0 & 0 \end{pmatrix},$$

$$\mathcal{B}_{13} = \begin{pmatrix} \bar{\mu}_2 \nabla_{xy}^2 \mathcal{L}_2 & -\bar{\mu}_2 \nabla_x v(\bar{x}, \bar{y}^2) \\ 0 & 0 \\ \lambda_2 D_y g(\bar{x}, \bar{y}^2) & 0 \end{pmatrix},$$

and with

$$\alpha_j^i = \begin{cases} 0, & j \notin Q_0(\bar{x}, \bar{y}^i) \\ 1, & j \in Q_0(\bar{x}, \bar{y}^i) \end{cases}, \quad \beta_j^i = \begin{cases} -1, & j \notin Q_0(\bar{x}, \bar{y}^i) \\ 0, & j \in Q_0(\bar{x}, \bar{y}^i) \end{cases},$$

$$\mathcal{B}_{i+1,1} = \begin{pmatrix} \nabla_{yx}^2 \mathcal{L}_i & 0 & 0 \\ \operatorname{diag}(\alpha^i) D_x v(\bar{x}, \bar{y}^i) & 0 & 0 \end{pmatrix}$$

as well as

$$\mathcal{B}_{i+1,i+1} = \begin{pmatrix} \nabla_y^2 \mathcal{L}_i & -\nabla_y v(\bar{x}, \bar{y}^i) \\ \operatorname{diag}(\alpha^i) D_y v(\bar{x}, \bar{y}^i) & \operatorname{diag}(\beta^i) \end{pmatrix}, \quad i = 1, 2.$$

Our aim is to show that $DT(\bar{z})$ is nonsingular under Assumption 4.1. The main idea of the proof is to consider an appropriate Schur complement in the block matrix $DT(\bar{z})$ (see Appendix A.2).

Note that the matrices $\mathcal{B}_{22}$ and $\mathcal{B}_{33}$ are nonsingular under the Reduction Ansatz. In fact, by the definitions of α^1 and β^1, $\mathcal{B}_{22}$ is nonsingular if and only if the matrix

$$\begin{pmatrix} \nabla_y^2 \mathcal{L}_1 & \nabla_y v_{Q_0^1}(\bar{x}, \bar{y}^1) \\ D_y v_{Q_0^1}(\bar{x}, \bar{y}^1) & 0 \end{pmatrix}$$

is nonsingular. Since under the Reduction Ansatz we have $(LI)_{Q(\bar{x})}$ and $(SOSC)_{Q(\bar{x})}$ at $\bar{y}^1$, the latter matrix is nonsingular by Lemma A.4. Analogously the nonsingularity of $\mathcal{B}_{33}$ is shown.

Consequently, according to Lemma A.3 the matrix $DT(\bar{z})$ is nonsingular if and only if the Schur complement

$$\begin{aligned} S = DT(\bar{z})/\begin{pmatrix} \mathcal{B}_{22} & 0 \\ 0 & \mathcal{B}_{33} \end{pmatrix} &= \mathcal{B}_{11} - (\mathcal{B}_{12},\ \mathcal{B}_{13}) \begin{pmatrix} \mathcal{B}_{22} & 0 \\ 0 & \mathcal{B}_{33} \end{pmatrix}^{-1} \begin{pmatrix} \mathcal{B}_{21} \\ \mathcal{B}_{31} \end{pmatrix} \\ &= \mathcal{B}_{11} - \mathcal{B}_{12}\mathcal{B}_{22}^{-1}\mathcal{B}_{21} - \mathcal{B}_{13}\mathcal{B}_{33}^{-1}\mathcal{B}_{31} \end{aligned}$$

is nonsingular. We will show that the latter is the case under the Robinson condition.

Let us calculate $\mathcal{B}_{12}\mathcal{B}_{22}^{-1}\mathcal{B}_{21}$. From the block structures of $\mathcal{B}_{12}$ and $\mathcal{B}_{21}$ it is clear that only the first and second block in the first block column of this matrix are nonzero, and we only have to calculate

$$\begin{pmatrix} \bar{\mu}_1 \nabla^2_{xy}\mathcal{L}_1 & -\bar{\mu}_1 \nabla_x v(\bar{x}, \bar{y}^1) \\ \lambda_1 D_y g(\bar{x}, \bar{y}^1) & 0 \end{pmatrix} \mathcal{B}_{22}^{-1} \begin{pmatrix} \nabla^2_{yx}\mathcal{L}_1 \\ \operatorname{diag}(\alpha^1) D_x v(\bar{x}, \bar{y}^1) \end{pmatrix}. \tag{14}$$

Now recall that the implicit functions $y^1(x)$ and $\gamma^1(x)$ from Theorem 3.3 satisfy

$$\begin{pmatrix} \nabla_y \mathcal{L}(x, y^1(x), \gamma^1(x)) \\ \psi(\gamma_1^1(x), -v_1(x, y^1(x))) \\ \vdots \\ \psi(\gamma_q^1(x), -v_q(x, y^1(x))) \end{pmatrix} \equiv 0. \tag{15}$$

