Heuristic Methods for Γ-Robust Mixed-Integer Linear Bilevel Problems

Yasmine Beck, Ivana Ljubić, and Martin Schmidt

Abstract. Due to their nested structure, bilevel problems are intrinsically hard to solve—even if all variables are continuous and all parameters of the problem are exactly known. In this paper, we study mixed-integer linear bilevel problems with lower-level objective uncertainty, which we address using the notion of Γ-robustness. To tackle the Γ-robust counterpart of the bilevel problem, we present heuristic methods that are based on the solution of a linear number of problems of the nominal type. Moreover, quality guarantees for heuristically obtained solutions as well as sufficient ex-post conditions for global optimality of the outcomes are provided. In an extensive computational study on 2240 instances, we assess the performance of our heuristics and compare them to alternative methods—both heuristic and exact—from the literature. We observe that the optimality gap is closed for a significant portion of the considered instances and that our methods often practically outperform alternative approaches in terms of the solution quality. Moreover, for the special case of Γ-robust interdiction problems, we report considerable speed-up factors when compared to recently published problem-tailored and exact solution approaches while also solving more instances to global optimality.

1. Introduction

Bilevel optimization is a rather young but very active field of research, having its game-theoretic roots dating back to the seminal publications of von Stackelberg (1932, 1954). Over the last years and decades, bilevel problems have gained increasing attention due to their ability to model hierarchical decision making processes. For an overview of the many applications of bilevel optimization, we refer to the annotated bibliography by Dempe (2020) as well as to the recent surveys by Kleinert et al. (2021) and by Beck et al. (2023b). The latter focuses on bilevel problems under uncertainty, which is also at the core of this paper.

Due to their hierarchical structure, bilevel problems are intrinsically hard to solve—even if all objective functions and constraints are linear, all variables are continuous, and all parameters of the problem are exactly known (Hansen et al. 1992). However, the situation becomes more challenging if, e.g., (i) discrete variables are introduced and (ii) problems under uncertainty are considered. In mathematical optimization, there are two main approaches to deal with uncertainties: stochastic optimization (Birge and Louveaux 2011; Kall and Wallace 1994) and robust optimization (Ben-Tal and Nemirovski 1998; Ben-Tal et al. 2009; Bertsimas et al. 2011; Soyster 1973). While, in the context of bilevel optimization, stochastic approaches to deal with uncertainties are more thoroughly studied, robust bilevel optimization is still in its infancy; see, e.g., Beck et al. (2022, 2023b) for more detailed discussions.

The contributions of this paper are the following. We consider mixed-integer linear bilevel problems with a binary lower-level problem that is affected by objective

Date: October 8, 2024.

2020 *Mathematics Subject Classification.* 90C11, 90C27, 90C57, 90C70.

Key words and phrases. Bilevel optimization, Robust optimization, Mixed-integer programming, Knapsack interdiction.

uncertainty. To deal with this kind of uncertainty, we pursue a Γ-robust approach (Bertsimas and Sim 2003; Sim 2004) in which the follower only hedges against a subset of the uncertain parameters that adversely influence the solution to the problem. In particular, we exploit the main result by Bertsimas and Sim (2003) and Sim (2004) for Γ-robust single-level problems—namely that the Γ-robust counterpart of a binary problem can be solved by solving a finite number of deterministic binary problems that is linear in the problem data. Exact approaches for Γ-robust min-max problems have been presented in our previous work (Beck et al. 2023a). Moreover, the heuristics proposed in DeNegre (2011) and Fischetti et al. (2018) can be applied to specific classes of Γ-robust interdiction problems after some modifications. However, we are not aware of any general-purpose methods in the literature that can tackle mixed-integer linear bilevel problems with a Γ-robust follower. Due to the overall hardness of the considered problems, which are Σ_p^2-hard in general (see Jeroslow (1985) for the first results on multilevel problems in the context of the polynomial hierarchy and Grüne and Wulf (2024) for very recent developments in this area), we thus study primal heuristics for these problems. We present such heuristics that have the following special properties: They (i) do not require problem-specific tailoring as they rely on solving a linear number of bilevel problems of the nominal type, they (ii) allow to use state-of-the-art as well as off-the-shelf solvers for the solution of these problems, they (iii) provide dual bounds from which ex-post quality guarantees can be derived, and they (iv) support a parallelization of the solution of the nominal problems. The latter aspects can make a huge difference when considering Γ-robust bilevel problems computationally. First, in our numerical study, we observe that our heuristics frequently outperform alternative approaches adapted from the literature in terms of the solution quality. In particular, our methods solve a considerable number of instances to global optimality. Second, for the special case of Γ-robust interdiction problems, we can find significant speed-up factors if our method is used. Finally, let us comment on another design principle of our heuristics. As mentioned above, the bilevel problems considered in this paper are Σ_p^2-hard in general. Usually, if one designs primal heuristics for hard problems, one aims to devise methods that produce primal feasible points quickly, i.e., one aims to resort to solving problems that are formally easier than the original problem. From a formal complexity-theoretical point of view, this is not the case for our heuristics since we iteratively solve mixed-integer bilevel problems of the nominal type. Although we suspect that the latter are on the same level of the polynomial hierarchy, they are easier to solve in a practical sense as they are of the nominal (and not of the robust) type anymore. In particular, this allows to exploit the sub-problems' structure and existing solution approaches for these sub-problems.

The remainder of this paper is organized as follows. In Section 2, we describe the overall problem statement and present the main result by Bertsimas and Sim (2003) and Sim (2004), which we apply to the Γ-robust lower-level problem. In Section 3, we focus on the special case of Γ-robust mixed-integer linear min-max problems for which we present a heuristic that is based on solving a linear number of bilevel problems of the nominal type. The latter is extended to the general Γ-robust bilevel setting in Section 4. In Section 5, we perform an extensive computational study to assess the performance of the heuristic methods presented in this paper. Finally, we derive conclusions in Section 6.

2. Problem Statement

In this paper, we consider mixed-integer linear bilevel problems of the form

$$\begin{aligned} \min_{x,y} \quad & c^\top x + d^\top y \\ \text{s.t.} \quad & x \in X, \\ & y \in \arg\max_{y'} \left\{ f^\top y' \colon y' \in Y(x) \right\}, \end{aligned} \tag{BMIP}$$

where $Y(x) \subseteq \{0,1\}^{n_y}$ and $X := \{x \in \mathbb{R}^{n_C} \times \mathbb{Z}^{n_D} \colon Ax \geq a\}$ with $n_x = n_C + n_D$, $c \in \mathbb{R}^{n_x}$, d, $f \in \mathbb{R}^{n_y}$, $A \in \mathbb{R}^{m \times n_x}$, and $a \in \mathbb{R}^m$. We refer to the first two lines of (BMIP) as the upper-level (or the leader's) problem. The last constraint in (BMIP) is the so-called lower-level (or follower's) problem. The variables x and y are the leader's and the follower's variables, respectively. Here, we consider the optimistic approach to bilevel optimization; see, e.g., Dempe (2002). This means that, whenever the set of optimal solutions to the lower-level problem is not a singleton, the follower decides such as to favor the leader w.r.t. her[1] objective function value. This is expressed in (BMIP) by optimizing not only over the leader's variables x but also over the follower's variables y. Throughout this paper, the following will be a standing assumption.

Assumption 1. *(i) The shared constraint set $\{(x, y) \colon x \in X, y \in Y(x)\}$ is nonempty and compact.*

(ii) All linking variables, i.e., all variables of the leader that appear in the lower-level constraints, are bounded integers.

Assumption 1 is necessary to ensure that (BMIP) has a solution; see, e.g., Section 5.1 in Kleinert et al. (2021) and the references therein for a detailed discussion. For $x \in X$, we further define the lower-level optimal-value function

$$\Phi(x) = \max_{y} \left\{ f^\top y \colon y \in Y(x) \right\} \tag{1}$$

to re-write (BMIP) as the single-level problem

$$\begin{aligned} \min_{x,y} \quad & c^\top x + d^\top y \\ \text{s.t.} \quad & x \in X,\ y \in Y(x), \\ & f^\top y \geq \Phi(x). \end{aligned}$$

In this paper, we are interested in bilevel problems of the above form, which are, however, affected by lower-level data uncertainty. We focus on uncertainties in the lower-level objective function coefficients, i.e., for all $i \in [n_y] := \{1, \dots, n_y\}$, we consider the coefficients $\bar{f}_i$ with $\bar{f}_i \in [f_i - \Delta f_i, f_i]$ instead of f_i. Here, we denote f_i as the nominal value of the ith lower-level objective function coefficient and Δf_i as its maximum deviation from the nominal value. For a discussion of the case with a certain objective function and uncertainties in a single packing-type constraint in the lower level, we refer to Beck et al. (2023a).

To deal with lower-level data uncertainty, we pursue a Γ-robust approach (Bertsimas and Sim 2003, 2004) in which the follower hedges against at most $\Gamma \in [n_y]$ deviations in his objective function coefficients. This leads us to considering the bilevel problem

$$\begin{aligned} \min_{x,y} \quad & c^\top x + d^\top y \\ \text{s.t.} \quad & x \in X,\ y \in S_\Gamma(x), \end{aligned} \tag{Rob-BMIP}$$

[1] Throughout this paper, we use "her" for the leader and "his" for the follower.

where $S_\Gamma(x)$ is the set of optimal solutions to the x-parameterized Γ-robust lower-level problem

$$\max_y \quad f^\top y - \max_{\{S\subseteq[n_y]:\, |S|\le\Gamma\}} \sum_{i\in S} \Delta f_i y_i \quad \text{s.t.} \quad y \in Y(x).$$

For a feasible upper-level decision $x \in X$, we define the optimal-value function of the Γ-robust lower level as

$$\Phi_{\text{rob}}(x) = \max_{y\in Y(x)} \left\{ f^\top y - \max_{\{S\subseteq[n_y]:\, |S|\le\Gamma\}} \sum_{i\in S} \Delta f_i y_i \right\} \tag{2}$$

such that the Γ-robust counterpart (Rob-BMIP) of the bilevel problem can be written as

$$\begin{aligned} \min_{x,y} \quad & c^\top x + d^\top y \\ \text{s.t.} \quad & x \in X,\, y \in Y(x), \\ & f^\top y - \max_{\{S\subseteq[n_y]:\, |S|\le\Gamma\}} \sum_{i\in S} \Delta f_i y_i \ge \Phi_{\text{rob}}(x). \end{aligned}$$

For the validity of the techniques we present in this paper, we further impose the following assumption throughout the remainder of the paper.

Assumption 2. *(i) The deviations are non-negative, i.e.,* $\Delta f_i \ge 0$ *for all* $i \in [n_y]$.
(ii) The indices are ordered such that the deviations are given in non-increasing order, i.e., $\Delta f_i \ge \Delta f_{i+1}$ *for all* $i \in [n_y]$ *with* $\Delta f_{n_y+1} = 0$.

Note that Assumption 2 is w.l.o.g. but necessary to exploit Theorem 3 in Bertsimas and Sim (2003), which is what we do in the next lemma.

Lemma 1. *Let* $x \in X$ *be a feasible upper-level decision. Then, solving the* Γ*-robust counterpart* (2) *of the lower-level problem is equivalent to solving* $n_y + 1$ *problems of the nominal type, i.e.,*

$$\Phi_{rob}(x) = \max_{\ell\in[n_y+1]} \{\Phi_\ell(x)\}$$

holds, where for all $\ell \in [n_y + 1]$, *we have*

$$\Phi_\ell(x) = -\Gamma\Delta f_\ell + \max_{y\in Y(x)} \left\{ \tilde f(\ell)^\top y \right\} \tag{3}$$

with

$$\tilde f(\ell)_i = \begin{cases} f_i - (\Delta f_i - \Delta f_\ell), & 1 \le i \le \ell, \\ f_i, & \ell + 1 \le i \le n_y. \end{cases}$$

Lemma 1 can be shown in analogy to the proof of Theorem 3 in Bertsimas and Sim (2003). In Miranda et al. (2013), the authors present an improvement of the Bertsimas–Sim result by reducing the number of problems of the nominal type to be solved to $n_y - \Gamma + 2$. Further reductions have been established in Theorem 1 in the paper by Lee and Kwon (2014) by showing that it suffices to solve

$$\Phi_{\text{rob}}(x) = \max_{\ell\in\mathcal{L}} \{\Phi_\ell(x)\}, \tag{4}$$

with

$$\mathcal{L} = \{\Gamma+1, \Gamma+3, \Gamma+5, \dots, \Gamma+\gamma, n_y+1\} \tag{5}$$

and γ being the largest odd integer such that $\Gamma + \gamma < n_y + 1$ holds. Hence, only $\lceil (n_y - \Gamma)/2 \rceil + 1$ problems of the nominal type need to be considered. We will hold on to the result of Theorem 1 by Lee and Kwon (2014) throughout this paper. As we will show in Proposition 3, we can further assume, w.l.o.g., that the index set $\mathcal{L}$ is given such that the deviations $(\Delta f_\ell)_{\ell\in\mathcal{L}}$ are pairwise distinct.

TABLE 1. Central Notation.

$\Phi : X \to \mathbb{R}$	Optimal-value function of the nominal lower level; see (1)
$\Phi_{\text{rob}} : X \to \mathbb{R}$	Optimal-value function of the Γ-robust lower level; see (2)
$\Phi_\ell : X \to \mathbb{R}$	Optimal-value function of the ℓth lower-level sub-problem; see (3)
$v_{\text{rob}} \in \mathbb{R}$	Optimal objective value of the Γ-robust min-max problem; see (Rob-Min-Max)
$v_\ell \in \mathbb{R}$	Optimal objective value of the ℓth deterministic min-max sub-problem; see (ℓ-Min-Max)

3. Mixed-Integer Linear Min-Max Problems

In this section, we focus on mixed-integer linear min-max problems as a special case of (BMIP). To this end, we set $d = f$, i.e., in its deterministic form, we consider the bilevel problem

$$\begin{aligned} \min_x \quad & c^\top x + f^\top y \\ \text{s.t.} \quad & x \in X, \\ & y \in \arg\max_{y'} \left\{ f^\top y' : y' \in Y(x) \right\}. \end{aligned} \tag{Min-Max}$$

Here, we do not need to distinguish between an optimistic and a pessimistic follower since the follower's response always yields the worst-possible outcome for the leader. Using the lower-level optimal-value function (1), we obtain a single-level reformulation of (Min-Max) that is given by

$$\min_{x \in X} \left\{ c^\top x + \Phi(x) \right\}.$$

The Γ-robust counterpart of the problem in which the follower hedges against at most Γ deviations in his uncertain objective function coefficients is obtained by replacing $\Phi(x)$ with $\Phi_{\text{rob}}(x)$ as stated in (1) and (4), i.e.,

$$v_{\text{rob}} := \min_{x \in X} \left\{ c^\top x + \Phi_{\text{rob}}(x) \right\} = \min_{x \in X} \left\{ c^\top x + \max_{\ell \in \mathcal{L}} \left\{ \Phi_\ell(x) \right\} \right\}. \tag{Rob-Min-Max}$$

In Section 3.1, we present a heuristic for (Rob-Min-Max) that follows the ideas of the main result by Bertsimas and Sim (2003) and Sim (2004). We provide quality guarantees for heuristically obtained solutions in Section 3.2. In Sections 3.3–3.5, we discuss algorithmic refinements and sufficient conditions for the exactness of our method (parallelization, reducing the number of sub-problems to be solved, and special techniques for interdiction problems). For the ease of presentation, a summary of the central notation used in this section can be found in Table 1.

3.1. **A Heuristic in the Spirit of Bertsimas and Sim.** To the best of our knowledge, there are currently no methods in the literature that can tackle (Rob-Min-Max) directly except for the problem-tailored exact approaches discussed in Beck et al. (2023a). The heuristic we present in this section does not require problem-specific tailoring so that any off-the-shelf solver for the nominal problem can be used within our framework. As a motivation for our method, we start with the following.

Proposition 1. *For all $\ell \in \mathcal{L}$, let*

$$v_\ell := \min_{x \in X} \left\{ c^\top x + \Phi_\ell(x) \right\}. \tag{ℓ-Min-Max}$$

Then, $v_{rob} \geq v_\ell$ holds, i.e., v_ℓ is a valid lower bound for the optimal objective function value of (Rob-Min-Max). In particular, $v_{rob} \geq \max\{v_{\ell'} : \ell' \in \mathcal{L}\}$.

