Comput Optim Appl (2013) 54:189–210
DOI 10.1007/s10589-012-9489-4

A lifting method for generalized semi-infinite programs based on lower level Wolfe duality

M. Diehl · B. Houska · O. Stein · P. Steuermann

Received: 16 December 2011 / Published online: 12 June 2012
© Springer Science+Business Media, LLC 2012

Abstract This paper introduces novel numerical solution strategies for generalized semi-infinite optimization problems (GSIP), a class of mathematical optimization problems which occur naturally in the context of design centering problems, robust optimization problems, and many fields of engineering science. GSIPs can be regarded as bilevel optimization problems, where a parametric lower-level maximization problem has to be solved in order to check feasibility of the upper level minimization problem. The current paper discusses several strategies to reformulate this class of problems into equivalent finite minimization problems by exploiting the concept of Wolfe duality for convex lower level problems. Here, the main contribution is the discussion of the non-degeneracy of the corresponding formulations under various assumptions. Finally, these non-degenerate reformulations of the original GSIP allow us to apply standard nonlinear optimization algorithms.

Keywords Semi-infinite optimization · Lower level duality · Lifting approach · Adaptive convexification · Mathematical program with complementarity constraints

M. Diehl · B. Houska
Electrical Engineering Department (ESAT) and Optimization in Engineering Center (OPTEC), K.U. Leuven, Leuven, Belgium

M. Diehl
e-mail: moritz.diehl@esat.kuleuven.be

B. Houska
e-mail: boris.houska@esat.kuleuven.be

O. Stein (✉) · P. Steuermann
Institute of Operations Research, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
e-mail: stein@kit.edu

P. Steuermann
e-mail: steuermann@kit.edu

1 Introduction

This article introduces a novel lifting approach for *generalized semi-infinite optimization problems*. These problems have the form

$$\min_x f(x) \quad \text{s.t.} \quad x \in M \tag{GSIP}$$

with

$$M = \left\{ x \in X \mid g(x, y) \leq 0 \text{ for all } y \in Y(x) \right\}$$

and

$$Y(x) = \left\{ y \in \mathbb{R}^m \;\middle|\; v_\ell(x, y) \leq 0,\ 1 \leq \ell \leq s \right\}.$$

We assume the set $X \subset \mathbb{R}^n$ to be closed, and all defining functions f, g, v_ℓ, $1 \leq \ell \leq s$, to be real-valued and at least continuously differentiable on their respective domains.

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.

1.1 Existing approaches

For theory, applications and many references on *standard* semi-infinite optimization we refer to the reviews [11, 18]. There exist a wide range of numerical solution methods for linear and nonlinear standard SIPs. Overviews can be found in [12, 19, 21].

However, *generalized* semi-infinite optimization problems turn out to be much more expensive to solve. Several applications of generalized semi-infinite optimization problems are described in [24], for example robust optimization, design centering, minimax problems, Chebyshev approximation, and disjunctive programming. More recent applications include the identification of regression models and of dynamics of networks in the presence of uncertainty [14, 31]. A survey on standard and generalized semi-infinite optimization is given in [25].

Some authors have discussed discretization strategies for the varying index set of GSIPs, in order to replace the constraints by a finite approximation [11, 29, 30], but such discretization strategies are rather conceptual. Implementable, but very expensive methods using bi-level strategies and a branch-and-bound framework combined with interval arithmetic may be found in [3, 4, 7, 16, 17].

The situation is very different, however, if additional convexity assumptions are available such that duality ideas can be used to generate tractable reformulations [1, 2, 24, 28]. In the context of semi-infinite programming, [15] presents such an approach using duality for convex quadratic problems. The article [20] shows how, even without convexity assumptions, augmented Lagrangians can be used to reduce generalized to standard min-max problems.

1.2 Contribution and overview

In this paper we discuss four lifting reformulation strategies for generalized semi-infinite programming problems under the assumption of lower level convexity on the

feasible domain. Section 2 introduces our framework of lifting reformulation strategies for generalized semi-infinite programming problems under the assumption of lower level convexity. After a short review of some well-known Wolfe duality results, which help us to reformulate GSIP problems into standard finite minimization problems, a preliminary non-degeneracy result is shown. In Sect. 3 a second existing reformulation strategy is reviewed: here, the GSIP is reformulated into an MPCC by regarding the optimality conditions of the lower level maximization problem. Both approaches are compared by analyzing in which cases degeneracies can occur.

In Sect. 4 a novel lifting based reformulation strategy is suggested which is then shown to be non-degenerate under a lifted version of the linear independence constraint qualification. In Sect. 5 this analysis is extended for a set of weaker assumptions leading to a lifted version of the Mangasarian Fromovitz constraint qualification. Finally, the fourth formulation is considered in Sect. 6. Here, we generalize the lifting approach to the case that all constraint qualifications for the lower level problem might be violated. The paper concludes with an illustration and numerical results in Sect. 7 and some final remarks in Sect. 8.

2 Lifting in the nondegenerate case

The n-parametric *lower level problem* of GSIP is given by

$$Q(x): \quad \max_y g(x,y) \quad \text{s.t.} \quad y \in Y(x).$$

The problem $Q(x)$ is called convex if $Y(x)$ is a convex set and if $g(x,\cdot)$ is concave on $Y(x)$. As it will turn out, it is crucial that the blanket assumption of the current section is slightly stronger:

Assumption 2.1 For each $x \in X$ all functions $-g(x,\cdot),\ v_\ell(x,\cdot),\ 1 \le \ell \le s$, are convex on $\mathbb{R}^m$.

The optimal value function of the parametric problem $Q(x)$ is

$$\varphi(x) = \sup_{y \in Y(x)} g(x,y),$$

where we put $\varphi(x) = -\infty$ in the case $Y(x) = \emptyset$. Note that $Y(x) = \emptyset$ corresponds to the 'absence of restrictions', so that x is feasible in this case. It is easily seen that φ admits the description

$$M = \left\{ x \in X \;\middle|\; \varphi(x) \le 0 \right\} \tag{1}$$

of the feasible set.

Example 2.2 ([24]) For $x \in \mathbb{R}^2$ and $y \in \mathbb{R}$ choose $g(x,y) = -y$, $v_1(x,y) = x_1 - y$, and $v_2(x,y) = y - x_2$. Then $Q(x)$ is convex for each $x \in \mathbb{R}^2$. However, for $x_1 > x_2$ the feasible set

$$Y(x) = \{ y \in \mathbb{R} \mid x_1 \le y \le x_2 \}$$

of $Q(x)$ is empty, and we obtain

$$\varphi(x) = \begin{cases} -x_1, & x_1 \le x_2, \\ -\infty, & x_1 > x_2. \end{cases}$$

The resulting set

$$M = \left\{x \in \mathbb{R}^2 \,\middle|\, \varphi(x) \le 0\right\} = \left\{x \in \mathbb{R}^2 \,\middle|\, x_1 \ge 0,\ x_1 \le x_2\right\} \cup \left\{x \in \mathbb{R}^2 \,\middle|\, x_1 > x_2\right\}$$

is the union of an open with a closed set.

In the following we put $v(x, y) = (v_1(x, y), \ldots, v_s(x, y))^\mathsf{T}$. With the Lagrange function of $Q(x)$,

$$\mathcal{L}(x, y, \gamma) = g(x, y) - \gamma^\mathsf{T} v(x, y),$$

we can state the Wolfe dual problem of $Q(x)$ as follows:

$$D(x): \quad \min_{y,\gamma} \mathcal{L}(x, y, \gamma) \quad \text{s.t.} \quad \nabla_y \mathcal{L}(x, y, \gamma) = 0, \quad \gamma \ge 0.$$

Let $Y_D(x)$ denote the feasible set of $D(x)$, and let

$$\psi(x) = \inf_{(y,\gamma) \in Y_D(x)} \mathcal{L}(x, y, \gamma)$$

be the optimal value function of $D(x)$.

As it is well-known, the existence of a Karush-Kuhn-Tucker point of $Q(x)$ implies solvability of both $Q(x)$ and $D(x)$ as well as strong duality. In view of (1) we thus obtain

$$M = \left\{x \in X \,\middle|\, \min_{(y,\gamma) \in Y_D(x)} \mathcal{L}(x, y, \gamma) \le 0\right\}$$

if $Q(x)$ possesses a Karush-Kuhn-Tucker point for each $x \in X$.