Taking derivatives with respect to x and evaluating at $\bar{x}$ yields

$$\begin{aligned} &\begin{pmatrix} \nabla^2_y \mathcal{L}_1 & -\nabla_y v(\bar{x}, \bar{y}^1) \\ \operatorname{diag}(\alpha^1) D_y v(\bar{x}, \bar{y}^1) & \operatorname{diag}(\beta^1) \end{pmatrix} \begin{pmatrix} Dy^1(\bar{x}) \\ D\gamma^1(\bar{x}) \end{pmatrix} \\ &+ \begin{pmatrix} \nabla^2_{yx}\mathcal{L}_1 \\ \operatorname{diag}(\alpha^1) D_x v(\bar{x}, \bar{y}^1) \end{pmatrix} = 0 \end{aligned}$$

and, thus,

$$\mathcal{B}_{22}^{-1} \begin{pmatrix} \nabla^2_{yx}\mathcal{L}_1 \\ \operatorname{diag}(\alpha^1) D_x v(\bar{x}, \bar{y}^1) \end{pmatrix} = - \begin{pmatrix} Dy^1(\bar{x}) \\ D\gamma^1(\bar{x}) \end{pmatrix}.$$

Consequently, the matrix in (14) becomes

$$\begin{pmatrix} -\bar{\mu}_1 \nabla^2_{xy}\mathcal{L}_1 Dy^1(\bar{x}) + \bar{\mu}_1 \nabla_x v(\bar{x}, \bar{y}^1) D\gamma^1(\bar{x}) \\ -\lambda_1 D_y g(\bar{x}, \bar{y}^1) Dy^1(\bar{x}) \end{pmatrix}.$$

With an analogous calculation for $\mathcal{B}_{13}\mathcal{B}_{33}^{-1}\mathcal{B}_{31}$ we arrive at

$$S = \begin{pmatrix} \widetilde{\nabla^2_x L} & \nabla_x \mathcal{L}_1 & \nabla_x \mathcal{L}_2 \\ \lambda_1 \left(D_x g(\bar{x}, \bar{y}^1) + D_y g(\bar{x}, \bar{y}^1) Dy^1(\bar{x}) \right) & \theta_1 & 0 \\ \lambda_2 \left(D_x g(\bar{x}, \bar{y}^2) + D_y g(\bar{x}, \bar{y}^2) Dy^2(\bar{x}) \right) & 0 & \theta_2 \end{pmatrix}$$

with

$$\widetilde{\nabla^2_x L} = D^2_x f(\bar{x}) + \sum_{i=1}^{2} \bar{\mu}_i \left(\nabla^2_x \mathcal{L}_i + \nabla^2_{xy} \mathcal{L}_i Dy^i(\bar{x}) - \nabla_x v(\bar{x}, \bar{y}^i) D\gamma^i(\bar{x}) \right).$$

Using (10) it is not hard to see that

$$\nabla^2\varphi_i(\bar{x}) \;=\; \nabla_x^2\mathcal{L}_i + \nabla_{xy}^2\mathcal{L}_i Dy^i(\bar{x}) - \nabla_x v(\bar{x},\bar{y}^i)D\gamma^i(\bar{x})$$

holds for $i = 1, 2$, so that the matrix $\widetilde{\nabla_x^2 L}$ coincides with $\nabla_x^2 L$ from condition (GSIP-II). Next we show

$$D_x g(\bar{x},\bar{y}^i) + D_y g(\bar{x},\bar{y}^i)Dy^i(\bar{x}) \;=\; D_x\mathcal{L}_i \tag{16}$$

for $i = 1, 2$. In fact, for $i = 1$ we have

$$D_y g(\bar{x},\bar{y}^1)Dy^1(\bar{x}) \;=\; \sum_{j=1}^{q} \bar{\gamma}_j^1 D_y v_j(\bar{x},\bar{y}^1)Dy^1(\bar{x}). \tag{17}$$

The implicit functions from Theorem 3.3 particularly satisfy the identity

$$\gamma_j^1(x)\cdot v_j(x,y^1(x)) \;\equiv\; 0,\;\; j\in Q.$$

Taking derivatives with respect to x yields

$$D\gamma_j^1(x)\, v_j(x,y^1(x)) + \gamma_j^1(x)\,\big(D_x v_j(x,y^1(x)) + D_y v_j(x,y^1(x))Dy^1(x)\big) \;\equiv\; 0$$

for $j \in Q$, where the first term vanishes at $\bar{x}$ for $j \in Q_0(\bar{x},\bar{y}^1)$ due to $v_j(\bar{x},\bar{y}^1) = 0$, and for $j \notin Q_0(\bar{x},\bar{y}^1)$ because of $\gamma_j^1(x) \equiv 0$ and, thus, $D\gamma_j^1(\bar{x}) = 0$. Evaluating the remaining terms at $\bar{x}$ implies

$$\bar{\gamma}_j^1 D_y v_j(\bar{x},\bar{y}^1)Dy^1(\bar{x}) \;=\; -\bar{\gamma}_j^1 D_x v_j(\bar{x},\bar{y}^1), \quad j\in Q, \tag{18}$$

and a combination of (17) and (18) yields

$$D_y g(\bar{x},\bar{y}^1)Dy^1(\bar{x}) \;=\; -\sum_{j=1}^{q} \bar{\gamma}_j^1 D_x v_j(\bar{x},\bar{y}^1).$$

This shows (16) for $i = 1$, and analogously for $i = 2$. As a consequence, the Schur complement further simplifies to

$$S \;=\; \begin{pmatrix} \nabla_x^2 L & \nabla_x\mathcal{L}_1 & \nabla_x\mathcal{L}_2 \\ \lambda_1 D_x\mathcal{L}_1 & \theta_1 & 0 \\ \lambda_2 D_x\mathcal{L}_2 & 0 & \theta_2 \end{pmatrix}.$$

By the definitions of λ_i and θ_i, $i = 1, 2$, S is nonsingular if and only if the matrix

$$\begin{pmatrix} \nabla_x^2 L & \nabla_x\mathcal{L}_{P_0(\bar{x})} \\ D_x\mathcal{L}_{P_0(\bar{x})} & 0 \end{pmatrix}$$

is nonsingular. Under the Robinson condition the latter is true in view of Lemma A.4. This completes the proof for Case 1.

