Math. Program., Ser. A (2013) 142:539–577
DOI 10.1007/s10107-012-0591-2

FULL LENGTH PAPER

Nonlinear robust optimization via sequential convex bilevel programming

Boris Houska · **Moritz Diehl**

Received: 23 July 2010 / Accepted: 10 August 2012 / Published online: 15 September 2012
© Springer and Mathematical Optimization Society 2012

Abstract In this paper, we present a novel sequential convex bilevel programming algorithm for the numerical solution of structured nonlinear min–max problems which arise in the context of semi-infinite programming. Here, our main motivation are nonlinear inequality constrained robust optimization problems. In the first part of the paper, we propose a conservative approximation strategy for such nonlinear and nonconvex robust optimization problems: under the assumption that an upper bound for the curvature of the inequality constraints with respect to the uncertainty is given, we show how to formulate a lower-level concave min–max problem which approximates the robust counterpart in a conservative way. This approximation turns out to be exact in some relevant special cases and can be proven to be less conservative than existing approximation techniques that are based on linearization with respect to the uncertainties. In the second part of the paper, we review existing theory on optimality conditions for nonlinear lower-level concave min–max problems which arise in the context of semi-infinite programming. Regarding the optimality conditions for the concave lower level maximization problems as a constraint of the upper level minimization problem, we end up with a structured mathematical program with complementarity constraints (MPCC). The special hierarchical structure of this MPCC can be exploited in a novel sequential convex bilevel programming algorithm. We discuss the surprisingly strong global and locally quadratic convergence properties of this method, which can in this form neither be obtained with existing SQP methods nor with interior point relaxation techniques for general MPCCs. Finally, we discuss the application fields and

B. Houska (✉)· M. Diehl
Optimization in Engineering Center (OPTEC), K. U. Leuven, Kasteelpark Arenberg 10, 3001 Leuven-Heverlee, Belgium
e-mail: boris.houska@esat.kuleuven.be

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

implementation details of the new method and demonstrate the performance with a numerical example.

Keywords Robust optimization · Mathematical programming with complementarity constraints · Bilevel optimization · Semi-infinite optimization · Sequential convex programming

Mathematics Subject Classification 90C55 · 90C47 · 90C34

1 Introduction

In this paper, we consider inequality constrained optimization problems of the form

$$\begin{aligned} \min_{x \in \mathbb{R}^{n_x}} \quad & F_0(x, w) \\ \text{subject to} \quad & F_i(x, w) \leq 0 \quad \text{for all} \quad i \in \{1, \ldots, n\} \end{aligned} \tag{1}$$

with an uncertainty $w \in \mathbb{R}^{n_w}$ entering both the continuous objective function F_0 as well as the continuous constraint functions F_1,..., F_n. The robust counterpart methodology, developed by Ben-Tal and Nemirovski [5–7] and El-Ghaoui et al. [17], assumes that we have additional knowledge about the uncertainty w, namely that it lies in a given compact uncertainty set $W(x) \subset \mathbb{R}^{n_w}$. We are interested in the following worst-case formulation which incorporates our knowledge about the uncertainty:

$$\begin{aligned} \min_{x \in \mathbb{R}^{n_x}} \quad & \max_{w \in W(x)} \quad F_0(x, w) \\ \text{subject to} \quad & \max_{w \in W(x)} \quad F_i(x, w) \leq 0 \quad \text{for all} \quad i \in \{1, \ldots, n\}. \end{aligned}$$

In general it is hard to solve such min–max problems. Most existing algorithms address special cases in the context of convex optimization: in [6], Ben-Tal and Nemirovski observe that an important point in the formulation of robust counterpart problems is their computational tractability. Under the assumption that the uncertainty set is ellipsoidal they were able to show that the robust counterpart of a linear programming problem (LP) with uncertain data can be formulated as a second order code programming problem (SOCP). Similarly, the robustification of quadratic- or second order cone programs (QPs or SOCPs) leads to semi-definite programming problems (SDP). The robust counterpart of an SDP is NP-hard to solve. These achievements show that, even if we are able to reformulate the robust counterpart problem, the robustification increases the difficulty of the problem in the sense that SDPs are harder to solve than SOCPs which are in turn more difficult to treat than LPs. Note that the field of research addressing robust convex optimization problems has expanded fast during the last years [4,8]. Although these developments tend more and more towards approximation techniques, where the robust counterpart problem is replaced by more tractable formulations, they also cover an increasing amount of applications.

For the non-convex case exist approaches in literature [15,27,29,35] which suggest approximation techniques based on the assumption that w lies in a "small" set $W(x)$ or equivalently that the curvatures of the objective function F_0 as well as the constraint functions $F_1, \ldots, F_n$ are bounded by given constants such that the dependence of $F_0, F_1, \ldots, F_n$ can be described by a Taylor expansion where the second order term is over-estimated such that a conservative approximation is obtained. This linearization allows in some cases to compute the maxima in an explicit way. As in the convex case, these approaches usually assume that the uncertainty sets are ellipsoidal (while the ellipsoids might however be nonlinearly parameterized in x) such that the sub maximization problems can easily be eliminated while the conservatively robustified minimization problem is solved with existing NLP algorithms. Note that in [35,36] this approach has also been considered for more general polynomial chaos expansions, i.e., higher order Taylor expansions with respect to the unknowns are regarded. However, in practice it is often already quite expensive to compute linearizations of the functions $F_0, F_1, \ldots, F_n$ with respect to the uncertainty—especially if we think of optimal control problems where such an evaluation requires to solve possibly nonlinear differential along with their associated variational differential equations. This cost might increase dramatically, if higher order expansions have to be computed while the polynomial sub-maximization problems can itself only approximately be solved which requires again a level of conservatism.

For the case that polynomial approximations of the problem functions with respect to the uncertainties are not acceptable, the completely nonlinear robust optimization problem must be regarded. This completely nonlinear case has been studied in the context of semi-infinite programming [25]. Here, the term *semi-infinite* arises from the observation that the constraints of an uncertainty have to be satisfied for all possible realizations of the variables w in the given uncertainty set $W(x)$, i.e., an infinite number of constraints must be regarded. Here, the problems in which the set W may depend on x are usually called generalized semi-infinite programming (GSIP) problems while the name semi-infinite programming (SIP) is reserved for the case that the uncertainty set W is constant. The growing interest of semi-infinite optimization problems over the last decades has resulted in various contributions about the feasible set of these problems [30,49,54]. Moreover, first and second order optimality conditions for SIP and GSIP problems have been studied intensively [26,30,61]. However, when it comes to numerical algorithms semi-infinite optimization problems turn out to be in their general form rather expensive to solve. Some authors have discussed discretization strategies for the uncertainty set in order to replace the infinite number of constraints by a finite approximation [25,58,59]. Although this approach works acceptably for very small dimensions n_w, the curse of dimensionality hurts for $n_w \gg 1$ such that discretization strategies are in this case rather conceptual. Note that the situation is very different if additional concavity assumptions are available. Indeed, as semi-infinite optimization problems can under mild assumptions [55] be regarded as a Stackelberg game [53], the lower level maximization problems can—in the case of concavity—equivalently be replaced by their first order optimality conditions, which leads to an mathematical program with complementarity constraints (MPCC). In this context, we also note that semi-infinite optimization problems can be regarded as a special bilevel optimization problem [3]. However, as we shall argue in this paper, semi-infinite programming problems should not be treated as if they were a general bilevel optimization problem as important structure is lost otherwise.

Being at this point, semi-infinite optimization problems give rise to convexification methods with the aim to equivalently replace or to conservatively approximate the lower level maximization problems with a concave optimization problem. As discussed above, one way to obtain a convexification is linearization. However, in the field of global optimization more general Lagrangian underestimation (or, for maximization problems, overestimation) techniques are well-known tools [51,52,60] for convexification which are often used as a starting point for the development of branch-and-bound algorithms. In the context of generalized semi-infinite programming such a concave overestimation technique has been suggested in [22] to deal with the problem of finding the global solution of the lower level maximization problems discussing the case where the uncertainty is assumed to be in a given one-dimensional interval. The corresponding technique is called α-relaxation and works in principle also for uncertainties with dimension $n_w > 1$ which are bounded by a box. For $n_w \gg 1$ the α-relaxation can be used as a conservative approximation while the authors in [22] suggest for the case of small n_w to combine this α-overestimation with a branch-and-bound technique (α-BB method) which converges to the exact solution.

Concerning algorithms for nonlinear min–max optimization problems there exist some approaches which use recursive quadratic programming [34,43]. Most of these approaches concentrate either on the case of one-dimensional uncertainty intervals or finite approximations of the semi-infinite constraint. In this context, we also highlight the superlinearly convergent min–max algorithms [40,42] as well as the first order min-max algorithms which have been proposed in [41]. Note that these algorithmic developments work with finite but adaptive discretizations of semi-infinite optimization problems.

In this paper, we employ convexification methods for the semi-infinite or robust optimization problems of our interest, which will be discussed within Sect. 2. In contrast to the considerations in [22], we directly concentrate on the case $n_w \gg 1$, i.e., on the case that the dimension of the uncertainty is much larger than one. This means in particular that branch-and-bound methods are out of scope and we are rather interested in conservative approximations. The contribution of Sect. 2.2 is that we show for the case of ellipsoidal uncertainty sets that Lagrangian based overestimation techniques are always less conservative than linear approximations. Moreover, we discuss some non-concave cases for which a particular Lagrangian based concave overestimation is exact. In Sect. 3 we review the existing achievements for generalized semi-infinite programming in terms of first and second order optimality conditions.

The main contribution of this paper is presented in Sect. 4, where a new method for structured mathematical programming problems with complementarity constraints (MPCCs) is developed. Here, the particular structure of the MPCCs arises from the nature of semi-infinite programming problems. As the presented method solves in each step a convex bilevel optimization problem, we suggest the name *sequential convex bilevel programming* which can be interpreted as a generalization of sequential quadratic programming (SQP) methods [39]. Here, we discuss the local and global convergence properties of the presented method which can—as far as the authors are aware—in this form not easily be obtained with any existing method for MPCCs.

In Sect. 5, we discuss application fields and implementation details of the presented sequential convex bilevel programming method. Moreover, we demonstrate the applicability of the method by testing it with a numerical example which arises from the field of robust optimal control. Finally, we conclude the paper in Sect. 6.

2 Nonlinear robust counterpart problems

In this section we introduce the standard notation for nonlinear min–max optimization problems that arise in the context of robust optimization problems. We regard problem (1) with twice continuously differentiable functions

$$F_0, F_1, \ldots, F_n : \mathbb{R}^{n_x} \times \mathbb{R}^{n_w} \to \mathbb{R}$$

depending on an optimization variable $x \in \mathbb{R}^{n_x}$ and on an uncertain parameter w, which is known to be in a compact and non-empty set

$$w \in W(x) \subseteq \mathbb{R}^{n_w}$$

which additionally satisfies $0 \in W(x)$. We assume that whatever x the optimizer chooses, the adverse player "nature" chooses the worst possible value $V_i(x)$ defined by

$$V_i(x) := \max_{w \in W(x)} F_i(x, w). \tag{2}$$

Our aim is now to solve the associated worst-case minimization or robust counterpart problem

$$\begin{aligned} \min_{x \in \mathbb{R}^{n_x}} \quad & V_0(x) \\ \text{subject to} \quad & V_i(x) \leq 0 \quad \text{for all} \quad i \in \{1, \ldots, n\}. \end{aligned} \tag{3}$$

Following the naming conventions for semi-infinite programming problems, we suggest to call the above problem (3) a generalized robust counterpart problem in the case that W depends on x, while we speak of a standard robust counterpart problem otherwise. Note that the above "min–max" optimization problem (3) requires in each evaluation of the functions $V_0, \ldots, V_n$ the solution of the associated sub maximization problem of the form (2). Thus, if the functions $F_0, F_1, \ldots, F_n$ are non-convex there is in general no numerically efficient algorithm possible: every time we need to evaluate the functions $V_0, \ldots, V_n$ we have to apply expensive global search routines (e.g., branch-and-bound) for solving the maximization problems (2)—even if we are only interested in conservative approximations. As this will be limited to small n_w we introduce the following assumption:

Assumption 1 Let us assume that we have for each $i \in \{0, \ldots, n\}$ a twice continuously differentiable and non-negative function $\overline{\lambda}_i : \mathbb{R}^{n_x} \to \mathbb{R}^+$ which satisfies the inequality

$$\forall w \in W(x) : \quad \lambda_{\max}\left(\frac{\partial^2}{\partial w^2} F_i(x, w)\right) \leq 2\,\overline{\lambda}_i(x), \tag{4}$$

i.e., the maximum eigenvalue of the Hessian of F_i with respect to w is for all $w \in W(x)$ bounded by the function $2\,\overline{\lambda}_i$.

Note that there exist numerical techniques from the field of global optimization [10, 21, 38] which are able to provide interval bounds on the eigenvalues of the Hessian matrix of a given function as required in the above assumption. Nevertheless, the above assumption is still questionable, as it is in practice often not clear how we can obtain such functions $\overline{\lambda}_i$ if the suggested global numerical interval methods are too expensive to be applied. However, once we accept this assumption, we are able to develop efficient, derivative based algorithms for approximate robust optimization in the case $n_w \gg 1$. This is the aim of the present paper.

2.1 Approximate robust counterpart formulations based on linearization

Note that the Assumption 1 enables us to construct a conservative approximation for the maximization problems (2) based on linearization: by Taylor expansion we find

$$\begin{aligned} V_i(x) &= \max_{w \in W(x)} \; F_i(x, w) \\ &\leq \max_{v, w \in W(x)} \; \left\{ F_i(x, 0) + \frac{\partial F_i(x, 0)}{\partial w} w + \frac{1}{2} w^T \left(\frac{\partial^2}{\partial w^2} F_i(x, v) \right) w \right\} \\ &\leq \max_{w \in W(x)} \; \left\{ F_i(x, 0) + \frac{\partial F_i(x, 0)}{\partial w} w + \overline{\lambda}_i(x)\, w^T w \right\} \end{aligned} \tag{5}$$

Note that the uncertainty set can in practice often be modeled as an ellipsoidal set $\mathcal{B}$. In order to briefly discuss this case, we assume here for simplicity that $\mathcal{B}$ is a unit ball:

$$w \in \mathcal{B} := \left\{ v \in \mathbb{R}^{n_w} \mid v^T v \leq 1 \right\}.$$

In this case, we can explicitly solve the concave problem (5) finding the overestimate:

$$V_i(x) \leq F_i(x, 0) + \left\| \frac{\partial F_i(x, 0)}{\partial w} \right\|_2 + \overline{\lambda}_i(x). \tag{6}$$

Here and in the following $\| \cdot \|_2 : \mathbb{R}^{n_w} \to \mathbb{R}$ denotes the Euclidean norm. Note that this linear overestimate has in the context of robust optimization been introduced in [15, 35].