Under the latter assumption we obviously have feasibility of $x \in X$ if and only if $\mathcal{L}(x, y, \gamma) \le 0$ holds for some $(y, \gamma) \in Y_D(x)$. Letting pr_x denote the orthogonal projection to the 'x-space' $\mathbb{R}^n$, we obtain $M = \mathrm{pr}_x M_P$ with

$$M_P = \left\{(x, y, \gamma) \in X \times \mathbb{R}^m \times \mathbb{R}^s \,\middle|\, \mathcal{L}(x, y, \gamma) \le 0,\ \nabla_y \mathcal{L}(x, y, \gamma) = 0,\ \gamma \ge 0\right\}.$$

This motivates us to introduce the lifted Wolfe problem

$$\boxed{\min_{x,y,\gamma} f(x) \quad \text{s.t.} \quad x \in X, \quad \mathcal{L}(x, y, \gamma) \le 0, \quad \nabla_y \mathcal{L}(x, y, \gamma) = 0, \quad \gamma \ge 0} \quad \text{(LWP)}$$

As f does not depend on the variables y and γ, we have shown the following result, which is known in similar form from different contexts (e.g. [1, 15, 20]).

Theorem 2.3 *Let $Q(x)$ possess a Karush-Kuhn-Tucker point for each $x \in X$. Then the minimizers of GSIP coincide with the x-components of the minimizers of LWP.*

Standard arguments then lead to the following corollary.

Corollary 2.4 *Let $Y(x)$ be compact and possess a Slater point for each $x \in X$. Then the minimizers of GSIP coincide with the x-components of the minimizers of LWP.*

For standard semi-infinite optimization problems, the situation becomes even simpler.

Corollary 2.5 *In a standard semi-infinite program SIP, let Y be compact and possess a Slater point. Then the minimizers of SIP coincide with the x-components of the minimizers of LWP.*

Example 2.6 Consider Example 2.2 with $X_\varepsilon = \{x \in \mathbb{R}^2 \mid x_1 + \varepsilon \le x_2\}$ and $\varepsilon \ge 0$. Then for all $\varepsilon > 0$ the assumptions of Corollary 2.4 are satisfied. Furthermore, for $x_1 = x_2$ the problem $Q(x)$ possesses Karush-Kuhn-Tucker points although $Y(x)$ violates the Slater condition. Hence, the assumptions of Theorem 2.3 even hold for $\varepsilon = 0$ and, for any objective function f, the problem

$$\text{GSIP:} \quad \min_x f(x) \quad \text{s.t.} \quad x_1 \le x_2, \quad y \ge 0 \quad \text{for all } y \text{ with } x_1 \le y \le x_2$$

is equivalent to the lifted problem

$$\text{LWP:} \quad \min_{(x,y,\gamma)} f(x) \quad \text{s.t.} \quad x_1 \le x_2, \quad -y - \gamma_1(x_1 - y) - \gamma_2(y - x_2) \le 0,$$
$$-1 + \gamma_1 - \gamma_2 = 0, \quad \gamma_1, \gamma_2 \ge 0.$$

Theorem 2.3 and its subsequent corollaries show that GSIPs with sufficiently regular convex lower level problems are equivalent to *smooth* optimization problems in some higher dimensional space. This raises the question whether the smooth problem P suffers from some inherent degeneracy which could make its numerical solution intractable. The following result shows that this is not the case.

As $\mathcal{L}(x, y, \gamma)$ is concave in y for any $\gamma \ge 0$, its Hessian $D_y^2\mathcal{L}(x, y, \gamma)$ is negative semidefinite everywhere. If either $D_y^2 g(x, y)$ is negative definite or $D_y^2 v_\ell(x, y)$ is negative definite for some $1 \le \ell \le s$ with $\gamma_\ell > 0$, then $D_y^2\mathcal{L}(x, y, \gamma)$ becomes even negative definite. In this sense, the nonsingularity assumption for $D_y^2\mathcal{L}(x, y, \gamma)$ in the following lemma is mild. Here and in the following, $\nabla_y\mathcal{L}$ will denote the vector of first partial derivatives of $\mathcal{L}$ with respect to the vector y in column form, whereas $D_y\mathcal{L}$ abbreviates the row vector $\nabla_y^T\mathcal{L}$.

Lemma 2.7 *Let $(\bar{x}, \bar{y}, \bar{\gamma}) \in M_P$ be given such that $D_y^2\mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma})$ is nonsingular and such that $\mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma}) = 0$ and $\nabla_x\mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma}) = 0$ do not hold simultaneously. Then the gradients of active constraints among*

$$\mathcal{L}(x, y, \gamma) \le 0, \quad \nabla_y\mathcal{L}(x, y, \gamma) = 0, \quad \gamma \ge 0$$

are linearly independent at $(\bar{x}, \bar{y}, \bar{\gamma})$.

Proof At $(\bar{x}, \bar{y}, \bar{\gamma}) \in M_P$ we have

$$D_{(x,y,\gamma)} \begin{pmatrix} \mathcal{L} \\ \nabla_y \mathcal{L} \\ \gamma \end{pmatrix} = \begin{pmatrix} D_x \mathcal{L} & 0 & -v^\intercal \\ D^2_{yx} \mathcal{L} & D^2_y \mathcal{L} & -\nabla_y v \\ 0 & 0 & -I \end{pmatrix}$$

where the arguments are omitted, I denotes the identity matrix of appropriate dimension, and where we used that $D_y \mathcal{L}$ vanishes at feasible points. Due to its block structure and the nonsingularity assumption on $D^2_y \mathcal{L}$, this matrix has full row rank if and only if $D_x \mathcal{L}$ does not vanish. By assumption, it does not vanish if the constraint $\mathcal{L} \leq 0$ is active. And if this constraint is inactive, the first row is not part of the active constraint gradients. The assertion immediately follows. □

3 A comparison to the MPCC lifting approach

A different lifting approach for GSIPs with convex lower level problems was introduced in [24, 27] and further developed in [26, 28]. It uses the Karush-Kuhn-Tucker conditions of $Q(x)$ more directly. This results in the lifted problem

$$\min_{x,y,\gamma} f(x) \quad \text{s.t.} \quad \begin{cases} x \in X, g(x, y) \leq 0, \\ \nabla_y \mathcal{L}(x, y, \gamma) = 0, \\ 0 \leq -v(x, y) \perp \gamma \geq 0. \end{cases} \tag{MPCC}$$

This so-called *mathematical program with complementarity constraints* is degenerate in the sense that the Mangasarian Fromovitz constraint qualification (MFCQ) is violated everywhere in its feasible set [22]. However, the degeneracy of MPCC can be treated with tailored numerical algorithms. In fact, in [24, 26–28] it was solved numerically by employing smoothed NCP functions, where NCP functions admit a nondegenerate, albeit nonsmooth reformulation of the problem.

An advantage of the MPCC lifting approach is that the lower level constraint $v(x, y) \leq 0$ is modeled explicitly. In the approach from Sect. 2, this relation holds due to the duality effect which works under Assumption 2.1. It may fail, however, if $g(x, \cdot)$ is only concave on $Y(x)$, but not on all $\mathbb{R}^m$. The following example illustrates this drawback of the duality approach.

Example 3.1 For $x \in X = \mathbb{R}$ and $y \in \mathbb{R}$ choose $g(x, y) = -x - y^3 - 3y^2 + 16y - 12$, $v_1(y) = -y$, and $v_2(y) = y - 1$. This leads to the standard semi-infinite constraint

$$-x - y^3 - 3y^2 + 16y - 12 \leq 0 \quad \text{for all } y \in [0, 1].$$

As, for any $x \in \mathbb{R}$, $D^2_y g(x, y) = -6(y + 1)$ is nonpositive for all $y \in [0, 1]$, the lower level problem $Q(x)$ is convex in the sense that $g(x, \cdot)$ is concave on the convex set $Y = [0, 1]$. However, Assumption 2.1 does *not* hold, as $g(x, \cdot)$ is not concave on $\mathbb{R}$.

One easily computes $\varphi(x) = -x$ so that the semi-infinite constraint defines the set $M = \{x \in \mathbb{R} | \, x \geq 0\}$. Using the lower level Lagrange function

$$\mathcal{L}(x, y, \gamma) = -x - y^3 - 3y^2 + 16y - 12 + \gamma_1 y - \gamma_2 (y - 1),$$

the lifted feasible set M_P is defined by the constraints

$$-x + 2y^3 + 3y^2 - 12 + \gamma_2 \le 0, \qquad -3y^2 - 6y + 16 + \gamma_1 - \gamma_2 = 0, \quad \gamma_1, \gamma_2 \ge 0.$$

For any $x \in \mathbb{R}$ we may choose a solution y of

$$x = 2y^3 + 3y^2 - 12 + \max\{0, -3y^2 - 6y + 16\}$$

since it is not hard to see that the right hand side of this equation maps $\mathbb{R}$ onto $\mathbb{R}$. Putting

$$\gamma_1 = \max\{0, 3y^2 + 6y - 16\} \quad \text{and} \quad \gamma_2 = \max\{0, -3y^2 - 6y + 16\}$$

then implies $\gamma_1, \gamma_2 \ge 0$, $-3y^2 - 6y + 16 + \gamma_1 - \gamma_2 = 0$ as well as $x = 2y^3 + 3y^2 - 12 + \gamma_2$, so that $(x, y, \gamma) \in M_P$ and, thus, $x \in \mathrm{pr}_x M_P$ hold. This shows $\mathrm{pr}_x M_P = \mathbb{R} \ne M$ and, in particular, that under violation of Assumption 2.1 the assertion of, e.g., Corollary 2.5 may be wrong.

