Mixed-Integer Optimization Techniques for Robust Bilevel Problems with Here-and-Now Followers

Vom Fachbereich IV der Universität Trier zur Verleihung des akademischen Grades
Doktorin der Naturwissenschaften (Dr. rer. nat.) genehmigte Dissertation

von

Yasmine Beck

Trier, 2025

Betreuer	Prof. Dr. Martin Schmidt
1. Berichterstatter	Prof. Dr. Martin Schmidt
2. Berichterstatterin	Prof. Dr. Ivana Ljubić
3. Berichterstatter	Prof. Miguel F. Anjos, PhD
4. Berichterstatter	Prof. Dr. Volker Schulz
Datum der Disputation	10.12.2024

Acknowledgment

I would like to express my heartfelt gratitude to my supervisor and mentor, Martin Schmidt, for his constant support, guidance, patience, and kindness. I am truly grateful for the opportunity to learn from Martin's scientific rigor and enthusiasm, as well as for the trust he has placed in me since the very beginning of my PhD. Thank you, Martin, for all the invaluable experiences that have shaped me into the researcher I am today!

I would also like to sincerely thank Ivana Ljubić, not only for thoroughly reading and reviewing this thesis, but also for her encouragement, guidance, and support throughout my PhD journey.

Moreover, a special thanks goes to all my co-authors, Daniel Bienstock, Martine Labbé, Ivana Ljubić, Martin Schmidt, and Johannes Thürauf, for many fruitful discussions and for sharing their knowledge and experience with me.

I would like to extend my sincere gratitude to my thesis committee, Martin Schmidt, Ivana Ljubić, Miguel Anjos, Volker Schulz, and Leonhard Frerick, for their time, valuable feedback, and for fostering such a friendly and encouraging atmosphere during the final phase of my PhD.

A huge thanks also goes to all my colleagues at Trier University. In particular, I would like to thank Johannes Thürauf, Henri Lefebvre, and Andreas Horländer for many enlightening, profound, and funny conversations over "a few" coffees, and for proofreading parts of this thesis.

Finally, I would like to thank my family and friends for their unwavering support, understanding, and encouragement.

Curriculum Vitae

Personal Information	
Name	Yasmine Beck
Date and Place of Birth	August 1, 1995, Trier, Germany
Academic Career	
11/2020 – 10/2024	**Research and Teaching Assistant** Chair of Nonlinear Optimization Trier University, Trier, Germany
Education	
09/2020	**Master of Science, Applied Mathematics**
10/2018 – 09/2020	Trier University, Trier, Germany
03/2019	**Bachelor of Science, Business Mathematics**
10/2015 – 03/2019	Trier University, Trier, Germany
03/2015	**Abitur**
08/2006 – 03/2015	Gymnasium Konz, Konz, Germany

Abstract

In bilevel optimization, some of the variables of an optimization problem have to be an optimal solution to another nested optimization problem. This specific structure renders bilevel optimization a powerful tool for modeling hierarchical decision-making processes, which arise in various real-world applications such as in critical infrastructure defense, transportation, or energy. Due to their nested structure, however, bilevel problems are also inherently hard to solve—both in theory and in practice. Further challenges arise if, e.g., bilevel problems under uncertainty are considered.

In this dissertation, we address different types of uncertainties in bilevel optimization using techniques from robust optimization. We study mixed-integer linear bilevel problems with lower-level objective uncertainty, which we tackle using the notion of Γ-robustness. We present two exact branch-and-cut approaches to solve these Γ-robust bilevel problems, along with cuts tailored to the important class of monotone interdiction problems. Given the overall hardness of the considered problems, we additionally propose heuristic approaches for mixed-integer, linear, and Γ-robust bilevel problems. The latter rely on solving a linear number of deterministic bilevel problems so that no problem-specific tailoring is required. We assess the performance of both the exact and the heuristic approaches through extensive computational studies.

In addition, we study the problem of determining optimal tolls in a traffic network in which the network users hedge against uncertain travel costs in a robust way. The overall toll-setting problem can be seen as a single-leader multi-follower problem with multiple robustified followers. We model this setting as a mathematical problem with equilibrium constraints, for which we present a mixed-integer, nonlinear, and nonconvex reformulation that can be tackled using state-of-the-art general-purpose solvers. We further illustrate the impact of considering robustified followers on the toll-setting policies through a case study.

Finally, we highlight that the sources of uncertainty in bilevel optimization are much richer compared to single-level optimization. To this end, we study two aspects related to so-called decision uncertainty. First, we propose a strictly robust approach in which the follower hedges against erroneous observations of the leader's decision. Second, we consider an exemplary bilevel problem with a continuous but nonconvex lower level in which algorithmic necessities prevent the follower from making a globally optimal decision in an exact sense. The example illustrates that even very small deviations in the follower's decision may lead to arbitrarily large discrepancies between exact and computationally obtained bilevel solutions.

Author's Contribution

This dissertation is based on five peer-reviewed journal articles that have already been published, one submitted article that is currently under review, and one preprint. Part I of this dissertation consists of an extended summary that covers the main ideas, key findings, and the contributions of each article. The reprints of the articles are provided in Part II. Since all articles are written in co-authorship, the contributions of the author of this dissertation to these articles is highlighted in the following. The authors are always given in alphabetic order.

[YB1] Y. Beck, I. Ljubić, and M. Schmidt. "A Brief Introduction to Robust Bilevel Optimization." In: *SIAG on Optimization Views and News* 30.2 (2022). URL: `https://siagoptimization.github.io/assets/views/ViewsAndNews-30-2.pdf`

This article is a result of joint discussions between all three authors. The author of this dissertation developed all examples and wrote most parts of the article under the supervision of the other authors.

[YB2] Y. Beck, I. Ljubić, and M. Schmidt. "A survey on bilevel optimization under uncertainty." In: *European Journal of Operational Research* 311.2 (2023), pp. 401–426. DOI: `10.1016/j.ejor.2023.01.008`

The idea for this article was proposed by Ivana Ljubić and Martin Schmidt in joint discussions with the author of this thesis. All three authors contributed equally to the development and the writing of this survey article. The author of this dissertation is the primary author of Section 2 as well as Sections 3.2–3.5.

[YB3] Y. Beck, I. Ljubić, and M. Schmidt. "Exact methods for discrete Γ-robust interdiction problems with an application to the bilevel knapsack problem." In: *Mathematical Programming Computation* 15.4 (2023), pp. 733–782. DOI: `10.1007/s12532-023-00244-6`

The idea for this article was proposed by Ivana Ljubić and Martin Schmidt in joint discussions with the author of this thesis. All authors jointly developed the theoretical results and the algorithmic frameworks of the article. The author of this dissertation proved all propositions, lemmas, and theorems under the supervision of the other authors. In addition, she was responsible for the implementation of the presented methods, the generation of the benchmark instances, and for conducting the computational study. She also primarily wrote the article.

[YB4] Y. Beck, I. Ljubić, and M. Schmidt. *Heuristic Methods for Γ-Robust Mixed-Integer Linear Bilevel Problems.* Preprint. Revised and resubmitted. 2024. URL: https://optimization-online.org/?p=26186

The main idea for this article is a result of joint discussions between all three authors. All authors jointly developed the theoretical results and the algorithmic frameworks of the article. The author of this dissertation proved all propositions, lemmas, and theorems under the supervision of the other authors. She was further responsible for the implementation of the presented methods, the generation of the benchmark instances, and for conducting the computational study. In addition, she primarily wrote the article.

[YB5] Y. Beck, M. Labbé, and M. Schmidt. *A Toll-Setting Problem with Robust Wardrop Equilibrium Conditions Under Budgeted Uncertainty.* Preprint. 2024. URL: https://optimization-online.org/?p=26949

The idea for this article was proposed by Martine Labbé and Martin Schmidt in joint discussions with the author of this thesis. All authors jointly derived the modeling framework and the theoretical results of the article. The author of this dissertation proved all propositions, lemmas, and theorems under the supervision of the other authors. She was further responsible for the implementation of the presented models and for conducting the case study. She also primarily wrote the article.

The following peer-reviewed article has also been written during the doctoral studies of the author of this dissertation. While the contributions of the article are not at the core of the topics of this dissertation, which is a robust treatment of uncertainties in the lower level of bilevel problems, it still aligns with the broader framework of bilevel optimization under uncertainty. Therefore, the following article is included in this dissertation as well.

[YB6] Y. Beck, D. Bienstock, M. Schmidt, and J. Thürauf. “On a Computationally Ill-Behaved Bilevel Problem with a Continuous and Nonconvex Lower Level.” In: *Journal of Optimization Theory and Applications* 198 (2023), pp. 428–447. DOI: 10.1007/s10957-023-02238-9

The author of this dissertation contributed significantly to the development of the academic example presented in this article. She was particularly responsible for the implementation of the example and for proving all statements in the appendix. Moreover, she performed the analysis of analytic and computational solutions to the presented exemplary bilevel problem in joint discussions and under the close supervision of the other authors. In this regard, the author of this dissertation primarily wrote Sections 2–4 as well as Appendices A and B. In addition, she contributed to the development and writing of the results presented in Section 5 of the article.

To conclude, the following article has also been written and published during the doctoral studies of the author of this dissertation. It should be noted, however, that the core idea for the article stems from her master thesis. Nevertheless, the contributions of the article, including the consideration of a more general framework, a new theoretical result on the relation between limited observability and lower-level data uncertainty, and the consideration of bilevel bimatrix games, form novel extensions and go beyond the scope of the master thesis. Therefore, the following article is included in this dissertation as well.

[YB7] Y. Beck and M. Schmidt. “A robust approach for modeling limited observability in bilevel optimization.” In: *Operations Research Letters* 49.5 (2021), pp. 752–758. DOI: 10.1016/j.orl.2021.07.010

The idea for this article resulted from joint discussions between Martin Schmidt and the author of this dissertation within the scope of her master thesis. The author of this dissertation developed the theoretical results, which generalize and extend the results of her master thesis. She was further responsible for the implementation of the exemplary bilevel bimatrix game as well as its analysis. Under the supervision of Martin Schmidt, she also primarily wrote the article.

Contents

Part I

Extended Summary

1. Introduction

Optimization models are valuable tools that assist decision-makers in their decision-making processes. Classic, or single-level, optimization is used to model situations in which a single person or entity makes all decisions. While this is appropriate in many situations, there are also various real-world applications that involve more than one decision-maker. Often, a decision-maker has to decide in anticipation of the reaction of another decision-maker. The individual decisions may influence each other and, in addition, they may be motivated by conflicting objectives. Since classic optimization models do not cover situations in which multiple decision-makers interact, this has contributed to the development of *hierarchical* or *bilevel optimization*.

Bilevel optimization has its roots in economics and dates back to the seminal works by von Stackelberg (1934, 1952). However, in the field of mathematical optimization, it has been introduced much later by Bracken and McGill (1973). Bracken and McGill (1973) study hierarchical decision-making processes within military contexts, which they formulate as "mathematical programs with optimization problems in the constraints". This notion particularly reflects the essence of bilevel optimization, where some of the variables of an optimization problem have to be an optimal solution to another nested optimization problem. On the one hand, this nested structure makes bilevel optimization a powerful tool for modeling hierarchical decision-making processes as it allows to combine two different decision-makers in one model. On the other hand, however, it renders bilevel problems intrinsically hard to solve—both in theory and in practice. The challenges associated with bilevel problems have already been recognized in the early publication by Candler and Norton (1977). Complexity questions, however, have not been formally addressed until the mid-1980s; see Jeroslow (1985). In this context, Hansen et al. (1992) have shown that even linear bilevel problems, i.e., those with continuous variables, linear objective functions, and linear constraints, are strongly NP-hard in general. Despite their intrinsic hardness, researchers are increasingly interested in more and more complicated instantiations of bilevel problems to model situations in real-world applications. Additional challenges in bilevel optimization arise if, e.g., (i) mixed-integer aspects are involved, (ii) further nonlinearities or nonconvexities are introduced, or (iii) problems under uncertainty are considered. While the latter is at the core of this dissertation, we also incorporate aspects from (i) and (ii) in this thesis.

1.1 A Primer on Bilevel Optimization Under Uncertainty

In many practical applications, decision-makers are forced to make decisions under uncertainty due to various reasons. One possible source of uncertainty is that the problem

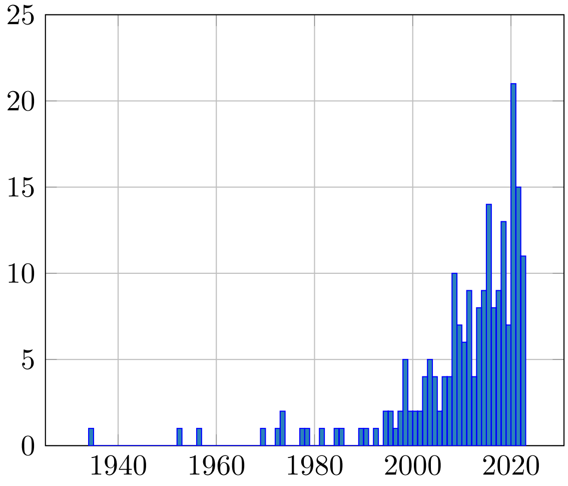

Figure 1.1: Papers per year as cited in the survey in [YB2]; see Figure 6 in [YB2].

data relies on predictions, which may not be accurate. For instance, travel times between two locations are based on estimates that may be subject to uncertainty due to unforeseen events such as accidents or changing weather conditions. Moreover, some parameters of the model may be practically impossible or too costly to measure exactly. In single-level optimization, there are two main approaches for dealing with such uncertainties: *stochastic optimization* (see, e.g., Birge and Louveaux (2011), Kall and Wallace (1994), and Shapiro et al. (2009)) and *robust optimization* (see, e.g., Ben-Tal and Nemirovski (1998), Ben-Tal et al. (2009), Bertsimas et al. (2011), and Soyster (1973)). The same two paths have been followed to address uncertainties in bilevel problems as well.

Although research on bilevel optimization under uncertainty has started rather recently in the 1990s, there have already been considerable contributions in this field. In Figure 1.1, we show the number of papers per year that have been cited in the survey in [YB2] on bilevel optimization under uncertainty, which illustrates the development of the field. Most of the papers discussed in [YB2] deal with bilevel problems that are subject to so-called *data uncertainty*, i.e., uncertainties in the data of the bilevel problem. To the best of our knowledge, the earliest papers that pursue a stochastic approach to address data uncertainty in bilevel problems are those by Cormican et al. (1998) and Patriksson and Wynter (1999, 1997). Especially over the last decade, however, a substantial amount of work has been conducted in the field of stochastic bilevel optimization. Notable contributions include, e.g., Bolusani et al. (2020), Ivanov (2018, 2014), Dempe et al. (2017), Yanıkoğlu and Kuhn (2018), Burtscheidt et al. (2020), or Burtscheidt and Claus (2020). While the majority of papers on bilevel optimization under uncertainty consider a stochastic setup, only a few works pursue a robust approach to deal with data uncertainty. Early works in this field include, e.g., Nikoofal and Zhuang (2011), Haghighat (2014), or Chuong and Jeyakumar (2017). Nevertheless, robust approaches to deal with uncertainties in bilevel optimization are still in their infancy.

In contrast to single-level optimization, the sources of uncertainty in bilevel optimization are much richer and, in particular, go beyond data uncertainty. Since bilevel optimization involves two decision-makers, there may be additional uncertainties associated with their respective decisions. This kind of uncertainty is referred to as *decision* *uncertainty*. We note that decision uncertainty does not play a role in single-level optimization, given that there is only one decision-maker involved. While data uncertainty has already received considerable attention in the context of bilevel optimization, decision uncertainty has been much less investigated, despite the various reasons contributing to it. One possible reason for decision uncertainty is that one decision-maker may not be able to observe or anticipate the reaction of the other decision-maker perfectly (see, e.g., Bagwell (1995), Molan and Schmidt (2023), Molan et al. (2023), Pita et al. (2010, 2008), and van Damme and Hurkens (1997)). Moreover, decision-makers are usually assumed to act as optimizers. However, limited intellectual or computational resources may prevent decision-makers from reacting optimally. Instead, approximately optimal or heuristic responses may be taken, which may lead to response uncertainty (see, e.g., Besançon et al. (2021, 2024), Shi et al. (2023), and Zare et al. (2020)). Another possible reason for decision uncertainty may stem from uncertainties regarding the level of cooperation between the decision-makers, who may act as adversaries, cooperators, or something in between (see, e.g., Aboussoror and Loridan (1995), Mallozzi and Morgan (1996), Salas and Svensson (2023), and Zeng (2020)). To conclude, let us mention that both data and decision uncertainty are naturally related to the broader notion of so-called *bounded rationality*. For an introduction to the topic, we refer to the seminal works by Simon (1955, 1956, 1972) and the book by Rubinstein (1998).

1.2 Contributions and Organization

In this dissertation, we summarize our contributions to the field of bilevel optimization under uncertainty in which we use techniques from robust optimization to address uncertainties in the lower-level problem. In particular, we focus on robust bilevel problems with a "here-and-now" follower, i.e., the follower has to decide before the uncertainty realizes. In Part I, we provide an extended summary of our contributions. All details on the presented approaches and all the proofs that we omit in Part I can be found in the respective original articles that are referenced throughout this thesis. The reprints of these articles are provided in Part II of this dissertation.

Part I of this dissertation is organized as follows. In Chapter 2, we review the foundations of bilevel and robust optimization, discuss reformulation techniques that are commonly used in practice, and elaborate on potential algorithmic challenges. Moreover, we discuss different modeling aspects for robust bilevel problems, which are at the intersection of bilevel and robust optimization problems. In Chapter 3, we present exact branch-and-cut approaches to solve mixed-integer linear bilevel problems with a Γ-robust treatment of lower-level objective uncertainty. In Chapter 4, we present heuristic approaches for these problems. In Chapter 5, we propose a robust approach to hedge against uncertain travel costs for the problem of determining optimal tolls in a traffic network. In Chapter 6, we show that the sources of uncertainty in bilevel optimization are much richer compared to single-level optimization using two exemplary instantiations related to decision uncertainty. Finally, in Chapter 7, we conclude Part I of this dissertation with a summary of our contributions and a discussion of possible future research directions.

2. Foundations of Bilevel and Robust Optimization

In this chapter, we elaborate on the basics of bilevel and robust optimization. In Section 2.1, which is based on Colson et al. (2007) and Dempe (2002), we introduce the concept of bilevel optimization. We discuss reformulation techniques for this class of optimization problems that are commonly used in practice and elaborate on potential challenges associated with them. Afterward, in Section 2.2, we briefly discuss the principles of robust optimization with a particular focus on the notions of strict and Γ-robustness. The content of this section is based on Ben-Tal et al. (2009), Bertsimas et al. (2011), and Bertsimas and Sim (2003). Finally, in Section 2.3, we address problems at the intersection of bilevel and robust optimization and summarize the discussions in [YB1].

2.1 Bilevel Optimization

In bilevel optimization, some of the variables of an optimization problem have to be an optimal solution to another nested optimization problem. This nested structure renders bilevel optimization a powerful tool for modeling hierarchical decision-making processes as it allows to combine two different decision-makers in one model. In bilevel optimization, one decision-maker (the leader) makes a decision while anticipating an optimal reaction of another decision-maker (the follower). The decisions of both the leader and the follower are interdependent, i.e., each influences the other. We consider a sequential interaction between the two decision-makers, i.e., we consider the timing

$$\text{leader} \quad \curvearrowright \quad \text{follower.}$$

This means that, in anticipation of the follower's response, the leader takes her decision first. Then, after observing the decision of the leader, the follower makes his own decision, taking the leader's choice into account. Throughout this dissertation, we use "her" for the leader and "his" for the follower.

2.1.1 Problem Statement

More formally, a bilevel problem is an optimization problem of the form

$$\begin{aligned} \text{“}\min_{x \in X}\text{”} \quad & F(x,y) && \text{(2.1a)} \\ \text{s.t.} \quad & G(x,y) \geq 0, && \text{(2.1b)} \\ & y \in S(x), && \text{(2.1c)} \end{aligned}$$

where $S(x)$ is the set of optimal solutions to the x-parameterized problem

$$\min_{y \in Y} \quad f(x, y) \tag{2.2a}$$

$$\text{s.t.} \quad g(x, y) \geq 0. \tag{2.2b}$$

The objective functions are given by $F, f : \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \to \mathbb{R}$ and the constraint functions are given by $G : \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \to \mathbb{R}^l$ and $g : \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \to \mathbb{R}^m$. The sets $X \subseteq \mathbb{R}^{n_x}$ and $Y \subseteq \mathbb{R}^{n_y}$ are used to impose integrality constraints on (some of) the variables x and y, respectively. We call

$$\{(x, y) \in X \times Y : G(x, y) \geq 0, \; y \in S(x)\}$$

the *bilevel feasible set*. We further define the so-called *shared constraint set* of the bilevel problem (2.1) as

$$\Omega := \{(x, y) \in X \times Y : G(x, y) \geq 0, \;\; g(x, y) \geq 0\}.$$

Its projection onto the x-space is given by

$$\Omega_x := \{x \in X : \exists y \text{ with } (x, y) \in \Omega\}.$$

We refer to Problem (2.1) as the leader's (or the upper-level) problem and to Problem (2.2) as the follower's (or the lower-level) problem. The variables x and y are called the leader's and the follower's variables, respectively. The set $S(x)$ is also called *rational reaction set* of the follower. In the case that $S(x)$ is not a singleton, i.e., the follower's response to the leader's decision is not unique, the bilevel problem (2.1) is ill-posed. This ambiguity is expressed by the quotation marks in (2.1a). To overcome this issue, it is common to pursue either an *optimistic* or a *pessimistic approach* to bilevel optimization. In the optimistic setting, it is assumed that the leader can influence the follower's response so that he selects a decision that favors the leader w.r.t. her objective function value. Formally, the cooperative nature of the follower is expressed by jointly optimizing over the variables x and y in the upper-level problem, i.e., the optimistic variant of the bilevel problem (2.1) is given by

$$\min_{x \in X, y} \quad F(x, y) \tag{2.3a}$$

$$\text{s.t.} \quad G(x, y) \geq 0, \tag{2.3b}$$

$$y \in S(x). \tag{2.3c}$$

Here, $S(x)$ again denotes the set of optimal solutions to the x-parameterized lower-level problem (2.2). In contrast, the leader anticipates that the follower will choose the worst possible solution for her in the pessimistic setting. The pessimistic approach to bilevel optimization has been addressed, e.g., in Aussel and Svensson (2019), Lampariello et al. (2019), Liu et al. (2020), Wiesemann et al. (2013), and Zeng (2020). In the case that we need to distinguish between an optimistic or a pessimistic follower, we focus on the optimistic setting in this dissertation.

2.1.2 Single-Level Reformulations

Most solution approaches in bilevel optimization rely on reformulating the bilevel problem as a classic, i.e., single-level, optimization problem. Then, after some potential further reformulations, these problems can be tackled using state-of-the-art general-purpose solvers. In this dissertation, we consider the following two approaches to obtain a single-level reformulation of the optimistic bilevel problem (2.3).

Value-Function Reformulation For a given leader's decision x, we define the *optimal-value function* of the lower-level problem (2.2) as

$$\Phi(x) := \min_{y \in Y} \left\{ f(x,y) \colon g(x,y) \geq 0 \right\}.$$

Using this notation, we can re-write the bilevel problem (2.3) as the single-level problem

$$\begin{aligned} \min_{x,y} \quad & F(x,y) && \text{(2.4a)} \\ \text{s.t.} \quad & (x,y) \in \Omega, && \text{(2.4b)} \\ & f(x,y) \leq \Phi(x). && \text{(2.4c)} \end{aligned}$$

Problem (2.4) is called the *value-function reformulation* of the bilevel problem (2.3). While Problem (2.4) may seem like an ordinary single-level problem, the lower-level optimal-value function poses considerable challenges. This is due to the fact that, even under strong assumptions, the optimal-value function is usually neither smooth nor known in a closed form. Hence, Constraint (2.4c), which ensures the optimality of the follower's response, still reflects the overall hardness of the bilevel problem. Despite this initial drawback, the value-function reformulation (2.4) is frequently used in practice, especially in mixed-integer bilevel optimization. For instance, many solution approaches for mixed-integer bilevel problems start by solving the so-called *high-point relaxation*

$$\min_{x,y} \quad F(x,y) \quad \text{s.t.} \quad (x,y) \in \Omega,$$

which is obtained from Problem (2.4) by omitting Constraint (2.4c). Afterward, bilevel-infeasible points are discarded, e.g., by branching, by adding cutting planes, or by approximating the lower-level optimal-value function; see, e.g., DeNegre and Ralphs (2009), Fischetti et al. (2017, 2019, 2018a), and Xu and Wang (2014).

KKT Reformulation Building on the seminal work by Fortuny-Amat and McCarl (1981), the most common approach in practice to reformulate Problem (2.3) as a single-level problem is to exploit optimality conditions of the lower level. This approach, however, is only viable if the follower's problem possesses a compact optimality certificate, which is both necessary and sufficient. For the remainder of this section, we thus assume that the lower-level problem (2.2) is a parametric convex optimization problem. The latter implies that, for a given leader's decisions x, the following holds:

(i) the objective function $y \mapsto f(x,y)$ is convex,

(ii) the constraint functions $y \mapsto g_i(x,y)$ are concave for all $i \in \{1,\dots,m\}$ with $g(x,y) = (g_i(x,y))_{i \in \{1,\dots,m\}}$, and

(iii) all follower's variables are continuous, i.e., $Y = \mathbb{R}^{n_y}$.

In addition, we assume that the objective and the constraint functions of the follower are continuously differentiable and that the lower-level problem satisfies a constraint qualification such as, e.g., Slater's constraint qualification for all $x \in \Omega_x$. Further details on constraint qualifications can, e.g., be found in Boyd and Vandenberghe (2004) and Nocedal and Wright (2006). Given the above assumptions, the Karush–Kuhn–Tucker (KKT) conditions are both necessary and sufficient optimality conditions for the x-parameterized lower-level problem. We can thus reformulate the bilevel problem (2.3) by replacing the lower level by its KKT conditions. The resulting single-level reformulation is then given by

$$
\begin{aligned}
\min_{x,y,\lambda} \quad & F(x,y) && \text{(2.5a)}\\
\text{s.t.} \quad & G(x,y) \geq 0,\ g(x,y) \geq 0, && \text{(2.5b)}\\
& \nabla_y \mathcal{L}(y,\lambda;x) = 0, && \text{(2.5c)}\\
& \lambda_i g_i(x,y) = 0, \quad i \in \{1,\dots,m\}, && \text{(2.5d)}\\
& x \in X,\ y \in Y,\ \lambda \in \mathbb{R}^m_{\geq 0}. && \text{(2.5e)}
\end{aligned}
$$

Here, we use $\nabla_y \mathcal{L}(y,\lambda;x)$ to denote the gradient of the Lagrangian function of the x-parameterized lower-level problem, i.e.,

$$\nabla_y \mathcal{L}(y,\lambda;x) = \nabla_y f(x,y) - \sum_{i=1}^{m} \lambda_i \nabla_y g_i(x,y).$$

Problem (2.5) is called the *KKT reformulation* of the bilevel problem (2.3). We emphasize that the bilevel problem (2.3) and its single-level KKT reformulation (2.5) are only equivalent on the level of globally optimal solutions; see, e.g., Theorems 2.1 and 2.3 in Dempe and Dutta (2012) for the details. Here, "equivalence" is to be understood in the following sense. For a global solution (x^*,y^*) to the bilevel problem (2.3), there exists λ^* such that (x^*,y^*,λ^*) is a global solution to Problem (2.5). Conversely, for a global solution (x^*,y^*,λ^*) to Problem (2.5), (x^*,y^*) is a global solution to the bilevel problem (2.3). In particular, to establish this relation, the assumption that the lower-level problem satisfies Slater's constraint qualification for all $x \in \Omega_x$ cannot be omitted.

Due to nonconvexities in Constraints (2.5c) and (2.5d), solving Problem (2.5) is not an easy task, even under the aforementioned convexity assumptions on the lower-level problem. While Constraints (2.5c) are linear equality constraints in the case of linear or convex-quadratic lower-level problems, the complementarity conditions (2.5d) render Problem (2.5) a nonlinear and nonconvex problem. Hence, computing global solutions to Problem (2.5) remains challenging even in this setting. Nevertheless, due to their intrinsic disjunctive nature, we can exploit techniques from mixed-integer optimization to tackle Constraints (2.5d). The latter is at the core of many enumeration-based algorithms such as, e.g., branch-and-bound (Land and Doig 1960). One approach that is frequently used in practice is to reformulate Constraints (2.5d) using auxiliary binary variables and sufficiently large big-M constants; see, e.g., Fortuny-Amat and McCarl (1981). However, a considerable drawback of this approach is the need for valid big-Ms. Choosing too small values for these constants may cut off lower-level optimal solutions, an issue that has been illustrated in Pineda and Morales (2019). Complexity questions related to finding and validating correct big-M constants have further been addressed in Buchheim (2023) and Kleinert et al. (2020), respectively. Another approach to tackle Constraints (2.5d), which does not require determining sufficiently large big-M constants, is the use of so-called special ordered sets of type 1 (SOS1). Further details on the latter two approaches as well as a computational comparison can, e.g., be found in Kleinert and Schmidt (2023).

Finally, we mention that one may also consider optimality certificates other than the KKT conditions to obtain a single-level reformulation of the bilevel problem (2.3). An alternative to the KKT approach, though less frequently used in practice, is to exploit a strong-duality theorem, if such a theorem is available for the lower-level problem at hand. Works that have pursued a strong-duality based approach include, e.g., Zare et al. (2019), Motto (2005), or Kleinert et al. (2021a).

2.2 Robust Optimization

In many practical applications, decision-making processes can be particularly challenging due to inherent uncertainties in real-world data. As a consequence, decision-makers are frequently forced to make decisions under uncertainty. Despite the fact that uncertainty can, in general, not be avoided when making decisions, the associated optimization models are usually considered under the assumption of certain data. However, even small perturbations in the data can render a decision sub-optimal or even infeasible for the problem at hand; see, e.g., the case study in Ben-Tal et al. (2009) for an illustrative example. Thus, the consideration of uncertainties in decision-making processes is of significant practical importance. One approach to deal with uncertainties in mathematical optimization is to exploit techniques from *robust optimization*, which has its origins in the seminal work by Soyster (1973). In robust optimization, it is assumed that the uncertainties only take values in a predefined *uncertainty set* $\mathcal{U}$, which is typically modeled using boxes, polyhedra, ellipsoids, or cones. The decision-maker adopts a worst-case-oriented approach by aiming for a decision that is feasible for all possible realizations of the uncertainty within the given uncertainty set. Among all these so-called *robust feasible* decisions, the decision-maker then selects the best one w.r.t. the objective function value.

Although the assumption of a predefined uncertainty set may seem paradoxical at first, it can indeed be reasonable in practical situations. Often, reliable estimates are available for the mean of the uncertain parameters or their variability so that, e.g., the uncertainty set could represent the confidence intervals of the uncertain parameters. Nevertheless, we note that other approaches are possible as well to deal with uncertainties. For instance, in *stochastic optimization*, it is assumed that the uncertainties can be described by probability distributions that are known in advance. The decision-maker then hedges against uncertainties in a probabilistic sense, e.g., by optimizing over expected values, by considering chance constraints, or some risk-averse models. Since we do not consider stochastic approaches in this dissertation, we refer to Birge and Louveaux (2011), Kall and Wallace (1994), and Shapiro et al. (2009) for further details. Finally, we mention that, in between stochastic and robust optimization, there is an additional approach known as *distributional robustness*; see, e.g., Goh and Sim (2010) and Wiesemann et al. (2014).

In this dissertation, we focus on strictly and Γ-robust approaches to deal with uncertainties in (mixed-integer) problems of the form

$$\max_{y} \quad f^\top y \quad \text{s.t.} \quad y \in Y \subseteq \mathbb{R}^{n_\mathrm{C}} \times \mathbb{Z}^{n_\mathrm{D}} \tag{2.6}$$

with $n_y = n_\mathrm{C} + n_\mathrm{D}$ and $f \in \mathbb{R}^{n_y}$. We refer to Problem (2.6) as the *nominal* or *deterministic* *problem*. In the following chapters, we commonly assume that the objective function coefficients of Problem (2.6) are uncertain but known to vary within a given uncertainty set $\mathcal{U}$. The feasible set Y is mainly considered to remain unaffected by uncertainties. Hence, to streamline the discussion, we focus on problems of the form given in (2.6) with objective uncertainty for the remainder of this section. Details on how to treat uncertainties in the constraints can, e.g., be found in Ben-Tal et al. (2009), Bertsimas et al. (2011), and Gorissen et al. (2015).