Definition 1 We define the best conservative first order approximation $\Lambda_i : \mathbb{R}^{n_x} \to \mathbb{R}$ associated with the ith maximization problem of the form (2) by

$$\forall x \in \mathbb{R}^{n_x} : \quad \Lambda_i(x) := F_i(x, 0) + \left\| \frac{\partial F_i(x, 0)}{\partial w} \right\|_2 + \overline{\lambda}_i(x). \tag{7}$$

The above definition is motivated by the observation that once we linearize the function F_i at $w = 0$ allowing neither to compute the gradient of F_i at any other point nor to compute any second order term, Λ_i is the smallest conservative approximation that we can obtain by using Assumption 1 only without having any further information on the function F_i.

Note that, e.g., in [15,27,29,35] it was suggested to solve the approximate robust counterpart problem

$$\begin{aligned} \min_{x\in\mathbb{R}^{n_x}} \quad & \Lambda_0(x) \\ \text{subject to} \quad & \Lambda_i(x) \le 0 \quad \text{for all} \quad i \in \{1,\dots,n\}. \end{aligned} \tag{8}$$

instead of the original problem (3). In this paper we are interested in the question whether we can find an alternative to the linear approximation approach which leads to a less conservative approximation of the worst case.

2.2 A worst case approximation based on the dual Lagrange function

In this section, we pick any $i \in \{0,\dots,n\}$ and ask once more the question how we can compute an upper bound on the function $V_i(x)$ which is needed in robust counterpart formulations. As in the previous consideration, we assume that $W(x) = \mathcal{B}$ is the unit ball. Recall that our only information about the function F_i is that Assumption 1 holds.

Let us consider the Lagrange dual function $d_i : \mathbb{R}^{n_x} \times \mathbb{R}^+ \to \mathbb{R}$, which is associated with the maximization problem (2):

$$\begin{aligned} d_i(x,\lambda_i) &:= \max_{w_i} \; G_i(x,\lambda_i,w_i) \quad \text{with} \\ G_i(x,\lambda_i,w_i) &:= F_i(x,w_i) - \lambda_i w_i^T w_i + \lambda_i. \end{aligned} \tag{9}$$

Note that d_i is an upper bound on $V_i(x)$, i.e., we have

$$\forall x \in \mathbb{R}^{n_x} : \quad V_i(x) \le \min_{\lambda_i \ge 0} \; d_i(x,\lambda_i). \tag{10}$$

For the case that the above inequality holds with equality we say that the strong duality condition is satisfied. This is for example the case if F_i is strictly concave in w.

So far, we have not solved the problem: we still need to solve the optimization problem (9) globally. However, an interesting observation is that we have

$$\forall x \in \mathbb{R}^{n_x} : \quad M_i(x) := \min_{\lambda_i \ge \overline{\lambda}_i(x)} \; d_i(x,\lambda_i) \;\ge\; \min_{\lambda_i \ge 0} \; d_i(x,\lambda_i), \tag{11}$$

since we assume that $\overline{\lambda}_i$ is a non-negative function. Note that $d_i(x,\lambda_i)$ is for $\lambda_i \ge \overline{\lambda}_i(x)$ easier to evaluate in the sense that the function $G_i(x,\lambda_i,\cdot)$ is concave, i.e., every local maximum of the function $G_i(x,\lambda_i,\cdot)$ is also a global maximum if $\lambda_i \ge \overline{\lambda}_i(x)$ is satisfied.

Lemma 1 *The function M_i is an upper bound on V_i which can never be more conservative than the best linear approximation Λ_i, i.e., we have*

$$\forall x \in \mathbb{R}^{n_x} : \quad V_i(x) \leq M_i(x) \leq \Lambda_i(x). \tag{12}$$

Proof Note that by using the Taylor expansion of the function G_i with respect to w_i there exists a $v \in \mathbb{R}^{n_w}$ such that

$$G_i(x, \lambda_i, w_i) = F_i(x, 0) + \frac{\partial}{\partial w} F_i(x, 0) w_i + \frac{1}{2} w_i^T \left(\frac{\partial^2}{\partial w^2} F_i(x, v) - 2\lambda_i \mathbf{1} \right) w_i + \lambda_i . \tag{13}$$

For the case $\lambda_i > \overline{\lambda}(x)$ the right-hand side expression of the above equation is for any fixed v concave and we can maximize over w_i finding

$$\begin{aligned} &\max_{w_i} \left\{ F_i(x, 0) + \tfrac{\partial}{\partial w} F_i(x, 0) w_i + \tfrac{1}{2} w_i^T \left(\tfrac{\partial^2}{\partial w^2} F_i(x, v) - 2\lambda_i \mathbf{1} \right) w_i + \lambda_i \right\} \\ &= F_i(x, 0) + \tfrac{1}{2} \tfrac{\partial F_i(x,0)}{\partial w} \left(2\lambda_i \mathbf{1} - \tfrac{\partial^2 F_i(x,v)}{\partial w^2} \right)^{-1} \tfrac{\partial F_i(x,0)}{\partial w}^T + \lambda_i . \end{aligned}$$

In the next step, we maximize over all v by using Assumption 1 in order to obtain for all $x \in \mathbb{R}^{n_x}$ and all $\lambda_i > \overline{\lambda}(x)$ the estimate

$$\begin{aligned} d_i(x, \lambda_i) &= \max_{w_i} \; G_i(x, \lambda_i, w_i) \\ &\leq F_i(x, 0) + \frac{1}{4} \frac{1}{\left(\lambda_i - \overline{\lambda}(x)\right)} \left\| \frac{\partial F_i(x, 0)}{\partial w} \right\|_2^2 + \lambda_i . \end{aligned} \tag{14}$$

Now, it follows that

$$\begin{aligned} M_i(x) &= \inf_{\lambda_i > \overline{\lambda}_i(x)} \; d_i(x, \lambda_i) \\ &\overset{(14)}{\leq} \inf_{\lambda_i > \overline{\lambda}_i(x)} \left\{ F_i(x, 0) + \frac{1}{4} \frac{1}{\left(\lambda_i - \overline{\lambda}(x)\right)} \left\| \frac{\partial F_i(x, 0)}{\partial w} \right\|_2^2 + \lambda_i \right\} \\ &= F_i(x, 0) + \left\| \frac{\partial F_i(x, 0)}{\partial w} \right\|_2 + \overline{\lambda}_i(x) \\ &= \Lambda_i(x). \end{aligned} \tag{15}$$

As the above consideration holds for all $x \in \mathbb{R}^{n_x}$ it follows with (11) that we have

$$\forall x \in \mathbb{R}^{n_x} : \quad V_i(x) \leq M_i(x) \leq \Lambda_i(x), \tag{16}$$

which is the statement of the Lemma. □

Note that for the case that F_i is already concave, we have $M_i = V_i$, i.e, there is no conservatism introduced. However, we might even in the non-concave case encounter the situation that the conservative approximation function M_i is exact. This happens for example in the non-concave quadratic case. As every function F_i is locally quadratic, this is an important observation:

Lemma 2 *Let the function F_i be a quadratic form in w such that*

$$F_i(x, w) = \frac{1}{2} w^T Q(x) w + q(x)^T w + r(x), \tag{17}$$

where $Q(x)$ is a symmetric matrix. If we use the eigenvalue bound

$$\overline{\lambda}_i(x) := \max \{0, \lambda_{\max}(Q(x))\},$$

then the approximation function M_i is exact, i.e., we have

$$\forall x \in \mathbb{R}^{n_x} : \quad V_i(x) = M_i(x). \tag{18}$$

Proof We start with an analysis of the following quadratically constrained quadratic program (QCQP):

$$\max_w \left\{ \frac{1}{2} w^T Q(x) w + q(x)^T w + r(x) \right\} \quad \text{s.t.} \quad w^T w \le 1. \tag{19}$$

If (w^*, λ^*) is a primal dual solution of the above maximization problem then (v^*, λ^*) is a primal dual solution of the problem

$$\max_v \left\{ \frac{1}{2} v^T D(x) v + p(x)^T v + r(x) \right\} \quad \text{s.t.} \quad v^T v \le 1, \tag{20}$$

where $p(x) := T(x) q(x)$ and $v^* := T(x) w^*$. Here, $T(x)$ denotes an orthonormal matrix and

$$D(x) := T(x)^T Q(x) T(x)$$

is a diagonal matrix which consists of the eigenvalues of $Q(x)$. Assume that the mth component of v^* satisfies the inequality $p_m(x)\, v_m^* < 0$. Then we can modify the vector v^* by exchanging the component v_m^* with $-v_m^*$ obtaining a contradiction to the assumption that v^* is a maximum of the problem (20). Thus, we have $p_m(x)\, v_m^* \ge 0$ for all components $m \in \{1, \ldots, n_w\}$ and consequently

$$\left(D_{mm}(x) - \lambda^*\right) v_m^* = -p_m(x) \Rightarrow \left(D_{mm}(x) - \lambda^*\right) (v_m^*)^2 \le 0. \tag{21}$$

Here, it should be noted that the LICQ condition for the problem (20) is always satisfied such that the above first order necessary conditions may indeed be applied. Now, the inequality (21) shows already that we have either $D_{mm}(x) \le \lambda^*$ or $v_m^* = 0$. Let us assume that we have a component m for which $v_m^* = 0$ and $D_{mm}(x) > \lambda^*$ holds. As, the LICQ condition implies not only the first order but also the second order necessary condition

$$\forall\, y \in \{z \mid z^T v^* = 0\} : \quad y^T \left(D(x) - \lambda^* I\right) y \le 0 \tag{22}$$

this case can directly be excluded. Thus, we have $D_{mm}(x) \leq \lambda^*$ for all $m \in \{1, \ldots, n_w\}$, i.e., the considered quadratic form is either concave or we have $\lambda^* \geq \overline{\lambda}_i(x)$ which implies the statement of the Lemma. □

Remark 1 The above Lemma is in a different version known in the context of trust region methods for exact Hessian SQP methods [14] as well as in the field of robust control [63].

Corollary 1 *Let $Q(x)$, $q(x)$ and $r(x)$ be given such that the associated QCQP (19) has a primal dual solution (w^*, λ^*). We assume that for all x either $Q(x)$ is negative definite or λ^* satisfies the regularity condition*

$$\lambda^* > \max \; \{0, \; \lambda_{\max}(Q(x))\} \, .$$

Moreover, let $g : \mathbb{R}^{n_x} \times \mathbb{R}^{n_w} \to \mathbb{R}$ be a twice continuously differentiable function. If the function F_i can be written as

$$F_i(x, w) = \frac{1}{2} w^T Q(x) w + q(x)^T w + r(x) + \epsilon \, g(x, w) \tag{23}$$

with $\overline{\lambda}_i(x) := \max \left\{ 0 \, , \; \max_{\|v\| \leq 1} \; \lambda_{\max}\left(\frac{\partial^2}{\partial w^2} F_i(x, v) \right) \right\}$ and $\epsilon > 0$ sufficiently small, then the approximation function M_i is still exact.

Proof The statement of the corollary follows immediately from Lemma 2 combined with the regularity of the solution under small data perturbations. However, in this argumentation we use the assumption that ϵ is sufficiently small. □

Summarizing our results so far, the functions M_i can be expected to yield a better conservative approximation than the linear approximations Λ_i. In particular, if F_i is quadratic in w or almost quadratic in the sense of Corollary 1, then the approximation function M_i is even exact. Thus, we are interested in solving an approximate robust counterpart problem of the form

$$\begin{aligned} \min_{x \in \mathbb{R}^{n_x}} \quad & M_0(x) \\ \text{subject to} \quad & M_i(x) \leq 0 \quad \text{for all} \quad i \in \{1, \ldots, n\}. \end{aligned} \tag{24}$$

However, solving the above problem with a standard NLP solver can not be recommended as each evaluation of the functions M_i is expensive and requires to solve a min–max problem. Moreover, the functions M_i are in general not differentiable. Rather, we plan to develop an algorithm to solve the problem (24) by taking the min–max structure explicitly into account. For this aim, we write the functions M_i for all $i \in \{0, \ldots, n\}$ in the form

$$M_i(x) = \max_{w_i} \;\; H_i(x, w_i) \quad \text{s.t.} \quad \|w_i\|_2^2 \leq 1, \tag{25}$$

where the functions $H_i : \mathbb{R}^{n_x} \times R^{n_w}$ are defined as

$$H_i(x, w_i) := G_i(x, \overline{\lambda}_i(x), w_i). \tag{26}$$

Note that Eq. (25) is equivalent to the definition (11) of M_i. In order to prove this, we recall that the function H_i is concave in w_i such that we have

$$\begin{aligned} M_i(x) &= \min_{\kappa_i \geq 0} \max_{w_i} \; H_i(x, w_i) - \kappa_i w_i^T w_i + \kappa_i \\ &= \min_{\kappa_i \geq 0} \max_{w_i} \; F_i(x, w_i) - \overline{\lambda}_i(x) w_i^T w_i + \overline{\lambda}_i(x) - \kappa_i w_i^T w_i + \kappa_i \\ &= \min_{\lambda_i \geq \overline{\lambda}_i(x)} \max_{w_i} \; F_i(x, w_i) - \lambda_i w_i^T w_i + \lambda_i \\ &= \min_{\lambda_i \geq \overline{\lambda}_i(x)} \; d_i(x, \lambda_i). \end{aligned}$$

Here, we denote with $\kappa_i > 0$ the multiplier of problem (25). This multiplier satisfies the equation $\overline{\lambda}_i(x) + \kappa_i = \lambda_i$.

The above consideration works with a quite natural set-up as uncertainty sets are often ellipsoidal [7] and can thus be rescaled as a unit ball. As we have discussed in this section, the associated Lagrangian overestimate functions M_i are for this case sometimes even exact. Another practically relevant case is when the uncertainty set is a box. As this case has been discussed in the literature within the context of the α-BB method we refer to [22], where concave overestimates for box constrained uncertainties are discussed. Note that in [22] the convexification method is combined with a branch-and-bound strategy and applied in the context of generalized semi-infinite programming for the case $n_w = 1$, i.e., for the case that the box is a one dimensional interval.

3 Optimality conditions for generalized robust counterpart problems

In this section we are interested in both necessary and sufficient optimality conditions for local minimizers of min–max optimization problems of the form

$$\begin{aligned} &\min_{x \in \mathbb{R}^{n_x}} \quad \max_{w_0 \in \mathcal{B}(x)} \quad H_0(x, w_0) \\ &\text{subject to} \quad \max_{w_i \in \mathcal{B}(x)} \quad H_i(x, w_i) \leq 0 \quad \text{for all} \quad i \in \{1, \ldots, n\}. \end{aligned} \tag{27}$$

This problem has the same form as the generalized robust counterpart problem (3), but we have switched notation to make entirely clear that we will from now on work with the following assumptions:

Assumption 2 We assume that the functions $H_0, \ldots, H_n$ are not only twice continuously differentiable but also (for all $x \in \mathbb{R}^{n_x}$) concave in w.

