PESSIMISTIC BILEVEL OPTIMIZATION[*]

WOLFRAM WIESEMANN[†], ANGELOS TSOUKALAS[‡], POLYXENI-MARGARITA
KLENIATI[§], AND BERÇ RUSTEM[†]

Abstract. We study a variant of the pessimistic bilevel optimization problem, which comprises constraints that must be satisfied for any optimal solution of a subordinate (lower-level) optimization problem. We present conditions that guarantee the existence of optimal solutions in such a problem, and we characterize the computational complexity of various subclasses of the problem. We then focus on problem instances that may lack convexity, but that satisfy a certain independence property. We develop convergent approximations for these instances, and we derive an iterative solution scheme that is reminiscent of the discretization techniques used in semi-infinite programming. We also present a computational study that illustrates the numerical behavior of our algorithm on standard benchmark instances.

Key words. global optimization, pessimistic bilevel problem, computational complexity

AMS subject classifications. 90C34, 90C26

DOI. 10.1137/120864015

1. Introduction. We study the pessimistic bilevel problem, which we define as

$$(\mathcal{PB}) \quad \begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & g(x,y) \leq 0 \qquad \forall y \in \mathcal{Y}(x) = \underset{z}{\arg\min} \{h(x,z) \,:\, z \in Y(x)\} \\ & x \in X, \end{array}$$

where $X \subseteq \mathbb{R}^n$, $Y(x) \subseteq Y \subseteq \mathbb{R}^m$ for all $x \in X$, $f : X \mapsto \mathbb{R}$, and $g, h : X \times Y \mapsto \mathbb{R}$. We call y the lower-level decision and refer to the embedded minimization problem as the lower-level problem, respectively. We refer to x as the upper-level decision. We stipulate that the decision $x \in X$ satisfies the bilevel constraint if the lower-level problem associated with x is infeasible. Note that the bilevel problem generalizes to multiple constraints $g_1, \ldots, g_p : X \times Y \mapsto \mathbb{R}$ if we set $g(x,y) = \max \{g_i(x,y) \,:\, i = 1, \ldots, p\}$. It also extends to certain classes of min-max problems with coupled constraints [82].

It is worthwhile to notice that the bilevel problem $(\mathcal{PB})$ deviates slightly from the standard formulation

$$(1.1) \qquad \min_{x \in X} \sup_{y \in M_2(x)} f_1(x,y), \quad \text{where } M_2(x) = \underset{y \in Y}{\arg\min} \; f_2(x,y);$$

[†]Department of Computing, Imperial College London, SW7 2AZ London, UK (wwiesema@ imperial.ac.uk, br@imperial.ac.uk).

[‡]Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139 (maurostsouk@googlemail.com). This work was done while the author was a doctoral student at the Department of Computing, Imperial College, London.

[§]Department of Chemical Engineering, Imperial College London, SW7 2AZ London, UK (polyxeni.kleniati03@imperial.ac.uk).

see, e.g., [24, 49]. In fact, the standard formulation (1.1) can be reformulated as an instance of ($\mathcal{PB}$):

$$\begin{aligned}
\underset{x,\tau}{\text{minimize}} \quad & \tau \\
\text{subject to} \quad & \tau \geq f_1(x,y) \qquad \forall y \in \mathcal{Y}(x) = \arg\min_{z}\{f_2(x,z) \,:\, z \in Y\} \\
& x \in X.
\end{aligned}$$

On the other hand, the bilevel problem ($\mathcal{PB}$) does not reduce to an instance of (1.1), unless we allow for extended real-valued functions f_1, f_2 in (1.1).

Bilevel problems have a long history that dates back to the investigation of market equilibria by von Stackelberg in the 1930s [35]. Bilevel problems have first been formalized as optimization problems in the early 1970s [17]. In recent years, bilevel optimization has been applied to various domains including revenue management [22], traffic planning [56], security [74], supply chain management [73], production planning [41], process design [20], market deregulation [13], optimal taxation [8], and parameter estimation [16, 63].

The bilevel problem $\mathcal{PB}$ has a natural interpretation as a noncooperative game between two players. Player A (the "leader") chooses her decision x first, and afterwards player B (the "follower") observes x and responds with a decision y. Both the objective function and the feasible region of the follower may depend on the leader's decision. Likewise, the leader has to satisfy a constraint that depends on the follower's decision. Since the leader cannot anticipate the follower's decision, the constraint must be satisfied for any rational decision of the follower, that is, for any decision $y \in \mathcal{Y}(x)$ that optimizes the follower's objective function.

The above-stated pessimistic bilevel problem is perceived to be very difficult to solve. As a result, most theoretical and algorithmic contributions to bilevel programming relate to the optimistic formulation, in which the universal quantifier "$\forall$" in the bilevel constraint is replaced with an existential quantifier "$\exists$." In a game-theoretic context, the optimistic problem can be justified in two ways. On one hand, there may be limited cooperation between the players to the extent that the follower altruistically chooses an optimal solution that also benefits the leader. On the other hand, the leader may be able to make small side payments that bias the follower's objective in her favor. Even though the optimistic and the pessimistic bilevel problem are very similar, their optimal solutions can differ considerably.

Example 1.1. Consider the following instance of the pessimistic bilevel problem:

$$\begin{aligned}
\underset{x}{\text{minimize}} \quad & x \\
\text{subject to} \quad & x \geq y \qquad \forall y \in \arg\min_{z}\left\{-z^2 \,:\, z \in [-1,1]\right\} \\
& x \in \mathbb{R}.
\end{aligned}$$

The lower-level problem is optimized by $z \in \{-1,1\}$, independent of the upper-level decision. The pessimistic bilevel problem therefore requires x to exceed 1, resulting in an optimal objective value of 1. In contrast, the optimistic bilevel problem requires x to exceed -1, which results in an optimal objective value of -1.

Many algorithms have been proposed for the bilevel problem. In the following, we review some of the methods that determine globally optimal solutions. For surveys of local optimization approaches and optimality conditions, see [4, 7, 21, 24, 76, 86] and the articles in [33]. We review the literature on the complexity of the bilevel problem in section 2.2. The relationship between bilevel problems and min-max

problems, generalized semi-infinite programs, mathematical programs with equilibrium constraints, and multicriteria optimization is explored in [32, 51, 61, 72, 77]. For more general multilevel optimization problems, see [19, 57].

The most benign class of bilevel problems concerns the optimistic variant and stipulates that the functions f, g, and h are affine, while the feasible regions X and $Y(x)$ are described by polyhedra. Although these linear optimistic bilevel problems are nonconvex, they are known to be optimized by an extreme point of the polyhedron $\Omega = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m : x \in X, y \in Y(x), g(x, y) \leq 0\}$. Several algorithms have been proposed that enumerate the extreme points of Ω; see [12]. If the lower-level problem of an optimistic bilevel problem is convex and some mild constraint qualifications are satisfied, then the Karush–Kuhn–Tucker conditions are necessary and sufficient for the global optimality of the lower-level problem. We can then replace the bilevel problem with a single-level problem that contains the Karush–Kuhn–Tucker conditions as constraints. The resulting problem is nonconvex and can be solved with DC (difference of convex functions) optimization techniques or tailored branch-and-bound algorithms; see [1, 3, 37, 38]. Alternatively, if the lower-level problem of an optimistic bilevel problem is linear or convex quadratic, then one can use multiparametric programming techniques to solve the lower-level problem parametrically for each upper-level decision x; see [29, 30, 70]. The resulting piecewise affine solution map $x \mapsto y^*(x)$ records the optimal lower-level decision $y^*(x)$ for each upper-level decision x. The optimal upper-level decision is then found by solving a single-level problem for each affine subregion of the solution map, where the lower-level decision y is replaced with the affine function $y^*(x)$.

Bilevel problems become much more challenging if the lower-level problem fails to be convex [61]. In particular, the Karush–Kuhn–Tucker conditions are no longer sufficient for the global optimality of the lower-level problem. Hence, a single-level formulation that replaces the lower-level problem with the Karush–Kuhn–Tucker conditions no longer results in an equivalent reformulation, but merely in a conservative approximation [58]. If an optimistic bilevel problem contains a nonconvex lower-level problem that satisfies certain monotonicity requirements, then one can apply algorithms from monotonic optimization [84] to globally solve the optimistic bilevel problem; see [85]. For optimistic bilevel problems with a generic nonconvex lower-level problem, a global optimization method is developed in [65]. The algorithm computes parametric upper bounds on the optimal value function of the lower-level problem as a function of the upper-level decisions. These parametric upper bounds can be used in an optimal value reformulation to construct a relaxation of the bilevel problem that can be solved as a single-level problem. Each of these single-level problems provides a lower bound on the optimal objective value of the bilevel problem. By solving the lower-level problem for a fixed value of the upper-level decisions and afterwards resolving the upper-level problem using an optimal value reformulation, one also obtains an upper bound on the optimal objective value. The authors present a method to iteratively tighten these bounds until the algorithm converges to the globally optimal solution. They also elaborate several extensions of the algorithm, such as tightened lower bounds using the Karush–Kuhn–Tucker conditions and a branching scheme for the upper-level decisions. The method is extended to efficiently deal with continuous and discrete variables in [59], and an extension to differential equations is proposed in [64]. A global optimization method for generalized semi-infinite, coupled min-max and optimistic bilevel problems without any convexity assumptions is developed in [81]. The algorithm relies on an "oracle" optimization problem that decides whether

a specified target objective value is achievable. If the target value can be attained, then the oracle problem also determines a feasible solution that attains the target value. The overall algorithm conducts a binary search over all target objective values to determine the globally optimal solution. Finally, the paper [62] compiles a test set of linear, convex, and nonconvex bilevel problems with optimal solutions to the optimistic formulation.

Typically, the algorithms for the optimistic bilevel problem do not extend to the pessimistic formulation. One may try to avoid this issue by solving the optimistic formulation and afterwards—should there be multiple optimal lower-level solutions—perturb the upper-level decision so that a unique optimal lower-level solution is induced; see [24, section 7.1] and [75]. However, Example 1.1 shows that such a perturbation may not exist.

Although the algorithms for the optimistic bilevel problem do not directly extend to the pessimistic formulation, several papers suggest to solve the pessimistic bilevel problem indirectly through a sequence of optimistic bilevel problems [49, 53, 54, 66]. In these papers, the objective function of the lower-level problem is amended with a penalty term which favors lower-level decisions that lead to higher costs in the upper-level objective. Under certain assumptions, the optimal solutions to these optimistic bilevel problems converge to the optimal solutions of the pessimistic formulation if the coefficient of the penalty term is decreased to zero. However, none of the papers provide numerical results for this scheme, and it remains unclear how well the penalization method would work as part of a global optimization procedure.

We summarize the contributions of this paper as follows.

1. We analyze the structural properties of the pessimistic bilevel problem, including the existence of optimal solutions and the computational complexity. In particular, we identify an "independence" property that facilitates the development of solution procedures.

2. We propose a solvable ϵ-approximation to the independent pessimistic bilevel problem, and we prove convergence to the original problem when ϵ approaches zero. While similar approximations have been suggested in the past, we provide a new condition that guarantees the convergence of our approximation.

3. We develop a solution procedure for the ϵ-approximations that does not require any convexity assumptions and that accommodates for integer lower-level and/or upper-level decisions. To the best of our knowledge, we propose the first direct solution scheme for the nonconvex pessimistic bilevel problem. We also provide a computational study that examines the numerical behavior of our algorithm.

In the related book chapter [82], we provide an introduction to bilevel optimization that illustrates some of the applications and computational challenges, and that outlines how bilevel problems can be solved. In this paper, we provide a formal justification for the conjectures made in [82], we examine the computational complexity of pessimistic bilevel problems, and we develop and analyze a solution scheme for these problems.

The remainder of this paper is structured as follows. In the next section, we study two structural properties of the pessimistic bilevel problem: the existence of optimal solutions and the computational complexity of the problem. In section 3 we develop a sequence of approximate problems that are solvable and whose optimal solutions converge to the optimal solutions of the pessimistic bilevel problem. We propose an iterative solution scheme for these approximations in section 4, and we demonstrate

the performance of our algorithm in section 5. We conclude in section 6. Extended numerical results can be found in the accompanying technical report [90].

2. Problem analysis. We start with some terminology. The *optimistic bilevel problem* results from problem $\mathcal{PB}$ if we replace the universal quantifier $\forall$ with an existential quantifier $\exists$. The optimistic bilevel problem can be equivalently formulated as follows:

$$\begin{aligned}
&\underset{x,y}{\text{minimize}} && f(x) \\
&\text{subject to} && g(x,y) \leq 0 \\
&&& y \in \mathcal{Y}(x) = \arg\min_{z} \{h(x,z) \,:\, z \in Y(x)\} \\
&&& x \in X.
\end{aligned}$$

If the set $Y(x)$ in an optimistic or pessimistic bilevel problem does not depend on x, that is, if $Y(x) = Y(x')$ for all $x, x' \in X$, then we call the problem *independent*. In this case, we denote by Y the set of feasible lower-level decisions. Note that the lower-level problem of an independent bilevel problem still depends on the upper-level decision x through the lower-level objective function h. We denote the independent pessimistic bilevel problem by $\mathcal{IPB}$. If $Y(x) \neq Y(x')$ for some $x, x' \in X$, then we call the bilevel problem *dependent*. Throughout this paper, we make the following regularity assumptions.