2.2.1 Strict Robustness

We are interested in a solution to the nominal problem (2.6) that is optimal for the worst possible realization of the uncertain objective function coefficients, which are known to take values in a predefined set $\mathcal{U}$. In particular, the decision-maker has to decide in a "here-and-now" fashion, i.e., before the uncertainty realizes. This leads us to considering the so-called *(strictly) robust counterpart* of Problem (2.6), which is given by

$$\max_{y \in Y} \min_{\tilde{f} \in \mathcal{U}} \tilde{f}^\top y. \tag{2.7}$$

Usually, the uncertainty set $\mathcal{U}$ contains infinitely many elements so that there may be infinitely many possible objective functions to consider. One of the main approaches in robust optimization to solve Problem (2.7) is thus to reformulate the problem as an equivalent computationally tractable one. In this context, we emphasize that the geometry of the uncertainty set $\mathcal{U}$ may significantly affect the tractability of the robust counterpart (2.7). For discussions on the tractability of robust counterparts w.r.t. different uncertainty set geometries, we refer to Ben-Tal et al. (2009) and Bertsimas et al. (2011). Often, a tractable reformulation of the robust counterpart (2.7) is obtained by using duality theory. We now briefly illustrate such a duality-based approach. To this end, we assume that the uncertainty set is parameterized in an affine way by

$$\mathcal{U} = \{f + P\zeta \colon \zeta \in \mathcal{Z} \subseteq \mathbb{R}^q\}$$

with the vector of *nominal values* $f \in \mathbb{R}^{n_y}$, a *perturbation matrix* $P \in \mathbb{R}^{n_y \times q}$, and a polyhedral *perturbation set* $\mathcal{Z} = \{\zeta \in \mathbb{R}^q \colon H\zeta \geq h\}$ with $H \in \mathbb{R}^{s \times q}$ and $h \in \mathbb{R}^s$. By the affine parameterization of the uncertainty set $\mathcal{U}$, we can re-write Problem (2.7) as

$$\max_{y \in Y} \left\{ f^\top y + \min_{\zeta \in \mathbb{R}^q} \left\{ (P^\top y)^\top \zeta \colon H\zeta \geq h \right\} \right\} \tag{2.8a}$$

$$= \max_{y \in Y} \left\{ f^\top y + \max_{z \in \mathbb{R}^s} \left\{ h^\top z \colon H^\top z = P^\top y,\ z \geq 0 \right\} \right\}. \tag{2.8b}$$

The equality in (2.8b) follows from the fact that the inner minimization problem in (2.8a) is a linear problem for fixed y and from exploiting strong duality. Taking the previous considerations into account, the robust counterpart (2.7) can be solved as the (mixed-integer) problem

$$\max_{y,z} \quad f^\top y + h^\top z \quad \text{s.t.} \quad y \in Y,\ H^\top z = P^\top y,\ z \in \mathbb{R}^s_{\geq 0}.$$

2.2.2 Γ-Robustness

A notable point of criticism regarding the strictly robust approach outlined in Section 2.2.1 is the potential over-conservatism of solutions. Since it seems unlikely that all uncertain parameters realize in a worst-case sense, there has been a growing interest in more flexible approaches to deal with uncertainties. One such approach is the so-called Γ-robust approach (Bertsimas and Sim 2003, 2004; Sim 2004), which we present in the following. To this end, we assume that the uncertainty set takes the form

$$\mathcal{U} = \mathcal{U}_1 \times \cdots \times \mathcal{U}_{n_y} \quad \text{with} \quad \mathcal{U}_i := [f_i - \Delta f_i, f_i] \text{ for all } i \in [n_y] := \{1, \dots, n_y\}\,.$$

Here, f_i is the nominal value of the ith objective function coefficient and $\Delta f_i \geq 0$ is its maximum deviation from the nominal value. Pursuing a Γ-robust approach, the decision-maker only hedges against a subset of at most Γ uncertain parameters that adversely affect the solution to the problem at hand. Here, the parameter $\Gamma \in \{0, \dots, n_y\}$ is used to control the decision-maker's level of conservatism regarding the solution. The so-called Γ-*robust counterpart* of Problem (2.6) is then given by

$$v := \max_{y \in Y} \left\{ f^\top y - \max_{\{S \subseteq [n_y] \colon |S| \leq \Gamma\}} \sum_{i \in S} \Delta f_i \, |y_i| \right\}. \tag{2.9}$$

In particular, Problem (2.9) captures the nominal and the strictly robust formulation of the problem as special cases. The remainder of this section is now devoted to obtaining tractable reformulations of the Γ-robust counterpart (2.9).

Theorem 2.1 (cf. Theorem 1 in Bertsimas and Sim (2003))**.** *The Γ-robust counterpart (2.9) of Problem (2.6) has an equivalent mixed-integer formulation given by*

$$\begin{aligned}
\max_{y,z,\theta} \quad & \sum_{i=1}^{n_y} f_i y_i - \Gamma\theta - \sum_{i=1}^{n_y} z_i && \text{(2.10a)}\\
\text{s.t.} \quad & \theta + z_i \geq \Delta f_i u_i, \quad i \in [n_y], && \text{(2.10b)}\\
& -u_i \leq y_i \leq u_i, \quad i \in [n_y], && \text{(2.10c)}\\
& y \in Y, \;\; z, u \in \mathbb{R}^{n_y}_{\geq 0}, \;\; \theta \in \mathbb{R}_{\geq 0}. && \text{(2.10d)}
\end{aligned}$$

Constraints (2.10c) are used to linearize the absolute value in Problem (2.9). Hence, Constraints (2.10c) as well as the variables u can be eliminated in Problem (2.10) if the variables y are restricted to be non-negative. We can further reformulate Problem (2.10) under the assumption of only binary variables y and by imposing the following.

Assumption 2.1. *The indices are ordered such that the deviations are given in non-increasing order, i.e.,* $\Delta f_i \geq \Delta f_{i+1}$ *holds for all* $i \in [n_y]$ *with* $\Delta f_{n_y+1} = 0$.

Assumption 2.1 is w.l.o.g. but necessary for the following theorem.

Theorem 2.2 (cf. Theorem 3 in Bertsimas and Sim (2003))**.** *Suppose that* $Y \subseteq \{0,1\}^{n_y}$ *holds. Then, under Assumption 2.1, the Γ-robust counterpart (2.9) of Problem (2.6) can be solved by solving* $n_y + 1$ *problems of the nominal type, i.e.,*

$$v = \max_{\ell \in [n_y+1]} \{v_\ell\}\,,$$

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

$$v_\ell := -\Gamma \Delta f_\ell + \max_{y \in Y} \left\{ \tilde{f}(\ell)^\top y \right\} \quad \textit{with} \quad \tilde{f}(\ell)_i = \begin{cases} f_i - (\Delta f_i - \Delta f_\ell), & 1 \le i \le \ell, \\ f_i, & \ell + 1 \le i \le n_y. \end{cases}$$

In Miranda et al. (2013), the authors present an improvement of the result in Theorem 2.2 by reducing the number of deterministic problems to be solved to $n_y - \Gamma + 2$. Further reductions have been established by Lee and Kwon (2014) so that it suffices to consider $\lceil (n_y - \Gamma)/2 \rceil + 1$ problems of the nominal type.

Theorem 2.3 (cf. Theorem 1 in Lee and Kwon (2014))**.** *Suppose that* $Y \subseteq \{0,1\}^{n_y}$ *holds. Then, under Assumption 2.1, the* Γ*-robust counterpart* (2.9) *of Problem* (2.6) *can be solved by solving*

$$v = \max_{\ell \in \mathcal{L}} \{v_\ell\},$$

where $\mathcal{L} = \{\Gamma + 1, \Gamma + 3, \Gamma + 5, \dots, \Gamma + \gamma, n_y + 1\}$, γ *is the largest odd integer such that* $\Gamma + \gamma < n_y + 1$, *and*

$$v_\ell := -\Gamma \Delta f_\ell + \max_{y \in Y} \left\{ \tilde{f}(\ell)^\top y \right\} \quad \textit{with} \quad \tilde{f}(\ell)_i = \begin{cases} f_i - (\Delta f_i - \Delta f_\ell), & 1 \le i \le \ell, \\ f_i, & \ell + 1 \le i \le n_y, \end{cases}$$

for all $\ell \in \mathcal{L}$.

2.3 Robust Bilevel Optimization

While typically being investigated in distinct communities, bilevel and robust optimization share several commonalities that strongly suggest a certain connection between these two problem classes. For instance, pessimistic bilevel optimization is rather naturally connected to the field of robust optimization; see, e.g., Wiesemann et al. (2013). Recently, further connections have been established in Goerigk et al. (2025) for certain instances of bilevel and robust optimization problems. In particular, the work in Goerigk et al. (2025) may be seen as a first systematic step towards bridging the gap between bilevel and robust optimization. In this dissertation and, in particular, in this section, however, we are interested in optimization problems at the *intersection* of bilevel and robust optimization. We refer to these problems as *robust bilevel problems*. The field of robust bilevel optimization is still in its infancy. Nevertheless, due to the growing interest in considering uncertainties in bilevel problems, it is a promising direction of current and future research. In this section, we outline the basic concepts discussed in [YB1] for bilevel problems with a robust treatment of uncertainties in the lower-level data. In Section 2.3.1, we present different variants of robust bilevel problems, which we illustrate in Section 2.3.2 using an academic example.

2.3.1 Variants of Robust Bilevel Problems

In [YB1], we focus on the bilevel setting without coupling constraints, i.e., there are no upper-level constraints that explicitly depend on the follower's variables y. This means that we consider deterministic bilevel problems of the form

$$\text{"}\min_{x \in X}\text{"} \quad F(x,y) \quad \text{s.t.} \quad G(x) \ge 0, \ y \in S(x) := \operatorname*{arg\,min}_{y' \in Y} \left\{ f(x,y') \colon g(x,y') \ge 0 \right\}. \tag{2.11}$$

In [YB1], we address bilevel problems of the form given in (2.11) with a (strictly) robust treatment of lower-level objective uncertainty. To be more precise, we assume that the lower-level objective function f of Problem (2.11) is affected by uncertainties u that are known to vary within a given and compact uncertainty set $\mathcal{U}$. In the single-level setting outlined in Section 2.2, we have considered a single robust decision-maker who makes a here-and-now decision before the uncertainty realizes. In bilevel optimization, there are two different possibilities regarding the timing of when the uncertainty is revealed.

In the first variant, the uncertainty realizes after the follower makes his decision. Hence, both the leader and the follower have to decide before the uncertainty is revealed, i.e., we consider the timing

$$\text{leader } x \quad \curvearrowright \quad \text{follower } y = y(x) \quad \curvearrowright \quad \text{uncertainty } u. \tag{2.12}$$

This means that the leader anticipates an optimal response of the follower who hedges against objective uncertainty in a robust way. We refer to this type of problem as a bilevel problem with a *here-and-now follower.* Formally, the problem is obtained by setting

$$S(x) := \operatorname*{arg\,min}_{y' \in Y} \left\{ \max_{u \in \mathcal{U}} \; \tilde{f}(x, y', u) \colon g(x, y') \geq 0 \right\} \tag{2.13}$$

in Problem (2.11). For a given leader's decision x, Problem (2.13) is a classic, i.e., single-level, robust problem. Hence, we can exploit any of the concepts known for single-level robust optimization to deal with the robust counterpart of the lower-level problem. However, the response of the robust follower may not be unique. To address such ambiguities, one also needs to distinguish between the optimistic and the pessimistic approach in the robust bilevel setting. For optimistic robust bilevel problems, the same reformulation techniques as for deterministic bilevel problems can be applied. If, e.g., Problem (2.13) can be reformulated as a problem for which the KKT conditions are necessary and sufficient, we can replace the robustified lower-level problem by its KKT conditions to obtain a single-level reformulation of the overall robust bilevel problem.

In the second timing variant, the uncertainty is revealed before the follower makes his decision. In this setting, the leader first decides in a here-and-now fashion without full knowledge of the lower-level problem. Then, the uncertainty realizes and, finally, the follower makes a "wait-and-see" decision, taking both the leader's decision and the uncertainty realization into account. Hence, one considers the timing

$$\text{leader } x \quad \curvearrowright \quad \text{uncertainty } u \quad \curvearrowright \quad \text{follower } y = y(x, u). \tag{2.14}$$

Bilevel problems that are subject to the timing in (2.14) are referred to as bilevel problems with a *wait-and-see follower.* Formally, these problems can be modeled as

$$\text{“}\min_{x \in X} \; \max_{u \in \mathcal{U}}\text{”} \quad F(x, y) \quad \text{s.t.} \quad G(x) \geq 0, \; y \in S(x, u), \tag{2.15}$$

where $S(x, u)$ is the set of optimal solutions to the (x, u)-parameterized lower-level problem

$$\min_{y \in Y} \quad \tilde{f}(x, y, u) \quad \text{s.t.} \quad g(x, y) \geq 0.$$

In Problem (2.15), the leader hedges against the worst possible response of the follower w.r.t. the uncertainties in the lower-level data. However, there may be an ambiguity in the case that the set $S(x, u)$ is not a singleton, which we express using quotation marks in (2.15). Hence, we also need to distinguish between the optimistic and the pessimistic approach in the robust bilevel setting with a wait-and-see follower.

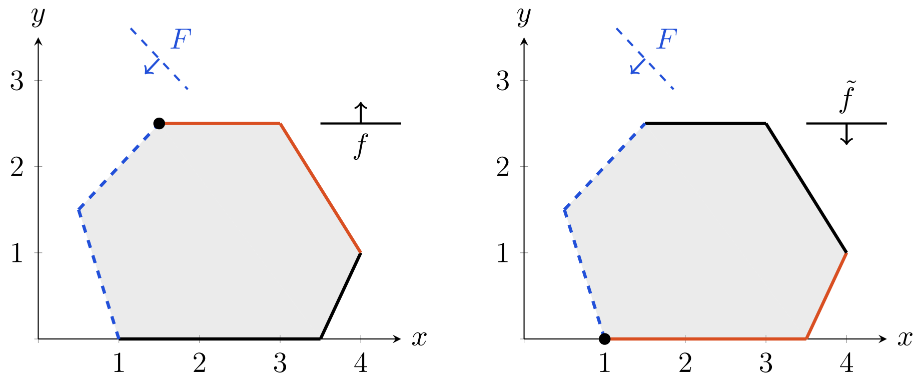

Figure 2.1: Both figures show the upper-level constraints (dashed blue lines), the lower-level constraints (solid black and orange lines), the shared constraint set (gray area), and the bilevel feasible set (solid orange lines) of the bilevel problem (2.16). The deterministic variant of the problem is depicted on the left and the variant with a here-and-now follower is given on the right; cf. Figure 1 in [YB1].

2.3.2 An Academic Example

We now illustrate how the timing (2.12) vs. (2.14) may affect solutions to robust bilevel problems. To this end, we consider the linear bilevel problem taken from [YB2] that is given by

$$\begin{aligned}
\text{“}\min_{x\in\mathbb{R}}\text{”} \quad & F(x,y) = x + y && \text{(2.16a)}\\
\text{s.t.} \quad & x - y \geq -1, && \text{(2.16b)}\\
& 3x + y \geq 3, && \text{(2.16c)}\\
& y \in S(x), && \text{(2.16d)}
\end{aligned}$$

where $S(x)$ denotes the set of optimal solutions to the x-parameterized lower level

$$\begin{aligned}
\min_{y\in\mathbb{R}} \quad & f(x,y) = -0.1y && \text{(2.17a)}\\
\text{s.t.} \quad & -2x + y \geq -7, && \text{(2.17b)}\\
& -3x - 2y \geq -14, && \text{(2.17c)}\\
& 0 \leq y \leq 2.5. && \text{(2.17d)}
\end{aligned}$$

The problem is shown in Figure 2.1 (left). Note that the lower-level problem (2.17) has a unique solution for every feasible decision x of the leader. Hence, there is no need to distinguish between the optimistic and the pessimistic case. The optimal solution to the bilevel problem (2.16), which is illustrated by the thick dot in Figure 2.1, is given by $(x^*, y^*) = (1.5, 2.5)$. We now introduce uncertainties in the lower-level objective. To this end, we consider

$$\tilde{f}(x,y,u) = (-0.1+u)y \quad \text{with} \quad u \in \mathcal{U} := \{u \in \mathbb{R} \colon -0.5 \leq u \leq 0.5\}.$$

Here-and-Now Follower We consider the timing in (2.12), i.e., the robust lower-level problem is given by

$$\min_{y\in\mathbb{R}} \max_{u\in\mathcal{U}} \quad \tilde{f}(x,u,y) = (-0.1+u)y \quad \text{s.t.} \quad (2.17b)–(2.17d).$$

Using classic techniques from robust optimization, we obtain a modified gradient of the lower-level objective function. The latter is shown in Figure 2.1 (right). Since the follower's response is unique for every feasible x, there is (again) no need to distinguish between the optimistic and the pessimistic setting. The optimal solution $(x^*, y^*) = (1, 0)$ to the overall robust bilevel problem is represented by the thick dot.

Wait-and-See Follower We now consider the timing in (2.14), i.e., the overall robust bilevel problem reads

$$\text{“}\min_{x\in\mathbb{R}} \max_{u\in\mathcal{U}}\text{”} \quad F(x,y) \quad \text{s.t.} \quad (2.16b)–(2.16c),\ y \in S(x,u),$$

where $S(x,u)$ is the set of optimal solutions to the (x,u)-parameterized lower-level problem

$$\min_{y\in\mathbb{R}} \quad \tilde{f}(x,u,y) = (-0.1+u)y \quad \text{s.t.} \quad (2.17b)–(2.17d).$$

To solve this problem, we need to distinguish the following three cases.

(a) $-0.5 \le u < 0.1$: This case corresponds to the setting depicted in Figure 2.1 (left). The optimal follower's response is given by

$$y(x,u) = \begin{cases} 2.5, & x \le 3, \\ -1.5x + 7, & 3 \le x \le 4. \end{cases} \tag{2.18}$$

(b) $u = 0.1$: In this case, we have $\tilde{f}(x,y,u) = 0$. Hence, any feasible decision of the follower, i.e., any $y \in \mathbb{R}$ satisfying (2.17b)–(2.17d), is optimal for the lower-level problem. This means that we need to distinguish between an optimistic and a pessimistic follower. In the optimistic setting, the follower reacts with

$$y(x,u) = \begin{cases} 0, & x \le 3.5, \\ 2x - 7, & 3.5 \le x \le 4. \end{cases} \tag{2.19}$$

The latter corresponds to the setting depicted in Figure 2.1 (right). A pessimistic follower responds with (2.18).

(c) $0.1 < u \le 0.5$: The optimal follower's response is given by (2.19).

An optimal solution to the bilevel problem (2.16) with a wait-and-see follower is now obtained by considering the worst among the previous three cases in dependence of the leader's decision x. Hence, we need to solve

$$\min_{x} \quad \hat{F}(x) \quad \text{s.t.} \quad 1.5 \le x \le 4 \tag{2.20}$$

with the piecewise-linear function

$$\hat{F}(x) = \begin{cases} x + 2.5, & 1.5 \leq x \leq 3, \\ -0.5x + 7, & 3 \leq x \leq 4. \end{cases}$$

The optimal solution to Problem (2.20) is given by $x^* = 1.5$. After observing both the leader's decision and the realization of the uncertainty, the follower then reacts with

$$y^*_{\mathrm{o}}(x^*, u) = \left\{ \begin{matrix} 2.5, \text{ if } u \in [-0.5, 0.1) \\ 0, \text{ if } u \in [0.1, 0.5] \end{matrix} \right\} \quad \text{and} \quad y^*_{\mathrm{p}}(x^*, u) = \left\{ \begin{matrix} 2.5, \text{ if } u \in [-0.5, 0.1] \\ 0, \text{ if } u \in (0.1, 0.5] \end{matrix} \right\}$$

in the optimistic and the pessimistic setting, respectively.

Comparison of Solutions We summarize the solutions obtained for the different variants of the bilevel problem (2.16) in Table 2.1. We point out that there is no need to consider the actual uncertainty realization when reporting solutions for the nominal and the robust here-and-now setting. In the wait-and-see setting, however, the follower's response depends on the actual uncertainty realization. Thus, we need to take the realization of the uncertainty into account when reporting robust solutions with a wait-and-see follower. When comparing the obtained solutions, there are three aspects that we find particularly worth mentioning. First, we note that the optimal leader's decision $x^* = 1$ for the setting with a here-and-now follower is neither feasible for the nominal problem nor for the robust variant with a wait-and-see follower. Second, we highlight that the nominal solution and the robust solution with a wait-and-see follower coincide if $u \in [-0.5, 0.1)$ realizes. Third and finally, we emphasize that, for $u \geq 0.1$, the optimal response of a wait-and-see follower may change significantly compared to the nominal solution. To sum up, the incorporation of uncertainties along with the considered timing ((2.12) vs. (2.14)) can make a huge difference when solving bilevel problems.

Table 2.1: Nominal and robust solutions with a here-and-now and a wait-and-see follower for the bilevel problem (2.16).

			follower	
	leader	u	optimistic	pessimistic
nominal	1.5		2.5	2.5
here-and-now	1.0		0.0	0.0
wait-and-see	1.5	< 0.1	2.5	2.5
		$= 0.1$	0.0	2.5
		> 0.1	0.0	0.0

3. Exact Methods for Mixed-Integer, Linear, and Γ-Robust Min-Max Problems

In this chapter, we focus on mixed-integer linear bilevel problems of the form

$$\begin{aligned} \min_{x} \quad & c^\top x + f^\top y \\ \text{s.t.} \quad & x \in X, \\ & y \in \arg\max_{y'} \left\{ f^\top y' \colon y' \in Y(x) \right\}, \end{aligned} \tag{3.1}$$

where $Y(x) \subseteq \mathbb{Z}^{n_y}_{\geq 0}$ and $X := \{x \in \mathbb{R}^{n_\mathrm{C}} \times \mathbb{Z}^{n_\mathrm{D}} \colon Ax \geq a\}$ with $n_x = n_\mathrm{C} + n_\mathrm{D}$, $c \in \mathbb{R}^{n_x}$, $f \in \mathbb{R}^{n_y}$, $A \in \mathbb{R}^{l \times n_x}$, and $a \in \mathbb{R}^l$. Problem (3.1) is a min-max problem. Hence, the follower's response always yields the worst-possible outcome for the leader, i.e., there is no need to distinguish between the optimistic and the pessimistic approach to bilevel optimization. Problems of the form given in (3.1) are of interest in various real-world applications, such as in critical infrastructure defense, network disruption, or marketing, as they particularly cover the important class of interdiction problems; see, e.g., Brown et al. (2006), Cormican et al. (1998), DeNegre (2011), Fischetti et al. (2019), Furini et al. (2021), Israeli and Wood (2002), and Wood (2011). For the remainder of this chapter, we make the following assumptions that are commonly used in mixed-integer bilevel optimization; see, e.g., DeNegre and Ralphs (2009), Fischetti et al. (2017), and Vicente et al. (1996).

Assumption 3.1.

(i) The shared constraint set $\{(x, y) \colon x \in X,\ y \in Y(x)\}$ *is non-empty and compact.*

(ii) For all $x \in X$*, the lower-level feasible set* $Y(x)$ *is non-empty.*

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

Assumption 3.1 ensures that Problem (3.1) has an optimal solution. Moreover, we mention that the reasons to impose Part (iii) of Assumption 3.1 are, in fact, two-fold. On the one hand, continuous linking variables may lead to situations in which the infimum of Problem (3.1) is not attained. An exemplary bilevel problem that illustrates this behavior can, e.g., be found in Example 1.1 in Köppe et al. (2010). On the other hand, Part (iii) of Assumption 3.1 implies a finite number of feasible choices for the linking variables so that the overall number of possible parameterized lower-level problems is finite as well. This allows for the application of enumeration-based algorithms, which is what we do in the remainder of this chapter. In Section 3.1, we review a generic branch-and-cut approach to solve Problem (3.1). Afterward, in Section 3.2, we consider a variant of Problem (3.1) that is subject to lower-level objective uncertainty. We exploit the notion of Γ-robustness to hedge against this kind of uncertainty and, in particular, we discuss how the generic branch-and-cut method for deterministic min-max problems needs to be adapted in this setting. Section 3.3 is devoted to tailored techniques for monotone interdiction problems with Γ-robust followers. Numerical results for solving this class of bilevel problems are discussed in Section 3.4. Finally, in Section 3.5, we elaborate on exact solution approaches for related classes of bilevel problems.

3.1 A Generic Branch-and-Cut Method for Deterministic Min-Max Problems

Problem (3.1) is a mixed-integer linear bilevel problem with a discrete lower level. Due to the overall hardness of the problem, which is Σ_2^p-hard in general, developing solution methods is known to be a challenging task. In particular, standard reformulation techniques such as, e.g., replacing the lower-level problem by its KKT conditions cannot be applied in this setting; see Chapter 2.1 for discussions on reformulation techniques in bilevel optimization. Instead, many solution approaches for mixed-integer bilevel problems solve the value-function reformulation of the problem by exploiting techniques from single-level mixed-integer optimization. In the seminal work by Moore and Bard (1990), the first branch-and-bound method for solving mixed-integer linear bilevel problems is discussed. The idea has been extended by DeNegre and Ralphs (2009), who provide a branch-and-cut approach for pure integer linear bilevel problems. In particular, the latter may be seen as a turning point regarding computational mixed-integer bilevel optimization. We also refer to DeNegre (2011) for further details. Over the past decade, many other influential works on solution approaches for mixed-integer bilevel problems have followed (Fischetti et al. 2017, 2018a; Tahernejad et al. 2020; Xu and Wang 2014) so that approaches commonly used today include branch-and-cut or (problem-tailored) decomposition methods. For a detailed discussion on further mixed-integer optimization techniques in bilevel optimization, we refer to the survey in Kleinert et al. (2021b).

In this section, we review a generic branch-and-cut framework for mixed-integer min-max problems of the form given in (3.1) that follows the ideas of DeNegre and Ralphs (2009) and Wood (2011). The method is based on solving the value-function reformulation of Problem (3.1), which is given by

$$\min_{x \in X} \quad c^\top x + \Phi(x) \tag{3.2}$$

with the lower-level optimal-value function

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

In Problem (3.1), we do not consider coupling constraints. This means that there are no upper-level constraints that explicitly depend on the follower's variables y. The latter is essential to project the variables y out of the problem as it is done in Problem (3.2) using the lower-level optimal-value function. The specific structure of Problem (3.2) now allows for the application of decomposition approaches similar to (generalized) Benders decomposition (Benders 1962; Geoffrion 1972), which is what we do in the following. In this context, we note that the optimal-value function $\Phi(x)$ is usually not known in closed form and that its evaluation corresponds to the solution of an x-parameterized discrete optimization problem. Following the notion of Geoffrion, we thus call the lower-level optimal-value function "complicating". We introduce an auxiliary variable $\eta \in \mathbb{R}$ to move the complicating part of the objective function into the constraints of the problem. An equivalent reformulation of Problem (3.2) is thus given by

$$\begin{aligned} \min_{x,\eta} \quad & c^\top x + \eta \\ \text{s.t.} \quad & \eta \geq \Phi(x), \\ & (x, \eta) \in X \times \mathbb{R}. \end{aligned} \tag{3.3}$$

Here, "equivalence" is to be understood in the following sense. Whenever x^* is an optimal solution to Problem (3.2), the pair (x^*, η^*) with $\eta^* := \Phi(x^*)$ is an optimal solution to Problem (3.3). Conversely, whenever (x^*, η^*) solves Problem (3.3), x^* is an optimal solution to Problem (3.2). To solve Problem (3.3), we start by considering the problem in which the integrality constraints on the variables x as well as $\eta \geq \Phi(x)$ are omitted, i.e., we solve the linear problem

$$\begin{aligned} \min_{x,\eta} \quad & c^\top x + \eta \\ \text{s.t.} \quad & (x, \eta) \in \Omega_0 := \left\{ (x', \eta') \in \bar{X} \times \mathbb{R} \colon \eta' \geq \eta^- \right\}. \end{aligned} \tag{3.4}$$

Here, the set $\bar{X}$ is a continuous relaxation of X, i.e., the integer points contained in $\bar{X}$ coincide with X. Moreover, $\eta^- \in \mathbb{R}$ is a lower bound on the optimal-value function $\Phi(x)$ for all $x \in X$. After considering Problem (3.4), we iteratively add valid inequalities or branch to cut off integer-infeasible points and we also add valid inequalities to cut off bilevel-infeasible points. At node j of the branch-and-cut search tree, we consider the problem

$$\begin{aligned} \min_{x,\eta} \quad & c^\top x + \eta \\ \text{s.t.} \quad & (x, \eta) \in \Omega_j \subseteq \bar{X} \times \mathbb{R}, \end{aligned} \tag{3.5}$$

where the set Ω_j contains all valid inequalities that have been added previously to cut off integer-infeasible and bilevel-infeasible points as well as all branching decisions. If either Problem (3.5) is infeasible or, if the objective function value corresponding to an optimal solution (x^j, η^j) exceeds the current upper bound U, we can fathom node j. Otherwise, we do the following. First, we check if the upper-level variables x^j satisfy the integrality constraints, i.e., we check if $x^j \in X$ holds. To separate a fractional point, we can either branch or exploit standard cutting planes from mixed-integer optimization, e.g., as elaborated in Cornuéjols (2008). Otherwise, we proceed by checking for bilevel-feasibility, i.e., we check if $\eta^j \geq \Phi(x^j)$ is satisfied. To this end, we solve the x^j-parameterized lower-level problem. If (x^j, η^j) is bilevel-infeasible, i.e., $\eta^j < \Phi(x^j)$ holds, we generate a cut that separates (x^j, η^j) and add it to the description of Ω_j. For an overview of various cutting planes that can be used for general classes of mixed-integer linear bilevel problems, we refer to Tahernejad and Ralphs (2020). Nevertheless, we emphasize that stronger formulations may be obtained for certain problems; see, e.g., Fischetti et al. (2019) and Furini et al. (2021). All steps required to process node j of the branch-and-cut search tree are summarized in Algorithm 3.1.

Algorithm 3.1 Processing node j of the branch-and-cut search tree

```
1:  Solve Problem (3.5).
2:  if Problem (3.5) is infeasible then
3:      Fathom the current node.
4:  Let $(x^j, \eta^j)$ denote the optimal solution to Problem (3.5).
5:  if $c^\top x^j + \eta^j \geq U$ then
6:      Fathom the current node.
7:  if $x^j \notin X$ then
8:      Generate cuts valid for $\Omega_j \cap (X \times \mathbb{R})$, augment $\Omega_j$, and go to Step 1 or branch.
9:  Determine $\Phi(x^j)$ and set $U \leftarrow \min\{U, c^\top x^j + \Phi(x^j)\}$.
10: if $\eta^j < \Phi(x^j)$ then
11:     Generate a valid cut that excludes $(x^j, \eta^j)$ from $\Omega_j$, augment $\Omega_j$, and go to Step 1.
```

3.2 Generic Branch-and-Cut Methods for Γ-Robust Min-Max Problems

In [YB3], we study mixed-integer linear min-max problems of the form given in (3.1), which are, however, affected by some kind of data uncertainty. The focus of [YB3] lies on lower-level objective uncertainty, which is also what we consider in the remainder of this chapter. To this end, we assume that the follower's objective function coefficients are uncertain but known to take values in the uncertainty set

$$\mathcal{U} = \mathcal{U}_1 \times \cdots \times \mathcal{U}_{n_y} \quad \text{with} \quad \mathcal{U}_i := [f_i - \Delta f_i, f_i] \quad \text{for all} \quad i \in [n_y] := \{1, \ldots, n_y\}.$$

Here, f_i denotes the nominal value of the ith lower-level objective function coefficient and $\Delta f_i \in \mathbb{R}_{\geq 0}$ denotes the maximum deviation from the nominal value. We pursue a Γ-robust approach (Bertsimas and Sim 2003; Sim 2004) so that the follower only hedges against a subset of deviations in the uncertain parameters that adversely affect the solution; see Chapter 2.2 for further details on Γ-robustness. For given $x \in X$ and $\Gamma \in \{0, \ldots, n_y\}$, the Γ-robust counterpart of the lower-level problem reads

$$\Phi_{\text{rob}}(x) = \max_y \left\{ f^\top y - \max_{\{S \subseteq [n_y] \colon |S| \leq \Gamma\}} \sum_{i \in S} \Delta f_i y_i \colon y \in Y(x) \right\}. \tag{3.6}$$

A robustification of the overall bilevel problem (3.1) with uncertainties in the lower-level objective function is now obtained by replacing $\Phi(x)$ with $\Phi_{\text{rob}}(x)$ in Problem (3.3). Hence, we consider the problem

$$\begin{aligned} \min_{x,\eta} \quad & c^\top x + \eta \\ \text{s.t.} \quad & \eta \geq \Phi_{\text{rob}}(x), \\ & (x, \eta) \in X \times \mathbb{R}. \end{aligned} \tag{3.7}$$

In [YB3], we present two generic branch-and-cut approaches to solve Problem (3.7). To the best of our knowledge, there are currently no other methods in the literature that can tackle Problem (3.7) directly. Both approaches discussed in [YB3] are based on the algorithmic framework outlined in Section 3.1 and work as follows. We start by solving the problem in which we relax the integrality constraints on the leader's variables x and we omit $\eta \geq \Phi_{\text{rob}}(x)$. Hence, also in the Γ-robust setting, we start by solving Problem (3.4). Afterward, we iteratively add valid inequalities or branch to separate integer- or bilevel-infeasible points, similar to how it is done in Algorithm 3.1. We emphasize that, in Algorithm 3.1, no structural assumptions are made on the lower-level problem for pruning nodes due to infeasibility (Line 3) or bounding (Line 6). Additionally, by exploiting techniques from single-level mixed-integer optimization, the separation of fractional points (Line 8) can be done independently of which specific problem is being considered at the lower level. Hence, only the steps related to ensure bilevel feasibility (Lines 9–11 of Algorithm 3.1) need to be adapted to account for the setting with a Γ-robust follower. In [YB3], we present two such adaptations—one based on an extended formulation and one based on a multi-follower formulation to solve the Γ-robust counterpart (3.6) of the lower-level problem. Both approaches pursue similar ideas in principle, however, they differ in the structural assumptions made regarding the lower-level feasible set and, thus, in the treatment of bilevel-infeasible points. We present both approaches in the remainder of this section.

3.2.1 An Extended Formulation

For fixed $x \in X$, the Γ-robust counterpart (3.6) of the lower-level problem is a classic, i.e., single-level, Γ-robust problem. Hence, we can apply Theorem 2.1 so that the Γ-robust counterpart (3.6) of the lower-level problem can be solved as a mixed-integer problem.