Proof. Let x^* be an optimal solution to (Rob-Min-Max), which exists by Assumption 1. Further, let $\ell \in \mathcal{L}$ be given arbitrarily. Then, we obtain

$$v_{\text{rob}} = c^\top x^* + \Phi_{\text{rob}}(x^*) = c^\top x^* + \max_{k \in \mathcal{L}} \{\Phi_k(x^*)\} \geq c^\top x^* + \Phi_\ell(x^*) \geq v_\ell.$$

Here, the first equality follows from the optimality of x^* for (Rob-Min-Max). Due to Assumption 2, we can apply Lemma 1 to obtain the second equality. The last inequality follows from $x^* \in X$, i.e., the feasibility of x^* for (ℓ-Min-Max). Finally, $v_{\text{rob}} \geq v_\ell$ for all $\ell \in \mathcal{L}$ is equivalent to $v_{\text{rob}} \geq \max\{v_{\ell'} : \ell' \in \mathcal{L}\}$. □

In Proposition 1, we state that a valid lower bound for the optimal objective function value of (Rob-Min-Max) can be obtained by solving appropriately chosen deterministic min-max problems. In particular, we show the minimax inequality

$$\min_{x \in X} \left\{ \max_{\ell \in \mathcal{L}} \left\{ c^\top x + \Phi_\ell(x) \right\} \right\} \geq \max_{\ell \in \mathcal{L}} \left\{ \min_{x \in X} \left\{ c^\top x + \Phi_\ell(x) \right\} \right\};$$

see, e.g., Section 3.4 in Bertsekas (2009) for further discussion of minimax theory. Our heuristic method for (Rob-Min-Max) is motivated by Proposition 1 and is formally stated in Algorithm 1.

Algorithm 1 A Heuristic for Γ-Robust Mixed-Integer Linear Min-Max Problems

Input: An instance of (Rob-Min-Max), an exact solution method for (Min-Max) and (2), an index set $\mathcal{L}$ as in (5)

Output: A feasible leader's decision x^*, a lower bound L, and an upper bound U for (Rob-Min-Max)

1: Set $x^* \leftarrow$ None, $L \leftarrow -\infty$, and $U \leftarrow \infty$.
2: **for all** $\ell \in \mathcal{L}$ **do**
3: Compute a solution x^ℓ to the deterministic min-max problem

$$v_\ell \leftarrow \min_{x \in X} \left\{ c^\top x + \Phi_\ell(x) \right\}. \qquad (\ell\text{-Min-Max})$$

4: Set $L \leftarrow \max\{L, v_\ell\}$.
5: **if** $U \leq L$ **then**
6: **return** x^*, L, U
7: Solve the x^ℓ-parameterized Γ-robust lower-level problem to obtain $\Phi_{\text{rob}}(x^\ell)$.
8: **if** $c^\top x^\ell + \Phi_{\text{rob}}(x^\ell) < U$ **then**
9: Set $x^* \leftarrow x^\ell$ and $U \leftarrow c^\top x^* + \Phi_{\text{rob}}(x^*)$.
10: **if** $U \leq L$ **then**
11: **return** x^*, L, U
12: **return** x^*, L, U

In Algorithm 1, we solve up to $|\mathcal{L}|$ deterministic bilevel problems, which yields a valid lower bound for (Rob-Min-Max). Hence, our method relates to the main result by Bertsimas and Sim (2003) and Sim (2004) in the sense that we solve a linear number of problems of the nominal type, i.e., deterministic min-max problems. To be more specific, the number of min-max problems to be solved is linear in the number of uncertain parameters in the lower level. In addition, we exploit the solutions to the sub-problems (ℓ-Min-Max) to obtain a feasible point for (Rob-Min-Max). For each leader's solution x^ℓ, $\ell \in \mathcal{L}$, we evaluate the objective function of the Γ-robust follower by solving the x^ℓ-parameterized Γ-robust counterpart of the lower-level problem. The latter yields a valid upper bound. Among the considered solutions, we then take the best w.r.t. the upper-level objective function value.

Note that any solver for deterministic mixed-integer linear min-max problems can be used for the solution of the problems considered in Line 3. Valid options include, but are not limited to, the MibS solver (Tahernejad et al. 2020) or the general branch-and-cut solver presented in Fischetti et al. (2017). Nevertheless, our approach differs from the Bertsimas–Sim result since, in addition to solving problems of the nominal type, we further solve the Γ-robust counterpart of the lower level in Line 7 of Algorithm 1. The latter can be tackled in two ways:

(i) We solve the problem as a mixed-integer linear problem; see, e.g., in Lemma 1 in Beck et al. (2023a).
(ii) We exploit Lemma 1 of this paper such that the problem can be solved by solving $|\mathcal{L}|$ lower-level sub-problems of the nominal type.

Let us emphasize that, regardless of the choice between (i) or (ii), any method for Γ-robust single-level problems can be used in Line 7 of Algorithm 1.

Theorem 1. *Algorithm 1 is correct, i.e., it returns a feasible leader's decision x^* as well as valid lower and upper bounds L and U for* (Rob-Min-Max).

Proof. Since $\ell \in \mathcal{L}$ does not affect the upper-level constraints, any $x^* \in X$ that is computed by Algorithm 1 is feasible for (Rob-Min-Max). Moreover, $c^\top x + \Phi_{\text{rob}}(x) \geq v_{\text{rob}}$ holds for all $x \in X$. By the updating rule in Line 9 of the algorithm, U is a valid upper bound for the optimal objective function value of (Rob-Min-Max). Finally, the validity of L as a lower bound follows from Proposition 1 and Line 4. □

By Assumption 1, an optimal solution to (ℓ-Min-Max) exists for all $\ell \in \mathcal{L}$. Hence, Line 3 of Algorithm 1 is well-defined. Moreover, we emphasize that also Line 7 of Algorithm 1 is well-defined since, due to x^ℓ being a solution to (ℓ-Min-Max), it holds $Y(x^\ell) \neq \emptyset$.

Remark 1. *If Assumption 1 were not satisfied, the infeasibility or unboundedness of* (Rob-Min-Max) *could be identified in Line 3 of Algorithm 1 as well. The reasons are the following. If a sub-problem* (ℓ-Min-Max) *were unbounded, Proposition 1 would imply that the overall problem* (Rob-Min-Max) *is unbounded. Moreover, since $\ell \in \mathcal{L}$ affects neither the upper- nor the lower-level constraints, the infeasibility of a sub-problem* (ℓ-Min-Max) *would imply the infeasibility of* (Rob-Min-Max).

3.2. **Quality Guarantees.** We now provide quality guarantees for a leader's decision x^* that is computed by Algorithm 1.

Remark 2. *If Algorithm 1 terminates with (x^*, L, U) in Line 6 or 11, $U - L = 0$ holds and x^* is an optimal solution to* (Rob-Min-Max).

By construction, if Algorithm 1 does not terminate in Line 6 or 11 with an optimal solution, it returns the best-known leader's decision x^* with a positive optimality gap. However, the latter does not necessarily imply that none of the $|\mathcal{L}|$ bilevel sub-problems produces a solution that is optimal for (Rob-Min-Max). The reasons are two-fold. On the one hand, this may be due to the multiplicity of solutions to the deterministic min-max problems (ℓ-Min-Max). On the other hand, we emphasize that the sub-problems (ℓ-Min-Max) are only relaxations of (Rob-Min-Max). Nevertheless, if Algorithm 1 does not terminate with a provably optimal solution, its output (x^*, L, U) can still be valuable for the exact branch-and-cut approach presented in Beck et al. (2023a). More specifically, the leader's decision x^* could, in principle, be used to warmstart the method, whereas L and U could provide bounding information to reduce the search space in the branch-and-cut method.

Next, we determine an upper bound for the optimality gap in the case in which Algorithm 1 terminates in Line 12. To this end, we start with the following technical lemmas. Note that all the proofs omitted here can be found in Appendix A.

Lemma 2. *For arbitrarily given $x \in X$ and $y \in Y(x)$, it holds*

$$f^\top y - \max_{\{S \subseteq [n_y] \colon |S| \leq \Gamma\}} \sum_{i \in S} \Delta f_i y_i = \max_{\ell \in \mathcal{L}} \left\{ -\Gamma \Delta f_\ell + \tilde{f}(\ell)^\top y \right\}.$$

Lemma 3. *Let ℓ, $k \in \mathcal{L}$ with $\ell \leq k$ be given arbitrarily. Then, $\tilde{f}(\ell) \geq \tilde{f}(k)$ holds.*

To conclude this section, we now provide an upper bound for the optimality gap of a point x^* that is computed by Algorithm 1.

Proposition 2. *Let (x^*, L, U) be the output of Algorithm 1. Then, it holds*

$$U - L \leq (2\Gamma + 1)\Delta f_{\Gamma+1} + \sum_{i=\Gamma+2}^{n_y} \Delta f_i.$$

While we acknowledge that the bound provided in Proposition 2 seems rather loose, it is important to note that no structural assumptions have been made regarding the lower-level feasible set $Y(x) \subseteq \{0,1\}^{n_y}$, $x \in X$. Tighter bounds may be obtained using specific knowledge of the application problem at hand.

3.3. **Parallelization.** We emphasize that the sub-problems (ℓ-Min-Max) that are solved in Line 3 of Algorithm 1 are independent. This means that, if the necessary capacities are available, they can be solved in parallel. Hence, instead of alternating between solving deterministic min-max problems and robustified lower-level problems as it is done in Algorithm 1, it may be beneficial to first solve all min-max problems (in parallel) and, afterward, perform the necessary computations to obtain a valid and ideally tight upper bound. The latter leads to a modification of Algorithm 1, which is summarized in Algorithm 2.

Algorithm 2 A Modification of Algorithm 1

Input: An instance of (Rob-Min-Max), an exact solution method for (Min-Max) and (2), an index set $\mathcal{L}$ as in (5)

Output: A feasible leader's decision x^*, a lower bound L, and an upper bound U for (Rob-Min-Max)

1: Set $x^* \leftarrow \textsf{None}$, $L \leftarrow -\infty$, and $U \leftarrow \infty$.
2: **for all** $\ell \in \mathcal{L}$ **do**
3: Compute a solution x^ℓ to the deterministic min-max problem

$$v_\ell \leftarrow \min_{x \in X} \left\{ c^\top x + \Phi_\ell(x) \right\}. \qquad (\ell\text{-Min-Max})$$

4: Sort the indices such that $v_{\ell_1} \leq v_{\ell_2} \leq \cdots \leq v_{\ell_{|\mathcal{L}|}}$ holds and set $L \leftarrow v_{\ell_{|\mathcal{L}|}}$.
5: Set $i \leftarrow 1$.
6: **while** $i \leq |\mathcal{L}|$ and $L < U$ **do**
7: Solve the x^{ℓ_i}-parameterized Γ-robust lower-level problem to obtain $\Phi_{\text{rob}}(x^{\ell_i})$.
8: **if** $c^\top x^{\ell_i} + \Phi_{\text{rob}}(x^{\ell_i}) < U$ **then**
9: Set $x^* \leftarrow x^{\ell_i}$ and $U \leftarrow c^\top x^* + \Phi_{\text{rob}}(x^*)$.
10: Set $i \leftarrow i + 1$.
11: **return** x^*, L, U

From what we have shown so far, it is evident that Algorithm 2 is correct. In Line 4 of Algorithm 2, we sort the indices so that the optimal objective function values of the sub-problems (ℓ-Min-Max) are given in non-decreasing order. While the latter is not necessary for the correctness of the method, we expect that it helps closing the optimality gap more quickly. Let us further emphasize that, if the necessary capacities are available, Lines 2 and 3 of the algorithm can be parallelized.

In addition, if we exploit the result from Lemma 1 to solve the Γ-robust counterpart of the lower level, we can further make use of parallelization in Line 7 of Algorithm 2. The reason is that the lower-level sub-problems

$$\Phi_\ell(x^{\ell_i}) = -\Gamma\Delta f_\ell + \max_{y\in Y(x^{\ell_i})} \left\{\tilde{f}(\ell)^\top y\right\}, \quad \ell \in \mathcal{L},$$

are independent for fixed $x^{\ell_i} \in X$ and can, thus, be solved in parallel as well. Let us mention, however, that other parallelization schemes than the one outlined above may be possible as well.

3.4. **Reduction of Sub-Problems to Be Solved.** By construction, Algorithms 1 and 2 terminate after solving (at most) $|\mathcal{L}|$ deterministic min-max problems and Γ-robust counterparts of the lower level, respectively. In particular, if Lemma 1 is exploited, this means that at most $|\mathcal{L}|^2$ lower-level problems of the nominal type are solved. Thus, it is evident that Algorithms 1 and 2 require a significant amount of resources—especially for large index sets $\mathcal{L}$. In what follows, we aim to reduce the computational burden by decreasing the number of sub-problems to be solved.

Proposition 3. *Let ℓ, $k \in \mathcal{L}$ with $\ell < k$ and $\Delta f_\ell = \Delta f_k$ be given arbitrarily. Then, the following holds:*

(i) For all $x \in X$ and $\ell \leq i \leq k$, we have $\Phi_\ell(x) = \Phi_i(x)$.
(ii) For all $\ell \leq i \leq k$, an optimal solution x^ℓ to (ℓ-Min-Max) is also an optimal solution to the ith deterministic min-max problem

$$\min_{x\in X} \left\{c^\top x + \Phi_i(x)\right\}$$

and vice versa.

Proof. For all $\ell \leq i \leq k$, we obtain $\Delta f_\ell = \Delta f_i$ from Assumption 2 and, thus, $\tilde{f}(\ell) = \tilde{f}(i)$ holds due to Lemma 1. Hence, for all $x \in X$ and all $\ell \leq i \leq k$, we have

$$\Phi_\ell(x) = -\Gamma\Delta f_\ell + \max_{y\in Y(x)} \left\{\tilde{f}(\ell)^\top y\right\} = -\Gamma\Delta f_i + \max_{y\in Y(x)} \left\{\tilde{f}(i)^\top y\right\} = \Phi_i(x).$$

This proves (i). In particular, we obtain

$$v_\ell = \min_{x\in X} \left\{c^\top x + \Phi_\ell(x)\right\} = \min_{x\in X} \left\{c^\top x + \Phi_i(x)\right\} = v_i$$

for all $\ell \leq i \leq k$. Thus, and since ℓ does not affect the upper-level constraints, an optimal solution x^ℓ to (ℓ-Min-Max) is also an optimal solution to the ith deterministic min-max problem, $\ell \leq i \leq k$, and vice versa. □

By Proposition 3, it suffices to only consider the sub-problems (ℓ-Min-Max) for which the associated deviations are pairwise distinct.

Remark 3. *For an arbitrarily given $\ell \in \mathcal{L}$, we already know $\Phi_\ell(x^\ell)$ from the solution of (ℓ-Min-Max) in Algorithm 1 or 2. Hence, when exploiting Lemma 1 to determine $\Phi_{rob}(x^\ell)$, it suffices to solve $|\mathcal{L}| - 1$ deterministic lower-level sub-problems. Consequently, the overall number of lower-level sub-problems to be solved in Algorithms 1 and 2 can be reduced to at most $|\mathcal{L}|\,(|\mathcal{L}| - 1)$.*

We conclude this section with a sufficient ex-post condition under which (Rob-Min-Max) can be solved by only solving problems of the nominal type.

Theorem 2. *Let $(x^\ell)_{\ell\in\mathcal{L}}$ be a family of solutions to the deterministic min-max problems (ℓ-Min-Max) and let $(v_\ell)_{\ell\in\mathcal{L}}$ be the vector of the associated objective function values. Further, let $k = \arg\max_{\ell\in\mathcal{L}}\{v_\ell\}$. If $x^k = x^\ell$ holds for all $\ell \in \mathcal{L}$, x^k is an optimal solution to (Rob-Min-Max) and $v_{rob} = v_k$ holds.*

Table 2. Summary of algorithmic refinements for Algorithms 1 and 2.