Case 2 Strict complementarity is violated in the upper level problem.

In this case we have $\bar{\mu}_i = g(\bar{x},\bar{y}^i) = 0$ for at least one $i \in \{1, 2\}$. Here we only consider the case that strict complementarity is violated at $i = 1$ with $P_0(\bar{x}) = \{1, 2\}$, the general case running along the same lines.

In the present case T is *not* differentiable at $\bar{z}$ since ψ is not differentiable at the origin. The Clarke (in fact, convex) subdifferentials of ψ_{FB} and ψ_{min} are given in Lemma A.1 in the Appendix. Moreover, the generalized Jacobian of $\psi(\mu_1, -g(x, y^1))$ can be computed by

the Chain Rule II and Proposition 2.3.6 from [2], noting that convex functions are regular. In fact, its generalized Jacobian with respect to (x, μ_1, y^1) is

$$\left\{\left(\lambda_1 D_x g(x, y^1),\ \theta_1\ ,\ \lambda_1 D_y g(x, y^1)\right) \mid (-\lambda_1, \theta_1) \in \partial\psi(0, 0)\right\}. \tag{19}$$

This means that the elements of $\partial T(\bar{z})$ can be parameterized by

$$\{W(\lambda_1, \theta_1) \mid (-\lambda_1, \theta_1) \in \partial\psi(0, 0)\},$$

where $W(\lambda_1, \theta_1)$ is a matrix of exactly the form from (13). Consequently, proving CD-regularity of T at $\bar{z}$ amounts to showing nonsingularity of all matrices $W(\lambda_1, \theta_1)$ with $(-\lambda_1, \theta_1) \in \partial\psi(0, 0)$.

Choose any (λ_1, θ_1) with $(-\lambda_1, \theta_1) \in \partial\psi(0, 0)$. With the same arguments as in Case 1 we find that $W(\lambda_1, \theta_1)$ is nonsingular if and only if the matrix

$$S(\lambda_1, \theta_1) = \begin{pmatrix} \nabla_x^2 L & \nabla_x \mathcal{L}_1 & \nabla_x \mathcal{L}_2 \\ \lambda_1 D_x \mathcal{L}_1 & \theta_1 & 0 \\ \lambda_2 D_x \mathcal{L}_2 & 0 & \theta_2 \end{pmatrix}$$

is nonsingular. The latter, however, is true by [16, Theorem 4.2] for finitely constrained programming problems. This completes the proof. □

Remark 4.3 In the special case of SIP, an explicit proof of Theorem 4.2 would be shorter due to the simplifications that $\nabla_x \mathcal{L}_i$ is replaced by $\nabla_x g(\bar{x}, \bar{y}^i)$, $\nabla_{xy} \mathcal{L}_i$ by $\nabla_{xy} g(\bar{x}, \bar{y}^i)$, and $\nabla_x v(\bar{x}, \bar{y}^i)$ vanishes.

Remark 4.4 Assumption 4.1 is a weak assumption in the following sense: generically, the Reduction Ansatz holds at all local minimizers of *GSIP*, and they are even nondegenerate critical points of the locally reduced problem $GSIP_{red}$. That is, generically even upper level strict complementarity and the Robinson condition hold. While this fact has been known for standard *SIP* for some time [21], it is a recent result for *GSIP* [5]. In view of this genericity, one can expect Assumption 4.1 to be satisfied in practical applications.

Altogether, in Sect. 4 we have shown that under the weak Assumption 4.1, in view of Theorems 2.1 and 4.2, the semismooth Newton method from (1) converges q-quadratically.

5 Numerical examples

In this section we report on some numerical experiments with the generalized damped semismooth Newton approach proposed in [18]. In fact, let

$$\theta(z) = \frac{1}{2} T(z)^T T(z)$$

be the merit function. If ψ_{FB} is used, then θ is continuously differentiable with the gradient given by

$$\nabla\theta(z) = W^T T(z),$$

where $W \in \partial T(z)$, the generalized Jacobian of T at z. In the case of ψ_{min}, the merit function is not continuously differentiable. We use ψ_{FB} whenever the gradient of the merit function is needed in the algorithm.

Algorithm 5.1 ([18])
Step 1. Let $z^0 \in \mathbb{R}^N,\ \sigma, \rho \in (0,1),\ \eta > 0,\ a > 2$ and $k = 0$.
Step 2. If $T(z^k) = 0$, stop. Otherwise, let d^k be a solution of

$$T(z^k) + W^k d = 0, \tag{20}$$

where $W^k \in \partial T(z^k)$.

If (20) is not solvable, or if

$$\nabla\theta(z^k)^T d^k > -\eta \left\| d^k \right\|^a,$$

set $d^k = -\nabla\theta(z^k)$
Step 3. Find a minimum nonnegative integer, say, m_k, such that

$$\theta(z^k + \rho^{m_k} d^k) \le \theta(z^k) + \sigma\rho^{m_k}\nabla\theta(z^k)^T d^k,$$

Let $\alpha_k = \rho^{m_k}$.
Step 4. Let $z^{k+1} = z^k + \alpha_k d^k$ and $k = k + 1$. Go to Step 2.