Corollary 3.1 (cf. Lemma 1 in [YB3]). *Let $x \in X$ be given arbitrarily. Then, the Γ-robust counterpart (3.6) of the lower-level problem can be solved as*

$$\begin{aligned} \max_{y,z,\theta} \quad & \sum_{i=1}^{n_y} f_i y_i - \Gamma\theta - \sum_{i=1}^{n_y} z_i \\ s.t. \quad & z_i + \theta \geq \Delta f_i y_i, \quad i \in [n_y], \\ & y \in Y(x) \subseteq \mathbb{Z}_{\geq 0}^{n_y}, \ z \in \mathbb{R}_{\geq 0}^{n_y}, \ \theta \in \mathbb{R}_{\geq 0}. \end{aligned} \tag{3.8}$$

We refer to Problem (3.8) as the *extended formulation* of the lower-level problem. To process node j of the branch-and-cut search tree using the extended formulation, we follow the steps outlined in Algorithm 3.1, wherein we replace Lines 9–11 with the separation procedure formally stated in Algorithm 3.2. In Line 3 of Algorithm 3.2, one may again use generic cuts (Tahernejad and Ralphs 2020) to exclude bilevel-infeasible points. Nevertheless, in [YB3], we present problem-tailored cuts for Γ-robust monotone interdiction problems that can be used to obtain stronger formulations. The derivation of these cuts is discussed in Section 3.3 of this dissertation.

Algorithm 3.2 Separating bilevel-infeasible points using the extended formulation

```
1: Solve Problem (3.8) to obtain $\Phi_{\text{rob}}(x^j)$ and set $U \leftarrow \min\{U, c^\top x^j + \Phi_{\text{rob}}(x^j)\}$.
2: if $\eta^j < \Phi_{\text{rob}}(x^j)$ then
3:     Generate a valid cut that excludes $(x^j, \eta^j)$ from $\Omega_j$, augment $\Omega_j$, and go to Step 1 of Algorithm 3.1.
```

3.2.2 A Multi-Follower Formulation

An alternative reformulation of the Γ-robust counterpart (3.6) can be obtained under the following additional assumption.

Assumption 3.2. *All lower-level variables* y *are binary, i.e.,* $Y(x) \subseteq \{0,1\}^{n_y}$ *holds for all* $x \in X$.

Under Assumption 3.2, we can exploit Theorem 2.2 to reformulate Problem (3.6).

Corollary 3.2 (cf. Lemma 3 in [YB3])**.** *Let* $x \in X$ *be given arbitrarily. Under Assumptions 2.1 and 3.2, solving the* Γ*-robust counterpart* (3.6) *of the lower-level problem is equivalent to solving* $n_y + 1$ *problems of the nominal type, i.e.,*

$$\Phi_{rob}(x) = \max_{\ell \in [n_y+1]} \{\Phi_\ell(x)\} \tag{3.9}$$

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

$$\Phi_\ell(x) = -\Gamma \Delta f_\ell + \max_{y \in Y(x)} \left\{ \tilde{f}(\ell)^\top y \right\} \tag{3.10}$$

with

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

In (3.9), the lower-level optimal-value function $\Phi_{\text{rob}}(x)$ is defined as the maximum of $n_y + 1$ value functions. Hence, Problem (3.7) can be interpreted as a single-leader multi-follower problem with $n_y + 1$ followers. We thus refer to (3.9) as the *multi-follower formulation.* Moreover, for a given $\ell \in [n_y + 1]$, we refer to (3.10) as the ℓth *lower-level (or follower) sub-problem.* As elaborated in Chapter 2.2, the number of nominal sub-problems to be considered in (3.9) can be reduced. In [YB3] and in what follows, we hold on to the result by Miranda et al. (2013) so that it suffices to consider

$$\Phi_{\text{rob}}(x) = \max_{\ell \in \mathcal{L}} \{\Phi_\ell(x)\} \quad \text{with} \quad \mathcal{L} := \{\Gamma, \dots, n_y + 1\}.$$

To process node j of the branch-and-cut search tree using the multi-follower formulation, we follow the steps outlined in Algorithm 3.1, wherein we replace Lines 9–11 with the separation procedure formally stated in Algorithm 3.3.

Algorithm 3.3 Separating bilevel-infeasible points using the multi-follower formulation

```
1: for all $\ell \in \mathcal{L}$ do
2:     Solve the $\ell$th lower-level sub-problem (3.10) to obtain $\Phi_\ell(x^j)$.
3:     if $\eta^j < \Phi_\ell(x^j)$ then
4:         Generate a valid cut that excludes $(x^j, \eta^j)$ from $\Omega_j$ and augment $\Omega_j$.
5: Set $\Phi_{\text{rob}}(x^j) \leftarrow \max_{\ell \in \mathcal{L}}\{\Phi_\ell(x^j)\}$ and $U \leftarrow \min\{U, c^\top x^j + \Phi_{\text{rob}}(x^j)\}$.
6: If at least one cut was added in Step 4, then go to Step 1 of Algorithm 3.1.
```

In contrast to the approach using the extended formulation, in which a single cut is added at each node of the branch-and-cut search tree in case of bilevel-infeasibility, a cut for each follower sub-problem $\ell \in \mathcal{L}$ that satisfies $\eta^j < \Phi_\ell(x^j)$ is added in Line 4 of Algorithm 3.3. This means that up to $|\mathcal{L}|$ cuts can be added at each node. However, it is also valid to consider, e.g., adding only the most violated cut for the given leader's decision x^j. In [YB3], we discuss and computationally assess various cut separation strategies. We further emphasize that the sub-problems solved in Line 2 of Algorithm 3.3 are independent. This means that, for each $\ell \in \mathcal{L}$, the objective function and the constraints of the ℓth lower-level sub-problem only include the upper-level decision x and the lower-level variables associated with the specific sub-problem. Hence, if the necessary capacities are available, the lower-level sub-problems can be solved in parallel.

In Table 3.1, we summarize the main differences between the two generic branch-and-cut approaches presented in this section. Despite these differences, both approaches have in common that they terminate after a finite number of iterations with an optimal solution to the Γ-robust min-max problem (3.7). Finite termination is due to the finiteness of the number of feasible choices for the linking variables (Part (iii) of Assumption 3.1), the finiteness of branch-and-cut methods to solve the lower-level problems (3.8) and (3.10), and the finiteness of the number of cuts possibly added to the problem formulation. More formally, we have the following theorem. For a proof of this theorem, we refer to [YB3].

Theorem 3.3 (cf. Theorem 1 in [YB3])**.** *If we embed Algorithm 3.1 with Lines 9–11 replaced by either Algorithm 3.2 or Algorithm 3.3 into a usual branch-and-bound framework, we obtain a correct method that terminates with an optimal solution (x^*, η^*) to Problem (3.7) after a finite number of nodes and after adding an overall finite number of cuts.*

Table 3.1: Comparison of the branch-and-cut approach using an extended formulation (Ext) and a multi-follower formulation (MF).

	Ext	MF
Follower variables	non-negative integers	binaries
Γ-robust problem type	mixed-integer problem	$\lvert\mathcal{L}\rvert$ binary sub-problems
Cuts per node	1	up to $\lvert\mathcal{L}\rvert$
Parallelization	✗	✓

3.3 Tailored Techniques for Monotone Interdiction Problems

We now focus on interdiction problems with a follower problem that satisfies a downward monotonicity property. In [YB3], we present tailored cuts and enhancement techniques for the Γ-robust variant of these problems, which we summarize in the following. The contributions of [YB3] extend the results that have been proposed in Fischetti et al. (2019) for the deterministic variant of the problem. Throughout the remainder of this section, we make the following assumptions.

Assumption 3.3.

(i) *The leader and the follower have the same number of variables, i.e.,* $n_x = n_y = n$.

(ii) *All variables of the leader are binary linking variables, i.e.,* $X \subseteq \{0,1\}^n$.

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

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

with $Y := \{y \in \mathbb{Z}^n_{\geq 0} \colon By \leq b\}$, $B \in \mathbb{R}^{m \times n}_{\geq 0}$, $b \in \mathbb{R}^m_{\geq 0}$, *and a vector of finite upper bounds* $u \in \mathbb{R}^n_{\geq 0}$.

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

(v) *All nominal lower-level objective function coefficients are positive, i.e.,* $f_i > 0$ *holds for all* $i \in [n]$.

Under Assumption 3.3, the deterministic min-max problem (3.1) is a discrete interdiction problem with a follower problem that satisfies a downward monotonicity property. We refer to this type of problem as a *monotone interdiction problem*. The downward monotonicity property of the lower-level problem is formally stated in the following.

Proposition 3.4 (Monotonicity Property; cf. Proposition 4 in [YB3]). *Let* $x \in X$ *be given arbitrarily. Further, let* $y \in Y(x)$ *and let* $y' \in Y$ *be such that* $y' \leq y$ *holds. Then,* y' *is a feasible follower's decision for the given leader's decision* x*, i.e.,* $y' \in Y(x)$ *holds.*

All the proofs that we omit in this section can be found in [YB3]. Let us point out that Part (iii) of Assumption 3.3 ensures that the deterministic lower-level problem satisfies the downward monotonicity property. Problems satisfying this property arise in many packing-type applications, e.g., in marketing or facility location. Parts (ii) and (iv) of Assumption 3.3 are commonly used in the interdiction setting. In particular, we highlight that the leader's variables x are linked to the lower-level problem only via the interdiction constraints $y_i \leq u_i(1-x_i)$. In this context, Parts (i) and (ii) of Assumption 3.3 are included for the ease of presentation. However, the following results can as well be adapted to account for non-interdicting, i.e., non-linking and, thus, possibly non-binary, variables of the leader. Moreover, the case in which the lower-level problem also includes variables that are not subject to interdiction can be handled by partitioning the follower's variable set into interdicted and non-interdicted variables as it is done in Fischetti et al. (2019). Finally, we mention that Part (v) of Assumption 3.3 is w.l.o.g. since items with non-positive objective function coefficients are not chosen in an optimal response of the follower. Consequently, the leader does not need to spend interdiction resources on these items and, thus, we can omit all items with non-positive objective function coefficients in the problem formulation.

Before we come to the core of this section, which is the derivation of cuts tailored to monotone interdiction problems with Γ-robust followers, we present additional inequalities for the leader's variables x, which prove to be highly effective when considering the problem computationally.

Theorem 3.5 (cf. Theorem 3 in [YB3]). *For all i, $j \in [n]$ with $i \neq j$ chosen such that $A_{\cdot i} > A_{\cdot j}$, $B_{\cdot i} \leq B_{\cdot j}$, $u_i \geq u_j$, $f_i \geq f_j$, and $f_i - \Delta f_i \geq f_j - \Delta f_j$, the dominance inequality*

$$x_j \leq x_i \tag{3.11}$$

is satisfied in at least one optimal solution to Problem (3.7).

In [YB3], we derive two variants of so-called *interdiction cuts* from the extended formulation (3.8) and the multi-follower formulation (3.9) of the lower-level problem. The latter can be used within the branch-and-cut frameworks presented in Section 3.2. We summarize the derivation of these interdiction cuts and present enhancements (lifted cuts and dominance results) in the remainder of this section.

3.3.1 Deriving Interdiction Cuts from the Extended Formulation

In [YB3], we derive cuts for monotone interdiction problems with a Γ-robust follower from a penalty reformulation of the extended formulation (3.8). To ensure the validity of this approach, two aspects are essential. First, we exploit that the leader's variables x are linked to the lower-level problem only via the interdiction constraints; see Part (iii) of Assumption 3.3. Second, we require that the Γ-robust lower-level problem satisfies the downward monotonicity property. With respect to the latter, we show that the downward monotonicity property is preserved from the deterministic setting when a Γ-robust follower is considered. More formally, we have the following result.

Proposition 3.6 (cf. Proposition 5 in [YB3]). *Let $x \in X$ be given arbitrarily. Further, let (y, z, θ) be feasible for the x-parameterized problem* (3.8) *and let $y' \in Y$ be such that $y' \leq y$ holds. Then, (y', z, θ) is feasible for the x-parameterized problem* (3.8) *as well.*

In [YB3], we show that, by Proposition 3.6, the extended formulation (3.8) and the mixed-integer linear problem

$$\begin{aligned} \max_{y,z,\theta} \quad & \sum_{i=1}^{n} f_i y_i (1 - x_i) - \Gamma\theta - \sum_{i=1}^{n} z_i \\ \text{s.t.} \quad & z_i + \theta \geq \Delta f_i y_i, \quad i \in [n], \\ & y_i \leq u_i, \qquad\qquad\; i \in [n], \\ & y \in Y, \ \ z \in \mathbb{R}^{n_y}_{\geq 0}, \ \ \theta \in \mathbb{R}_{\geq 0}, \end{aligned} \tag{3.12}$$

admit the same optimal value for arbitrarily given $x \in X$. In Problem (3.12), we replace the interdiction constraints $y_i \leq u_i(1 - x_i)$ by the simple bound constraints $y_i \leq u_i$. Moreover, we add penalty terms $-f_i y_i x_i$ for the violation of interdiction constraints to the objective function. These modifications result in the feasible set of the problem being independent of the leader's decision x. In addition, the objective function is linear for given x. Hence, an optimal solution to Problem (3.12) is attained at a vertex of the convex hull of the feasible set. In what follows, we use $\hat{\Psi}$ to denote the set containing the finite number of vertices of the convex hull of the feasible set of Problem (3.12). Taking the previous considerations into account, we obtain

$$\eta \geq \Phi_{\text{rob}}(x) \geq \sum_{i=1}^{n} f_i \hat{y}_i (1 - x_i) - \Gamma\hat{\theta} - \sum_{i=1}^{n} \hat{z}_i$$

for any $(\hat{y}, \hat{z}, \hat{\theta}) \in \hat{\Psi}$ and (x, η) that is feasible for Problem (3.7). As a consequence, the interdiction cuts

$$\eta \geq \sum_{i=1}^{n} f_i \hat{y}_i (1 - x_i) - \Gamma \hat{\theta} - \sum_{i=1}^{n} \hat{z}_i \quad \text{for all } \left(\hat{y}, \hat{z}, \hat{\theta}\right) \in \hat{\Psi} \tag{3.13}$$

are valid for Problem (3.7). While the set $\hat{\Psi}$ may still be large, we show in [YB3] that the number of cuts to generate can be reduced by exploiting the notion of *maximal packings*. We call $(\hat{y}, \hat{z}, \hat{\theta}) \in \hat{\Psi}$ a maximal packing w.r.t. the extended formulation (3.12) if no $(y', \hat{z}, \hat{\theta}) \in \hat{\Psi} \setminus \{\hat{y}\}$ exists such that $\hat{y} \leq y'$ holds.

Proposition 3.7 (Proposition 8 in [YB3]). *Let $(\hat{y}, \hat{z}, \hat{\theta}) \in \hat{\Psi}$ be a non-maximal packing for Problem (3.12) and let $(y', \hat{z}, \hat{\theta}) \in \hat{\Psi} \setminus \{\hat{y}\}$ be chosen such that $\hat{y} \leq y'$ holds. Then, the interdiction cut (3.13) associated with $(\hat{y}, \hat{z}, \hat{\theta})$ is dominated by the interdiction cut associated with $(y', \hat{z}, \hat{\theta})$.*

Here and in what follows, domination between two cuts is understood in the sense that the feasible set induced by the dominating cut is contained in the feasible set induced by the dominated one. Next, we present lifted cuts that dominate the interdiction cuts (3.13).

Theorem 3.8 (cf. Theorem 4 in [YB3]). *For an arbitrarily given $(\hat{y}, \hat{z}, \hat{\theta}) \in \hat{\Psi}$, let $K \in [n]$, $S_a = \{a_1, \dots, a_K\} \subset [n]$, and $S_b = \{b_1, \dots, b_K\} \subset [n]$ be chosen such that $S_a \cap S_b = \emptyset$, $\hat{y}_{a_k} \geq 1$, $\hat{y}_{b_k} = 0$, $B_{\cdot a_k} \geq B_{\cdot b_k}$, $u_{a_k} \leq u_{b_k}$, $\Delta f_{a_k} < \Delta f_{b_k}$, and $f_{a_k} - \Delta f_{a_k} < f_{b_k} - \Delta f_{b_k}$ holds for all $k \in [K]$. Then, the following lifted interdiction cut is valid for Problem (3.7):*

$$\eta \geq \sum_{i=1}^{n} f_i \hat{y}_i (1 - x_i) + \sum_{k=1}^{K} \left((f_{b_k} - \Delta f_{b_k}) - (f_{a_k} - \Delta f_{a_k}) \right) \hat{y}_{a_k} (1 - x_{b_k}) - \Gamma \hat{\theta} - \sum_{i=1}^{n} \hat{z}_i. \tag{3.14}$$

3.3.2 Deriving Interdiction Cuts from the Multi-Follower Formulation

In [YB3], we pursue similar ideas compared to those in Fischetti et al. (2019) to derive interdiction cuts valid for Problem (3.7) using the multi-follower formulation (3.9). Under Assumption 3.3, the downward monotonicity property of the lower-level problem is naturally preserved in each follower sub-problem (3.10). This is due to the fact that the robustification of the lower-level objective function does not affect the constraints of Problem (3.10). Exploiting the latter, we show in [YB3] that, for all $\ell \in \mathcal{L}$ and arbitrarily given $x \in X$, the ℓth lower-level sub-problem (3.10) and the problem

$$-\Gamma \Delta f_\ell + \max_{y \in \mathcal{Y}} \left\{ \sum_{i=1}^{n} \tilde{f}(\ell)_i y_i (1 - x_i) \right\} \tag{3.15}$$

with $\mathcal{Y} := Y \cap \{0, 1\}^n$ admit the same optimal objective function value. The objective function of Problem (3.15) is linear for fixed leader's variables x and the feasible set $\mathcal{Y}$ is independent of x. Hence, an optimal solution to Problem (3.15) is attained at a vertex of the convex hull of $\mathcal{Y}$ and we obtain

$$\eta \geq \Phi_{\text{rob}}(x) \geq -\Gamma \Delta f_\ell + \sum_{i=1}^{n} \tilde{f}(\ell)_i \hat{y}_i (1 - x_i)$$

for any $\hat{y} \in \hat{\mathcal{Y}}$, $\ell \in \mathcal{L}$, and (x, η) that feasible for Problem (3.7). Here, we use $\hat{\mathcal{Y}}$ to denote the set containing the finite number of vertices of the convex hull of $\mathcal{Y}$. Taking this into account, the interdiction cuts

$$\eta \geq -\Gamma \Delta f_\ell + \sum_{i=1}^{n} \tilde{f}(\ell)_i \hat{y}_i (1 - x_i) \quad \text{for all } \hat{y} \in \hat{\mathcal{Y}}, \ \ell \in \mathcal{L}, \tag{3.16}$$

are valid for Problem (3.7). As it is done in Section 3.3.1, we aim to reduce the number of cuts to generate by exploiting the notion of maximal packings. In the multi-follower setting, we call $\hat{y} \in \hat{\mathcal{Y}}$ a maximal packing if no $y' \in \hat{\mathcal{Y}} \setminus \{\hat{y}\}$ exists such that $\hat{y} \leq y'$ holds. In particular, the latter does not require the specification of $\ell \in \mathcal{L}$ since we consider the same feasible set in each follower sub-problem. In contrast to the situation considered in Section 3.3.1, obtaining a dominance result for interdiction cuts associated with maximal packings using the multi-follower formulation is not as straightforward. In the multi-follower setting, we need to further study the properties of the modified objective function coefficients $\tilde{f}(\ell)$, $\ell \in \mathcal{L}$, which may be non-positive for certain items in some sub-problems. If $\tilde{f}(\ell)_i \leq 0$ holds for all $\ell \in \mathcal{L}$, the ith item will not be chosen in any of the follower sub-problems. Thus, the leader does not need to spend interdiction resources on the ith item so that we can omit this item in the problem formulation. However, if there is an item $i \in [n]$ with non-positive modified objective function coefficients only for some follower sub-problems, i.e., $\tilde{f}(s)_i \leq 0$ for all $s \in \mathcal{S} \subset \mathcal{L}$ and $\tilde{f}(t)_i > 0$ for all $t \in \mathcal{L} \setminus \mathcal{S}$, the ith item may be part of an optimal solution. To take such situations into account, we define the set

$$\mathcal{F}_+^\ell := \{i \in [n] \colon \tilde{f}(\ell)_i > 0\}, \quad \ell \in \mathcal{L},$$

so that the interdiction cuts (3.16) can be replaced with

$$\eta \geq -\Gamma \Delta f_\ell + \sum_{i \in \mathcal{F}_+^\ell} \tilde{f}(\ell)_i \hat{y}_i (1 - x_i) \quad \text{for all } \hat{y} \in \hat{\mathcal{Y}}, \ \ell \in \mathcal{L}. \tag{3.17}$$

In particular, the cuts (3.17) dominate the basic interdiction cuts (3.16). Moreover, we obtain the following dominance result.

Proposition 3.9 (cf. Proposition 10 in [YB3]). *Let $\hat{y} \in \hat{\mathcal{Y}}$ be a non-maximal packing for Problem (3.15) and let $y' \in \hat{\mathcal{Y}} \setminus \{\hat{y}\}$ be such that $\hat{y} \leq y'$ holds. Then, the interdiction cuts (3.17) associated with $\hat{y}$ are dominated by the interdiction cuts associated with y'.*

We emphasize that Proposition 3.9 does not apply to the interdiction cuts stated in (3.16) as $\tilde{f}(\ell)_i \hat{y}_i (1 - x_i) \leq \tilde{f}(\ell)_i y'_i (1 - x_i)$ may be violated for $i \notin \mathcal{F}_+^\ell$. We now conclude this section by presenting lifted interdiction cuts that dominate the ones in (3.17).

Theorem 3.10 (cf. Theorem 5 in [YB3]). *For arbitrarily given $\hat{y} \in \hat{\mathcal{Y}}$ and $\ell \in \mathcal{L}$, let $K \in [n]$, $S_a^\ell = \{a_1, \dots, a_K\} \subset \mathcal{F}_+^\ell$, and $S_b^\ell = \{b_1, \dots, b_K\} \subset \mathcal{F}_+^\ell$ be chosen such that $S_a^\ell \cap S_b^\ell = \emptyset$, $\hat{y}_{a_k} = 1$, $\hat{y}_{b_k} = 0$, $B_{\cdot a_k} \geq B_{\cdot b_k}$, and $\tilde{f}(\ell)_{a_k} < \tilde{f}(\ell)_{b_k}$ holds for all $k \in [K]$. Under Assumptions 2.1 and 3.2, the lifted interdiction cut*

$$\eta \geq -\Gamma \Delta f_\ell + \sum_{i \in \mathcal{F}_+^\ell} \tilde{f}(\ell)_i \hat{y}_i (1 - x_i) + \sum_{k=1}^{K} (\tilde{f}(\ell)_{b_k} - \tilde{f}(\ell)_{a_k})(1 - x_{b_k}) \tag{3.18}$$

is valid for Problem (3.7).

3.4 Computational Results

In [YB3], we computationally evaluate and compare the branch-and-cut methods including the enhancement techniques presented in Section 3.3 for Γ-robust monotone interdiction problems. We label the approaches based on the extended formulation (cf. Algorithm 3.2) and the multi-follower formulation (cf. Algorithm 3.3) as Ext and MF, respectively. To asses the performance of both methods, we consider the bilevel knapsack interdiction problem studied in Caprara et al. (2016), which is a prominent example for an interdiction problem with a monotone follower. We have adapted the deterministic instances considered in Caprara et al. (2016) to account for a Γ-robust follower. Overall, our test set contains 560 robustified knapsack interdiction instances. For detailed information regarding the generation of the test instances and the computational setup, we refer to Section 4 in [YB3]. In this extended summary, we only present the main results and conclusions of the computational study in [YB3]. Before we summarize our findings, let us briefly mention that the results shown in the following differ slightly from those presented in the published article.[1] First, we present new figures showing empirical cumulative distribution functions (ECDFs) w.r.t. the running times in which instances that cannot be solved within the time limit of 1 h are excluded. The latter has not been done for the plots published in [YB3]. Second, we exclude the instances that cannot be solved within the time limit when generating the statistical data for our tables. In the tables published in [YB3], these instances have been accounted for by taking the respective statistical quantities at the time limit. Third, we correct the labeling of the log-scaled x-axis in the ECDF plots w.r.t. runtimes.

3.4.1 Lifted Cuts and Dominance Inequalities

In [YB3], we observe that Ext and MF significantly benefit from the tailored techniques presented in Section 3.3. To illustrate this, we show ECDF plots w.r.t. the running times for the variants of our methods with and without lifted cuts and dominance inequalities in Figure 3.1. The ECDFs can be interpreted as the percentage of instances (y-axis) that can be solved within a certain amount of time (log-scaled x-axis). Here, Ext and MF refer to the setting in which only basic interdiction cuts, i.e., Inequalities (3.13) and (3.16), are added without any further refinements. For MF, we consider the cut separation strategy in which all violated interdiction cuts are added to the problem formulation. Ext-D and MF-D are used to denote the extensions of Ext and MF in which we add the dominance inequalities (3.11) to the problem formulation. We use Ext-L and MF-L to abbreviate the setting without dominance inequalities but in which we add lifted interdiction cuts instead of the basic interdiction cuts; cf. Theorems 3.8 and 3.10. Finally, the methods that include both lifted cuts and dominance inequalities are referred to as Ext-LD and MF-LD. In Figure 3.1, we consider a total of 512 instances, excluding those that none of the considered variants can solve within the time limit of 1 h as well as those that every variant can solve within 5 s. First, it can be seen that lifted interdiction cuts slightly enhance the performance of Ext and MF. Second, we observe significant improvements for both approaches when adding dominance inequalities. Third, Figure 3.1 shows that the combination of both lifted cuts and dominance inequalities yields the best performance

---

[1] All details, including new figures and tables for all results in [YB3], can be found at `https://github.com/YasmineBeck/gamma-robust-knapsack-interdiction-solver`.

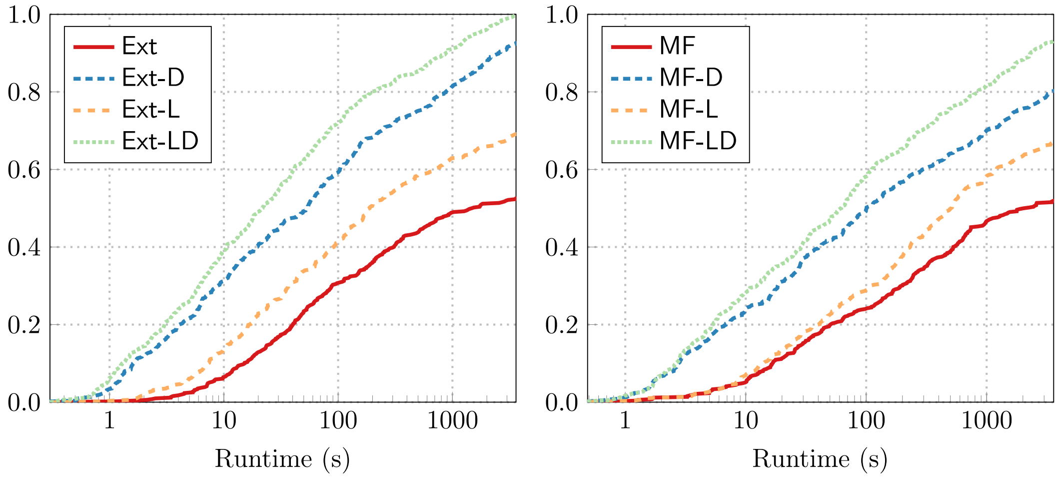

Figure 3.1: Log-scaled ECDF plots of the runtimes (in s) for Ext (left) and MF (right) using lifted cuts and/or dominance inequalities; cf. Figure 1 in [YB3].

among the considered variants of the respective solution approach. In particular, it can be seen that Ext-LD and MF-LD solve almost twice as many instances as the plain variants Ext and MF. In addition, we observe in [YB3] that adding only the interdiction cuts corresponding to maximal packings of the follower (cf. Propositions 3.7 and 3.9) further improves the performance of both Ext-LD and MF-LD.

3.4.2 Cut Separation Strategies and the Potential of Parallelization

Up to $|\mathcal{L}|$ interdiction cuts can be added at each node of the branch-and-cut search tree when using the multi-follower formulation (3.9). Nevertheless, instead of adding all violated cuts at each node, it is also valid to consider different cut separation strategies. In [YB3], we observe that the choice of the cut separation strategy may significantly affect the performance of the overall solution method. We illustrate this in Figure 3.2 (left) using ECDF plots w.r.t. the running times. Here, we consider MF with lifted cuts corresponding to maximal packings of the follower, dominance inequalities, and the cut separation strategies summarized in Table 3.2. For notational convenience, we omit MF as a prefix when comparing different cut separation strategies in Figure 3.2. We consider a total of 496 instances, excluding those that none of the considered variants can solve within the time limit of 1 h as well as those that every variant can solve within 5 s. Based on Figure 3.2 (left), it can be seen that Random outperforms all other cut separation strategies. The potential reasons for the effectiveness of Random are two-fold. On the one hand, it seems beneficial to add a single cut in each iteration of the algorithm; otherwise, as for the strategy All-In, the leader's problem can get extremely large w.r.t. the number of constraints. On the other hand, Random has comparatively low computational costs. For Sorting, we update the information regarding the cumulative violation of the interdiction cuts and the frequency of generation in each iteration of the algorithm. We then order the follower sub-problems in non-increasing order w.r.t. the ratio of cumulative violation to frequency of generation. Overall, this causes higher computational costs

Table 3.2: Cut separation strategies considered in [YB3].

All-In	All cuts that are violated by the current leader's decision are added to the problem formulation.
Most-Violated	A single cut that is maximally violated by the current leader's decision is added to the problem formulation.
Sorting	Sub-problems that produce potentially good cuts are identified using a sorting mechanism similar to the one presented in Miranda et al. (2015). A single potentially good cut is added to the problem formulation.
First-In	The first cut that is violated by the current leader's decision is added to the problem formulation.
Random	Among the violated cuts, we randomly choose a single cut and add it to the problem formulation.

compared to Random. Moreover, when considering Most-Violated, we need to solve all follower sub-problems to determine the most violated cut, which seems to be rather expensive. However, the situation changes when we account for parallelization within MF. Since the follower sub-problems (3.10) that are solved to generate interdiction cuts are independent, these sub-problems can be solved in parallel if the necessary capacities are available. In [YB3], we use so-called *idealized parallel runtimes* to assess the potential of parallelization. The latter reflect the overall runtime of a solution 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 the sub-problems sequentially by taking the maximum of all runtimes for the sub-problems. This means that, if a problem could not be solved within the time limit of 1 h in the sequential setting, it is also considered as unsolved in the idealized parallel setting. Figure 3.2 (right) shows that, in the idealized parallel setting, Most-Violated dominates all other cut separation strategies. The latter is not surprising as we can benefit the most from parallelization for Most-Violated, where we indeed need to solve all of the follower sub-problems to determine the most violated cut.

3.4.3 Comparison of Solution Approaches

We conclude this section by comparing the "winning" parameterizations of Ext and MF, i.e., we consider the variants of our approaches with those enhancement techniques that lead to the overall best performance of our methods. This means that, for both Ext and MF, we consider the variants with lifted cuts corresponding to maximal packings of the follower and dominance inequalities. For MF, we further distinguish between the sequential and the idealized parallel setting. In the sequential setting, we consider the variant that uses the cut separation strategy Random. We label this approach as MF-seq. In the idealized parallel setting, we consider the variant that exploits the strategy Most-Violated, which we abbreviate with MF-ideal. In Figure 3.3, we show ECDF plots w.r.t. the running times for the three considered "winning" approaches. We emphasize that we consider sequential

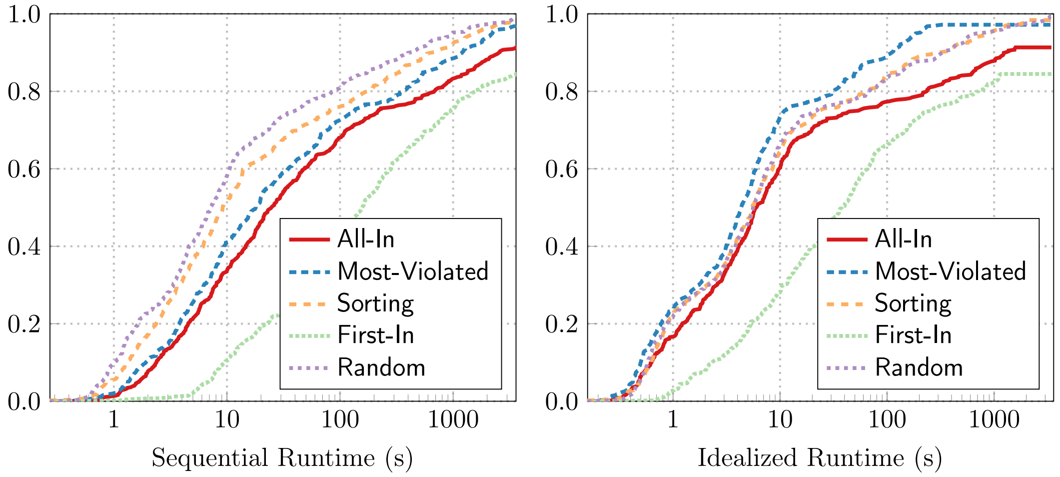

Figure 3.2: Log-scaled ECDF plots of the sequential runtimes (left) and the idealized parallel runtimes (right) (in s) for different cut separation strategies; cf. Figure 3 in [YB3].

and idealized parallel runtimes for MF-seq and MF-ideal, respectively. Figure 3.3 is based on a total of 407 instances, excluding those that none of the considered variants can solve within the time limit of 1 h as well as those that every variant can solve within 5 s. We observe that MF-ideal clearly outperforms the remaining two approaches w.r.t. running times. The latter is also underlined by the runtime results shown in Table 3.3. The previous observations particularly demonstrate that the strength of the multi-follower approach lies in the possibility to parallelize the solution of the follower sub-problems. However, if the capacities are not available to have an idealized parallelization, the multi-follower approach MF-seq still performs slightly better than Ext. Nevertheless, reflected by the number of solved instances shown in Table 3.3, Ext seems to have an advantage over MF-seq on computationally challenging instances. We further mention that, using the extended formulation, one could transform the Γ-robust interdiction problem into a standard mixed-integer linear bilevel problem. The latter can, in general, be solved using the MibS solver (Tahernejad et al. 2020) or the general branch-and-cut solver presented in Fischetti et al. (2017). We have tested both solvers for 40 robustified knapsack interdiction instances with 35 items in a preliminary computational study. The results revealed that our tailored methods outperform both solvers by significant orders of magnitude, which is why we omit a more detailed comparison with these general-purpose solvers.