A situation like in Example 3.1 routinely occurs in the adaptive convexification algorithm from [7, 26] for the solution of standard semi-infinite optimization problems with nonconvex lower level problems. There, convex relaxations are constructed for nonconvex functions on boxes. Since outside these boxes the relaxation functions are not necessarily convex, Assumption 2.1 may be violated, and treating the relaxed convex problems by the duality approach from Sect. 2 may fail. Numerical experiments indicate, in fact, that the latter duality approach does not work in the framework of adaptive convexification, as it produces solutions which violate the constraint $v(y) \le 0$ (cf. also Sect. 7). The MPCC lifting approach, on the other hand, works well.

Still, the following strong connection between the problems LWP and MPCC even holds without Assumption 2.1.

Proposition 3.2 *For $X = \mathbb{R}^n$ let $(\bar{x}, \bar{y}, \bar{\gamma})$ be a Karush-Kuhn-Tucker point of LWP with active constraint $\mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma}) \le 0$, let the multiplier corresponding to the latter constraint be positive, and let $D_y^2 \mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma})$ be nonsingular. Then $(\bar{x}, \bar{y}, \bar{\gamma})$ is feasible for MPCC.*

Proof Under our assumptions, there are multipliers $\bar{\chi}$, $\bar{\rho}$ and $\bar{\kappa}$ with

$$\nabla f(\bar{x}) + \bar{\chi} \nabla_x \mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma}) + D_{yx}^2 \mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma}) \bar{\rho} = 0, \tag{2}$$

$$D_y^2 \mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma}) \bar{\rho} = 0, \tag{3}$$

$$-\bar{\chi} v(\bar{x}, \bar{y}) - D_y v(\bar{x}, \bar{y}) \bar{\rho} - \bar{\kappa} = 0, \tag{4}$$

$$\nabla_y \mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma}) = 0, \tag{5}$$

$$\mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma}) = 0, \tag{6}$$

$$\bar{\chi} > 0, \tag{7}$$

$$\bar{\gamma} \ge 0, \tag{8}$$

$$\bar{\kappa} \geq 0, \tag{9}$$

$$\bar{\gamma}^\mathsf{T} \bar{\kappa} = 0. \tag{10}$$

The nonsingularity of $D_y^2 \mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma})$ and (3) yield $\bar{\rho} = 0$. In view of (7) and (9), equation (4) implies $v(\bar{x}, \bar{y}) \leq 0$. Moreover, taking the inner product of (4) with $\bar{\gamma}$ and using (10) leads to $\bar{\gamma}^\mathsf{T} v(\bar{x}, \bar{y}) = 0$. Together with (5) and (8) this yields the assertion. □

We also note that, under mild additional assumptions (cf., e.g., [27]) which do not include Assumption 2.1, each local minimizer of GSIP on the boundary of the feasible set satisfies a KKT condition of the form $\nabla f(\bar{x}) + \bar{\chi} \nabla_x \mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma}) = 0$ with $\bar{\chi} \geq 0$, with some $\bar{y} \in \{y \in Y(\bar{x}) |\ g(\bar{x}, \bar{y}) = 0\}$ and with a multiplier vector $\bar{\gamma}$ satisfying $\nabla_y \mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma}) = 0$ and $0 \leq -v(\bar{x}, \bar{y}) \perp \bar{\gamma} \geq 0$. In particular, the point $(\bar{x}, \bar{y}, \bar{\gamma})$ is feasible for MPCC, and with the choices $\bar{\rho} = 0$ and $\bar{\kappa} = -\bar{\chi} v(\bar{x}, \bar{y})$ it is also a KKT point of LWP. In Example 3.1 this leads to a degeneracy effect: for the objective function $f(x) = x$ the point $\bar{x} = 0$ is a solution point of GSIP, and $(\bar{x}, \bar{y}, \bar{\gamma}_1, \bar{\gamma}_2) = (0, 1, 0, 7)$ is the corresponding KKT point of LWP, although $\bar{x} = 0$ lies in the interior of $\mathrm{pr}_x M_P = \mathbb{R}$.

4 Lifting including the index set constraint

The previous section has shown that solving the problem LWP might not lead to correct solutions, if Assumption 2.1 is violated, but the lower level problems are only convex on their respective domains. On the other hand, the MPCC lifting approach treats this case correctly, at the expense of entering the numerically demanding class of mathematical programs with complementarity constraints.

Note that the MPCC formulation from Sect. 3 incorporates two types of additional constraints: first, it includes the lower level constraint $v(x, y) \leq 0$ explicitly. This entails that the function $g(x, \cdot)$ needs only be concave on $Y(x)$, but not on all $\mathbb{R}^m$. Second, it imposes the orthogonality condition $\gamma^\mathsf{T} v(x, y) = 0$ such that the constraints $g(x, y) \leq 0$ and $\mathcal{L}(x, y, \gamma) \leq 0$ are equivalent.

At this point the question arises whether the complementarity constraint of the form $v(x, y)^\mathsf{T} \gamma = 0$ needs to be stated in the MPCC formulation, or if it may simply be skipped. In order to give an answer to this question, we consider an optimization problem of the form

$$\min_{x,y,\gamma} f(x) \quad \text{s.t.} \quad \begin{cases} x \in X, \\ g(x, y) - \gamma^\mathsf{T} v(x, y) = \mathcal{L}(x, y, \gamma) \leq 0, \\ \nabla_y \mathcal{L}(x, y, \gamma) = 0, \\ v(x, y) \leq 0, \quad \gamma \geq 0. \end{cases} \tag{LWPC}$$

The feasible set of LWPC is 'intermediate' between the feasible sets of LWP and of MPCC in the sense that, in comparison to LWP, the additional constraint $v(x, y) \leq 0$ appears and, in comparison to MPCC, the orthogonality constraint $\gamma^\mathsf{T} v(x, y) = 0$

is missing. The following result shows that in Theorem 2.3 the problem LWP may be replaced by LWPC.

Lemma 4.1 *Let $Q(x)$ possess a Karush-Kuhn-Tucker point and let the function $g(x,\cdot)$ be concave on $Y(x)$ for each $x \in X$. Then the minimizers of GSIP coincide with the x-components of the minimizers of LWPC.*

Proof The argumentation is analogous to Theorem 2.3 with the only difference that strong duality now holds for the lower level problems under the slightly weaker requirement that $g(x,\cdot)$ is concave on $Y(x)$ but not necessarily on the whole domain $\mathbb{R}^m$. This is due to the fact that the inequality $v(x,y) \leq 0$ is explicitly required as a constraint in problem LWPC. □

Remark 4.2 Lemma 4.1 even holds under weaker requirements: as long as we have strong duality in the lower level problem the above argumentation remains correct. In this case, the function $g(x,\cdot)$ does not have to be concave on $Y(x)$.

Comparing the problems LWP and LWPC, the only difference is the explicit appearance of the constraint $v(x,y) \leq 0$. Thus, once more the question arises whether this additional constraint leads to an inherent degeneracy of the formulation. In the following, it is shown that this is not the case.

For simplicity of presentation, we may assume $X = \mathbb{R}^n$. The Lagrangian for the minimization problem LWPC is then given by

$$\mathcal{K} := f(x) + \chi \mathcal{L}(x,y,\gamma) + \rho^\mathsf{T} \nabla_y \mathcal{L}(x,y,\gamma) + \mu^\mathsf{T} v(x,y) - \kappa^\mathsf{T} \gamma. \tag{11}$$

Here, the variables $\chi \in \mathbb{R}_+$, $\rho \in \mathbb{R}^m$, $\mu \in \mathbb{R}^s_+$, and $\kappa \in \mathbb{R}^s_+$ are multipliers which are associated with the constraints in the minimization problem LWPC.

Definition 4.3 For a point $(\bar{x},\bar{y})$ let $T = \{p \in \mathbb{R}^m \mid D_y v^{\mathrm{act}}(\bar{x},\bar{y})p = 0\}$, where v^{act} denotes the components of the function v which are active at $(\bar{x},\bar{y})$. We say that a feasible point $(\bar{x},\bar{y},\bar{\gamma})$ of LWPC satisfies the lifted LICQ condition (LLICQ) if the following conditions hold:

1. The matrix $D_y v^{\mathrm{act}}(\bar{x},\bar{y})$ has full rank (lower level LICQ).
2. The condition $v(\bar{x},\bar{y}) - \bar{\gamma} < 0$ holds (lower level strict complementarity).
3. The inequality $p^\mathsf{T} D_y^2 \mathcal{L}(\bar{x},\bar{y},\bar{\gamma})p < 0$ holds for all $p \in T \setminus \{0\}$ (lower level second order condition).
4. $\mathcal{L}(\bar{x},\bar{y},\bar{\gamma}) = 0$ and $D_x \mathcal{L}(\bar{x},\bar{y},\bar{\gamma}) = 0$ do not hold simultaneously (upper level LICQ).