(A1) The sets X and $Y(x)$, $x \in X$, are compact.

(A2) The functions f, g, and h are continuous over their domains.

We allow the sets X and $Y(x)$, $x \in X$, as well as the functions f, g, and h, to be nonconvex. This implies that some or all of the upper-level and/or lower-level decisions may be restricted to integer values.

Some of our results hold under weaker conditions. However, the aim of this paper is to develop a numerical solution scheme for the pessimistic bilevel problem. Our algorithm requires the solution of global optimization subproblems, and assumptions (A1) and (A2) are required by virtually all global optimization procedures. For ease of exposition, we therefore do not present the most general statement of our results.

For our analysis in this section, we also define the *linear dependent pessimistic bilevel problem*:

(2.1)

$$\begin{aligned}
&\underset{x}{\text{minimize}} && c^\top x \\
&\text{subject to} && Ax + By \geq b && \forall y \in \mathcal{Y}(x) = \arg\min_{z \in \mathbb{R}_+^m} \{f^\top z \,:\, Cx + Dz \geq g\} \\
&&& x \in \mathbb{R}_+^n,
\end{aligned}$$

where $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{p \times n}$, $B \in \mathbb{R}^{p \times m}$, $b \in \mathbb{R}^p$, $f \in \mathbb{R}^m$, $C \in \mathbb{R}^{q \times n}$, $D \in \mathbb{R}^{q \times m}$, and $g \in \mathbb{R}^q$. In problem (2.1), the lower-level objective could additionally contain a linear term $d^\top x$ that depends on the upper-level decision x. Such a term would not change the set of lower-level minimizers, however, and we omit it for ease of exposition. We obtain the *linear dependent optimistic bilevel problem* if we replace the universal quantifier $\forall$ in the bilevel constraint with an existential quantifier $\exists$. We say that the linear optimistic or pessimistic bilevel problem is *independent* if the lower-level problem satisfies $C = 0$.

Bilevel problems are closely related to (generalized) semi-infinite programs [40, 78]. Under some mild conditions, a generalized semi-infinite program can be formulated as a dependent (optimistic or pessimistic) bilevel problem in which the lower-level objective maximizes the violation of the generalized semi-infinite constraint [77]. Similarly, under some mild conditions, a semi-infinite program can be formulated as an independent (optimistic or pessimistic) bilevel problem in which the lower-level objective maximizes the violation of the semi-infinite constraint. On the other hand, an optimistic (dependent or independent) bilevel problem can be formulated as a generalized semi-infinite program in which the generalized semi-infinite constraint ensures global optimality of the lower-level decision [77, 81]. Moreover, if the lower-level problem is convex and satisfies a constraint qualification, then the generalized semi-infinite constraint can be replaced with the optimality conditions for the lower-level problem, which results in a finite-dimensional single-level problem [1, 3, 37, 38]. The situation changes significantly when we consider the pessimistic bilevel problem. If the lower-level problem is convex and satisfies a constraint qualification, then the pessimistic bilevel problem can be formulated as a generalized semi-infinite program in which the upper-level constraint has to hold for all lower-level decisions that satisfy the optimality conditions for the lower-level problem. If the lower-level problem is strictly convex, then the optimality conditions have a unique solution, and we can reformulate the pessimistic bilevel problem as a finite-dimensional single-level problem [1, 3, 37, 38]. We are not aware of any reformulations that allow us to reduce nonconvex pessimistic bilevel problems to (generalized) semi-infinite programs, apart from the trivial case where the lower-level objective is constant.

In the remainder of this section, we first investigate under which conditions bilevel problems have optimal solutions. Afterwards, we analyze the computational complexity of bilevel problems.

2.1. Existence of optimal solutions. Unlike other optimization problems, a bilevel problem may not possess an optimal solution even though it is feasible and satisfies (A1) and (A2). This happens if the feasible region of the bilevel problem fails to be closed. In this section, we show under which conditions the existence of optimal solutions is guaranteed.

In the most benign setting, a bilevel problem has at most one optimal lower-level decision associated with each upper-level decision. In this case, a linear dependent bilevel problem is solvable if its feasible region is nonempty and bounded [24, Theorem 3.2]. Equally, a nonlinear dependent bilevel problem is solvable if its feasible region is nonempty and compact, and if the Mangasarian–Fromowitz constraint qualification holds at all feasible solutions [24, Theorem 5.1].

In general, however, a bilevel problem can have (infinitely) many optimal lower-level decision associated with some or all of the upper-level decisions. In this case, a linear dependent optimistic bilevel problem is solvable if its feasible region is nonempty and bounded [24, Theorem 3.3]. This result includes linear independent optimistic bilevel problems as a special case. On the other hand, the linear independent pessimistic bilevel problem is not solvable in general; see, e.g., the example presented in [24, section 3.3]. This implies that, in general, the linear dependent pessimistic bilevel problem is not solvable either.

A nonlinear dependent optimistic bilevel problem is solvable if its feasible region is nonempty and compact, and if the Mangasarian–Fromowitz constraint qualification holds at all feasible solutions [24, Theorem 5.2]. On the other hand, a nonlinear dependent pessimistic bilevel problem in the standard formulation (1.1) is solvable if its

feasible region is nonempty and compact, and if the set of lower-level optimal solutions is lower semicontinuous for all upper-level decisions; see, e.g., [24, Theorem 5.3] or [50]. To ensure lower semicontinuity of the set of lower-level optimal solutions, it is sufficient to assume that the feasible region of the bilevel problem is polyhedral, the upper-level objective function is continuous and the lower-level objective function is convex and weakly analytic [50]. Alternative sufficient conditions are presented in [43, 71, 88] and the references therein. As pointed out in [44], however, these assumptions are very strong, and they are not even satisfied for linear independent pessimistic bilevel problems.

We now present a set of alternative existence conditions that focus on the property of independence.

THEOREM 2.1 (existence of optimal solutions). *Assume that the assumptions* (A1) *and* (A2) *are satisfied and that the bilevel problem is feasible. Then the following properties hold:*

1. *The independent optimistic bilevel problem has an optimal solution.*
2. *The independent pessimistic bilevel problem has an optimal solution if the objective function h of the lower-level problem is (additively) separable. Otherwise, it does not have an optimal solution in general* [24].
3. *The dependent optimistic and pessimistic bilevel problems do not have optimal solutions in general, even if the objective function h of the lower-level problem is separable.*

Proof. We prove the three properties separately.

1. We note that the lower-level problem must be feasible for *some* upper-level decision x since it is assumed that the bilevel problem is feasible. Since the bilevel problem is independent, however, this implies that the lower-level problem must indeed be feasible for *all* upper-level decisions $x \in X$. The lower-level problem attains its optimal objective value for all $x \in X$ due to the extreme value theorem since h is continuous and Y is compact. For a fixed $x \in X$, we can therefore denote the optimal objective value of the lower-level problem by $h^*(x) = \min \{h(x,y) : y \in Y\}$. Using the well-known optimal value reformulation [26, 65, 68], the independent optimistic bilevel problem can then be expressed as follows:

$$
(2.2) \quad
\begin{aligned}
\underset{x,y}{\text{minimize}} \quad & f(x) \\
\text{subject to} \quad & g(x,y) \leq 0 \\
& h(x,y) \leq h^*(x) \\
& x \in X, \ y \in Y.
\end{aligned}
$$

The function h^* is continuous since it constitutes the minimum of continuous functions. Hence, the feasible region of problem (2.2) is compact, and due to the continuity of f we can employ the extreme value theorem to conclude that the independent optimistic bilevel problem has an optimal solution.

2. Assume that h is separable, that is, $h(x,y) = h_1(x) + h_2(y)$ for continuous functions $h_1 : X \mapsto \mathbb{R}$ and $h_2 : Y \mapsto \mathbb{R}$. The lower-level problem is feasible for *all* upper-level decisions $x \in X$ if and only if it is feasible for *some* upper-level decision x. If the lower-level problem is infeasible, then the existence of an optimal solution to the bilevel problem follows from the extreme value theorem since X is compact and f is continuous. Assume now that the lower-level problem is feasible. In this case, the lower-level problem attains its optimal objective value for all $x \in X$ due to the extreme value theorem since h_2 is continuous and Y is compact. For a fixed $x \in X$, we can therefore denote the optimal objective value of the lower-level problem by

$h^*(x) = h_1(x) + h_2^*$, where $h_2^* = \min\{h_2(y) : y \in Y\}$. Using again an optimal value reformulation, the set of optimal solutions to the lower-level problem can then be described as

$$\{y \in Y : h_1(x) + h_2(y) \le h^*(x)\} \;=\; \{y \in Y : h_2(y) \le h_2^*\}.$$

The set on the right-hand side of this identity does not depend on x, and it is compact since h_2 is continuous and Y is compact. If we denote this set by $\mathcal{Y}$, then we can reformulate the bilevel constraint as

$$\max\{g(x,y) : y \in \mathcal{Y}\} \le 0.$$

The expression on the left-hand side of this inequality is a continuous function of x since it constitutes the maximum of continuous functions. Hence, the feasible region of the independent pessimistic bilevel problem is compact, and due to the continuity of f we can employ the extreme value theorem to conclude that the independent pessimistic bilevel problem has an optimal solution if h is separable.

Section 3.3 in [24] presents an instance of the independent pessimistic bilevel problem with a nonseparable function h that has no optimal solution. This concludes the second part of the proof.

3. Consider the following dependent pessimistic bilevel problem:

$$\begin{aligned}
&\underset{x}{\text{maximize}} && x \\
&\text{subject to} && x \le y && \forall y \in \underset{z}{\arg\min}\{z : xz \le 0,\, z \in [-1,1]\} \\
& && x \in [-1,1].
\end{aligned}$$

For $x < 0$, the unique optimizer of the lower-level problem is $y = 0$, which implies that the upper-level constraint is satisfied. For $x \ge 0$, on the other hand, the unique optimizer of the lower-level problem is $y = -1$, which implies that the upper-level constraint is violated. As a result, the feasible region is $[-1,0)$, and the problem has no optimal solution. Moreover, since the lower-level problem always has a unique optimizer, the pessimistic and the optimistic formulation of this problem are equivalent, that is, the optimistic formulation has no optimal solution either. $\quad\square$

Theorem 2.1 confirms the conventional wisdom that pessimistic bilevel problems are less well-behaved than optimistic bilevel problems. It also shows that independent bilevel problems are more well-behaved than their dependent counterparts. In the next section we will complement these results with an analysis of the computational complexity of various formulations of the bilevel problem.

2.2. Computational complexity. Clearly, we expect the bilevel problem to be difficult to solve if the objective function f or the feasible regions X or $Y(x)$ are nonconvex. Unlike other optimization problems, however, the bilevel problem remains computationally challenging even in seemingly benign cases. In particular, the linear dependent bilevel problem is well known to be $\mathcal{NP}$-hard. Amongst others, this has been proven in [6, 9, 14]. These results are strengthened in [27, 39, 42, 55], where it is shown that the linear dependent bilevel problem is indeed strongly $\mathcal{NP}$-hard. In fact, verifying local optimality of a given candidate solution for the linear dependent bilevel problem is $\mathcal{NP}$-hard [55, 87], and it has been shown in [18] that linear dependent bilevel problems can possess exponentially many local optima. Nevertheless, linear bilevel problems can be solved in polynomial time if the number m of lower-level decision variables is considered constant. This result, which is repeated

below in Theorem 2.2, is discussed in [27, 28, 45, 46, 47]. The complexity of multilevel problems, which contain several layers of nested lower-level problems, is studied in [14, 27, 42]. More detailed reviews of the complexity of bilevel problems can be found in [19, 27, 69].

The following theorem compiles several known as well as some new results about the complexity of the linear bilevel problem. Our objective is to highlight the differences between the independent and the dependent formulation of the problem.

THEOREM 2.2 (linear bilevel problem). *Assume that* (A1) *holds, and consider the linear bilevel problem* (2.1).

1. *The independent optimistic and independent pessimistic formulation of problem* (2.1) *can be solved in polynomial time.*
2. *The dependent optimistic and dependent pessimistic formulation of* (2.1) *can be solved in polynomial time if* m, *the number of lower-level decision variables, is constant* [27]. *Otherwise, both formulations are strongly* $\mathcal{NP}$-*hard* [42].

Remark 2.3. The polynomial-time solvability of the independent formulation of problem (2.1) may seem trivial since the set of optimal lower-level solutions does not depend on the upper-level decision. However, it is easy to construct problem instances that possess infinitely many optimal lower-level solutions that are described by a polyhedron with exponentially many vertices. This is the case, for example, if $f = 0$, which implies that all solutions of the inequality system $Dz \geq g$ are optimizers in the lower-level problem. It is therefore not a priori clear whether the independent formulation of problem (2.1) can be solved efficiently.