Finally, we briefly elaborate on the *price of robustness* (Bertsimas and Sim 2004), which reflects the impact of robustification on the optimal objective function value of the problem. We observe that robustification consistently leads to smaller objective function values compared to their nominal counterparts. In particular, the optimal objective function value decreases with increasing values of Δf and Γ. In [YB3], we further address the so-called *computational cost of robustness*, which reflects the effect of robustification, e.g., on the overall runtimes of the solution methods. For Ext, we observe that the time required to solve the Γ-robust knapsack interdiction problem increases with increasing values of Δf and Γ. For MF-seq and MF-ideal, this does not seem to be the case in principle due to the following reasons. As the value of Γ increases, the number

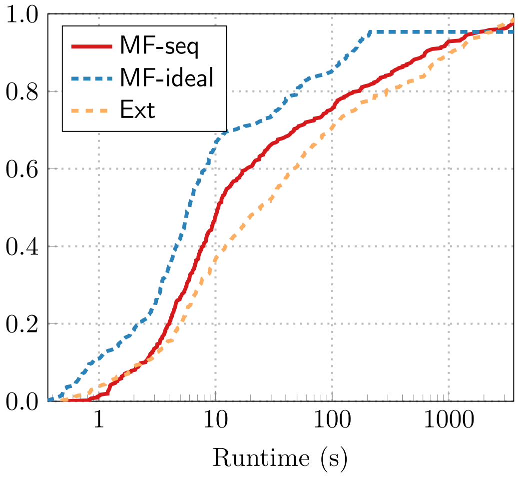

Figure 3.3: Log-scaled ECDF plots of the runtimes (in s) for the "winner settings" of Ext and MF; cf. Figure 5 in [YB3].

Table 3.3: Mean and median runtimes (in s) as well as the number of solved instances (out of 407 considered instances) for the "winner settings" of Ext and MF; cf. Table 5 in [YB3].

	runtimes		
	mean	median	solved
Ext	267.04	23.31	400
MF-seq	192.18	10.32	398
MF-ideal	33.47	5.62	387

of follower sub-problems that need to be solved to generate interdiction cuts using the multi-follower formulation (3.9) decreases. This leads to overall reduced computational costs in the sequential multi-follower setting. In the idealized parallel setting, in which all sub-problems are solved in parallel, the number of sub-problems to solve becomes less critical as only the time to solve one sub-problem contributes to the overall runtime of the method. Hence, the performance of MF-ideal is comparable to that of the method in the nominal setting. Interestingly, we also observe in this context that robustification does not always lead to increased computational costs. This means that, for some instances, the robust counterpart can be solved faster than the nominal problem. In general, however, Γ-robust solutions are obtained at the expense of increased computational difficulty of the problem. Nevertheless, provided that there are sufficient capacities available to solve all of the follower sub-problems in parallel, the price of robustness w.r.t. running times is comparatively small.

3.5 Approaches for Related Bilevel Problems

To conclude this chapter on exact methods for mixed-integer, linear, and Γ-robust min-max problems, we shift our focus to exact branch-and-cut approaches for related bilevel problems. The approaches we present in the following have been considered as benchmark solvers to assess the performance of the heuristics presented in [YB4]. As these approaches have only been briefly addressed in [YB4], we discuss them in more detail in this dissertation. In Section 3.5.1, we present a branch-and-cut framework for so-called *generalized monotone interdiction problems*; see, e.g., Fischetti et al. (2018b) for heuristic approaches for this class of optimization problems. While the considered bilevel problems are deterministic ones, we demonstrate in Chapter 4 that solving these problems can also be relevant when considering Γ-robust bilevel problems computationally. In Section 3.5.2, we present a branch-and-cut framework for the Γ-robust variant of generalized monotone interdiction problems.

3.5.1 A Generalization of Monotone Interdiction Problems

In [YB4], we consider generalizations of deterministic monotone interdiction problems (cf. Section 3.3) in which the leader and the follower do not consider the same objective function that is optimized in opposing directions but, instead, both players may have different objectives. More formally, we consider bilevel problems of the form

$$\begin{aligned} \min_{x,y} \quad & c^\top x + d^\top y \\ \text{s.t.} \quad & x \in X := \{x \in \{0,1\}^n : Ax \geq a\}, \\ & y \in \arg\max_{y'} \left\{ f^\top y' \colon y' \in Y(x) \right\}, \end{aligned} \tag{3.19}$$

where $X \neq \emptyset$ and $Y(x) := \{y \in \{0,1\}^n \colon By \leq b,\ y_i \leq 1 - x_i,\ i \in [n]\}$ with $c, d, f \in \mathbb{R}^n$, $A \in \mathbb{R}^{l \times n}$, $a \in \mathbb{R}^l$, $B \in \mathbb{R}^{m \times n}_{\geq 0}$, and $b \in \mathbb{R}^m_{\geq 0}$. We consider the optimistic approach to bilevel optimization. Hence, whenever the optimal response of the follower is not unique, he selects the one that favors the leader w.r.t. her objective function value. By construction, Problem (3.19) satisfies Assumption 3.1, i.e., Problem (3.19) has an optimal solution. Moreover, we emphasize that the lower-level problem satisfies the downward monotonicity property; cf. Proposition 3.4. The value-function reformulation of Problem (3.19) is given by

$$\begin{aligned} \min_{x,y} \quad & c^\top x + d^\top y \\ \text{s.t.} \quad & x \in X,\ \ y \in Y(x), \\ & f^\top y \geq \Phi(x) \end{aligned} \tag{3.20}$$

with the lower-level optimal-value function

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

We solve Problem (3.20) using a branch-and-cut framework similar to the one presented in DeNegre and Ralphs (2009). We summarize the method in the following. We start by solving the linear problem

$$\min_{x,y} \quad c^\top x + d^\top y \quad \text{s.t.} \quad (x,y) \in \Omega_0 \tag{3.21}$$

with

$$\Omega_0 := \{(x,y) \in [0,1]^n \times [0,1]^n \colon Ax \geq a, \ By \leq b, \ y_i \leq 1 - x_i, \ i \in [n]\}.$$

After considering Problem (3.21), we iteratively add valid inequalities or branch to separate integer- or bilevel-infeasible points. At node j of the branch-and-cut search tree, we consider the problem

$$\min_{x,y} \quad c^\top x + d^\top y \quad \text{s.t.} \quad (x,y) \in \Omega_j \subseteq [0,1]^n \times [0,1]^n, \tag{3.22}$$

where the set Ω_j contains all valid inequalities that have been added previously to cut off integer- or bilevel-infeasible points as well as all branching decisions. The method to process node j of the branch-and-cut search tree is formally stated in Algorithm 3.4. Here, U denotes the current upper bound on the optimal objective function value of Problem (3.19). To separate bilevel-infeasible points in Line 12 of Algorithm 3.4, one can use generic cuts such as, e.g., (generalized) no-good cuts; see DeNegre (2011), DeNegre and Ralphs (2009), and Tahernejad et al. (2020). The remainder of this section, however, is devoted to obtaining stronger formulations by exploiting the ideas of Fischetti et al. (2019) and Section 3.3 of this dissertation.

Algorithm 3.4 Processing node j of the branch-and-cut search tree

```
 1: Solve Problem (3.22).
 2: if Problem (3.22) is infeasible then
 3:     Fathom the current node.
 4: Let $(x^j, y^j)$ denote the optimal solution to Problem (3.22).
 5: if $c^\top x^j + d^\top y^j \geq U$ then
 6:     Fathom the current node.
 7: if $x^j \notin \{0,1\}^n$ or $y^j \notin \{0,1\}^n$ then
 8:     Generate cuts valid for $\Omega_j \cap (\{0,1\}^n \times \{0,1\}^n)$, augment $\Omega_j$, and go to Step 1
        or branch.
 9: Compute a solution $\hat{y}$ to the $x^j$-parameterized lower-level problem
        $\max_y \quad f^\top y \quad \text{s.t.} \quad Y(x^j).$
10: Set $\Phi(x^j) \leftarrow f^\top \hat{y}$ and $U \leftarrow \min\{U, c^\top x^j + d^\top \hat{y}\}$.
11: if $f^\top y^j < \Phi(x^j)$ then
12:     Generate a valid cut that excludes $(x^j, y^j)$ from $\Omega_j$, augment $\Omega_j$, and go to Step 1.
```

Proposition 3.11. *Let $x \in X$ be given arbitrarily. Then, the x-parameterized lower-level problem*

$$\max_y \quad f^\top y \quad s.t. \quad Y(x) \tag{3.23}$$

and the problem

$$\max_y \quad \sum_{i=1}^n f_i y_i (1 - x_i) \quad s.t. \quad y \in \mathcal{Y} := \left\{y' \in \{0,1\}^n : By' \leq b\right\} \tag{3.24}$$

admit the same optimal value.

Proof. Let y^* be an optimal solution to Problem (3.23). Then, $y_i^* x_i = 0$ holds for all $i \in [n]$ and, in particular, y^* is feasible for Problem (3.24). Thus, we obtain

$$\max_{y \in Y(x)} \left\{ f^\top y \right\} = f^\top y^* = \sum_{i=1}^n f_i y_i^* (1 - x_i) \leq \max_{y \in \mathcal{Y}} \left\{ \sum_{i=1}^n f_i y_i (1 - x_i) \right\}.$$

Let now $\hat{y}$ be an optimal solution to Problem (3.24). Suppose that there is an item $k \in [n]$ for which the interdiction constraint $\hat{y}_k \leq 1 - x_k$ is violated, i.e., $\hat{y}_k = 1 = x_k$ holds. Then, we consider the alternative follower's decision

$$y_i' = \begin{cases} \hat{y}_i, & i \in [n] \setminus \{k\}, \\ 0, & i = k. \end{cases}$$

By construction, y' is feasible for Problem (3.24) and it satisfies the interdiction constraint $y_k' \leq 1 - x_k$. Moreover, we have

$$\max_{y \in \mathcal{Y}} \left\{ \sum_{i=1}^n f_i y_i (1 - x_i) \right\} = \sum_{i=1}^n f_i \hat{y}_i (1 - x_i) = \sum_{i=1}^n f_i y_i' (1 - x_i).$$

Here, the last equality follows from the construction of y' as well as from $f_k \hat{y}_k (1 - x_k) = 0 = f_k y_k' (1 - x_k)$ for $x_k = 1$. As a consequence, y' is an optimal solution to Problem (3.24) as well. We repeat the previous procedure until there are no items remaining that violate the corresponding interdiction constraint. Ultimately, we obtain a follower's decision y' that solves Problem (3.24) and that satisfies all interdiction constraints. Hence, y' is feasible for Problem (3.23) and we obtain

$$\max_{y \in \mathcal{Y}} \left\{ \sum_{i=1}^n f_i y_i (1 - x_i) \right\} = \sum_{i=1}^n f_i y_i' (1 - x_i) = \sum_{i=1}^n f_i y_i' \leq \max_{y \in Y(x)} \left\{ f^\top y \right\}.$$

This concludes the proof. □

For given $x \in X$, the feasible set of Problem (3.24) is independent of x and the objective function is linear. Hence, an optimal solution to Problem (3.24) is attained at a vertex of the convex hull of $\mathcal{Y}$. As a consequence, we obtain the following.

Theorem 3.12. *Problem (3.20) can equivalently be reformulated by replacing the constraint $f^\top y \geq \Phi(x)$ with*

$$f^\top y \geq \sum_{i=1}^n f_i \hat{y}_i (1 - x_i) \quad \text{for all } \hat{y} \in \hat{\mathcal{Y}},$$

where $\hat{\mathcal{Y}}$ denotes the set containing the finite number of vertices of the convex hull of $\mathcal{Y} := \{y \in \{0,1\}^n : By \leq b\}$.

Proof. From Proposition 3.11, we obtain $f^\top y \geq \Phi(x) \geq \sum_{i=1}^n f_i \hat{y}_i (1 - x_i)$ for any $\hat{y} \in \hat{\mathcal{Y}}$ and (x, y) that is feasible for Problem (3.20). This concludes the proof. □

3.5.2 A Generalization of Γ-Robust Monotone Interdiction Problems

In [YB4], we also consider the Γ-robust variant of Problem (3.19), which is given by

$$\begin{aligned} \min_{x,y} \quad & c^\top x + d^\top y \\ \text{s.t.} \quad & x \in X, \ y \in S_\Gamma(x). \end{aligned} \tag{3.25}$$

Here, we use $S_\Gamma(x)$ to denote the set of optimal solutions to the x-parameterized Γ-robust counterpart of the lower-level problem. The latter is formally stated in (3.6). The Γ-robust counterpart of the overall bilevel problem (3.25) can be written as

$$\begin{aligned} \min_{x,y} \quad & c^\top x + d^\top y \\ \text{s.t.} \quad & x \in X, \ y \in Y(x), \\ & f^\top y - \max_{\{S \subseteq [n] \colon |S| \leq \Gamma\}} \sum_{i \in S} \Delta f_i y_i \geq \Phi_{\text{rob}}(x). \end{aligned} \tag{3.26}$$

Using the same arguments as in the proof of Theorem 3 in Bertsimas and Sim (2003), we can replace the last inequality in Problem (3.26) by

$$\sum_{i=1}^n f_i y_i - \Gamma\theta - \sum_{i=1}^n z_i \geq \Phi_{\text{rob}}(x), \quad z_i + \theta \geq \Delta f_i y_i, \ i \in [n], \quad z \in \mathbb{R}^n_{\geq 0}, \quad \theta \in \mathbb{R}_{\geq 0},$$

i.e., we consider the problem

$$\begin{aligned} \min_{x,y,z,\theta} \quad & c^\top x + d^\top y \\ \text{s.t.} \quad & z_i + \theta \geq \Delta f_i y_i, \quad i \in [n], \\ & \sum_{i=1}^n f_i y_i - \Gamma\theta - \sum_{i=1}^n z_i \geq \Phi_{\text{rob}}(x), \\ & x \in X, \ y \in Y(x), \ z \in \mathbb{R}^n_{\geq 0}, \ \theta \in \mathbb{R}_{\geq 0}. \end{aligned} \tag{3.27}$$

In [YB4], we use a branch-and-cut framework similar to the one outlined in Section 3.3.1 to solve Problem (3.27). We initialize the method by solving the linear problem

$$\begin{aligned} \min_{x,y,z,\theta} \quad & c^\top x + d^\top y \\ \text{s.t.} \quad & Ax \geq a, \ By \leq b, \\ & y_i \leq 1 - x_i, \quad i \in [n], \\ & z_i + \theta \geq \Delta f_i y_i, \quad i \in [n], \\ & (x, y, z, \theta) \in [0,1]^n \times [0,1]^n \times \mathbb{R}^n_{\geq 0} \times \mathbb{R}_{\geq 0}. \end{aligned}$$

Afterward, we iteratively add valid inequalities or branch to separate integer- or bilevel-infeasible points. At node j of the branch-and-cut tree, we consider the problem

$$\begin{aligned} \min_{x,y,z,\theta} \quad & c^\top x + d^\top y \\ \text{s.t.} \quad & Ax \geq a, \ By \leq b, \\ & y_i \leq 1 - x_i, \quad i \in [n], \\ & z_i + \theta \geq \Delta f_i y_i, \quad i \in [n], \\ & (x, y, z, \theta) \in \Omega_j \subseteq [0,1]^n \times [0,1]^n \times \mathbb{R}^n_{\geq 0} \times \mathbb{R}_{\geq 0}, \end{aligned} \tag{3.28}$$

where the set Ω_j contains all valid inequalities that have been added previously to cut off integer- or bilevel-infeasible points as well as all branching decisions. The method to process node j of the branch-and-cut search tree is formally stated in Algorithm 3.5. Here, U denotes the current upper bound on the optimal objective function value of Problem (3.25). To derive problem-tailored cuts that can be used in Line 12 of Algorithm 3.5 for separating bilevel-infeasible points, we exploit Problem (3.12) in which we set $u = (1, \dots, 1)^\top \in \mathbb{R}^n$. In [YB3], it is shown that Problem (3.12) and the extended formulation (3.8) admit the same optimal value. From the latter, we obtain the following result.

Theorem 3.13. *Problem (3.27) can equivalently be reformulated by replacing the constraint*

$$\sum_{i=1}^{n} f_i y_i - \Gamma\theta - \sum_{i=1}^{n} z_i \geq \Phi_{rob}(x)$$

with

$$\sum_{i=1}^{n} f_i y_i - \Gamma\theta - \sum_{i=1}^{n} z_i \geq \sum_{i=1}^{n} f_i \hat{y}_i (1 - x_i) - \Gamma\hat{\theta} - \sum_{i=1}^{n} \hat{z}_i \quad \text{for all } \left(\hat{y}, \hat{z}, \hat{\theta}\right) \in \hat{\Psi},$$

where $\hat{\Psi}$ denotes the set containing the finite number of vertices of the convex hull of

$$\Psi := \left\{(y, z, \theta) \in \{0,1\}^n \times \mathbb{R}^n_{\geq 0} \times \mathbb{R}_{\geq 0} \colon By \leq b,\ z_i + \theta \geq \Delta f_i y_i,\ i \in [n]\right\}.$$

Proof. For any $(\hat{y}, \hat{z}, \hat{\theta}) \in \hat{\Psi}$ and (x, y, z, θ) feasible for Problem (3.27), we obtain

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

from Proposition 6 in [YB3]. This concludes the proof. □

Computational results for the approaches presented in this section are provided in Chapter 4 of this dissertation.

Algorithm 3.5 Processing node j of the branch-and-cut search tree

```
1: Solve Problem (3.28).
2: if Problem (3.28) is infeasible then
3:     Fathom the current node.
4: Let $(x^j, y^j, z^j, \theta^j)$ denote the optimal solution to Problem (3.28) and set
       $$v_j \leftarrow \sum_{i=1}^{n} f_i y_i^j - \Gamma\theta^j - \sum_{i=1}^{n} z_i^j.$$
5: if $c^\top x^j + d^\top y^j \geq U$ then
6:     Fathom the current node.
7: if $x^j \notin \{0,1\}^n$ or $y^j \notin \{0,1\}^n$ then
8:     Generate cuts valid for $\Omega_j \cap (\{0,1\}^n \times \{0,1\}^n \times \mathbb{R}^n_{\geq 0} \times \mathbb{R}_{\geq 0})$, augment $\Omega_j$, and go
       to Step 1 or branch.
9: Compute a solution $(\hat{y}, \hat{z}, \hat{\theta})$ to the $x^j$-parameterized extended formulation of the
   lower-level problem (3.8).
10: Set $U \leftarrow \min\{U, c^\top x^j + d^\top \hat{y}\}$ and
       $$\Phi_{\text{rob}}(x^j) \leftarrow \sum_{i=1}^{n} f_i \hat{y}_i - \Gamma\hat{\theta} - \sum_{i=1}^{n} \hat{z}_i.$$
11: if $v_j < \Phi_{\text{rob}}(x^j)$ then
12:     Generate a valid cut that excludes $(x^j, y^j, z^j, \theta^j)$ from $\Omega_j$, augment $\Omega_j$, and go to
        Step 1.
```

4. Heuristic Methods for Mixed-Integer, Linear, and Γ-Robust Bilevel Problems

Developing solution methods for mixed-integer linear bilevel problems is a challenging task, especially if problems under uncertainty are considered. To the best of our knowledge, there are currently no methods in the literature that can directly tackle mixed-integer, linear, and Γ-robust bilevel problems, except for the problem-tailored and exact approaches presented in [YB3]. In addition to exact approaches, it thus also seems reasonable to consider primal heuristics for these problems. In this chapter, we present such heuristics for mixed-integer linear bilevel problems with a binary lower-level problem and a Γ-robust treatment of lower-level objective uncertainty. These heuristics, which have been proposed in [YB4], 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. In the remainder of this chapter, we summarize the main contributions of [YB4]. In Section 4.1, we present heuristics for the special case of mixed-integer, linear, and Γ-robust min-max problems. In Section 4.2, we present a heuristic for more general mixed-integer, linear, and Γ-robust bilevel problems.

4.1 Heuristics for Γ-Robust Min-Max Problems

We start by considering mixed-integer linear min-max problems, which have been formally stated in their deterministic form in Problem (3.1). Again, we use the lower-level optimal-value function to obtain an equivalent single-level reformulation, which is given by

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

Building on Section 3.2, we consider the robustified variant of the problem in which the follower hedges against a subset of at most $\Gamma \in \{0, \dots, n_y\}$ deviations in his uncertain objective function coefficients. The Γ-robust counterpart of the min-max problem can be obtained by replacing $\Phi(x)$ with $\Phi_{\text{rob}}(x)$. The latter is formally stated in (3.6). Throughout this chapter, we impose Assumptions 2.1, 3.1, and 3.2. Assumption 3.1 ensures that the considered bilevel problems have an optimal solution. Assumptions 2.1 and 3.2 are necessary to exploit Corollary 3.2 and Theorem 2.3 so that we can re-write the Γ-robust counterpart of the min-max problem (3.1) as

$$v_{\text{rob}} := \min_{x \in X} \left\{ c^\top x + \Phi_{\text{rob}}(x) \right\} = \min_{x \in X} \left\{ c^\top x + \max_{\ell \in \mathcal{L}} \{\Phi_\ell(x)\} \right\} \tag{4.1}$$

with

$$\mathcal{L} = \{\Gamma + 1, \Gamma + 3, \Gamma + 5, \ldots, \Gamma + \gamma, n_y + 1\} \tag{4.2}$$

and γ being the largest odd integer such that $\Gamma + \gamma < n_y + 1$ holds. In Section 4.1.1, we present a heuristic for Problem (4.1), along with a quality guarantee for heuristically obtained solutions. Afterward, in Section 4.1.2, we elaborate on the potential of parallelization and present a modified variant of our method. In Section 4.1.3, we discuss algorithmic refinements and provide sufficient ex-post conditions for global optimality of the outcomes of our methods. Finally, we conclude in Section 4.1.4 with computational results.

4.1.1 A Heuristic in the Spirit of Bertsimas and Sim

The heuristic we present in the following builds on a lower bounding scheme that exploits the ideas of Theorem 2.2 (Bertsimas and Sim 2003) and Theorem 2.3 (Lee and Kwon 2014). The following proposition is central to the algorithmic developments and serves as a motivation for our approach. A proof of this proposition, along with all the proofs that we omit in this chapter, can be found in [YB4].

Proposition 4.1 (Proposition 1 in [YB4]). *For all $\ell \in \mathcal{L}$, let*

$$v_\ell := \min_{x \in X} \left\{ c^\top x + \Phi_\ell(x) \right\}. \tag{4.3}$$

Then, $v_{rob} \geq v_\ell$ holds, i.e., v_ℓ is a valid lower bound for the optimal objective function value of (4.1). In particular, $v_{rob} \geq \max\{v_{\ell'} : \ell' \in \mathcal{L}\}$.

We formally state the heuristic for Problem (4.1) that has been proposed in [YB4] in Algorithm 4.1. In Lines 3 and 4 of the algorithm, we exploit the result of Proposition 4.1 so that a valid lower bound for the optimal objective function value of Problem (4.1) is obtained by solving appropriately chosen problems of the nominal type. Under Assumption 3.1, an optimal solution to Problem (4.3) exists for all $\ell \in \mathcal{L}$ so that Line 3 of Algorithm 4.1 is well-defined. For each leader's solution x^ℓ, $\ell \in \mathcal{L}$, we solve the x^ℓ-parameterized Γ-robust counterpart (3.6) of the lower-level problem to obtain a valid upper bound for (4.1). Since x^ℓ is a solution to Problem (4.3), we have $Y(x^\ell) \neq \emptyset$, i.e., Line 7 of the algorithm is also well-defined. We emphasize that any solver for deterministic mixed-integer linear min-max problems can be used for the solution of the problems considered in Line 3 of Algorithm 4.1. 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). Moreover, any method for Γ-robust single-level problems can be used in Line 7 of the algorithm. In particular, we can tackle the Γ-robust counterpart (3.6) of the lower-level problem by exploiting Corollary 3.1 or 3.2. Overall, the heuristic presented in Algorithm 4.1 relates to the main result by Bertsimas and Sim (2003) and Sim (2004) in the sense that we solve a linear number of (up to $|\mathcal{L}|$) problems of the nominal type. Nevertheless, our approach differs from the Bertsimas–Sim result since, in addition to solving these deterministic problems, we further solve robustified lower-level problems.

Algorithm 4.1 A Heuristic for Γ-Robust Mixed-Integer Linear Min-Max Problems

Input: An instance of Problem (4.1), an exact solution method for Problems (3.1) and (3.6), an index set $\mathcal{L}$ as in (4.2)

Output: A feasible leader's decision x^*, a lower bound L, and an upper bound U for Problem (4.1)

```
 1: Set $x^* \leftarrow$ None, $L \leftarrow -\infty$, and $U \leftarrow \infty$.
 2: for all $\ell \in \mathcal{L}$ do
 3:     Compute a solution $x^\ell$ to the deterministic min-max problem
            $$v_\ell \leftarrow \min_{x \in X} \left\{ c^\top x + \Phi_\ell(x) \right\}.$$
 4:     Set $L \leftarrow \max\{L, v_\ell\}$.
 5:     if $U \leq L$ then
 6:         return $x^*$, $L$, $U$
 7:     Solve the $x^\ell$-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 $x^*$, $L$, $U$
```

Theorem 4.2 (Theorem 1 in [YB4]). *Algorithm 4.1 is correct, i.e., it returns a feasible leader's decision x^* as well as valid lower and upper bounds L and U for (4.1).*

Remark 4.3 (Remark 2 in [YB4]). *If Algorithm 4.1 terminates with (x^*, L, U) in Line 6 or 11, $U - L = 0$ holds and x^* is an optimal solution to Problem (4.1).*

By construction, if Algorithm 4.1 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. The latter, however, does not necessarily imply that none of the $|\mathcal{L}|$ bilevel sub-problems produces a solution that is optimal for Problem (4.1). The reasons are two-fold. On the one hand, this may be due to the multiplicity of solutions to the deterministic min-max problems (4.3). On the other hand, we emphasize that the sub-problems (4.3) are only relaxations of (4.1).

4.1.2 Parallelization

The deterministic min-max problems (4.3) that are solved in Line 3 of Algorithm 4.1 are independent. Hence, if the necessary capacities are available, they can be solved in parallel. Instead of alternating between solving deterministic min-max problems and robustified lower-level problems as it is done in Algorithm 4.1, it may thus 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 4.1, which we summarize in Algorithm 4.2. The correctness of the method immediately follows from Theorem 4.2.

If the necessary capacities are available, Lines 2 and 3 of Algorithm 4.2 can be parallelized. Moreover, we can make use of parallelization in Line 7 of the algorithm by exploiting Theorem 2.3 to solve the Γ-robust counterpart of the lower-level problem.

Algorithm 4.2 A Modification of Algorithm 4.1

```
Input: An instance of Problem (4.1), an exact solution method for Problems (3.1)
    and (3.6), an index set $\mathcal{L}$ as in (4.2)
Output: A feasible leader's decision $x^*$, a lower bound $L$, and an upper bound $U$ for
    Problem (4.1)
 1: Set $x^* \leftarrow$ None, $L \leftarrow -\infty$, and $U \leftarrow \infty$.
 2: for all $\ell \in \mathcal{L}$ do
 3:     Compute a solution $x^\ell$ to the deterministic min-max problem
            $$v_\ell \leftarrow \min_{x \in X} \left\{ c^\top x + \Phi_\ell(x) \right\}.$$
 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 $x^{\ell_i}$-parameterized $\Gamma$-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 $x^*$, $L$, $U$
```

4.1.3 Reduction of Sub-Problems to Be Solved

Algorithms 4.1 and 4.2 terminate after solving at most $|\mathcal{L}|$ deterministic min-max problems and Γ-robust counterparts of the lower-level problem, respectively. If we exploit Theorem 2.3, this means that up to $|\mathcal{L}|^2$ deterministic lower-level problems are solved. Thus, it is evident that Algorithms 4.1 and 4.2 require a significant amount of resources—especially for large index sets $\mathcal{L}$. In [YB4], we provide conditions under which the number of sub-problems can be decreased to reduce the computational burden of our methods.

Proposition 4.4 (Proposition 3 in [YB4])**.** *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 (4.3) 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.

By Proposition 4.4, we only need to consider the sub-problems (4.3) for which the associated deviations are pairwise distinct. In the remainder of this section, we further provide conditions under which no additional lower-level problems need to be solved in Line 7 of Algorithm 4.2. In particular, these conditions are sufficient ex-post conditions for global optimality of heuristically obtained solutions.

Theorem 4.5 (Theorem 2 in [YB4])**.** *Let $(x^\ell)_{\ell \in \mathcal{L}}$ be a family of solutions to the deterministic min-max problems (4.3) 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 (4.1) and $v_{rob} = v_k$ holds.*

Theorem 4.5 indicates that there may be situations in which the Bertsimas–Sim result extends to the min-max setting, i.e., Problem (4.1) can be solved by only solving problems of the nominal type. However, the result does not carry over completely as the requirements of Theorem 4.5 can only be checked ex post, i.e., after solving the deterministic min-max problems (4.3). We emphasize that, in Theorem 4.5, we do not make any structural assumptions about the lower-level feasible set $Y(x)$, $x \in X$, except for the follower's variables being binary. However, further sufficient ex-post conditions can be obtained by exploiting the specific properties of the application problem at hand.

Assumption 4.1.

(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 of 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 4.1, the min-max problem (3.1) is an interdiction problem. Let us point out that the assumptions made here are less restrictive than those in Assumption 3.3. In particular, we do not require that the lower-level problem satisfies the downward monotonicity property; cf. Proposition 3.4.

Theorem 4.6 (Theorem 3 in [YB4])**.** *Suppose that Assumption 4.1 holds. Let $(x^\ell)_{\ell \in \mathcal{L}}$ be a family of solutions to the deterministic min-max problems (4.3) 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 (4.1) and $v_{rob} = v_k$ holds.*

Theorem 4.6 states that any feasible leader's decision $x \in X$ that dominates the solutions to all deterministic min-max problems (4.3) is an optimal solution to Problem (4.1).

4.1.4 Computational Results

To assess the performance of the heuristics presented in this section, we consider the bilevel knapsack interdiction problem, which, in its deterministic form, has been studied in Caprara et al. (2016). For the computational study in [YB4], we have generated a total of 1120 robustified knapsack interdiction instances. Our test set includes 560 instances with deviations taking uniformly distributed integer values and 560 instances with deviations taking continuous, uniformly distributed values. All details regarding the generation of the test instances and the computational setup can be found in Section 5 in [YB4].

In Algorithms 4.1 and 4.2, any solver for mixed-integer linear min-max problems can be used for the solution of the deterministic bilevel problems (4.3). In [YB4], we consider the following two solver options:

(i) the problem-tailored branch-and-cut approach presented in Fischetti et al. (2019)[1],

(ii) the bkpsolver (Weninger and Fukasawa 2023), which is publicly available at https://github.com/nwoeanhinnogaehr/bkpsolver.

The heuristic presented in Algorithm 4.1, which incorporates solver options (i) and (ii), is referred to as H-IC and H-BKP, respectively. For Algorithm 4.2, which is a modification of Algorithm 4.1, we add the suffix "-M", i.e., we refer to the two variants of the method as H-IC-M and H-BKP-M. In [YB4], only the results for H-IC-M and H-BKP-M are shown. In this dissertation, we show computational results for both Algorithm 4.1 and Algorithm 4.2. Consequently, the following figures and tables differ slightly from those presented in [YB4]. To establish a benchmark for evaluating our heuristics' performance, we compare our methods to the exact single-leader multi-follower approach presented in [YB3]. The latter, which we refer to as E-MF, has been discussed in detail in Chapter 3 of this dissertation.

In Tables 4.1 and 4.2, we show the number of instances that the considered approaches can tackle, i.e., they find at least a feasible point with finite optimality gap. It can be seen that the heuristics, which incorporate the bkpsolver for the solution of the deterministic bilevel problems, as well as the exact approach E-MF find at least a feasible point with finite gap for all considered instances. The heuristics H-IC and H-IC-M, however, could not compute a finite gap within the time limit of 1 h for some of the instances. An infinite gap is obtained in the case in which the solution of deterministic bilevel problems exceeds the time limit so that the upper bound, initially being set to infinity, is not updated. Based on Tables 4.1 and 4.2, we further observe that the computational burden of our methods seems to be lower in the setting with integer deviations. The latter is reflected by the number of instances that H-IC/H-IC-M can tackle ("feasible") in the settings with integer and continuous deviations, respectively. As discussed in detail in [YB4], the computational burden of the heuristics is lower in the setting with integer deviations since, just due to the generation of the instances, the result of Proposition 4.4 can be applied more frequently than in the setting with continuous deviations. Taking all previous observations into account leads to one of the main observations of the computational study in [YB4].