	Algorithm 1	Algorithm 2
Reduction of sub-problems to be solved according to Remark 3, Proposition 3, and Theorems 2 and 3	✓	✓
Possibility to parallelize the solution of ...		
min-max problems (ℓ-Min-Max)	✗	✓
lower-level sub-problems (using Lemma 1)	✓	✓
Possibility to terminate early without ...		
solving all min-max problems (ℓ-Min-Max)	✓	✗
solving additional lower-level problems	✗	✓

Proof. Suppose that $x^k = x^\ell$ holds for all $\ell \in \mathcal{L}$. Then, we have $\Phi_\ell(x^\ell) = \Phi_\ell(x^k)$ and, thus,

$$c^\top x^k + \Phi_k(x^k) = v_k \geq v_\ell = c^\top x^\ell + \Phi_\ell(x^\ell) = c^\top x^k + \Phi_\ell(x^k)$$

holds for all $\ell \in \mathcal{L}$. The latter is equivalent to $\Phi_k(x^k) \geq \Phi_\ell(x^k)$ for all $\ell \in \mathcal{L}$, i.e., $\Phi_k(x^k) = \Phi_{\text{rob}}(x^k)$ holds due to Lemma 1. Hence, we obtain

$$v_k = c^\top x^k + \Phi_k(x^k) = c^\top x^k + \Phi_{\text{rob}}(x^k) \geq v_{\text{rob}}.$$

Here, the last inequality follows from $x^k \in X$, i.e., the feasibility of x^k for Problem (Rob-Min-Max). In addition, we obtain $v_{\text{rob}} \geq v_k$ from Proposition 1. To sum up, we have $v_{\text{rob}} = v_k$, which concludes the proof. □

Theorem 2 indicates that there may be situations in which the Bertsimas–Sim result extends to the min-max setting. However, the result does not carry over completely as the requirements of Theorem 2 can only be checked ex post, i.e., after solving the deterministic min-max problems. Moreover, we emphasize that the requirements of Theorem 2 are rather strong. Nevertheless, the result of Theorem 2 has the following implications for the presented heuristics:

(i) If the solutions $(x^\ell)_{\ell \in \mathcal{L}}$ obtained in Line 3 of Algorithm 2 are the same, the algorithm can terminate early with an optimal solution to (Rob-Min-Max). No additional lower-level problems need to be solved, i.e., Lines 4–11 of Algorithm 2 can be omitted. Verifying the requirements of Theorem 2 is simple and only requires one additional line of pseudo-code.

(ii) As per the algorithm's design, exploiting Theorem 2 in Algorithm 1 is not as straightforward as in Algorithm 2. The Γ-robust counterpart of the lower level is solved at least once, namely in the first iteration of the for-loop. Afterward, Line 7 only needs to be executed if the current fixed leader's decision x^ℓ differs from those obtained in the algorithm so far.

A summary of all algorithmic refinements discussed in this section is given in Table 2.

3.5. **Tailored Techniques for Interdiction Problems.** Let us emphasize that, up to now, we have not made any structural assumptions about the lower-level feasible set $Y(x)$, $x \in X$, except for the follower's variables being binary. However, further results on the reduction of sub-problems can be obtained by exploiting the specific properties of the application problem at hand. An important problem class that is covered by the min-max setting considered in this section is interdiction (Beck et al. 2023a; Brown et al. 2006; Cormican et al. 1998; DeNegre 2011; Fischetti et al. 2019; Furini et al. 2021; Israeli and Wood 2002; Wood 2011).

Assumption 3. *(i) All linking variables are binary.*

(ii) *For all $x \in X$, the lower-level feasible set is of the form*

$$Y(x) := Y \cap \{y\colon y_i \leq 1 - x_i,\ i \in I \subseteq [n_y]\}$$

with $Y \subseteq \{0,1\}^{n_y}$ being independent from the leader's variables.

(iii) *There are no terms depending on the leader's variables in the upper-level objective, i.e., $c = 0$.*

Under Assumption 3, Problem (Min-Max) is an interdiction problem.

Proposition 4. *Suppose that Assumption 3 holds and let x^ℓ be an optimal solution to (ℓ-Min-Max) for an arbitrarily given $\ell \in \mathcal{L}$. If there exists $x \in X$ with $x \geq x^\ell$, then x is an optimal solution to (ℓ-Min-Max) as well.*

Proposition 4 states that any feasible leader's decision $x \in X$ dominating an optimal solution to (ℓ-Min-Max) is an optimal solution to the problem as well.

Theorem 3. *Suppose that Assumption 3 holds. Let $(x^\ell)_{\ell \in \mathcal{L}}$ be a family of solutions to the deterministic min-max problems (ℓ-Min-Max) and let $(v_\ell)_{\ell \in \mathcal{L}}$ be the vector of the associated objective function values. Further, let $k = \arg\max_{\ell \in \mathcal{L}}\{v_\ell\}$. If there exists $x \in X$ with $x \geq x^\ell$ for all $\ell \in \mathcal{L}$, then x is an optimal solution to (Rob-Min-Max) and $v_{rob} = v_k$ holds.*

Proof. Let $x \in X$ be such that $x \geq x^\ell$ holds for all $\ell \in \mathcal{L}$. Then, x solves all deterministic min-max problems (ℓ-Min-Max) due to Proposition 4. Finally, the claim follows from applying Theorem 2. □

Theorem 3 extends the result of Theorem 2 such that, under Assumption 3, no additional lower-level problems need to be solved to obtain $\Phi_{\text{rob}}(x)$.

4. General Mixed-Integer Linear Bilevel Problems

We now return to the more general setting of Γ-robust mixed-integer linear bilevel problems as stated in (Rob-BMIP). Here, the objective function coefficients for the follower's variables y may differ in the upper- and the lower-level problem, i.e., $d = f$ does not need to hold anymore. In what follows, we illustrate that this setting is considerably more challenging than its min-max counterpart. Nevertheless, we present a heuristic for (Rob-BMIP) that builds on the ideas of Section 3. We formally state the method in Section 4.1 and provide quality guarantees for heuristically obtained solutions in Section 4.2. In Section 4.3, we discuss algorithmic refinements.

4.1. A Heuristic for General Γ-Robust Bilevel Problems.

In this section, we, again, build on a lower bounding scheme that follows the ideas of the Bertsimas–Sim result. For this purpose, we start with the following technical observation.

Lemma 4. *Let (x^*, y^*) be an optimal solution to (Rob-BMIP). Then, there exists an index $\ell \in \mathcal{L}$ such that*

$$f^\top y^* - \max_{\{S \subseteq [n_y]\colon |S| \leq \Gamma\}} \sum_{i \in S} \Delta f_i y_i^* = \Phi_\ell(x^*) = -\Gamma \Delta f_\ell + \tilde{f}(\ell)^\top y^*$$

holds.

Lemma 4 can be used to provide a lower bound for the optimal objective function value of (Rob-BMIP), which is what we do in the following.

Proposition 5. *There exists an index $\ell \in \mathcal{L}$ such that the optimal objective function value of the bilevel problem*

$$\begin{aligned} \min_{x,y} \quad & c^\top x + d^\top y \\ s.t. \quad & x \in X, \\ & y \in \arg\max_{y' \in Y(x)} \left\{ -\Gamma \Delta f_\ell + \tilde{f}(\ell)^\top y' \right\} \end{aligned} \tag{ℓ-BMIP}$$

yields a valid lower bound for the optimal objective function value of (Rob-BMIP).

Proof. Let (x^*, y^*) denote an optimal solution to (Rob-BMIP), which exists by Assumption 1. Due to Lemma 4, there is an index $\ell \in \mathcal{L}$ such that

$$f^\top y^* - \max_{\{S \subseteq [n_y] : \, |S| \leq \Gamma\}} \sum_{i \in S} \Delta f_i y_i^* = \Phi_\ell(x^*) = -\Gamma \Delta f_\ell + \tilde{f}(\ell)^\top y^*$$

holds. Hence, and since $x^* \in X$ as well as $y^* \in Y(x^*)$ hold by assumption, (x^*, y^*) is feasible for (ℓ-BMIP). Consequently, we obtain $c^\top x^* + d^\top y^* \geq c^\top x^\ell + d^\top y^\ell$ with (x^ℓ, y^ℓ) being an optimal solution to (ℓ-BMIP). This concludes the proof. □

Let us point out that Proposition 5 only yields an ex-post result since it requires the knowledge of an optimal solution to (Rob-BMIP) in advance. Nevertheless, it can be exploited to obtain an overall valid lower bound for (Rob-BMIP).

Corollary 1. *For all $\ell \in \mathcal{L}$, let (x^ℓ, y^ℓ) be an optimal solution to* (ℓ-BMIP). *Then,*

$$\min_{\ell \in \mathcal{L}} \left\{ c^\top x^\ell + d^\top y^\ell \right\}$$

is a valid lower bound for the optimal objective function value of (Rob-BMIP).

Corollary 1 motivates our heuristic for (Rob-BMIP). Before we discuss the method in detail, let us briefly comment on two main reasons why the setting considered in this section is more challenging than the one of Section 3:

(i) Obtaining a valid lower bound for (Rob-BMIP) is significantly more involved than in the min-max setting; cf. Proposition 1 and Corollary 1. In particular, Corollary 1 implies that the set of deterministic bilevel subproblems (ℓ-BMIP) needs to be considered holistically, i.e., an iterative refinement of the lower bound for (Rob-BMIP) such as in Line 4 of Algorithm 1 in the min-max setting can, in general, not be obtained.

(ii) In Section 3, we solve the Γ-robust counterpart of the lower level to obtain a valid upper bound, while the feasibility of a sub-problem's solution x^ℓ, $\ell \in \mathcal{L}$, for (Rob-Min-Max) is already guaranteed. A solution (x^ℓ, y^ℓ) to (ℓ-BMIP), however, may not be feasible for the Γ-robust bilevel problem (Rob-BMIP). Hence, we need to perform a correction step to restore feasibility. The latter may involve further challenges, which we address in detail in Section 4.3.

The heuristic for (Rob-BMIP) is formally stated in Algorithm 3. We start by solving $|\mathcal{L}|$ bilevel problems of the nominal type to obtain a valid lower bound; see Lines 2 and 3. By Part (i) of Proposition 3, we can assume, w.l.o.g., that the index set $\mathcal{L}$ is given such that the deviations $(\Delta f_\ell)_{\ell \in \mathcal{L}}$ are pairwise distinct. As in Algorithm 2, we sort the indices so that the optimal objective function values of the deterministic bilevel problems are given in non-decreasing order to potentially close the optimality gap more quickly; see Line 4. Note that the solutions $(x^\ell, y^\ell)_{\ell \in \mathcal{L}}$ may not be feasible for (Rob-BMIP). It may even be the case that none of the solutions to the deterministic bilevel problems (ℓ-BMIP) is feasible for (Rob-BMIP).[2] To

[2]An instance showing this behavior can be found at https://github.com/YasmineBeck/gamma-robust-bilevel-heuristics/tree/main/counterexample.

Algorithm 3 A Heuristic for Γ-Robust Mixed-Integer Linear Bilevel Problems

Input: An instance of (Rob-BMIP), an exact solution method for (BMIP) and (2), an index set $\mathcal{L}$ as in (5)

Output: A feasible pair (x^*, y^*), a lower bound L, and an upper bound U for (Rob-BMIP)

1: Set $(x^*, y^*) \leftarrow (\mathsf{None}, \mathsf{None})$, $L \leftarrow -\infty$, and $U \leftarrow \infty$.
2: **for all** $\ell \in \mathcal{L}$ **do**
3: Compute a solution (x^ℓ, y^ℓ) to the bilevel problem

$$\begin{aligned} \min_{x,y} \quad & c^\top x + d^\top y \\ \text{s.t.} \quad & x \in X, \\ & y \in \arg\max_{y' \in Y(x)} \left\{ -\Gamma \Delta f_\ell + \tilde{f}(\ell)^\top y' \right\}. \end{aligned} \tag{ℓ-BMIP}$$

4: Sort the indices such that

$$c^\top x^{\ell_1} + d^\top y^{\ell_1} \leq c^\top x^{\ell_2} + d^\top y^{\ell_2} \leq \cdots \leq c^\top x^{\ell_{|\mathcal{L}|}} + d^\top y^{\ell_{|\mathcal{L}|}}$$

holds and set $L \leftarrow c^\top x^{\ell_1} + d^\top y^{\ell_1}$.
5: Set $i \leftarrow 1$.
6: **while** $i \leq |\mathcal{L}|$ and $L < U$ **do**
7: Solve the x^{ℓ_i}-parameterized Γ-robust lower-level problem to obtain $\Phi_{\text{rob}}(x^{\ell_i})$ and let $\hat{y}$ denote its optimal solution.
8: **if** $c^\top x^{\ell_i} + d^\top \hat{y} < U$ **then**
9: Set $(x^*, y^*) \leftarrow (x^{\ell_i}, \hat{y})$ and $U \leftarrow c^\top x^* + d^\top y^*$.
10: Set $i \leftarrow i + 1$.
11: **return** (x^*, y^*), L, U

obtain a feasible point, we thus perform a correction step that involves solving the Γ-robust counterpart of the lower level; see Lines 7–9. We emphasize that any suitable solver can be used for the solution of the sub-problems (ℓ-BMIP) such as, e.g., the MibS solver (Tahernejad et al. 2020) or the general branch-and-cut solver presented in Fischetti et al. (2017). Since the sub-problems (ℓ-BMIP) are independent, they can be solved in parallel if the necessary capacities are available.

Theorem 4. *Algorithm 3 is correct, i.e., it returns a feasible pair (x^*, y^*) as well as valid lower and upper bounds L and U for (Rob-BMIP).*

Proof. Let $(x^\ell, y^\ell)_{\ell \in \mathcal{L}}$ be the family of solutions to the deterministic bilevel problems solved in Line 3 of Algorithm 3. Note that any pair $(x^{\ell_i}, \hat{y})$, $i \in \{1, \ldots, |\mathcal{L}|\}$, obtained from Line 7 of the algorithm satisfies $x^{\ell_i} \in X$, $\hat{y} \in Y(x^{\ell_i})$, and

$$f^\top \hat{y} - \max_{\{S \subseteq [n_y] \colon |S| \leq \Gamma\}} \sum_{i \in S} \Delta f_i \hat{y}_i = \Phi_{\text{rob}}(x^{\ell_i}),$$

i.e., $(x^{\ell_i}, \hat{y})$ is feasible for (Rob-BMIP). Consequently, $c^\top x^{\ell_i} + d^\top \hat{y}$ is a valid upper bound for the optimal objective function value of (Rob-BMIP). Let $((x^*, y^*), L, U)$ be the output of Algorithm 3. Then, by our previous considerations, (x^*, y^*) is feasible for (Rob-BMIP) and $U := c^\top x^* + d^\top y^*$ is a valid upper bound. Finally, the validity of L as a lower bound follows from Corollary 1. □

4.2. **Quality Guarantees.** We now provide quality guarantees for heuristically obtained solutions to (Rob-BMIP).

Proposition 6. *Let $(x^\ell, y^\ell)_{\ell \in \mathcal{L}}$ be the family of solutions to the bilevel problems solved in Line 3 of Algorithm 3. Further, let $((x^*, y^*), L, U)$ be the output of*

Algorithm 3 and suppose that $c^\top x^* + d^\top y^* \le c^\top x^\ell + d^\top y^\ell$ *holds for all* $\ell \in \mathcal{L}$*. Then,* $U - L = 0$ *holds and* (x^*, y^*) *is an optimal solution to* (Rob-BMIP).

Proof. Due to Theorem 4, (x^*, y^*) is feasible for (Rob-BMIP) and $L \le U$ holds. By assumption, we further have $c^\top x^* + d^\top y^* \le c^\top x^\ell + d^\top y^\ell$ for all $\ell \in \mathcal{L}$, which is equivalent to

$$U = c^\top x^* + d^\top y^* \le \min_{\ell \in \mathcal{L}} \left\{ c^\top x^\ell + d^\top y^\ell \right\} = L.$$

Here, the equalities follow from Lines 9 and 4 of Algorithm 3, respectively. To sum up, we have $U - L = 0$, which concludes the proof. □

Next, we provide a sufficient ex-post condition under which (Rob-BMIP) can be solved by only solving bilevel problems of the nominal type.

Theorem 5. *Let* $(x^\ell, y^\ell)_{\ell \in \mathcal{L}}$ *be the family of solutions to the bilevel problems solved in Line 3 of Algorithm 3. If there exists an index* $k \in \mathcal{L}$ *with* $x^k = x^\ell$*,* $d^\top y^k \le d^\top y^\ell$ *and* $\tilde{f}(\ell)^\top y^k \ge \tilde{f}(\ell)^\top y^\ell$ *for all* $\ell \in \mathcal{L}$*,* (x^k, y^k) *is an optimal solution to* (Rob-BMIP).