For the implementation of the algorithm, at iterates which are differentiability points of T we do not use the Jacobian matrix DT in the form from the proof of Theorem 4.2, since the iterates cannot be expected to be zeros of T. Thus we may not use the simplified gradients of the NCP functions from (25), but the ones from (23), (24). For ψ_{FB} this results in replacing $\lambda_i,\ \theta_i,\ \alpha_j^i,\ \beta_j^i,\ i \in P,\ j \in Q$, by

$$\lambda_i = \frac{g(\bar{x}, \bar{y}^i)}{\sqrt{\bar{\mu}_i^2 + g(\bar{x}, \bar{y}^i)^2}} + 1, \quad \theta_i = \frac{\bar{\mu}_i}{\sqrt{\bar{\mu}_i^2 + g(\bar{x}, \bar{y}^i)^2}} - 1, \tag{21}$$

and

$$\alpha_j^i = \frac{v_j(\bar{x}, \bar{y}^i)}{\sqrt{(\bar{\gamma}_j^i)^2 + v_j(\bar{x}, \bar{y}^i)^2}} + 1, \quad \beta_j^i = \frac{\bar{\gamma}_j^i}{\sqrt{(\bar{\gamma}_j^i)^2 + v_j(\bar{x}, \bar{y}^i)^2}} - 1. \tag{22}$$

whereas for ψ_{min} the gradients in (24) and (25) coincide.

At nondifferentiability points of T we choose the element W from the generalized Jacobian of T which corresponds to the midpoints of the subdifferentials of the NCP functions. In view of (19) and Lemma A.1 this means that for ψ_{FB} we use $\lambda_i = 1,\ \theta_i = -1$, and for ψ_{min} we use $\lambda_i = 1/2,\ \theta_i = -1/2$.

The Algorithm 5.1 is implemented in MATLAB 7.3. Throughout the computational experiments, the parameters used in the algorithm are $\rho = 0.5,\ a = 2.1,\ \eta = 10^{-8}$ and $\sigma = 0.1$. The algorithm is terminated when $||T(z^k)|| < 10^{-6}$.

In the numerical examples we test both ψ_{FB} and ψ_{min} as the NCP function. However, due to the mentioned smoothness properties, in the merit function we use ψ_{FB} for both cases.

The test problems in Examples 5.2 and 5.3 are taken from [22].

Example 5.2 Design Centering:

In a general design centering problem, the aim is to maximize some measure (e.g., the volume $\mathrm{Vol}(B(x))$) of a body $B(x)$ depending on a parameter under the constraint that $B(x)$ is contained in a given fixed body G:

$$\max_{x \in \mathbb{R}^n} \mathrm{Vol}(B(x)) \quad \text{s.t.} \quad B(x) \subset G$$

Let $G = \{y \in \mathbb{R}^2 \mid g(y) \leq 0\}$ with

$$g(y) = \begin{pmatrix} -y_1 - y_2^2 \\ y_1/4 + y_2 - 3/4 \\ -y_2 - 1 \end{pmatrix}.$$

The *GSIP* formulation of the general design centering problem is as follows:

$$\max_{x \in \mathbb{R}^n} \text{Vol}(B(x)) \text{ s.t. } g(y) \leq 0 \, forall \, y \in B(x).$$

Problem 1 The aim is to find the largest disc with free center and radius inscribed in G. We then have $n = 3$ and

$$B(x) = \{y \in \mathbb{R}^2 \mid (y_1 - x_1)^2 + (y_2 - x_2)^2 - x_3^2 \leq 0\}, \quad \text{Vol}(B(x)) = \pi x_3^2.$$

In the ψ_{FB} case, with the starting point $x^0 = (0, 0, 1)$ the semismooth Newton method obtains the optimal value 1.8606 with $\bar{x}$=(0.749, −0.230, 0.770) and $\bar{y}^1$=(−0.008, −0.091), $\bar{y}^2$ = (0.935, 0.516), $\bar{y}^3 = (0.749, -1)$ for the optimal point. We have $||T(\bar{z})|| = 7.1239 \cdot 10^{-10}$ after 4 iterations within 0.23 seconds of CPU time. In the ψ_{min} case, the optimal point and the optimal value are obtained in 4 iterations with $||T(\bar{z})|| = 3.3466 \cdot 10^{-13}$ within 0.42 seconds of CPU time.

Problem 2 The aim is to find the largest ellipse with free center and axis lengths inscribed in G. We have $n = 4$ and

$$B(x) = \left\{ y \in \mathbb{R}^2 \mid \frac{(y_1 - x_1)^2}{x_3^2} + \frac{(y_2 - x_2)^2}{x_4^2} - 1 \leq 0 \right\}, \quad \text{Vol}(B(x)) = \pi x_3 x_4.$$

In the ψ_{FB} case, with the starting point $x^0 = (0, 0, 1, 1)$ the semismooth Newton method obtains the optimal value 3.484 with $\bar{x} = (2.013, -0.5, 2.217, 0.5)$ and $\bar{y}^1$=(−0.167, −0.408), $\bar{y}^2 = (3.658, -0.165)$, $\bar{y}^3 = (2.013, -1)$ for the optimal point. We have $||T(\bar{z})|| = 3.3603 \cdot 10^{-10}$ after 6 iterations within 0.34 seconds of CPU time. In the ψ_{min} case, the optimal point and the optimal value are obtained in 6 iterations with $||T(\bar{z})|| = 2.9269 \cdot 10^{-11}$ within 0.57 seconds of CPU time.