Theorem 4.4 *Let $(\bar{x},\bar{y},\bar{\gamma})$ be a minimizer of problem LWPC at which the LLICQ condition from Definition 4.3 is satisfied. Then the gradients of active constraints among*

$$\mathcal{L}(x,y,\gamma) \leq 0, \quad \nabla_y \mathcal{L}(x,y,\gamma) = 0, \qquad v(x,y) \leq 0, \quad \gamma \geq 0$$

are linearly independent at $(\bar{x},\bar{y},\bar{\gamma})$. In particular, $(\bar{x},\bar{y},\bar{\gamma})$ is a KKT point of LWPC.

Proof As $(\bar{x}, \bar{y}, \bar{\gamma})$ is a minimizer, it satisfies the Fritz John conditions of problem LWPC. The first step of the proof is to regard these conditions with respect to the lower level variables y and γ:

$$0 = \rho^\intercal D_y^2 \mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma}) + \mu^\intercal D_y v(\bar{x}, \bar{y}) \quad \text{and} \tag{12}$$

$$0 = \chi v(\bar{x}, \bar{y}) + D_y v(\bar{x}, \bar{y})\rho + \kappa. \tag{13}$$

Here, the feasibility condition $D_y \mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma}) = 0$ has been used to simplify the stationarity condition with respect to y. The main idea is to multiply equation (12) by ρ, multiply equation (13) by μ, and subtract the second from the first resulting equation. This leads to

$$0 = \rho^\intercal D_y^2 \mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma})\rho - \chi \mu^\intercal v(\bar{x}, \bar{y}) - \mu^\intercal \kappa.$$

As we may use the complementarity between μ and $v(\bar{x}, \bar{y})$, this equation further simplifies to

$$\rho^\intercal D_y^2 \mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma})\rho = \mu^\intercal \kappa \geq 0. \tag{14}$$

Recall that by definition we have $v^{\text{act}}(\bar{x}, \bar{y}) = 0$ which implies $\gamma^{\text{act}} > 0$ as we assume strict complementarity in the lower level problem. Thus, the complementarity condition $\kappa^\intercal \gamma = 0$ can be used to conclude $\kappa^{\text{act}} = 0$. Consequently, we have

$$0 = \chi v^{\text{act}}(\bar{x}, \bar{y}) + D_y v^{\text{act}}(\bar{x}, \bar{y})\rho + \kappa^{\text{act}} = D_y v^{\text{act}}(\bar{x}, \bar{y})\rho = 0,$$

so that $\rho \in T$. Assume that $\rho \in T \setminus \{0\}$ holds. Then the lower level second order condition leads to

$$\rho^\intercal D_y^2 \mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma})\rho < 0$$

which contradicts (14). Hence, we have $\rho = 0$. In view of the lower level LICQ, equation (12) now yields $\mu = 0$.

It remains to show that χ is unique. This is analogous to the argumentation in Lemma 2.7 as, in view of the upper level LICQ, the stationarity condition with respect to x simplifies to

$$0 = D_x f(\bar{x}) - \chi D_x \mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma}).$$

The latter equation is the familiar first order optimality condition for problem GSIP from, e.g., [10, 13]. □

Remark 4.5 If $(\bar{x}, \bar{y}, \bar{\gamma})$ is a minimizer of the problem MPCC at which the lower level second order condition and lower level strict complementarity are satisfied, then the MPCC-LICQ condition from [6] for the problem MPCC is satisfied if and only if $(\bar{x}, \bar{y}, \bar{\gamma})$ satisfies the LLICQ condition from Definition 4.3. This immediately follows from the above theorem. In other words, the standard MPCC-LICQ condition and the LLICQ condition are equivalent for our class of problems.

5 Lifting without strict complementarity in the lower level problem

In this section we discuss a non-degeneracy result for the problem LWPC if the lower level strict complementarity condition is violated, i.e., if there are weakly active constraints in the maximization problem. Unfortunately, the result of Theorem 4.4 cannot be rescued in this case, as we can construct the following counter example:

Example 5.1 Consider the following semi-infinite optimization problem in min-max form:

$$\min_x -x \quad \text{s.t.} \quad \max_{y\geq 0} x - y^2 \leq 0.$$

This problem is both lower- and upper-level convex. The associated lifted problem can be written as

$$\text{LWPC:} \quad \min_{x,y,\gamma} -x \quad \text{s.t.} \quad \begin{cases} 0 \geq x - y^2 + \gamma y, \\ 0 = -2y + \gamma, \\ 0 \geq -y, \\ 0 \geq -\gamma. \end{cases} \tag{15}$$

The solution is obviously at $(\bar{x}, \bar{y}, \bar{\gamma})^\mathsf{T} = 0$. Note that the lower level constraint $y \geq 0$ is weakly active in $Q(0)$. However, the associated constraint Jacobian J of the active constraints at the solution is given by

$$J = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -2 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Thus, the LICQ condition for the problem (15) cannot possibly be satisfied, as there are simply too many constraints active. However, an interesting observation is that for example the vector $\xi := (-1, 1, 1)$ satisfies

$$J\xi = (-1, -1, -1)^\mathsf{T} < 0.$$

Consequently, the Mangasarian Fromovitz constraint qualification (MFCQ) for problem (15) is satisfied.

With the above example in mind, we ask the question whether we can prove MFCQ for the problem LWPC under more general conditions without requiring strict complementarity in the lower level problem. The aim of the following considerations is to show that this is indeed possible. Let us start with an introduction of the following extension of Definition 4.3:

Definition 5.2 For a point $(\bar{x}, \bar{y}, \bar{\gamma})$ we define the set $\mathbb{A}_\text{s}$ of strongly active constraints as well as the set $\mathbb{A}_\text{w}$ of weakly active constraints as

$$\mathbb{A}_s = \left\{ \ell \mid v_\ell(\bar{x}, \bar{y}) = 0 \wedge \bar{\gamma}_\ell > 0 \right\} \quad \text{and}$$

$$\mathbb{A}_w = \big\{\ell \mid v_\ell(\bar{x}, \bar{y}) = 0 \wedge \bar{\gamma}_\ell = 0\big\}, \quad \text{respectively.}$$

Moreover, we use the notation

$$T := \big\{p \in \mathbb{R}^m \mid \forall \ell \in \mathbb{A}_\mathrm{s} : D_y v_\ell(\bar{x}, \bar{y})p = 0 \wedge \forall \ell \in \mathbb{A}_\mathrm{w} : D_y v_\ell(\bar{x}, \bar{y})p \le 0\big\}.$$

Now, we say that the point $(\bar{x}, \bar{y}, \bar{\gamma})$ satisfies the lifted Mangasarian Fromovitz constraint qualification (LMFCQ) if the following requirements are satisfied:

1. There exist a vector $\xi_1 \in \mathbb{R}^{n_x}$ with $D_y v^{\mathrm{act}}(\bar{x}, \bar{y})\xi_1 < 0$.
2. The inequality $p^\mathsf{T} D_y^2 \mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma})p < 0$ holds for all $p \in T \setminus \{0\}$.
3. $\mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma}) = 0$ and $\nabla_x \mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma}) = 0$ do not hold simultaneously.

Note that the three conditions in Definition 5.2 correspond to the lower level MFCQ, lower level SOSC, and upper level MFCQ condition.

Before the main result of this section is stated, it is helpful to recall the following well-known theorem of alternatives for positive linear independent matrices [5]:

Lemma 5.3 (Gordan 1873) *Let $A \in \mathbb{R}^{n\times m}$ be a given matrix. Then exactly one of the following two alternatives holds*:

1. *There exists a vector $x \in \mathbb{R}^m$ with $Ax < 0$.*
2. *There exists a nonzero vector $y \in \mathbb{R}^n$ with $A^\mathsf{T} y = 0$ and $y \ge 0$.*

Using this technical result we can prove the following nondegeneracy result:

Theorem 5.4 *Let $(\bar{x}, \bar{y}, \bar{\gamma})$ be a minimizer of problem LWPC at which the lifted MFCQ condition from Definition 5.2 is satisfied. Then the gradients of active constraints among*

$$\mathcal{L}(x, y, \gamma) \le 0, \quad \nabla_y \mathcal{L}(x, y, \gamma) = 0, \qquad v(x, y) \le 0, \quad \gamma \ge 0$$

are positive linearly independent at $(\bar{x}, \bar{y}, \bar{\gamma})$, i.e., the MFCQ condition for the problem LWPC is satisfied. In particular, $(\bar{x}, \bar{y}, \bar{\gamma})$ is a KKT point of LWPC.