Proof. Suppose that there exists an index $k \in \mathcal{L}$ such that the requirements are satisfied, i.e., we have $x^k \in X$, $y^k \in Y(x^\ell)$, as well as

$$\Phi_\ell(x^k) = \Phi_\ell(x^\ell) = -\Gamma \Delta f_\ell + \tilde{f}(\ell)^\top y^\ell \le -\Gamma \Delta f_\ell + \tilde{f}(\ell)^\top y^k$$

for all $\ell \in \mathcal{L}$. The latter implies

$$\Phi_{\text{rob}}(x^k) = \max_{\ell \in \mathcal{L}} \left\{ \Phi_\ell(x^k) \right\} = \max_{\ell \in \mathcal{L}} \left\{ -\Gamma \Delta f_\ell + \tilde{f}(\ell)^\top y^k \right\}.$$

Thus, by Lemma 1, y^k solves the x^k-parameterized Γ-robust counterpart (2) of the lower level. In particular, the optimality of y^k is proven without solving additional lower-level problems. To sum up, (x^k, y^k) is feasible for (Rob-BMIP) and we have $c^\top x^k + d^\top y^k \le c^\top x^\ell + d^\top y^\ell$ for all $\ell \in \mathcal{L}$. Hence, using the same arguments as in the proof of Proposition 6, the optimality gap is closed. □

We now provide an upper bound for the optimality gap for the case in which Algorithm 3 does not terminate with an optimal solution.

Proposition 7. *Let* $((x^*, y^*), L, U)$ *be the output of Algorithm 3. Then, it holds*

$$U - L \le \|d\|_1.$$

Let us emphasize that the bound for the optimality gap given in Proposition 7 only depends on the upper-level objective function coefficients d for the follower's variables. Hence, as the influence of the follower on the leader's objective function value decreases, i.e., by diminishing $\|d\|_1$, the optimality gap of the pair (x^*, y^*) decreases as well. Note, however, that $d = 0$ would imply that the upper level is completely decoupled from the lower level.

4.3. **Algorithmic Refinements.** To conclude this section, we discuss further techniques that can be incorporated in Line 7 of Algorithm 3 to obtain refined upper bounds for (Rob-BMIP). For the ease of presentation, we focus on the case in which the Γ-robust counterpart of the lower level is solved by exploiting the result of Lemma 1. The case in which the problem is solved as a mixed-integer linear problem as in Lemma 1 of Beck et al. (2023a) can be treated similarly.

In Algorithm 4, we provide a detailed description of the steps involved to solve the Γ-robust counterpart of the lower level, i.e., Algorithm 4 may be used to replace Line 7 of Algorithm 3. Here, we include a so-called refinement step in which we solve further binary problems in addition to the lower-level sub-problems of the nominal type. The reasons are the following. In this paper, we consider the optimistic approach to bilevel optimization. Hence, whenever the set of optimal solutions to the Γ-robust lower level is not a singleton, a follower's response that favors the leader is chosen. When solving the lower-level sub-problems in Line 3 of Algorithm 4, no information about the upper-level objective is used. This means that, if these sub-problems do not have a unique solution, the cooperative nature of the follower may not be taken into account. Line 4 of Algorithm 4 is intended to remedy this situation so that we obtain a pair $(\hat{x}, \hat{y})$ that is more likely to correspond to an optimal solution to (Rob-BMIP). It would also be sufficient to consider

$$k \leftarrow \arg\max_{\ell \in \mathcal{L}} \{\Phi_\ell(\hat{x})\}$$

instead of the selection rule presented in Lines 5 and 6 of Algorithm 4 to obtain a feasible pair $(\hat{x}, \hat{y})$ for (Rob-BMIP). Then, however, the upper-level objective and, thus, the cooperative nature of the follower would not be taken into account again. In particular, the latter may lead to ambiguities if the choice of $k \in \mathcal{L}$ is not unique.

Algorithm 4 Correct-and-Refine

Input: A family of solutions $(x^\ell, y^\ell)_{\ell \in \mathcal{L}}$ to the deterministic bilevel sub-problems (ℓ-BMIP), an index set $\mathcal{L}$ as in (5), an index $\ell_i \in \mathcal{L}$

Output: A solution $\hat{y}$ to the x^{ℓ_i}-parameterized Γ-robust lower level (2)

1: Set $\hat{x} \leftarrow x^{\ell_i}$.

2: **for** $\ell \in \mathcal{L} \setminus \{\ell_i\}$ with $x^\ell \neq \hat{x}$ **do**

3: Correction step: Solve the $\hat{x}$-parameterized ℓth lower-level sub-problem

$$\Phi_\ell(\hat{x}) = -\Gamma \Delta f_\ell + \max_{y \in Y(\hat{x})} \left\{ \tilde{f}(\ell)^\top y \right\}.$$

4: Refinement step: Compute a solution $\hat{y}^\ell$ to the problem

$$\min_{y \in Y(\hat{x})} \quad d^\top y \quad \text{s.t.} \quad -\Gamma \Delta f_\ell + \tilde{f}(\ell)^\top y \geq \Phi_\ell(\hat{x})$$

and set $y^\ell \leftarrow \hat{y}^\ell$.

5: Set $\Phi_{\text{rob}}(\hat{x}) \leftarrow \max_{\ell \in \mathcal{L}} \{\Phi_\ell(\hat{x})\}$ and determine $\mathcal{C} := \{\ell \in \mathcal{L} \colon \Phi_\ell(\hat{x}) = \Phi_{\text{rob}}(\hat{x})\}$.

6: Set $k \leftarrow \arg\min_{\ell \in \mathcal{C}} \left\{ c^\top \hat{x} + d^\top y^\ell \right\}$ and $\hat{y} \leftarrow y^k$.

7: **return** $\hat{y}$

Proposition 8. *Let $(x^\ell, y^\ell)_{\ell \in \mathcal{L}}$ be a given family of solutions to the deterministic bilevel sub-problems (ℓ-BMIP). Further, let $\ell_i \in \mathcal{L}$, $i \in \{1, \dots, |\mathcal{L}|\}$, be given arbitrarily. Then, Algorithm 4 is correct, i.e., it returns an optimal solution to the x^{ℓ_i}-parameterized Γ-robust counterpart (2) of the lower-level problem.*

Proof. For notational convenience, we set $\hat{x} = x^{\ell_i}$. By Remark 3, it suffices to solve

$$\Phi_\ell(\hat{x}) = -\Gamma \Delta f_\ell + \max_{y \in Y(\hat{x})} \left\{ \tilde{f}(\ell)^\top y \right\} \quad \text{for all } \ell \in \mathcal{L} \setminus \{\ell_i\} \tag{6}$$

to determine $\Phi_{\text{rob}}(\hat{x})$. We now show that Lines 3 and 4 of Algorithm 4 only need to be executed if $\hat{x} \neq x^\ell$ holds for some $\ell \in \mathcal{L} \setminus \{\ell_i\}$. To this end, suppose that there exists an index $\ell \in \mathcal{L} \setminus \{\ell_i\}$ with $\hat{x} = x^\ell$ and let it be given arbitrarily. Then, y^ℓ solves the ℓth $\hat{x}$-parameterized lower-level sub-problem in (6) as well. Moreover, $(\hat{x}, y^\ell) = (x^\ell, y^\ell)$ solves (ℓ-BMIP) by assumption, i.e., y^ℓ also solves the corresponding $\hat{x}$-parameterized binary problem considered in Line 4 of Algorithm 4. Hence, there is no need to solve the problems in Lines 3 and 4 to obtain $\Phi_{\text{rob}}(\hat{x})$. Overall, $(\hat{x}, \hat{y})$ obtained from Algorithm 4 thus satisfies $\hat{x} \in X$, $\hat{y} \in Y(\hat{x})$, and

$$f^\top \hat{y} - \max_{\{S \subseteq [n_y] \colon |S| \leq \Gamma\}} \sum_{i \in S} \Delta f_i \hat{y}_i = \Phi_{\text{rob}}(\hat{x}),$$

i.e., $(\hat{x}, \hat{y})$ is feasible for (Rob-BMIP). □

5. Computational Results

In this section, we computationally assess the performance of the heuristics presented in this paper by comparing them with exact methods as well as alternative heuristics adapted from the related literature. Before we discuss the numerical results for each method in detail, we briefly describe the generation of the test instances as well as the computational setup in Sections 5.1 and 5.2, respectively. In Section 5.3, we focus on the min-max setting considered in Section 3. Afterward, in Section 5.4, we discuss general Γ-robust mixed-integer linear bilevel problems; cf. Section 4.

The evaluations of the proposed heuristic methods rely on (i) the running times, (ii) the number of instances solved to global optimality, as well as on (iii) optimality gaps. Moreover, as it is mentioned in Sections 3 and 4, the proposed methods can be partially parallelized. To assess the potential of parallelization, we further use so-called idealized parallel runtimes. The latter reflect the overall runtime of a method provided that there are sufficient capacities available to solve all arising sub-problems in parallel. For each instance, we compute the idealized parallel runtime after solving all sub-problems sequentially by taking the maximum of all runtimes for the sub-problems. Hence, if an instance could not be tackled within a reasonable amount of time in the sequential setting, we consider it as unsolved in the idealized parallel setting as well.

5.1. **Generation of Knapsack Test Instances.** For our computational study, we consider modifications of the deterministic knapsack interdiction problem that has been considered in Caprara et al. (2016). The deterministic problem is formally stated as

$$\begin{aligned} \min_{x \in \{0,1\}^n} \quad & f^\top y \\ \text{s.t.} \quad & v^\top x \leq B, \\ & y \in \operatorname*{arg\,max}_{y' \in \{0,1\}^n} \left\{ f^\top y' \colon w^\top y' \leq C,\, y'_i \leq 1 - x_i,\, i \in [n] \right\}, \end{aligned}$$

in which all parameters are assumed to be non-negative integers, i.e., $B, C \in \mathbb{Z}_{\geq 0}$, and $f, v, w \in \mathbb{Z}^n_{\geq 0}$. For each instance size $n \in \{35, 40, 45, 50, 55, \dots, 100\}$, 10 instances have been generated according to Martello et al. (1999). A detailed description of the generation of the deterministic test instances can also be found in Section 4.1 of Beck et al. (2023a). To account for a Γ-robust follower, we adapt the deterministic instances in the following way. The parameter Γ is set to either 10 % or 50 % of the instance size n. In the case of a fractional value for Γ, the closest integer is considered. For the deviations in the objective function coefficients, we include four different settings. To this end, we choose $\delta \in \{0.1, 0.25\}$ and generate the deviations as follows:

(i) Integer deviations: The deviations Δf_i take uniformly distributed integer values from the interval $[0, \lceil \delta f_i \rceil]$.
(ii) Continuous deviations: We generate a continuous and uniformly distributed value α_i from the interval $[0, \delta)$ and set $\Delta f_i = \alpha_i f_i$.

In summary, we consider 80 instances per size such that our overall test set contains 1120 robustified knapsack interdiction instances. The latter are used to compare the approaches in the min-max setting discussed in Section 5.3.

For the more general bilevel setting evaluated in Section 5.4, we do the following. We re-consider the previously generated 1120 robustified knapsack interdiction instances, maintaining the uncertainty parameterization as well as the structure of the lower-level problem. An instance of the general form (Rob-BMIP) is then obtained by replacing upper-level objective function coefficients for the leader's and the follower's variables. These coefficients take uniformly distributed integer values from the interval $[0, 100]$. Hence, we also consider 1120 robustified instances in the more general setting.

5.2. **Computational Setup.** All tests have been realized on an Intel XEON SP 6126 at 2.6 GHz (with up to 16 cores) and 32 GB RAM. The approaches considered in our computational study use Gurobi 11.0.0 to solve all arising optimization problems. For the solution of each instance, we set a time limit of 1 h. We now comment on the implementation for each setting.

5.2.1. *Mixed-Integer Linear Min-Max Problems.* In Algorithms 1 and 2, a linear number of min-max problems of the nominal type is solved. In particular, any solver for mixed-integer linear min-max problems can be used for the solution of these problems. To assess the performance of our heuristics on instances of the Γ-robust knapsack interdiction problem, we consider the following two choices for the black-box solver.

First, we consider the problem-tailored branch-and-cut method presented in Fischetti et al. (2019) for the solution of the deterministic problems. In this method, the authors exploit so-called interdiction cuts to separate bilevel infeasible points.

Second, we consider the bkpsolver (Weninger and Fukasawa 2023) for the solution of the deterministic knapsack interdiction problems. The method is based on a branch-and-bound framework that incorporates ideas from dynamic programming to obtain strong lower bounds and is publicly available at `https://github.com/nwoeanhinnogaehr/bkpsolver`. We emphasize that the bkpsolver requires the parameters of the considered deterministic problems to satisfy the following properties:

(i) the leader's (the follower's) item weights do not exceed the leader's (the follower's) budget,
(ii) all parameters of the problem are integer,
(iii) all parameters of the problem are non-negative.

While (i) is satisfied for all of our considered instances by construction, (ii) and (iii) may be violated in some cases. For continuous deviations Δf, the modified objective function coefficients $\tilde{f}(\ell)$, $\ell \in \mathcal{L}$, may be continuous as well. In this case, we scale the data accordingly. Moreover, the modified objective function coefficients $\tilde{f}(\ell)$, $\ell \in \mathcal{L}$, may be negative for certain items in some sub-problems. To ensure the applicability of the bkpsolver within our framework, we have thus incorporated a presolve step into Algorithms 1 and 2 that is based on the following. If $\tilde{f}(\ell)_i < 0$ holds for some $i \in [n]$ and $\ell \in \mathcal{L}$, the ith item will not be chosen by the follower; see, e.g., Pisinger and Toth (1998). Thus, the leader does not need to spend interdiction resources on the ith item and $x_i = y_i = 0$ can be fixed for this sub-problem.

Due to their similarity and for the ease of presentation, we only discuss the results for one variant of our heuristic. Preliminary computational results revealed that Algorithm 2 seems to have an advantage over Algorithm 1, which is why we will focus on the following two variants.

H-BKP: Algorithm 2 in which we incorporate the bkpsolver (Weninger and Fukasawa 2023) for the solution of the deterministic interdiction problems.

H-IC: Algorithm 2 in which we incorporate the branch-and-cut approach proposed by Fischetti et al. (2019) for the solution of the deterministic interdiction problems.

To assess the performance of our methods, we compare H-BKP and H-IC with the following two benchmark approaches from the literature.

H-GI: The "Greedy Interdiction" heuristic proposed by DeNegre (2011). The method generates a feasible leader's decision x in a greedy-like fashion and, afterward, a valid upper bound is computed by solving the x-parameterized lower-level problem. The original method has been proposed for deterministic interdiction problems. Hence, we have adapted the method to account for a Γ-robust follower. Moreover, since the original method does not provide dual information, we further solve the so-called high-point relaxation (HPR; see, e.g., Definition 1.9 in Schmidt and Beck (2023)) of the problem to obtain a valid lower bound.

E-MF: The exact single-leader multi-follower approach presented in our previous work in Beck et al. (2023a). The method relies on a branch-and-cut framework in which interdiction cuts tailored to the Γ-robust setting are added. The code is publicly available at `https://github.com/YasmineBeck/gamma-robust-knapsack-interdiction-solver`. In Beck et al. (2023a), various cut separation strategies are studied. In our computational study, we consider the setting in which a single most-violated cut is added at each node of the branch-and-cut search tree. To generate these cuts, all lower-level sub-problems need to be solved, which can be done in parallel if the necessary capacities are available. We account for this feature by considering idealized parallel runtimes for E-MF in our evaluations.

Algorithms 1 and 2 as well as the re-implementation of the "Greedy Interdiction" heuristic (DeNegre 2011) are implemented in Python 3.7.11. Moreover, since the original branch-and-cut method proposed in Fischetti et al. (2019) uses CPLEX 12.7 to solve all arising optimization problems, we have also re-implemented this method using Python 3.7.11 and Gurobi to have a fair comparison between the considered approaches. Our re-implementation exploits Gurobi's lazy constraint callbacks to add interdiction cuts, which requires to set the parameter LazyConstraints to 1. All other parameters have been left at their default settings. For the solution of the Γ-robust counterpart of the lower level, we exploit the result of Lemma 1 so that we solve a linear number of lower-level sub-problems of the nominal type.[3] If the necessary capacities are available, the independence of these sub-problems allows for a parallelization of their solution. Hence, we also consider idealized parallel runtimes for H-BKP, H-IC, and H-GI.

5.2.2. *General Mixed-Integer Linear Bilevel Problems.* We now briefly describe the implementation of the heuristic for general Γ-robust mixed-integer linear bilevel problems presented in Algorithms 3 and 4. Again, any solver for deterministic mixed-integer linear bilevel problems can be used for the solution of the problems of the nominal type in Line 3 of Algorithm 3. In our computational study, we consider the following.

H: Algorithm 3 in which we incorporate a problem-tailored branch-and-cut approach for the solution of the deterministic bilevel problems. The method is based on H-IC, which we have adapted accordingly to account for the more general setting. As elaborated in Section 4, our heuristic can be partially parallelized. In what follows, we thus abbreviate the heuristic in the sequential and the idealized parallel setting by H-seq and H-ideal, respectively.