Observation 4.7. *The solution of the deterministic bilevel problems is a bottleneck of Algorithms 4.1 and 4.2. Hence, the algorithmic choice for solving these problems is crucial.*

---

[1] To ensure a fair comparison with all other approaches considered in the computational study in [YB4], which use Gurobi, we have re-implemented the method originally developed for CPLEX 12.7 using Gurobi as well. The code is publicly available at https://github.com/YasmineBeck/gamma-robust-bilevel-heuristics.

Table 4.1: 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) and the number of instances with finite but non-zero gap ("open gap") for the approaches H-BKP, H-IC, and E-MF. Additionally, the number of instances solved to global optimality ("optimal"), along with the number of instances satisfying the sufficient optimality condition in Rem. 4.3, is shown.

Δf		feasible	optimal	Rem. 4.3	open gap
integer	H-BKP	560	555	244	5
	H-IC	529	525	230	4
	E-MF	560	526	–	34
continuous	H-BKP	560	554	268	6
	H-IC	499	494	240	5
	E-MF	560	524	–	36

Table 4.2: 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) and the number of instances with finite but non-zero gap ("open gap") for the approaches H-BKP-M, H-IC-M, 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. 4.5 or Thm. 4.6), is shown; cf. Table 5 in [YB4].

Δf		feasible	optimal	Thm. 4.5	Thm. 4.6	open gap
integer	H-BKP-M	560	555	340	340	5
	H-IC-M	517	513	277	315	4
	E-MF	560	526	–	–	34
continuous	H-BKP-M	560	554	359	359	6
	H-IC-M	481	476	266	309	5
	E-MF	560	524	–	–	36

In Tables 4.1 and 4.2, we particularly show the number of instances solved to global optimality ("optimal") and the number of instances satisfying sufficient optimality conditions for the considered approaches. We emphasize that E-MF is an exact approach, which solves a single problem of the form given in (4.1) using branch-and-cut. Hence, optimality of a solution obtained from E-MF is proven by a closed gap. The sufficient condition in Remark 4.3 is only applicable to H-IC and H-BKP, whereas the sufficient conditions in Theorems 4.5 and 4.6 are only applicable to H-IC-M and H-BKP-M.

Observation 4.8. *A substantial number of the considered instances either has a closed optimality gap or satisfies one of the sufficient optimality conditions stated in Remark 4.3 and Theorems 4.5 and 4.6.*

We emphasize that those instances that satisfy the requirements of Remark 4.3 can be solved to global optimality without solving all $|\mathcal{L}|$ bilevel sub-problems. Since the solution of the bilevel problems is a bottleneck of our methods, the latter poses a significant advantage of the heuristic presented in Algorithm 4.1 over the one presented in Algorithm 4.2. Note that, using Algorithm 4.2, exactly $|\mathcal{L}|$ deterministic bilevel problems need to be solved. The advantage of Algorithm 4.1 over Algorithm 4.2 is also underlined by the number of instances solved to global optimality ("optimal") by H-IC compared to the number of instances solved to global optimality by H-IC-M. For the variants H-IC-M and H-BKP-M, we further observe that the majority of the considered instances satisfies one of the sufficient conditions in Theorems 4.5 and 4.6 so that a globally optimal solution to Problem (4.1) is obtained by only solving bilevel problems of the nominal type. Moreover, it can be seen that more instances satisfy the sufficient conditions for optimality in Theorems 4.5 and 4.6 when using H-BKP-M instead of H-IC-M. For those instances for which our heuristics only find feasible (but not provably optimal) points with finite optimality gap, our methods still provide favorable results in terms of the solution quality. The largest finite optimality gap we report in [YB4] for our heuristics is 0.18 %, while, for E-MF, the largest gap observed is 10.98 %. Nevertheless, reflected by the number of instances that H-IC and H-IC-M cannot tackle within the time limit of 1 h, we still acknowledge that the computational burden remains a drawback of our methods.

In Figure 4.1, we show box-plots for the sequential and idealized parallel runtimes of our methods. The latter reflect the overall runtime of a solution method provided that there are sufficient capacities available to solve all arising sub-problems in parallel. For further details on idealized parallel runtimes and their computation, we refer to Section 3.4 of this dissertation. In the sequential setting, H-IC and H-IC-M perform slightly better than E-MF w.r.t. running times. Moreover, H-IC seems to have a minor advantage over H-IC-M due to the possibility to terminate early without solving all $|\mathcal{L}|$ deterministic bilevel problems; cf. Remark 4.3. Nevertheless, we observe that H-BKP and H-BKP-M significantly outperform all other methods. When comparing the box-plots of H-IC and H-IC-M in the idealized parallel setting, we can clearly see the potential of parallelization. Note that, for the variants H-IC and H-BKP, only the solution of the lower-level problems solved in Line 7 of Algorithm 4.1 can be parallelized. Since, however, the solution of the deterministic bilevel problems is a bottleneck of our methods, H-IC-M benefits more from parallelization than H-IC. Due to its overall small runtime required for solving bilevel problems, parallelization affects the performance of H-BKP-M only slightly. Overall, we also observe significant speed-up factors in the idealized parallel setting when comparing the runtimes of H-BKP/H-BKP-M to those of H-IC, H-IC-M, as well as the exact approach E-MF. Summarizing the previous discussion leads to the following final main observation of the computational study in [YB4].

Observation 4.9. *When efficient black-box methods such as the* bkpsolver *(Weninger and Fukasawa 2023) are available to tackle the deterministic bilevel problems, our heuristic outperforms the exact branch-and-cut method proposed in [YB3] both in terms of runtimes and solution quality.*

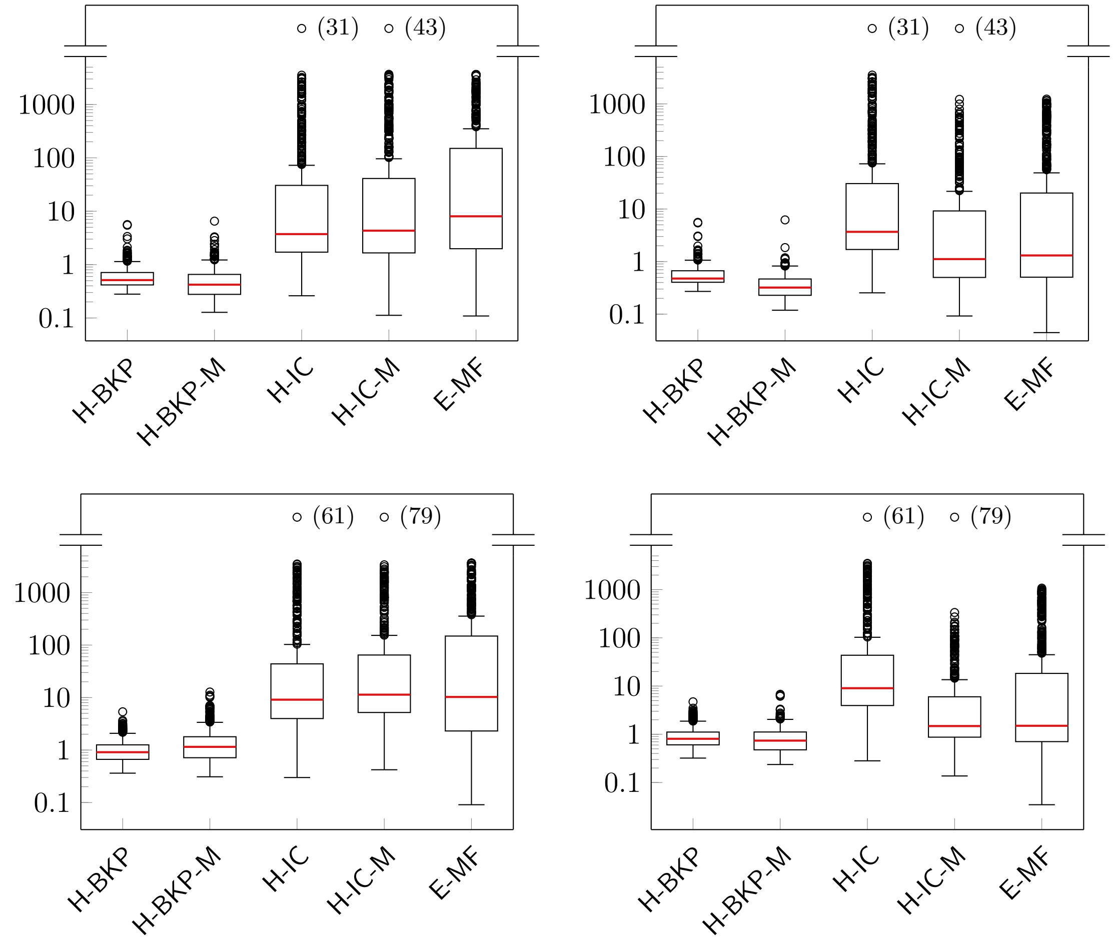

Figure 4.1: Box-plots of the sequential (left) and the idealized parallel runtimes (right) for the approaches H-BKP, H-BKP-M, H-IC, H-IC-M, and E-MF in the min-max setting with integer (top) and continuous deviations (bottom). Runtimes (in s) are depicted on a log-scaled y-axis.

4.2 A Heuristic for General Γ-Robust Bilevel Problems

We now focus on bilevel problems, which, in their deterministic form, are given by

$$\begin{aligned} \min_{x,y} \quad & c^\top x + d^\top y \\ \text{s.t.} \quad & x \in X, \\ & y \in \operatorname*{arg\,max}_{y'} \left\{ f^\top y' \colon y' \in Y(x) \right\}, \end{aligned} \tag{4.4}$$

where $Y(x) \subseteq \{0,1\}^{n_y}$ and $X := \{x \in \mathbb{R}^{n_\mathrm{C}} \times \mathbb{Z}^{n_\mathrm{D}} \colon Ax \geq a\}$ with $n_x = n_\mathrm{C} + n_\mathrm{D}$, $c \in \mathbb{R}^{n_x}$, $d, f \in \mathbb{R}^{n_y}$, $A \in \mathbb{R}^{l \times n_x}$, and $a \in \mathbb{R}^l$. In Problem (4.4), the objective function coefficients for the follower's variables y may differ in the upper- and the lower-level problem. This is in contrast to the min-max setting considered in Section 4.1, where we assume $d = f$.

The Γ-robust counterpart of the bilevel problem (4.4), in which the follower hedges against at most Γ deviations in his uncertain objective function coefficients, is given by

$$\min_{x,y} \quad c^\top x + d^\top y \quad \text{s.t.} \quad x \in X,\ y \in S_\Gamma(x). \tag{4.5}$$

Here, we use $S_\Gamma(x)$ to denote the set of optimal solutions to the x-parameterized Γ-robust counterpart of the lower-level problem; see Problem (3.6). In [YB4], we develop a heuristic for Problem (4.5), which we present in Section 4.2.1. In Section 4.2.2, we illustrate that the setting considered in this section is considerably more challenging than its min-max counterpart; cf. Section 4.1. In addition, we discuss algorithmic refinements for the heuristic presented in Section 4.2.1. Finally, we show computational results in Section 4.2.3.

4.2.1 Presentation of the Heuristic

The heuristic we present in this section builds on a lower bounding scheme that exploits the solution of appropriately chosen bilevel problems of the nominal type. More formally, we have the following result.

Proposition 4.10 (Proposition 5 in [YB4]). *There exists an index $\ell \in \mathcal{L}$ such that the optimal objective function value of the bilevel problem*

$$\begin{aligned} \min_{x,y} \quad & c^\top x + d^\top y \\ s.t. \quad & x \in X, \\ & y \in \operatorname*{arg\,max}_{y' \in Y(x)} \left\{ -\Gamma \Delta f_\ell + \tilde{f}(\ell)^\top y' \right\} \end{aligned} \tag{4.6}$$

yields a valid lower bound for the optimal objective function value of (4.5).

We emphasize that Proposition 4.10 only yields an ex-post result. To determine the sub-problem (4.6) that yields a valid lower bound, we need to know an optimal solution to Problem (4.5) in advance. Nevertheless, we can exploit Proposition 4.10 to obtain an overall valid lower bound for Problem (4.5).

Corollary 4.11 (Corollary 1 in [YB4]). *For all $\ell \in \mathcal{L}$, let (x^ℓ, y^ℓ) be an optimal solution to* (4.6)*. 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 (4.5).

We formally state the heuristic for Problem (4.5) in Algorithm 4.3. In Lines 2 and 3 of the algorithm, we solve $|\mathcal{L}|$ bilevel problems of the nominal type to obtain a valid lower bound for Problem (4.5); cf. Corollary 4.11. We emphasize that any suitable solver can be used for the solution of these problems such as, e.g., the MibS solver (Tahernejad et al. 2020) or the general branch-and-cut solver presented in Fischetti et al. (2017). Moreover, since the sub-problems (4.6) are independent, we can parallelize Lines 2 and 3 of Algorithm 4.3 if the necessary capacities are available. In Line 4, we sort the indices so that the optimal objective function values of the deterministic bilevel problems are given in non-decreasing order. This allows to potentially close the optimality gap more quickly. In Lines 7–9, we perform a correction step to obtain a feasible point and, thus, a valid upper bound for Problem (4.5). This correction step involves solving the Γ-robust counterpart of the lower-level problem, which can be done using Corollary 3.1 or 3.2.

Algorithm 4.3 A Heuristic for Γ-Robust Mixed-Integer Linear Bilevel Problems

```
Input: An instance of Problem (4.5), an exact solution method for Problems (4.4)
    and (3.6), an index set $\mathcal{L}$ as in (4.2)
Output: A feasible pair $(x^*, y^*)$, a lower bound $L$, and an upper bound $U$ for Prob-
    lem (4.5)
 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 $(x^\ell, y^\ell)$ to the bilevel problem
        $$\min_{x,y} \; c^\top x + d^\top y \quad \text{s.t.} \; x \in X, \; y \in \arg\max_{y' \in Y(x)} \left\{ -\Gamma \Delta f_\ell + \tilde{f}(\ell)^\top y' \right\}.$$
 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 $x^{\ell_i}$-parameterized $\Gamma$-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 $(x^*, y^*)$, $L$, $U$
```

Theorem 4.12 (Theorem 4 in [YB4]). *Algorithm 4.3 is correct, i.e., it returns a feasible pair (x^*, y^*) as well as valid lower and upper bounds L and U for (4.5).*

If Algorithm 4.3 terminates with $((x^*, y^*), L, U)$ in Line 6, $U - L = 0$ holds and (x^*, y^*) is an optimal solution to Problem (4.5). More formally, we have the following result.

Proposition 4.13 (Proposition 6 in [YB4]). *Let $(x^\ell, y^\ell)_{\ell \in \mathcal{L}}$ be the family of solutions to the bilevel problems solved in Line 3 of Algorithm 4.3. Further, let $((x^*, y^*), L, U)$ be the output of Algorithm 4.3 and suppose that $c^\top x^* + d^\top y^* \leq c^\top x^\ell + d^\top y^\ell$ holds for all $\ell \in \mathcal{L}$. Then, $U - L = 0$ holds and (x^*, y^*) is an optimal solution to (4.5).*

4.2.2 Challenges and Algorithmic Refinements

Before we discuss algorithmic refinements for the heuristic presented in Algorithm 4.3, we briefly comment on two main reasons why the setting considered in this section is considerably more challenging than the min-max setting of Section 4.1:

(i) Obtaining a valid lower bound for Problem (4.5) is significantly more involved than in the min-max setting; cf. Proposition 4.1 and Corollary 4.11. In particular, Corollary 4.11 implies that the set of deterministic bilevel sub-problems (4.6) needs to be considered holistically, i.e., an iterative refinement of the lower bound for Problem (4.5) such as in Line 4 of Algorithm 4.1 in the min-max setting can, in general, not be obtained.

(ii) In Section 4.1, the Γ-robust counterpart of the lower-level problem is solved to obtain a valid upper bound, while the feasibility of a sub-problem's solution x^ℓ, $\ell \in \mathcal{L}$, for Problem (4.1) is already guaranteed. A solution (x^ℓ, y^ℓ) to (4.6), however, may not be feasible for the Γ-robust bilevel problem (4.5). Hence, we need to perform a correction step to restore feasibility; see Lines 7–9 of Algorithm 4.3.

With respect to (ii), we point out that an optimal solution to the Γ-robust counterpart (3.6) of the lower-level problem may not be unique. However, when solving the problem in Line 7 of Algorithm 4.3, no information about the upper-level objective is used. This means that, if the Γ-robust counterpart (3.6) of the lower-level problem does not have a unique solution, the follower's level of cooperation is not taken into account. In [YB4], we consider the optimistic approach to bilevel optimization. To obtain good upper bounds for the optimal objective function value of Problem (4.5), it may thus be beneficial to include a so-called refinement step to account for the cooperative nature of the optimistic follower. In Algorithm 4.4, we provide a detailed description of the steps involved to solve the Γ-robust counterpart of the lower-level problem (correction step) and to obtain refined upper bounds for Problem (4.5) by solving further binary single-level problems (refinement step). Algorithm 4.4 may be used to replace Line 7 of Algorithm 4.3. For the correction step in Line 3 of Algorithm 4.4, we exploit the result of Corollary 3.2 and Theorem 2.3. In Line 4 of Algorithm 4.4, we solve a refinement problem so that, among multiple optimal responses of the follower, one that favors the leader is chosen. The cooperative nature of the follower is also taken into account in Lines 5 and 6 of Algorithm 4.4 so that we obtain a pair $(\hat{x}, \hat{y})$ that is more likely to correspond to an optimal solution to Problem (4.5).

Proposition 4.14 (Proposition 8 in [YB4])**.** *Let $(x^\ell, y^\ell)_{\ell \in \mathcal{L}}$ be a given family of solutions to the deterministic bilevel sub-problems (4.6). Further, let $\ell_i \in \mathcal{L}$, $i \in \{1, \dots, |\mathcal{L}|\}$, be given arbitrarily. Then, Algorithm 4.4 is correct, i.e., it returns an optimal solution to the x^{ℓ_i}-parameterized Γ-robust counterpart (3.6) of the lower-level problem.*

We now conclude this section with results on the reduction of the overall number of sub-problems to be solved.

Proposition 4.15. *Let $\ell, k \in \mathcal{L}$ with $\ell < k$ and $\Delta f_\ell = \Delta f_k$ be given arbitrarily. Then, for all $\ell \leq i \leq k$, an optimal solution (x^ℓ, y^ℓ) to the ℓth deterministic bilevel sub-problem (4.6) is also an optimal solution to the ith deterministic bilevel sub-problem (4.6), and vice versa.*

Algorithm 4.4 Correct-and-Refine

```
Input: A family of solutions $(x^\ell, y^\ell)_{\ell \in \mathcal{L}}$ to the deterministic bilevel sub-problems (4.6), an index set $\mathcal{L}$ as in (4.2), an index $\ell_i \in \mathcal{L}$
Output: A solution $\hat{y}$ to the $x^{\ell_i}$-parameterized $\Gamma$-robust counterpart (3.6) of the lower-level problem
Set $\hat{x} \leftarrow x^{\ell_i}$.
for $\ell \in \mathcal{L} \setminus \{\ell_i\}$ with $x^\ell \neq \hat{x}$ do
    Correction step: Solve the $\hat{x}$-parameterized $\ell$th lower-level sub-problem
        $$\Phi_\ell(\hat{x}) = -\Gamma \Delta f_\ell + \max_{y \in Y(\hat{x})} \left\{ \tilde{f}(\ell)^\top y \right\}.$$
    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 $y^\ell \leftarrow \hat{y}^\ell$.
Set $\Phi_{\text{rob}}(\hat{x}) \leftarrow \max_{\ell \in \mathcal{L}} \{\Phi_\ell(\hat{x})\}$ and determine $\mathcal{C} := \{\ell \in \mathcal{L} \colon \Phi_\ell(\hat{x}) = \Phi_{\text{rob}}(\hat{x})\}$.
Set $k \leftarrow \arg\min_{\ell \in \mathcal{C}} \left\{ c^\top \hat{x} + d^\top y^\ell \right\}$ and $\hat{y} \leftarrow y^k$.
return $\hat{y}$
```

Proof. From Assumption 2.1, we obtain $\Delta f_\ell = \Delta f_i$ as well as $\tilde{f}(\ell) = \tilde{f}(i)$ for all $\ell \leq i \leq k$. Let (x^ℓ, y^ℓ) be an optimal solution to the ℓth sub-problem (4.6). Then,

$$\begin{aligned} -\Gamma \Delta f_i + \tilde{f}(i)^\top y^\ell &= -\Gamma \Delta f_\ell + \tilde{f}(\ell)^\top y^\ell \\ &= \max_{y \in Y(x^\ell)} \left\{ -\Gamma \Delta f_\ell + \tilde{f}(\ell)^\top y \right\} \\ &= \max_{y \in Y(x^\ell)} \left\{ -\Gamma \Delta f_i + \tilde{f}(i)^\top y \right\} \end{aligned}$$

holds for all $\ell \leq i \leq k$. Thus, and since the upper- and lower-level constraints do not depend on ℓ, the pair (x^ℓ, y^ℓ) is feasible for the ith sub-problem (4.6) with $\ell \leq i \leq k$. This yields

$$v_\ell := c^\top x^\ell + d^\top y^\ell \geq c^\top x^i + d^\top y^i =: v_i,$$

where (x^i, y^i) denotes an optimal solution to the ith sub-problem (4.6) with $\ell \leq i \leq k$. Analogously, (x^i, y^i) is feasible for the ℓth sub-problem (4.6) and, thus, $v_i \geq v_\ell$ holds. As a consequence, we have $v_\ell = v_i$, which concludes the proof. □

By Proposition 4.15, it suffices to only consider the sub-problems for which the associated deviations are pairwise distinct. Finally, we provide a sufficient ex-post condition under which Problem (4.5) can be solved by only solving bilevel problems of the nominal type, i.e., without solving additional lower-level problems.

Theorem 4.16 (Theorem 5 in [YB4]). *Let $(x^\ell, y^\ell)_{\ell \in \mathcal{L}}$ be the family of solutions to the bilevel problems solved in Line 3 of Algorithm 4.3. If there exists an index $k \in \mathcal{L}$ with $x^k = x^\ell$, $d^\top y^k \leq d^\top y^\ell$ and $\tilde{f}(\ell)^\top y^k \geq \tilde{f}(\ell)^\top y^\ell$ for all $\ell \in \mathcal{L}$, (x^k, y^k) is an optimal solution to (4.5).*

Table 4.3: 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) and the number of instances with finite but non-zero gap ("open gap") for the approaches E and H. Additionally, the number of instances solved to global optimality ("optimal"), along with the number of instances satisfying the sufficient optimality condition in Thm. 4.16, is shown; see Table 6 in [YB4].

Δf		feasible	optimal	Thm. 4.16	open gap
integer	H	249	186	70	63
	E	480	236	–	244
continuous	H	217	172	58	45
	E	474	230	–	244

4.2.3 Computational Results

To computationally assess the performance of the presented heuristic, we consider problems that generalize the knapsack interdiction problem studied in Caprara et al. (2016). In its deterministic form, the so-called *generalized knapsack interdiction problem* reads

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

with $B, C \in \mathbb{Z}_{\geq 0}$, and $c, d, f, v, w \in \mathbb{Z}^n_{\geq 0}$. For the computational study in [YB4], a total of 1120 robustified instances of the generalized knapsack interdiction problem has been generated. These include 560 instances with deviations taking uniformly distributed integer values and 560 instances with deviations taking continuous, uniformly distributed values. For detailed information regarding the generation of the test instances and the computational setup, we refer to Section 5 in [YB4].

In Line 3 of Algorithm 4.3, any solver for mixed-integer linear bilevel problems can be used. In [YB4], we use the problem-tailored branch-and-cut approach outlined in Section 3.5.1. Preliminary computational tests revealed that this 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 branch-and-cut solver presented in Fischetti et al. (2017). In what follows, we abbreviate the heuristic by H. We further mention that, to the best of our knowledge, there is no other method in the literature that can directly tackle general mixed-integer bilevel problems with a Γ-robust treatment of lower-level objective uncertainty, neither globally nor heuristically. To provide a point of reference for comparison, we have thus implemented an exact branch-and-cut approach tailored to our setting. We refer to this method, which is outlined in Section 3.5.2, as E.

In Table 4.3, we show the number of instances that both E and H can tackle, i.e., they find at least a feasible point with finite optimality gap. We emphasize that E is an exact approach that solves a single problem of the form given in (3.26) using branch-and-cut. This means that optimality of a solution obtained from E is proven by a closed gap. Based on Table 4.3, we observe that the heuristic could not compute a finite gap within the time limit of 1 h for a significant portion of the considered instances. An infinite gap is obtained if the solution of the deterministic bilevel problems exceeds the time limit so that the upper bound, initially being set to infinity, is not updated. Reflected by the results shown in Table 4.3, we acknowledge that the computational burden of our heuristic is quite large. Nevertheless, it can be seen that, for the majority of the instances that the heuristic can tackle, global optimality is proven either using the sufficient ex-post condition in Theorem 4.16 or by a closed optimality gap (cf. Proposition 4.13). In particular, for the 70 and 58 instances in the settings with integer and continuous deviations that satisfy the requirements of Theorem 4.16, optimality is proven by only solving bilevel problems of the nominal type. To further assess the quality of heuristically obtained solutions, we show ECDF plots w.r.t. the optimality gaps in Figure 4.2. The ECDFs can be interpreted as the percentage of instances (y-axis) for which a certain optimality gap (x-axis) can be achieved. Here, we exclude 80 instances with integer deviations and 86 instances with continuous deviations that neither E nor H can tackle. Figure 4.2 shows that we obtain an optimality gap of at most 8.85 % and 11.35 % (the values from which on the curves stay horizontal) for those instances that can be tackled using the heuristic in the settings with integer and continuous deviations, respectively. The largest gaps we observe for E are 57.63 % and 58.10 % in the settings with integer and continuous deviations, respectively.

In [YB4], we further observe that the heuristic H is faster than the exact solution approach E on the instances that can be tackled. This is particularly true if the solution of the deterministic bilevel problems in Line 3 of Algorithm 4.3 and, if necessary, the solution of the additional lower-level problems in Line 7, can be parallelized. Moreover, the computational burden of the heuristic can significantly be reduced by applying the result of Proposition 4.15. In [YB4], we observe that, just due to the generation of the instances, a considerable number of deterministic bilevel problems can be eliminated in the setting with integer deviations. Nevertheless, the large portion of instances for which the heuristic cannot compute a finite gap still reflects the significant computational burden of our method. Solving deterministic bilevel problems remains a bottleneck of Algorithm 4.3 and, thus, the algorithmic choice for solving these problems is crucial.

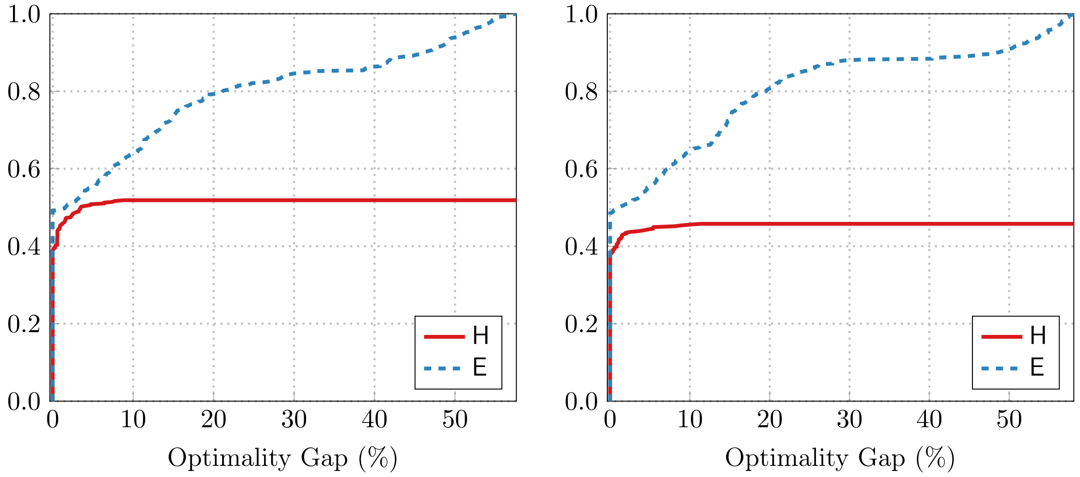

Figure 4.2: ECDF plots of the optimality gaps (in %) for the approaches in the general bilevel setting with integer (left) and continuous deviations (right).

5. A Toll-Setting Problem with Robust Wardrop Equilibrium Conditions

Bilevel optimization is a powerful tool for modeling hierarchical decision-making processes, which frequently arise in the context of transportation; see, e.g., Ben-Ayed et al. (1992), Brotcorne et al. (2001), Dempe and Zemkoho (2012), Dewez et al. (2008), Kalashnikov et al. (2020), Labbé et al. (1998, 2000), and Migdalas (1995). In this chapter, we summarize the contributions of [YB5], which address one important aspect of transportation science: the problem of determining optimal tolls in a traffic network. We consider the setting in which a toll-setting authority aims to maximize revenues by imposing tolls on certain arcs of the traffic network. These revenues can, e.g., be used to support the maintenance of existing infrastructure or fund the construction of new roads. Regarding the users of the traffic network, we assume that they act according to Wardrop's user equilibrium (Wardrop 1952; Wardrop and Whitehead 1952), minimizing their individual travel costs that are parameterized by the imposed tolls. Overall, this renders the considered toll-setting problem a mathematical problem with equilibrium constraints (MPEC); see, e.g., Luo et al. (1996) for a general overview. In Section 5.1, we formally state the toll-setting problem, reformulate it as a mixed-integer nonlinear problem (MINLP), and provide valid inequalities. Afterward, in Section 5.2, we consider a robustified variant of the toll-setting problem in which the users of the traffic network hedge against uncertain travel costs. We also present an MINLP reformulation of the toll-setting problem in this setting. In Section 5.3, we elaborate on the impact of considering robust travel decisions on the computational difficulty and on the revenues realized by the toll-setting authority.

5.1 A Deterministic Toll-Setting Problem

In [YB5], we consider a traffic network that is modeled using a directed graph $\mathcal{G} = (\mathcal{N}, \mathcal{A})$ with node set $\mathcal{N}$ and arc set $\mathcal{A} \subseteq \mathcal{N} \times \mathcal{N}$. In what follows, we denote $f = (f_a)_{a \in \mathcal{A}}$ as the vector of all arc flows and $\tau = (\tau_a)_{a \in \mathcal{A}}$ as the vector of tolls imposed on the arcs of the network. The aim of the toll-setting authority is to maximize the revenues realized by imposing tolls on certain arcs of the network. The overall toll-setting problem is given by

$$\max_{\tau, f, x} \quad \sum_{a \in \mathcal{A}} \tau_a f_a \quad \text{s.t.} \quad \tau \in \mathcal{T}, \ (f, x) \in S(\tau). \tag{5.1}$$

Here, $S(\tau)$ denotes the set of Wardrop equilibria that are parameterized by the tolls τ. The set $\mathcal{T}$ is used to model lower and upper bounds on the tolls or toll-free arcs. More formally, we impose the following.

Assumption 5.1. *The set $\mathcal{T}$ induces a finite upper bound τ_a^+ as well as a lower bound of zero for the toll τ_a on each arc $a \in \mathcal{A}$.*

All arcs $a \in \mathcal{A}$ for which the set $\mathcal{T}$ imposes the upper bound $\tau_a^+ = 0$ are called toll-free arcs. The remaining arcs are called toll arcs. Before we elaborate on the set of Wardrop equilibria $S(\tau)$, let us mention that Problem (5.1) can be interpreted as a single-leader multi-follower problem. Here, the toll-setting authority acts as the leader and the users of the traffic network act as the followers. In this context, optimizing over the tolls τ as well as over the variables f and x particularly relates to the optimistic approach to bilevel optimization. This means that, whenever there are multiple optimal route choices for the network users, they choose the ones that favor the leader the most w.r.t. the associated revenues. The latter is a common assumption in the literature; see, e.g., Brotcorne et al. (2001) and Labbé et al. (1998).

5.1.1 Wardrop Equilibrium Conditions

For node subsets $\mathcal{O}, \mathcal{D} \subseteq \mathcal{N}$, we denote the set of all origin-destination (OD) pairs of the network as $\mathcal{K} \subseteq \mathcal{O} \times \mathcal{D}$. We consider a single commodity for each OD pair $k \in \mathcal{K}$ and use $x^k = (x_a^k)_{a \in \mathcal{A}} \in \mathbb{R}^{|\mathcal{A}|}$ to denote the flow vector of commodity k. The vector of arc flows is then given by

$$f = \sum_{k \in \mathcal{K}} x^k \in \mathbb{R}^{|\mathcal{A}|}. \tag{5.2}$$

Throughout this chapter, we make the following assumptions regarding the connectivity of the traffic network and the travel demand of its users.