Example 5.3 Robust Optimization:

In robust optimization problems the data are uncertain and only known to belong to some uncertainty set which may be taken as infinite index set in semi-infinite programming. The following robust portfolio optimization problem is originally taken from [1] and treated in different variants in [22].

Let 1 € be invested in a portfolio comprised of K shares. At the end of a given period the return of share i is $y_i > 0$. The goal is to determine the amount x_i to be invested in share i, $i = 1, \ldots, K$, so as to maximize the end-of-period portfolio value $y^T x$.

Since y is uncertain, the assumption that y varies in some non-empty compact set $Y \subset \mathbb{R}^K$ leads us to the following semi-infinite programming problem:

$$\max_{x,z} z \text{ s.t. } z - y^T x \leq 0 \quad \text{for all} y \in Y, \quad \sum_{i=1}^{K} x_i = 1, x \geq 0.$$

In fact, this is a linear semi-infinite programming problem.

Problem 1 Let the uncertainty set Y be in the form:

$$Y = \left\{ y \in \mathbb{R}^K \;\middle|\; \sum_{i=1}^{K} \frac{(y_i - \bar{y}_i)^2}{\sigma_i^2} \leq \theta^2 \right\}$$

where $\bar{y}_i$ is some nominal value of y_i, σ_i is scaling parameter and θ measures the risk aversion. With the particular choices from [1]

$$\bar{y}_i = 1.15 + i\frac{0.05}{K},\ i = 1, \ldots, K$$

$$\sigma_i = \frac{0.05}{\theta K}\sqrt{\frac{K(K+1)i}{2}},\ i = 1, \ldots, K$$

$$\theta = 1.5$$

the optimal value is 1.15 for any K (see [1]). The optimal policy in this situation is to invest equally in all shares and $x_i = 1/K$, $i = 1, \ldots, K$. We use the starting point $x^0 = (1, 0, \ldots, 0)$ in $\mathbb{R}^{K+1}$.

The columns of the following tables are labeled as follows: `K` is the number of shares, `ov` is the optimal value, $\|\mathtt{T}(\bar{z})\|$ is the Euclidean norm of `T` at the last iteration point, `CPU time` is the CPU time for iterations in seconds, `iter` is the number of iterations. Note that this optimization problem is convex so that the computed KKT point even is a global maximizer Numerical results are given in Tables 1 and 2.

Problem 2 Let the uncertainty set Y depend on x in which the risk aversion of the decision maker depends on the point x. Replacing θ by $\Theta(x)$, $Y(x)$ is given in the form:

$$Y(x) = \left\{ y \in \mathbb{R}^K \;\middle|\; \sum_{i=1}^{K} \frac{(y_i - \bar{y}_i)^2}{\sigma_i^2} \leq \Theta(x)^2 \right\}$$

with

$$\Theta(x) = \theta\left(1 + \sum_{i=1}^{K}\left(x_i - \frac{1}{N}\right)^2\right)$$

In this case we have an example for *generalized* semi-infinite programming problems. We use the starting point $x^0 = (1, 0, \ldots, 0)$ in $\mathbb{R}^{K+1}$. Since the optimization problem is not convex, we have no guarantee that the computed KKT point is a global maximizer Tables 3 and 4.

It can be checked that in the problems in Example 5.2 and 5.3 strict complementarity holds in the upper and lower level problems, so that we actually have a smooth system. Now, for an illustration of the case that strict complementarity is violated in the upper level, we give the following example.

Table 1 Optimal portfolio, Problem 1 with ψ_{FB}

K	ov	$\|\|\mathtt{T}(\bar{z})\|\|$	CPU time	iter
10	1.15	$5.2928 \cdot 10^{-8}$	0.36	11
50	1.15	$5.7463 \cdot 10^{-7}$	1.43	11
100	1.15	$1.7676 \cdot 10^{-10}$	8.61	11
150	1.15	$4.2121 \cdot 10^{-10}$	25.17	12

Table 2 Optimal portfolio, Problem 1 with $\psi_{\min}$

K	ov	$\|\|T(\bar{z})\|\|$	CPU time	iter
10	1.15	$3.2058 \cdot 10^{-7}$	0.54	10
50	1.15	$2.5966 \cdot 10^{-13}$	1.39	11
100	1.15	$7.5748 \cdot 10^{-10}$	5.58	11
150	1.15	$1.0056 \cdot 10^{-12}$	15.35	12

Table 3 Optimal portfolio, Problem 2 with ψ_{FB}

K	ov	$\|\|T(\bar{z})\|\|$	CPU time	iter
10	0.7033	$4.2379 \cdot 10^{-8}$	0.28	5
50	0.9638	$2.3920 \cdot 10^{-9}$	0.72	7
100	1.0259	$4.0606 \cdot 10^{-7}$	2.87	7
150	1.0535	$8.7426 \cdot 10^{-10}$	8.76	8