Proof Let us assume that the minimizer $(\bar{x}, \bar{y}, \bar{\gamma})$ does not satisfy the MFCQ condition of the problem LWPC. The aim of the proof is to show that this assumption leads to a contradiction. We distinguish two cases:

Case 1: The upper level constraint is active.

In this case, we can use Lemma 5.3 to conclude that there exist vectors $y_1 \in \mathbb{R}_+$, $y_2 \in \mathbb{R}^{n_y}$, $y_3 \in \mathbb{R}_+^{|\mathbb{A}_\mathrm{s}|+|\mathbb{A}_\mathrm{w}|}$, and $y_4 \in \mathbb{R}_+^{s-|\mathbb{A}_\mathrm{s}|}$ which are not all simultaneously zero and which satisfy the conditions

$$\begin{pmatrix} D_x \mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma}) & 0 & -v(\bar{x}, \bar{y})^\mathsf{T} \\ D_{yx}^2 \mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma}) & D_y^2 \mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma}) & -(D_y v)^\mathsf{T} \\ D_x v^{\mathrm{act}}(\bar{x}, \bar{y}) & D_y v^{\mathrm{act}}(\bar{x}, \bar{y}) & 0 \\ 0 & 0 & -E \end{pmatrix}^\mathsf{T} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = 0. \tag{16}$$

Here, we use the notation $E := D_\gamma \gamma^{\text{act}} \in \mathbb{R}^{(s-|\mathbb{A}_s|)\times s}$, where γ^{act} consists of the active components of γ with regard to the constraint $\gamma \geq 0$. Let us first analyze the third row of (16) in more detail which can also be written as

$$-y_1 v(\bar{x}, \bar{y}) - D_y v(\bar{x}, \bar{y}) y_2 - E^\mathsf{T} y_4 = 0. \tag{17}$$

This equation implies that we must have $D_y v_\ell(\bar{x}, \bar{y}) y_2 = 0$ for all $\ell \in \mathbb{A}_s$, as we have $v_\ell(\bar{x}, \bar{y}) = 0$ as well as $(E^\mathsf{T} y_4)_\ell = 0$ for all strongly active components. For the weakly active components $\ell \in \mathbb{A}_w$ we also have $v_\ell(\bar{x}, \bar{y}) = 0$, but we only know that $(E^\mathsf{T} y_4)_\ell \geq 0$. Consequently, we have $D_y v_\ell(\bar{x}, \bar{y}) y_2 \leq 0$ for all $\ell \in \mathbb{A}_w$. Both conditions together can be summarized as $y_2 \in T$.

In the next step of the proof, we multiply the second row in (16) with y_2^T from the left finding

$$y_2^\mathsf{T} D_y^2 \mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma}) y_2 - y_3^\mathsf{T} D_y v^{\text{act}}(\bar{x}, \bar{y}) y_2 = 0.$$

As we have already proven that $y_2 \in T$, we must have $D_y v^{\text{act}}(\bar{x}, \bar{y}) y_2 \leq 0$ and thus the above equation yields

$$y_2^\mathsf{T} D_y^2 \mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma}) y_2 \geq 0. \tag{18}$$

Now, as the second order sufficient condition of the lower level maximization problem is satisfied, we may conclude $y_2 = 0$. Note that the second row in equation (16) simplifies with $y_2 = 0$ to

$$\left(D_y v^{\text{act}}\right)^\mathsf{T} y_3 = 0.$$

Due to the fact that the lower level MFCQ condition is satisfied, this equation cannot have a positive nonzero solution as Lemma 5.3 would be violated otherwise. Thus, we may conclude $y_3 = 0$. Moreover, the first row in (16) yields

$$D_x \mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma})^\mathsf{T} y_1 = 0.$$

As the upper level MFCQ condition is satisfied, we may conclude $y_1 = 0$. Finally, we substitute $y_2 = 0$ and $y_1 = 0$ in (17) which leads to the relation

$$\forall \ell \in \{1, \ldots, s\} \setminus \mathbb{A}_s : \quad \left(E^\mathsf{T} y_4\right)_\ell = 0 \quad \Longrightarrow \quad y_4 = 0. \tag{19}$$

Thus, we have a contradiction as (16) has no nonnegative solution which is not zero.

Case 2: The upper level constraint is not active.

In this case we can show that the equation

$$\begin{pmatrix} D_{yx}^2 \mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma}) & D_y^2 \mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma}) & -(D_y v)^\mathsf{T} \\ D_x v(\bar{x}, \bar{y}) & D_y v(\bar{x}, \bar{y}) & 0 \\ 0 & 0 & -E \end{pmatrix}^\mathsf{T} \begin{pmatrix} y_2 \\ y_3 \\ y_4 \end{pmatrix} = 0 \tag{20}$$

has no nonzero solution with (y_2, y_3, y_4) with $y_3 \geq 0$ and $y_4 \geq 0$ by applying an analogous argumentation as in the case 1. □

6 Lifting in the degenerate case

As Example 2.2 shows, the case $Y(x) = \emptyset$ may occur at points $x \in X$, so that the assumptions of Theorem 2.3 are violated. Moreover, under the violated Slater condition, $Q(x)$ may not possess a Karush-Kuhn-Tucker point although $Y(x)$ is nonempty.

In this section we will show that, under Assumption 2.1 and mild additional assumptions, such degenerate cases can be tackled by the lifted problem

$$\widetilde{P}: \quad \min_{x,y,\alpha,\gamma} f(x) \quad \text{s.t.} \quad x \in X, \quad \widetilde{\mathcal{L}}(x,y,\alpha,\gamma) \le 0, \quad \nabla_y \widetilde{\mathcal{L}}(x,y,\alpha,\gamma) = 0,$$
$$(\alpha,\gamma) \in S^s,$$

where

$$\widetilde{\mathcal{L}}(x,y,\alpha,\gamma) = \alpha g(x,y) - \gamma^\mathsf{T} v(x,y)$$

is a 'Fritz John type' Lagrangian of the lower level problem, and where

$$S^s = \left\{ (\alpha,\gamma) \in \mathbb{R} \times \mathbb{R}^s \;\middle|\; (\alpha,\gamma) \ge 0,\ \alpha + e^\mathsf{T}\gamma = 1 \right\}$$

denotes the s-dimensional unit simplex.

To study situations in which Karush-Kuhn-Tucker points of $Q(x)$ do not exist, we first introduce a slight modification of an auxiliary lower level problem from [23]:

$$\widetilde{Q}(x): \quad \max_{(y,z)} z \quad \text{s.t.} \quad (y,z) \in \widetilde{Y}(x)$$

with $(y,z) \in \widetilde{Y}(x)$ if and only if

$$-g(x,y) + z \le 0, \qquad v_\ell(x,y) + z \le 0, \quad 1 \le \ell \le s.$$

By $\widetilde{\varphi}(x)$ we denote the optimal value of $\widetilde{Q}(x)$, and we put

$$\widetilde{M} = \left\{ x \in X \;\middle|\; \widetilde{\varphi}(x) \le 0 \right\}.$$

Upon defining

$$\sigma(x,y) = \min\left\{ g(x,y),\ -\max_{1\le\ell\le s} v_\ell(x,y) \right\}$$

the problem $\widetilde{Q}(x)$ may be interpreted as the epigraph reformulation of the unconstrained convex problem

$$\max_{y\in\mathbb{R}^m} \sigma(x,y).$$

To ensure solvability of the latter problem and, thus, of $\widetilde{Q}(x)$, we impose the following coercivity assumption.

Assumption 6.1 For each $x \in X$ we have

$$\lim_{\|y\|\to\infty} \sigma(x,y) = -\infty.$$

Example 6.2 In Example 2.2 the function $\sigma(x, y) = \min\{-y, y - x_1, x_2 - y\}$ satisfies Assumption 6.1 for $X = \mathbb{R}^2$. Note that this covers the case $Y(x) = \emptyset$. We obtain

$$\widetilde{\varphi}(x) = \max_{y \in \mathbb{R}} \sigma(x, y) = \begin{cases} -\frac{x_1}{2}, & x_2 \geq 0, \\ \frac{x_2 - x_1}{2}, & x_2 < 0 \end{cases}$$

and

$$\widetilde{M} = \left\{x \in \mathbb{R}^2 \,\middle|\, \widetilde{\varphi}(x) \leq 0\right\} = \left\{x \in \mathbb{R}^2 \,\middle|\, x \geq 0\right\} \cup \left\{x \in \mathbb{R}^2 \,\middle|\, x_2 \leq x_1, x_2 < 0\right\}.$$

Note that M is a proper subset of $\widetilde{M}$, but that $\widetilde{M}$ coincides with the topological closure $\overline{M}$ of M.