Assumption 5.2. *For every node $i \in \mathcal{N}$, there is at least one path that connects node i to each destination node $j \in \mathcal{D}$.*

Assumption 5.3. *For every commodity $k \in \mathcal{K}$, the travel demand $d_k \in \mathbb{R}$ is non-negative and fixed.*

Assumption 5.2 is a standard assumption in the context of traffic assignment; cf., e.g., Assumption 2.A in Patriksson (2015). Moreover, we can impose Assumption 5.3 w.l.o.g. as any elastic-demand problem can equivalently be reformulated as a fixed-demand problem; see, e.g., Dantzig et al. (1976) and Gartner (1980). For each commodity $k = (\alpha_k, \omega_k) \in \mathcal{K}$, flow conservation can now be modeled via

$$\sum_{a \in \delta^{\text{in}}(i)} x_a^k - \sum_{a \in \delta^{\text{out}}(i)} x_a^k = d_i^k, \quad i \in \mathcal{N}, \tag{5.3}$$

with

$$d_i^k = \begin{cases} +d_k, & i = \omega_k, \\ 0, & i \in \mathcal{N} \setminus \{\alpha_k, \omega_k\}, \\ -d_k, & i = \alpha_k. \end{cases}$$

Here, $\delta^{\text{in}}(i)$ and $\delta^{\text{out}}(i)$ denote the sets of in- and outgoing arcs of node $i \in \mathcal{N}$, respectively. In [YB5], we consider the setting in which network users act according to Wardrop's user equilibrium, minimizing their individual travel costs. This behavior can be modeled as

$$0 \leq c_a^k(f; \tau_a) + t_j^k - t_i^k \perp x_a^k \geq 0, \quad a = (i,j) \in \mathcal{A}, \ k \in \mathcal{K}. \tag{5.4}$$

In (5.4), the cost for commodity $k \in \mathcal{K}$ to travel along an arc $a \in \mathcal{A}$ is given by the function $c_a^k(f;\tau_a)$ that depends on the overall flows f and that is parameterized by the imposed toll τ_a. Moreover, t_i^k denotes the minimum cost to reach the destination of commodity $k \in \mathcal{K}$ from node $i \in \mathcal{N}$. We abbreviate $t = (t^k)_{k\in\mathcal{K}}$ with $t^k = (t_i^k)_{i\in\mathcal{N}} \in \mathbb{R}^{|\mathcal{N}|}$. To sum up, the τ-parameterized set of Wardrop equilibria is given by

$$S(\tau) := \{(f,x) \colon \exists t \text{ such that } (f,x,t) \text{ solves (5.2)–(5.4)}\}.$$

5.1.2 An MINLP Reformulation

In [YB5], we introduce additional binary variables $z \in \{0,1\}^{|\mathcal{A}|\cdot|\mathcal{K}|}$ and sufficiently large big-M constants to reformulate the toll-setting problem (5.1) as the problem

$$\max_{\tau,f,x,t,z} \quad \sum_{a\in\mathcal{A}} \tau_a f_a \tag{5.5a}$$

$$\text{s.t.} \quad \tau \in \mathcal{T}, \quad f = \sum_{k\in\mathcal{K}} x^k, \tag{5.5b}$$

$$\sum_{a\in\delta^{\text{in}}(i)} x_a^k - \sum_{a\in\delta^{\text{out}}(i)} x_a^k = d_i^k, \qquad i \in \mathcal{N},\ k \in \mathcal{K}, \tag{5.5c}$$

$$x_a^k \ge 0, \quad c_a^k(f;\tau_a) + t_j^k - t_i^k \ge 0, \qquad a = (i,j) \in \mathcal{A},\ k \in \mathcal{K}, \tag{5.5d}$$

$$c_a^k(f;\tau_a) + t_j^k - t_i^k \le M_a^k(1 - z_a^k), \qquad a = (i,j) \in \mathcal{A},\ k \in \mathcal{K}, \tag{5.5e}$$

$$x_a^k \le M_a^k z_a^k, \qquad a \in \mathcal{A},\ k \in \mathcal{K}, \tag{5.5f}$$

$$z_a^k \in \{0,1\}, \qquad a \in \mathcal{A},\ k \in \mathcal{K}. \tag{5.5g}$$

Problem (5.5) is as an MINLP due to bilinearities in the objective function and possible nonlinearities in the travel cost functions $c_a^k(f;\tau_a)$, $a \in \mathcal{A}$, $k \in \mathcal{K}$. By construction, Problem (5.5) is equivalent to the toll-setting problem (5.1) if the big-M constants M_a^k, $a \in \mathcal{A}$, $k \in \mathcal{K}$, are sufficiently large. However, to obtain such constants, further knowledge about the travel cost functions $c_a^k(f;\tau_a)$ is needed. To this end, we assume in [YB5] that the travel costs are affine-linear in the flows f. For the remainder of this chapter, we thus impose the following.

Assumption 5.4. *For every commodity $k \in \mathcal{K}$, the travel cost functions $c^k(f;\tau) = \left(c_a^k(f;\tau_a)\right)_{a\in\mathcal{A}}$ are affine-linear in the flows, i.e., there exists a matrix $C^k \in \mathbb{R}_{\ge 0}^{|\mathcal{A}|\times|\mathcal{A}|}$ and a vector $c^{\mathit{fix},k} \in \mathbb{R}_{>0}^{|\mathcal{A}|}$ with $c^k(f;\tau) = C^k f + c^{\mathit{fix},k} + \tau$.*

We acknowledge that Assumption 5.4 is rather strong. Even under this simplifying assumption, however, solving the toll-setting problem computationally is a highly challenging task. We elaborate on this in more detail in Section 5.3. In [YB5], we exploit Assumptions 5.1–5.4 to derive bounds for the variables f, x, and t, from which sufficiently large big-Ms for Problem (5.5) can be obtained. Detailed derivations can be found in Section 3.1 in [YB5]. In addition to determining correct big-M constants, we further exploit the derived variable bounds for proving the existence of solutions to the toll-setting problem (5.1). More formally, we have the following theorem. A proof of this theorem, along with all proofs that we omit in this chapter, can be found in [YB5].

Theorem 5.1 (Theorem 1 in [YB5]). *Under Assumptions 5.1–5.4, the toll-setting problem (5.1) has an optimal solution (τ, f, x).*

5.1.3 Valid Inequalities

In [YB5], we derive valid inequalities for the feasible set and optimal solutions of Problem (5.5). Preliminary computational tests revealed that including the following inequalities significantly enhances the solution process.

Proposition 5.2 (Proposition 4 in [YB5])**.** *Let* $\tau \in \mathcal{T}$ *be given arbitrarily. Further, let* $i, j \in \mathcal{N}$ *be such that* $(i, j), (j, i) \in \mathcal{A}$ *holds. Then, under Assumptions 5.1 and 5.4, the inequalities*

$$z^k_{(i,j)} + z^k_{(j,i)} \leq 1, \quad k \in \mathcal{K},$$

are valid for the feasible set of Problem (5.5)*.*

Proposition 5.2 implies that, under the assumption of positive travel costs, there cannot be positive commodity flow on both an arc and its reversed arc.

Proposition 5.3 (Proposition 5 in [YB5])**.** *Under Assumptions 5.1–5.4, there exists an optimal solution* (τ, f, x, t, z) *to Problem* (5.5) *that satisfies*

$$\tau_a \geq \tau^+_a \left(1 - \sum_{k \in \mathcal{K}} z^k_a \right), \quad a \in \mathcal{A}.$$

Proposition 5.3 is used to fix a toll τ_a, $a \in \mathcal{A}$, to its upper bound τ^+_a in the case that $z^k_a = 0$ holds for all commodities $k \in \mathcal{K}$. The latter implies that there is no flow on arc a and, thus, no revenues are generated by imposing tolls on this arc.

5.2 Robust Toll-Setting Under Budgeted Uncertainty

In Section 5.1, we have considered the setting in which the network users act under perfect information. In real-world applications, however, travelers often face uncertainties when making their decisions. For instance, the travel costs may be subject to uncertainty due to unforeseen events such as accidents, maintenance work, or changing weather conditions. Hence, the assumption of perfect information seems to be rather strong. In [YB5], we account for uncertainties regarding the travel costs by considering so-called robust Wardrop equilibria under budgeted uncertainty, which we discuss in the following. In the context of traffic assignment, robust Wardrop equilibria have also been studied in Ito (2011) and Ordóñez and Stier-Moses (2007, 2010).

5.2.1 Robust Wardrop Equilibrium Conditions

We start from the nominal Wardrop equilibrium model given by Conditions (5.2)–(5.4), for which we now assume that the travel costs of each arc $a \in \mathcal{A}$ and each commodity $k \in \mathcal{K}$ are not known exactly. More formally, we impose the following.

Assumption 5.5. *For all* $a \in \mathcal{A}$ *and* $k \in \mathcal{K}$*, the travel costs* $c^k_a(f; \tau_a)$ *are subject to additive deviations* $Y^k_a \Delta c^k_a$ *with* Y^k_a *being a random variable with support in* $[0, 1]$ *and* $\Delta c^k_a \geq 0$*.*

The parameters $\Delta c_a^k \geq 0$ denote upper bounds on the possible deviation from the nominal travel costs. Since it is unlikely that the costs realize in a worst-case sense on every arc of the network and, hence, to avoid being overly conservative, we assume that each commodity $k \in \mathcal{K}$ hedges against deviations of up to $\Gamma^k \in \{0, \dots, |\mathcal{A}|\}$. The robustified version of Wardrop's user equilibrium (5.4) then reads

$$0 \leq c_a^k(f; \tau_a) + y_a^k \Delta c_a^k + t_j^k - t_i^k \perp x_a^k \geq 0, \quad a = (i,j) \in \mathcal{A},\ k \in \mathcal{K}. \tag{5.6}$$

Here, for a commodity $k \in \mathcal{K}$ and a given flow vector x^k, the vector y^k solves

$$\begin{aligned} \max_{y^k} \quad & \sum_{a \in \mathcal{A}} (\Delta c_a^k x_a^k) y_a^k && \text{(5.7a)} \\ \text{s.t.} \quad & \sum_{a \in \mathcal{A}} y_a^k \leq \Gamma^k, && \text{(5.7b)} \\ & 0 \leq y_a^k \leq 1, \quad a \in \mathcal{A}. && \text{(5.7c)} \end{aligned}$$

Problem (5.7) is a linear problem for fixed x^k, $k \in \mathcal{K}$. Hence, the KKT conditions are necessary and sufficient optimality conditions, i.e., replacing Problem (5.7) by its KKT conditions yields an equivalent reformulation of (5.6) and (5.7) that is given by

$$\begin{aligned} 0 \leq c_a^k(f; \tau_a) + y_a^k \Delta c_a^k + t_j^k - t_i^k \perp x_a^k \geq 0, & \qquad a = (i,j) \in \mathcal{A},\ k \in \mathcal{K}, && \text{(5.8a)} \\ 0 \leq \xi^k + \zeta_a^k - \Delta c_a^k x_a^k \perp y_a^k \geq 0, & \qquad a \in \mathcal{A},\ k \in \mathcal{K}, && \text{(5.8b)} \\ 0 \leq 1 - y_a^k \perp \zeta_a^k \geq 0, & \qquad a \in \mathcal{A},\ k \in \mathcal{K}, && \text{(5.8c)} \\ 0 \leq \Gamma^k - \sum_{a \in \mathcal{A}} y_a^k \perp \xi^k \geq 0, & \qquad k \in \mathcal{K}. && \text{(5.8d)} \end{aligned}$$

Since Conditions (5.2) and (5.3) do not depend on the (robustified) travel costs, the set of robust Wardrop equilibria for given tolls τ and fixed $\Gamma = (\Gamma^k)_{k \in \mathcal{K}}$ can be stated as

$$S_{\text{rob}}(\tau) = \{(f, x) \colon \exists (t, y, \xi, \zeta) \text{ such that } (f, x, t, y, \xi, \zeta) \text{ solves } (5.2), (5.3), \text{ and } (5.8)\}.$$

The overall robustified toll-setting problem then reads

$$\max_{\tau, f, x} \quad \sum_{a \in \mathcal{A}} \tau_a f_a \quad \text{s.t.} \quad \tau \in \mathcal{T},\ (f, x) \in S_{\text{rob}}(\tau). \tag{5.9}$$

In [YB5], the existence of solutions to Problem (5.9) and, in particular, the existence of robust Wardrop equilibria under budgeted uncertainty is shown using Theorem 5.5 in Aashtiani and Magnanti (1981). More formally, we have the following theorem.

Theorem 5.4 (Theorem 3 in [YB5]). *Under Assumptions 5.1–5.5, the robustified toll-setting problem (5.9) has an optimal solution (τ, f, x).*

5.2.2 An MINLP Reformulation

Similar as it is done in Section 5.1.2, we exploit sufficiently large big-M constants and additional binary variables to linearize the complementarity constraints in the robustified version of Wardrop's user equilibrium (5.8). After some further reformulations (cf. Lemma 2 and the respective discussion in Section 4.2 in [YB5]), an equivalent MINLP reformulation of the robustified toll-setting problem (5.9) is then given by

$$
\begin{aligned}
\max_{\tau,f,x,t,r} \quad & \sum_{a\in\mathcal{A}} \tau_a f_a && && (5.10a)\\
\text{s.t.} \quad & \tau \in \mathcal{T}, \quad f = \sum_{k\in\mathcal{K}} x^k, && && (5.10b)\\
& \sum_{a\in\delta^{\text{in}}(i)} x_a^k - \sum_{a\in\delta^{\text{out}}(i)} x_a^k = d_i^k, && i\in\mathcal{N},\ k\in\mathcal{K}, && (5.10c)\\
& \tilde{c}_a^k(f;\tau_a) + t_j^k - t_i^k \geq 0, && a=(i,j)\in\mathcal{A},\ k\in\mathcal{K}, && (5.10d)\\
& \tilde{c}_a^k(f;\tau_a) + t_j^k - t_i^k \leq M_a^k(1-z_a^k), && a=(i,j)\in\mathcal{A},\ k\in\mathcal{K}, && (5.10e)\\
& x_a^k \geq 0, \quad x_a^k \leq M_a^k z_a^k, && a\in\mathcal{A},\ k\in\mathcal{K}, && (5.10f)\\
& \xi^k + \zeta_a^k - \Delta c_a^k x_a^k \geq 0, && a\in\mathcal{A},\ k\in\mathcal{K}, && (5.10g)\\
& \xi^k + \zeta_a^k - \Delta c_a^k x_a^k \leq N_a^k w_a^k, && a\in\mathcal{A},\ k\in\mathcal{K}, && (5.10h)\\
& y_a^k \geq 0, \quad y_a^k \leq 1 - w_a^k, && a\in\mathcal{A},\ k\in\mathcal{K}, && (5.10i)\\
& y_a^k \leq 1, \quad y_a^k \geq v_a^k, && a\in\mathcal{A},\ k\in\mathcal{K}, && (5.10j)\\
& \zeta_a^k \geq 0, \quad \zeta_a^k \leq L_a^k v_a^k, && a\in\mathcal{A},\ k\in\mathcal{K}, && (5.10k)\\
& \xi^k \geq 0, \quad \xi^k \leq R^k, && k\in\mathcal{K}, && (5.10l)\\
& \sum_{a\in\mathcal{A}} y_a^k = \Gamma^k, && k\in\mathcal{K}, && (5.10m)\\
& v_a^k,\ w_a^k,\ z_a^k \in \{0,1\}, && a\in\mathcal{A},\ k\in\mathcal{K}, && (5.10n)
\end{aligned}
$$

for sufficiently large constants L_a^k, M_a^k, N_a^k, and R^k for all $a \in \mathcal{A}$ and $k \in \mathcal{K}$. Here, we use $\tilde{c}_a^k(f;\tau_a) := c_a^k(f;\tau_a) + y_a^k \Delta c_a^k$ to denote the robustified travel costs of commodity $k \in \mathcal{K}$ on arc $a \in \mathcal{A}$. In addition, we abbreviate $r := (y, \xi, \zeta, q, v, w, z)$. Under Assumptions 5.1–5.5, sufficiently large constants L_a^k, M_a^k, N_a^k, and R^k, $a \in \mathcal{A}$, $k \in \mathcal{K}$, can be obtained from the variable bounds provided in Section 4.3 in [YB5].

In Problem (5.10), we introduce $|\mathcal{K}|(8|\mathcal{A}|+3)$ additional constraints and $|\mathcal{K}|(4|\mathcal{A}|+1)$ additional variables for the robustification and the linearization of robustified constraints. Problem (5.10) is thus significantly larger than Problem (5.5) in terms of the number of variables and constraints. Therefore, it seems reasonable to aim for a more compact formulation of the toll-setting problem that maintains a similar level of flexibility regarding the conservatism of robust travel decisions. In this context, we emphasize that the budgeted uncertainty modeling considered in Problem (5.7) is closely related to the notion of Γ-robustness according to Bertsimas and Sim (2003) and Sim (2004). In our setting, pursuing a Γ-robust approach would imply that the users of the traffic network hedge against uncertain travel costs on at most Γ^k many arcs of the network. This requires imposing integrality on the variables y^k in Problem (5.10). In principle, this integrality can be exploited to significantly reduce the number of variables and constraints in Problem (5.10). However, we emphasize that integrality of y^k leads to robustified travel cost functions $\tilde{c}_a^k(f;\tau_a) := c_a^k(f;\tau_a) + y_a^k \Delta c_a^k$ that are no longer continuous. Continuity is necessary to prove the existence of robust Wardrop equilibria using Theorem 5.5 in Aashtiani and Magnanti (1981). Hence, proving existence of Γ-robust Wardrop equilibria in the sense of Bertsimas and Sim (2003) and Sim (2004) most likely requires different techniques compared to those used in the proof of Theorem 5.4.

5.2.3 Valid Inequalities

Under Assumptions 5.1, 5.4, and 5.5, the robustified travel costs $\tilde{c}_a^k(f;\tau_a)$ are positive for all $a \in \mathcal{A}$ and $k \in \mathcal{K}$. Hence, the inequalities provided in Proposition 5.2 are also valid for Problem (5.10). Moreover, no information about the travel costs is used to derive the inequalities presented in Proposition 5.3. Hence, the latter are valid for optimal solutions to the robustified toll-setting problem as well. To conclude, we provide additional valid inequalities for the feasible set of Problem (5.10).

Proposition 5.5 (Proposition 6 in [YB5]). *The inequalities*

$$v_a^k + w_a^k \leq 1, \quad a \in \mathcal{A}, \ k \in \mathcal{K},$$

are valid for the feasible set of Problem (5.10).

5.3 Case Study

In [YB5], we illustrate how the consideration of robust Wardrop equilibria may impact toll-setting policies through a case study on a subnetwork of the well-known Sioux Falls network (LeBlanc et al. 1975).[1] The subnetwork considered in the case study in [YB5], which we refer to as “Sioux Falls East”, consists of 12 nodes and 36 arcs. Moreover, we vary the number of toll arcs and OD pairs between 4 and 8, respectively. We emphasize that Problems (5.5) and (5.10) are nonconvex MINLPs, which can be tackled using state-of-the-art general-purpose solvers such as, e.g., Gurobi. For detailed information regarding the considered network instances and the computational setup, we refer to Section 5 in [YB5]. We now summarize the main observations of the case study in [YB5].

Observation 5.6. *Increasing the number of OD pairs and toll arcs contributes to an increased computational difficulty and an increased amount of resources required to solve the respective toll-setting problems.*

The number of OD pairs directly influences the size of the toll-setting problem, which represents a considerable challenge when solving these problems computationally. Since the robustification of uncertain travel costs leads to significantly larger models than in the nominal setting, these computational challenges are particularly pronounced for the robustified toll-setting problem. In addition, more toll arcs lead to more nonconvex terms in the objective function of the toll-setting problem, which require a special algorithmic treatment. General-purpose solvers such as Gurobi tackle these nonconvexities using spatial branching based on convex envelopes. Overall, solving Problems (5.5) and (5.10) is thus a highly challenging task. This is also reflected in the size of the instances solved in [YB5]. For the nominal toll-setting problem, the largest instances of the considered “Sioux Falls East” network, which can be tackled using Gurobi within 1 h, include up to 8 OD pairs with up to 6 toll arcs. In the robust setting, only the smallest of the considered instances, i.e., the one with 4 OD pairs and 4 toll arcs, can be solved within 1 h. In

---

[1] The instance data is publicly available at `https://github.com/bstabler/TransportationNetworks`.

Figures 5.1 and 5.2, we show the "Sioux Falls East" network with 4 OD pairs and 4 toll arcs, along with the flows in an equilibrium and the imposed tolls for the nominal and the robust setting, respectively. Based on Figure 5.1, it can be seen that revenues are only generated by imposing tolls on arc $(15, 22)$ in the nominal setting. In the robust setting depicted in Figure 5.2, revenues are additionally generated by imposing tolls on the arcs that connect nodes 16 and 17. In [YB5], we make the following overall observation.

Observation 5.7. *For the "Sioux Falls East" network with* 4 *OD pairs and* 4 *toll arcs depicted in Figures 5.1 and 5.2, the revenues generated by imposing tolls are significantly higher in the robust setting compared to the nominal one. Nevertheless, network users always face increased travel costs to reach their destination when hedging against uncertainties in a robust way.*

We emphasize that it is not the toll-setting authority that hedges against uncertain travel costs, but the users of the traffic network. In particular, the network users decide on their route choices in a "here-and-now" fashion, i.e., before the uncertainty realizes. Viewing the overall toll-setting problem as a single-leader multi-follower problem, this means that we consider multiple "here-and-now" followers. Since this problem is considered from the leader's perspective, having higher revenues in the robust setting is thus not in contrast to classic robust optimization theory. For given tolls $\tau \in \mathcal{T}$, the previous observations indicate that neither $S_{\text{rob}}(\tau) \subseteq S(\tau)$ nor $S_{\text{rob}}(\tau) \supseteq S(\tau)$ holds in general. Regarding the behavior of robustified network users, we make the following final observation in [YB5].

Observation 5.8. *Users of the traffic network, who hedge against uncertain travel cost in a robust way, may be indifferent to uncertainties, change their travel decisions completely, or decide on something in between.*

For the "Sioux Falls East" network with 4 OD pairs and 4 toll arcs depicted in Figures 5.1 and 5.2, we observe that the green and the orange commodities, i.e., OD pairs $(9, 21)$ and $(8, 20)$, do not change their travel decisions when hedging against uncertain travel costs in a robust way. However, the travel decision of the purple commodity, i.e., OD pair $(16, 21)$, changes completely in the robust setting. Instead of taking a toll-free detour, it now takes the most direct route, which includes a toll arc. Moreover, compared to the nominal setting, the flow of the blue commodity, i.e., OD pair $(17, 22)$, is divided between the most direct tolled route and toll-free detours in the robust setting.

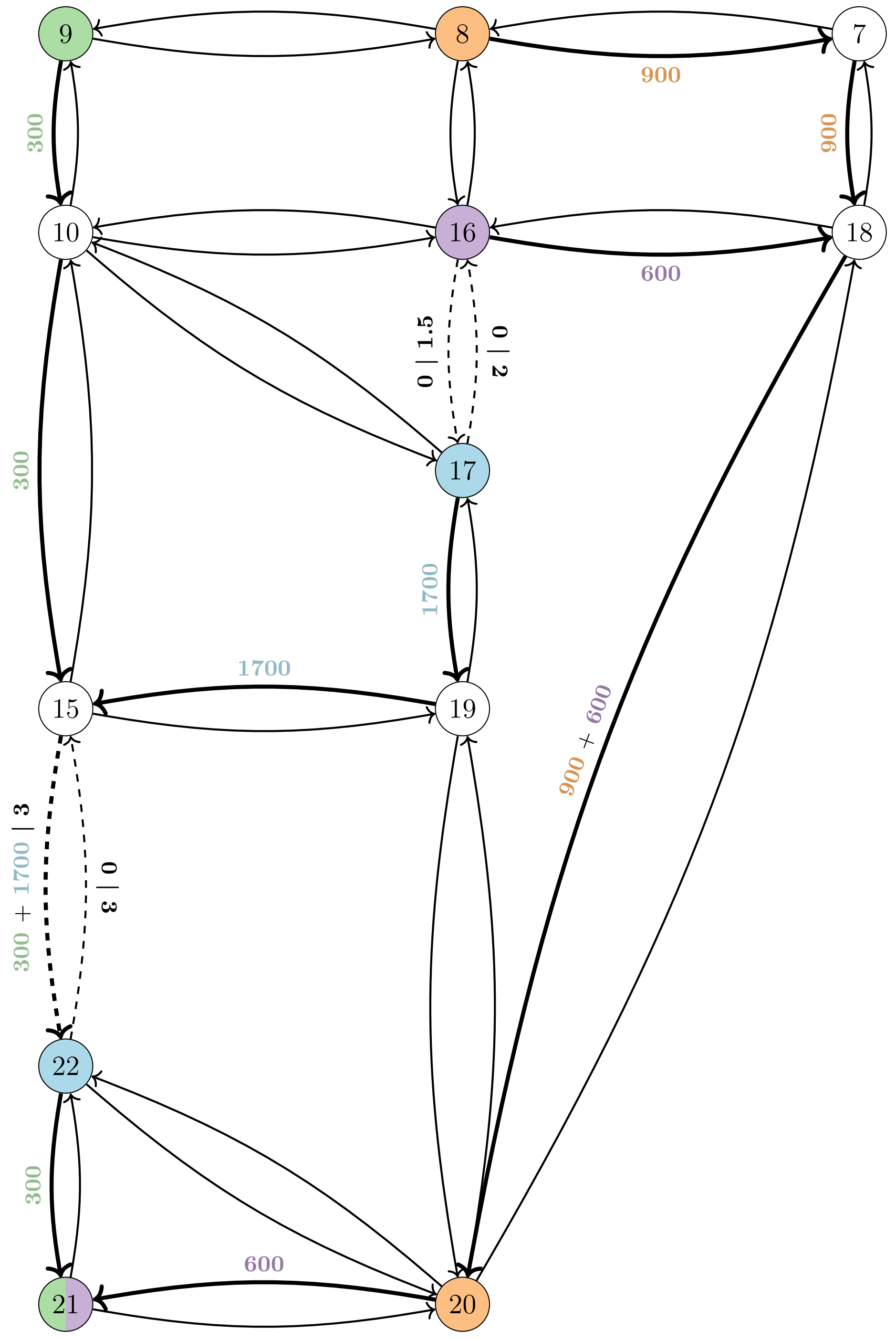

Figure 5.1: The "Sioux Falls East" network with 4 OD pairs and 4 toll arcs. Each OD pair is color-coded (orange, green, blue, purple). Dashed arcs represent toll arcs and solid arcs represent toll-free arcs. Edge labels correspond to commodity flows. For toll arcs, edge labels are given in the format "flow | toll". If no label is shown, there is no flow on that edge. See Figure 2 in [YB5].

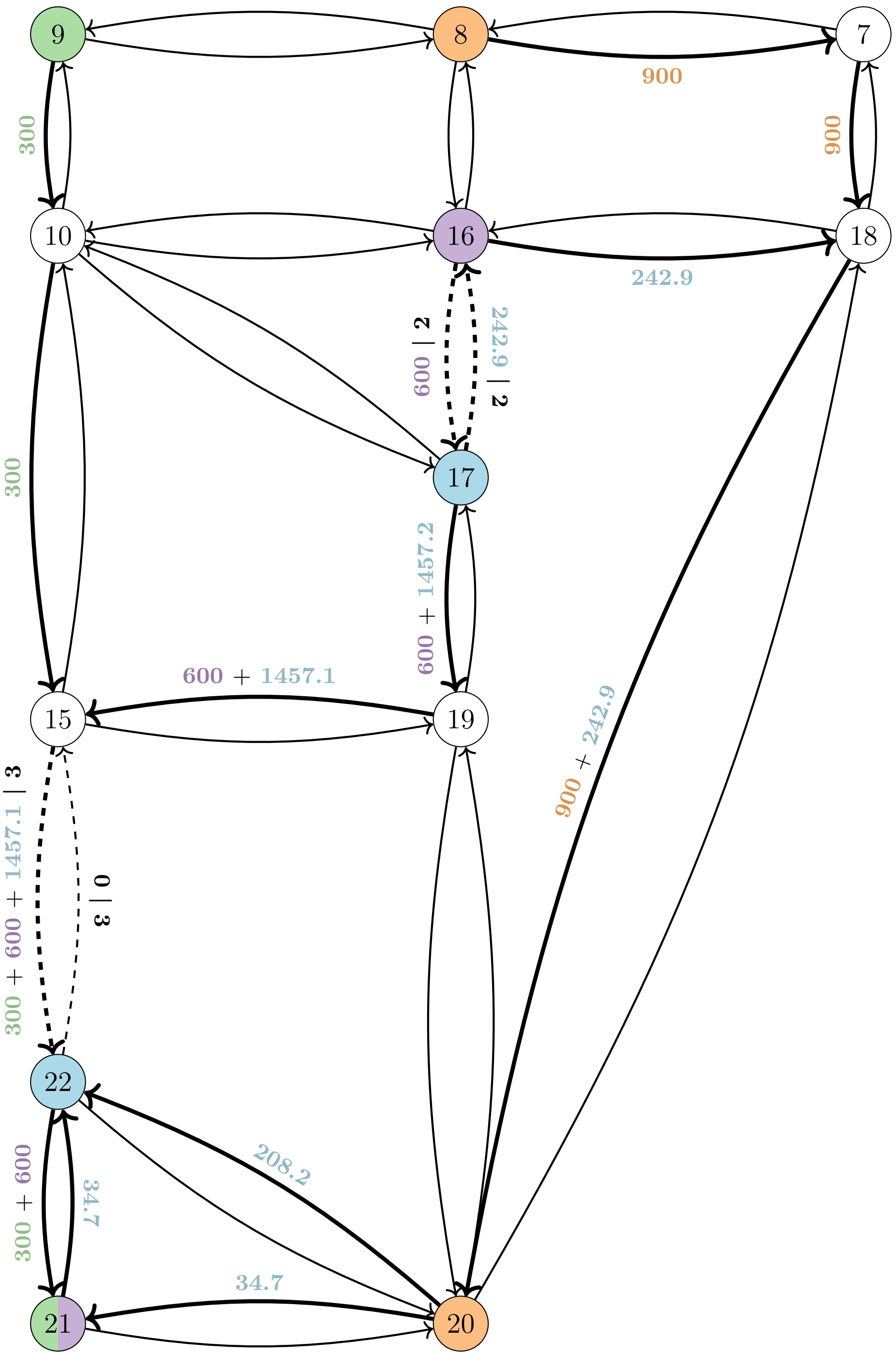

Figure 5.2: The robust "Sioux Falls East" network with 4 OD pairs and 4 toll arcs. Each OD pair is color-coded (orange, green, blue, purple). Dashed arcs represent toll arcs and solid arcs represent toll-free arcs. Edge labels correspond to commodity flows. For toll arcs, edge labels are given in the format "flow | toll". If no label is shown, there is no flow on that edge. See Figure 3 in [YB5].

6. Bilevel-Specific Sources of Uncertainty

In bilevel optimization, the decision-makers are usually assumed to have perfect information and to act perfectly rational. This means that both the leader and the follower

(i) have full knowledge about the problem data,

(ii) can perfectly observe or anticipate the decision of the other,

(iii) solve their respective optimization problems to global optimality in an exact sense.

In many real-world applications, however, the required cognitive, intellectual, or computational skills and resources to fulfill Conditions (i)–(iii) may be limited, preventing decision-makers from making fully rational decisions. As a consequence, decision-makers are frequently faced with what economists refer to as *bounded rationality* (Rubinstein 1998; Simon 1972). While *data uncertainty*, which concerns limitations related to (i), can be present in both single-level and bilevel optimization, we emphasize that (ii) and (iii) may have implications that are particularly prominent in bilevel optimization. Due to their nature of combining two different decision-makers in one model, bilevel problems may be subject to uncertainties associated with the respective decisions of these decision-makers. This type of uncertainty, which we refer to as *decision uncertainty*, relates to limitations regarding Conditions (ii) and (iii). It is evident that decision uncertainty does not play a role in single-level optimization, given that there is only one decision-maker involved. Overall, the sources of uncertainty in bilevel optimization are thus much richer than in single-level optimization and, in particular, they go beyond data uncertainty. In contrast to the case of uncertain data, decision uncertainty has been much less investigated in the bilevel literature. For an overview of the recent advances in this area, we refer to Section 5 in [YB1] and Sections 3.3–3.5 in [YB2].

In this chapter, we illustrate two aspects related to decision uncertainty. In Section 6.1, we summarize the results of [YB7] in which we address the follower's limited ability to perfectly observe the leader's decision in a robust way. Section 6.2 is devoted to the exemplary bilevel problem with a continuous but nonconvex lower level presented in [YB6]. The example illustrates the potential severe consequences of only considering nearly feasible solutions to the lower-level problem.