Table 4 Optimal portfolio, Problem 2 with $\psi_{\min}$

K	ov	$\|\|T(\bar{z})\|\|$	CPU time	iter
10	0.7033	$1.8731 \cdot 10^{-11}$	0.51	4
50	0.9638	$9.2578 \cdot 10^{-9}$	1.43	6
100	1.0259	$4.2917 \cdot 10^{-10}$	7.06	7
150	1.0535	$9.9549 \cdot 10^{-10}$	34.4	11

Example 5.4 Strict complementarity violated in the upper level:

Let us consider the following problem:

$$\min \ f(x) = (x_1 - 1)^2 + (x_2 - 1)^2 \ \text{ s.t. } g(x, y) \leq 0 \quad \forall\, y \in Y$$

where

$$g(x, y) = (y_1 - x_1) + (y_2 - x_2)$$

and

$$Y = \{y \in \mathbb{R}^2 \mid v_1(y) = y_1^2 - 1 \leq 0,\ v_2(y) = y_2^2 - 1 \leq 0\}.$$

The feasible set is $M = \{x \in \mathbb{R}^2 |\ x_1 + x_2 \geq 2\}$, so that strict complementarity is violated at the solution $\bar{x} = (1, 1)$. In the ψ_{FB} case with the starting point $x^0 = (1, 2)$ the semismooth Newton method obtains the optimal value 0 with $\bar{x} = (1, 1)$ and $\bar{y} = (1, 1)$ for the optimal point. We have $||T(\bar{z})|| = 1.6391 \cdot 10^{-10}$ after 7 iterations within 0.14 s of CPU time. In the ψ_{min} case, the optimal point and the optimal value are obtained in 6 iterations with $||T(\bar{z})|| = 5.3765 \cdot 10^{-10}$ within 0.28 s of CPU time. This shows that the method also works well for this problem where strict complementarity is violated in the upper level.

Finally, although we have not provided a theoretical foundation for this case, we also test the method for problems with violated strict complementarity in the lower level. At the corresponding nondifferentiability points the values of α_j^i and β_j^i are chosen according to the same rule as explained above for λ_i and θ_i.

Example 5.5 Strict complementarity violated in the lower level:

Problem 1 Let us consider the following problem:

$$\min \ f(x) = x_1^2 + x_2^2 \ \text{ s.t. } g(x, y) \leq 0 \quad \forall\, y \in Y$$

where

$$g(x, y) = -(y_1 - x_1)^2 - (y_2 - x_2)^2$$

and

$$Y = \{y \in \mathbb{R}^2 \mid v_1(y) = y_1 \leq 0,\ v_2(y) = y_2 \leq 0\}$$

The unique unconstrained minimum of the objective function $\bar{x} = (0, 0)$ is feasible and therefore optimal for this problem. Its active index set only contains the point $\bar{y} = (0, 0)$, and lower level strict complementarity is violated there. In the ψ_{FB} case with the starting point $x^0 = (1, 1)$ the semismooth Newton method obtains the optimal value 0 with $\bar{x} = (0, 0)$ and $\bar{y} = (0, 0)$ for the optimal point. We have $||T(\bar{z})|| = 9.7946 \cdot 10^{-7}$ after 115 iterations within 1.09 s of CPU time. In the ψ_{min} case, the optimal point and the optimal value are obtained in 2 iterations with $||T(\bar{z})|| = 0$ within 0.25 s of CPU time. This shows that the method also works well for this problem where strict complementarity is violated in the lower level.

Problem 2 Another example for the case of violated strict complementarity in the lower level is given in [12]:

$$\min\ f(x) = x_1^2 + 3x_2^2 + x_3 \text{ s.t. } g(x, y) \leq 0 \quad \forall\, y \in Y$$

where

$$g(x, y) = -\frac{1}{2}(y_1 - x_1)^2 - (y_2 - x_2)^2 - x_3$$

and

$$Y = \{y \in \mathbb{R}^2 \mid v_1(y) = -y_1 - y_2 \leq 0, \quad v_2(y) = -y_2 \leq 0, \quad v_3(y) = y_1^2 + y_2^2 - 1 \leq 0\}.$$

In the ψ_{FB} case with the starting point $x^0 = (1, 1, 1)$ the semismooth Newton method obtains the optimal value 0 with $\bar{x} = (0, 0, 0)$ and $\bar{y} = (0, 0)$ for the optimal point. We have $||T(\bar{z})|| = 5.4371 \cdot 10^{-9}$ after 10 iterations within 0.16 s of CPU time. In the ψ_{min} case, the optimal point and the optimal value are obtained in 16 iterations with $||T(\bar{z})|| = 2.3642 \cdot 10^{-17}$ within 0.44 s of CPU time.

Problem 3 For the case of violated strict complementarity simultaneously in the upper and lower level problems, consider the following variation of Problem 2

$$\min\ f(x) = x_1^2 + x_2^2 + x_3^2 \text{ s.t. } g(x, y) \leq 0 \quad \forall\, y \in Y$$

where

$$g(x, y) = -\frac{1}{2}(y_1 - x_1)^2 - (y_2 - x_2)^2 - x_3$$

and

$$Y = \{y \in \mathbb{R}^2 \mid v_1(y) = -y_1 - y_2 \leq 0,\ v_2(y) = -y_2 \leq 0,\ v_3(y) = y_1^2 + y_2^2 - 1 \leq 0\}.$$

In the ψ_{FB} case with the starting point $x^0 = (1, 1, 1)$ the semismooth Newton method obtains the optimal value 0 with $\bar{x} = (0, 0, 0)$ and $\bar{y} = (0, 0)$ for the optimal point. We have $||T(\bar{z})|| = 2.0792 \cdot 10^{-7}$ after 8 iterations within 0.17 s of CPU time. In the ψ_{min} case, the optimal point and the optimal value are obtained in 5 iterations with $||T(\bar{z})|| = 0$ within 0.29 s of CPU time. The method works also for this problem where strict complementarity is violated simultaneously in the upper and lower level problems.