The closure effect from Example 6.2 can be reproduced in general under mild assumptions. As $\widetilde{\varphi}(x) \leq 0$ holds if and only if for all $y \in \mathbb{R}^m$ we either have $\max_{1 \leq \ell \leq s} v_\ell(x, y) \geq 0$ or $g(x, y) \leq 0$, we conclude

$$\widetilde{M} = \left\{x \in X \,\middle|\, \widetilde{\varphi}(x) \leq 0\right\} = \left\{x \in \mathbb{R}^n \,\middle|\, g(x, y) \leq 0 \text{ for all } y \in Y^<(x)\right\}$$

with

$$Y^<(x) = \left\{y \in \mathbb{R}^m \,\middle|\, v_\ell(x, y) < 0,\ 1 \leq \ell \leq s\right\}.$$

In view of $Y^<(x) \subset Y(x)$ we clearly have $M \subset \widetilde{M}$. As $Y^<(x)$ is 'only slightly smaller' than $Y(x)$, the set $\widetilde{M}$ can be expected to be 'only slightly larger' than M. Under mild additional assumptions one can actually show that $\widetilde{M}$ coincides with the topological closure $\overline{M}$ of M [8, 9].

To introduce these additional assumptions, recall that the set-valued mapping $Y : X \rightrightarrows \mathbb{R}^m$ is called locally bounded, if for each $\bar{x} \in X$ there exists a neighborhood U of $\bar{x}$ such that $\bigcup_{x \in U \cap X} Y(x)$ is bounded.

Assumption 6.3 The set-valued mapping $Y : X \rightrightarrows \mathbb{R}^m$ is locally bounded.

Assumption 6.3 is a standard assumption in generalized semi-infinite programming. It implies, for example, that $\widetilde{M}$ is a closed set [23].

The crucial assumption under which the set M enjoys nice topological properties is the so-called *symmetric Mangasarian-Fromovitz constraint qualification (Sym-MFCQ)* from [9]. For its statement we introduce the following notation: define $\sigma_0(x, y) = g(x, y)$, $\sigma_\ell(x, y) = -v_\ell(x, y)$, $1 \leq \ell \leq s$, as well as $L = \{0, \ldots, s\}$, so that we have

$$\sigma(x, y) = \min_{\ell \in L} \sigma_\ell(x, y).$$

For any $\bar{x} \in \widetilde{M}$ we define

$$\Sigma_0(\bar{x}) = \left\{y \in \mathbb{R}^m \,\middle|\, \sigma(\bar{x}, y) = 0\right\},$$

and for any $\bar{y} \in \Sigma_0(\bar{x})$ we put

$$L_0(\bar{x}, \bar{y}) = \left\{\ell \in L \,\middle|\, \sigma_\ell(\bar{x}, \bar{y}) = 0\right\}.$$

In view of [9, Proposition 1.2] one can define Sym-MFCQ at $\bar{x} \in \widetilde{M}$ by the existence of a vector $d \in \mathbb{R}^{n+m}$ with

$$D\sigma_\ell(\bar{x}, y)d < 0 \quad \text{for all } \ell \in L_0(\bar{x}, y),\ y \in \Sigma_0(\bar{x}).$$

Assumption 6.4 Sym-MFCQ holds at all $x \in \widetilde{M}$.

Under Assumption 6.3, Theorem 1.3 in [9] shows that Assumption 6.4 is mild. Moreover, under Assumptions 6.3 and 6.4, the set $\widetilde{M}$ coincides with the topological closure $\overline{M}$ of M ([9, Theorem 1.4]). Short calculations show that in Example 2.2 both, Assumption 6.3 and 6.4, are satisfied.

The tight connection between the sets M and $\widetilde{M}$ motivates us to study the problem

$$\widetilde{GSIP}: \quad \min f(x) \quad \text{s.t.} \quad \widetilde{\varphi}(x) \le 0$$

along the lines of Sect. 2.

With the Lagrange function of $\widetilde{Q}(x)$,

$$\begin{aligned}\widehat{\mathcal{L}}(x, y, z, \alpha, \gamma) &= z - \alpha\big(-g(x, y) + z\big) - \gamma^\mathsf{T}\big(v(x, y) + ze\big)\\ &= \big(1 - \alpha - \gamma^\mathsf{T} e\big)z + \alpha g(x, y) - \gamma^\mathsf{T} v(x, y),\end{aligned}$$

where $e \in \mathbb{R}^s$ denotes the all ones vector, we can state the Wolfe dual problem of $\widetilde{Q}(x)$ as follows:

$$\widehat{D}(x): \quad \min_{y,z,\alpha,\gamma} \widehat{\mathcal{L}}(x, y, z, \alpha, \gamma) \quad \text{s.t.} \quad \nabla_{(y,z)}\widehat{\mathcal{L}}(x, y, z, \alpha, \gamma) = 0, \quad \alpha, \gamma \ge 0.$$

Let $Y_{\widehat{D}}(x)$ denote the feasible set of $\widehat{D}(x)$. In view of $\nabla_z \widehat{\mathcal{L}}(x, y, z, \alpha, \gamma) = 0$, for each $(y, z, \alpha, \gamma) \in Y_{\widehat{D}}(x)$ we have $(\alpha, \gamma) \in S^s$ with the s-dimensional unit simplex

$$S^s = \big\{(\alpha, \gamma) \in \mathbb{R} \times \mathbb{R}^s \;\big|\; (\alpha, \gamma) \ge 0,\ \alpha + e^\mathsf{T}\gamma = 1\big\}.$$

As a consequence, the restriction of $\widehat{\mathcal{L}}(x, y, z, \alpha, \gamma)$ to $Y_{\widehat{D}}(x)$ does not depend on z and, with the function

$$\widetilde{\mathcal{L}}(x, y, \alpha, \gamma) = \alpha g(x, y) - \gamma^\mathsf{T} v(x, y),$$

the problem $\widehat{D}(x)$ is equivalent to

$$\widetilde{D}(x): \quad \min_{y,\alpha,\gamma} \widetilde{\mathcal{L}}(x, y, \alpha, \gamma) \quad \text{s.t.} \quad \nabla_y \widetilde{\mathcal{L}}(x, y, \alpha, \gamma) = 0, \quad (\alpha, \gamma) \in S^s.$$

Let $Y_{\widetilde{D}}(x)$ denote the feasible set of $\widetilde{D}(x)$, and let

$$\widetilde{\psi}(x) = \inf_{(y,\alpha,\gamma) \in Y_{\widetilde{D}}(x)} \widetilde{\mathcal{L}}(x, y, \alpha, \gamma)$$

be the optimal value function of $\widetilde{D}(x)$.

Under Assumption 6.1 the problem $\widetilde{Q}(x)$ is solvable for each $x \in X$, and $\widetilde{Y}(x)$ obviously satisfies the Slater condition. Thus, $\widetilde{Q}(x)$ possesses a Karush-Kuhn-Tucker point for each $x \in X$, and we obtain the following result.

Lemma 6.5 *Under Assumption 6.1 we have*

$$\widetilde{M} = \left\{ x \in X \;\middle|\; \min_{(y,\alpha,\gamma)\in Y_{\widetilde{D}}(x)} \widetilde{\mathcal{L}}(x, y, \alpha, \gamma) \leq 0 \right\}.$$

This finally motivates the introduction of the lifted problem

$$\widetilde{P}: \quad \min_{x,y,\alpha,\gamma} f(x) \quad \text{s.t.} \quad x \in X, \quad \widetilde{\mathcal{L}}(x, y, \alpha, \gamma) \leq 0, \quad \nabla_y \widetilde{\mathcal{L}}(x, y, \alpha, \gamma) = 0,$$
$$(\alpha, \gamma) \in S^s.$$

As f does not depend on the variables y, α, and γ, we have shown the following result.

Theorem 6.6 *Under Assumption 6.1 the minimizers of $\widetilde{GSIP}$ coincide with the x-components of the minimizers of $\widetilde{P}$.*

Corollary 6.7 *Under Assumptions 6.1, 6.3, and 6.4, the minimizers of f on $\overline{M}$ coincide with the x-components of the minimizers of $\widetilde{P}$.*

Example 6.8 Consider Example 2.2 with $X = \mathbb{R}^2$. In Example 6.2 we have seen that Assumption 6.1 is satisfied, and that $\overline{M}$ coincides with $\widetilde{M}$. By Theorem 6.6, the x-components of the minimizers of the lifted problem

$$\widetilde{P}: \quad \min_{(x,y,\alpha,\gamma)} f(x) \quad \text{s.t.} \quad -\alpha y - \gamma_1(x_1 - y) - \gamma_2(y - x_2) \leq 0,$$
$$-\alpha + \gamma_1 - \gamma_2 = 0, \quad \alpha, \gamma_1, \gamma_2 \geq 0, \quad \alpha + \gamma_1 + \gamma_2 = 1$$

coincide with the minimizers of f on $\overline{M}$.