Preliminary computational tests revealed that our problem-tailored branch-and-cut approach outperforms general-purpose solvers, which is why we refrain from using solvers such as, e.g., the MibS solver (Tahernejad et al. 2020) or the general

[3]The code for the methods presented in this paper, along with the nominal instance data used for our computational study, is publicly available at `https://github.com/YasmineBeck/gamma-robust-bilevel-heuristics`.

branch-and-cut solver presented in Fischetti et al. (2017) for the solution of the deterministic bilevel problems within our framework.

To the best of our knowledge, there is no other method in the literature that can tackle general mixed-integer linear bilevel problems with a Γ-robust treatment of lower-level objective uncertainty directly, neither globally nor heuristically. Nevertheless, the heuristics that have been proposed in Fischetti et al. (2018) can be applied to our considered instances by reformulating (Rob-BMIP) as a "generalized interdiction problem". The latter can be done using the ideas of Section 2.1 in Beck et al. (2023a); see Appendix B for the details. To assess the performance of H, we thus consider the following three benchmark approaches.

H-OS: The ONE-SHOT heuristic presented in Fischetti et al. (2018). The method builds on solving a single-level mixed-integer linear problem, which is obtained from the generalized interdiction problem by relaxing the integrality of the follower's variables and by exploiting strong duality. The original method computes a bilevel-feasible point and an upper bound for the optimal objective function value of (Rob-BMIP), but no lower bound is provided. To have at least some basis for evaluating the quality of the obtained solutions, we thus compute a valid lower bound by solving the HPR of the original bilevel problem.

H-IT: The ITERATE heuristic presented in Fischetti et al. (2018). The method iteratively adds no-good cuts to the single-level mixed-integer problem considered for ONE-SHOT and terminates with a bilevel-feasible pair and a valid upper bound once the time limit is reached. As before, we solve the HPR of the original bilevel problem to obtain a valid lower bound.

E: An exact branch-and-cut approach tailored to the mixed-integer linear reformulation of the considered generalized knapsack interdiction instances. The method is outlined in Appendix B.

E and the branch-and-cut method used within H are implemented in Python 3.7.11 and we use Gurobi's lazy constraint callbacks to add cuts by setting the parameter LazyConstraints to 1. All remaining parameters have been left at their default settings. Since the code for the heuristics presented in Fischetti et al. (2018) is not publicly available, we have re-implemented ONE-SHOT and ITERATE using Python 3.7.11 and Gurobi.[4] Finally, we point out that no parallelization can be exploited for H-OS, H-IT, or E, which is why we do not distinguish between a sequential and an idealized parallel setting for these approaches.

5.3. **Evaluation of the Heuristic for Min-Max Problems.** We now evaluate the heuristic for Γ-robust min-max problems. To this end, we apply the methods H-BKP, H-IC, H-GI, and E-MF to the Γ-robust knapsack interdiction problem, which is a special case of a Γ-robust min-max problem. Before we discuss the performance of the considered methods in detail, let us briefly highlight the main differences between them. Note that E-MF is an exact solution method that solves a single problem of the form given in (Rob-Min-Max), whereas H-BKP, H-IC, and H-GI are heuristic approaches for this problem. The considered heuristics have in common that they all solve the Γ-robust counterpart of the lower level for a fixed leader's decision. However, they differ in the sense that H-BKP and H-IC additionally solve a linear number of bilevel problems of the nominal type. Tables 3–5 as well as Figures 1–4 provide a comprehensive summary of the numerical results for the overall 1120 considered instances for all four approaches. In the following, we discuss the settings with integer and continuous deviations separately.

[4]The code for the methods presented in this paper, along with the nominal instance data used for our computational study, is publicly available at https://github.com/YasmineBeck/gamma-robust-bilevel-heuristics.

Table 3. The number of instances (out of the 560 considered instances in the min-max setting with integer and continuous deviations, respectively) for which our presolve techniques (see Section 5.2.1) have been applied.

	presolved	
Δf	sub-problems	variables
integer	560	500
continuous	440	520

Table 4. Statistics for the number of eliminated sub-problems and the number of fixed variables due to presolve (both in %) for the min-max setting with integer and continuous deviations.

Δf	presolved	min	1st quartile	median	3rd quartile	max
integer	sub-problems	40.91	63.84	73.21	80.00	88.00
	variables	0.00	3.48	5.00	6.04	8.00
continuous	sub-problems	0.00	3.13	8.89	21.08	44.00
	variables	0.00	3.48	5.00	6.67	9.09

5.3.1. *Instances with Integer Deviations.* We focus on the 560 Γ-robust knapsack interdiction instances for which the deviations take uniformly distributed integer values; cf. Section 5.1. In Table 3, we show the number of instances for which (i) the result of Proposition 3 has been applied and (ii) variables have been fixed due to negative modified profits; cf. the presolve techniques discussed in Section 5.2.1. As per the generation of our instances, there are multiple items that have the same deviation across all instances. By Proposition 3, the number of sub-problems to be solved can thus be reduced. Moreover, it can be seen that at least one variable is fixed in at least one sub-problem for a significant portion of our instances (500 out of 560 instances). In Table 4, we summarize the statistics for the number of sub-problems and variables that can be eliminated due to our presolve techniques proposed in Section 5.2.1. It can be seen that at least 40.91 % and up to 88 % of the sub-problems can be eliminated, which significantly reduces the computational burden of the heuristic presented in this paper. In addition, a maximum of 8 % of the number of variables is fixed due to negative modified profits $\tilde{f}(\ell)$, $\ell \in \mathcal{L}$. According to preliminary computational results, presolving variables affects the performance of H-IC only slightly. We emphasize, however, that the latter is necessary to apply H-BKP.

In Table 5, we summarize the number of instances for which (i) a feasible point with finite gap is found ("feasible"), (ii) global optimality is proven either by a closed gap or using one of the sufficient optimality conditions presented in Theorems 2 and 3 ("optimal"), (iii) a sufficient optimality condition is satisfied ("Thm. 2" or "Thm. 3"), and (iv) the computed solution has a finite but non-zero gap ("open gap"). For those instances with an open gap, we further provide the average gap ("average gap"). In addition, we show box-plots of the optimality gaps and the running times for all four considered approaches in Figures 1 and 4, respectively. In Figure 2, we further show box-plots of the ex-post optimality gaps for H-BKP, H-IC, and H-GI. The ex-post optimality gaps are derived by comparing the heuristic solutions with the exact solution obtained from E-MF. Moreover, we provide box-plots of the percentages of solved sub-problems (out of the total number of sub-problems to be

TABLE 5. The number of instances for which a feasible point with finite gap is found ("feasible"; out of the 560 considered instances in the min-max setting with integer and continuous deviations, respectively) for the approaches H-BKP, H-IC, H-GI, and E-MF. Additionally, the number of instances solved to global optimality ("optimal"), along with the number of instances satisfying a sufficient optimality condition (Thm. 2 or Thm. 3), is shown. For those instances with finite but non-zero gap ("open gap"), also the average gap ("average gap"; in %) is shown.

Δf		feasible	optimal	Thm. 2	Thm. 3	open gap	average gap
integer	H-BKP	560	555	340	340	5	0.12
	H-IC	517	513	277	315	4	0.14
	H-GI	560	4	–	–	556	100.00
	E-MF	560	526	–	–	34	6.08
continuous	H-BKP	560	554	359	359	6	0.08
	H-IC	481	476	266	309	5	0.10
	H-GI	560	4	–	–	556	100.00
	E-MF	560	524	–	–	36	7.03

solved) for H-BKP and H-IC in Figure 3. We now assess the performance of the four considered approaches.

We start with a comparison of H-IC and E-MF. Based on Table 5, it can be seen that a feasible point with finite gap is obtained using E-MF for all 560 considered instances. In particular, E-MF solves 526 of the 560 instances (93.93 %) to global optimality. Again, we emphasize that E-MF is an exact approach, which solves a single problem of the form given in (Rob-Min-Max) using branch-and-cut. Hence, optimality of a solution obtained from E-MF is proven by a closed gap. The sufficient conditions in Theorems 2 and 3 are only applicable to the heuristic approaches H-BKP and H-IC. H-IC finds a feasible point with finite gap for 517 of the 560 considered instances, while proving global optimality for 513 of them (91.61 %). Table 5 shows that the optimality of solutions obtained from H-IC is proven using the result of Theorem 2 for 277 of the 513 instances (54.00 %). Note that Theorem 2 is a special case of Theorem 3 so that, overall, 315 instances (61.40 %) satisfy the requirements of Theorem 3. Hence, the majority of the considered instances satisfies one of the sufficient conditions so that H-IC computes a globally optimal solution to (Rob-Min-Max) by only solving bilevel problems of the nominal type. For the remaining instances solved to global optimality by H-IC, a closed gap is obtained after solving additional lower-level problems. For those 4 instances for which H-IC has found a feasible but not provably optimal point with finite gap ("open gap"), the method still provides favorable results in terms of the solution quality. In Figure 1 (left), we show box-plots for the optimality gaps obtained from the four considered methods. The largest finite optimality gap we report for H-IC is 0.18 %, while, for E-MF, the outliers for the optimality gaps are widely scattered with the largest gap observed being 10.98 %. It is worth mentioning, however, that global optimality of the primal solutions found by H-IC has been verified ex post for all instances solved by E-MF; see Figure 2 (left). Nevertheless, we acknowledge that the computational burden of H-IC remains a drawback of our method. The latter is particularly reflected by the 43 instances for which H-IC could not compute a finite gap within the time limit of 1 h, as indicated by the labeled node in Figure 1 (left). An infinite gap occurs if the solution of the linear number of bilevel sub-problems exceeds the time limit, preventing the upper bound from being updated. In Figure 3 (left), we show box-plots of the percentages of sub-problems solved within the time limit (out of the total number of sub-problems to be solved) to provide further insight into the time consumption of H-IC. When evaluating the instances that each method can handle, H-IC seems to perform slightly better than E-MF both w.r.t. sequential and idealized parallel runtimes. The latter can be seen from the box-plots shown in Figure 4 (top). Here, we observe smaller median running times for H-IC compared to E-MF, along with a reduction of the overall variability of runtimes. The latter indicates that H-IC tends to have a slightly more consistent performance than E-MF. Nevertheless, despite its fairly promising results in terms of performance and solution quality, it is worth mentioning again that E-MF is an exact approach while H-IC is a heuristic. E-MF may thus still be considered as the overall better method.

However, the situation changes significantly if we use the bkpsolver for the solution of the deterministic bilevel problems. Figure 4 (top) clearly illustrates the benefits of using the heuristic. In the sequential as well as in the idealized parallel setting, H-BKP significantly outperforms H-IC and E-MF. We further observe that, compared to E-MF, the sequential runtime of H-BKP is more than a factor of 15 smaller in the median. The same qualitative behavior can be observed for the idealized parallel setting. In terms of the solution quality, we note that more instances satisfy the sufficient conditions for optimality in Theorems 2 and 3 when using H-BKP instead of H-IC. Moreover, H-BKP solves 33 instances that have not been solved by E-MF, resulting in 99.11 % of the 560 considered instances being solved to global optimality. For the remaining 5 instances for which H-BKP has found a feasible point with finite but non-zero gap, we report a gap of at most 0.18 %; cf. Figure 1 (left). Comparing to the results obtained from E-MF, however, we could verify ex post that H-BKP indeed solves all 560 considered instances to global optimality; cf. Figure 2 (left).

Finally, let us comment on the performance of H-GI (DeNegre 2011). Figure 4 (top) clearly shows that H-GI dominates all other approaches both w.r.t. sequential and idealized parallel runtimes. The latter is not surprising given that H-GI only considers single-level problems, whereas the remaining approaches (additionally) tackle bilevel problems that are harder to solve in general. Despite its favorable results in terms of runtimes, however, the quality of the solutions obtained from H-GI is rather poor. The latter can be seen from the results depicted in Table 5 as well as Figures 1 and 2 (left). In this context, we mention that the 4 instances that have been solved to global optimality by H-GI are trivial in the sense that the leader can interdict all items for the follower. In this case, the lower bound obtained from solving the HPR of the problem is tight. In general, however, it is well-known that lower bounds obtained from solving the HPR are very loose in an interdiction setting.

5.3.2. *Instances with Continuous Deviations.* We now focus on the 560 Γ-robust knapsack interdiction instances for which the deviations take continuous and uniformly distributed values; cf. (ii) in Section 5.1. The main observations discussed in the previous section become even more pronounced in this setting. We emphasize that, just due to the generation of the deviations, the result of Proposition 3 cannot be exploited as often as in the setting with integer deviations; see Tables 3 and 4. Consequently, more deterministic bilevel problems need to be solved within the heuristic framework presented in this paper. The latter results in an overall increased computational burden for H-BKP and H-IC as it can be inferred from the box-plots shown in Figure 3 (right) and Figure 4 (bottom). Nevertheless, due to its overall small runtime required for solving bilevel problems, the latter affects the performance of H-BKP only slightly. Again, we observe significant speed-up factors when comparing the runtimes of H-BKP to those of the exact approach

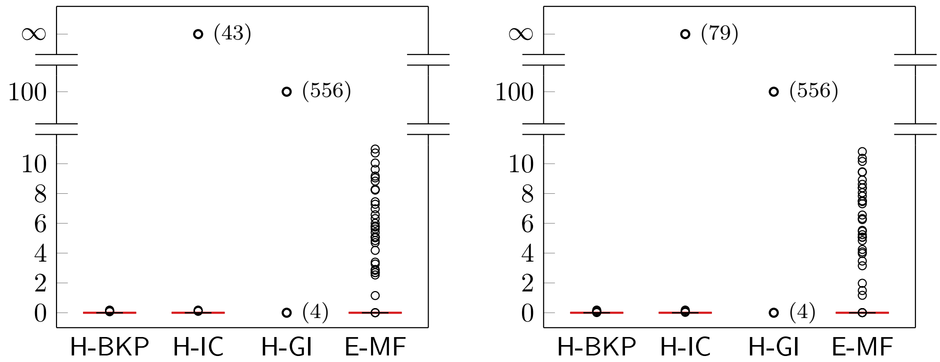

Figure 1. Box-plots of the optimality gaps (in %) for the approaches H-BKP, H-IC, H-GI, and E-MF in the min-max setting with integer (left) and continuous deviations (right).

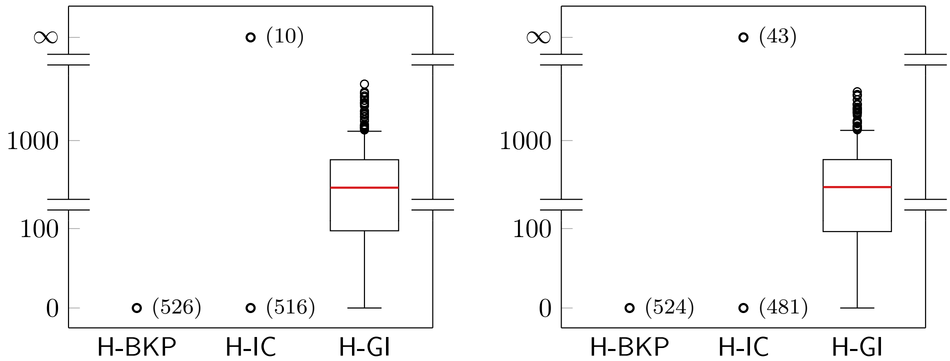

Figure 2. Box-plots of the ex-post optimality gaps (in %) for the approaches H-BKP, H-IC, and H-GI for the 526 instances with integer deviations (left) and the 524 instances with continuous deviations (right) that have been solved to global optimality by E-MF. Values above 100 % are shown on a log-scaled y-axis.

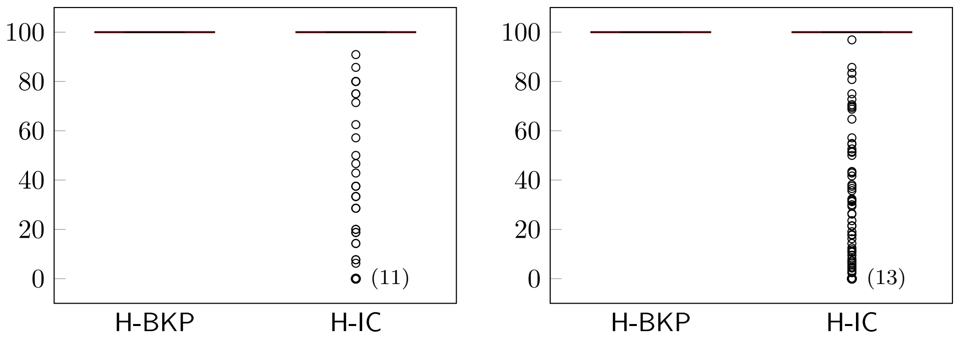

Figure 3. Box-plots of the percentages of solved sub-problems (out of the total number of sub-problems to be solved) for the approaches H-BKP and H-IC in the min-max setting with integer deviations (left) and continuous deviations (right).