Proof of Theorem 2.2. Consider the bilevel constraint of the linear independent optimistic bilevel problem:

$$\exists y \in \underset{z \in \mathbb{R}_+^m}{\arg\min} \left\{ f^\top z \,:\, Dz \geq g \right\} \,:\, Ax + By \geq b.$$

Define $f^\star = \min_{z \in \mathbb{R}_+^m} \left\{ f^\top z \,:\, Dz \geq g \right\}$ as the optimal value of the lower-level problem. Then $y \in \mathbb{R}_+^m$ is an optimal solution to the lower-level problem if and only if $Dy \geq g$ and $f^\top y \leq f^\star$. We therefore conclude that the constraints of the linear independent optimistic bilevel problem are satisfied if and only if

$$\exists y \in \mathbb{R}_+^m \,:\, Dy \geq g, \ f^\top y \leq f^\star, \ Ax + By \geq b.$$

We can then reformulate the linear independent optimistic bilevel problem as follows:

$$\begin{array}{ll} \underset{x,y}{\text{minimize}} & c^\top x \\ \text{subject to} & Ax + By \geq b \\ & Dy \geq g, \ f^\top y \leq f^\star \\ & x \in \mathbb{R}_+^n, \ y \in \mathbb{R}_+^m. \end{array}$$

This is a linear program whose size is polynomial in the length of the input data, that is, the vectors b, c, f, and g, as well as the matrices A, B, and D. We can use the ellipsoid method or interior point techniques to solve this problem in polynomial time [67].

Consider now the constraints of the linear independent pessimistic bilevel problem:

$$Ax + By \geq b \qquad \forall y \in \underset{z \in \mathbb{R}_+^m}{\arg\min} \left\{ f^\top z \,:\, Dz \geq g \right\}.$$

From the first part of this proof we know that $y \in \mathbb{R}_+^m$ is an optimal solution to the lower-level problem if and only if $Dy \geq g$ and $f^\top y \leq f^\star$. Hence, the bilevel constraint of the linear independent pessimistic bilevel problem is equivalent to

$$Ax + By \geq b \qquad \forall y \in \left\{ z \in \mathbb{R}_+^m \, : \, Dz \geq g, \ f^\top z \leq f^\star \right\}.$$

Let $A_i^\top$ and $B_i^\top$ denote the ith row of matrix A and B, respectively. We can then reformulate the constraints of the linear independent pessimistic bilevel problem as

$$A_i^\top x + \min_{y \in \mathbb{R}_+^m} \left\{ B_i^\top y \, : \, Dy \geq g, \ f^\top y \leq f^\star \right\} \geq b_i \qquad \forall i = 1, \dots, p.$$

Due to linear programming duality, the embedded minimization problem equals

$$\max_{\gamma, \delta} \left\{ g^\top \gamma - f^\star \delta \, : \, D^\top \gamma \leq B_i + f\delta, \ \gamma \in \mathbb{R}_+^q, \ \delta \in \mathbb{R}_+ \right\}.$$

We thus conclude that the ith constraint of the linear independent pessimistic bilevel problem is satisfied if and only if there is $\gamma \in \mathbb{R}_+^q$ and $\delta \in \mathbb{R}_+$ such that

$$A_i^\top x + g^\top \gamma - f^\star \delta \geq b_i, \ \ D^\top \gamma \leq B_i + f\delta.$$

This allows us to reformulate the linear independent pessimistic bilevel problem as

$$\begin{aligned}
\underset{x, \Gamma, \delta}{\text{minimize}} \quad & c^\top x \\
\text{subject to} \quad & Ax + \Gamma g - f^\star \delta \geq b \\
& \Gamma D \leq B + \delta f^\top \\
& x \in \mathbb{R}_+^n, \ \Gamma \in \mathbb{R}_+^{p \times q}, \ \delta \in \mathbb{R}_+^p.
\end{aligned}$$

Here, all inequalities are understood elementwise. Again, this is a linear program whose size is polynomial in the length of the input data, that is, the vectors b, c, f, and g, as well as the matrices A, B, and D. We thus conclude that the linear independent pessimistic bilevel problem can be solved in polynomial time.

The polynomial-time solvability of the linear dependent optimistic and pessimistic bilevel problem for constant m follows from [27], and the strong $\mathcal{NP}$-hardness for nonconstant m is shown in [42]. $\quad\square$

The proof of the first part of Theorem 2.2 deserves further attention as it employs duality theory to solve variants of the linear bilevel problem. Duality theory has been used previously to construct exact penalty functions for linear and quadratic bilevel problems; see, e.g., [2, 52, 89, 91]. In these methods, the duality gap of the lower-level problem is included in the upper-level objective function to determine an optimal upper-level solution that simultaneously optimizes the lower-level objective. In contrast, we use duality theory to equivalently reformulate the independent bilevel problem as a polynomial-time solvable single-level optimization problem. To this end, the property of independence turns out to be crucial.

Clearly, Theorem 2.2 also implies the strong $\mathcal{NP}$-hardness of the *nonlinear* dependent bilevel problem if the number m of lower-level variables is nonconstant. It turns out, however, that even the most well-behaved independent formulation of the nonlinear bilevel problem is already strongly $\mathcal{NP}$-hard.

THEOREM 2.4 (nonlinear bilevel problem). *Assume that* (A1) *holds, and consider the independent bilevel problem where f and g are affine, X and Y are polyhedral, and the objective function h of the lower-level problem is quadratic and strictly convex. The optimistic and the pessimistic formulation of this problem are strongly $\mathcal{NP}$-hard.*

Proof. If h is strictly convex, then the lower-level problem has a unique optimal solution for each upper-level decision $x \in X$, and the optimistic and pessimistic formulation of the bilevel problem are equivalent. We can therefore restrict ourselves to the pessimistic formulation in the following.

Our proof is based on a polynomial-time reduction to the strongly $\mathcal{NP}$-hard KERNEL problem [36]:

KERNEL Problem
INSTANCE An undirected graph $G = (V, E)$ with nodes $V = \{1, \ldots, n\}$ and edges $E \subseteq \{\{j, k\} : j, k \in V\}$.
QUESTION Does G contain a *kernel*, that is, is there a subset of nodes $K \subseteq V$ such that no nodes in K are adjacent and all nodes in $V \setminus K$ are adjacent to nodes in K?

We assume that G does not contain any isolated nodes, that is, $\{k \in V : \{j, k\} \in E\} \neq \emptyset$ for all $j \in V$. Indeed, if G only has isolated nodes, that is, if $E = \emptyset$, then $K = V$ is a trivial kernel. Otherwise, if some of the nodes in G are isolated, then the answer to KERNEL does not change if we remove those nodes.

Consider the following instance of the independent pessimistic bilevel problem:

(2.3)

$$
\begin{aligned}
\underset{x, \tau}{\text{minimize}} \quad & \tau \\
\text{subject to} \quad & x_j + x_k \leq 1 \qquad \forall \{j, k\} \in E \\
& \tau \geq \sum_{j \in V} y_{2j} \qquad \forall (y_1, y_2) \in \underset{(z_1, z_2)}{\arg\min} \\
& \qquad \qquad \qquad \left\{ \|x - z_1\|_2^2 + \tfrac{1}{2n} \|\mathrm{e} - z_2\|_2^2 \ : \ (z_1, z_2) \in Y \right\} \\
& x \in \mathbb{R}_+^n, \ \ \tau \in \mathbb{R}_+,
\end{aligned}
$$

where e denotes the n-dimensional vector of ones and

$$
Y = \left\{ (z_1, z_2) \in \mathbb{R}_+^n \times \mathbb{R}_+^n \ : \ z_{2j} \leq 1 - z_{1j} \ \ \forall j \in V, \ \ z_{2j} \leq 1 - z_{1k} \ \ \forall \{j, k\} \in E \right\}.
$$

This problem satisfies the assumptions of the theorem. We claim that the problem attains an optimal objective value of zero if and only if the graph G has a kernel. Moreover, we show that if G has a kernel, then there is an optimal solution x with $x_j = 1$ if $j \in V$ is part of the kernel and $x_j = 0$ otherwise. Intuitively, the first upper-level constraint ensures that no two adjacent nodes can be part of the kernel, whereas the bilevel constraint requires τ to exceed the number of nodes outside the kernel that are *not* adjacent to some kernel node. Indeed, we will see that any optimal solution (y_1, y_2) to the lower-level problem satisfies $y_1 = x$ and $y_{2j} = 1$ if $j \in V$ is a node outside the kernel that is not adjacent to any kernel node. This follows from the definition of Y: the first constraint requires y_{2j} to be zero if j is a kernel node, and the second constraint ensures that y_{2j} is zero if j is adjacent to any kernel node. Note that problem (2.3) is feasible for any graph G since $(x, \tau) = (0, n)$ always constitutes a feasible solution, the associated optimal lower-level solution being $(y_1, y_2) = (0, \mathrm{e})$. Also, the optimal objective value of problem (2.3) is nonnegative by construction.

Our proof consists of three steps. First, we show that for a given upper-level decision x, every optimal lower-level decision (y_1, y_2) satisfies $y_1 = x$. Afterwards,

we prove that every optimal (y_1, y_2) satisfies

$$(2.4) \qquad y_{2j} = \min\left\{1 - y_{1j}, \ \min_{\{j,k\} \in E}\{1 - y_{1k}\}\right\} \qquad \forall j \in V.$$

Finally, we combine (2.4) with an argument presented in [39] to show that the graph G has a kernel if and only if the optimal objective value of the bilevel problem (2.3) is zero.

In view of the first step, assume that (y_1, y_2) is an optimal lower-level decision that satisfies $y_{1l} = x_l + \delta$ for some $l \in V$ and $\delta \neq 0$. Consider the lower-level decision (y_1', y_2') defined through $y_{1l}' = x_l$, $y_{1j}' = y_{1j}$ for $j \in V \setminus \{l\}$, as well as $y_{2j}' = \max\{0, y_{2j} - |\delta|\}$ for $j \in V$. This decision satisfies $y_2' \geq 0$ by construction, as well as $y_1' \geq 0$ since $x \geq 0$ and $y_1 \geq 0$. One readily verifies that (y_1', y_2') also satisfies the other constraints of Y. Hence, $(y_1', y_2') \in Y$, that is, (y_1', y_2') is feasible in the lower-level problem. Moreover, we obtain that

$$h(x; y_1, y_2) - h(x; y_1', y_2') \ \geq \ \delta^2 - \frac{n}{2n}\delta^2 \ = \ \frac{1}{2}\delta^2 \ > \ 0,$$

where

$$h(x; y_1, y_2) = \|x - y_1\|_2^2 + \frac{1}{2n}\|e - y_2\|_2^2 \qquad \text{for } x \in X, \ (y_1, y_2) \in Y$$

denotes the objective function of the lower-level problem. We conclude that (y_1, y_2) is not optimal, and hence every optimal lower-level solution indeed satisfies $y_1 = x$.

As for the second step, assume that (y_1, y_2) is an optimal lower-level decision that does not satisfy (2.4). Since $(y_1, y_2) \in Y$, this implies that there is $l \in V$ and $\delta > 0$ such that

$$y_{2l} \leq \min\left\{1 - y_{1l}, \ \min_{\{l,k\} \in E}\{1 - y_{1k}\}\right\} - \delta.$$

Consider the lower-level decision (y_1, y_2') defined through $y_{2l}' = y_{2l} + \delta$ and $y_{2j}' = y_{2j}$ for $j \in V \setminus \{l\}$. By construction, we have that $(y_1, y_2') \in Y$. Moreover, since any $(z_1, z_2) \in Y$ satisfies $z_2 \leq e$ due to the first constraint of Y, the lower-level objective h is strictly decreasing in its last component vector over Y. Hence, we obtain that $h(x; y_1, y_2') < h(x; y_1, y_2)$, which contradicts the optimality of (y_1, y_2). We therefore conclude that every optimal lower-level decision (y_1, y_2) satisfies (2.4).

In view of the third step, assume that the graph G has a kernel $K \subseteq V$. In this case, x with $x_j = 1$ if $j \in K$ and $x_j = 0$ for $j \in V \setminus K$ satisfies the first upper-level constraint. Moreover, this choice of x ensures that $(y_1, y_2) = (x, 0)$ is feasible in the lower-level problem. Indeed, from the previous two steps we conclude that (y_1, y_2) is optimal in the lower-level problem. Thus, the optimal value of τ associated with our choice of x is zero, which implies that problem (2.3) attains an objective value of zero. Assume now that an optimal solution (x, τ) to problem (2.3) attains an objective value of zero. From the bilevel constraint we know that the optimal lower-level solution satisfies $y_2 = 0$. Employing (2.4), we see that this entails the existence of a set $K = \{j \in V : z_{1j} = 1\}$ such that every node in $V \setminus K$ is adjacent to at least one node of K. From the first part of this proof we furthermore know that $x = z_1$. Thus, the first upper-level constraint ensures that no nodes in K are adjacent to each other. In this case, however, K must constitute a kernel of G. $\qquad \square$

Theorems 2.2 and 2.4 complement the existing complexity results for bilevel problems by distinguishing between independent and dependent problems, as well as linear

and nonlinear problems. In particular, Theorem 2.2 shows that the linear independent bilevel problem can be solved in polynomial time, whereas its dependent counterparts is strongly $\mathcal{NP}$-hard. However, Theorem 2.4 shows that the tractability of the linear independent bilevel problem does not carry over to nonlinear variants.