Assumption 3 Similarly, we assume from now on that the set $\mathcal{B}(x)$ is not only (for all $x \in \mathbb{R}^{n_x}$) compact but also convex. Moreover, we assume that we can write the set $\mathcal{B}(x)$ in the form

$$\mathcal{B}(x) := \left\{ w \in \mathbb{R}^{n_w} \mid B(x, w) \leq 0 \right\},$$

where the function $B : \mathbb{R}^{n_x} \times \mathbb{R}^{n_w} \to \mathbb{R}^{n_B}$ is twice continuously differentiable and (for all $x \in \mathbb{R}^{n_x}$) component-wise convex in w.

Recall from the last section that such a convex set $\mathcal{B}(x)$ might in a conservative approximation setting be obtained by taking the convex hull of $W(x)$ while the functions $H_0, \ldots, H_n$ are concave over-estimators of the original functions $F_0, \ldots, F_n$. However, note that Lemma 2 does only apply to the case of special nonconvex min–max problems with ball constraints on the uncertainty. The above assumptions are on the one hand more restrictive as convexity is required but on the other hand less restrictive as they include a more general class of uncertainty sets. Nevertheless, our main motivation are the examples

$$B_{\text{ball}}(x, w) = \|w\|_2^2 - 1 \qquad \text{and} \qquad B_{\text{box}}(x, w) = \begin{pmatrix} w - u(x) \\ l(x) - w \end{pmatrix} \tag{28}$$

from the previous section. Here, $l, u : \mathbb{R}^{n_x} \to \mathbb{R}^{n_w}$ denote twice continously differentiable functions representing the upper and lower bounds of a parametric uncertainty box.

Definition 2 A point (x^*, w^*) is said to be a local min–max point if the components of the variable $w^* := \left(w_0^*, \ldots, w_n^*\right)$ are global maximizers of the functions

$$H_0(x^*, \cdot), \ldots, H_n(x^*, \cdot)$$

while x^* is a local minimizer of the problem (27).

Assuming that the lower level maximizers in (27) are KKT-points the Assumptions 2 and 3 enable us to equivalently replace the condition "$w \in \mathcal{B}(x^*)$ maximizes $H(x^*, w)$" (with x^* being a local minimizer of (27)) by the first order KKT conditions of the form

$$\begin{aligned} 0 &= \nabla_w L_j(x^*, w_j^*, \lambda_j^*) \\ 0 &\geq B(x^*, w_j^*) \\ 0 &\leq \lambda_j^* \\ 0 &= \sum_{k=0}^{n_B} \lambda_k^T B_k(x^*, w_k^*) \end{aligned} \tag{29}$$

for all $j \in \{0, \ldots, n\}$. Here, we have used the notation

$$L_j(x, w, \lambda) := H_j(x, w) - \lambda^T B(x, w)$$

to denote the Lagrangian $L_j : \mathbb{R}^{n_x} \times \mathbb{R}^{n_w} \times \mathbb{R}^{n_B} \to \mathbb{R}$ which is associated with the jth lower level concave maximization problem.

We make the assumption that at least the Mangasarian-Fromovitz constraint qualification (MFCQ) for the lower level maximization problems holds such that the existence of the multipliers λ_j^* can be guaranteed. In this case the KKT conditions (29) are both necessary and sufficient to guarantee that w^* denotes the maximizers of the concave lower level problems. Under the stronger linear independence constraint qualification (LICQ) λ^* is also unique. Following the classical framework [56,57], we introduce two other assumptions on the maximizers w_j^* of the lower level problems: first we assume that the strict complementarity condition (SCC) is satisfied, i.e., we assume ($i \in \{0, \ldots, n\}$)

$$B(x^*, w_i^*) - \lambda_i^* < 0 \tag{30}$$

at the local min–max point (x^*, w^*) of our interest. And second, we assume that the second order sufficient condition (SOSC)

$$\forall p_i \in \mathcal{T}_i \backslash \{0\} : \quad p_i^T \left(\frac{\partial^2}{\partial w_i^2} H_i(x^*, w^*) - 2\lambda_i^* \right) p_i < 0 \tag{31}$$

is satisfied, where the set $\mathcal{T}_i$ is defined as

$$\mathcal{T}_i := \left\{ p \in R^{n_w} \mid \frac{\partial}{\partial w} B^{i,\mathrm{act}}(x^*, w^*) p = 0 \right\}, \tag{32}$$

where $B^{i,\mathrm{act}}$ denotes the constraint components of the function B which are active for the ith lower level maximization problem.

Now, we use the language from the semi-infinite programming literature:

Definition 3 A point w^* is nondegenerate if it satisfies the LICQ, SCC, and SOSC condition for all lower level maximization problems in (27).

The corresponding assumption that a point w^* is nondegenerate is in the context of generalized semi-infinite programming (GSIP) also known under the name reduction ansatz [57,24]. It can be used to guarantee that the primal and dual solution $\hat{w}_j(x)$ and $\hat{\lambda}_j(x)$ of the jth parameterized lower level problems of the form

$$\min_{w_j \in \mathcal{B}(x)} H_j(x, w_j) \tag{33}$$

can be regarded as differentiable functions in x. In fact, if $w_j^* = \hat{w}_j(x^*)$ is a non-degenerate maximizer, the functions $\hat{w}_j$ and $\hat{\lambda}_j$ exist in an open neighborhood $D_x \subset R^{n_x}$ of x^* and are differentiable in this neighborhood D_x—this is a well-known result [47,48] which follows immediately from the implicit function theorem.

Definition 4 We say that a point (x, w, λ) satisfies the extended Mangasarian Fromovitz constraint qualification (EMFCQ) if there exists a vector $\xi \in \mathbb{R}^{n_x}$ with

$$\frac{\partial}{\partial x} L_i(x, w, \lambda)\, \xi < 0 \qquad \forall i \in \mathcal{A}. \tag{34}$$

Here, $\mathcal{A} := \{k \mid H(x, w) = 0\}$ denotes the active set of the higher level minimization problem. Moreover, we say that (x, w, λ) satisfies the extended linear independence constraint qualification (ELICQ) if the vectors

$$\frac{\partial}{\partial x} L_i(x, w, \lambda) \qquad \forall i \in \mathcal{A} \tag{35}$$

are linearly independent.

The result of the following theorem has in a more general form (without even using any reduction ansatz) for the first time been proven in [30], where first order optimality conditions for generalized semi-infinite programming problems are discussed. In this paper, we summarize this result being on the one hand less general, as we require the reduction ansatz, but on the other hand we can give a shorter proof:

Theorem 1 *Let (x^*, w^*, λ^*) be a local min–max solution of the problem (27) with w^* being a nondegenerate maximizer of the lower level concave maximization problems at x^* and λ^* the associated dual solution. Now, the following statements hold:*

1. *If (x^*, w^*, λ^*) satisfies the EMFCQ condition, then there exists a multiplier $\chi^* \in \mathbb{R}^n$ such that the KKT-type conditions*

$$\begin{array}{ll} 0 = \frac{\partial}{\partial x} K(x^*, \chi^*, w^*, \lambda^*) & 0 = \frac{\partial}{\partial w} L_j(x^*, w_j^*, \lambda_j^*) \\ 0 \geq L_i(x^*, w_i^*, \lambda_i^*) & 0 \geq B(x^*, w_j^*) \\ 0 \geq \chi_i^* & 0 \leq \lambda_j^* \\ 0 = \sum_{k=1}^{n} \chi_k^* L_k(x^*, w_k^*, \lambda_k^*) & 0 = \sum_{k=0}^{n} \lambda_k^{*T} B(x^*, w_k^*) \end{array} \tag{36}$$

   are satisfied for all $i \in \{1, \ldots, n\}$ and all $j \in \{0, \ldots, n\}$. Here, we use the notation

$$K(x, \chi, w^*, \lambda^*) := L_0(x, w_0^*, \lambda_0^*) - \sum_{k=1}^{n} \chi_k L_k(x, w_k^*, \lambda_k^*). \tag{37}$$

2. *If (x^*, w^*, λ^*) satisfies also the ELICQ condition, then the multiplier χ in the necessary conditions (36) is unique.*

Proof Due to the complementarity relation for the lower level maximization problems we have

$$\forall x \in D_x : \quad H_j(x, \hat{w}_j(x)) = L_j(x, \hat{w}_j(x), \hat{\lambda}_j(x)),$$

where $\hat{w}_j$ and $\hat{\lambda}_j$ denote the parameterized primal-dual solution of the lower level maximization problems as a function of $x \in D_x$ as introduced above. Thus, the min–max problem (27) is locally equivalent to the following auxiliary problem

$$\min_{x \in D_x} \quad L_0(x, \hat{w}_0(x), \hat{\lambda}_0(x)) \quad \text{s.t.} \quad L_i(x, \hat{w}_i(x), \hat{\lambda}_i(x)) \leq 0 \quad \text{for all} \quad i \in \{1, \ldots, n\}. \tag{38}$$

Using the optimality and feasibility condition for the lower level maximizer $\hat{w}_j(x^*)$ we find

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}L_j(x^*,\hat{w}_j(x^*),\hat{\lambda}_j(x^*)) &= \frac{\partial}{\partial x}L_j(x^*,w_j^*,\lambda_j^*) \\ &\quad + \frac{\partial}{\partial w}L_j(x^*,w_j^*,\lambda_j^*)\frac{\partial \hat{w}_j(x^*)}{\partial x} - B^{j,\mathrm{act}}(x^*,w_j^*)\frac{\partial \hat{\lambda}_j(x^*)}{\partial x} \\ &= \frac{\partial}{\partial x}L_j(x^*,w_j^*,\lambda_j^*). \end{aligned} \tag{39}$$

for all $j \in \{0,\ldots,n\}$. Thus, the EMFCQ (or ELICQ) condition from Definition 4 boils down to the MFCQ (or LICQ) condition for the auxiliary problem (38). The statements of the Theorem are now equivalent to the standard KKT theorem for the problem (38) under the MFCQ and LICQ condition respectively. □

Remark 2 The proof for the above theorem can be modified in the sense that the optimization problem (38) can also be considered without any constraint qualification. In this case, we can only consider Fritz John optimality conditions for the auxiliary problem (38), as discussed in [57].

Remark 3 The above proof can be generalized for the case that the lower level problems comprise not only convex inequalities but also linear equalities. Furthermore, we could consider the case that the problem (27) has additional equality and/or inequality constraints which only depend on x etc.. Please note that such generalizations are straightforward and omitted here for the ease of notation.

In order to complete our review of optimality conditions for min–max problems, we note that it is also possible to write the KKT conditions for the lower level maximization problems into the constraints of the higher-level problem, considering a mathematical program with complementarity constraints (MPCC) of the form

$$\begin{aligned} \underset{x,w,\lambda}{\text{minimize}} \quad & H_0(x,w_0) \\ \text{subject to} \quad & 0 \geq H_i(x,w_i) \\ & 0 = \nabla_w L_j(x,w_j,\lambda_j) \\ & 0 \geq B(x,w_j) \\ & 0 \leq \lambda_j \\ & 0 = \sum_{k=0}^{n} \lambda_k^T B_k(x,w_k) \end{aligned} \tag{40}$$

for all $i \in \{1,\ldots,n\}$ and all $j \in \{0,\ldots,n\}$. Note that this MPCC formulation is known and discussed in the literature [57].

Unfortunately, it is without further precaution not trivial to discuss KKT points of an MPCC. In order to understand the problem, we note that the Mangasarian Fromovitz constraint qualification (MFCQ) for the minimization problem (40) is violated at all feasible points of an MPCC. This can directly be seen by looking at the complementarity conditions but we also refer to [50] for a discussion of the details of this statement.

As the LICQ condition implies the MFCQ condition, both constraint qualifications are rendered useless for mathematical programs with complementarity constraints.

The degeneracy of the MPCC (40) seems to be the main motivation for the development of smoothing techniques for numerical approaches. In [57] and also in [18] such smoothing techniques for MPCCs have been discussed. In this paper, we will discuss a very different and new approach to numerically deal with min–max problems. While in [57] the MPCC (40) was the starting point for the development of numerical algorithms that find Fritz John points based on smoothing techniques, we are in the following section interested in numerical algorithms which use the necessary KKT-type conditions (36) directly as a starting point.

4 Sequential convex bilevel programming

The aim of this section is to develop an algorithm which globally converges to local minimizers of the problem (27). The question of how such an algorithm should be designed depends heavily on the functions H_i and B. For example the dimensions of these functions, the costs for an evaluation as well as the cost of computing derivatives will mainly influence our choice of numerical techniques. If the function evaluation is cheap while the difficulty is in determining the active sets, an application of interior point techniques might come to our mind. However, in this paper, we are interested in the opposite situation, i.e., in the case that the evaluation of the functions and their derivatives is the most expensive part. Recall that for standard nonlinear programs SQP methods have turned out to perform very well in such situations.