In all of the examples, the performance of the method does not change significantly if the NCP function ψ_{FB} is replaced by ψ_{min}, in both systems (7) for *SIP* and (12) for *GSIP*.

6 Final remarks

The problems given in Example 5.5 show that Algorithm 5.1 may even converge if lower level strict complementarity is violated. The canonical lower level regularity condition which takes care of this situation is the Robinson condition. However, the auxiliary functions φ_i, $i \in P$, of the locally reduced problem $GSIP_{red}$ are then not twice continuously differentiable anymore, and the relation of an appropriate upper level regularity condition to CD-regularity becomes essentially more complicated. This situation is currently under investigation and will be subject of a future publication.

Acknowledgements We would like to thank Georg Still for many helpful comments on a previous version of this manuscript, and the anonymous referees for their precise and substantial remarks which lead to a significantly improved version of this article. The second author gratefully acknowledges support through TÜBITAK. (Turkish Scientific and Technical Research Council).

Appendix A

A.1 NCP functions

Both functions $\psi_{FB}(a,b) = \sqrt{a^2+b^2} - a - b$ and $\psi_{min}(a,b) = -\min\{a,b\}$ are convex on $\mathbb{R}^2$, and they are differentiable on $\mathbb{R}^2$ except for the origin and the diagonal $\{(a,b) \in \mathbb{R}^2 \mid a = b\}$, respectively. For $(a,b) \neq 0$ we have

$$D\psi_{FB}(a,b) = \left(\frac{a}{\sqrt{a^2+b^2}} - 1,\ \frac{b}{\sqrt{a^2+b^2}} - 1 \right) \tag{23}$$

and for $a \neq b$

$$D\psi_{min}(a,b) = \begin{cases} (-1,0), & a < b \\ (0,-1), & a > b \end{cases}. \tag{24}$$

In their (identical) zero set, the only point of nondifferentiability for either of the two functions is the origin, and their gradients for $(a,b) \neq 0$ coincide:

$$D\psi_{FB}(a,b) = D\psi_{min}(a,b) = \begin{cases} (-1,0), & a = 0 \\ (0,-1), & b = 0 \end{cases}. \tag{25}$$

At the origin we calculate subdifferentials of the convex functions ψ_{FB} and ψ_{min} as follows. We denote the usual *directional derivative* of ψ in the direction d at $\bar{x}$ by $\psi'(\bar{x}; d)$ and the *generalized* directional derivative of ψ (in the sense of Clarke) in the direction d at $\bar{x}$ by $\psi^0(\bar{x}; d)$. For completeness, we give the proofs of the following well-known results.

Lemma A.1 *The following assertions hold:*

(i) $\psi'_{FB}(0;d) = \psi^0_{FB}(0;d) = \psi_{FB}(d)$ *for any* $d \in \mathbb{R}^2$,
(ii) $\psi'_{min}(0;d) = \psi^0_{min}(0;d) = \psi_{min}(d)$ *for any* $d \in \mathbb{R}^2$,
(iii) $\partial\psi_{FB}(0) = \{s \in \mathbb{R}^2 \mid (s_1+1)^2 + (s_2+1)^2 \leq 1\}$,
(iv) $\partial\psi_{min}(0) = \mathrm{conv}\{(-1,0)^T,\ (0,-1)^T\}$.

Proof Parts (i) and(ii) can easily be checked by calculation. In part (iii), by definition of the Clarke subdifferential for real-valued functions, we have

$$\begin{aligned}\partial\psi_{FB}(0) &= \{s \in \mathbb{R}^2 |\ \psi^0(0; d) \geq d^T s \quad \text{for all } d \in \mathbb{R}^2\} \\ &= \{\tilde{s} \in \mathbb{R}^2 |\ d^T\tilde{s} \leq ||d||_2, \quad \text{for all } d \in \mathbb{R}^2\} - (1, 1)^T \\ &= \{\tilde{s} \in \mathbb{R}^2 |\ \max_{d\in\partial B(0,1)} d^T\tilde{s} \leq 1\} - (1, 1)^T\end{aligned}$$

where $\tilde{s} = s+(1, 1)^T$ and where $\partial B(0, 1)$ denotes the boundary of the unit ball. The Cauchy-Schwarz inequality $d^T\tilde{s} \leq ||d||_2||\tilde{s}||_2$ implies that $\max_{d\in\partial B(0,1)} d^T\tilde{s} \leq ||\tilde{s}||_2$ for all $\tilde{s} \in \mathbb{R}^2$. On the other hand, with $\bar{d} = \tilde{s}/||\tilde{s}||_2 \in \partial B(0, 1)$, we find $\max_{d\in\partial B(0,1)} d^T\tilde{s} \geq \bar{d}^T\tilde{s} = ||\tilde{s}||_2$ for all $\tilde{s} \in \mathbb{R}^2 \setminus \{0\}$, the case $\tilde{s} = 0$ being trivial. This shows

$$\partial\psi_{FB}(0) = \{\tilde{s} \in \mathbb{R}^2 |\ ||\tilde{s}||_2 \leq 1\} - (1, 1)^T$$

and completes the proof of part (iii).