3. Convergent ϵ-approximations. From now on, we will focus on the independent pessimistic bilevel problem $\mathcal{IPB}$. We know from Theorem 2.1 that the feasible region of this problem may not be closed, which implies that the problem may not be solvable. In order to derive a solvable optimization problem, we now consider an ϵ-approximation of $\mathcal{IPB}$:

$(\mathcal{IPB}(\epsilon))$

$$\begin{aligned}
&\underset{x}{\text{minimize}} && f(x) \\
&\text{subject to} && g(x,y) \leq 0 \quad \forall y \in \mathcal{Y}_\epsilon(x) = \{z \in Y \, : \, h(x,z) < h(x,z') + \epsilon \ \ \forall z' \in Y\} \\
& && x \in X.
\end{aligned}$$

Here, $\epsilon > 0$ is a parameter. If Y is empty, then the bilevel constraint is vacuously satisfied for all $x \in X$, and we can find an optimal solution to both the independent pessimistic bilevel problem $\mathcal{IPB}$ and any ϵ-approximation $\mathcal{IPB}(\epsilon)$ by minimizing $f(x)$ over X. In the following, we assume that Y is nonempty.

From a game-theoretic perspective, the approximation $\mathcal{IPB}(\epsilon)$ can be interpreted as a conservative version of a Stackelberg leader-follower game in which the leader accounts for all ϵ-optimal decisions of the follower. Apart from tractability considerations, problem $\mathcal{IPB}(\epsilon)$ is of interest in its own right for at least two reasons. First, the leader in a Stackelberg game may have *incomplete information*, that is, she may not know the values of all parameters in the follower's optimization problem. In this case, the leader may want to include some safety margin to hedge against deviations from the anticipated follower's decision. Second, the follower in a Stackelberg game may be constrained by *bounded rationality*, which implies that she may not be able to solve the lower-level optimization problem to global optimality. In that case, the leader may want to implement a decision that performs best in view of all ϵ-optimal decisions of the follower.

We now show that the ϵ-approximation $\mathcal{IPB}(\epsilon)$ has a closed feasible region for any $\epsilon > 0$.

PROPOSITION 3.1 (closedness of ϵ-approximation). *Assume that Y is nonempty and that* (A1) *and* (A2) *are satisfied. Then the feasible region of $\mathcal{IPB}(\epsilon)$ is closed for any $\epsilon > 0$.*

Proof. The Tietze extension theorem allows us to assume that the functions f, g, and h are continuous on the extended domains $f : \mathbb{R}^n \mapsto \mathbb{R}$ and $g, h : \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}$. To avoid notational clutter, we use the same symbols f, g, and h for the extended functions in this proof. We denote by $h^* : \mathbb{R}^n \mapsto \mathbb{R}$ the function that maps an upper-level decision to the value of an optimal lower-level decision, that is,

$$h^*(x) = \min_{y \in Y} h(x,y) \qquad \text{for } x \in \mathbb{R}^n.$$

The compactness of Y and the continuity of h guarantee that $h^*(x)$ is well-defined for all $x \in \mathbb{R}^n$. The set $\mathcal{Y}_\epsilon(x)$ of ϵ-optimal lower-level decisions in problem $\mathcal{IPB}(\epsilon)$ can now be reformulated as

$$\mathcal{Y}_\epsilon(x) = \{z \in Y \, : \, h(x,z) < h^*(x) + \epsilon\} \, .$$

Since the set X is closed, the feasible region of the ϵ-approximation $\mathcal{IPB}(\epsilon)$ is closed if the set

$$\mathcal{X}_\epsilon = \{x \in \mathbb{R}^n \,:\, g(x,y) \leq 0 \;\; \forall y \in \mathcal{Y}_\epsilon(x)\}$$

is closed. This set is closed if and only if its complement set

$$\overline{\mathcal{X}}_\epsilon = \{x \in \mathbb{R}^n \,:\, g(x,y) > 0 \;\; \text{for some } y \in \mathcal{Y}_\epsilon(x)\}$$

is open. To show that $\overline{\mathcal{X}}_\epsilon$ is indeed open, we take an element $\widehat{x}$ of this set and show that $x \in \overline{\mathcal{X}}_\epsilon$ for all elements x in the δ-ball $\mathcal{B}_\delta(\widehat{x})$ around $\widehat{x}$, where $\delta > 0$ is a constant that we will specify shortly.

Since $\widehat{x} \in \overline{\mathcal{X}}_\epsilon$, there is $\widehat{y} \in \mathcal{Y}_\epsilon(\widehat{x})$ such that $g(\widehat{x}, \widehat{y}) \geq \lambda_g$ for some constant $\lambda_g > 0$, as well as $h(\widehat{x}, \widehat{y}) \leq h^*(\widehat{x}) + \epsilon - \lambda_h$ for some constant $\lambda_h > 0$. We show that $g(x, \widehat{y}) > 0$ and $h(x, \widehat{y}) < h^*(x) + \epsilon$ for all $x \in \mathcal{B}_\delta(\widehat{x})$.

The continuity of g implies that there is indeed a constant $\delta_g > 0$ such that $g(x, \widehat{y}) > 0$ for all $x \in \mathcal{B}_{\delta_g}(\widehat{x})$. Similarly, the composite function $(x,y) \mapsto h(x,y) - h^*(x)$ is continuous because h^* inherits continuity from h. Hence, there is a constant $\delta_h > 0$ such that $h(x, \widehat{y}) < h^*(x) + \epsilon$ for all $x \in \mathcal{B}_{\delta_h}(\widehat{x})$, that is, $\widehat{y} \in \mathcal{Y}_\epsilon(x)$ for all $x \in \mathcal{B}_{\delta_h}(\widehat{x})$. We thus conclude that for $\delta \leq \min\{\delta_g, \delta_h\}$, we have $x \in \overline{\mathcal{X}}_\epsilon$ for all $x \in \mathcal{B}_\delta(\widehat{x})$. $\quad\square$

It is straightforward to generalize the ϵ-approximation $\mathcal{IPB}(\epsilon)$ to dependent pessimistic bilevel problems. However, the feasible region of the resulting approximate problems would not be closed in general. The property of independence therefore turns out to be central to the development of our approximation scheme.

We now investigate whether the ϵ-approximation $\mathcal{IPB}(\epsilon)$ converges in some suitable sense to the independent pessimistic bilevel problem $\mathcal{IPB}$. To this end, we define $\mathcal{Y}_0(x) = \mathcal{Y}(x)$ for $x \in X$, and we show that that mapping $\epsilon \mapsto \mathcal{Y}_\epsilon(x)$ is upper semicontinuous at zero, that is, the set of ϵ-optimal lower-level decisions in $\mathcal{IPB}(\epsilon)$ converges to the set $\mathcal{Y}(x)$ of optimal lower-level decisions in $\mathcal{IPB}$ as ϵ goes to zero.

LEMMA 3.2. *Assume that Y is nonempty and that* (A1) *and* (A2) *are satisfied. Then, for all $x \in X$ the set-valued mapping $\epsilon \mapsto \mathcal{Y}_\epsilon(x)$ is Hausdorff upper semicontinuous at zero:*

$$\forall \kappa > 0 \;\; \exists \overline{\epsilon} > 0 \quad such\ that \quad \forall \epsilon \in (0, \overline{\epsilon}], \; y \in \mathcal{Y}_\epsilon(x) \;\; \exists y' \in \mathcal{Y}(x) \,:\, \|y - y'\| \leq \kappa.$$

Proof. Fix some $x \in X$, and assume to the contrary that for some $\kappa > 0$ we have

$$\forall \overline{\epsilon} > 0 \;\; \exists \epsilon \in (0, \overline{\epsilon}], \; y \in \mathcal{Y}_\epsilon(x) \,:\, \|y - y'\| > \kappa \;\; \forall y' \in \mathcal{Y}(x).$$

For the sequence $\overline{\epsilon}_k = 1/k$, we can then construct sequences $\epsilon_k \in (0, \overline{\epsilon}_k]$ and $y_k \in \mathcal{Y}_{\epsilon_k}(x)$ such that

$$\|y_k - y'\| > \kappa \quad \forall y' \in \mathcal{Y}(x).$$

Since $\mathcal{Y}_{\epsilon_k}(x) \subseteq Y$ for all k and Y is bounded, we can apply the Bolzano–Weierstrass theorem to conclude that y_k has a convergent subsequence. Without loss of generality, we assume that y_k itself converges to y^*. By construction, the limit y^* satisfies

$$\|y^* - y'\| \geq \kappa \quad \forall y' \in \mathcal{Y}(x).$$

However, from the closedness of Y and the continuity of h we conclude that

$$y^* \in Y \quad \text{and} \quad h(x, y^*) \leq h(x, y) \;\; \forall y \in Y,$$

that is, $y^* \in \mathcal{Y}(x)$. This contradicts the fact that there is $\kappa > 0$ such that $\|y^* - y'\| \geq \kappa$ for all $y' \in \mathcal{Y}(x)$. We thus conclude that our assumption was wrong, that is, the assertion of the lemma is indeed valid. $\square$

We now show that the upper semicontinuity of the mapping $\epsilon \mapsto \mathcal{Y}_\epsilon(x)$ carries over to the bilevel constraint function. This is a not a new observation; a similar result can be found in [5, Theorem 4.2.2]. To keep the paper self-contained, however, we provide an independent proof in the following.

LEMMA 3.3. *Assume that Y is nonempty and that* (A1) *and* (A2) *are satisfied. Then, for all $x \in X$ the mapping $\epsilon \mapsto \sup\{g(x,y) : y \in \mathcal{Y}_\epsilon(x)\}$ is upper semicontinuous at zero:*

$$\forall \kappa > 0 \ \exists \overline{\epsilon} > 0 \quad such \ that \quad \sup_{y \in \mathcal{Y}_\epsilon(x)} g(x,y) \leq \sup_{y \in \mathcal{Y}(x)} g(x,y) + \kappa \quad \forall \epsilon \in (0, \overline{\epsilon}].$$

Proof. Fix some $x \in X$ and $\kappa > 0$. Since X and Y are compact, we can apply the Heine–Cantor theorem to conclude that g is uniformly continuous over its support $X \times Y$. Hence, there is $\delta_g > 0$ such that

$$\forall y, y' \in Y \ : \ \|y - y'\| \leq \delta_g \implies |g(x,y) - g(x,y')| \leq \kappa.$$

For this δ_g, we can now apply Lemma 3.2 to ensure that

$$\exists \overline{\epsilon} > 0 \quad such \ that \quad \forall \epsilon \in (0, \overline{\epsilon}], \ y \in \mathcal{Y}_\epsilon(x) \ \exists y' \in \mathcal{Y}(x) \ : \ \|y - y'\| \leq \delta_g.$$

Taken together, the last two statements imply that there is $\overline{\epsilon} > 0$ such that for all $\epsilon \in (0, \overline{\epsilon}]$, we have

$$\forall y \in \mathcal{Y}_\epsilon(x) \ \exists y' \in \mathcal{Y}(x) \quad such \ that \quad |g(x,y) - g(x,y')| \leq \kappa.$$

The assertion now follows from the continuity of g. $\square$

We now show that under certain conditions, the upper semicontinuity of the mapping $\epsilon \mapsto \sup\{g(x,y) : y \in \mathcal{Y}_\epsilon(x)\}$ implies that the optimal value of the ϵ-approximation $\mathcal{IPB}(\epsilon)$ "converges" (in some well-defined sense) to the optimal value of the independent pessimistic bilevel problem $\mathcal{IPB}$. We intentionally use quotation marks in this statement because $\mathcal{IPB}$ may not have an optimal value; see Theorem 2.1. We therefore consider a variant of $\mathcal{IPB}$ in which we replace the feasible region of $\mathcal{IPB}$ with its closure. By construction, this problem—which we refer to as cl($\mathcal{IPB}$)—has an optimal solution whenever its feasible region is nonempty. The following observation summarizes some basic relationships between $\mathcal{IPB}$ and cl($\mathcal{IPB}$).

Observation 3.4. Assume that (A1) and (A2) are satisfied. Then the two problems $\mathcal{IPB}$ and cl($\mathcal{IPB}$) satisfy the following properties:
1. cl($\mathcal{IPB}$) has a feasible solution if and only if $\mathcal{IPB}$ is feasible.
2. If $\mathcal{IPB}$ has an optimal solution x^*, then x^* is also optimal in cl($\mathcal{IPB}$).
3. If cl($\mathcal{IPB}$) has an optimal solution x^*, then there exists a sequence of feasible solutions $\{x_k\}_{k \in \mathbb{N}}$ for $\mathcal{IPB}$ that converges to x^*.

By a slight abuse of notation, we denote by cl($\mathcal{IPB}$) both the closed independent pessimistic bilevel problem and its optimal value (with the convention that cl($\mathcal{IPB}$) = $+\infty$ if the problem is infeasible). Likewise, we use $\mathcal{IPB}(\epsilon)$ to refer to either the ϵ-approximation of $\mathcal{IPB}$ or the optimal value of this approximation. In both cases, the meaning will be clear from the context.

If the independent pessimistic bilevel problem $\mathcal{IPB}$ is infeasible, then both the closed problem cl($\mathcal{IPB}$) and all approximate problems $\{\mathcal{IPB}(\epsilon)\}_{\epsilon > 0}$ are infeasible as

well. We now show that the approximate problems $\mathcal{IPB}(\epsilon)$ converge to $\mathrm{cl}(\mathcal{IPB})$ if the independent pessimistic bilevel problem $\mathcal{IPB}$ is feasible.