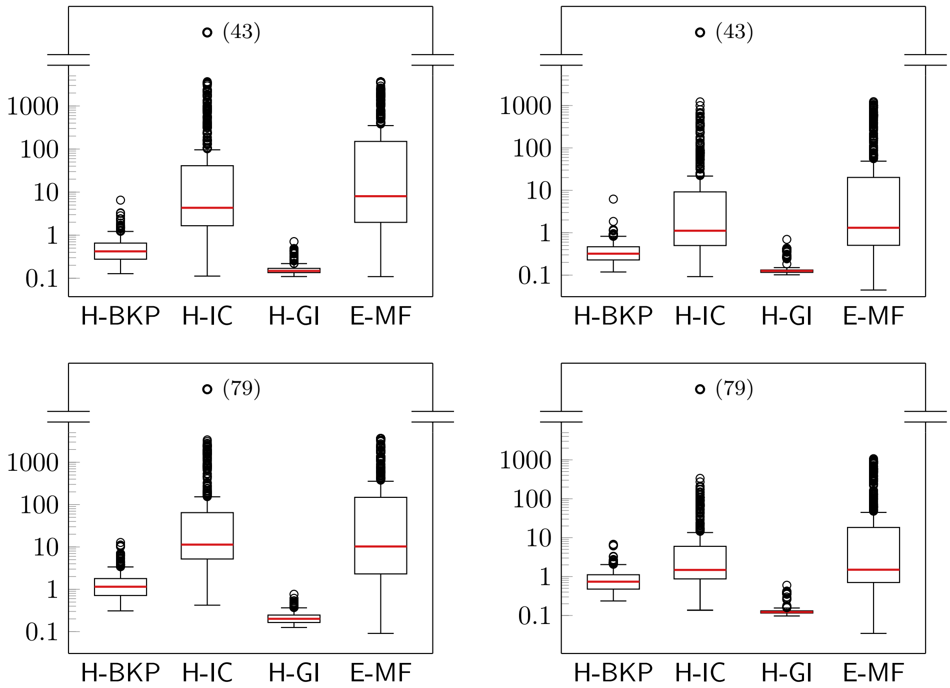

FIGURE 4. Box-plots of the sequential (left) and the idealized parallel runtimes (right) for the approaches H-BKP, H-IC, H-GI, and E-MF in the min-max setting with integer deviations (top) and continuous deviations (bottom). Sequential and idealized parallel runtimes (in s) are depicted on a log-scaled y-axis.

E-MF. Moreover, H-BKP optimally solves 35 instances that have not been solved to global optimality by E-MF. Overall, H-BKP proves global optimality for 98.93 % of the 560 instances; cf. Table 5. For those instances that have not been solved by H-BKP, we report a gap of at most 0.18 %. Nevertheless, global optimality can be verified ex post for all considered instances by comparing with the results obtained from E-MF; cf. Figure 2 (right).

The increased number of deterministic bilevel problems to solve significantly affects the performance of H-IC. The latter is particularly reflected by the number of instances for which a feasible point with finite optimality gap is found (560 instances for E-MF vs. 481 for H-IC); see Table 5. Hence, the solution of $|\mathcal{L}|$ bilevel problems (as it is done by H-IC) seems to be computationally more expensive in general than solving $|\mathcal{L}|$ lower-level, i.e., single-level, problems at each node of the branch-and-cut search tree as it is done by E-MF. Nevertheless, based on the instances that each method can handle, it is worth mentioning that both H-IC and E-MF exhibit similar median running times, which can be seen from the box-plots depicted in Figure 4 (bottom). Moreover, H-IC again seems to perform slightly more consistently than E-MF due to its overall smaller variability in running times.

For H-GI, we observe the same qualitative behavior as in the setting with integer deviations. Again, H-GI outperforms all other considered approaches in terms of running times but the solution quality is rather poor; see Figures 1 and 2 (right) as well as Figure 4 (bottom).

To conclude, the heuristic presented in this paper frequently proves optimality on the considered benchmark instances and, in particular, outperforms the "Greedy Interdiction" heuristic (DeNegre 2011) in terms of solution quality. However, the solution of the deterministic bilevel problems remains a bottleneck of our method and, thus, the algorithmic choice for solving these problems is crucial. When efficient black-box methods such as the bkpsolver (Weninger and Fukasawa 2023) are available to tackle the deterministic bilevel problems, our heuristic further outperforms the exact branch-and-cut method proposed in Beck et al. (2023a) both in terms of runtimes and solution quality.

5.4. **Evaluation of the Heuristic for General Bilevel Problems.** We now evaluate the performance of the heuristic for general Γ-robust bilevel problems. Before we start, let us mention that the structure of the lower-level problem from the min-max setting is preserved by the construction of our instances; see Section 5.1. Hence, the results for the reduction of sub-problems applied to the min-max setting summarized in Tables 3 and 4 are exactly the same in the more general setting. In particular, this means that a considerable number of deterministic bilevel problems can be eliminated in the setting with integer deviations, which significantly reduces the computational burden of our method.

In Table 6, we show the number of instances for which (i) a feasible point with finite gap is found ("feasible"), (ii) global optimality is proven either by a closed gap or using the sufficient optimality condition in Theorem 5 ("optimal"), (iii) the sufficient optimality condition is satisfied ("Thm. 5"), and (iv) the computed solution has a finite but non-zero gap ("open gap"). For those instances with an open gap, we further provide the average gap ("average gap"). In Figures 6 and 7, we show box-plots of the running times and the optimality gaps for all four considered approaches, respectively. In addition, we provide box-plots of the percentages of solved sub-problems (out of the total number of sub-problems to be solved) within H as well as box-plots of the ex-post optimality gaps for H, H-OS, and H-IT in Figures 5 and 8, respectively. As before, the ex-post optimality gaps are derived by comparing the heuristic solutions with the solution obtained from the exact branch-and-cut method. We now assess the performance of the four considered approaches.

While a feasible point with finite gap has been found by H-OS and H-IT for all considered instances, H could not compute a finite gap within the time limit of 1 h for around 55.54 % and 61.25 % of the considered instances in the setting with integer and continuous deviations, respectively; see Table 6. As before, we obtain an infinite gap in the case in which the solution of the deterministic bilevel problems exceeds the time limit so that the upper bound, initially being set to infinity, is not updated. To provide further insight into the time consumption of H, we additionally show box-plots of the percentages of sub-problems solved within the time limit (out of the total number of sub-problems to be solved) in Figure 5. Overall, the previous observations underline that the computational burden of our heuristic is quite large in the more general bilevel setting.

Despite this drawback, however, the heuristic presented in this paper still offers the advantage to parallelize the solution process of the deterministic bilevel problems and, if necessary, the solution of the additional lower-level problems. In Figure 6, we show box-plots for the running times for the instances that each method can handle, i.e., they find a feasible point with finite gap. Comparing the box-plots of H-seq and H-ideal clearly illustrates the potential of parallelization. Moreover, on the instances that the methods can tackle, H performs significantly better than E and H-IT both w.r.t. sequential and idealized parallel runtimes. It is important to note, however, that H-IT only terminates when reaching the time limit of 1 h, which is due to the method's design. Figure 6 further shows that H-OS performs significantly

Table 6. The number of instances for which a feasible point with finite gap is found ("feasible"; out of the 560 considered instances in the general bilevel setting with integer and continuous deviations, respectively) for the approaches H, H-OS, H-IT, and E. Additionally, the number of instances solved to global optimality ("optimal"), along with the number of instances satisfying the sufficient optimality condition in Thm. 5, is shown. For those instances with finite but non-zero gap ("open gap"), also the average gap ("average gap"; in %) is shown.

Δf		feasible	optimal	Thm. 5	open gap	average gap
integer	H	249	186	70	63	1.94
	H-OS	560	0	–	560	100.00
	H-IT	560	0	–	560	100.00
	E	480	236	–	244	23.05
continuous	H	217	172	58	45	2.10
	H-OS	560	0	–	560	100.00
	H-IT	560	0	–	560	100.00
	E	474	230	–	244	22.01

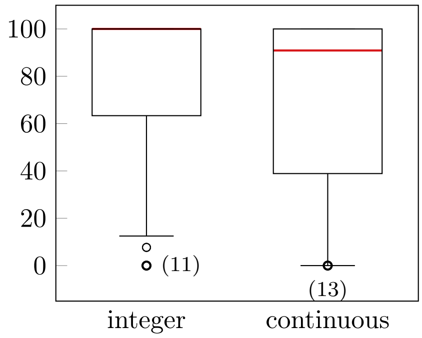

Figure 5. Box-plots of the percentages of solved sub-problems (out of the total number of sub-problems to be solved) for H in the general bilevel setting with integer and continuous deviations.

better than all other approaches in terms of running times. Nevertheless, its outliers are widely scattered. Hence, the heuristic presented in this paper seems to have a slightly more consistent performance.

Let us now comment on the quality of obtained solutions. Table 6 shows that H solves 33.21 % of the instances to global optimality in the setting with integer deviations. Here, optimality is proven using the sufficient condition in Theorem 5 for 70 of the 186 solved instances (37.63 %). In particular, this means that optimality is guaranteed by only solving bilevel problems of the nominal type. For the majority of the instances solved to global optimality, however, this is not the case so that additional lower-level problems need to be considered; cf. Proposition 6. For the setting with continuous deviations, we obtain similar results. Among the 172 instances that H solves to global optimality, 58 instances (33.72 %) satisfy the requirements of Theorem 5, whereas Proposition 6 is used to prove optimality for the remaining ones. For those instances for which H provides a feasible but not provably optimal solution, we still observe favorable results in terms of solution quality. Based on

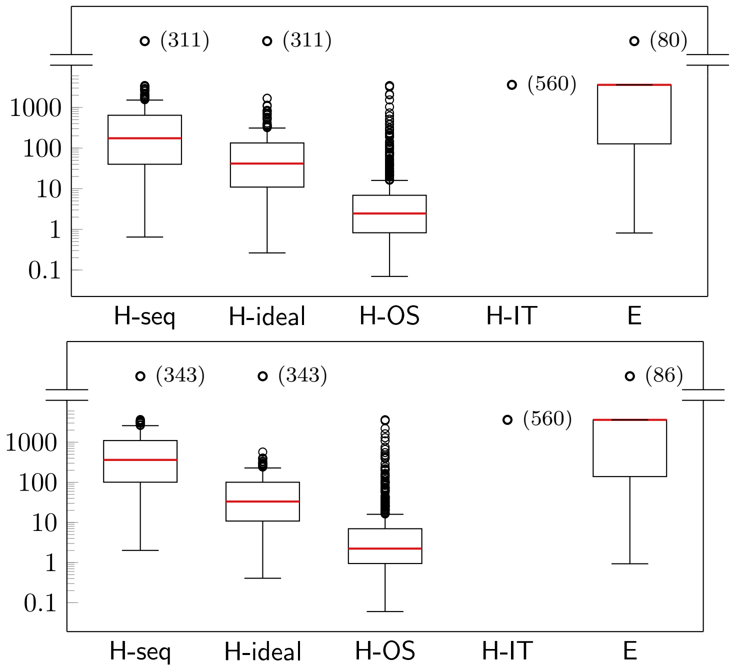

FIGURE 6. Box-plots of the runtimes for the approaches in the general bilevel setting with integer (top) and continuous deviations (bottom). Runtimes (in s) are depicted on a log-scaled y-axis.

Figure 7, it can be seen that we obtain a gap of at most 8.85 % and 11.35 % using H in the setting with integer and continuous deviations, respectively. The largest gaps we observe for E are 57.63 % and 58.10 % in the setting with integer and continuous deviations, respectively. However, the gaps obtained by H-OS and H-IT are quite poor. The latter is due to the, in general, rather weak lower bound that can be obtained from solving the HPR of the original bilevel problem. To further assess the solution quality, Figure 8 thus also shows box-plots of the ex-post optimality gaps, which can be obtained by comparing with the exact solution computed by E. The advantage of H-OS and H-IT is their ability to find a feasible point with finite gap for all instances, whereas this is not the case for H. Whenever H finds a feasible point with finite gap, however, its solution quality is slightly better than that of H-IT and significantly better than that of H-OS. Nevertheless, H-OS and H-IT still provide promising results in terms of solution quality.

To sum up, we observe that the heuristic presented in this paper is faster than the exact branch-and-cut approach on those instances for which it finds a feasible point with finite gap. Nevertheless, reflected by the large portion of instances for which the heuristic cannot compute a finite gap, we acknowledge the significant computational burden of our method for the setting of general mixed-integer linear bilevel problems. In this context, the heuristic approaches H-OS and H-IT seem to provide a reasonable trade-off between time consumption and solution quality.

Overall, we see two significant differences between the min-max and the general bilevel setting. First, our ex-post optimality criteria are stronger in the min-max setting (cf. Tables 5 and 6). This mainly influences the number of instances for which we can decide ex post that we have indeed computed an optimal solution. Second, our methods in both settings heavily rely on the respective solvers for the corresponding deterministic setting. While the bkpsolver by Weninger and Fukasawa

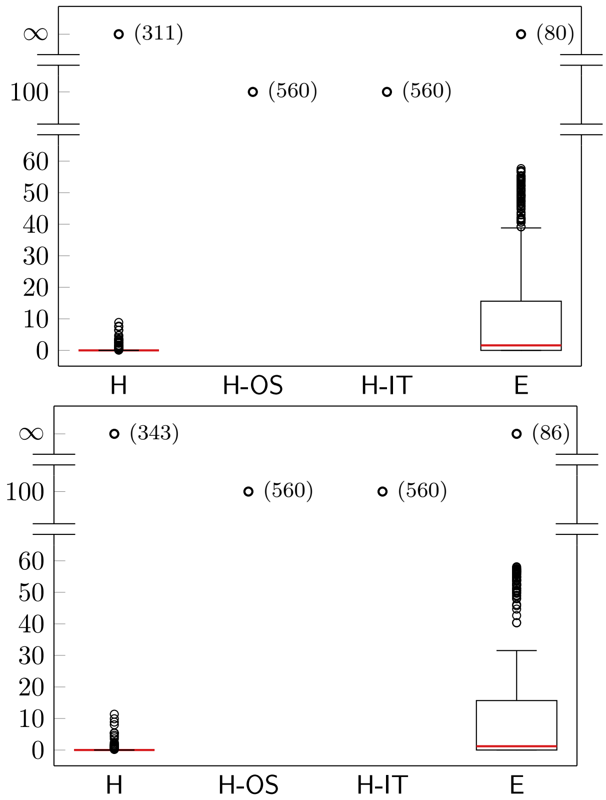

FIGURE 7. Box-plots of the optimality gaps (in %) for the approaches in the general bilevel setting with integer (top) and continuous deviations (bottom).

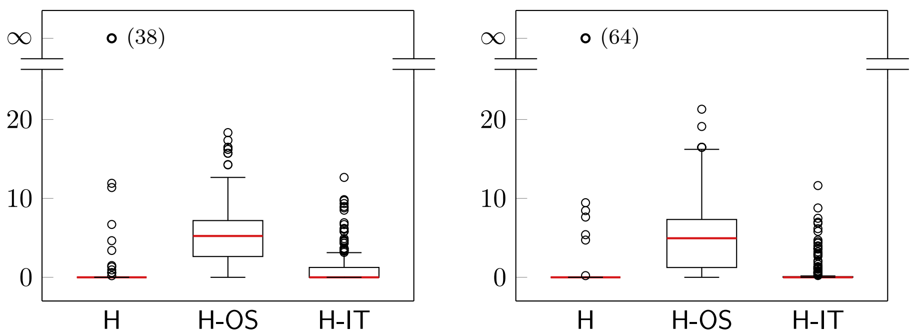

FIGURE 8. Box-plots of the ex-post optimality gaps (in %) for the approaches H, H-OS, and H-IT for the 236 instances with integer (left) and the 230 instances with continuous deviations (right) that have been solved to global optimality by E.

(2023) significantly speeds up our methods for the min-max setting, any future advancements in the field of general mixed-integer linear bilevel problems may be beneficial for our heuristic for the more general setting as well.