6.1 A Robust Approach for Modeling Limited Observability

In this section, we discuss the results of [YB7]. To this end, we consider optimistic bilinear bilevel problems of the form

$$\begin{aligned} \min_{x,y} \quad & c^\top x + d^\top y + x^\top R y \\ \text{s.t.} \quad & Ax + By \geq a, \\ & y \in S(x) \end{aligned} \tag{6.1}$$

with $x, c \in \mathbb{R}^{n_x}$, $y, d \in \mathbb{R}^{n_y}$, $R \in \mathbb{R}^{n_x \times n_y}$, $A \in \mathbb{R}^{l \times n_x}$, $B \in \mathbb{R}^{l \times n_y}$, and $a \in \mathbb{R}^l$. Here, $S(x)$ denotes the set of optimal solutions to the x-parameterized lower-level problem

$$\begin{aligned} \min_{y} \quad & f^\top x + g^\top y + x^\top Q y \\ \text{s.t.} \quad & Cx + Dy \geq b \end{aligned} \tag{6.2}$$

with $f \in \mathbb{R}^{n_x}$, $g \in \mathbb{R}^{n_y}$, $Q \in \mathbb{R}^{n_x \times n_y}$, $C \in \mathbb{R}^{m \times n_x}$, $D \in \mathbb{R}^{m \times n_y}$, and $b \in \mathbb{R}^m$. In the bilevel problem (6.1), we make the strong assumption that both the leader and the follower act perfectly rational. In real-world applications, however, this assumption rarely holds as both players may face bounded rationality. In [YB7], we pursue similar ideas compared to those in Pita et al. (2009, 2010, 2008), who consider limited observability regarding the leader's decision as one important aspect of bounded rationality. This means that the follower cannot perfectly observe the actual leader's decision x but, instead, observes a decision $\bar{x}$. As the follower assumes that $\bar{x}$ is the strategy the leader actually plays, the follower's response is then based on $\bar{x}$ instead of x. In Pita et al. (2009, 2010, 2008), the authors consider the setting in which the leader hedges against the worst-possible reaction of the follower due to his erroneous observation. Our contributions in [YB7] differ from those in Pita et al. (2009, 2010, 2008) in the sense that we focus on how limited observability regarding the leader's decision affects the lower-level problem. In this context, it seems reasonable to assume that the follower is fully aware of his inability to perfectly observe the actual decision of the leader and, thus, tries to hedge against his uncertainty regarding the leader's decision. In [YB7], we exploit robust optimization to model the decision-making of the follower. To this end, we assume that the perceived leader's decision $\bar{x}$ belongs to the predefined uncertainty set

$$\mathcal{U}(x) = \{x + P\zeta \colon \zeta \in \mathcal{Z} \subseteq \mathbb{R}^q\} \quad \text{with} \quad \mathcal{Z} = \{\zeta \in \mathbb{R}^q \colon H\zeta \geq h\}$$

with matrices P and H and a vector h of appropriate dimension. To ensure that the actual leader's decision x is still part of the uncertainty set $\mathcal{U}(x)$, one may additionally assume that $0 \in \mathcal{Z}$ holds. The robust counterpart of the lower-level problem, in which the follower hedges against the worst-possible realization of his uncertainty regarding the leader's decision, reads

$$\begin{aligned} \min_{y} \quad & g^\top y + \max_{\bar{x} \in \mathcal{U}(x)} \left\{ f^\top \bar{x} + \bar{x}^\top Q y \right\} \\ \text{s.t.} \quad & C\bar{x} + Dy \geq b \quad \text{for all } \bar{x} \in \mathcal{U}(x). \end{aligned} \tag{6.3}$$

In [YB7], we apply a duality-based approach (cf. Section 2.2.1) to obtain a tractable reformulation of Problem (6.3). The latter is given by

$$
\begin{aligned}
\min_{y,\sigma,\lambda} \quad & f^\top x + g^\top y + x^\top Q y - h^\top \sigma && && (6.4a)\\
\text{s.t.} \quad & C_{j\cdot} x + D_{j\cdot} y + h^\top \lambda^j \geq b_j, && j \in [m], && (6.4b)\\
& H^\top \sigma = -P^\top (f + Qy), && && (6.4c)\\
& H^\top \lambda^j = (C_{j\cdot} P)^\top, && j \in [m], && (6.4d)\\
& \lambda^j \geq 0, \ \sigma \geq 0, && j \in [m]; && (6.4e)
\end{aligned}
$$

see Section 4 in [YB7] for a detailed derivation. For a given leader's decision x, Problem (6.4) is a linear problem. Hence, the KKT conditions are both necessary and sufficient optimality conditions for Problem (6.4). A single-level reformulation of the overall bilevel problem, in which the follower's limited observability regarding the leader's decision is treated in a robust way, can thus be obtained by replacing the lower level by the KKT conditions of Problem (6.4). This yields

$$
\begin{aligned}
\min_{x,y,z} \quad & c^\top x + d^\top y + x^\top R y && && (6.5a)\\
\text{s.t.} \quad & Ax + By \geq a, && && (6.5b)\\
& C_{j\cdot} x + D_{j\cdot} y + h^\top \lambda^j \geq b_j, && j \in [m], && (6.5c)\\
& H^\top \sigma = -P^\top (f + Qy), && && (6.5d)\\
& H^\top \lambda^j = (C_{j\cdot} P)^\top, && j \in [m], && (6.5e)\\
& g + Q^\top x - D^\top \alpha - Q^\top P \beta = 0, && && (6.5f)\\
& h + H\beta + \delta = 0, && && (6.5g)\\
& \alpha_j h + H\gamma^j + \varepsilon^j = 0, && j \in [m], && (6.5h)\\
& \alpha_j \left(C_{j\cdot} x + D_{j\cdot} y + h^\top \lambda^j - b_j \right) = 0, && j \in [m], && (6.5i)\\
& \delta^\top \sigma = 0, && && (6.5j)\\
& \left(\varepsilon^j\right)^\top \lambda^j = 0, && j \in [m], && (6.5k)\\
& \lambda^j, \varepsilon^j \geq 0, \ \sigma, \alpha, \delta \geq 0, && j \in [m], && (6.5l)
\end{aligned}
$$

where z contains all primal variables used for the robust treatment of limited observability as well as all dual variables of the follower. Problem (6.5) is a nonlinear and nonconvex problem due to the complementarity constraints in (6.5i)–(6.5k). Nevertheless, Problem (6.5) can be solved using state-of-the-art general-purpose solvers, such as Gurobi, after some transformations; see Section 2.1.2 of this dissertation for further details. In [YB7], we tackle Constraints (6.5i)–(6.5k) using special ordered sets of type 1 (SOS1).

Example 6.1 (cf. Sections 3 and 5 in [YB7])**.** *We consider the bilinear bilevel problem* (6.1) *defined by* $c, d, f, g = 0 \in \mathbb{R}^2$, $B = C = 0 \in \mathbb{R}^{4\times 2}$, *as well as*

$$
R = \begin{bmatrix} 2 & 4 \\ 3 & 4 \end{bmatrix}, \quad Q = \begin{bmatrix} 5 & 0 \\ 1 & 2 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 1 \\ -1 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = D, \quad \textit{and} \quad a = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix} = b.
$$

This renders the bilevel problem a sequential bimatrix game, which is of interest in many security applications; see, e.g., Brown et al. (2006), Gatti (2008), Jain et al. (2010), Kiekintveld et al. (2009), Shieh et al. (2012), and Yang et al. (2014). If the follower can perfectly observe the leader's decision, an optimal solution is given by $x = (1/6, 5/6)$ *and* $y = (1, 0)$*. Now, we assume that the follower faces limited observability regarding the leader's decision and that the perceived leader's decision* $\bar{x}$ *is known to take values in the uncertainty set*

$$\mathcal{U}(x) = \{\bar{x} = x + P\zeta \colon \zeta \in \mathcal{Z}\} \cap \{\bar{x} \colon \bar{x}_1 + \bar{x}_2 = 1\}$$

with

$$P = \begin{bmatrix} p_1 & 0 \\ 0 & p_2 \end{bmatrix} \in \mathbb{R}^{2\times 2} \quad \text{and} \quad \mathcal{Z} = \left\{\zeta \in \mathbb{R}^2 \colon -1 \leq \zeta_i \leq 1,\, i = 1, 2\right\}.$$

Here, the follower particularly anticipates the resulting simplex structure of the leader's strategy. In [YB7], we compare robust solutions obtained for different values of p_1 *and* p_2*; cf. Tables 2–4 in [YB7]. For the considered academic example, we make the following main observations:*

(i) The follower's strategy shifts entirely when he faces limited observability regarding the leader's decision.

(ii) Also the leader's strategy may change significantly compared to the one under perfect rationality.

(iii) The leader's anticipation of a follower's response under limited observability always results in significantly increased costs for the leader.

The example illustrates that limited observability may significantly impact the solution to the underlying bilevel problem. Hence, this aspect should not be ignored if the application problem at hand involves a follower who cannot perfectly observe the decision of the leader.

Compared to the problem under perfect rationality, the consideration of limited observability regarding the leader's decision yields significantly larger optimization problems in terms of the number of variables and constraints. In [YB7], we thus study the relation of Problem (6.5) to bilevel problems with lower-level right-hand side uncertainty with the aim of obtaining a more compact formulation to model limited observability. To this end, we consider the problem

$$\min_{x,y} \quad c^\top x + d^\top y + x^\top R y \tag{6.6a}$$

$$\text{s.t.} \quad Ax + By \geq a, \tag{6.6b}$$

$$y \in \arg\min_{y'} \left\{ f^\top x + g^\top y' + x^\top Q y' \colon Cx + Dy' \geq \bar{b} \text{ for all } \bar{b} \in \mathcal{U}(b) \right\}. \tag{6.6c}$$

Here, b denotes the nominal right-hand side vector and the uncertainty set is given by $\mathcal{U}(b) = \mathcal{U}_1(b_1) \times \cdots \times \mathcal{U}_m(b_m)$ with

$$\mathcal{U}_j(b_j) = \left\{ b_j + \left(\tilde{p}^j\right)^\top \zeta^j \colon \tilde{H}^j \zeta^j \geq \tilde{h}^j \right\} \quad \text{for all } j \in [m],$$

where $\tilde{H}^j$ are matrices and $\tilde{p}^j$, ζ^j, and $\tilde{h}^j$ are vectors of appropriate dimension. A tractable reformulation of the lower-level problem (6.6c) can be obtained using strong duality. The resulting robustified lower-level problem is again a linear problem for a given leader's decision x. Hence, we can exploit the KKT conditions to obtain a single-level reformulation of Problem (6.6), which is given by

$$
\begin{aligned}
\min_{x,y,\tilde{z}} \quad & c^\top x + d^\top y + x^\top R y && && \text{(6.7a)}\\
\text{s.t.} \quad & Ax + By \geq a, && && \text{(6.7b)}\\
& C_{j\cdot}x + D_{j\cdot}y + \left(\tilde{h}^j\right)^\top \tilde{\lambda}^j \geq b_j, && j \in [m]\,, && \text{(6.7c)}\\
& \left(\tilde{H}^j\right)^\top \tilde{\lambda}^j = -\tilde{p}^j, && j \in [m]\,, && \text{(6.7d)}\\
& g + Q^\top x - D^\top \tilde{\alpha} = 0, && && \text{(6.7e)}\\
& \tilde{\alpha}_j \tilde{h}^j + \tilde{H}^j \tilde{\beta}^j + \tilde{\gamma}^j = 0, && j \in [m]\,, && \text{(6.7f)}\\
& \tilde{\alpha}_j \left(C_{j\cdot}x + D_{j\cdot}y + \left(\tilde{h}^j\right)^\top \tilde{\lambda}^j - b_j \right) = 0, && j \in [m]\,, && \text{(6.7g)}\\
& \left(\tilde{\gamma}^j\right)^\top \tilde{\lambda}^j = 0, && j \in [m]\,, && \text{(6.7h)}\\
& \tilde{\lambda}^j, \tilde{\gamma}^j \geq 0,\ \tilde{\alpha} \geq 0, && j \in [m]\,. && \text{(6.7i)}
\end{aligned}
$$

Here, $\tilde{z}$ contains all primal variables used for the robustification of the uncertain right-hand sides as well as all resulting dual lower-level variables. In the following theorem, we formally address the connection between Problem (6.5) and Problem (6.7). A proof of this theorem can be found in [YB7].

Theorem 6.2 (cf. Theorem 2 in [YB7])**.** *Let (x, y, z) be a solution to Problem (6.5) with parameters P, H, and h modeling the uncertainty set. Furthermore, let $(x, y, \tilde{z})$ be a solution to Problem (6.7) with the parameters $\tilde{p}^j$, $\tilde{H}^j$, and $\tilde{h}^j$, $j \in [m]$, modeling the uncertainty sets. Third, suppose that the lower-level's objective function does not contain bilinear terms, i.e., $Q = 0$, and that* $\operatorname{rank}(D^\top) = \operatorname{rank}([D^\top, g])$ *holds. Then, the uncertainty modeling parameters satisfy*

$$
\begin{aligned}
\left(\tilde{h}^j\right)^\top \tilde{\lambda}^j &= h^\top \lambda^j, && j \in [m]\,,\\
\left(\tilde{\beta}^j\right)^\top \tilde{p}^j &= -\left(\gamma^j\right)^\top \left(C_{j\cdot}P\right)^\top, && j \in [m]\,,\\
\left(\tilde{\beta}^j\right)^\top \left(\tilde{H}^j\right)^\top \tilde{\lambda}^j &= -\left(\gamma^j\right)^\top H^\top \lambda^j, && j \in [m]\,,\\
\left(\tilde{\beta}^j\right)^\top \left(\tilde{H}^j\right)^\top \tilde{\lambda}^j &= \left(\gamma^j\right)^\top \left(H^\top \sigma - P^\top \left(C_{j\cdot}^\top + f\right)\right), && j \in [m]\,.
\end{aligned}
$$

In Theorem 6.2, we state that there exists a connection between Problem (6.5) and a suitably chosen bilevel problem with lower-level right-hand side uncertainty. We obtain this relation under the rather strong assumptions that $Q = 0$ and $\operatorname{rank}(D^\top) = \operatorname{rank}([D^\top, g])$ hold. The assumption $Q = 0$ is used to reduce Constraints (6.5f) and (6.7e) to the linear system $D^\top \alpha = g$, which, given the rank condition, has a solution. Nevertheless, despite these rather strong assumptions, the established relation requires the knowledge of the lower-level primal and dual variables in advance. Hence, the established result is only an ex-post relation.

6.2 A Computationally Ill-Behaved Bilevel Problem

In this section, we illustrate that even if (i) all problem data is known, (ii) the leader can fully anticipate the follower's response, and (iii) the follower can perfectly observe the leader's decision, limited computational resources may still prevent the decision-makers from reacting optimally. More specifically, we show that even extremely small violations of the lower-level constraints can have severe consequences when solving bilevel problems with a continuous but nonconvex lower level computationally. To this end, we consider the exemplary bilevel problem that has been presented in [YB6]. The problem is given by

$$\begin{aligned} \max_{x \in \mathbb{R}^2} \quad & F(x,y) = x_1 - 2y_{n+1} + y_{n+2} \\ \text{s.t.} \quad & (x_1, x_2) \in [\underline{x}_1, \bar{x}_1] \times [\underline{x}_2, \bar{x}_2], \\ & y \in S(x), \end{aligned} \tag{6.8}$$

where $\underline{x}$, $\bar{x} \in \mathbb{R}^2$ with $1 \leq \underline{x}_i < \bar{x}_i$, $i \in \{1,2\}$, denote lower and upper bounds on the variables x. Here, $S(x)$ denotes the set of optimal solutions to the x-parameterized problem

$$\max_{y \in \mathbb{R}^{n+2}} \quad f(x,y) = y_1 - y_n \left(x_1 + x_2 - y_{n+1} - y_{n+2}\right) \tag{6.9a}$$

$$\text{s.t.} \quad y_1 + y_n = \frac{1}{2}, \tag{6.9b}$$

$$y_i^2 \leq y_{i+1}, \quad i \in \{1, \ldots, n-1\}, \tag{6.9c}$$

$$y_i \geq 0, \quad i \in \{1, \ldots, n\}, \tag{6.9d}$$

$$y_{n+1} \in [0, x_1], \tag{6.9e}$$

$$y_{n+2} \in [-x_2, x_2]. \tag{6.9f}$$

In the bilevel problem (6.8), all coefficients are ±1 and we only use quadratic or linear terms. The upper-level problem (6.8) is a linear problem in both the leader's and the follower's variables. Moreover, the only constraints that occur in this problem are variable bounds for the leader's variables x. Hence, there are no upper-level constraints that explicitly depend on the follower's variables y, i.e., there are no coupling constraints. The main idea for the construction of the lower-level problem is based on a constraint set presented first by Bienstock et al. (2021). In particular, we emphasize that all lower-level constraints are linear except for the quadratic but convex inequality constraints in (6.9c). Nevertheless, the overall lower-level problem (6.9) is nonconvex due to bilinear terms in the follower's objective function.

Proposition 6.3 (Sections 2–3 and Appendix in [YB6]). *Let* $(x_1, x_2) \in [\underline{x}_1, \bar{x}_1] \times [\underline{x}_2, \bar{x}_2]$ *be given arbitrarily. Then, the lower-level problem* (6.9) *has the following properties:*

(i) The feasible set of the lower-level problem is non-empty, convex, and compact.

(ii) A feasible follower's decision $y \in \mathbb{R}^{n+2}$ *satisfies* $y_n > 0$.

(iii) The lower-level problem satisfies Slater's constraint qualification.

(iv) The Mangasarian–Fromovitz constraint qualification (MFCQ) is satisfied at every feasible decision of the follower.

(v) The unique optimal solution y^* to the lower-level problem is given by $(y^*_{n+1}, y^*_{n+2}) = (x_1, x_2)$ *and*

$$y^*_i = (y^*_1)^{2^{i-1}} \qquad \text{for all } i \in \{2, \ldots, n\},$$

where y^*_1 is the unique root of the function

$$h : \left[0, \frac{1}{2}\right] \to \mathbb{R}, \quad z \mapsto z + z^{2^{n-1}} - \frac{1}{2}.$$

(vi) The linear independence constraint qualification (LICQ) is satisfied at the optimal solution y^*.

(vii) The strict complementarity condition is satisfied at the optimal solution y^*.

By Part (v) of Proposition 6.3, there is no need to distinguish between the optimistic and the pessimistic approach to bilevel optimization. Hence, it suffices to solve

$$\max_x \quad -x_1 + x_2 \quad \text{s.t.} \quad (x_1, x_2) \in [\underline{x}_1, \bar{x}_1] \times [\underline{x}_2, \bar{x}_2]$$

to determine an optimal leader's decision to the overall bilevel problem (6.8). This leads to the following result.

Proposition 6.4 (cf. Result 3 in [YB6]). *The bilevel problem (6.8) has a unique solution given by $x^* = (\underline{x}_1, \bar{x}_2)$ with an optimal objective function value of $F^* = -\underline{x}_1 + \bar{x}_2$.*

While, in theory, we can solve the bilevel problem (6.8) in an exact sense, this is usually not possible when solving bilevel problems with a continuous but nonconvex lower-level problem computationally. The reason is that the lower-level problem can, in general, not be solved to global optimality anymore in an exact sense in finite time since we need to exploit techniques such as spatial branching to tackle nonconvexities. These techniques only lead to finite algorithms for prescribed and strictly positive feasibility tolerances; see, e.g., Locatelli and Schoen (2013). Hence, just due to algorithmic necessities, we cannot expect to get an exact feasible follower's solution when solving continuous but nonconvex lower-level problems computationally. This limitation of the follower's computational resources may be seen as an instantiation of bounded rationality. In what follows, we determine an optimal solution to the bilevel problem (6.8) under the assumption that we allow for small violations of the nonlinear lower-level constraints according to the following notion.

Definition 6.5 (cf. Definition 1 in [YB6]). *Let $0 < \varepsilon \in \mathbb{R}$, $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, and $h : \mathbb{R}^n \to \mathbb{R}^p$ be given. A point $x \in \mathbb{R}^n$ is called* ε-feasible *for the problem*

$$\max_{x \in \mathbb{R}^n} \quad f(x) \quad \text{s.t.} \quad g(x) \le 0, \; h(x) = 0,$$

if the following conditions are satisfied:

(i) $g_i(x) \le 0$ *for all* $i \in \{1, \ldots, m\} \setminus N$,

(ii) $h_j(x) = 0$ *for all* $j \in \{1, \ldots, p\} \setminus M$, *and*

(iii) $\max\{\max\{g_i(x) \colon i \in N\}, \max\{|h_j(x)| \colon j \in M\}\} \le \varepsilon$.

Here, $N \subseteq \{1, \dots, m\}$ and $M \subseteq \{1, \dots, p\}$ denote the index sets of all nonlinear inequality and equality constraints, respectively.

Definition 6.5 is motivated by the necessary special treatment of nonlinear and, in particular, nonconvex constraints in computational global optimization. In the following proposition, we state the set of ε-feasible follower's solutions according to the notion of Definition 6.5.

Proposition 6.6 (Section 4 in [YB6]). *Let $(x_1, x_2) \in [\underline{x}_1, \bar{x}_1] \times [\underline{x}_2, \bar{x}_2]$ be given arbitrarily and let $\varepsilon \geq 2^{-2^{n-1}}$. Then, the set of ε-feasible solutions to the lower-level problem (6.9) is given by*

$$\left\{ y \in \mathbb{R}^{n+2} \colon y_i = 2^{-2^{i-1}},\ i \in \{1, \dots, n-1\},\ y_n = 0,\ y_{n+1} \in [0, x_1],\ y_{n+2} \in [-x_2, x_2] \right\}.$$

In contrast to the exact case, the set of ε-feasible solutions to the lower-level problem (6.9) is not a singleton anymore since the variables y_{n+1} and y_{n+2} can be chosen arbitrarily within their variable bounds. Hence, we need to distinguish between the optimistic and the pessimistic approach to bilevel optimization. In the optimistic setting, the follower chooses $y_{n+1} = 0$ and $y_{n+2} = x_2$ to favor the leader w.r.t. her objective function value. To determine an optimistic optimal leader's decision for the bilevel problem with an ε-feasible follower, it thus suffices to solve the linear problem

$$\max_{x} \quad x_1 + x_2 \quad \text{s.t.} \quad (x_1, x_2) \in [\underline{x}_1, \bar{x}_1] \times [\underline{x}_2, \bar{x}_2].$$

In the pessimistic setting, the follower chooses $y_{n+1} = x_1$ and $y_{n+2} = -x_2$ to adversely affect the leader's decision. Hence, a pessimistic optimal leader's decision is obtained by solving the linear problem

$$\max_{x} \quad -x_1 - x_2 \quad \text{s.t.} \quad (x_1, x_2) \in [\underline{x}_1, \bar{x}_1] \times [\underline{x}_2, \bar{x}_2].$$

Proposition 6.7 (cf. Result 6 in [YB6]). *Let $\varepsilon \geq 2^{-2^{n-1}}$ and suppose that we allow for ε-feasible follower's solutions. Then, the optimistic optimal solution to the bilevel problem (6.8) is given by $x_o^* = (\bar{x}_1, \bar{x}_2)$ with an optimal objective function value of $F_o^* = \bar{x}_1 + \bar{x}_2$. The pessimistic optimal solution is given by $x_p^* = (\underline{x}_1, \underline{x}_2)$ with an optimal objective function value of $F_p^* = -\underline{x}_1 - \underline{x}_2$.*

In Figure 6.1, we summarize all previous results by showing the unique exact solution to the bilevel problem (6.8) as well as the optimistic and the pessimistic solution for the problem with an ε-feasible follower. Here, three aspects are remarkable. First, by enlarging the feasible interval for the leader's variables x, a solution with an ε-feasible follower can be arbitrarily far away from the overall exact bilevel solution. Second, we also obtain an arbitrarily large error regarding the optimal objective function value of the leader. Third, the obtained errors are independent of the feasibility tolerance ε. Hence, the results hold true for arbitrarily small values of $\varepsilon \geq 2^{-2^{n-1}}$. Note that we obtain arbitrarily small values for ε by increasing the parameter n. In particular, due to

$$\varepsilon \geq 2^{-2^{n-1}} \iff n \geq \log_2 \left(\log_2 \left(\frac{1}{\varepsilon^2} \right) \right),$$

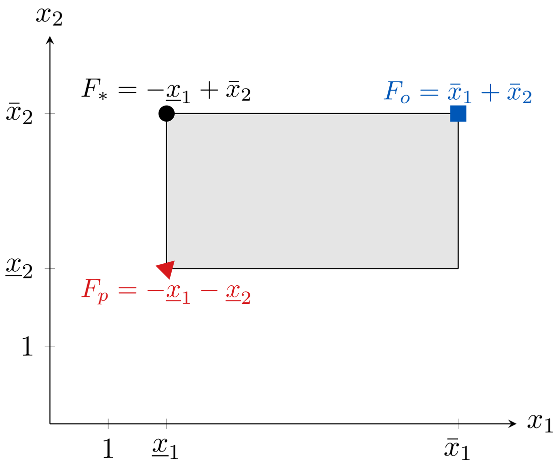

Figure 6.1: The upper-level feasible region (gray area), the unique exact solution to the bilevel problem (6.8) (black dot), as well as the optimistic (blue square) and the pessimistic solution (red triangle) for the problem with an ε-feasible follower. Additionally, the corresponding objective function values of the leader are shown.

only very moderate values of n are required to observe a discrepancy between the exact and computationally obtained solutions for the bilevel problem (6.8) for a given tolerance ε. For instance, a tolerance of $\varepsilon = 10^{-8}$ already leads to a wrong result for $n = 6$. Hence, the considered bilevel problem remains moderate in size w.r.t. the number of constraints and variables. In particular, the used constraint coefficients do not depend on n or the tolerance ε. Taking all previous observations into account leads to the following main result of [YB6].

Observation 6.8. *Even if the feasibility tolerance ε can be made extremely small, the exact solution to the bilevel problem* (6.8) *can be arbitrarily far away from the solution obtained for the problem with ε-feasibility in the lower level. The same holds true for the respective optimal objective function values.*

The result of Observation 6.8 is very much in contrast to the situation in single-level optimization, for which sensitivity results are available; see, e.g., Proposition 4.2.2 in Bertsekas (2016). In particular, Observation 6.8 highlights that computational bilevel optimization with continuous but nonconvex lower levels needs to be done with great care. If the application problem at hand involves a follower who, due to his bounded rationality, cannot solve the lower-level problem in an exact sense (in finite time), ex-post checks may be needed to avoid considering arbitrarily bad points as “solutions”.

In [YB6], we further show that linear bilevel problems behave better on the level of feasible points, which indicates that the pathological behavior observed for the bilevel problem (6.8) is due to the nonlinearities at the lower level. We summarize the results of [YB6] for the linear case in the remainder of this section. To this end, we consider linear bilevel problems of the form

$$
\begin{aligned}
\min_{x,y} \quad & c_x^\top x + c_y^\top y && \text{(6.10a)}\\
\text{s.t.} \quad & Ax \geq a, && \text{(6.10b)}\\
& y \in \arg\min_{y'} \left\{ d^\top y' \colon Cx + Dy' \geq b \right\} && \text{(6.10c)}
\end{aligned}
$$

with $c_x \in \mathbb{Q}^{n_x}$, $c_y, d \in \mathbb{Q}^{n_y}$, $A \in \mathbb{Q}^{l \times n_x}$, $a \in \mathbb{Q}^l$, $C \in \mathbb{Q}^{m \times n_x}$, $D \in \mathbb{Q}^{m \times n_y}$, and $b \in \mathbb{Q}^m$. We assume that the shared constraint set $\{(x, y) \in \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \colon Ax \geq a,\ Cx + Dy \geq b\}$ is non-empty and compact and that for every feasible upper-level decision x, there exists a feasible lower-level decision y. The latter assumptions are commonly used in bilevel optimization to ensure that Problem (6.10) has a solution. In particular, they imply that the lower-level problem is bounded for every feasible decision of the leader and that the dual problem of the lower level is feasible.

Definition 6.9 (cf. Definition 3 in [YB6]). *Let $z \in \mathbb{R}^m$ satisfy $D^\top z = d$ and define $B = B(z) = \{j \colon z_j \neq 0\}$. We say that z is* dual basic *if the submatrix $D_B^\top$ of $D^\top$ corresponding to the columns in B has rank $|B|$.*

We further assume that the set $\{x \in \mathbb{R}^{n_x} \colon Ax \geq a\}$ is bounded and that, for a given tolerance $\varepsilon \in (0, 1)$ and e_k being the k-dimensional vector of all ones, our underlying solver ensures the properties

$$
A\hat{x} \geq a - \varepsilon e_l, \quad C\hat{x} + D\hat{y} \geq b - \varepsilon e_m, \quad \text{and} \quad d^\top \hat{y} \geq \min_y \left\{ d^\top y \colon C\hat{x} + Dy \geq b \right\} - \varepsilon.
$$

This means that $(\hat{x}, \hat{y})$ is ε-feasible w.r.t. the shared constraints of the bilevel problem (6.10) and the follower's decision $\hat{y}$ may be superoptimal up to a degree of ε. The following theorem provides the main result of [YB6] on the distance to feasibility and the superoptimality of an ε-feasible point $(\hat{x}, \hat{y})$ for the linear bilevel problem (6.10).

Theorem 6.10 (cf. Theorem 2 in [YB6]). *Let $\hat{x} \in \mathbb{R}^{n_x}$, $\hat{y} \in \mathbb{R}^{n_y}$, and $\hat{z} \in \mathbb{R}^m$ be such that*

(i) $A\hat{x} \geq a - \varepsilon e_l$,

(ii) $D\hat{y} \geq b - C\hat{x} - \varepsilon e_m$,

(iii) $\|D^\top \hat{z} - d\|_\infty \leq \varepsilon$, $\hat{z} \geq -\varepsilon e_m$,

(iv) $d^\top \hat{y} - (b - C\hat{x})^\top \hat{z} \leq \varepsilon$,

(v) $\|\hat{z} - \tilde{z}\|_\infty \leq \varepsilon$ *for some dual basic* $\tilde{z}$.

Then, there exists a pair (x^*, y^*) that is feasible for the bilevel problem (6.10) such that

$$
\begin{aligned}
\|(x^*, y^*)^\top - (\hat{x}, \hat{y})^\top\|_\infty &\leq \varepsilon \kappa_1(A, C, D, a, b, d),\\
|c_x^\top x^* + c_y^\top y^* - (c_x^\top \hat{x} + c_y^\top \hat{y})| &\leq \varepsilon \kappa_2(A, C, D, a, b, c, d)
\end{aligned}
$$

hold for certain constants $\kappa_1(A, C, D, a, b, d)$ and $\kappa_2(A, C, D, a, b, c, d) > 0$, whose sizes are polynomial in the size of the input data.

In Theorem 6.10, we state that the distance to feasibility and the superoptimality of a nearly feasible point $(\hat{x}, \hat{y})$ for the linear bilevel problem (6.10) is linear in ε with coefficients κ that have polynomial size in the input data. Here, "size" refers to the corresponding encoding length of the constants κ or of the input data. This type of guarantee with polynomially sized coefficients is not available in the nonlinear case; cf. Observation 6.8.

7. Conclusion and Outlook

In this dissertation, we have addressed some of the various challenges that may arise in the context of bilevel optimization under uncertainty. In particular, we have highlighted that the sources of uncertainty in bilevel optimization are much richer compared to classic, i.e., single-level, optimization. In bilevel optimization, not only the problem data but also the (observation of the) decisions of the decision-makers may be subject to uncertainty. We have presented robust approaches to deal with both data and decision uncertainty in the lower level of bilevel problems. In this dissertation, we have focused on robust bilevel problems with "here-and-now" followers, i.e., bilevel problems in which the follower has to decide before the uncertainty realizes. We have presented two exact branch-and-cut approaches to solve mixed-integer linear bilevel problems with a Γ-robust treatment of lower-level objective uncertainty, along with cuts tailored to the class of monotone interdiction problems. In addition, we have presented heuristics for mixed-integer, linear, and Γ-robust bilevel problems, which rely on solving a linear number of bilevel problems of the nominal type. These heuristics do not require problem-specific tailoring so that any state-of-the-art or off-the-shelf solver can be used within our frameworks. The algorithmic approaches presented in this dissertation are the first to tackle mixed-integer, linear, and Γ-robust bilevel problems directly. Moreover, we have proposed a robust approach to hedge against uncertain travel costs for the problem of determining optimal tolls in a traffic network. In this problem, which can be seen as a single-leader multi-follower problem with multiple robustified followers, the travelers act in the sense of Wardrop's user equilibrium. We have demonstrated the impact of robustified followers on the toll-setting policies through a case study. Finally, we have illustrated the importance of considering the decision-makers' bounded rationality in bilevel optimization. To this end, we have presented a strictly robust approach in which the follower hedges against uncertainties arising from his limited ability to perfectly observe the leader's decision. Moreover, we have shown on an exemplary bilevel problem with a continuous but nonconvex lower level that even very small deviations in the follower's decision may lead to arbitrarily large discrepancies between exact and computationally obtained bilevel solutions. Nevertheless, despite our considerable contributions, there are still many topics open for future research.

One possible direction of future research could be the derivation of further valid cutting planes for mixed-integer bilevel problems with robustified followers. The cuts presented in Chapter 3 of this dissertation crucially rely on the assumption of a downward monotone lower-level problem. Thus, deriving valid cuts for problems that do not satisfy the downward monotonicity property may be a reasonable aspect of future work.

Another research question that naturally arises is whether we can identify properties that guarantee ex ante that mixed-integer linear bilevel problems with a Γ-robust treatment of lower-level objective uncertainty can be solved by only solving bilevel problems of the nominal type. Up to now, it is unclear whether such properties exist. The sufficient conditions presented in Chapter 4 of this dissertation can only be checked ex post, i.e., after solving all deterministic bilevel problems.

In Chapter 5 of this dissertation, we have seen that determining optimal tolls in a traffic network, in which travelers act in the sense of Wardrop's user equilibrium, is a highly challenging task. One promising avenue of future research in this area may be the development of tailored solution approaches that exploit piecewise-linear approximations of the bilinearities in the upper-level objective function. Moreover, for the robust toll-setting problem, an interesting research question is whether we can identify properties that ensure the existence of Γ-robust Wardrop equilibria in the sense of Bertsimas and Sim (2003) and Sim (2004).

Overall, robust approaches to deal with uncertainties in bilevel optimization are still in their infancy. The majority of works focuses on strictly or Γ-robust approaches to deal with uncertainties that are mainly assumed to take values in interval or polyhedral uncertainty sets. Hence, the consideration of (i) other uncertainty set geometries and (ii) other robustness concepts, such as light robustness (Fischetti and Monaci 2009), adjustable robustness (Ben-Tal et al. 2004), or distributional robustness (Goh and Sim 2010; Wiesemann et al. 2014), may be reasonable aspects of future research.