THEOREM 3.5 (convergence of ϵ-approximation). *Assume that Y is nonempty, that* (A1) *and* (A2) *are satisfied, that $\mathcal{IPB}$ is feasible, and that $\mathrm{cl}(\mathcal{IPB})$ has an optimal solution x^* that is not a local minimizer of the function $x \mapsto \sup\{g(x,y) : y \in \mathcal{Y}(x)\}$ over X with value zero. Then we have*

$$\lim_{\epsilon \searrow 0} \mathcal{IPB}(\epsilon) = \mathrm{cl}(\mathcal{IPB}).$$

Remark 3.6. The assumption that x^* is not a local minimizer of the function $x \mapsto \sup\{g(x,y) : y \in \mathcal{Y}(x)\}$ over X with value zero is reminiscent of a stability condition developed for global optimization problems; see [83]. It is equivalent to the requirement that x^* is the limit of a sequence of Slater points.

Proof of Theorem 3.5. By construction, the feasible region of $\mathcal{IPB}(\epsilon)$ is a subset of the feasible region of $\mathcal{IPB}$ and, a fortiori, $\mathrm{cl}(\mathcal{IPB})$. Hence, $\mathcal{IPB}(\epsilon) \geq \mathrm{cl}(\mathcal{IPB})$ for all $\epsilon > 0$, and we have only to show that

$$\forall \kappa > 0 \ \exists \overline{\epsilon} > 0 \quad \text{such that} \quad \mathcal{IPB}(\epsilon) \leq \mathrm{cl}(\mathcal{IPB}) + \kappa \quad \forall \epsilon \in (0, \overline{\epsilon}].$$

We distinguish between three different cases, depending on the value of the bilevel constraint in $\mathrm{cl}(\mathcal{IPB})$ at x^*. Assume first that $\sup\{g(x^*,y) : y \in \mathcal{Y}(x)\} < 0$, and fix some $\kappa > 0$. In this case, there is $\lambda > 0$ such that $\sup\{g(x^*,y) : y \in \mathcal{Y}(x)\} \leq -\lambda$, and we can invoke Lemma 3.3 to conclude that there is $\overline{\epsilon} > 0$ such that

$$\sup_{y \in \mathcal{Y}_\epsilon(x^*)} g(x^*,y) \leq 0 \quad \forall \epsilon \in (0, \overline{\epsilon}],$$

that is, x^* is also feasible in $\mathcal{IPB}(\epsilon)$ for all $\epsilon \in (0, \overline{\epsilon}]$. Since the problems $\{\mathcal{IPB}(\epsilon)\}_{\epsilon > 0}$ share the same objective function f with $\mathrm{cl}(\mathcal{IPB})$, we conclude that $\mathcal{IPB}(\epsilon) = \mathrm{cl}(\mathcal{IPB})$ for all $\epsilon \in (0, \overline{\epsilon}]$ if $\sup\{g(x^*,y) : y \in \mathcal{Y}(x)\} < 0$.

Assume now that $\sup\{g(x^*,y) : y \in \mathcal{Y}(x)\} = 0$, and fix some $\kappa > 0$. Since the objective function f in $\mathrm{cl}(\mathcal{IPB})$ and $\{\mathcal{IPB}(\epsilon)\}_{\epsilon > 0}$ is continuous, there is $\delta > 0$ such that $f(x) \leq f(x^*) + \kappa$ for all x in the δ-ball $\mathcal{B}_\delta(x^*)$ around x^*. Moreover, since x^* is not a local minimizer of the function $x \mapsto \sup\{g(x,y) : y \in \mathcal{Y}(x)\}$ over X, there is $\widehat{x} \in \mathcal{B}_\delta(x^*) \cap X$ such that $\sup\{g(\widehat{x},y) : y \in \mathcal{Y}(x)\} \leq -\lambda$ for some $\lambda > 0$. We can again invoke Lemma 3.3, this time to conclude that there is $\overline{\epsilon} > 0$ such that

$$\sup_{y \in \mathcal{Y}_\epsilon(\widehat{x})} g(\widehat{x},y) \leq 0 \quad \forall \epsilon \in (0, \overline{\epsilon}],$$

that is, $\widehat{x}$ is feasible in $\mathcal{IPB}(\epsilon)$ for all $\epsilon \in (0, \overline{\epsilon}]$. We therefore know that $\mathcal{IPB}(\epsilon) \in [\mathrm{cl}(\mathcal{IPB}), \mathrm{cl}(\mathcal{IPB}) + \kappa]$ for all $\epsilon \in (0, \overline{\epsilon}]$ if $\sup\{g(x^*,y) : y \in \mathcal{Y}(x)\}=0$. Since $\kappa > 0$ was chosen arbitrarily, the assertion follows.

Finally, the case $\sup\{g(x^*,y) : y \in \mathcal{Y}(x)\} > 0$ cannot arise because x^* is assumed to be feasible. $\quad\square$

Theorem 3.5 also provides a condition under which the ϵ-approximations $\mathcal{IPB}(\epsilon)$ are feasible.

COROLLARY 3.7. *Under the assumptions of Theorem* 3.5*, there is $\overline{\epsilon} > 0$ such that $\mathcal{IPB}(\epsilon)$ is feasible for all $\epsilon \in (0, \overline{\epsilon}]$.*

The assumptions in Theorem 3.5 are both necessary and sufficient. In fact, one can easily construct counterexamples where $\mathcal{IPB}(\epsilon)$ does not converge to $\mathrm{cl}(\mathcal{IPB})$ if any of the assumptions is violated.

Example 3.8. Consider the following instance of the independent pessimistic bilevel problem:

$$\begin{array}{ll} \underset{x}{\text{minimize}} & x \\ \text{subject to} & \max\left\{y - x - 2,\ 4 - x^2\right\} \le 0 \qquad \forall y \in \underset{y}{\arg\min}\left\{y\ :\ y \in [0,1]\right\} \\ & x \in \mathbb{R}. \end{array}$$

The function $x \mapsto \sup\left\{g(x,y)\ :\ y \in \mathcal{Y}(x)\right\}$ is nonpositive for $x \ge 2$, positive for $x \in (-\infty, 2) \setminus \{-2\}$ and zero for $x = -2$. Hence, the formulation $\mathcal{IPB}$ has its unique optimal solution at $x = -2$, which is a local minimizer of the function $x \mapsto \sup\left\{g(x,y)\ :\ y \in \mathcal{Y}(x)\right\}$ with value zero. In contrast, $x \mapsto \sup\left\{g(x,y)\ :\ y \in \mathcal{Y}_\epsilon(x)\right\}$ is nonpositive for $x \ge 2$ and positive for $x \in (-\infty, 2)$ for any $\epsilon > 0$. Hence, the approximate problems $\mathcal{IPB}(\epsilon)$ have their unique optimal solutions at $x = 2$, and they do not converge to the independent pessimistic bilevel problem $\mathcal{IPB}$ as ϵ goes to zero.

To sum up, the feasible region of the independent pessimistic bilevel problem $\mathcal{IPB}$ may not be closed, which implies that $\mathcal{IPB}$ does not have an optimal solution in general. We considered a variant cl($\mathcal{IPB}$) of $\mathcal{IPB}$ in which we replaced the feasible region of $\mathcal{IPB}$ with its closure. Problem cl($\mathcal{IPB}$) is feasible whenever $\mathcal{IPB}$ is, but cl($\mathcal{IPB}$) is guaranteed to have an optimal solution whenever it is feasible. We have also presented an ϵ-approximation $\mathcal{IPB}(\epsilon)$ of $\mathcal{IPB}$. Similar to cl($\mathcal{IPB}$), the approximate problems $\mathcal{IPB}(\epsilon)$ have optimal solutions whenever they are feasible, and we have shown that the optimal values of $\mathcal{IPB}(\epsilon)$ converge to the optimal value of cl($\mathcal{IPB}$) under some technical condition.

The idea of replacing bilevel problems with ϵ-approximations is not new; see, e.g., [23, 25, 31]. A Stackelberg game with polyhedral feasible regions for both players, quadratic leader objective function and linear follower objective function is studied in [53], and the authors present an ϵ-approximation to the problem. Approximations to generic nonconvex bilevel problems are developed by Molodtsov [66], as well as Loridan and Morgan; see, e.g., [48]. Our main contribution in this section is to provide a new condition that guarantees the convergence of our ϵ-approximation.

4. Iterative solution procedure. In this section, we fix a value $\epsilon > 0$ for the approximate problem $\mathcal{IPB}(\epsilon)$ and develop an iterative solution procedure for $\mathcal{IPB}(\epsilon)$ that is reminiscent of the discretization schemes used in the solution of semi-infinite programs. To this end, we first reformulate $\mathcal{IPB}(\epsilon)$ as an infinite-dimensional single-level problem.

PROPOSITION 4.1. *The approximate problem $\mathcal{IPB}(\epsilon)$ is equivalent to the infinite-dimensional problem*

(4.1)

$$\begin{array}{ll} \underset{x,z,\lambda}{\text{minimize}} & f(x) \\ \text{subject to} & \lambda(y) \cdot [h(x,z) - h(x,y) + \epsilon] + (1 - \lambda(y)) \cdot [g(x,y)] \le 0 \qquad \forall y \in Y \\ & x \in X,\ z \in Y,\ \lambda : Y \mapsto [0,1]\,, \end{array}$$

where the function $\lambda : Y \mapsto [0,1]$ is a decision variable.

Proof. By definition, the bilevel constraint in $\mathcal{IPB}(\epsilon)$ is equivalent to the following semi-infinite constraint:[1]

$$
\begin{aligned}
& [y \in \mathcal{Y}_\epsilon(x) \;\Rightarrow\; g(x,y) \le 0] && \forall y \in Y \\
(4.2) \qquad \Longleftrightarrow \quad & [y \notin \mathcal{Y}_\epsilon(x)] \;\vee\; [g(x,y) \le 0] && \forall y \in Y.
\end{aligned}
$$

From the definition of the set $\mathcal{Y}_\epsilon(x)$, we conclude that $y \notin \mathcal{Y}_\epsilon(x)$ if and only if

$$
\exists z \in Y \,:\, h(x,y) \ge h(x,z) + \epsilon.
$$

Satisfaction of the semi-infinite constraint (4.2) is therefore equivalent to the existence of $z \in Y$ such that

$$
[h(x,y) \ge h(x,z) + \epsilon] \;\vee\; [g(x,y) \le 0] \qquad \forall y \in Y.
$$

For a fixed lower-level decision $y \in Y$, this constraint can be reformulated as

$$
\exists \lambda \in [0,1] \,:\, \lambda\,[h(x,z) - h(x,y) + \epsilon] + (1-\lambda)\,[g(x,y)] \le 0.
$$

The assertion now follows if we introduce a different variable $\lambda(y)$ for each $y \in Y$. $\quad\square$

We propose to solve the infinite-dimensional optimization problem (4.1) with an iterative solution scheme that is inspired by the discretization techniques used in semi-infinite programming [15, 40]. Our solution scheme is described in Algorithm 1.

At its core, the algorithm solves a sequence of finite-dimensional approximations (4.1_k) of problem (4.1) that involve subsets of the constraints parametrized by $y \in Y$, as well as subsets of the decision variables $\lambda(y)$, $y \in Y$. Each of these approximations constitutes a relaxation of problem (4.1) in the sense that any feasible solution to (4.1) can be reduced to a feasible solution in (4.1_k), whereas it may not be possible to extend a feasible solution to (4.1_k) to a feasible solution in (4.1). Step 3 of Algorithm 1 aims to identify a constraint in problem (4.1) that cannot be satisfied by any extension of the optimal solution to the relaxation (4.1_k). If no such constraint exists, then the optimal solution to (4.1_k) can be extended to an optimal solution in (4.1), and the algorithm terminates. Otherwise, we refine the finite-dimensional approximation (4.1_k) and enter the next iteration. We now prove the correctness of the algorithm.

THEOREM 4.2. *Assume that* (A1) *and* (A2) *are satisfied. If Algorithm* 1 *terminates in step* 4 *of the kth iteration, then x_k can be extended to an optimal solution of problem* (4.1). *If Algorithm* 1 *does not terminate, then the sequence $\{x_k\}_{k \in \mathbb{N}}$ contains accumulation points, and any accumulation point of $\{x_k\}_{k \in \mathbb{N}}$ can be extended to an optimal solution of* (4.1).

Proof. Assume that Algorithm 1 terminates in step 4 of the kth iteration. In that case, we have

$$
\begin{aligned}
\tau_k \le 0 \quad \Longleftrightarrow \quad & \max_{y \in Y} \; \min \left\{ \min_{z \in Y} h(x_k, z) - h(x_k, y) + \epsilon, \; g(x_k, y) \right\} \le 0 \\
\Longleftrightarrow \quad & \max_{y \in Y} \; \min_{\lambda \in [0,1]} \; \lambda \cdot \left[\min_{z \in Y} h(x_k, z) - h(x_k, y) + \epsilon \right] + (1-\lambda) \cdot [g(x_k, y)] \le 0.
\end{aligned}
$$

---

[1] The semi-infinite disjunctive constraint (4.2) bears some similarity to the exposition in [65]. In that paper, however, disjunctive constraints are used to enforce dependent second-stage constraints, whereas we employ them to enforce the bilevel constraint for all approximate second-stage solutions $y \in \mathcal{Y}_\epsilon(x)$.

Algorithm 1.