7 Numerical results

In order to illustrate how the Wolfe dual based lifting strategy can be applied in practice, this section discusses some numerical examples.

7.1 Design centering

This first example is a classical application of generalized semi-infinite programming. Here, the test problem is given in the form

$$\min_{x\in R^3} -x_3$$
$$\text{s.t.} \quad \begin{cases} 0 \geq 1 - (x_1 + y_1)^2 - (x_2 + y_2)^2 & \forall y \in Y(x), \\ 0 \geq \frac{1}{4} - (x_1 - \frac{3}{2} + y_1)^2 - (x_2 + y_2)^2 & \forall y \in Y(x), \\ 0 \geq (x_1 - \frac{1}{2} + y_1)^2 - (x_2 + y_2)^2 - \frac{9}{4} & \forall y \in Y(x), \\ 0 \geq -x_2, \end{cases} \tag{21}$$

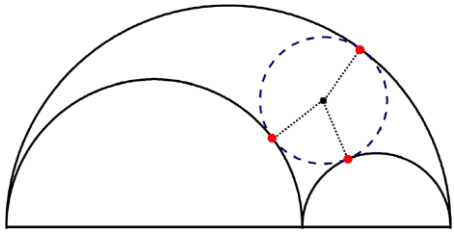


Fig. 1 A geometrical interpretation of the generalized semi-infinite optimization problem (21)

where the set $Y(x)$ depends explicitly on the third component of the optimization variable x:

$$Y(x) := \left\{ y \in \mathbb{R}^2 \;\middle|\; y^\mathsf{T} y \leq x_3 \right\}.$$

Note that the above GSIP has a simple geometrical interpretation which is visualized in Fig. 1. The first two components of the variable x can be interpreted as the center (x_1, x_2) of the dashed (blue) circle while the third component x_3 is the square of the corresponding radius. The three generalized semi-infinite constraints in the optimization problem require this circle to be contained in a circle with center $(0, \frac{1}{2})$ and radius $\frac{3}{2}$ but outside of two other circles with centers $(0, 0)$ and $(0, \frac{3}{2})$ and radius 1 and $\frac{1}{2}$, respectively. The objective is to maximize the radius of the dashed circle.

In this example, we may assume $x_3 > 0$, i.e., the set $Y(x)$ has a non-empty interior and Slater's constraint qualification is satisfied. The first two constraints are concave in y and we may directly apply the dual formulation strategy. For the third constraint, which requires the dashed circle to be contained in the large circle with radius $\frac{3}{2}$, it is less obvious that the dual reformulation strategy is correct as this constraint is convex in y. Fortunately, for this special case that the constraint function is a quadratic form in w while the set $Y(x)$ is a parameterized ellipsoid, the S-procedure [32] can be applied, i.e., there is no duality gap.

In the following, we formulate the above GSIP once as an MPCC, once as an LWP, and once as an LWPC. Here, we discuss the performance of a standard full-step SQP algorithm with BFGS Hessian updates which is initialized at

$$\begin{aligned} x^0 &:= (1, 1, 0.1)^\mathsf{T}, \qquad \gamma^0 := (1, 1, 1)^\mathsf{T}, \\ y_1^0 &:= (-0.1, -0.1)^\mathsf{T}, \qquad y_2^0 := (0.1, -0.1)^\mathsf{T}, \qquad y_3^0 := (0.1, 0.1)^\mathsf{T}. \end{aligned}$$

Moreover, we stop the SQP solver whenever it finds the optimal solution

$$\begin{aligned} x^* &\approx (1.14, 0.86, 0.18)^\mathsf{T}, \qquad \gamma^* \approx (2.33, 1.16, 3.50)^\mathsf{T}, \\ y_1^* &\approx (-0.34, -0.26)^\mathsf{T}, \qquad y_2^* \approx (0.16, -0.40)^\mathsf{T}, \qquad y_3^* \approx (0.26, 0.34)^\mathsf{T} \end{aligned}$$

achieving a KKT-Tolerance of less than 10^{-9}. Note that the optimal solutions for the lower level maximizers y_1^*, y_2^*, and y_3^* are visualized in Fig. 1 in form of the red points which are the touching points of the dashed circle with the three constraints.

If we employ the above SQP solver settings, the following numerical results can be obtained:

1. In order to apply our standard SQP solver to the MPCC reformulation of the problem (21) we replace the lower feasibility and complementarity constraints by a smoothed Fisher-Burmeister function of the form

$$\forall i \in \{1, \ldots, s\}: \quad \gamma_i - v_i(x, y) - \sqrt{\gamma_i^2 + v_i(x, y) + 2\tau^2} = 0,$$

   where $\tau > 0$ is a homotopy parameter. Now, we solve the corresponding smoothed MPCC with the SQP solver for $\tau = 10^i$ for the homotopy sequence $i = 0$, $i = 1, \ldots,$ and $i = 9$, respectively, where the SQP solver is always warm-started with the previous solution. In this case, a total number of 61 SQP iterations are needed to solve the original GSIP with a numerical accuracy of $\approx 10^{-9}$.
2. If we reformulate the problem (21) into an LWP, the SQP solver needs 42 iterations to converge to the optimal solution.
3. Finally, if we reformulate the problem (21) into an LWPC, the SQP solver needs 56 iterations to achieve a comparable accuracy.

Summarizing these results, the lifted problem reformulations LWP and LWPC lead in general to a faster convergence. In particular, these lifted reformulations avoid the introduction of smoothing heuristics which are needed for solving an MPCC with standard optimization algorithms. In this example, we have observed that the formulation based on the LWP leads to less SQP iterations than the formulation based on the LWPC does. This might be explained by the fact that in this particular example strong duality holds independently of the choice of the upper level variable x, i.e., the additional index constraints in the LWPC formulation are not necessarily needed.

7.2 Chebyshev approximation

The next example is taken from Chebychev approximation. The problem we investigate is

$$\text{CA:} \quad \min_{x \in \mathbb{R}^5} \max_{y \in Y} \left| \sin(y_1) + \sin(y_2) - a(x, y) \right|,$$

with $a(x, y) = x_1 y_1^2 + x_2 y_1 + x_3 y_2^2 + x_4 y_2 + x_5$ and $Y = [0, \pi] \times [0, \pi]$.

One can rewrite CA as

$$\begin{aligned} \text{SIP}_{\text{CA}}: \quad \min_{x \in \mathbb{R}^6} x_6 \quad \text{s.t.} \quad & \sin(y_1) + \sin(y_2) - a(x, y) - x_6 \leq 0 \quad \text{for all } y \in Y, \\ & -\sin(y_1) - \sin(y_2) + a(x, y) - x_6 \leq 0 \quad \text{for all } y \in Y, \end{aligned}$$

which is a standard semi-infinite programming problem. It is not hard to see that SIP_{CA} violates Assumption 2.1.

Hence, our algorithm uses the techniques discussed in [7, 26] as convexification strategies for the constraints.

The main idea of these techniques is to partition the index set Y adaptively into smaller sets, and to construct convex relaxations of the original constraints on these sets.

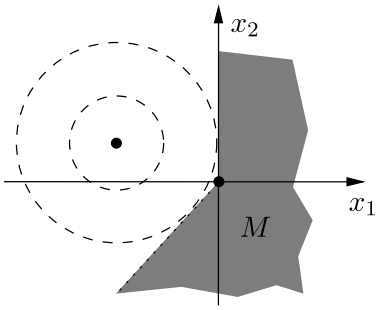


Fig. 2 Feasible set and level lines of objective function

As in the former example we reformulate the problem once as an MPCC, once as an LWP, and once as an LWPC. In all formulations we used $x^0 = (-3, 5, -3, 5, 2, 2)^\mathsf{T}$ as a starting point for x in the algorithm. To obtain starting values for y and γ we applied a phase 1 of the algorithm that is similar to the one discussed in [7].

For the MPCC reformulation the method identified

$$x^* = (-4, 4, -4, 4, -0.056, 0.056)^\mathsf{T}$$

as an approximately stationary point after 147 iterations.

In x^* the norm of the stationarity condition is less then 10^{-16}. During the iterations the algorithm generated a partition of Y with an overall number of 1197 boxes. For the LWP reformulation we had to abort the algorithms in the fourth iteration since some of the computed indices y on the generated partition of the index set Y were out of the valid range. This result is not surprising since the relaxed constraints are not convex on the entire partition of the set Y, and, thus, one has to apply the techniques discussed in Sect. 4 to avoid this outcome. With the LWPC reformulation we obtain the same results as for the MPCC reformulation, but, after 103 iterations and an overall number of 735 boxes in the generated partition of Y.