6. Conclusion

In this paper, we consider mixed-integer linear bilevel problems with a follower facing uncertainties regarding his objective function coefficients. To deal with this kind of uncertainty, we pursue a Γ-robust approach in which the follower hedges against a subset of the uncertain parameters that adversely influence the solution to the problem. More specifically, we exploit the main result by Bertsimas and Sim (2003) and Sim (2004) for Γ-robust single-level optimization—namely that the Γ-robust counterpart of a binary problem can be solved by solving a linear number of binary problems of the nominal type. We present heuristic methods for Γ-robust bilevel problems in the spirit of the Bertsimas–Sim result, wherein a linear number of bilevel problems of the nominal type is solved. Moreover, quality guarantees for heuristically obtained solutions as well as sufficient ex-post conditions for global optimality are provided. To assess the performance of our approaches, we conduct an extensive computational study on a total number of 2240 instances, comprising 1120 instances of the Γ-robust knapsack interdiction problem and 1120 more general Γ-robust bilevel instances. We observe that our heuristics often practically outperform alternative approaches adapted from the literature, including both heuristic and exact methods, in terms of the solution quality. In particular, the optimality gap is closed for a substantial part of the considered instances using the heuristics presented in this paper. A bottleneck of our methods, however, is the solution of the deterministic bilevel problems. Thus, the algorithmic choice for solving these problems is crucial. When efficient black-box methods are available to tackle the deterministic bilevel problems, our heuristic can outperform generic exact branch-and-cut methods. In particular, for Γ-robust knapsack interdiction problems, we report significant speed-up factors when compared to recently published problem-tailored and exact solution approaches. Nevertheless, more general Γ-robust bilevel problems remain challenging so that any algorithmic advances for general mixed-integer linear bilevel problems may be beneficial for the methods presented in this paper.

Acknowledgment

The computations have been executed on the high performance cluster "Elwetritsch" at TU Kaiserslautern within the "Alliance of High Performance Computing Rheinland-Pfalz" (AHRP). We kindly acknowledge the support of RHRK (https://rz.rptu.de/en/). Moreover, we are very grateful to two anonymous reviewers for thoroughly reading our work. Their observations and valuable comments have significantly contributed to improving the quality of our paper.

Code and Data Availability

The code for the methods presented in this paper, along with the nominal instance data used for our computational study, is publicly available at https://github.com/YasmineBeck/gamma-robust-bilevel-heuristics.

References

Beck, Y., I. Ljubić, and M. Schmidt (2022). "A Brief Introduction to Robust Bilevel Optimization." In: *SIAG on Optimization Views and News* 30.2, pp. 1–10. URL: https://siagoptimization.github.io/assets/views/ViewsAndNews-30-2.pdf.

Beck, Y., I. Ljubić, and M. Schmidt (2023a). "Exact methods for discrete Γ-robust interdiction problems with an application to the bilevel knapsack problem." In: *Mathematical Programming Computation* 15, pp. 733–782. DOI: 10.1007/s12532-023-00244-6.

Beck, Y., I. Ljubić, and M. Schmidt (2023b). "A survey on bilevel optimization under uncertainty." In: *European Journal of Operational Research* 311.2, pp. 401–426. DOI: 10.1016/j.ejor.2023.01.008.

Ben-Tal, A. and A. Nemirovski (1998). "Robust Convex Optimization." In: *Mathematics of Operations Research* 23.4, pp. 769–805. DOI: 10.1287/moor.23.4.769.

Ben-Tal, A., L. Ghaoui, and A. Nemirovski (2009). *Robust Optimization*. DOI: 10.1515/9781400831050.

Bertsekas, D. P. (2009). *Convex Optimization Theory*. Belmont, Massachusetts: Athena Scientific.

Bertsimas, D., D. B. Brown, and C. Caramanis (2011). "Theory and Applications of Robust Optimization." In: *SIAM Review* 53.3, pp. 464–501. DOI: 10.1137/080734510.

Bertsimas, D. and M. Sim (2003). "Robust discrete optimization and network flows." In: *Mathematical Programming* 98, pp. 49–71. DOI: 10.1007/s10107-003-0396-4.

Bertsimas, D. and M. Sim (2004). "The Price of Robustness." In: *Operations Research* 52.1, pp. 35–53. DOI: 10.1287/opre.1030.0065.

Birge, J. R. and F. Louveaux (2011). *Introduction to Stochastic Programming*. Springer-Verlag New York. DOI: 10.1007/978-1-4614-0237-4.

Brown, G., M. Carlyle, J. Salmerón, and R. Wood (2006). "Defending Critical Infrastructure." In: *Interfaces* 36, pp. 530–544. DOI: 10.1287/inte.1060.0252.

Caprara, A., M. Carvalho, A. Lodi, and G. J. Woeginger (2016). "Bilevel Knapsack with Interdiction Constraints." In: *INFORMS Journal on Computing* 28.2, pp. 319–333. DOI: 10.1287/ijoc.2015.0676.

Cormican, K. J., D. P. Morton, and R. K. Wood (1998). "Stochastic Network Interdiction." In: *Operations Research* 46.2, pp. 184–197. DOI: 10.1287/opre.46.2.184.

Cornuéjols, G. (2008). "Valid inequalities for mixed integer linear programs." In: *Mathematical Programming* 112, pp. 3–44. DOI: 10.1007/s10107-006-0086-0.

Dempe, S. (2002). *Foundations of Bilevel Programming*. Springer US. DOI: 10.1007/b101970.

Dempe, S. (2020). "Bilevel Optimization: Theory, Algorithms, Applications and a Bibliography." In: *Bilevel Optimization: Advances and Next Challenges*. Ed. by S. Dempe and A. Zemkoho. Springer International Publishing, pp. 581–672. DOI: 10.1007/978-3-030-52119-6_20.

DeNegre, S. T. (2011). "Interdiction and Discrete Bilevel Linear Programming." PhD thesis. Lehigh University. URL: https://coral.ise.lehigh.edu/~ted/files/papers/ScottDeNegreDissertation11.pdf.

Fischetti, M., I. Ljubić, M. Monaci, and M. Sinnl (2017). "A New General-Purpose Algorithm for Mixed-Integer Bilevel Linear Programs." In: *Operations Research* 65.6, pp. 1615–1637. DOI: 10.1287/opre.2017.1650.

Fischetti, M., I. Ljubić, M. Monaci, and M. Sinnl (2019). "Interdiction Games and Monotonicity, with Application to Knapsack Problems." In: *INFORMS Journal on Computing* 31.2, pp. 390–410. DOI: 10.1287/ijoc.2018.0831.

Fischetti, M., M. Monaci, and M. Sinnl (2018). "A dynamic reformulation heuristic for Generalized Interdiction Problems." In: *European Journal of Operational Research* 267.16, pp. 40–51. DOI: 10.1016/j.ejor.2017.11.043.

Furini, F., I. Ljubić, E. Malaguti, and P. Paronuzzi (2021). "Casting light on the hidden bilevel combinatorial structure of the capacitated vertex separator problem." In: *Operations Research* 70.4, pp. 2399–2420. DOI: 10.1287/opre.2021.2110.

Grüne, C. and L. Wulf (2024). *Completeness in the Polynomial Hierarchy for many natural Problems in Bilevel and Robust Optimization*. Tech. rep. URL: https://arxiv.org/abs/2311.10540.

Hansen, P., B. Jaumard, and G. Savard (1992). "New branch-and-bound rules for linear bilevel programming." In: *SIAM Journal on Scientific and Statistical Computing* 13.5, pp. 1194–1217. DOI: 10.1137/0913069.

Israeli, E. and R. K. Wood (2002). "Shortest-Path Network Interdiction." In: *Networks* 40.2, pp. 97–111. DOI: 10.1002/net.10039.

Jeroslow, R. G. (1985). "The polynomial hierarchy and a simple model for competitive analysis." In: *Mathematical Programming* 32.2, pp. 146–164. DOI: 10.1007/BF01586088.

Kall, P. and S. W. Wallace (1994). *Stochastic Programming*. Wiley-Interscience Series in Systems and Optimization. New York: Wiley.

Kleinert, T., M. Labbé, I. Ljubić, and M. Schmidt (2021). "A Survey on Mixed-Integer Programming Techniques in Bilevel Optimization." In: *EURO Journal on Computational Optimization* 9. DOI: 10.1016/j.ejco.2021.100007.

Lee, T. and C. Kwon (2014). "A short note on the robust combinatorial optimization problems with cardinality constrained uncertainty." In: *4OR* 12, pp. 373–378. DOI: 10.1007/s10288-014-0270-7.

Martello, S., D. Pisinger, and P. Toth (1999). "Dynamic Programming and Strong Bounds for the 0-1 Knapsack Problem." In: *Management Science* 45.3, pp. 414–424. DOI: 10.1287/mnsc.45.3.414.

Miranda, E. Álvarez, I. Ljubić, and P. Toth (2013). "A note on the Bertsimas & Sim algorithm for robust combinatorial optimization problems." In: *4OR* 11, pp. 349–360. DOI: 10.1007/s10288-013-0231-6.

Pisinger, D. and P. Toth (1998). "Knapsack Problems." In: *Handbook of Combinatorial Optimization*. Springer, Boston, MA. DOI: 10.1007/978-1-4613-0303-9_5.

Schmidt, M. and Y. Beck (2023). *A Gentle and Incomplete Introduction to Bilevel Optimization*. URL: https://optimization-online.org/?p=17182.

Sim, M. (2004). "Robust Optimization." PhD thesis. Massachusetts Institute of Technology, Sloan School of Management. URL: https://dspace.mit.edu/handle/1721.1/17725.

Soyster, A. L. (1973). "Technical Note—Convex Programming with Set-Inclusive Constraints and Applications to Inexact Linear Programming." In: *Operations Research* 21.5, pp. 1154–1157. DOI: 10.1287/opre.21.5.1154.

Tahernejad, S., T. K. Ralphs, and S. T. DeNegre (2020). "A branch-and-cut algorithm for mixed integer bilevel linear optimization problems and its implementation." In: *Mathematical Programming Computation* 12, pp. 529–568. DOI: 10.1007/s12532-020-00183-6.

von Stackelberg, H. (1932). *Marktform und Gleichgewicht*. Springer.

von Stackelberg, H. (1954). *The Theory of the Market Economy*. Oxford University Press.

Weninger, N. and R. Fukasawa (2023). "A Fast Combinatorial Algorithm for the Bilevel Knapsack Problem with Interdiction Constraints." In: *Integer Programming and Combinatorial Optimization*. Ed. by A. Del Pia and V. Kaibel. Vol. 13904. Lecture Notes in Computer Science, pp. 438–452. DOI: 10.1007/978-3-031-32726-1_31.

Wood, R. K. (2011). "Bilevel Network Interdiction Models: Formulations and Solutions." In: *Wiley Encyclopedia of Operations Research and Management Science*. DOI: 10.1002/9780470400531.eorms0932.

Appendix A. Omitted Proofs

Proof of Lemma 2. Along the lines of the proof of Theorem 3 by Bertsimas and Sim (2003), we obtain

$$\max_{\{S \subseteq [n_y] \colon |S| \leq \Gamma\}} \sum_{i \in S} \Delta f_i y_i = \min_{\ell \in \{1,\dots,n_y+1\}} \left\{ \Gamma \Delta f_\ell + \sum_{i=1}^{\ell} (\Delta f_i - \Delta f_\ell) y_i \right\}.$$

The latter yields

$$\begin{aligned} & f^\top y - \max_{\{S \subseteq [n_y] \colon |S| \leq \Gamma\}} \sum_{i \in S} \Delta f_i y_i \\ = {} & f^\top y - \min_{\ell \in \{1,\dots,n_y+1\}} \left\{ \Gamma \Delta f_\ell + \sum_{i=1}^{\ell} (\Delta f_i - \Delta f_\ell) y_i \right\} \\ = {} & f^\top y + \max_{\ell \in \{1,\dots,n_y+1\}} \left\{ -\Gamma \Delta f_\ell - \sum_{i=1}^{\ell} (\Delta f_i - \Delta f_\ell) y_i \right\} \\ = {} & \max_{\ell \in \{1,\dots,n_y+1\}} \left\{ -\Gamma \Delta f_\ell + \tilde{f}(\ell)^\top y \right\}, \end{aligned}$$

where the last equality follows from the definition of the robustified lower-level objective function coefficients in Lemma 1. The remainder of the proof now follows the one of Lemma 1 by Lee and Kwon (2014). To this end, we define

$$\varphi_\ell(y) := -\Gamma \Delta f_\ell + \tilde{f}(\ell)^\top y, \quad \ell \in \{1, \dots, n_y + 1\}.$$

Using this definition, we obtain

$$\varphi_{\ell+1}(y) - \varphi_\ell(y) = (\Delta f_\ell - \Delta f_{\ell+1}) \left(\Gamma - \sum_{i=1}^{\ell} y_i \right), \quad \ell \in \{1, \dots, n_y\}, \tag{7}$$

as well as

$$\varphi_\ell(y) - \varphi_{\ell-1}(y) = (\Delta f_{\ell-1} - \Delta f_\ell) \left(\Gamma - \sum_{i=1}^{\ell-1} y_i \right), \quad \ell \in \{2, \dots, n_y + 1\}. \tag{8}$$

In what follows, let $\ell \in \{2, \dots, n_y\}$ be given arbitrarily. We distinguish two cases. First, suppose that $\sum_{i=1}^{\ell} y_i \leq \Gamma$ holds. Hence, due to $y \geq 0$, we further have $\sum_{i=1}^{\ell-1} y_i \leq \Gamma$. From (7) and (8), we thus obtain

$$\varphi_{\ell-1}(y) \leq \varphi_\ell(y) \leq \varphi_{\ell+1}(y).$$

Second, let us assume that $\sum_{i=1}^{\ell} y_i > \Gamma$ holds, i.e., $\sum_{i=1}^{\ell} y_i \geq \Gamma + 1$ due to the integrality of y and Γ. In particular, we have $\sum_{i=1}^{\ell-1} y_i \geq \Gamma$. Thus, again by (7) and (8), we obtain

$$\varphi_{\ell-1}(y) \geq \varphi_\ell(y) \geq \varphi_{\ell+1}(y).$$

Let us finally note that, since y is binary, $\sum_{i=1}^{\ell} y_i \leq \Gamma$ holds for all $\ell \in \{1, \dots, \Gamma\}$. By our previous observations, we thus have $\varphi_\ell(y) \leq \varphi_{\Gamma+1}(y)$ for all $\ell \in \{1, \dots, \Gamma\}$. This concludes the proof. ☐

Proof of Lemma 3. By Assumption 2, we have $\Delta f_\ell \geq \Delta f_{\ell+1} \geq \cdots \geq \Delta f_k$. Hence, we obtain

$$\tilde{f}(\ell)_i = \begin{cases} f_i - \Delta f_i + \Delta f_\ell \geq f_i - \Delta f_i + \Delta f_k = \tilde{f}(k)_i, & 1 \leq i \leq \ell, \\ f_i \geq f_i - \Delta f_i + \Delta f_k = \tilde{f}(k)_i, & \ell + 1 \leq i \leq k, \\ f_i = \tilde{f}(k)_i, & k+1 \leq i \leq n_y, \end{cases}$$

which concludes the proof. □

Proof of Proposition 2. Due to Assumption 2 and Remark 2, it suffices to consider the case in which Algorithm 1 terminates in Line 12. To this end, let $(x^\ell)_{\ell \in \mathcal{L}}$ be the family of solutions to the min-max problems solved in Line 3 of the algorithm. Further, choose $\ell^* \in \mathcal{L}$ so that $L = c^\top x^{\ell^*} + \Phi_{\ell^*}(x^{\ell^*})$ holds. From Line 9, we thus obtain $U \leq c^\top x^{\ell^*} + \Phi_{\text{rob}}(x^{\ell^*})$. Let $y^* \in Y(x^{\ell^*})$ be an optimal solution to the x^{ℓ^*}-parameterized Γ-robust counterpart of the lower level. By Lemma 2, we get

$$\Phi_{\text{rob}}(x^{\ell^*}) = f^\top y^* - \max_{\{S \subseteq [n_y] :\, |S| \leq \Gamma\}} \sum_{i \in S} \Delta f_i y_i^* = \max_{\ell \in \mathcal{L}} \left\{ -\Gamma \Delta f_\ell + \tilde{f}(\ell)^\top y^* \right\}.$$

In addition, y^* is feasible for the ℓ^*-th lower-level sub-problem, i.e., we have

$$-\Gamma \Delta f_{\ell^*} + \tilde{f}(\ell^*)^\top y^* \leq \Phi_{\ell^*}(x^{\ell^*}).$$