1. Initialization. Set $Y_0 = \emptyset$ (considered lower-level decisions) and $k = 0$ (iteration counter).
2. Master problem. Solve the following finite-dimensional approximation of problem (4.1):

   (4.1_k)

   $$\begin{aligned}
   \underset{x,z,\lambda}{\text{minimize}} \quad & f(x) \\
   \text{subject to} \quad & \lambda(y_k) \cdot [h(x,z) - h(x,y_k) + \epsilon] \\
   & \quad + (1 - \lambda(y_k)) \cdot [g(x,y_k)] \leq 0 \qquad \forall y_k \in Y_k \\
   & x \in X, \ \ z \in Y, \ \ \lambda : Y_k \mapsto [0,1].
   \end{aligned}$$

   Let (x_k, z_k, λ_k) denote an optimal solution to this problem.
3. Subproblem. Calculate the value $h_k = \min\{h(x_k,z) : z \in Y\}$ of the lower-level problem associated with x_k, and solve the problem

   $$\begin{aligned}
   \underset{\tau,y}{\text{maximize}} \quad & \tau \\
   \text{subject to} \quad & \tau \leq h_k - h(x_k,y) + \epsilon \\
   & \tau \leq g(x_k,y) \\
   & \tau \in \mathbb{R}, \ \ y \in Y.
   \end{aligned}$$

   Let (τ_k, y_k) denote an optimal solution to this problem.
4. Termination criterion. If $\tau_k \leq 0$, terminate: x_k solves the approximate problem $\mathcal{IPB}(\epsilon)$. Otherwise, set $Y_{k+1} = Y_k \cup \{y_k\}$, $k \to k+1$ and go back to step 2.

By construction, the last inequality is equivalent to

$$\exists \lambda : Y \mapsto [0,1] : \ \lambda(y) \cdot \left[\min_{z \in Y} h(x_k,z) - h(x_k,y) + \epsilon \right] + (1 - \lambda(y)) \cdot [g(x_k,y)] \leq 0 \quad \forall y \in Y,$$

that is, x_k can be extended to a feasible solution (x_k, z, λ) to problem (4.1) if we choose $z \in \arg\min\{h(x_k,y) : y \in Y\}$. Since problem (4.1_k) constitutes a relaxation of problem (4.1) and both problems share the same objective function, this implies that the solution (x_k, z, λ) is indeed optimal in problem (4.1).

Assume now that Algorithm 1 does not terminate. Since the sets X and Y are bounded, we can apply the Bolzano–Weierstrass theorem to conclude that the sequence $\{(x_k, y_k)\}_{k \in \mathbb{N}}$ generated by Algorithm 1 contains accumulation points. By selecting any accumulation point and possibly going over to subsequences, we can assume that the sequence $\{(x_k, y_k)\}_{k \in \mathbb{N}}$ itself converges to (x^*, y^*). The closedness of X and Y guarantees that $x^* \in X$ and $y^* \in Y$. We apply a similar reasoning as in Theorem 2.1 in [15] to show that x^* can be extended to a feasible solution (x^*, z, λ) to problem (4.1). Choosing again $z \in \arg\min\{h(x^*, y) : y \in Y\}$, we need to show that there is a function $\lambda : Y \mapsto [0,1]$ such that

$$\lambda(y) \cdot \left[\min_{z \in Y} h(x^*,z) - h(x^*,y) + \epsilon \right] + (1 - \lambda(y)) \cdot [g(x^*,y)] \leq 0 \quad \forall y \in Y.$$

Assume to the contrary that there is $\widehat{y} \in Y$ such that

$$\lambda \cdot \left[\min_{z \in Y} h(x^*, z) - h(x^*, \widehat{y}) + \epsilon\right] + (1 - \lambda) \cdot [g(x^*, \widehat{y})] \geq \delta$$

for all $\lambda \in [0, 1]$ and some $\delta > 0$. The continuity of g and h implies that for k sufficiently large, we have

$$\lambda \cdot \left[\min_{z \in Y} h(x_k, z) - h(x_k, \widehat{y}) + \epsilon\right] + (1 - \lambda) \cdot [g(x_k, \widehat{y})] \geq \delta'$$

for all $\lambda \in [0, 1]$ and some $\delta' > 0$. From the subproblem in step 3 of Algorithm 1, we can see that

$$\min \left\{ \min_{z \in Y} h(x_k, z) - h(x_k, y_k) + \epsilon, \; g(x_k, y_k) \right\}$$
$$\geq \min \left\{ \min_{z \in Y} h(x_k, z) - h(x_k, \widehat{y}) + \epsilon, \; g(x_k, \widehat{y}) \right\},$$

that is,

$$\lambda \cdot \left[\min_{z \in Y} h(x_k, z) - h(x_k, y_k) + \epsilon\right] + (1 - \lambda) \cdot [g(x_k, y_k)] \geq \delta'$$

for all $\lambda \in [0, 1]$. Taking the limit as k goes to infinity, the continuity of g and h implies that

$$\lambda \cdot \left[\min_{z \in Y} h(x^*, z) - h(x^*, y^*) + \epsilon\right] + (1 - \lambda) \cdot [g(x^*, y^*)] \geq \delta'$$

for all $\lambda \in [0, 1]$. However, by construction of (4.1_k), there is $\lambda \in [0, 1]$ such that

$$\lambda \cdot \left[\min_{z \in Y} h(x_{k+1}, z) - h(x_{k+1}, y_k) + \epsilon\right] + (1 - \lambda) \cdot [g(x_{k+1}, y_k)] \leq 0$$

in iteration $k + 1$ of the algorithm. Taking the limit as k goes to infinity, we have

$$\lambda \cdot \left[\min_{z \in Y} h(x^*, z) - h(x^*, y^*) + \epsilon\right] + (1 - \lambda) \cdot [g(x^*, y^*)] \leq 0$$

for some $\lambda \in [0, 1]$ since the sequence $\{x_{k+1}\}_{k \in \mathbb{N}}$ also converges to x^*. This yields a contradiction, and we conclude that there is a function $\lambda : Y \mapsto [0, 1]$ such that

$$\lambda(y) \cdot \left[\min_{z \in Y} h(x^*, z) - h(x^*, y) + \epsilon\right] + (1 - \lambda(y)) \cdot [g(x^*, y)] \leq 0 \quad \forall y \in Y,$$

that is, x^* can indeed be extended to a feasible solution (x^*, z, λ) to problem (4.1). We can now apply the same reasoning as in the case of finite termination to conclude that (x^*, z, λ) actually solves (4.1). $\quad\square$

Remark 4.3. Algorithm 1 is an example of an *outer approximation scheme* since it constructs a sequence of infeasible solutions to problem $\mathcal{IPB}(\epsilon)$ that converges to an optimal solution of $\mathcal{IPB}(\epsilon)$. In practice, Algorithm 1 is stopped as soon as the termination criterion in step 4 is met approximately (i.e., up to some tolerance).

Theorem 4.2 provides a formal justification for this approach in the sense that after sufficiently many iterations of the algorithm, the resulting solution can be expected to be close to an optimal solution of $\mathcal{IPB}(\epsilon)$. Stronger guarantees can be obtained if problem $\mathcal{IPB}(\epsilon)$ is solved with an *inner approximation scheme.* Under the assumptions of Theorem 3.5, $\mathcal{IPB}(\epsilon)$ inherits Slater points from $\mathcal{IPB}$ if ϵ is sufficiently small; see also Remark 3.6. We can then employ the reasoning in the proof of Proposition 4.1 to equivalently reformulate problem $\mathcal{IPB}(\epsilon)$ as the following semi-infinite program:

$$
\begin{aligned}
& \underset{x,z}{\text{minimize}} && f(x) \\
& \text{subject to} && \min\left\{h(x,z) - h(x,y) + \epsilon,\ g(x,y)\right\} \leq 0 && \forall y \in Y \\
& && x \in X,\ \ z \in Y,\ \ \lambda : Y \mapsto [0,1].
\end{aligned}
$$

One could try to solve this semi-infinite program with inner approximation schemes such as the adaptive convexification algorithm presented in [34], solution approaches using interval analysis [10, 11], or methods relying on clever restrictions of the constraint right-hand sides [60, 80]. Note, however, that additional complications may arise due to the nonsmoothness of the constraint function. For simplicity, we restrict ourselves in the following to the outer approximation scheme presented in Algorithm 1.

We close with an extension of Algorithm 1 that we will use in our numerical study.

Remark 4.4 (tightened master problem). If we evaluate the semi-infinite constraint in problem (4.1) at $y = z$, then we obtain

$$
\lambda(z) \cdot \epsilon + (1 - \lambda(z)) \cdot [g(x,z)] \leq 0.
$$

Since ϵ is strictly positive, any feasible solution to (4.1) must satisfy $\lambda(z) < 1$, that is, $g(x,z) \leq 0$. In contrast, an optimal solution (x_k, z_k, λ_k) to the finite-dimensional approximation (4.1_k) may not satisfy $g(x_k, z_k) \leq 0$ if $z_k \notin Y_k$. We therefore obtain a tighter approximation if we include the constraint $g(x,z) \leq 0$ in problem (4.1_k). Similar constraints are used in Karush–Kuhn–Tucker methods for generalized semi-infinite programs; see [79].

This seemingly insignificant extension of Algorithm 1 can lead to substantial performance improvements. To illustrate this, consider the following instance of the independent pessimistic bilevel problem $\mathcal{IPB}$:

$$
\begin{aligned}
& \underset{x}{\text{minimize}} && x \\
& \text{subject to} && x \geq y && \forall y \in \underset{z}{\arg\min}\left\{|z - 2x| \ :\ z \in [0,10]\right\} \\
& && x \geq 1.
\end{aligned}
$$

If we solve the approximate problem $\mathcal{IPB}(\epsilon)$ for $\epsilon = 0.1$, then Algorithm 1 terminates in iteration $k = 81$ with the optimal solution $x_{81} = 10$. Up to the penultimate iteration, the algorithm generates the sequence $x_k = 1 + 0.05k$, $z_k = 2x_k$, $Y_k = \{2x_l\ :\ l = 0, \ldots, k-1\}$, and $\lambda_k(y_k) = 1$ for all $y_k \in Y_k$. In each of these iterations, x_k is increased just enough so that the choice $z_k = 2x_k$ satisfies $h(x_k, z_k) - h(x_k, y_k) + \epsilon \leq 0$ for all lower-level decisions $y_k \in Y_k$ in the master problem (4.1_k). Intuitively, this choice of z_k ensures that none of the elements $y_k \in Y_k$ is identified as an ϵ-optimal lower-level decision for x_k. However, the resulting pairs (x_k, z_k) are not feasible in $\mathcal{IPB}(\epsilon)$ since they violate the constraint $g(x_k, z_k) \leq 0$. As a result, up to the penultimate iteration x_k violates the bilevel constraint $g(x_k, y_k) \leq 0$ for all considered lower-level decisions $y_k \in Y_k$, and the algorithm performs an exhaustive exploration of the set Y of feasible lower-level decisions.

If we include the constraint $g(x, z) \leq 0$ in problem (4.1_k), then Algorithm 1 terminates in iteration $k = 4$ with the same optimal solution $x_4 = 10$. In that case, the algorithm generates the sequence $x_k = \min \{2^k, 10\}$, $z_k \in [0, x_k]$, $Y_k = \{2x_l : l = 0, \ldots, k-1\}$, and $\lambda_k(y_k) = 0$ for all $y_k \in Y_k$. The algorithm now ensures that the resulting pairs (x_k, z_k) satisfy $g(x_k, z_k) \leq 0$ in $\mathcal{IPB}(\epsilon)$, that is, the choice $z_k = 2x_k$ is no longer feasible in (4.1_k). The upper-level decision x_k now satisfies $g(x_k, y_k) \leq 0$ for all $y_k \in Y_k$ in the master problem (4.1_k). Hence, the additional constraint $g(x, z) \leq 0$ avoids an exhaustive exploration of the set Y in this case.

5. Computational study. We applied an implementation of Algorithm 1 to the bilevel programming benchmark instances published by Mitsos and Barton [62]. The test set comprises 36 problem instances. Out of these, six instances do not contain upper-level decisions, and the respective lower-level problems have unique global minima. While our algorithm can be applied to these problems, we do not report the results since all of these instances are solved in the first iteration. In another 12 instances, the feasible region of the lower-level problem depends on the upper-level decision, which implies that our solution scheme cannot be used to solve these instances. In the following, we apply Algorithm 1 to the remaining 18 benchmark instances. The intermediate master problems and subproblems were solved to global optimality on an 8-core Intel Xeon machine with 2.33GHz clock speed and 8GB memory using GAMS 23.9 and BARON 11.1.[2,3] We used the standard settings of BARON, with the exception that we changed the optimality tolerance to 10^{-5}.

The numerical results are summarized in Table 5.1. From left to right, the columns of the table describe the name of the problem instance, the number of upper-level and lower-level decision variables, the optimal solution to the optimistic and the pessimistic bilevel problem, the parameter ϵ that specifies the approximation quality of problem $\mathcal{IPB}(\epsilon)$, as well as the number of iterations and the runtime in CPU seconds required by Algorithm 1. For each problem instance, we report the analytical solutions in the first one or two rows (italicized), while the numerical results determined by our solution scheme are given in the last one or two rows (in roman font).