7.3 The degenerate case

Our last example briefly illustrates that the lifting approach for the degenerate case from Sect. 6 works numerically.

In fact, we consider the GSIP

$$\min_{x \in \mathbb{R}^2} \left(x_1 + 1 + \sqrt{2}\right)^2 + (x_2 - 1)^2 \quad \text{s.t.} \quad y \geq 0 \quad \text{for all } y \text{ with } x_1 \leq y \leq x_2$$

and its reformulation $\widetilde{P}$ from Example 6.8.

The feasible set M and some level lines of the objective function are depicted in Fig. 2.

It is not hard to see that the set of globally minimal points of f on the closure of M consists of the two points $(0, 1)^\mathsf{T}$ and $-1/\sqrt{2}(1, 1)^\mathsf{T}$. A standard SQP solver with initial points $x^0 := (1, 0)^\mathsf{T}$, $y^0 = 0$, $\alpha^0 = 0$, $\gamma^0 = (0, 1)$ then identifies the optimal point $(0, 1)^\mathsf{T}$, whereas replacing x^0 by $(0, -1)^\mathsf{T}$ leads to the optimal point $-1/\sqrt{2}(1, 1)^\mathsf{T}$.

8 Conclusions

In this article we have shown that lifting approaches for generalized semi-infinite optimization problems do not necessarily lead to mathematical programs with complementarity constraints, that is, to nonsmooth finite optimization problems. Under appropriate assumptions we rather lifted GSIPs to smooth and nondegenerate finite optimization problems so that, as opposed to MPCCs, there is no need to employ tailored numerical algorithms for their solution. The main contribution of this article is a novel problem formulation that can be regarded a mixture between the numerically badly behaved but well understood MPCC approach and the dual representation of the infinite constraints which however cannot deal with the practically relevant case when the lower level problems are non-convex outside their feasible domain. We have shown that using lower level Wolfe duality, but keeping the lower level constraints in the problem, leads to a numerically well behaved formulation. A slight modification of our lifting approach even works for degenerate situations in which the feasible set of GSIP is not closed. Our numerical experience so far indicates that the approach is implementable and successful.

Acknowledgements We thank two anonymous referees and the associated editor for their precise and substantial remarks which significantly improved this manuscript.

The research was supported by the Research Council KUL via GOA/11/05 Ambiorics, GOA/10/09 MaNet, CoE EF/05/006 Optimization in Engineering (OPTEC), IOF-SCORES4CHEM and PhD/postdoc/fellow grants, the Flemish Government via FWO (PhD/postdoc grants, projects G0226.06, G0321.06, G.0302.07, G.0320.08, G.0558.08, G.0557.08, G.0588.09, G.0377.09, research communities ICCoS, ANMMM, MLDM) and via IWT (PhD Grants, Eureka-Flite+, SBO LeCoPro, SBO Climaqs, SBO POM, O&O-Dsquare), the Belgian Federal Science Policy Office: IUAP P6/04 (DYSCO, Dynamical systems, control and optimization, 2007–2011), the IBBT, the EU (ERNSI; FP7-HD-MPC (INFSO-ICT-223854), COST intelliCIS, FP7-EMBOCON (ICT-248940), FP7-SADCO (MC ITN-264735), ERC HIGHWIND (259 166)), the Contract Research (AMINAL), the Helmholtz Gemeinschaft via viCERP and the ACCM.

References

1. Ben-Tal, A., Nemirovski, A.: Robust convex optimization. Math. Oper. Res. **23**, 769–805 (1998)
2. Ben-Tal, A., El Ghaoui, L., Nemirovski, A.: Robust Optimization. Princeton University Press, Princeton (2009)
3. Bhattacharjee, B., Green, W.H., Barton, P.I.: Interval methods for semi-infinite programs. Comput. Optim. Appl. **30**, 63–93 (2005)
4. Bhattacharjee, B., Lemonidis, P., Green, W.H., Barton, P.I.: Global solution of semi-infinite programs. Math. Program. **103**, 283–307 (2005)
5. Dantzig, G.B., Thapa, M.N.: Linear Programming. Springer, Berlin (2003)
6. Flegel, M.L., Kanzow, C.: A Fritz John approach to first order optimality conditions for mathematical programs with equilibrium constraints. Optimization **52**, 277–286 (2003)
7. Floudas, C.A., Stein, O.: The adaptive convexification algorithm: a feasible point method for semi-infinite programming. SIAM J. Optim. **18**, 1187–1208 (2007)
8. Günzel, H., Jongen, H.Th., Stein, O.: On the closure of the feasible set in generalized semi-infinite programming. Cent. Eur. J. Oper. Res. **15**, 271–280 (2007)
9. Guerra-Vázquez, F., Jongen, H.Th., Shikhman, V.: General semi-infinite programming: symmetric Mangasarian-Fromovitz constraint qualification and the closure of the feasible set. SIAM J. Optim. **20**, 2487–2503 (2010)
10. Hettich, R., Jongen, H.Th.: Semi-infinite programming: conditions of optimality and applications. In: Stoer, J. (ed.) Optimization Techniques, Part 2. Lecture Notes in Control and Information Sciences, vol. 7, pp. 1–11. Springer, Berlin (1978)

11. Hettich, R., Kortanek, K.: Semi infinite programming: theory, methods, and applications. SIAM Rev. **35**, 380–429 (1993)
12. Hettich, R., Zencke, P.: Numerische Methoden der Approximation und Semi-Infiniten Optimierung. Teubner, Stuttgart (1982)
13. Jongen, H.Th., Rückmann, J.J., Stein, O.: Generalized semi-infinite optimization: a first order optimality condition and examples. Math. Program. **83**, 145–158 (1998)
14. Kropat, E., Weber, G.W.: Robust regression analysis for gene-environment and eco-finance networks under polyhedral and ellipsoidal uncertainty. Preprint at IAM, METU
15. Levitin, E., Tichatschke, R.: A branch-and-bound approach for solving a class of generalized semi-infinite programming problems. J. Glob. Optim. **13**, 299–315 (1998)
16. Mitsos, A., Lemonidis, P., Lee, C.K., Barton, P.I.: Relaxation-based bounds for semi-infinite programs. SIAM J. Optim. **19**, 77–113 (2007)
17. Mitsos, A., Lemonidis, P., Barton, P.I.: Global solution of bilevel programs with a nonconvex inner program. J. Glob. Optim. **42**, 475–513 (2008)
18. Polak, E.: On the mathematical foundation of nondifferentiable optimization in engineering design. SIAM Rev. **29**, 21–89 (1987)
19. Polak, E.: Optimization. Algorithms and Consistent Approximations. Springer, Berlin (1997)
20. Polak, E., Royset, J.O.: On the use of augmented Lagrangians in the solution of generalized semi-infinite min-max problems. Comput. Optim. Appl. **31**, 173–192 (2005)
21. Reemtsen, R., Görner, S.: Numerical methods for semi-infinite programming: a survey. In: Reemtsen, R., Rückmann, J.J. (eds.) Semi-Infinite Programming, pp. 195–275. Kluwer, Boston (1998)
22. Scheel, H., Scholtes, S.: Mathematical programs with complementarity constraints: stationarity, optimality, and sensitivity. Math. Oper. Res. **25**, 1–22 (2000)
23. Stein, O.: First order optimality conditions for degenerate index sets in generalized semi-infinite programming. Math. Oper. Res. **26**, 565–582 (2001)
24. Stein, O.: Bi-level Strategies in Semi-infinite Programming. Kluwer Academic, Boston (2003)
25. Stein, O.: How to solve a semi-infinite optimization problem. Eur. J. Oper. Res. (2012). doi:10.1016/j.ejor.2012.06.009
26. Stein, O., Steuermann, P.: The adaptive convexification algorithm for semi-infinite programming with arbitrary index sets. Math. Program. B (2012). doi:10.1007/s10107-012-0556-5
27. Stein, O., Still, G.: Solving semi-infinite optimization problems with interior point techniques. SIAM J. Control Optim. **42**, 769–788 (2003)
28. Stein, O., Winterfeld, A.: A feasible method for generalized semi-infinite programming. J. Optim. Theory Appl. **146**, 419–443 (2010)
29. Still, G.: Discretization in semi-infinite programming: the rate of convergence. Math. Program. **91**, 53–69 (2001)
30. Still, G.: Generalized semi-infinite programming: numerical aspects. Optimization **49**, 223–242 (2001)
31. Weber, G.W., Tezel, A.: On generalized semi-infinite optimization of genetic networks. Top **15**, 65–77 (2007)
32. Yakubovich, V.A.: S-procedure in nonlinear control theory. Vestn. Leningr. Univ. **4**, 73–93 (1977)