Thus, taking all previous observations into account, we obtain

$$\begin{aligned} U - L &\leq c^\top x^{\ell^*} + \Phi_{\text{rob}}(x^{\ell^*}) - (c^\top x^{\ell^*} + \Phi_{\ell^*}(x^{\ell^*})) \\ &= \Phi_{\text{rob}}(x^{\ell^*}) - \Phi_{\ell^*}(x^{\ell^*}) \\ &= \max_{\ell \in \mathcal{L}} \left\{ -\Gamma \Delta f_\ell + \tilde{f}(\ell)^\top y^* \right\} - \Phi_{\ell^*}(x^{\ell^*}) \\ &\leq \max_{\ell \in \mathcal{L}} \left\{ \Gamma(\Delta f_{\ell^*} - \Delta f_\ell) + (\tilde{f}(\ell) - \tilde{f}(\ell^*))^\top y^* \right\} \\ &\leq \max_{\ell \in \mathcal{L}} \left\{ |\Gamma(\Delta f_{\ell^*} - \Delta f_\ell)| + \left| (\tilde{f}(\ell) - \tilde{f}(\ell^*))^\top y^* \right| \right\}. \end{aligned}$$

For all $\ell \in \mathcal{L}$, Assumption 2 as well as $\ell^* \in \mathcal{L}$ yield

$$\Gamma \left| \Delta f_{\ell^*} - \Delta f_\ell \right| = \begin{Bmatrix} \Gamma(\Delta f_{\ell^*} - \Delta f_\ell), \text{ if } \ell^* \leq \ell \\ \Gamma(\Delta f_\ell - \Delta f_{\ell^*}), \text{ if } \ell^* \geq \ell \end{Bmatrix} \leq \Gamma(\Delta f_{\Gamma+1} - \Delta f_{n_y+1}) = \Gamma \Delta f_{\Gamma+1}.$$

Moreover, we obtain

$$\begin{aligned} \left| (\tilde{f}(\ell) - \tilde{f}(\ell^*))^\top y^* \right| &\leq \|\tilde{f}(\ell) - \tilde{f}(\ell^*)\|_1 \|y^*\|_\infty \leq \|\tilde{f}(\ell) - \tilde{f}(\ell^*)\|_1 \\ &= \sum_{i=1}^{\Gamma+1} \left| \tilde{f}(\ell)_i - \tilde{f}(\ell^*)_i \right| + \sum_{i=\Gamma+2}^{n_y} \left| \tilde{f}(\ell)_i - \tilde{f}(\ell^*)_i \right| \\ &\leq (\Gamma+1)\Delta f_{\Gamma+1} + \sum_{i=\Gamma+2}^{n_y} \Delta f_i. \end{aligned}$$

Here, the first inequality follows from Hölder's inequality, whereas the second one follows from $y^* \in \{0,1\}^{n_y}$. The last inequality is due to the following. First, for all $\ell \in \mathcal{L}$ and $i \in [n_y]$, Lemma 3 yields

$$\left| \tilde{f}(\ell)_i - \tilde{f}(\ell^*)_i \right| = \begin{Bmatrix} \tilde{f}(\ell)_i - \tilde{f}(\ell^*)_i, \text{ if } \ell \leq \ell^* \\ \tilde{f}(\ell^*)_i - \tilde{f}(\ell)_i, \text{ if } \ell \geq \ell^* \end{Bmatrix} \leq \tilde{f}(\Gamma+1)_i - \tilde{f}(n_y+1)_i.$$

Second and lastly, Assumption 2 and Lemma 1 yield

$$\tilde{f}(\Gamma+1)_i - \tilde{f}(n_y+1)_i = \begin{cases} \Delta f_{\Gamma+1}, & 1 \leq i \leq \Gamma+1, \\ \Delta f_i, & \Gamma+1 \leq i \leq n_y. \end{cases}$$

To sum up,

$$|\Gamma(\Delta f_{\ell^*} - \Delta f_\ell)| + \left|(\tilde f(\ell) - \tilde f(\ell^*))^\top y^*\right| \le \Gamma \Delta f_{\Gamma+1} + (\Gamma+1)\Delta f_{\Gamma+1} + \sum_{i=\Gamma+2}^{n_y} \Delta f_i$$

holds for all $\ell \in \mathcal{L}$. Note that the right-hand side of the last inequality does not depend on the sub-problem index $\ell \in \mathcal{L}$. This concludes the proof.

Proof of Proposition 4. Let $x \in X$ be such that $x \ge x^\ell$ holds. Evidently, the claim is true for $x = x^\ell$. Hence, we assume $x \neq x^\ell$, i.e., there exists at least one index $i \in [n_x]$ with $x_i > x_i^\ell$. We define $X_>^\ell := \{i \in [n_x] \colon x_i > x_i^\ell\}$, i.e., $X_>^\ell \neq \emptyset$. Let y be an optimal solution to the x-parameterized ℓth lower-level sub-problem

$$\Phi_\ell(x) = -\Gamma \Delta f_\ell + \max_{y' \in Y(x)} \left\{ \tilde f(\ell)^\top y' \right\}.$$

Clearly, $y \in Y \subseteq \{0,1\}^{n_y}$ holds due to Part (ii) of Assumption 3. Moreover, we have $y \in Y(x^\ell)$ due to the following. First, suppose $X_>^\ell \cap I = \emptyset$. This means that the indices $i \in [n_x]$ with $x_i > x_i^\ell$ correspond to non-linking variables. Hence, we have $y_i \le 1 - x_i = 1 - x_i^\ell$ for all $i \in I$. Second, we assume that $X_>^\ell \cap I \neq \emptyset$ holds. From Part (i) of Assumption 3, we then have $1 = x_i > x_i^\ell = 0$ for all $i \in X_>^\ell \cap I$. Thus, we obtain

$$y_i \le 1 - x_i = \begin{cases} 0 < 1 = 1 - x_i^\ell, & i \in X_>^\ell \cap I, \\ 1 - x_i^\ell, & i \in I \setminus X_>^\ell. \end{cases}$$

Taking all previous considerations and Part (iii) of Assumption 3 into account yields

$$c^\top x + \Phi_\ell(x) = \Phi_\ell(x) = -\Gamma \Delta f_\ell + \sum_{i=1}^{n_y} \tilde f(\ell)_i y_i \le \Phi_\ell(x^\ell) = c^\top x^\ell + \Phi_\ell(x^\ell),$$

i.e., x solves (ℓ-Min-Max) as well.

Proof of Lemma 4. By Lemma 1, we have

$$f^\top y^* - \max_{\{S \subseteq [n_y] \colon |S| \le \Gamma\}} \sum_{i \in S} \Delta f_i y_i^* = \Phi_{\text{rob}}(x^*) = \max_{k \in \mathcal{L}} \{\Phi_k(x^*)\}.$$

Let ℓ_1, $\ell_2 \in \mathcal{L}$ be chosen such that

$$\ell_1 = \arg\max_{k \in \mathcal{L}} \{\Phi_k(x^*)\} \quad \text{and} \quad \ell_2 = \arg\max_{k \in \mathcal{L}} \left\{ -\Gamma \Delta f_k + \tilde f(k)^\top y^* \right\} \tag{9}$$

hold. Further, let y^{ℓ_i} denote an optimal solution to the x^*-parameterized ℓ_i-th lower-level sub-problem with $i \in \{1,2\}$, i.e., y^{ℓ_i} solves

$$\Phi_{\ell_i}(x^*) = -\Gamma \Delta f_{\ell_i} + \max_{y \in Y(x^*)} \left\{ \tilde f(\ell_i)^\top y \right\}.$$

Using (9), we thus obtain

$$\begin{aligned} -\Gamma \Delta f_{\ell_2} + \tilde f(\ell_2)^\top y^{\ell_2} &\le -\Gamma \Delta f_{\ell_1} + \tilde f(\ell_1)^\top y^{\ell_1} \\ &= \Phi_{\ell_1}(x^*) \\ &= f^\top y^* - \max_{\{S \subseteq [n_y] \colon |S| \le \Gamma\}} \sum_{i \in S} \Delta f_i y_i^*. \end{aligned}$$

Moreover, we have

$$f^\top y^* - \max_{\{S \subseteq [n_y] \colon |S| \le \Gamma\}} \sum_{i \in S} \Delta f_i y_i^* = -\Gamma \Delta f_{\ell_2} + \tilde f(\ell_2)^\top y^* \le -\Gamma \Delta f_{\ell_2} + \tilde f(\ell_2)^\top y^{\ell_2}.$$

Here, the equality follows from Lemma 2, whereas the inequality follows from the optimality of y^{ℓ_2} for the x^*-parameterized ℓ_2th lower-level sub-problem. The last two displayed formulas yield

$$-\Gamma\Delta f_{\ell_1} + \tilde{f}(\ell_1)^\top y^{\ell_1} = -\Gamma\Delta f_{\ell_2} + \tilde{f}(\ell_2)^\top y^* = -\Gamma\Delta f_{\ell_2} + \tilde{f}(\ell_2)^\top y^{\ell_2}.$$

This implies that, in (9), we could have chosen

$$\arg\max_{k\in\mathcal{L}}\{\Phi_k(x^*)\} = \ell_2 = \arg\max_{k\in\mathcal{L}}\left\{-\Gamma\Delta f_k + \tilde{f}(k)^\top y^*\right\},$$

which concludes the proof.

Proof of Proposition 7. Let $(x^\ell, y^\ell)_{\ell\in\mathcal{L}}$ be the family of solutions to the bilevel problems solved in Line 3 of Algorithm 3. Further, choose $k \in \mathcal{L}$ so that $L = c^\top x^k + d^\top y^k$ holds. According to the updating rule in Line 9 of Algorithm 3, we have $U \leq c^\top x^k + d^\top \hat{y}$ with $\hat{y}$ being an optimal solution to the Γ-robust counterpart of the x^k-parameterized lower level. Due to Hölder's inequality as well as $\hat{y}_i - y_i^k \in \{-1, 0, 1\}$ for all $i \in [n_y]$, we thus obtain

$$\begin{aligned} U - L &\leq c^\top x^k + d^\top \hat{y} - (c^\top x^k + d^\top y^k) \\ &= d^\top(\hat{y} - y^k) \\ &\leq \left|d^\top(\hat{y} - y^k)\right| \\ &\leq \|d\|_1 \|\hat{y} - y^k\|_\infty \leq \|d\|_1, \end{aligned}$$

which concludes the proof.

Appendix B. An Exact Branch-and-Cut Approach for Generalized Γ-Robust Knapsack Interdiction Problems

We consider a generalization of the knapsack interdiction problem studied in Caprara et al. (2016). The deterministic problem reads

$$\begin{aligned} \min_{x\in\{0,1\}^n, y} \quad & c^\top x + d^\top y \\ \text{s.t.} \quad & v^\top x \leq B, \\ & y \in \arg\max_{y'\in\{0,1\}^n} \left\{f^\top y' \colon w^\top y' \leq C,\ y'_i \leq 1 - x_i,\ i \in [n]\right\} \end{aligned}$$

with $B, C \in \mathbb{Z}_{\geq 0}$, and $c, d, f, v, w \in \mathbb{Z}^n_{\geq 0}$. For a given $x \in X$, the Γ-robust counterpart of the lower-level problem, in which the follower hedges against his uncertain objective function coefficients, is given by

$$\max_{y\in\{0,1\}^n} \quad f^\top y - \max_{\{S\subseteq[n]\colon |S|\leq\Gamma\}} \sum_{i\in S} \Delta f_i y_i \quad \text{s.t.} \quad w^\top y \leq C,\ y_i \leq 1 - x_i,\ i \in [n].$$

Along the lines of the proof of Theorem 3 by Bertsimas and Sim (2003), the latter can be re-written as

$$\begin{aligned} \max_{y,z,\theta} \quad & f^\top y - \Gamma\theta - \sum_{i=1}^n z_i \\ \text{s.t.} \quad & w^\top y \leq C, \\ & y_i \leq 1 - x_i, \quad i \in [n], \\ & z_i + \theta \geq \Delta f_i y_i, \quad i \in [n], \\ & y \in \{0,1\}^n,\ z \in \mathbb{R}^n_{\geq 0},\ \theta \in \mathbb{R}_{\geq 0}; \end{aligned} \tag{10}$$

cf. Lemma 1 in Beck et al. (2023a). The Γ-robust counterpart of the overall generalized knapsack interdiction problem thus reads

$$\min_{x\in\{0,1\}^n,y,z,\theta} \quad c^\top x + d^\top y \quad \text{s.t.} \quad v^\top x \le B,\ (y,z,\theta) \in S(x), \tag{11}$$

where $S(x)$ denotes the set of optimal solutions to the x-parameterized mixed-integer linear problem (10). Problem (11) is a standard mixed-integer linear bilevel problem. Using the lower-level optimal-value function, Problem (11) can be stated as the single-level problem

$$\begin{aligned} \min_{x,y,z,\theta} \quad & c^\top x + d^\top y \\ \text{s.t.} \quad & v^\top x \le B, \quad w^\top y \le C, \\ & y_i \le 1 - x_i, \quad i \in [n], \\ & z_i + \theta \ge \Delta f_i y_i, \quad i \in [n], \\ & f^\top y - \Gamma\theta - \sum_{i=1}^n z_i \ge \Phi_{\text{rob}}(x), \\ & x,\, y \in \{0,1\}^n, z \in \mathbb{R}^n_{\ge 0},\, \theta \in \mathbb{R}_{\ge 0}. \end{aligned} \tag{12}$$

Here, $\Phi_{\text{rob}}(x)$ is used to denote the optimal-value function associated with Problem (10). We can solve Problem (12) using a branch-and-cut framework. At node j of the branch-and-cut tree, we consider the problem

$$\begin{aligned} \min_{x,y,z,\theta} \quad & c^\top x + d^\top y \\ \text{s.t.} \quad & v^\top x \le B, \quad w^\top y \le C, \\ & y_i \le 1 - x_i, \quad i \in [n], \\ & z_i + \theta \ge \Delta f_i y_i, \quad i \in [n], \\ & (x,y,z,\theta) \in \Omega_j \subseteq [0,1]^n \times [0,1]^n \times \mathbb{R}^n_{\ge 0} \times \mathbb{R}_{\ge 0}, \end{aligned} \tag{P$_j$}$$

where the set Ω_j contains all valid inequalities that have been added previously to cut off integer-infeasible or bilevel-infeasible points as well as all branching decisions. If Problem (P$_j$) is infeasible for node j or if the objective function value corresponding to an optimal solution $(x^j, y^j, z^j, \theta^j)$ exceeds the current upper bound, we can fathom the node. Otherwise, we check for integer and bilevel feasibility. To separate a fractional solution, we can either branch or exploit standard cutting planes from mixed-integer linear optimization, e.g., as elaborated in Cornuéjols (2008). To check for bilevel feasibility, we compute the optimal objective function value of the x^j-parameterized Γ-robust counterpart of the lower-level problem. Using Proposition 6 in Beck et al. (2023a), this can be done by solving the mixed-integer linear problem

$$\begin{aligned} \max_{y,z,\theta} \quad & \sum_{i=1}^n f_i y_i (1 - x_i^j) - \Gamma\theta - \sum_{i=1}^n z_i \\ \text{s.t.} \quad & w^\top y \le C, \\ & y_i \le 1 - x_i^j, \quad i \in [n], \\ & z_i + \theta \ge \Delta f_i y_i, \quad i \in [n], \\ & y \in \{0,1\}^n,\, z \in \mathbb{R}^n_{\ge 0},\, \theta \in \mathbb{R}_{\ge 0}. \end{aligned} \tag{13}$$

Let $(\hat{y}, \hat{z}, \hat{\theta})$ denote an optimal solution to (13) and let $\hat{\Phi}$ denote the corresponding objective function value. If $\hat{\Phi} < \Phi_{\text{rob}}(x^j)$ holds, the point $(x^j, y^j, z^j, \theta^j)$ is bilevel-infeasible. To separate bilevel-infeasible points, we generate a cut of the form

$$\sum_{i=1}^{n} f_i y_i - \Gamma\theta - \sum_{i=1}^{n} z_i \geq \sum_{i=1}^{n} f_i \hat{y}_i (1 - x_i) - \Gamma\hat{\theta} - \sum_{i=1}^{n} \hat{z}_i$$

and add it to the description of the set Ω_j.

(Y. Beck, M. Schmidt) Trier University, Department of Mathematics, Universitätsring 15, 54296 Trier, Germany

Email address: `yasmine.beck@uni-trier.de`

Email address: `martin.schmidt@uni-trier.de`

(I. Ljubić) ESSEC Business School of Paris, Cergy-Pontoise, France

Email address: `ljubic@essec.edu`