The table shows that the optimal solutions to the optimistic and the pessimistic bilevel formulation coincide in seven instances: `mb_1_1_02`, `mb_1_1_03`, `mb_1_1_05`, `mb_1_1_08`, `mb_1_1_10`, `mb_1_1_14`, and `gf_4`. In the remaining 11 instances, the optimistic and the pessimistic formulation lead to different solutions. We discuss these instances in further detail in the accompanying technical report [90].

To further study the scalability of Algorithm 1, we consider a stylized production planning problem in which two companies A and B manufacture a set of products indexed by $i = 1, \ldots, n$. Company A is the market leader, and as such it has to choose its production quantities $x \in X = [0, 10]^n$ first. Company B is a market follower, which means that it can observe the production quantities x before choosing its production levels $y \in Y = [0, 10]^n$. The product prices are given by a function $p : X \times Y \mapsto \mathbb{R}_+$, which is defined through

$$p_i(x, y) = 10 - \frac{x_i + y_i}{4} - \sum_{j \neq i} \frac{x_j + y_j}{8(n-1)}, \qquad i = 1, \ldots, n.$$

This expression reflects the assumption that the market clearing price for product i is decreasing in the cumulative supply $x_i + y_i$ of product i as well as the cumulative

---

[2]GAMS homepage: http://www.gams.com.

[3]BARON homepage: http://archimedes.cheme.cmu.edu/baron/baron.html.

TABLE 5.1

Numerical results for Algorithm 1. From left to right, the columns describe the problem instance, the number of upper-level and lower-level decisions, the optimal solution to the optimistic and the pessimistic bilevel problem, the parameter ϵ of the approximation $\mathcal{IPB}(\epsilon)$, as well as the number of iterations and the runtime in CPU seconds required by Algorithm 1. For each problem instance, the first one or two rows report the analytical solutions (italicized), while the last one or two rows document the numerical results for different values of ϵ (in roman font). Variables of value "∞" indicate infeasible problems, and α represents an arbitrarily small positive number.

Instance	n	m	Optimistic Obj.	Optimistic x	Optimistic y	Pessimistic Obj.	Pessimistic x	Pessimistic y	ϵ	Iter.	CPU
mb_0.1_03	0	1	*-1*	*-*	*-1*	*1*	*-*	*1*	-	-	-
			-	-	-	1	-	1	0.001	1	0.12
mb_1.1_01	1	1	*0*	*-0.567*	*0*	*∞*	*$\emptyset$*	*$\emptyset$*	-	-	-
			-	-	-	∞	$\emptyset$	$\emptyset$	0.001	5	1.86
mb_1.1_02	1	1	*-1*	*-1*	*-1*	*-1*	*-1*	*-1*	-	-	-
			-	-	-	-1	-1	-1	0.001	2	0.92
mb_1.1_03	1	1	*0.5*	*[0.1,1]*	*0.5*	*0.5*	*[0.1,1]*	*0.5*	-	-	-
			-	-	-	0.5	0.7256	0.5	0.001	3	1.54
mb_1.1_04	1	1	*-0.8*	*0*	*-0.8*	*0.5*	*(0,1]*	*0.5*	-	-	-
			-	-	-	0.5	1	0.5	0.001	3	0.95
mb_1.1_05	1	1	*0*	*0*	*0*	*0*	*0*	*0*	-	-	-
			-	-	-	0.0327	-0.0327	0	0.001	8	3.21
			-	-	-	0.0201	-0.0201	0	0.0001	10	4.44
mb_1.1_06	1	1	*-1*	*0*	*1*	*0*	*1*	*1*	-	-	-
			-	-	-	0	-1	-1	-	-	-
			-	-	-	0	-1	-1	0.001	8	2.96
mb_1.1_07	1	1	*0.25*	*0.25*	*0.5*	*0.25 + α² + α*	*0.25 + α*	*√0.25 + α*	-	-	-
			-	-	-	0.2507	0.2507	0.5007	0.001	4	1.87
mb_1.1_08	1	1	*0*	*-1*	*1*	*0*	*-1*	*-1*	-	-	-
			-	-	-	0	-1	-1	0.001	2	1.04
mb_1.1_09	1	1	*-2*	*-1*	*0*	*-2*	*-1*	*0*	-	-	-
			-2	*-0.5*	*-1*	-2	-1	0	0.001	3	1.50
mb_1.1_10	1	1	*0.1875*	*-0.25*	*±0.5*	*0.1875*	*-0.25*	*±0.5*	-	-	-
			-	-	-	0.1875	-0.2481	-0.4980	0.001	3	1.52
mb_1.1_11	1	1	*0.25*	*0.5*	*0*	*0.25 + α² + α*	*0.5 + α*	*0*	-	-	-
			-	-	-	0.2510	0.5010	0	0.001	2	1.96
mb_1.1_12	1	1	*-0.258*	*0.189*	*0.4347*	*0*	*[-1,0]*	*0*	-	-	-
			-	-	-	0	-0.2214	0	0.001	4	2.67
mb_1.1_13	1	1	*0.3125*	*0.5*	*0.5*	*0.3125 + 2α² + 1.5α*	*0.5 + α*	*0.5 + α*	-	-	-
			-	-	-	0.3135	0.5007	0.5007	0.001	4	1.90
mb_1.1_14	1	1	*0.2095*	*-0.5545*	*0.4554*	*0.2095*	*-0.5545*	*0.4554*	-	-	-
			-	-	-	0.2115	-0.5993	0.4599	0.001	4	1.47
			-	-	-	0.2100	-0.5320	0.4532	0.0001	7	2.39
mb_1.1_17	1	1	*-1.7550*	*0.2106*	*1.7990*	*-0.2929*	*0*	*0.2929*	-	-	-
			-	-	-	-0.2929	0	0.2929	0.001	10	4.37
mb_2.3_02	2	3	*-2.35*	*(-1,-1)*	*(1,1,-0.707)*	*-0.6838*	*(-√0.1,-1)*	*(-1,±1,0)*	-	-	-
			-	-	-	-0.6838	(-0.3162,-1)	(-1,-1,0)	0.001	10	4.26
gf_4	1	1	*9*	*3*	*5*	*9*	*3*	*5*	-	-	-
			-	-	-	9	3	5	0.001	2	0.94

TABLE 5.2

Numerical results for the scalability experiment. The tables show the numbers of iterations (left) and the runtimes in CPU seconds (right) for different numbers of products manufactured by the companies. All results constitute median values over 25 randomly generated problem instances. For the lower right problem ("both companies produce five products"), more than 50% of the generated instances exceeded the per-iteration time limit of 1,000 CPU seconds.

		# follower						# follower			
	1	2	3	4	5		1	2	3	4	5
# leader 1	2	2	2	3	3	# leader 1	0.13	0.17	0.18	0.23	0.27
2	3	5	6	5	6	2	0.68	0.75	0.71	0.91	0.98
3	7	11	14	14	18	3	1.02	7.66	56.53	178.80	220.91
4	6	15	17	19	21	4	1.19	58.17	709.90	1,423.72	1,602.41
5	7	20	17	22	–	5	1.06	192.08	1,501.79	2,325.03	–

supply $\sum_{j \neq i}(x_j + y_j)$ of the other products, that is, the goods are weak substitutes. Assuming that the production cost vectors for companies A and B are given by $c_A \in \mathbb{R}^n_+$ and $c_B \in \mathbb{R}^n_+$, respectively, the problem can then be formulated as follows:

$$
\begin{aligned}
&\underset{x,\tau}{\text{maximize}} && \tau \\
&\text{subject to} && \tau \leq [p(x,y) - c_A]^\top x && \forall y \in \mathcal{Y}(x) \\
& && x \in [0,10]^n, \\
&\text{where} && \mathcal{Y}(x) = \underset{z}{\arg\max}\left\{[p(x,z) - c_B]^\top z \,:\, z \in [0,10]^n\right\}.
\end{aligned}
$$

In this problem, both companies choose production quantities for the products that maximize the overall profit. In the following, we want study the impact of the number of upper-level and lower-level decisions on the tractability of the bilevel problem. To this end, we consider a variant of the production planning problem where we vary the number of products manufactured by each company independently. The results over 25 problem instances with uniformly distributed costs $c_{A,i}, c_{B,i} \sim \mathcal{U}[3,5]$ are shown in Table 5.2. While small instances of the problem can be solved with reasonable effort, the table reveals that the runtime of Algorithm 1 grows very quickly, in particular if the number of upper-level variables is increased.

6. Conclusions. We studied the pessimistic bilevel problem without convexity assumptions. We derived conditions that guarantee the existence of optimal solutions, and we investigated the computational complexity of various problem formulations. We then examined a sequence of approximate problems that are solvable and that converge (under some technical condition) to the pessimistic bilevel problem, and we developed an iterative solution scheme for these approximations. We identified an independence property that is essential for both the formulation and the solution of the approximate problems. To the best of our knowledge, we presented the first direct solution method for the pessimistic bilevel problem.

It would be interesting to see how our approach can be extended to instances of the pessimistic bilevel problem that do not possess the independence property. A promising starting point for such a study would be to investigate how our solution scheme can be combined with the algorithm presented in [65] for the dependent optimistic bilevel problem. Another avenue for future research is the development of local optimization procedures that avoid a discretization of the bilevel constraint and thus scale more gracefully with problem size. In particular, it would be instructive to investigate how inner approximation schemes such as [10, 11, 34, 60, 80] can be applied to our approximations of the pessimistic bilevel problem.

Acknowledgment. The authors wish to express their gratitude to the referees for their constructive criticism, which led to substantial improvements of the paper.

REFERENCES

[1] F. A. AL-KHAYYAL, R. HORST, AND P. M. PARDALOS, *Global optimization of concave functions subject to quadratic constraints: An application in nonlinear bilevel programming*, Ann. Oper. Res., 34 (1992), pp. 125–147.

[2] M. AMOUZEGAR AND K. MOSHIRVAZIRI, *A penalty method for linear bilevel programming problems*, in Multilevel Optimization: Algorithms and Applications, A. Migdalas, P. M. Pardalos, and P. Värbrand, eds., Kluwer Academic Publishers, Dordrecht, 1998, pp. 251–271.

[3] L. T. H. AN, P. D. TAO, N. N. CANH, AND N. V. THOAI, *DC programming techniques for solving a class of nonlinear bilevel programs*, J. Global Optim., 44 (2009), pp. 313–337.

[4] G. ANANDALINGAM AND T. L. FRIESZ, *Hierarchical optimization: An introduction*, Ann. Oper. Res., 34 (1992), pp. 1–11.

[5] B. BANK, J. GUDDAT, D. KLATTE, B. KUMMER, AND K. TAMMER, *Nonlinear Parametric Optimization*, Birkhäuser Verlag, Basel, Boston, Stuttgart, 1983.

[6] J. F. BARD, *Some properties of the bilevel programming problem*, J. Optim. Theory Appl., 68 (1991), pp. 371–378.

[7] J. F. BARD, *Practical Bilevel Optimization: Algorithms and Applications*, Kluwer Academic Publishers, Dordrecht, 1998.

[8] J. F. BARD, J. PLUMMER, AND J. C. SOURIE, *A bilevel programming approach to determining tax credits for biofuel production*, European J. Oper. Res., 120 (2000), pp. 30–46.

[9] O. BEN-AYED AND C. E. BLAIR, *Computational difficulties of bilevel linear programming*, Oper. Res., 38 (1990), pp. 556–560.

[10] B. BHATTACHARJEE, W. H. GREEN JR., AND P. I. BARTON, *Interval methods for semi-infinite programs*, Comput. Optim. Appl., 30 (2005), pp. 63–93.

[11] B. BHATTACHARJEE, P. LEMONIDIS, W. H. GREEN JR., AND P. I. BARTON, *Global solution of semi-infinite programs*, Math. Program., 103 (2005), pp. 283–307.

[12] W. F. BIALAS AND M. H. KARWAN, *Two-level linear programming*, Management Sci., 30 (1984), pp. 1004–1020.

[13] M. BJONDAL AND K. JORNSTEN, *The deregulated electricity market viewed as a bilevel programming problem*, J. Global Optim., 33 (2005), pp. 465–475.

[14] C. E. BLAIR, *The computational complexity of multi-level linear programs*, Ann. Oper. Res., 34 (1992), pp. 13–19.

[15] J. W. BLANKENSHIP AND J. E. FALK, *Infinitely constrained optimization problems*, J. Optim. Theory Appl., 19 (1976), pp. 261–281.

[16] G. M. BOLLAS, P. I. BARTON, AND A. MITSOS, *Bilevel optimization formulation for parameter estimation in vapor–liquid(–liquid) phase equilibrium problems*, Chemical Engineering Science, 64 (2009), pp. 1768–1783.

[17] J. BRACKEN AND J. T. McGILL, *Mathematical programs with optimization problems in the constraints*, Oper. Res., 21 (1973), pp. 37–44.

[18] P. H. CALAMAI AND L. N. VICENTE, *Generating linear and linear-quadratic bilevel programming problems*, SIAM J. Sci. Comput., 14 (1993), pp. 770–782.

[19] A. CHINCHULUUN, P. M. PARDALOS, AND H.-X. HUANG, *Multilevel (hierarchical) optimization: Complexity issues, optimality conditions, algorithms*, in Advances in Applied Mathematics and Global Optimization, D. Y. Gao and H. D. Sherali, eds., Springer, New York, 2009.