In the following, we will constrain ourself to the semi-infinite case, i.e., we assume that the function B is independent of x. The aim of the algorithm is to find a point $z^* := (x^*, \chi^*, w^*, \lambda^*)$ which satisfies the necessary KKT-type conditions (36) with x^* being a minimizer of the problem (27). In order to apply the idea of SQP methods to our situation, we assume that we have an initial guess z^0 for the point z^* and plan to perform iterates of the form

$$z^+ = z + \alpha\, \Delta z := (x + \alpha \Delta x, \chi + \alpha \Delta \chi, w + \alpha \Delta w, \lambda + \alpha\, \Delta\lambda)$$

with $\alpha \in (0, 1]$ being a damping parameter while the steps Δx, $\Delta \chi$, and Δw and $\Delta\lambda$ are assumed to be the primal dual local min–max point of the following convex bilevel quadratic program (min–max QCQP):

$$\begin{aligned} \min_{\Delta x} \max_{\Delta w_0 \in \mathcal{B}_0^{\text{lin}}} & \left\{ H^0 + L_x^0 \Delta x + \left(\frac{\Delta w_0^T L_{ww}^0}{2} + \Delta x^T L_{xw}^0 + H_w^0 \right) \Delta w_0 + \frac{\Delta x^T K_{xx} \Delta x}{2} \right\} \\ \text{s.t.} \max_{\Delta w_i \in \mathcal{B}_i^{\text{lin}}} & \left\{ H^i + L_x^i \Delta x + \left(\frac{\Delta w_i^T L_{ww}^i}{2} + \Delta x^T L_{xw}^i + H_w^i \right) \Delta w_i \right\} \leq 0 \end{aligned} \tag{41}$$

with $i \in \{1, \ldots, n\}$ and

$$\mathcal{B}_j^{\text{lin}} := \left\{ \Delta w_j \;\mid\; B_w^i \Delta w_i + B^i \leq 0 \right\} \tag{42}$$

for all $j \in \{0, \ldots, n\}$. Here, it should be explained that we use the notation $\Delta\lambda_j := \lambda_j^{\dagger} - \lambda_j$ to denote the steps to be taken in the multipliers of the lower level maximization problems, while $\Delta\chi := \chi^{\dagger} - \chi$ depends on the dual solution $\chi^{\dagger}$ which is associated with the inequality constraints in the minimization problem (41). Moreover, we use the following short hands:

$$
\begin{array}{lll}
L_{ww}^j := \frac{\partial^2}{\partial w^2} L_j(x, w_j, \lambda_j), & L_{wx}^j := \frac{\partial^2}{\partial w \partial x} L_j(x, w_j, \lambda_j), & L_{xw}^j := \left(L_{wx}^j\right)^T, \\
L_w^j := \frac{\partial}{\partial w} L_j(x, w_j, \lambda_j), & L_x^j := \frac{\partial}{\partial x} L_j(x, w_j, \lambda_j), & L^j := L_j(x, w_j, \lambda_j), \\
H_w^j := \frac{\partial}{\partial w} H_j(x, w_j), & H_x^j := \frac{\partial}{\partial x} H_j(x, w_j), & H^j := H_j(x, w_j), \\
B_w^j := \frac{\partial}{\partial w} B(x, w_j), & & B^j := B(x, w_j).
\end{array}
$$

At this point we have to remark that the iteration index is suppressed for ease of notation, i.e., once a step has been performed we set the variable z to z^+ in order to continue with the next step. In particular, the symmetric and positive definite matrix

$$K_{xx} \in \mathbb{R}^{n_x \times n_x}$$

may change from iteration to iteration although this is in our notation not indicated by an iteration index. Possible choices of this matrix K_{xx} will be discussed later, but we mention already at this point that K_{xx} should be a suitable approximation of the Hessian matrix

$$L_{xx}^0 - \sum_{k=1}^{n} \chi_k L_{xx}^k,$$

where we use the short hand $L_{xx}^j := \frac{\partial^2}{\partial x^2} L_j(x, w_j, \lambda_j)$.

Note that the sub-maximization problems within the min–max problem (41) can be regarded as concave quadratic programs (QPs) of the form

$$
\begin{aligned}
V_i(\Delta x) := \max_{\Delta w_i} & \left\{ \frac{1}{2} \Delta w_i^T L_{ww}^i \Delta w_i + \left(\Delta x^T L_{xw}^i + H_w^i \right) \Delta w_i \right\} \\
\text{s.t. } & B_w^i \Delta w_i + B^i \leq 0,
\end{aligned} \tag{43}
$$

as L_{ww}^i is assumed to be negative semi-definite (cf. Assumption 2 and 3). Moreover, the upper level minimization problem takes the form

$$
\begin{aligned}
\min_{\Delta x} & \left\{ H^0 + L_x^0 \Delta x + V_0(\Delta x) + \frac{1}{2} \Delta x^T K_{xx} \Delta x \right\} \\
\text{s.t. } & H^i + L_x^i \Delta x + V_i(\Delta x) \leq 0,
\end{aligned} \tag{44}
$$

which is a strictly convex optimization problem if K_{xx} is positive definite. Here, we have used the fact that the functions V_j are convex in Δx as the maximum over linear functions is convex. As for SQP methods, the existence of Δz is not guaranteed as the sub-problems might be infeasible. However, assuming that the sub-problems are feasible and that the concave quadratic programs (43) have unique solutions, we have a guarantee that the step Δz is unique. Moreover, the convexity has the practical advantage that the sub-problem can efficiently be solved with existing convex optimization tools.

In the case that L^i_{ww} is strictly negative definite, we can analyze the dual minimization problem which is associated with the concave maximization problem (43). Provided that the QPs (43) admit strictly feasible points (Slater's condition) the problem (44) is equivalent to a convex QCQP of the form

$$\begin{aligned} \min_{\Delta x, \lambda^\dagger} \; & \left\{ H^0 + L^0_x \Delta x - \tfrac{1}{2} g_0(\Delta x, \lambda^\dagger) \left(L^0_{ww}\right)^{-1} g_0(\Delta x, \lambda^\dagger)^T - B_0^T \lambda^\dagger + \tfrac{1}{2} \Delta x^T K_{xx} \Delta x \right\} \\ \text{s.t. } & H^i + L^i_x \Delta x - \tfrac{1}{2} g_i(\Delta x, \lambda^\dagger) \left(L^i_{ww}\right)^{-1} g_i(\Delta x, \lambda^\dagger)^T - B_i^T \lambda^\dagger \leq 0. \end{aligned}$$

where we have used the short hand

$$g_j(\Delta x, \lambda^\dagger) := \frac{1}{2} \left(\Delta x^T L^j_{xw} - \left(\lambda^\dagger_j\right)^T B^j_w + H^j_w \right).$$

Note that this problem can solved with any suitable convex QCQP solver.

Definition 5 We define for each $j \in \{0, \ldots, n\}$ the lower level working set $A_j(\lambda^\dagger)$ by

$$\mathcal{A}_j(\lambda^\dagger) \; := \; \left\{ k \in \{0, \ldots, n\} \mid \left(\lambda^\dagger_j\right)_k > 0 \right\}. \tag{45}$$

Moreover, we denote the number of elements in $\mathcal{A}_j(\lambda^\dagger)$ by $m_j := \left|\mathcal{A}_j(\lambda^\dagger)\right|$.

We use the above notation to introduce the lower level KKT matrices

$$\Omega_j \; := \; \begin{pmatrix} L^j_{ww} & \left(B^{j,\text{act}}_w\right)^T \\ B^{j,\text{act}}_w & 0 \end{pmatrix}, \tag{46}$$

where $B^{j,\text{act}}_w \in \mathbb{R}^{m_j \times n_w}$ is a matrix which consists of the rows of B^j_w, whose index is in the working set $\mathcal{A}_j(\lambda^\dagger)$.

Assumption 4 We assume that the matrix Ω_j is invertible for all $j \in \{0, \ldots, n\}$.

Note that the above assumption seems reasonable in our context as we are interested in the case that the lower level optimization problems are convex while a non-degeneracy assumption (or reduction ansatz) holds in the optimal solution. In this sense, the above assumption is not excessively restrictive requiring a kind of regularity condition to be satisfied during the iterations.

Proposition 1 *If the Assumption 4 holds, the bilevel optimization problem* (41) *can equivalently be regarded as an MPCC, i.e., the condition that the pairs* $(\Delta w_j, \lambda^\dagger_j)$ *are* *primal-dual maximizers can for all $j \in \{0, \dots, n\}$ equivalently be replaced by the corresponding KKT conditions*

$$0 = \Delta x^T L^j_{xw} + \Delta w_j^T L^j_{ww} - \Delta\lambda_j^T B^j_w + L^j_w \tag{47}$$

$$0 \geq B^j_w \Delta w_j + B^j \tag{48}$$

$$0 \leq \lambda_j + \Delta\lambda_j = \lambda_j^\dagger \tag{49}$$

$$0 = \left(B^j_w \Delta w_j + B^j\right)^T \lambda_j^\dagger \tag{50}$$

using the notation $L^j_w := H^j_w - \lambda_j^T B^j_w$.

Proof The above Proposition should be self-explaining: the conditions (47)–(50) are simply the necessary KKT optimality conditions for the lower-level QPs (43). Here, Assumption 4 guarantees that the linear independence constraint qualification is satisfied justifying an application of the KKT theorem. □

Remark 4 The above Proposition shows that the bilevel optimization problem (41) can be regarded as a mathematical program with linear complementarity constraints (MPLCC), which are in their general form rather expensive and difficult to solve [12, 31]. Note that the special structure arising from the semi-infinite programming context as well as the convexity of the bilevel problem (41) are the foundation of the presented sequential convex bilevel programming method, which make it efficient. This aspect is also the main difference of the presented method in comparison to techniques like piecewise sequential quadratic programming methods for general MPCCs [32,46,64], where a quadratic program with linear complementarity constraints (QPLCC) must be solved in each step of the sequential method.

In the next step we work out the optimality conditions for the bilevel QP (41). For this aim, we introduce the matrices $R_j \in \mathbb{R}^{n_x \times (n_w + m_j)}$ as well as the vectors $s_j \in \mathbb{R}^{n_w + m_j}$ (with $j \in \{0, \dots, n\}$) which are defined as

$$R_j := \begin{pmatrix} L^j_{w,x} \\ 0 \end{pmatrix} \quad \text{and} \quad s_j := \begin{pmatrix} \left(H^j_w\right)^T \\ B^{j,\text{act}} \end{pmatrix}, \tag{51}$$

respectively. Here, the matrix $B^{j,\text{act}}$ consists of all components of B^j, whose index is in the working set $\mathcal{A}_j(\lambda^\dagger)$. Moreover, we use the notation $T_j := R_j^T \Omega_j^{-1} R_j$.

Definition 6 Requiring that Assumption 4 is satisfied, we say that the QP (43) is nondegenerate for a given Δx if the strict complementarity condition (SCC)

$$B^j_w \Delta w_j + B^j - \lambda_j^\dagger < 0. \tag{52}$$

holds at the primal dual solution $(\Delta w_j, \lambda_j^\dagger)$ of the QP (43).

Assumption 5 We assume that all lower level QPs of the form (43) are non-degenerate at the solution $(\Delta x, \Delta w, \lambda^\dagger)$ of the problem (41), i.e., the strict inequality (52) is for all indices $j \in \{0, \dots, n\}$ satisfied at this point.

Note that the non-degeneracy of the jth lower level QP at a given Δx implies that the variables Δw_j and $\lambda_j^\dagger$ can in a neighborhood of Δx be regarded as a locally linear function. This is due to the fact that Assumption 4 is equivalent to the LICQ and SOSC condition for the lower level QPs while the SCC condition is required by the above definition.

Definition 7 Let the point $(\Delta x, \Delta w, \lambda^\dagger)$ be a feasible point of the bilevel problem (41). Providing that Assumption 4 is satisfied, we say that the extended LICQ condition is satisfied at $(\Delta x, \Delta w, \lambda^\dagger)$ if the vectors

$$L_x^k - s_k^T \Omega_k^{-1} R_k + \Delta x^T T_k^T \qquad \forall k \in \mathcal{W} \tag{53}$$

are linearly independent. Here,

$$\mathcal{W} := \left\{ k \mid \left(L_x^i \Delta x + L_w^i \Delta w_i - \Delta \lambda_i^T B^i + L^i \right)_k = 0 \right\}$$

denotes the active set which is associated with the upper level constraints.

Lemma 3 *Let the Assumptions 4 and 5 be satisfied. Furthermore, let* $(\Delta x, \Delta w, \lambda^\dagger)$ *be a minimizer of problem* (41) *for which the extended LICQ-condition holds. Now, we have necessarily*

$$0 = \left(K_{xx} - T_0 + \sum_{k=1}^{n} \chi_k^\dagger T_k \right) \Delta x + \left(L_x^0 - s_0^T \Omega_0^{-1} R_0 \right)^T - \sum_{k=1}^{n} \chi_k^\dagger \left(L_x^k - s_k^T \Omega_k^{-1} R_k \right)^T \tag{54}$$

$$0 \geq H^i + L_x^i \Delta x - \frac{1}{2} \left(R_i \Delta x + s_i \right)^T \Omega_i^{-1} \left(R_i \Delta x + s_i \right) \tag{55}$$

$$0 \geq \chi + \Delta \chi \ := \ \chi^\dagger \tag{56}$$

$$0 = \left(H^i + L_x^i \Delta x - \frac{1}{2} \left(R_i \Delta x + s_i \right)^T \Omega_i^{-1} \left(R_i \Delta x + s_i \right) \right) \chi_i^\dagger \tag{57}$$

for all $i \in \{1, \ldots, n\}$. *Here, the multiplier* $\chi^\dagger$ *is unique.*

Proof Due to the non-degeneracy Assumption 5 for the lower level QPs (43) the bilevel problem (41) is locally equivalent to an auxiliary quadratically constrained quadratic program of the form

$$\begin{aligned} \min_{\Delta x} \quad & \left\{ \tfrac{1}{2} \Delta x^T K_{xx} \Delta x + L_x^0 \Delta x - \tfrac{1}{2} \left(R_0 \Delta x + s_0 \right)^T \Omega_0^{-1} \left(R_0 \Delta x + s_0 \right) \right\} \\ \text{s.t.} \quad & H^i + L_x^i \Delta x - \tfrac{1}{2} \left(R_i \Delta x + s_i \right)^T \Omega_i^{-1} \left(R_i \Delta x + s_i \right) \leq 0 \end{aligned} . \tag{58}$$

This follows immediately from a local elimination of the variable Δw on dependence on Δx, i.e., we know that the active set of the lower level QPs remains locally constant in Δx such that we can exploit the relation

$$R_j \Delta x + \Omega_j \begin{pmatrix} \Delta w_j \\ -\lambda_j^{\dagger,\mathrm{act}} \end{pmatrix} + s_j = 0, \tag{59}$$

which summarizes the parameterized stationarity as well as the primal feasibility condition of the active constraints associated with the jth sub-QP (43). In this notation, $\lambda_j^{\dagger,\mathrm{act}}$ is the vector which consists of the non-zero components of $\lambda_j^{\dagger}$. Now, the extended LICQ condition for the bilevel problem (41) reduces to a standard LICQ condition for the auxiliary problem (58). Consequently, an application of the KKT theorem yields the statement of the Lemma. □

4.1 Global convergence analysis

A crucial point in the discussion of global convergence of any SQP type method is the availability of a merit function which measures the progress of the iterations $z^+ = z + \alpha \Delta z$ towards a local minimum. This can for example be achieved via line search techniques [39] adjusting the damping parameter α if necessary but also trust region methods [14] make use of merit functions. In standard SQP methods with suitable regularity assumptions Han's exact l_1-penalty function [23] is a traditional choice but there are also other choices [39].

Note that for general MPCCs it is not straightforward to apply the idea of penalty functions as most of the techniques, as, e.g., discussed in [39], are based on the assumption that a suitable constraint qualification holds. As MPCCs do often not satisfy these constraint qualifications, standard proof techniques typically fail. Global convergence of SQP methods for general MPCCs is an active field of research and we refer to [1,9] for further reading on global convergence of methods and a discussion of penalty functions for general MPCCs.