To see part (iv), note that by definition of the Clarke subdifferential for vector-valued functions we have

$$\partial\psi_{min}(0) \ = \ \text{conv}\{\lim_{x\to 0} \nabla\psi_{min}(x) \mid x \in D_\psi\} \ = \ \text{conv}\{(-1, 0)^T,\ (0, -1)^T\}$$

where D_ψ is set of differentiability points of ψ_{min}. □

A.2 Block matrices

Definition A.2 Consider the quadratic block matrix

$$A = \begin{pmatrix} E & F \\ G & H \end{pmatrix}$$

where H is quadratic and nonsingular. Then the matrix

$$S = A/H = E - FH^{-1}G$$

is called the *Schur complement* of H in A.

Lemma A.3 ([15]) *Let a block matrix A be given as in Definition A.2 and let H be nonsingular. Then*

$$det(A) = det(H) \cdot det(A/H).$$

In particular, A is nonsingular if and only if A/H is nonsingular.

Lemma A.4 ([11]) *For $A \in S^N$ and $B \in \mathbb{R}_K^{N\times M}$ we have*

$$In\begin{pmatrix} A & B \\ B^T & 0_{M\times M} \end{pmatrix} = In(A|_{Ker(B^T)}) + (K, K, M - K). \tag{26}$$

Here, In(A) denotes the *inertia-triple* of A, S^N denotes the set of symmetric matrices in $\mathbb{R}^{N\times N}$ and $\mathbb{R}_K^{N\times M}$ denotes the set of matrices in $\mathbb{R}^{N\times M}$ with rank K. The null space of a matrix A is denoted by Ker(A).

By Lemma A.4, we have that $A|_{\text{Ker}(B^T)}$ is nonsingular and the columns of B are linearly independent if and only if the matrix

$$\begin{pmatrix} A & B \\ B^T & 0 \end{pmatrix}$$

is nonsingular.

References

1. Ben-Tal, A., Nemirovski, A.: Robust solutions of uncertain linear programs. Opera. Res. Lett. **25**, 1–13 (1999)
2. Clarke, F.H.: Optimization and Nonsmooth Analysis. Wiley, New York (1983)
3. Floudas, C. A., Stein, O.: The adaptive convexification algorithm: a feasible point method for semi-infinite programming. SIAM J Optim (to appear)
4. Goberna, M.A., López, M.A.: Linear Semi-infinite Optimization. Wiley, Chichester (1998)
5. Günzel, H., Jongen, H. Th., Stein, O.: Generalized semi-infinite programming: on generic local minimizers, AOR-Preprint 2/2007, Chair for Applications of OR, University of Karlsruhe (2007)
6. Guerra Vazquez, F., Rückmann, J.-J., Stein, O., Still, G.: Generalized semi-infinite programming: a tutorial. J. Comput. Appl. Math. (to appear)
7. Hettich, R., Kortanek, K.: Semi-infinite programming: theory, methods and applications. SIAM Rev. **35**, 380–429 (1993)
8. Hettich, R., Still, G.: Second order optimality conditions for generalized semi-infinite programming problems. Optimization. **34**, 195–211 (1995)
9. John, F.: Extremum problems with inequalities as subsidiary conditions, In: studies and essays, R. Courant Anniversary Volume, Interscience, New York 187–204 (1948)
10. Jongen, H.Th., Jonker, P., Twilt, F.: Critical sets in parametric optimization. Math. Programming **34**, 333–353 (1986)
11. Jongen, H.Th., Möbert, T., Rückmann, J.-J., Tammer, T.: On inertia and Schur complement in optimization. Lin. Alg. Appl. **95**, 97–109 (1987)
12. Jongen, H.Th., Wetterling, W., Zwier, G.: On sufficient conditions for local optimality in semi-infinite programming. Optimization **18**, 165–178 (1987)
13. Klatte, D., Kummer, B.: Nonsmooth Equations in Optimization. Kluwer, Dordrecht (2002)
14. Mifflin, R.: Semismooth and semiconvex functions in constrained optimization. SIAM J. Control Optim. **15**, 959–972 (1997)
15. Ouellette, D.V.: Schur complements and statistics. Lin. Alg. Appl. **36**, 187–295 (1981)
16. Qi, L., Jiang, H.: Semismooth KKT equations and convergence analysis of Newton and quasi-Newton methods for solving these equations. Math. Op. Res. **22**, 301–325 (1997)
17. Qi, L., Sun, J.: A nonsmooth version of Newton's Method. Math. Programming **58**, 353–367 (1993)
18. Qi, L., Wu, S.-Y., Zhou, G.: Semismooth Newton Methods for Solving Semi-infinite Problems. J. Global Optim. **27**, 215–232 (2003)
19. Reemtsen, R., Görner, S.: Numerical methods for semi-infinite programming: a survey. In: Reemtsen, R., Rückmann, J.-J. (eds) Semi-Infinite Programming, Kluwer, Boston, pp. 195–275 (1998)
20. Robinson, S.M.: Generalized Equations and their solutions, Part II, Applications to nonlinear programming. Math. Programming Stud. **19**, 200–221 (1982)
21. Stein, O.: On parametric semi-infinite optimization. PhD Thesis, University of Trier, Department of Mathematics (1997)
22. Stein, O.: Bi-level strategies in semi-infinite programming. Kluwer (2003)