[20] P. A. CLARK, *Bilevel programming for steady-state chemical process design. 2. Performance study for nondegenerate problems*, Computers & Chemical Engineering, 14 (1990), pp. 99–109.

[21] B. COLSON, P. MARCOTTE, AND G. SAVARD, *An overview of bilevel optimization*, Ann. Oper. Res., 153 (2007), pp. 235–256.

[22] J.-P. CÔTÉ, P. MARCOTTE, AND G. SAVARD, *A bilevel modelling approach to pricing and fare optimisation in the airline industry*, J. Revenue and Pricing Management, 2 (2003), pp. 23–36.

[23] S. DASSANAYAKA, *Methods of Variational Analysis in Pessimistic Bilevel Programming*, Ph.D. thesis, Wayne State University, Detroit, MI, 2010.

[24] S. DEMPE, *Foundations of Bilevel Programming*, Kluwer Academic Publishers, Dordrecht, 2002.

[25] S. DEMPE, *Bilevel programming*, in Essays and Surveys in Global Optimization, C. Audet, P. Hansen, and G. Savard, eds., Springer, New York, 2005, pp. 165–193.

[26] S. DEMPE AND A. B. ZEMKOHO, *The bilevel programming problem: Reformulations, constraint qualifications and optimality conditions*, Math. Program., 2012, available online via Online First.

[27] X. DENG, *Complexity issues in bilevel linear programming*, in Multilevel Optimization: Algorithms and Applications, A. Migdalas, P. M. Pardalos, and P. Värbrand, eds., Kluwer Academic Publishers, Dordrecht, 1998, pp. 149–164.

[28] X. DENG, Q. WANG, AND S. WANG, *On the complexity of linear bilevel programming*, in Proceedings of the First International Symposium on Operations Research with Applications, 1995, pp. 205–212.

[29] N. P. FAÍSCA, V. DUA, B. RUSTEM, P. M. SARAIVA, AND E. N. PISTIKOPOULOS, *Parametric global optimisation for bilevel programming*, J. Global Optim., 38 (2007), pp. 609–623.

[30] N. P. FAÍSCA, P. M. SARAIVA, B. RUSTEM, AND E. N. PISTIKOPOULOS, *A multi-parametric programming approach for multilevel hierarchical and decentralised optimisation problems*, Comput. Manag. Sci., 6 (2009), pp. 377–397.

[31] D. FANGHÄNEL AND S. DEMPE, *Bilevel programming with discrete lower level problems*, Optimization, 58 (2009), pp. 1029–1047.

[32] J. FLIEGE AND L. N. VICENTE, *A multicriteria approach to bilevel optimization*, J. Optim. Theory Appl., 131 (2006), pp. 209–225.

[33] C. A. FLOUDAS AND P. M. PARDALOS, *Encyclopedia of Optimization*, 2nd ed., Springer, New York, 2009.

[34] C. A. FLOUDAS AND O. STEIN, *The adaptive convexification algorithm: A feasible point method for semi-infinite programming*, SIAM J. Optim., 18 (2007), pp. 1187–1208.

[35] D. FUDENBERG AND J. TIROLE, *Game Theory*, MIT Press, Cambridge, MA, 1991.

[36] M. R. GAREY AND D. S. JOHNSON, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W. H. Freeman and Co., San Francisco, 1979.

[37] Z. H. GÜMÜŞ AND C. A. FLOUDAS, *Global optimization of nonlinear bilevel programming problems*, J. Global Optim., 20 (2001), pp. 1–31.

[38] Z. H. GÜMÜŞ AND C. A. FLOUDAS, *Global optimization of mixed-integer bilevel programming problems*, Comput. Manag. Sci., 2 (2005), pp. 181–212.

[39] P. HANSEN, B. JAUMARD, AND G. SAVARD, *New branch-and-bound rules for linear bilevel programming*, SIAM J. Sci. Statist. Comput., 13 (1992), pp. 1194–1217.

[40] R. HETTICH AND K. O. KORTANEK, *Semi-infinite programming: Theory, methods, and applications*, SIAM Rev., 35 (1993), pp. 380–429.

[41] R. R. IYER AND I. E. GROSSMANN, *A bilevel decomposition algorithm for long-range planning of process networks*, Industrial & Engineering Chemistry Research, 37 (1998), pp. 474–481.

[42] R. G. JEROSLOW, *The polynomial hierarchy and a simple model for competitive analysis*, Math. Programming, 32 (1985), pp. 146–164.

[43] D. KLATTE, *On the lower semicontinuity of optimal sets in convex parametric optimization*, Math. Programming Stud., 10 (1979), pp. 104–109.

[44] D. KLATTE, *A sufficient condition for lower semicontinuity of solution sets of systems of convex inequalities*, Math. Programming Stud., 21 (1984), pp. 139–149.

[45] M. LABBÉ, P. MARCOTTE, AND G. SAVARD, *A bilevel model of taxation and its application to optimal highway pricing*, Management Sci., 44 (1998), pp. 1608–1622.

[46] M. LABBÉ, P. MARCOTTE, AND G. SAVARD, *On a class of bilevel programs*, in Nonlinear Optimization and Applications, Volume 2, D. di Pillo and F. Gianessi, eds., Kluwer Academic Publishers, Dordrecht, 2000, pp. 1–24.

[47] Y.-H. LIU AND T. H. SPENCER, *Solving a bilevel linear program when the inner decision maker controls few variables*, European J. Oper. Res., 81 (1995), pp. 644–651.

[48] P. LORIDAN AND J. MORGAN, *ε-regularized two-level optimization problems: Approximation and existence results*, in Proceedings of the Fifth French-German Optimization Conference S. Dolecki, ed., Lecture Notes in Math. 1405, Springer, Berlin, 1989, pp. 99–113.

[49] P. LORIDAN AND J. MORGAN, *Weak via strong Stackelberg problem: New results*, J. Global Optim., 8 (1996), pp. 263–287.

[50] R. LUCCHETTI, F. MIGNANEGO, AND G. PIERI, *Existence theorems of equilibrium points in Stackelberg games with constraints*, Optimization, 18 (1987), pp. 857–866.

[51] Z.-Q. LUO, J.-S. PANG, AND D. RALPH, *Mathematical Programs with Equilibrium Constraints*, Cambridge University Press, Cambridge, UK, 2008.

[52] Y. LV, *An exact penalty function method for solving a class of nonlinear bilevel programs*, J. Appl. Math. Inform., 29 (2011), pp. 1533–1539.

[53] A. M. MALYSHEV AND A. S. STREKALOVSKIĬ, *On a connection between a bilevel optimization and a nonlinear optimization problems*, Russian Math., 55 (2011), pp. 83–86.

[54] A. V. Malyshev and A. S. Strekalovskii, *Global search for pessimistic solution in bilevel problems*, in Proceedings of the Toulouse Global Optimization Workshop, S. Cafieri, B. G.-Tóth, E. M. T. Hendrix, L. Liberti, and F. Messine, eds., 2010, pp. 77–80.

[55] P. Marcotte and G. Savard, *Bilevel programming: A combinatorial perspective*, in Graph Theory and Combinatorial Optimization, D. Avis, A. Hertz, and O. Marcotte, eds., Springer, New York, 2005, pp. 191–217.

[56] A. Migdalas, *Bilevel programming in traffic planning: Models, methods and challenge*, J. Global Optim., 7 (1995), pp. 381–405.

[57] A. Migdalas, P. M. Pardalos, and P. Värbrand, eds., *Multilevel Optimization: Algorithms and Applications*, Springer, New York, 1998.

[58] J. A. Mirrlees, *The theory of moral hazard and unobservable behaviour: Part I*, Rev. of Econom. Stud., 66 (1999), pp. 3–21.

[59] A. Mitsos, *Global solution of nonlinear mixed-integer bilevel programs*, J. Global Optim., 47 (2010), pp. 557–582.

[60] A. Mitsos, *Global optimization of semi-infinite programs via restriction of the right-hand side*, Optimization, 60 (2011), pp. 1291–1308.

[61] A. Mitsos and P. I. Barton, *Issues in the Development of Global Optimization Algorithms for Bilevel Programs with a Nonconvex Inner Program*, Technical report, Massachusetts Institute of Technology, Cambridge, MA, 2006.

[62] A. Mitsos and P. I. Barton, *A Test Set for Bilevel Programs*, Technical report, Massachusetts Institute of Technology, Cambridge, MA, 2006.

[63] A. Mitsos, G. M. Bollas, and P. I. Barton, *Bilevel optimization formulation for parameter estimation in liquid–liquid phase equilibrium problems*, Chemical Engineering Science, 64 (2009), pp. 548–559.

[64] A. Mitsos, B. Chachuat, and P. I. Barton, *Towards global bilevel dynamic optimization*, J. Global Optim., 45 (2009), pp. 63–93.

[65] A. Mitsos, P. Lemonidis, and P. I. Barton, *Global solution of bilevel programs with a nonconvex inner program*, J. Global Optim., 42 (2008), pp. 475–513.

[66] D. A. Molodtsov, *The solution of a class of non-antagonistic games*, USSR Comput. Math. Math. Phys., 16 (1976), pp. 67–72.

[67] Y. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*, SIAM, Philadelphia, 1994.

[68] J. V. Outrata, *A note on the usage of nondifferentiable exact penalties in some special optimization problems*, Kybernetika, 24 (1988), pp. 251–258.

[69] P. M. Pardalos and X. Deng, *Complexity issues in hierarchical optimization*, in DIMACS Workshop on Mathematical Hierarchies and Biology B. Mirkin, F. R. McMorris, F. S. Roberts, and A. Rzhetsky, eds., American Mathematical Society, Providence, RI, 1997, pp. 219–224.

[70] E. N. Pistikopoulos, M. C. Georgiadis, and V. Dua, eds., *Multi-Parametric Programming: Theory, Algorithms, and Applications*, Vol. 1, Wiley-VCH, Weinheim, 2007.

[71] S. M. Robinson, *Stability theory for systems of inequalities. I, Linear systems*, SIAM J. Numer. Anal., 12 (1975), pp. 754–769.

[72] B. Rustem and M. Howe, *Algorithms for Worst-Case Design and Applications to Risk Management*, Princeton University Press, Princeton, NJ, 2002.

[73] J. H. Ryu, V. Dua, and E. N. Pistikopoulos, *A bilevel programming framework for enterprise-wide process networks under uncertainty*, Computers & Chemical Engineering, 28 (2004), pp. 1121–1129.

[74] M. P. Scaparra and R. L. Church, *A bilevel mixed-integer program for critical infrastructure protection planning*, Comput. Oper. Res., 35 (2008), pp. 1905–1923.

[75] A. Shapiro, *Perturbation theory of nonlinear programs when the set of optimal solutions is not a singleton*, Appl. Math. Optim., 18 (1988), pp. 215–229.

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

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

[78] G. Still, *Generalized semi-infinite programming: Theory and methods*, European J. Oper. Res., 119 (1999), pp. 301–313.

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

[80] A. Tsoukalas and B. Rustem, *A feasible point adaptation of the Blankenship and Falk algorithm for semi-infinite programming*, Optim. Lett., 5 (2011), pp. 705–716.

[81] A. Tsoukalas, B. Rustem, and E. N. Pistikopoulos, *A global optimization algorithm for generalized semi-infinite, continuous minimax with coupled constraints and bi-level problems*, J. Global Optim., 44 (2009), pp. 235–250.

[82] A. Tsoukalas, W. Wiesemann, and B. Rustem, *Global optimisation of pessimistic bi-level problems*, in Lectures on Global Optimization P. M. Pardalos and T. F. Coleman, eds., Fields Inst. Commun. 55, Americal Mathematical Society, Providence, RI, 2009, pp. 215–243.

[83] H. Tuy, *Convex programs with an additional reverse convex constraint*, J. Optim. Theory Appl., 52 (1987), pp. 463–486.

[84] H. Tuy, *Monotonic optimization: Problems and solution approaches*, SIAM J. Optim., 11 (2000), pp. 464–494.

[85] H. Tuy, A. Migdalas, and N. T. Hoai-Phuong, *A novel approach to bilevel nonlinear programming*, J. Global Optim., 38 (2007), pp. 527–554.

[86] L. N. Vicente and P. H. Calamai, *Bilevel and multilevel programming: A bibliographic review*, J. Global Optim., 5 (1994), pp. 291–306.

[87] L. N. Vicente, G. Savard, and J. Judice, *Descent approaches for quadratic bilevel programming*, J. Optim. Theory Appl., 81 (1994), pp. 379–399.

[88] D. W. Walkup and R. J.-B. Wets, *A Lipschitzian characterization of convex polyhedra*, Proc. Amer. Math. Soc., 23 (1969), pp. 167–173.

[89] D. J. White and G. Anandalingam, *A penalty function approach for solving bi-level linear programs*, J. Global Optim., 3 (1993), pp. 397–419.

[90] W. Wiesemann, A. Tsoukalas, P.-M. Kleniati, and B. Rustem, *Pessimistic Bi-level Optimization*, Technical report DTR12-4, Imperial College London, UK, 2012.

[91] Z. K. Xu, *Deriving the properties of linear bilevel programming via a penalty function approach*, J. Optim. Theory Appl., 103 (1999), pp. 441–456.