Fortunately, as the MPCC (40) arises from the context of semi-infinite programming it has a special structure which is exploited in the method presented in this paper and which helps us also to construct a suitable merit function for our needs. Let us start by defining an upper level merit function $\Phi_U : \mathbb{R}^{n_x} \times \mathbb{R}^{(n+1)n_w} \times \mathbb{R}^{(n+1)n_B} \to \mathbb{R}$ planning to measure the progress in terms of upper level feasibility and optimality in the form

$$\Phi_U(x, w, \lambda) := L_0(x, w_0, \lambda_0) + \sum_{k=1}^{n} \hat{\chi}_k \, \pi_k(L_k(x, w_k, \lambda_k)), \tag{60}$$

where $\hat{\chi} \in \mathbb{R}^n_{++}$ is a constant vector. Here, it should be explained that the positive projection $\pi : \mathbb{R}^d \to \mathbb{R}^d_+$ is defined for arbitrary dimensions d while the components of π satisfy

$$\forall s \in \mathbb{R}^d, \forall k \in \{1, \ldots, d\} : \quad \pi_k(s) := \max\{0, \; s_k\}.$$

Similarly, $|\cdot| : \mathbb{R}^d \to \mathbb{R}^d_+$ is also defined for arbitrary d where $|s|$ denotes the component-wise absolute value of a vector $s \in \mathbb{R}^d$.

Besides the upper-level feasibility, we also need to measure the violation of the stationarity and primal feasibility condition for the lower level optimization problems. In this context, we observe that the dual feasibility condition $\lambda^+ \geq 0$ is automatically satisfied for the iterates, since λ^+ satisfies by construction the optimality conditions of the lower level maximization problems in problem (41). Thus, a violation of dual feasibility in the lower level problems does not need to be detected motivating the introduction of primal lower level merit functions of the form $\Phi_L^j : \mathbb{R}^{n_x} \times \mathbb{R}^{n_w} \times \mathbb{R}^{n_B} \to \mathbb{R}$ which are defined as

$$\Phi_L^j(x, w_j, \lambda_j) := \left| \frac{\partial L_j(x, w_j, \lambda_j)}{\partial w} \right| \hat{\rho}_j + \hat{\lambda}_j^T \pi(B(x, w_j))$$

for all $j \in \{0, \ldots, n\}$. Here, $\hat{\rho}_j \in \mathbb{R}_{++}^{n_w}$ and $\hat{\lambda}_j \in \mathbb{R}_{++}^{n_B}$ are positive constants. The final step is to compose a merit function $\Phi : \mathbb{R}^{n_x} \times \mathbb{R}^{(n+1)n_w} \times \mathbb{R}^{(n+1)n_B} \to \mathbb{R}$ as

$$\Phi(x, w, \lambda) := \Phi_U(x, w, \lambda) + \Phi_L^0(x, w_0, \lambda_0) + \sum_{k=1}^{n} \hat{\chi}_k \Phi_L^k(x, w_k, \lambda_k). \tag{61}$$

In the following, we prepare the proof of Theorem 2 where a condition for a descent direction of the merit function Φ will be discussed. In this context we make use of the following assumption:

Assumption 6 The matrix L_{ww} is negative definite.

Let us introduce the short-hand "∂_α" to denote one sided directional derivatives in the step direction, i.e., we define for example

$$\partial_\alpha L_0(x, w_0, \lambda_0) := \lim_{\alpha \to 0^+} \frac{L_0(x+\alpha \Delta x, w_0+\alpha \Delta w_0, \lambda_0+\alpha \Delta \lambda_0) - L_0(x, w_0, \lambda_0)}{\alpha}. \tag{62}$$

This abstract notation for one sided derivatives can analogously be transferred for the other terms in the merit function. Let us summarize the following technical result:

Proposition 2 *Transferring the notation (62) to denote one-sided directional derivatives, the following expressions exist (i.e., the corresponding limits for $\alpha \to 0^+$ exist) and satisfy*

$$\partial_\alpha L_0(x, w_0, \lambda_0) = L_x^0 \Delta x + L_w^0 \Delta w_0 - \Delta \lambda_0^T B^0 \tag{63}$$

$$\partial_\alpha \pi(L_i(x, w_i, \lambda_i)) \leq -\pi(L^i) - \frac{1}{2} L_w^i \left(L_{ww}^i \right)^{-1} {L_w^i}^T \tag{64}$$

$$\partial_\alpha \pi(B(x, w_j)) \leq -\pi(B^j) \tag{65}$$

$$\partial_\alpha \left| \frac{\partial}{\partial w} L_j(x, w_j, \lambda_j) \right| = - \left| L_w^j \right| \tag{66}$$

for all $i \in \{1, \ldots, n\}$ and all $j \in \{0, \ldots, n\}$. Here, the formula (64) requires the Assumption 6 to be satisfied.

Proof The first formula (63) follows immediately from the definition (62). In order to derive the remaining formulas, we first recall that we have for any differentiable function $\varphi : \mathbb{R}^r \to \mathbb{R}$ with derivative function $\varphi' := \partial_\xi \varphi$ the following rules

$$\partial_\alpha \left|\varphi(\xi + \alpha \Delta\xi)\right| = \begin{cases} \varphi'(\xi)\Delta\xi & \text{if } \varphi(\xi) > 0 \\ \left|\varphi'(\xi)\Delta\xi\right| & \text{if } \varphi(\xi) = 0 \\ -\varphi'(\xi)\Delta\xi & \text{if } \varphi(\xi) < 0 \end{cases} \tag{67}$$

$$\text{and} \quad \partial_\alpha \pi\left(\varphi(\xi + \alpha \Delta\xi)\right) = \begin{cases} \varphi'(\xi)\Delta\xi & \text{if } \varphi(\xi) > 0 \\ \pi\left(\varphi'(\xi)\Delta\xi\right) & \text{if } \varphi(\xi) = 0 \\ 0 & \text{if } \varphi(\xi) < 0 \end{cases} \tag{68}$$

for all $\xi \in \mathbb{R}^r$, as long as $\varphi'(\xi)\Delta\xi \neq 0$ whenever $\varphi(\xi) = 0$. Moreover, the conditions from the lower level QP optimality

$$B_w^j \Delta w_j \leq -B^j, \tag{69}$$

$$\text{and} \quad \Delta x^T L_{xw}^j + \Delta w_j^T L_{ww}^j - \Delta\lambda_j^T B_w^j = -L_w^j \tag{70}$$

can be combined with Eqs. (67) and (68) to estimate the directional derivatives (65) and (66) respectively. It remains to verify the estimate (64). For this aim, we first compute for all $i \in \{1, \ldots, n\}$ the term

$$\begin{aligned}
\partial_\alpha L_i &= L_x^i \Delta x + L_w^i \Delta w_i - \Delta\lambda_i^T B^i \\
&\leq -H^i - \left(\frac{1}{2}\Delta w_i^T L_{ww}^i + \Delta x^T L_{xw}^i + H_w^i\right)\Delta w_i + L_w^i \Delta w_i - \Delta\lambda_i^T B^i \\
&= -L^i - \left(\frac{1}{2}\Delta w_i^T L_{ww}^i + \Delta x^T L_{xw}^i + H_w^i\right)\Delta w_i + L_w^i \Delta w_i - (\Delta\lambda_i^\dagger)^T B^i \\
&\overset{(50)}{=} -L^i - \left(\frac{1}{2}\Delta w_i^T L_{ww}^i + \Delta x^T L_{xw}^i + H_w^i - (\Delta\lambda_i^\dagger)^T B_w^i\right)\Delta w_i + L_w^i \Delta w_i \\
&\overset{(47)}{=} -L^i + \frac{1}{2}\Delta w_i^T L_{ww}^i \Delta w_i + L_w^i \Delta w_i \\
&= -L^i + \frac{1}{2}\left(L_{ww}^i \Delta w_i + (L_w^i)^T\right)^T \left(L_{ww}^i\right)^{-1} \left(L_{ww}^i \Delta w_i + (L_w^i)^T\right) \\
&\quad -\frac{1}{2} L_w \left(L_{ww}^i\right)^{-1} L_w^{i\,T} \\
&\leq -L^i - \frac{1}{2} L_w^i \left(L_{ww}^i\right)^{-1} L_w^{i\,T}.
\end{aligned} \tag{71}$$

In the last step, we have used that L_{ww}^i is negative definite. Estimate (64) is now a direct consequence, as we can use the above estimate in combination with Eq. (68). □

Definition 8 Provided Assumption 6 is satisfied, we introduce the notation

$$\rho_j := \left(L_{ww}^j\right)^{-1} L_w^{j\,T}$$

for all $j \in \{0, \ldots, n\}$.

Assumption 7 We assume that the matrix K_{xx} is symmetric and positive definite.

In the following Theorem we discuss that the presented sequential convex bilevel programming method generates descent directions of the function Φ:

Theorem 2 *Let us assume that z is a given iterate of the above sequential bilinear programming method for which the bilevel quadratic optimization problem (41) admits a feasible solution Δz while the Assumptions 4, 5, 6 and 7 are satisfied. Furthermore, we assume that the weights in the merit function Φ are sufficiently large such that we have*

$$\forall j \in \{0, \ldots, n\} : \quad \hat{\chi} > \left|\chi^{\dagger}\right|, \; \hat{\rho}_k > \frac{3}{2}\left|\rho_k\right|, \; \hat{\lambda}_j > 0. \tag{72}$$

Then, we have either

$$\begin{aligned} \Delta x = 0, \; \pi(B^j) = 0, \; \pi(L^i) = 0, \; \left|L_w^j\right| = 0, \\ \rho_j = 0, \; \Delta w_j = 0, \; \text{and} \; \lambda_j^T B^j = 0 \end{aligned} \tag{73}$$

for all $i \in \{1, \ldots, n\}$ and all $j \in \{0, \ldots, n\}$ or Δz is a descent direction of the merit function Φ, i.e., we have

$$\partial_\alpha \Phi := \lim_{\alpha \to 0^+} \frac{\Phi(\, x + \alpha \Delta x, \; w + \alpha \Delta w, \; \lambda + \alpha \Delta \lambda \,) - \Phi(x, w, \lambda)}{\alpha} < 0. \tag{74}$$

Proof In the first step of this proof, we use the formula (63) in combination with the linearized stationarity conditions (54) to compute

$$\begin{aligned} \partial_\alpha L_0(x, w_0, \lambda_0) \; &= \; L_x^0 \Delta x + L_w^0 \Delta w_0 - \Delta\lambda_0^T B^0 \\ &\overset{(54)}{=} -\Delta x^T K_{xx} \Delta x + \Delta x^T T_0 \Delta x + s_0^T \Omega_0^{-1} R_0 \Delta x + L_w^0 \Delta w_0 - \Delta\lambda_0^T B^0 \\ &\quad - \sum_{k=1}^{n} \chi_k^{\dagger} \left(-L_x^k \Delta x + \Delta x^T T_k \Delta x + s_k^T \Omega_k^{-1} R_k \Delta x \right) \end{aligned} \tag{75}$$

By collecting terms, the above equation can also be summarized in the form

$$\partial_\alpha L_0(x, w_0, \lambda_0) = -\Delta x^T K_{xx} \Delta x + X_0 - \sum_{k=1}^{n} \chi_k^{\dagger} X_k - \sum_{k=1}^{n} \chi_k^{\dagger} L^k, \tag{76}$$

where we use the short hands

$$X_0 := \Delta x^T T_0 \Delta x^T + s_0^T \Omega_0^{-1} R_0 \Delta x + L_w^0 \Delta w_0 - \Delta\lambda_0^T B^0 \tag{77}$$

and

$$X_k := -L_k - L_x^k \Delta x + \Delta x^T T_k \Delta x^T + s_k^T \Omega_k^{-1} R_k \Delta x. \tag{78}$$

for $k \in \{1, \ldots, n\}$. Now, the basic strategy is to use the necessary optimality conditions to transform the expressions for X_0 and X_k and completing squares in such a way that we can find suitable estimates for them. We start with the term for X_0:

$$\begin{aligned} X_0 &:= \Delta x^T T_0 \Delta x + s_0^T \Omega_0^{-1} R_0 \Delta x + L_w^0 \Delta w_0 - \Delta\lambda_0^T B^0 \\ &= -\Delta x^T L_{xw}^0 \Delta w_0 + L_w^0 \Delta w_0 - \Delta\lambda_0^T B^0. \end{aligned} \tag{79}$$

The latter equality can be verified by multiplying Eq. (59) with $\Delta x^T R_0^T \Omega_0^{-1}$ from the left. In the next step we use the stationarity condition for the lower QP to further transform to

$$\begin{aligned} X_0 &= \Delta w_0 L_{ww}^0 \Delta w_0 + 2L_w^0 \Delta w_o - \Delta\lambda_0^T B_w^0 \Delta w - \Delta\lambda_0^T B^0 \\ &= \left(L_{ww}^0 \Delta w_0 + L_w^0\right)\left(L_{ww}^0\right)^{-1}\left(L_{ww}^0 \Delta w_0 + L_w^0\right) \\ &\quad - L_w^0 \left(L_{ww}^0\right)^{-1} L_w^{0\,T} + \lambda_0^T \left(B_w^0 \Delta w_0 + B^0\right). \end{aligned} \tag{80}$$

The first term in the right side of the above transformation is negative as L_{ww}^0 is negative definite. Similarly, we have $\lambda_0^T \left(B_w^0 \Delta w + B^0\right) \le 0$ as $\lambda_0 \ge 0$ and $B_w^0 \Delta w + B^0 \le 0$. Thus, we find

$$X_0 \le -L_w^0 \left(L_{ww}^0\right)^{-1} L_w^{0\,T}. \tag{81}$$

In order to obtain a similar estimate for X_k with $k \in \{1, \ldots, n\}$ we use the complementarity relation (57) to find

$$\begin{aligned} X_k &= -L_k - L_x^k \Delta x + \Delta x^T T_k \Delta x^T + s_k^T \Omega_k^{-1} R_k \Delta x \\ &\overset{(57)}{=} -\frac{1}{2}\Delta w_k^T L_{ww}^k \Delta w_k + \Delta w_k^T B_w^{k\,T} \lambda^\dagger + \lambda_k^T B_k - \Delta x^T L_{xw} \Delta w \\ &= \frac{1}{2}\Delta w_k^T L_{ww}^k \Delta w_k + L_w^k \Delta w_k - \Delta\lambda_k B_w^k \Delta w_k + \lambda^{\dagger T} B_w^k \Delta w_k + \lambda_k^T B_k \\ &= \frac{1}{2}\left(L_{ww}^k \Delta w_k + L_w^k\right)\left(L_{ww}^k\right)^{-1}\left(L_{ww}^k \Delta w_k + L_w^k\right) \\ &\quad - \frac{1}{2} L_w^k \left(L_{ww}^k\right)^{-1} L_w^{k\,T} + \lambda_k^T \left(B_w^k \Delta w_k + B^k\right) \end{aligned} \tag{82}$$

Thus, we have

$$X_k \le -\frac{1}{2} L_w^k \left(L_{ww}^k\right)^{-1} L_w^{k\,T}. \tag{83}$$

In the next step, we are interested in computing the directional derivative of the upper-level merit function Φ_U. For this aim, we use Eq. (76) to find

$$\begin{aligned}\partial_\alpha \Phi_U &\le -\Delta x^T K_{xx} \Delta x + X_0 - \sum_{k=1}^{n} \chi_k^\dagger X_k - \sum_{k=1}^{n} \left(\hat{\chi} + \chi_k^\dagger\right) \pi_k(L^k) \\ &\quad - \sum_{k=1}^{n} \hat{\chi} L_w^k \left(L_{ww}^k\right)^{-1} L_w^{k\,T} \\ &\le X_0 - \sum_{k=1}^{n} \chi_k^\dagger X_k - \sum_{k=1}^{n} \hat{\chi}_k L_w^k \left(L_{ww}^k\right)^{-1} L_w^{k\,T}, \end{aligned} \tag{84}$$

where the last inequality holds strictly if $\Delta x \neq 0$ as K_{xx} is assumed to be positive definite and $0 \le \left|\chi^\dagger\right| < \hat{\chi}$. Similarly, we compute the directional derivative of the lower level merit functions using the formulas from Proposition 2 to find

$$X_0 + \partial_\alpha \Phi_L^0 \;(81) \le -\left|L_w^0\right| \left(\hat{\rho}_0 - |\rho_0|\right) - \hat{\lambda}_0 \pi(B^0) \le 0 \tag{85}$$

as well as

$$X_k + \frac{1}{3}\partial_\alpha \Phi_L^k \;(83) \le -\left|L_w^k\right| \left(\frac{1}{3}\hat{\rho}_k - \frac{1}{2}|\rho_k|\right) - \frac{1}{3}\hat{\lambda}_k \pi(B^k) \le 0. \tag{86}$$

as we assume $\hat{\rho}_k > \frac{3}{2}|\rho_k|$. Both estimates together yield

$$\partial_\alpha \Phi \le \sum_{k=1}^{n} \left(-\frac{2}{3}\hat{\chi}_k \left|L_w^k\right| \hat{\rho}_k + \hat{\chi}_k \left|L_w^k\right| |\rho_k|\right) \le 0, \tag{87}$$

where we use again the assumption $\hat{\rho}_k > \frac{3}{2}|\rho_k|$. For the case that we have $\partial_\alpha \Phi = 0$ all the above inequalities must be tight. Collecting the corresponding conditions, we find that this can only be the case if we have

$$\Delta x = 0, \;\; \pi(B_j) = 0, \;\; \pi(L^i) = 0, \;\; \left|L_w^j\right| = 0 \tag{88}$$

$$\Delta w_j = 0, \;\; \lambda_j^T B^j = 0, \;\; \text{and} \;\; \rho_j = 0. \tag{89}$$

for all $i \in \{1, \ldots, n\}$ and all $j \in \{0, \ldots, n\}$. Thus, we conclude the statement of the Theorem. □

Note that the above Theorem shows that we get either a descent direction of the merit function Φ or $\lambda_j^T B_j = 0$ is implied. This is surprising in the sense that we did not penalize the complementarity condition in the function Φ. Indeed, this observation leads to the following corollary:

Corollary 2 *Let us assume that the penalty weights in the merit function Φ are sufficiently large. Then every local solution of the unconstrained optimization problem*

$$\min_{x,w,\lambda} \;\; \Phi(x, w, \lambda) \tag{90}$$

at which the regularity Assumptions 4, 5, and 6 are satisfied, is either an infeasible stationary point or a KKT-point of the MPCC (40). Moreover, if there exists a solution $(\hat{x}, \hat{w}, \hat{\lambda})$ of the unconstrained optimization problem (90) at which the regularity Assumptions 4, 5, and 6 hold, then every solution of the MPCC (40) is also a solution of the unconstrained optimization problem (90), i.e., the merit function Φ is an exact penalty function.

Proof Let us assume that we have a solution (x^*, w^*, λ^*) of the unconstrained penalty problem (90) which is not a KKT point of the MPCC (40). Provided that (x^*, w^*, λ^*) not an infeasible point, an application of the above sequential convex bilevel programming method is well defined in the sense that a feasible step Δz must exist—independent on how we choose the positive definite matrix K_{xx}. As (x^*, w^*, λ^*) is assumed to be not a KKT point it can easily be seen that we can not possibly satisfy all the conditions (73), i.e., we get a descent direction of Φ, which is obviously a contradiction to our assumption that (x^*, w^*, λ^*) is a local solution of the unconstrained penalty problem (90). Thus, every local solution of the unconstrained optimization problem (90) must either be an infeasible stationary point or a KKT point of the MPCC (40).

The other way round, let us assume that (x^*, w^*, λ^*) is a solution of the MPCC (40) achieving the minimum objective value $H_0(x^*, w_0^*)$. If this point is not a solution of the unconstrained optimization problem (90) and not an infeasible stationary point, then the solution $(\hat{x}, \hat{w}, \hat{\lambda})$ of (90) satisfies $H_0(\hat{x}, \hat{w}_0) < H_0(x^*, w_0^*)$, i.e., we can use the above argumentation to show that $(\hat{x}, \hat{w}, \hat{\lambda})$ is a feasible KKT point of the MPCC (40) with a lower objective value than the assumed solution (x^*, w^*, λ^*). This is a contradiction to the assumption that (x^*, w^*, λ^*) is a solution of the MPCC (40). Consequently, Φ is an exact penalty function. □

Note that the Theorem 2 and the corresponding Corollary 2 enable us to transfer the traditional argumentation for the globalization of SQP methods [23,39], i.e., we can require an Armijo-Goldstein condition of the form

$$\tilde{\Phi}(\alpha) \leq \tilde{\Phi}(0) + \epsilon\, \alpha\, \partial_\alpha \tilde{\Phi}(0) \quad \text{with} \quad \Phi(\alpha) \ := \ \Phi(x + \alpha \Delta x, w + \alpha \Delta w, \lambda + \alpha \Delta \lambda) \tag{91}$$

to be satisfied with some $\epsilon > 0$, adjusting α via a line search such that a descent of the iterations is guaranteed. Under some additional assumptions, i.e., feasibility of the sub-problems, uniform boundedness of the multipliers χ, ρ, and λ, and the uniform boundedness of K_{xx} and K_{xx}^{-1}, the traditional global convergence statements from the SQP theory transfer [23].

4.2 Local convergence analysis

The local convergence properties of the presented sequential convex bilevel programming method are much easier to discuss than the global convergence. Basically, we can transfer the classical concepts for the local analysis of standard SQP theory. Thus, we will in this section present the local convergence theory on an adequate advanced level aiming at remarks on the details which are specific for sequential convex bilevel programming methods.

Let us directly constrain ourselves to the assumption that the active set during the local phase of the algorithm is already correctly detected and stable. Here, the stability of the active set can in our context be guaranteed as follows:

Assumption 8 We assume that at the local MPCC minimizer (x^*, w^*, λ^*) of our interest the following strong regularity conditions are satisfied:

1. The solution w^* of the lower level maximization problems is nondegenerate.
2. The ELICQ (or equivalently the MPCC-LICQ) condition is satisfied at (x^*, w^*, λ^*).
3. The second order sufficient condition for the auxiliary problem (38) is satisfied.
4. The upper level strict complementarity condition

$$L_i(x^*, w_i^*, \lambda_i^*) - \chi_i^* < 0 \tag{92}$$

holds for all $i \in \{1, \ldots, n\}$.

Lemma 4 *Let (x^*, w^*, λ^*) be a local minimizer of the MPCC (40) at which the regularity Assumption 8 is satisfied. Then there exists a neighborhood of (x^*, w^*, λ^*) in which the bilevel optimization admits a feasible solution Δz which has the same active set as the local minimizer (x^*, w^*, λ^*), i.e., we have $\mathcal{A}_j(\lambda^\dagger) = \mathcal{A}_j^*$ for all $j \in \{0, \ldots, n\}$ as well as $\mathcal{A}(\chi^\dagger) := \{k \mid \chi_k > 0\} = \mathcal{A}^*$ for all iterates in this neighborhood.*

Proof The feasibility as well as the stability of the active set for the lower level QPs follows immediately from Robinson's theorem [47,48]. Similarly, we can also apply Robinson's theorem to the upper level auxiliary problem (38), i.e., we obtain the feasibility and active set stability of the local QP-type necessary conditions from Lemma 3. Here, we use that the ELICQ condition boils down to an LICQ condition for the problem (38). As the fourth requirement of Assumption 8 guarantees the SCC condition for the problem (38), we have all the necessary regularity conditions for the problem (38) such that an application of Robinson's theorem is justified. Thus, we conclude the statement of the theorem. $\square$

A question which we have not discussed so far is how we should choose the matrix K_{xx}. In the previous section we have assumed that K_{xx} is positive definite as this was needed to guarantee a descent in the merit function during the global phase. This assumption is in principle not necessary for the discussion of local convergence properties, although it is still desirable in the sense that it guarantees the convexity of the sub-problems. However, in the context of local convergence, we are rather interested in a Dennis-Moré condition of the form

$$\left\| \left(K_{xx}^m - \frac{\partial^2}{\partial x^2} L_0(x^*, w_0^*, \lambda_0^*) + \sum_{k=1}^n \chi_k^* \frac{\partial^2}{\partial x^2} L_k(x^*, w_k^*, \lambda_k^*) \right) \Delta x^m \right\| \leq c_m \| \Delta x^m \| \tag{93}$$

where $(c_m)_{m\in\mathbb{N}}$ is a non-negative real valued sequence. Note that—with quite some abuse of notation—the iteration index m has been recovered in this formulation recalling that the Hessian approximation $K_{xx} = K_{xx}^m$ may change from iteration to iteration.

Theorem 3 *Let the Assumption (8) be satisfied while the Hessian approximation sequence K_{xx}^m satisfies the Dennis-Moré estimate (93) for a sequence $(c_m)_{m\in\mathbb{N}}$. Moreover, we assume that the sequential convex bilevel programming method takes—at least close to the solution—always full-steps while the functions H_i and B have Lipschitz continuous Hessians. Now, the following statements hold:*

- *If the sequence $(c_m)_{m\in\mathbb{N}}$ satisfies $\lim_{m\to\infty}\ c_m = 0$, then the local convergence of the sequential convex bilevel programming method is r-superlinear.*
- *If the sequence $(c_m)_{m\in\mathbb{N}}$ satisfies $c_{m+1} \le \kappa\, c_m$ for some $\kappa < 1$, then the local convergence of the sequential convex bilevel programming method is r-quadratic.*

Proof Using the Lemma 4 our aim is to show that the sequential convex bilevel programming method is locally equivalent to a Newton type method applied to the necessary conditions (36) from Theorem 1 under the assumption that the active set is fixed. As the Proposition 1 show already that the sequential convex bilevel programming method linearizes the primal feasibility condition of the lower level problem in every step exactly, we discuss directly the linearization of the active upper level constraint:

$$\begin{aligned} L_i + \frac{\partial}{\partial z} L_i\, \Delta z &= L^i + L_x^i \Delta x + L_w^i \Delta w - \Delta\lambda_i^T B_i \\ &= H^i + L_x^i \Delta x + \left(\frac{1}{2} \Delta w_i^T L_{ww}^i + \Delta x^T L_{xw}^i + H_w^i \right) \Delta w_i \\ &\quad + \frac{1}{2} \Delta w_i L_{ww}^i \Delta w_i + L_w^i \Delta w_i \\ &= H^i + L_x^i \Delta x + \left(\frac{1}{2} \Delta w_i^T L_{ww}^i + \Delta x^T L_{xw}^i + H_w^i \right) \Delta w_i \\ &\quad - \frac{1}{2} \Delta w_i L_{ww}^i \Delta w_i - \Delta x L_{xw}^i \Delta w_i + \Delta\lambda_i^T B_w^i \Delta w, \end{aligned} \tag{94}$$

which leads to

$$\left\| L_i + \frac{\partial}{\partial z} L_i\, \Delta z \right\| \le O(\|\Delta z\|^2) \tag{95}$$

for all i in the active set, i.e., for all i with

$$H^i + L_x^i \Delta x + \left(\frac{1}{2} \Delta w_i^T L_{ww}^i + \Delta x^T L_{xw}^i + H_w^i \right) \Delta w_i = 0.$$

It remains to discuss the Newton step with regard to the stationarity equation

$$0 = \frac{\partial}{\partial x} \hat{K}(x, w, \lambda) = \frac{\partial}{\partial x} L_0(x, w_0, \lambda_0) - \sum_{k=1}^{n} \chi_k \frac{\partial}{\partial x} L_k(x, w_k, \lambda_k). \tag{96}$$

A linearization of the above expression for $\frac{\partial}{\partial x}\hat{K}$ leads to

$$\frac{\partial}{\partial x}\hat{K} + \frac{\partial}{\partial z}\left[\frac{\partial}{\partial x}\hat{K}\right]\Delta z = \left(L_{xx}^0 \Delta x + L_{xw}^0 \Delta w_0 + L_x^0\right) - \sum_{k=1}^{n} \chi_k \left(L_{xx}^k \Delta x + L_{xw}^k \Delta w_k + L_x^k\right) - \sum_{k=1}^{n} \Delta \chi_k L_x^k. \tag{97}$$

Note that we may assume $\Delta\lambda_j^{\mathrm{inact}} = 0$ during the local phase as we consider the case that the correct active set has already settled. Combining this knowledge with the relation

$$R_j \Delta x + \Omega_j \begin{pmatrix} \Delta w_j \\ -\lambda_j^{\dagger,\,\mathrm{act}} \end{pmatrix} + s_j = 0$$

we can further transform to

$$\frac{\partial}{\partial x}\hat{K} + \frac{\partial}{\partial z}\left[\frac{\partial}{\partial x}\hat{K}\right]\Delta z = \left(L_{xx}^0 - \sum_{k=1}^{n} \chi_k L_{xx}^k\right)\Delta x - T_0 \Delta x + \sum_{k=1}^{n} \chi_k T_k \Delta x - R_0^T \Omega_0^{-1} s_0 + \sum_{k=1}^{n} \chi_k R_0^T \Omega_0^{-1} s_0 + L_x^0 - \sum_{k=1}^{n} \chi_k^{\dagger} L_x^k. \tag{98}$$

Using the result of Lemma 3 in combination with the Lipschitz continuity of the Hessians terms as well as the Dennis-Moré estimate (93) we obtain

$$\left\| \frac{\partial}{\partial x}\hat{K} + \frac{\partial}{\partial z}\left[\frac{\partial}{\partial x}\hat{K}\right]\Delta z \right\| \leq \left(c_m + O(\|z - z^*\|)\right) \|\Delta x\| + \left\| -\sum_{k} \Delta \chi_k T_k \Delta x - \sum_{k} \Delta \chi_k R_k \Omega_k^{-1} s_k \right\| \tag{99}$$

Now, we use that

$$\left\| R_k \Omega_k^{-1} s_k \right\| = \|-T_k \Delta x - L_{xw} \Delta w\| \leq O(\|\Delta z\|) \tag{100}$$

to finally conclude

$$\left\| \frac{\partial}{\partial x}\hat{K} + \frac{\partial}{\partial z}\left[\frac{\partial}{\partial x}\hat{K}\right]\Delta z \right\| \leq \left(c_m + O(\|z - z^*\|)\right) \|\Delta x\| + O(\|\Delta z\|^2). \tag{101}$$

Note that this last estimate (101) together with the estimate (95) boil down to a standard Dennis-Moré convergence criterion for the Newton method applied to the optimality conditions with respect to the fixed active set. Both statements of the Theorem are a direct consequence. □

Remark 5 If we use, e.g., a line-search globalization routine based on the proposed exact non-smooth penalty, the Maratos effect [33,45] might prevent the method from taking full-steps. Note that there exists mature literature on how to avoid the Maratos effect in standard SQP algorithms [11,13,19,34]. These techniques can also be used for modifying the proposed sequential convex bilevel programming algorithm.

Remark 6 Note that the above Theorem covers the case that K_{xx}^m is generated by BFGS updates, for which superlinear convergence is obtained. In the case of exact Hessian approximations we have even quadratic convergence. This is in analogy to standard SQP methods.

Remark 7 Note that for a direct application of a general purpose SQP method to the MPCC (40) local convergence is much more difficult to analyze [20], as the KKT points of the MPCC (40) do not satisfy the MFCQ condition. Moreover, globalization results for general purpose SQP methods applied to MPCCs are—due to the unbounded multiplier solution set of an MPCC—difficult to obtain [20], but they are subject of current research [2,62]. From this perspective, the presented sequential convex bilevel programming method is a mathematically sound alternative to standard SQP methods for structured MPCCs.

Remark 8 Note that the above local convergence result could be generalized to the case that the second order matrices L_{ww}, L_{wx}, and L_{xw} do not exactly coincide with their associated second order terms as long as they are suitable approximations. However, for such an ”inexact“ sequential convex bilevel programming method, the global convergence argumentation from the previous section would fail, as an approximation of these second order terms would amount to an inexact linearization of the lower level stationarity conditions, which are in the MPCC (40) formulated as equality constraints.

4.2.1 A stopping criterion

Note that within an implementation of the proposed method, we need a stopping criterion to decide numerically when convergence is achieved. For this aim, we define the KKT-tolerance ϵ of the sequential convex bilevel programming method analogous to SQP methods as

$$\epsilon_{\mathrm{L}}^j := \Phi_{\mathrm{L}}^j(x, w_j, \lambda_j) = \left| \frac{\partial L_j(x, w_j, \lambda_j)}{\partial w} \right| \hat{\rho}_j + \hat{\lambda}_j^T \pi(B(x, w_j)),$$

$$\epsilon_{\mathrm{U}} := |\partial_\alpha L_0(x, w_0, \lambda_0)| + \sum_{k=1}^{n} \hat{\chi}_k \, \pi_k(L_k(x, w_k, \lambda_k)),$$

$$\text{and} \qquad \epsilon := \epsilon_{\mathrm{U}} + \epsilon_{\mathrm{L}}^0 + \sum_{k=1}^{n} \hat{\chi}_k \, \epsilon_{\mathrm{L}}^k. \tag{102}$$

We can stop the method if $\epsilon < \text{TOL}$ is satisfied for a user-specified tolerance TOL, as the above definition of the KKT tolerance ϵ measures the violation of the KKT conditions for optimality.

4.2.2 The possible loss of superlinear convergence for non-convex problems and positive definite Hessian approximation

Being at this point, we have discussed the local and global convergence of the method rather independently obtaining consistent results. However, the question which we have not addressed so far is whether we can always satisfy the Dennis-Moré condition for superlinear or quadratic convergence which is needed in the above Theorem 3. This is certainly possible if we work with exact Hessians. For the case that the upper level problems are convex these exact Hessian matrices will be positive semi-definite and we cannot encounter problems with convexity of the sub-problems. The question is now whether we can work with bounded and positive semi-definite Hessian approximations K_{xx} even if the exact Hessian

$$\frac{\partial^2}{\partial x^2} L_0(x^*, w_0^*, \lambda_0^*) + \sum_{k=1}^{n} \chi_k^* \frac{\partial^2}{\partial x^2} L_k\left(x^*, w_k^*, \lambda_k^*\right) \tag{103}$$

is indefinite or negative definite still obtaining superlinear convergence. Although this is for standard SQP methods the case [44], this will in general not be possible for our sequential convex bilevel programming method. The corresponding effect has been worked out for sequential linear conic programming methods [16]. It was shown that sequential linear conic programming methods with bounded positive definite Hessian may not converge superlinearly for some non-convex problems.

In the following we will show that there exists an example for which the proposed sequential convex bilevel programming method suffers from the so-called Diehl-Jarre-Vogelbusch effect. For this aim we consider the problem

$$\begin{aligned} &\min_{x\in\mathbb{R}^2} \quad \left\{-x_1^2 - (x_2-1)^2\right\} \\ &\text{s.t.} \quad \max_{w\in\mathbb{R}^2} \left\{2x^T w - 1 - w^T w\right\} \leq 0 \end{aligned} \tag{104}$$

Applying the presented sequential convex bilevel programming strategy with the exact Hessian $K_{xx} = -2I_{2\times 2}$, the method converges independent of the starting point in one step to the unique solution $x^* = w^* = (0, -1)^T$.

The closest positive semi-definite approximation of the exact Hessian $-2I_{2\times 2}$ would be given by $K_{xx} = 0$. If we use this approximation the method converges linearly with convergence rate $\frac{1}{2}$. Note that this example is completely analogous to the one proposed in [16] and thus the corresponding argumentation transfers.

The Diehl-Jarre-Vogelbusch effect can never cause a problem if the original optimization problem is convex as the exact Hessian is positive semi-definite in this case. However, for general non-convex optimization problems we should be aware of the fact that there exist non-convex cases in which the superlinear convergence is lost if we want to work with positive semi-definite Hessian approximations.

5 Applications, implementation details, and numerical examples

Looking for applications of the presented sequential convex bilevel programming method in the context of robust optimization problems we should be aware of the fact that it is hard to find nonlinear min–max optimization problems in practice which are—at least without further transformation—concave with respect to their lower level maximization problems. There is one important exception: the problems for which the functions H_i are affine in w. As mentioned in the introduction, the case that H_i is affine in w has been discussed by many authors [5,6,15,17,29,35] using the assumption that the uncertainty sets are ellipsoidal or intersections of ellipsoids. In this case the lower-level maximization problems can explicitly be eliminated which leads typically to a (possibly nonlinear) second order cone program (SOCP) on the upper level. As this explicit elimination might make the problem also more nonlinear, the presented sequential convex bilevel programming method could be an alternative, although we do not expect major improvements when applying it to this well-elaborated problem class.

Looking for the case that the functions H_i are not convex in w, we have to accept that we need an estimate for the second order terms to achieve conservative reformulations, which are however still less conservative than linear approximation approaches as we have extensively discussed within Sect. 2. The corresponding nonlinear robust or semi-infinite optimization problems are the main application of the presented method. Here, the efficiency of the sequential convex bilevel programming method is mainly influenced by two factors:

First, the functions H_i and B as well as their first and some of their second order derivatives need to be evaluated. In most practical situations the evaluation of the function B will be cheap. Recall that we have the cases that B describes a simple box or ball in mind. Thus, the aim is to reduce the cost of the evaluation of sensitivities of the functions H_i. Note that the presented method needs even in exact Hessian mode only second order derivatives. This is in contrast to linear adjoint based approximation techniques [15] which would need third order derivatives to achieve quadratic convergence. Thus, the proposed sequential convex bilevel programming method can outperform the adjoint based linearization technique with respect to both: the function and sensitivity evaluation cost per iteration of a method as well as the conservatism of the robust counterpart approximation.

Second, the cost of the sequential convex bilevel programming method is influenced by the cost of solving the sub-problems, which are itself structured min–max QCQPs. The tractability of these bilevel sub-problems is on the one hand due to their convexity and on the other hand implied by the structure which comes from the fact that for any given Δx the lower level QPs are decoupled. For practical purposes, the min–max QCQPs can for example be transformed into equivalent convex QCQPs which can be solved with existing algorithms.

5.1 A numerical test example

In this section, we demonstrate the applicability of the proposed method with a numerical example. For this aim, we consider a control task for an elastic rope which is modeled by 11 mass points and 10 connecting springs. The states of the system are the position coordinates $z_0, z_1, \ldots, z_{10} \in \mathbb{R}^2$ as well as the associated velocities $v_0, \ldots, v_{10}$. Consequently, we have 44 states in total which satisfy the following model equations

$$\forall i \in \{0, \ldots, 10\}: \quad \dot{z}_i(t) = v_i(t) \tag{105}$$

$$\dot{v}_0(t) = (u(t), 0)^T \tag{106}$$

$$\forall i \in \{1, \ldots, 9\}: \quad \dot{v}_i(t) = (0, -g)^T + \frac{D_i}{m} \frac{\|z_i - z_{i-1}\| - a}{\|z_i - z_{i-1}\|} (z_{i-1} - z_i) + \frac{D_{i+1}}{m} \frac{\|z_i - z_{i+1}\| - a}{\|z_i - z_{i+1}\|} (z_{i+1} - z_i) \tag{107}$$

$$\dot{v}_{10}(t) = (0, -g)^T + \frac{D_{10}}{m} \frac{\|z_{10} - z_9\| - a}{\|z_{10} - z_9\|} (z_9 - z_{10}) \, . \tag{108}$$

Note that the mass point with index 0 is assumed to be directly controllable in horizontal direction, which is indicated by the control input u. Furthermore, $g = 9.81$ denotes the gravitational constant. The above dynamic system is nonlinear if the length a of the spring is not equal to 0. We use $a = 1$ in our example. Moreover, the spring constants are assumed to be unknown but given in the form

$$D_i = \overline{D} + w_i \, , \tag{109}$$

where $w := (w_1, \ldots, w_{10})$ is only known to satisfy $\|w\|_2 \leq 1$ while $\overline{D} = 10$ is given.

In order to discretize the control input we replace the function u by a piecewise constant approximation

$$\sim u(t) := \sum_{i=0}^{N-1} x_i \, I_{[t_i, t_{i+1}]}(t),$$

where $I_{[a,b]}(t)$ is equal to 1 if $t \in [a, b]$ and equal to 0 otherwise and $x_i \in \mathbb{R}^2$ are the new control parameters. The time sequence $0 = t_0 < t_1 < \cdots < t_N = T = 10$ is in our example assumed to be equidistant with $N = 10$. In the following, we summarize all decision variables in the vector

$$x := (x_0, \ldots, x_{N-1})^T \, .$$

If we start the above system at the equilibrium position for $z_0 = 0$ and $w = 0$ and simulate from $t_0 = 0$ to $T = 10$, the position and the velocity of the mass point with index 10 can be regarded as a function of x and w which we denote by $z_{10}(T, x, w)$ and $v_{10}(T, x, w)$ such that we can define the functions $F_1, F_2 : \mathbb{R}^N \times \mathbb{R}^{10} \to \mathbb{R}$ as

$$F_1(x, w) := -(1, 0)^T \, z_{10}(T, x, w) + 5$$
$$\text{and} \quad F_2(x, w) := (1, 0)^T \, v_{10}(T, x, w) - 1$$

Our objective is now to minimize the control input $F_0(x) := \sum_{i=0}^{N-1} x_i^2$ while satisfying constraints on F_1 and F_2 in a robust way. The corresponding robust optimization problem can be written as

$$\boxed{\begin{array}{lll} \min\limits_{x} & F_0(x) & \\ \text{s.t.} & \max\limits_{\|w\|_2 \le 1} F_i(x, w) \le 0 & i \in \{1, 2\} \end{array}} . \tag{110}$$

Note that each evaluation of the function F_1 or F_2 requires a simulation of the nonlinear dynamic system with its 44 states which is rather expensive. In the implementation which we use for this paper the ACADO BDF integrator [28] is used as this integrator provides internal automatic differentiation for first and second order adjoint derivatives of the adaptively discretized differential equation. Running the integrator with the default tolerance of 10^{-6}, one evaluation of F_1 and F_2 as well as all required first and second order sensitivities takes all together approximately 0.5 seconds. Compared to this time, the computational cost for evaluating the function $B(x, w) = \|w\|_2^2 - 1$ as well as the the objective F_0 is negligible.

Now, we have a problem: evaluating the matrices $\frac{\partial^2 F_i(x,w)}{\partial w^2}$ for some randomly chosen points (x, w) and $i \in \{1, 2\}$ returns some indefinite matrices. The only thing we know from this test is that there exist points x for which the functions F_1 and F_2 are both definitely not convex in w. If we state now that $\overline{\lambda} := 0.05$ is an upper bound on the eigenvalues of the Hessian matrices of F_1 and F_2 with respect to w, this is an empirical statement, which was only validated by computing these Hessians at randomly chosen points (x, w). This is the aspect which we have to accept here recalling that we would have the same problem if we would work with linear approximation strategies. Assuming that this value for $\overline{\lambda}$ is valid we reformulate the above problem into a conservative lower level concave min–max problem following the techniques from Sect. 2.2 obtaining the functions $H_i(x, w) := F_i(x, w) + \overline{\lambda} - \overline{\lambda} w^T w$.

Being at this point, it remains to choose a suitable initial guess for x^0, w_1^0, and w_2^0 in order to finally start our sequential convex bilevel programming algorithm. In order to obtain such a guess it is advisable to solve the nominal optimization problem first, i.e., we use a standard optimal control tool to determine x^0 as the optimal solution for the above optimization problem in its non-robustified version, i.e., with $w_1 = w_2 = 0$. Having this point we compute the vectors $r_i := \frac{\partial}{\partial w} H_i(x_0, 0)$ which help us to use the heuristic $w_i^0 := \frac{r_i^T}{\|r_i\|}$ as an initialization which is possible as the vectors r_i (with $i \in \{1, 2\}$) are in our example not equal to zero—otherwise we suggest to start with $w_i^0 = 0$.

In order to analyze the iterations we list the KKT tolerance of the sequential convex bilevel programming method, as defined in Eq. (102), against the iteration number:

k	1	2	3	4	5
ϵ	1.5×10^1	1.4×10^0	3.1×10^{-1}	1.2×10^{-3}	6.1×10^{-10}

From the above table, we can observe the quadratic convergence behavior of the method as full steps were taken. Solving the convex bilevel sub-problem took in our example $\approx$ 6 ms. Compared to the 0.5 s which are needed for the evaluation of the functions H_i and their sensitivities, the time for solving the min–max QCQP sub-problems is negligibly small.

In the optimal solution, we found that $\|w_1^*\|_2^2 = 1$ and $\|w_2^*\|_2^2 = 1$ is satisfied, i.e., we have $F_i(x^*, w_i^*) = H_i(x^*, w_i^*)$. This shows a posteriori that our reformulation based on the estimate $\overline{\lambda}$ cannot have possibly introduced any conservatism. Thus, we know a posteriori that we have found a stationary point of the original non-convex problem with w_1^* and w_2^* being local maximizers of the original non-concave functions F_1 and F_2 respectively. If we would have a guarantee that our estimate for $\overline{\lambda}$ is not only empirically but also verifiably an upper bound on the Hessian matrices of the functions F_1 and F_2 with respect to w, we could even guarantee that w_1^* and w_2^* are global maximizers of the functions F_1 and F_2 at the optimal solution x^*. However, as we have in this paper not provided a proof that $\overline{\lambda}$ is such an upper bound, the global optimality of w_1^* and w_2^* remains in our example a conjecture.

6 Conclusions

In this paper, we have presented a sequential convex bilevel programming algorithm for solving semi-infinite optimization problems arising in the context of robust optimization. We have started with a discussion on how to approximate nonlinear and non-convex robust counterpart problems with lower level concave min–max optimization problems. Here, we have shown that the proposed approximation strategies are always less conservative than existing linear approximation techniques. Moreover, we have concentrated on optimality conditions for min–max problems working out relations between the theory for semi-infinite optimization and the theory for mathematical programs with complementarity constraints.

The main contribution of this paper is the introduction of the sequential convex bilevel programming method. The main advantage of this method is that it exploits the specific structure of the lower-level concave min–max problems. This is in contrast to existing methods for mathematical programs with complementarity constraints which are more general but also more difficult to use and analyze as they suffer from the degeneracy of MPCCs. This problem is avoided by the sequential convex bilevel technique leading to strong local and global convergence results. Furthermore, we have discussed the implementation details for the presented method including the transformation of the min–max QCQP in an equivalent standard QCQP. The applicability of the method has successfully been tested with a numerical example observing that the convex bilevel sub-problems can efficiently and reliably be solved such that the proposed method converges—at least in the discussed example—without numerical problems towards the optimal solution. Note that a comparison of different algorithms for semi-infinite optimization problems is beyond the scope of this paper, but might be an interesting direction for future research.

Acknowledgments The authors would like to thank Quoc Tran Dinh for fruitful discussions on sequential convex programming [37]. Additionally, we thank two anonymous referees and the associate editor for substantial remarks which significantly improved this paper. This research was supported by the Research Council KUL via the Center of Excellence on Optimization in Engineering EF/05/006 (OPTEC, http://www.

kuleuven.be/optec/), GOA AMBioRICS, IOF-SCORES4CHEM and PhD/postdoc/fellow grants, the Flemish Government via FWO (PhD/postdoc grants, projects G.0452.04, G.0499.04, G.0211.05,G.0226.06, G.0321.06, G.0302.07, G.0320.08, G.0558.08, G.0557.08, research communities ICCoS, ANMMM, MLDM) and via IWT (PhD Grants, McKnow-E, Eureka-Flite+), Helmholtz Gemeinschaft via vICeRP, the EU via ERNSI, Contract Research AMINAL, as well as the Belgian Federal Science Policy Office: IUAP P6/04 (DYSCO, Dynamical systems, control and optimization, 2007–2011).

References

1. Anitescu, M.: Nonlinear programs with unbounded Lagrange multiplier sets. Technical report, Preprint ANL/MCS-P796-0200, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL (2000)
2. Anitescu, M.: Global convergence of an elastic mode approach for a class of mathematical programs with complementarity constraints. SIAM J. Optim. **16**, 120–145 (2005)
3. Bard, J.F.: Practical Bilevel Optimization: Algorithms and Applications. Kluwer, Boston MA (1999)
4. Ben-Tal, A., Boyd, S., Nemirovski, A.: Extending scope of robust optimization: comprehensive robust counterparts of uncertain problems. Math. Program. (Ser. B) **107**, 63–89 (2006)
5. Ben-Tal, A., Nemirovski, A.: Robust truss topology design via semidefinite programming. SIAM J. Optim. **7**, 991–1016 (1997)
6. Ben-Tal, A., Nemirovski, A.: Robust convex optimization. Math. Oper. Res **23**, 769–805 (1998)
7. Ben-Tal, A., Nemirovski, A.: Robust solutions of uncertain linear programs. Oper. Res. **25**, 1–13 (1999)
8. Ben-Tal, A., Nemirovski, A., Roos, C.: Robust solutions of uncertain quadratic and conic-quadratic problems. SIAM J. Optim. **13**(2), 535–560 (2002)
9. Benson, H.Y., Sen, A., Shanno, D.F., Vanderbei, R.J.: Interior-point algorithms, penalty methods and equilibrium problems. Computat. Optim. Appl. **34**, 155–182 (2006)
10. Bhattacharjee, B., Lemonidis, P., Green, W.H., Barton, P.I.: Global solution of semi-infinite programs. Math. Program. (Ser. B) **103**(2), 283–307 (2005)
11. Chamberlain, R., Lemaréchal, C., Pedersen, H.C., Powell, M.J.D.: The watchdog technique for forcing convergence in algorithms for constrained optimization. Math. Program. **16**, 1–17 (1982)
12. Chen, L., Goldfarb, D.: An active set method for mathematical programs with linear complementarity constraint. Manuscript, Department of Industrial Engineering and Operations Research, Columbia University (2007)
13. Coleman, T.F., Conn, A.R.: Non-linear programming via an exact penalty function: asymptotic analysis. Math. Program. **24**, 123–136 (1982)
14. Conn, A.R., Gould, N., Toint, P.L.: Trust-Region Methods. MPS/SIAM Series on Optimization. SIAM, Philadelphia, USA (2000)
15. Diehl, M., Bock, H.G., Kostina, E.: An approximation technique for robust nonlinear optimization. Math. Program. **107**, 213–230 (2006)
16. Diehl, M., Jarre, F., Vogelbusch, C.: Loss of superlinear convergence for an SQP-type method with conic constraints. SIAM J. Optim. **16**(4), 1201–1210 (2006)
17. El-Ghaoui, L., Lebret, H.: Robust solutions to least-square problems to uncertain data matrices. SIAM J. Matrix Anal. **18**, 1035–1064 (1997)
18. Facchinei, F., Jiang, H., Qi, L.: A smoothing method for mathematical programs with equilibrium constraints. Math. Program. **85**, 107–134 (1999)
19. Fletcher, R.: Second order corrections for non-differentiable optimization. In: Griffiths, D. (ed.) Numerical Analysis. Springer, Berlin, pp. 85–114 (1982)
20. Fletcher, R., Leyffer, S., Ralph, D., Scholtes, S.: Local convergence of SQP methods for mathematical programs with equilibrium constraints. SIAM J. Optim. **17**, 259–286 (2006)
21. Floudas, C.A.: Deterministic Global Optimization: Theory, Methods, and Applications. Kluwer, Dordrecht (1999)
22. Floudas, C.A., Stein, O.: The adaptative convexification algorithm: a feasible point method for semi-infinite programming. SIAM J. Optim. **18**(4), 1187–1208 (2007)
23. Han, S.P.: A globally convergent method for nonlinear programming. JOTA **22**, 297–310 (1977)
24. Hettich, R., Jongen, H.T.: Semi-infinite programming: conditions of optimality and applications. In: Stoer, J. (ed.) Optimization Techniques, Part 2, Lecture Notes in Control and Inform. Sci., vol. 7. Springer, Berlin (1978)

25. Hettich, R., Kortanek, K.: Semi infinite programming: theory, methods, and application. SIAM Rev. **35**(3), 380–429 (1993)
26. Hettich, R., Still, G.: Second order optimality conditions for generalized semi-infinite programming problems. Optimization **34**, 195–211 (1995)
27. Houska, B.: Robustness and stability optimization of open-loop controlled power generating kites. University of Heidelberg, Master's thesis (2007)
28. Houska, B., Ferreau, H.J., Diehl, M.: ACADO toolkit—an open source framework for automatic control and dynamic optimization. Optim. Control Appl. Methods **32**(3), 298–312 (2011)
29. Houska, B., Logist, F., Van Impe, J., Diehl, M.: Approximate robust optimization of time-periodic stationary states with application to biochemical processes. In: Proceedings of the 48th Conference on Decision and Control, pp. 6280–6285. Shanghai, China (2009)
30. Jongen, H.T., Rückmann, J.J., Stein, O.: Generalized semi-infinite optimization: a first order optimality condition and examples. Math. Program. **83**, 145–158 (1998)
31. Liu, G.S., Zhang, J.Z.: A new branch and bound algorithm for solving quadratic programs with linear complementarity constraints. J. Comput. Appl. Math. **146**, 77–87 (2002)
32. Luo, Z., Pang, J.S., Ralph, D.: Piecewise sequential quadratic programming for mathematical programs with nonlinear complementarity constraints. In: Migdalas, A., Pardalos, P.M., Värbrand, P. (eds.) Multilevel Optimization: Algorithms and Applications, pp. 209–229. Kluwer, Dordrecht, The Netherlands (1998)
33. Maratos, N.: Exact penaly function algorithms for finite-dimensional and control optimization problems. PhD thesis, Imperial College, London (1978)
34. Mayne, D.Q., Polak, E.: A quadratically convergent algorithm for solving infinite dimensional inequalities. J. Appl. Math. Optim. **9**, 25–40 (1982)
35. Nagy, Z.K., Braatz, R.D.: Open-loop and closed-loop robust optimal control of batch processes using distributional and worst-case analysis. J. Process Control **14**, 411–422 (2004)
36. Nagy, Z.K., Braatz, R.D.: Distributional uncertainty analysis using power series and polynomial chaos expansions. J. Process Control **17**, 229–240 (2007)
37. Necoara, I., Savorgnan, C., Tran Dinh, Q., Suykens, J.A.K., Diehl, M.: Distributed nonlinear optimal control using sequential convex programming and smoothing techniques. In: Proceedings of the 48th IEEE Conference on Decision and Control, Shanghai, China (2009) (Accepted for publication)
38. Neumaier, A.: Complete Search in Continuous Global Optimization and Constraint Satisfaction. Cambridge University Press, Cambridge, pp. 271–369 (2004)
39. Nocedal, J., Wright, S.J.: Numerical Optimization. Springer Series in Operations Research and Financial Engineering, 2nd edn. Springer, Berlin (2006)
40. Polak, E., Mayne, D.Q., Higgins, J.E.: Superlinearly convergent algorithm for min-max problems. J. Optim. Theory Appl. **69**, 407–439 (1991)
41. Polak, E., Qi, L., Sun, D.: First-order algorithms for generalized semi-infinite min-max problems. Comput. Optim. Appl. **13**, 137–161 (1999)
42. Polak, E., Qi, L., Sun, D.: Second-order algorithms for generalized finite and semi-infinite min-max problems. SIAM J. Optim. **11**, 937–961 (2001)
43. Polak, E., Tits, A.: A recursive quadratic programming algorithm for semi-infinite optimization problems. J. Appl. Math. Optim. **8**, 325–349 (1982)
44. Powell, M.J.D.: The convergence of variable metric methods for nonlinear constrained optimization calculations. Academic Press, New York (1978)
45. Powell, M.J.D.: Convergence properties of algorithms for nonlinear optimization. SIAM Rev. **28**, 487–500 (1984)
46. Ralph, D.: Sequential quadratic programming for mathematical programs with linear complementarity constraints. In: May, R.L., Eston, A.K. (eds). Proceedings of the seventh conference on computational techniques and applications, pp. 663–668. Scienctific Press, Singapore (1996)
47. Robinson, S.M.: First order conditions for general nonlinear optimization. SIAM J. Appl. Math. **30**(4), 597–607 (1976)
48. Robinson, S.M.: Strongly regular generalized equations. Math. Oper. Res. **5**(1), 43–62 (1980)
49. Rückmann, J.J., Stein, O.: On linear and linearized generalized semi-infinite optimization problems. Ann. Oper. Res., **101**, 191–208 (2001)
50. Scheel, H., Scholtes, S.: Mathematical programs with complementarity constraints: stationarity, optimality, and sensitivity. Math. Oper. Res. **25**, 1–22 (2000)

51. Sherali, H.D., Tuncbilek, C.H.: A Reformulation-Convexification Approach for Solving Nonconvex Quadratic Programming Problems. J. Glob. Optim. **7**, 1–31 (1995)
52. Sherali, H.D., Tuncbilek, C.H.: New reformulation linearization/convexification relaxations for univariate and multivariate polynomial programming problems. Oper. Res. Lett. **21**, 1–9 (1997)
53. Shimizu, K., Lu, M.: A global optimization method for the Stackelberg problem with convex functions via problem transformation and concave programming. IEEE Trans. Syst. Man Cybern. **25**(23), 1635–1640 (1995)
54. Stein, O.: The feasible set in generalized semi-infinite programming. In: Lassonde, M. (ed.) Approximation, Optimization, and Mathematical Economics, pp. 313–331. Physica-Verlag, Heidelberg (2001)
55. Stein, O.: Bilevel Strategies in Semi Infinite Optimization. Kluwer, Boston (2003)
56. Stein, O., Still, G.: On optimality conditions for generalized semi-infinite programming problems. J. Optim. Theory Appl. **104**(2), 443–458 (2000)
57. Stein, O., Still, G.: Solving semi-infinite optimization problems with interior point techniques. SIAM J. Control Optim. **42**(3), 769–788 (2003)
58. Still, G.: Discretization in semi-infinite programming: the rate of convergence. Math. Program. **91**, 53–69 (2001)
59. Still, G.: Generalized semi-infinite programming: numerical aspects. Optimization **49**, 223–242 (2001)
60. Voorhis, T.V.: A global optimization algorithm using Lagrangian underestimates and the interval Newton method. J. Glob. Optim. **24**, 349–370 (2002)
61. Weber, G.W.: Generalized semi-infinite optimization and related topics. Habilitation Thesis , Darmstadt University of Technology, Darmstadt, Germany (1999)
62. Wright, S.J.: Modifying SQP for degenerate problems. SIAM J. Optim. **13**, 470–497 (2002)
63. Yakubovich, V.A.: S-procedure in nonlinear control theory. Vestnik Leningrad University **4**, 73–93 (1977)
64. Zhang, J., Liu, G.: A new extreme point algorithm and its application in PSQP algorithms for solving mathematical programs with linear complementarity constraints. J. Glob. Optim. **19**(4), 345–361 (2001)

Copyright of Mathematical Programming is the property of Springer Science & Business Media B.V. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.