Finally, we point out that the illustrative bilevel problem discussed in Section 6.2 of this dissertation leaves many open research questions regarding the sensitivity of bilevel problems. Future research could, e.g., address methods to check the quality of computationally obtained bilevel solutions (ex post) to avoid considering arbitrarily bad points as "solutions". Moreover, an important research direction includes identifying classes of bilevel problems for which we can guarantee that the observed pathological behavior cannot occur.

Bibliography

Aashtiani, H. Z. and T. L. Magnanti. “Equilibria on a Congested Transportation Network.” In: *SIAM Journal on Algebraic Discrete Methods* 2.3 (1981). DOI: `10.1137/0602024`.

Aboussoror, A. and P. Loridan. “Strong-weak Stackelberg Problems in Finite Dimensional Spaces.” In: *Serdica. Mathematical Journal* 21 (1995), pp. 151–170. URL: `https://api.semanticscholar.org/CorpusID:54708699`.

Aussel, D. and A. Svensson. “Is Pessimistic Bilevel Programming a Special Case of a Mathematical Program with Complementarity Constraints?” In: *Journal of Optimization Theory and Applications* 181 (2019), pp. 504–520. DOI: `10.1007/s10957-018-01467-7`.

Bagwell, K. “Commitment and observability in games.” In: *Games and Economic Behavior* 8.2 (1995), pp. 271–280. DOI: `10.1016/S0899-8256(05)80001-6`.

Ben-Ayed, O., C. Blair, D. Boyce, and L. LeBlanc. “Construction of a real-world bilevel linear programming model of the highway network design problem.” In: *Annals of Operations Research* 34 (1992), pp. 219–254. DOI: `10.1007/BF02098181`.

Ben-Tal, A., A. Goryashko, E. Guslitzer, and A. Nemirovski. “Adjustable robust solutions of uncertain linear programs.” In: *Mathematical Programming* 99.2 (2004), pp. 351–376. DOI: `10.1007/s10107-003-0454-y`.

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

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

Benders, J. F. “Partitioning procedures for solving mixed-variables programming problems.” In: *Numerische Mathematik* 4 (1962), pp. 238–252. DOI: `10.1007/BF01386316`.

Bertsekas, D. P. *Nonlinear Programming.* Belmont, Massachusetts: Athena Scientific, 2016.

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

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

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

Besançon, M., M. F. Anjos, and L. Brotcorne. “Complexity of near-optimal robust versions of multilevel optimization problems.” In: *Optimization Letters* 15 (2021), pp. 2597–2610. DOI: `10.1007/s11590-021-01754-9`.

Besançon, M., M. F. Anjos, and L. Brotcorne. “Robust bilevel optimization for near-optimal lower-level solutions.” In: *Journal of Global Optimization* (2024), pp. 813–842. DOI: `10.1007/s10898-024-01422-z`.

Bienstock, D., A. D. Pia, and R. Hildebrand. “Complexity, Exactness, and Rationality in Polynomial Optimization.” In: *Integer Programming and Combinatorial Optimization*. Ed. by M. Singh and D. P. Williamson. Vol. 12707. IPCO 2021. Springer, Cham., 2021, pp. 58–72. DOI: `10.1007/978-3-030-73879-2_5`.

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

Bolusani, S., S. Coniglio, T. K. Ralphs, and S. Tahernejad. “A Unified Framework for Multistage Mixed Integer Linear Optimization.” In: *Bilevel Optimization: Advances and Next Challenges*. Ed. by S. Dempe and A. Zemkoho. Cham: Springer International Publishing, 2020, pp. 513–560. DOI: `10.1007/978-3-030-52119-6_18`.

Boyd, S. and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

Bracken, J. and J. T. McGill. “Mathematical programs with optimization problems in the constraints.” In: *Operations Research* 21.1 (1973), pp. 37–44. DOI: `10.1287/opre.21.1.37`.

Brotcorne, L., M. Labbé, P. Marcotte, and G. Savard. “A Bilevel Model for Toll Optimization on a Multicommodity Transportation Network.” In: *Transportation Science* 35.4 (2001). DOI: `10.1287/trsc.35.4.345.10433`.

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

Buchheim, C. “Bilevel linear optimization belongs to NP and admits polynomial-size KKT-based reformulations.” In: *Operations Research Letters* 51.6 (2023), pp. 618–622. DOI: `10.1016/j.orl.2023.10.006`.

Burtscheidt, J. and M. Claus. “Bilevel Linear Optimization Under Uncertainty.” In: *Bilevel Optimization: Advances and Next Challenges*. Ed. by S. Dempe and A. Zemkoho. Springer International Publishing, 2020, pp. 485–511. DOI: `10.1007/978-3-030-52119-6_17`.

Burtscheidt, J., M. Claus, and S. Dempe. “Risk-Averse Models in Bilevel Stochastic Linear Programming.” In: *SIAM Journal on Optimization* 30.1 (2020), pp. 377–406. DOI: `10.1137/19M1242240`.

Candler, W. and R. Norton. *Multi-level Programming*. Discussion Papers, Development Research Center, International Bank for Reconstruction and Development. World Bank, 1977. URL: `https://books.google.de/books?id=TicmAQAAMAAJ`.

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

Chuong, T. D. and V. Jeyakumar. “Finding Robust Global Optimal Values of Bilevel Polynomial Programs with Uncertain Linear Constraints.” In: *Journal of Optimization Theory and Applications* 173.2 (2017), pp. 683–703. DOI: `10.1007/s10957-017-1069-4`.

Colson, B., P. Marcotte, and G. Savard. “An overview of bilevel optimization.” In: *Annals of Operations Research* 153 (2007), pp. 235–256. DOI: `10.1007/s10479-007-0176-2`.

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

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

Dantzig, G. B., S. F. Maier, and Z. F. Lansdown. *Application of Decomposition to Transportation Network Analysis.* Tech. rep. 1976. URL: https://rosap.ntl.bts.gov/view/dot/10514.

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

Dempe, S. and J. Dutta. "Is bilevel programming a special case of a mathematical program with complementarity constraints?" In: *Mathematical Programming* 131.1-2 (2012), pp. 37–48. DOI: 10.1007/s10107-010-0342-1.

Dempe, S., S. Ivanov, and A. Naumov. "Reduction of the bilevel stochastic optimization problem with quantile objective function to a mixed-integer problem." In: *Applied Stochastic Models in Business and Industry* 33.5 (2017), pp. 544–554. DOI: 10.1002/asmb.2254.

Dempe, S. and A. Zemkoho. "Bilevel road pricing: theoretical analysis and optimality conditions." In: *Annals of Operations Research* 196 (2012), pp. 223–240. DOI: 10.1007/s10479-011-1023-z.

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

DeNegre, S. T. and T. K. Ralphs. "A Branch-and-cut Algorithm for Integer Bilevel Linear Programs." In: *Operations Research and Cyber-Infrastructure.* Springer, 2009, pp. 65–78. DOI: 10.1007/978-0-387-88843-9_4.

Dewez, S., M. Labbé, P. Marcotte, and G. Savard. "New formulations and valid inequalities for a bilevel pricing problem." In: *Operations Research Letters* 36.2 (2008), pp. 141–149. DOI: 10.1016/j.orl.2007.03.005.

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

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

Fischetti, M., I. Ljubić, M. Monaci, and M. Sinnl. "On the use of intersection cuts for bilevel optimization." In: *Mathematical Programming* 172 (2018), pp. 77–103. DOI: 10.1007/s10107-017-1189-5.

Fischetti, M. and M. Monaci. "Light Robustness." In: *Robust and Online Large-Scale Optimization: Models and Techniques for Transportation Systems.* Ed. by R. K. Ahuja, R. H. Möhring, and C. D. Zaroliagis. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 61–84. DOI: 10.1007/978-3-642-05465-5_3.

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

Fortuny-Amat, J. and B. McCarl. "A Representation and Economic Interpretation of a Two-Level Programming Problem." In: *The Journal of the Operational Research Society* 32.9 (1981), pp. 783–792. DOI: 10.1057/jors.1981.156.

Furini, F., I. Ljubić, P. S. Segundo, and Y. Zhao. "A branch-and-cut algorithm for the Edge Interdiction Clique Problem." In: *European Journal of Operational Research* 294.1 (2021), pp. 54–69. DOI: 10.1016/j.ejor.2021.01.030.

Gartner, N. H. "Optimal Traffic Assignment with Elastic Demands: A Review Part II. Algorithmic Approaches." In: *Transportation Science* 14.2 (1980), pp. 192–208. URL: http://www.jstor.org/stable/25767978.

Gatti, N. "Game Theoretical Insights in Strategic Patrolling: Model and Algorithm in Normal-Form." In: *ECAI 2008—Proceedings of the 18th European Conference on Artificial Intelligence*. 2008, pp. 403–407. DOI: 10.3233/978-1-58603-891-5-403.

Geoffrion, A. M. "Generalized Benders decomposition." In: *Journal of Optimization Theory and Applications* 10 (1972), pp. 237–260. DOI: 10.1007/BF00934810.

Goerigk, M., J. Kurtz, M. Schmidt, and J. Thürauf. "Connections between Robust and Bilevel Optimization." In: *Open Journal of Mathematical Optimization* 6.2 (2025), pp. 1–17. DOI: 10.5802/ojmo.38.

Goh, J. and M. Sim. "Distributionally Robust Optimization and Its Tractable Approximations." In: *Operations Research* 58.4-part-1 (2010), pp. 902–917. DOI: 10.1287/opre.1090.0795.

Gorissen, B. L., İhsan Yanıkoğlu, and D. den Hertog. "A practical guide to robust optimization." In: *Omega* 53 (2015), pp. 124–137. DOI: 10.1016/j.omega.2014.12.006.

Haghighat, H. "Strategic offering under uncertainty in power markets." In: *International Journal of Electrical Power & Energy Systems* 63 (2014), pp. 1070–1077. DOI: 10.1016/j.ijepes.2014.05.049.

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

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

Ito, Y. "Robust Wardrop Equilibria in the Traffic Assignment Problem with Uncertain Data." Master's thesis. Graduate School of Informatics, Kyoto University, 2011. URL: http://www-optima.amp.i.kyoto-u.ac.jp/papers/master/2011_master_ito.pdf.

Ivanov, S. V. "A Bilevel Stochastic Programming Problem with Random Parameters in the Follower's Objective Function." In: *Journal of Applied and Industrial Mathematics* 12.4 (2018), pp. 658–667. DOI: 10.1134/S1990478918040063.

Ivanov, S. V. "Bilevel Stochastic Linear Programming Problems with Quantile Criterion." In: *Automation and Remote Control* 75 (2014), pp. 107–118. DOI: 10.1134/S0005117914010081.

Jain, M., J. Tsai, J. Pita, C. Kiekintveld, S. Rathi, M. Tambe, and F. Ordóñez. "Software Assistants for Randomized Patrol Planning for the LAX Airport Police and the Federal Air Marshal Service." In: *INFORMS Journal on Applied Analytics* 40.4 (2010), pp. 267–290. DOI: 10.1287/inte.1100.0505.

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

Kalashnikov, V., J. G. Flores Muñiz, and N. Kalashnykova. "Bilevel Optimal Tolls Problems with Nonlinear Costs: A Heuristic Solution Method." In: *Beyond Traditional Probabilistic Data Processing Techniques: Interval, Fuzzy etc. Methods and Their Applications*. Ed. by O. Kosheleva, S. P. Shary, G. Xiang, and R. Zapatrin. Cham: Springer International Publishing, 2020, pp. 481–516. DOI: 10.1007/978-3-030-31041-7_28.

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

Kiekintveld, C., M. Jain, J. Tsai, J. Pita, F. Ordóñez, and M. Tambe. "Computing optimal randomized resource allocations for massive security games." In: *AAMAS 2009—Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems.* Vol. 1. 2009, pp. 689–696. URL: `https://dl.acm.org/doi/pdf/10.5555/1558013.1558108`.

Kleinert, T., V. Grimm, and M. Schmidt. "Outer approximation for global optimization of mixed-integer quadratic bilevel problems." In: *Mathematical Programming* (2021), pp. 461–521. DOI: `10.1007/s10107-020-01601-2`.

Kleinert, T., M. Labbé, I. Ljubić, and M. Schmidt. "A Survey on Mixed-Integer Programming Techniques in Bilevel Optimization." In: *EURO Journal of Computational Optimization* 9 (2021), p. 100007. DOI: `10.1016/j.ejco.2021.100007`.

Kleinert, T., M. Labbé, F. Plein, and M. Schmidt. "There's No Free Lunch: On the Hardness of Choosing a Correct Big-M in Bilevel Optimization." In: *Operations Research* 68.6 (2020), pp. 1716–1721. DOI: `10.1287/opre.2019.1944`.

Kleinert, T. and M. Schmidt. "Why there is no need to use a big-M in linear bilevel optimization: a computational study of two ready-to-use approaches." In: *Computational Management Science* 3 (2023). DOI: `10.1007/s10287-023-00435-5`.

Köppe, M., M. Queyranne, and C. T. Ryan. "Parametric Integer Programming Algorithm for Bilevel Mixed Integer Programs." In: *Journal of Optimization Theory and Applications* 146.1 (2010), pp. 137–150. DOI: `10.1007/s10957-010-9668-3`.

Labbé, M., P. Marcotte, and G. Savard. "A Bilevel Model of Taxation and Its Application to Optimal Highway Pricing." In: *Management Science* 44.12.1 (1998). DOI: `10.1287/mnsc.44.12.1608`.

Labbé, M., P. Marcotte, and G. Savard. "On a class of bilevel programs." In: *Nonlinear Optimization and Related Topics.* Ed. by G. D. Pillo and F. Giannessi. Boston, MA: Springer US, 2000, pp. 183–206. DOI: `10.1007/978-1-4757-3226-9_10`.

Lampariello, L., S. Sagratella, and O. Stein. "The Standard Pessimistic Bilevel Problem." In: *SIAM Journal on Optimization* 29.2 (2019), pp. 1634–1656. DOI: `10.1137/18M119759X`.

Land, A. H. and A. G. Doig. "An Automatic Method of Solving Discrete Programming Problems." In: *Econometrica* 28.3 (1960), pp. 497–520. URL: `http://www.jstor.org/stable/1910129`.

LeBlanc, L. J., E. K. Morlok, and W. P. Pierskalla. "An efficient approach to solving the road network equilibrium traffic assignment problem." In: *Transportation Research* 9.5 (1975), pp. 309–318. DOI: `10.1016/0041-1647(75)90030-1`.

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

Liu, J., Y. Fan, Z. Chen, and Y. Zheng. "Methods for Pessimistic Bilevel Optimization." In: *Bilevel Optimization: Advances and Next Challenges.* Ed. by S. Dempe and A. Zemkoho. Springer International Publishing, 2020, pp. 403–420. DOI: `10.1007/978-3-030-52119-6_14`.

Locatelli, M. and F. Schoen. *Global Optimization.* Philadelphia, PA: Society for Industrial and Applied Mathematics, 2013. DOI: `10.1137/1.9781611972672`.

Luo, Z.-Q., J.-S. Pang, and D. Ralph. *Mathematical Programs with Equilibrium Constraints.* Cambridge University Press, 1996. DOI: 10.1017/CBO9780511983658.

Mallozzi, L. and J. Morgan. "Hierarchical Systems with Weighted Reaction Set." In: *Nonlinear Optimization and Applications.* Ed. by G. D. Pillo and F. Giannessi. Springer, Boston, MA, 1996, pp. 271–282. DOI: 10.1007/978-1-4899-0289-4_19.

Migdalas, A. "Bilevel programming in traffic planning: Models, methods and challenge." In: *Journal of Global Optimization* 7 (1995), 381—405. DOI: 10.1007/BF01099649.

Miranda, E. Álvarez, E. Fernández, and I. Ljubić. "The Recoverable Robust Facility Location Problem." In: *Transportation Research Part B: Methodological* 79 (2015), pp. 93–120. DOI: 10.1016/j.trb.2015.06.001.

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

Molan, I. and M. Schmidt. "Using neural networks to solve linear bilevel problems with unknown lower level." In: *Optimization Letters* 17 (2023), pp. 1083–1103. DOI: 10.1007/s11590-022-01958-7.

Molan, I., M. Schmidt, and J. Thürauf. *Learning the Follower's Objective Function in Sequential Bilevel Games.* 2023. URL: https://optimization-online.org/?p=23988.

Moore, J. T. and J. F. Bard. "The mixed integer linear bilevel programming problem." In: *Operations Research* 38.5 (1990), pp. 911–921. DOI: 10.1287/opre.38.5.911.

Motto, A. "On the exact solution of a class of Stackelberg games." In: *Proceedings of the 2005, American Control Conference, 2005.* Vol. 1. 2005, pp. 249–250. DOI: 10.1109/ACC.2005.1469940.

Nikoofal, M. E. and J. Zhuang. "Robust Allocation of a Defensive Budget Considering an Attacker's Private Information." In: *Risk Analysis* 32.5 (2011). DOI: 10.1111/j.1539-6924.2011.01702.x.

Nocedal, J. and S. J. Wright. *Numerical Optimization.* Springer, 2006. DOI: 10.1007/978-0-387-40065-5.

Ordóñez, F. and N. E. Stier-Moses. "Robust Wardrop Equilibrium." In: *Network Control and Optimization.* Ed. by T. Chahed and B. Tuffin. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 247–256. DOI: 10.1007/978-3-540-72709-5_26.

Ordóñez, F. and N. E. Stier-Moses. "Wardrop Equilibria with Risk-Averse Users." In: *Transportation Science* 44.1 (2010), pp. 63–86. DOI: 10.1287/trsc.1090.0292.

Patriksson, M. *The Traffic Assignment Problem: Models and Methods.* Courier Dover Publications, 2015.

Patriksson, M. and L. Wynter. "Stochastic mathematical programs with equilibrium constraints." In: *Operations Research Letters* 25.4 (1999), pp. 159–167. DOI: 10.1016/S0167-6377(99)00052-8.

Patriksson, M. and L. Wynter. *Stochastic Nonlinear Bilevel Programming.* 1997. URL: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.16.5155.

Pineda, S. and J. M. Morales. "Solving Linear Bilevel Problems Using Big-Ms: Not All That Glitters Is Gold." In: *IEEE Transactions on Power Systems* (2019). DOI: 10.1109/TPWRS.2019.2892607.

Pita, J., M. Jain, F. Ordóñez, M. Tambe, S. Kraus, and R. Magori-Cohen. "Effective solutions for real-world Stackelberg games: When agents must deal with human uncertainties." In: *AAMAS 2009—Proceedings of the 8th International Conference* *on Autonomous Agents and Multiagent Systems*. Vol. 1. 2009, pp. 369–376. DOI: 10.1017/CBO9780511973031.010.

Pita, J., M. Jain, M. Tambe, F. Ordóñez, and S. Kraus. “Robust solutions to Stackelberg games: Addressing bounded rationality and limited observations in human cognition.” In: *Artificial Intelligence* 174.15 (2010), pp. 1142–1171. DOI: 10.1016/j.artint.2010.07.002.

Pita, J., C. Portway, M. Tambe, P. Paruchuri, and S. Kraus. “Robust Solutions in Stackelberg Games: Addressing Boundedly Rational Human Preference Models.” In: *AAAI 2008—4th Multidisciplinary Workshop on Advances in Preference Handling*. 2008. URL: https://www.aaai.org/Papers/Workshops/2008/WS-08-09/WS08-09-010.pdf.

Rubinstein, A. *Modeling Bounded Rationality*. The MIT Press, 1998. URL: https://mitpress.mit.edu/books/modeling-bounded-rationality.

Salas, D. and A. Svensson. “Existence of Solutions for Deterministic Bilevel Games under a General Bayesian Approach.” In: *SIAM Journal on Optimization* 33.3 (2023), pp. 2311–2340. DOI: 10.1137/21M1442164.

Shapiro, A., D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming*. Society for Industrial and Applied Mathematics, 2009. DOI: 10.1137/1.9780898718751.

Shi, X., O. Prokopyev, and T. K. Ralphs. “Mixed integer bilevel optimization with a k-optimal follower: a hierarchy of bounds.” In: *Mathematical Programming Computation* 15 (2023), pp. 1–51. DOI: 10.1007/s12532-022-00227-z.

Shieh, E., B. An, R. Yang, M. Tambe, C. Baldwin, J. DiRenzo, B. Maule, and G. Meyer. “PROTECT: An application of computational game theory for the security of the ports of the United States.” In: *AAAI 2012—Proceedings of the 26th National Conference on Artificial Intelligence*. 2012, pp. 2173–2179. URL: http://www.intsci.ac.cn/en/boan/papers/AAAI122.pdf.

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

Simon, H. A. “A Behavioral Model of Rational Choice.” In: *The Quarterly Journal of Economics* 69.1 (1955), pp. 99–118. URL: http://www.jstor.org/stable/1884852.

Simon, H. A. “Rational choice and the structure of the environment.” In: *Psychological Review* 63.2 (1956), pp. 129–138. DOI: 10.1037/h0042769.

Simon, H. A. “Theories of Bounded Rationality.” In: *Decision and Organization*. Ed. by C. B. McGuire and R. Radner. North Holland Publishing Company, 1972, pp. 161–176. URL: http://innovbfa.viabloga.com/files/Herbert_Simon___theories_of_bounded_rationality___1972.pdf.

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

Tahernejad, S. and T. K. Ralphs. *Valid Inequalities for Mixed Integer Bilevel Linear Optimization Problems*. Tech. rep. 2020. URL: https://engineering.lehigh.edu/sites/engineering.lehigh.edu/files/_DEPARTMENTS/ise/pdf/tech-papers/20/20T_013.pdf.

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

van Damme, E. and S. Hurkens. “Games with Imperfectly Observable Commitment.” In: *Games and Economic Behavior* 21.1–2 (1997), pp. 282–308. DOI: 10.1006/game.1997.0524.

Vicente, L., G. Savard, and J. Júdice. “Discrete Linear Bilevel Programming Problem.” In: *Journal of Optimization Theory and Applications* 89.3 (1996), pp. 597–614. DOI: 10.1007/BF02275351.

von Stackelberg, H. *Marktform und Gleichgewicht.* Springer, 1934.

von Stackelberg, H. *Theory of the market economy.* Oxford University Press, 1952.

Wardrop, J. G. “Some Theoretical Aspects of Road Traffic Research.” In: *Proceedings of the Institution of Civil Engineers* 1.3 (1952), 325—362. DOI: 10.1680/ipeds.1952.11259.

Wardrop, J. G. and J. I. Whitehead. “Correspondence. Some Theoretical Aspects of Road Traffic Research.” In: *ICE Proceedings: Engineering Divisions* 1.5 (1952), p. 767. DOI: 10.1680/ipeds.1952.11362.

Weninger, N. and R. Fukasawa. “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. 2023, pp. 438–452. DOI: 10.1007/978-3-031-32726-1_31.

Wiesemann, W., D. Kuhn, and M. Sim. “Distributionally Robust Convex Optimization.” In: *Operations Research* 6 (2014), pp. 1358–1376. DOI: 10.1287/opre.2014.1314.

Wiesemann, W., A. Tsoukalas, P. Kleniati, and B. Rustem. “Pessimistic Bilevel Optimization.” In: *SIAM Journal on Optimization* 23.1 (2013), pp. 353–380. DOI: 10.1137/120864015.

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

Xu, P. and L. Wang. “An exact algorithm for the bilevel mixed integer linear programming problem under three simplifying assumptions.” In: *Computers & Operations Research* 41 (2014), pp. 309–318. DOI: 10.1016/j.cor.2013.07.016.

Yang, R., B. Ford, M. Tambe, and A. Lemieux. “Adaptive Resource Allocation for Wildlife Protection against Illegal Poachers.” In: *AAMAS 2014—Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems.* 2014, 453—460. URL: https://teamcore.seas.harvard.edu/files/teamcore/files/2014_7_teamcore_aamas2014_yang.pdf.

Yanıkoğlu, I. and D. Kuhn. “Decision Rule Bounds for Two-Stage Stochastic Bilevel Programs.” In: *SIAM Journal on Optimization* 28.1 (2018), pp. 198–222. DOI: 10.1137/16M1098486.

Zare, M. H., J. S. Borrero, and B. Zeng. “A note on linearized reformulations for a class of bilevel linear integer problems.” In: *Annals of Operations Research* 272 (2019), pp. 99–117. DOI: 10.1007/s10479-017-2694-x.

Zare, M. H., O. A. Prokopyev, and D. Sauré. “On Bilevel Optimization with Inexact Follower.” In: *Decision Analysis* 17.1 (2020). DOI: 10.1287/deca.2019.0392.

Zeng, B. “A Practical Scheme to Compute the Pessimistic Bilevel Optimization Problem.” In: *INFORMS Journal on Computing* 32.4 (2020), pp. 1128–1142. DOI: 10.1287/ijoc.2019.0927.

Part II

Reprints of Published Journal Articles and Preprints

Article 1

A Brief Introduction to Robust Bilevel Optimization

Yasmine Beck, Ivana Ljubić, and Martin Schmidt
SIAG on Optimization Views and News (2022)
https://siagoptimization.github.io/assets/views/ViewsAndNews-30-2.pdf

A Brief Introduction to Robust Bilevel Optimization

Yasmine Beck, Ivana Ljubić, and Martin Schmidt

Abstract. Bilevel optimization is a powerful tool for modeling hierarchical decision making processes. However, the resulting problems are challenging to solve—both in theory and practice. Fortunately, there have been significant algorithmic advances in the field so that we can solve much larger and also more complicated problems today compared to what was possible to solve two decades ago. This results in more and more challenging bilevel problems that researchers try to solve today. In this article, we give a brief introduction to one of these more challenging classes of bilevel problems: bilevel optimization under uncertainty using robust optimization techniques. To this end, we briefly state different versions of uncertain bilevel problems that result from different levels of cooperation of the follower as well as on when the uncertainty is revealed. We highlight these concepts using an academic example and discuss recent results from the literature concerning complexity as well as solution approaches. Finally, we discuss that the sources of uncertainty in bilevel optimization are much richer than in single-level optimization and, to this end, introduce the concept of decision uncertainty.

1. Introduction

Bilevel optimization has its roots in economics and dates back to the seminal works by von Stackelberg (1934, 1952). It has been introduced in the field of mathematical optimization much later in the publications by Bracken and McGill (1973) as well as Candler and Norton (1977). We use bilevel optimization to model hierarchical decision making processes, typically with two players, which we refer to as the leader and the follower. Despite its intrinsic hardness (Hansen et al. 1992; Jeroslow 1985), several innovative works pushed the boundaries of computational bilevel optimization so that we can tackle some relevant practical applications today; see, e.g., Kleinert et al. (2021) for a recent survey on computational bilevel optimization as well as the annotated bibliography by Dempe (2020).

The main goal of this article is to give a brief introduction to some basic concepts of bilevel optimization problems under uncertainty. The field is still in its infancy but, nevertheless, due to its relevance in many practical applications, it is developing very fast. In classic, i.e., single-level, optimization, there are two major approaches to address uncertainty: stochastic optimization (Birge and Louveaux 2011; Kall and Wallace 1994) and robust optimization (Ben-Tal, El Ghaoui, et al. 2009; Ben-Tal and Nemirovski 1998; Bertsimas, Brown, et al. 2011; Soyster 1973). The same two paths have been followed as well in bilevel optimization starting from the 1990s on. However, the sources of uncertainty are much richer in bilevel optimization compared to single-level optimization. To make this more concrete, let us consider the linear optimization problem $\min_x \{c^\top x \colon Ax \geq b\}$. It can "only" be subject to uncertainty due to uncertainties in the problem's data c, A, and b. Throughout this article, we will refer to this setting as *data uncertainty*. Moreover, a bilevel

---

Date: January 24, 2023.

2020 *Mathematics Subject Classification.* 90-02, 90C70, 90C90, 91A65.

Key words and phrases. Bilevel optimization, Robust optimization, Optimization under uncertainty.

optimization problem may also be subject to an additional source of uncertainty, which is due to its nature that it combines two different decision makers in one model. Hence, there can be further uncertainty involved either if the leader is not sure about the reaction of the follower or if the follower is not certain about the observed leader's decision. We will denote this additional type of uncertainty as *decision uncertainty*. Obviously, decision uncertainty does not play any role in single-level optimization since only one decision maker is involved.

In this introductory article, we will solely focus on data uncertainty that is tackled using concepts from robust optimization. For more details regarding stochastic bilevel optimization, decision uncertainty, etc. we refer to our recent survey (Beck, Ljubić, et al. 2022a).

2. Problem Statement

We start by considering the deterministic bilevel problem (we explain the quotation marks below)

$$\text{“}\min_{x\in X}\text{”} \quad F(x,y) \tag{1a}$$

$$\text{s.t.} \quad G(x,y) \geq 0, \tag{1b}$$

$$y \in S(x), \tag{1c}$$

where $S(x)$ denotes the set of optimal solutions of the x-parameterized problem

$$\min_{y\in Y} \quad f(x,y) \tag{2a}$$

$$\text{s.t.} \quad g(x,y) \geq 0. \tag{2b}$$

Problem (1) is referred to as the upper-level (or the leader's) problem and Problem (2) is the so-called lower-level (or the follower's) problem. Moreover, we refer to $x \in X$ and $y \in Y$ as the leader's and the follower's variables, respectively. The sets $X \subseteq \mathbb{R}^{n_x}$ and $Y \subseteq \mathbb{R}^{n_y}$ can be used to include possible integrality constraints. The objective functions are given by $F, f : \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \to \mathbb{R}$ and the constraint functions by $G : \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \to \mathbb{R}^m$ as well as $g : \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \to \mathbb{R}^\ell$. In the case that the lower-level problem does not have a unique solution, the bilevel problem (1) and (2) is ill-posed. This ambiguity is expressed by the quotation marks in (1a). To overcome this issue, it is common to pursue either an optimistic or a pessimistic approach to bilevel optimization; see, e.g., Dempe (2002). In the optimistic setting, the leader chooses the follower's response among multiple optimal solutions of the lower-level problem such that it favors the leader's objective function value. Hence, the leader also minimizes her[1] objective in the y variables, i.e., we consider the problem

$$\min_{x\in\bar{X}} \ \min_{y\in S(x)} \quad F(x,y) \tag{3}$$

with $\bar{X} := \{x \in X : G(x) \geq 0\}$ and $G : \mathbb{R}^{n_x} \to \mathbb{R}^m$. Here and in what follows, we focus on the setting without coupling constraints, i.e., without upper-level constraints that depend on the variables y. In the pessimistic setting, the leader anticipates that, among multiple optimal solutions of the follower, the worst possible response w.r.t. the upper-level objective function will be chosen by the follower. Thus, one studies the problem

$$\min_{x\in\bar{X}} \ \max_{y\in S(x)} \quad F(x,y).$$

In this article, we focus on bilevel problems of the above form which are additionally affected by data uncertainty.

---

[1] Throughout this article, we use “her” for the leader and “his” for the follower.

2.1. **Data Uncertainty.** Data uncertainty arises when some of the players only have access to inaccurate or incomplete data. In robust optimization, it is assumed that these uncertainties take values in a given, and usually compact, uncertainty set $\mathcal{U}$. The uncertainty sets are typically modeled using boxes, polyhedra, ellipsoids, or cones; see, e.g., Ben-Tal, El Ghaoui, et al. (2009), Ben-Tal, Goryashko, et al. (2004), Ben-Tal and Nemirovski (1998), Bertsimas, Brown, et al. (2011), and Soyster (1973). In the context of single-level robust optimization, there are two possibilities to hedge against data uncertainty.

First, assuming that the coefficients of the objective function are uncertain, one searches for a solution that is optimal for the worst-case realization of the uncertain parameters. The problem can be modeled as

$$\min_{x \in \bar{X}} \max_{u \in \mathcal{U}} \quad F(x, u), \tag{4}$$

where the objective function $F : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \to \mathbb{R}$ is continuous and the sets $\mathcal{U} \subseteq \mathbb{R}^{n_u}$ and $\bar{X}$ are defined as above.

Second, in the case that the uncertainty affects the coefficients of the constraints, one is interested in a solution that is feasible for all possible realizations of the uncertainty. This problem can be stated as

$$\min_{x \in X} \quad F(x) \quad \text{s.t.} \quad G(x, u) \geq 0 \quad \text{for all } u \in \mathcal{U}, \tag{5}$$

where both the objective function $F : \mathbb{R}^{n_x} \to \mathbb{R}$ and the constraint function $G : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \to \mathbb{R}^m$ are continuous. Problem (5) can be reformulated as

$$\min_{x \in X} \quad F(x) \quad \text{s.t.} \quad \min \{G(x, u) \colon u \in \mathcal{U}\} \geq 0. \tag{6}$$

In particular, Problem (4) can be restated as an instance of Problem (6) using an epigraph reformulation, i.e.,

$$\min_{x \in \bar{X}, t \in \mathbb{R}} \quad t \quad \text{s.t.} \quad t \geq \max \{F(x, u) \colon u \in \mathcal{U}\}.$$

Note that for the two settings discussed so far, a single decision maker has to take a here-and-now decision before the uncertainty is revealed. In bilevel optimization, however, there are two different timings that are possible—one in which the uncertainty realizes after and one in which the uncertainty realizes before the follower takes his decision.

2.1.1. *Here-and-Now Follower.* In this case, both the leader and the follower have to make their decisions before the uncertainty is revealed, i.e., one considers the timing

$$\text{leader } x \quad \curvearrowright \quad \text{follower } y = y(x) \quad \curvearrowright \quad \text{uncertainty } u. \tag{7}$$

This means that the leader anticipates an optimal response of the follower who hedges against data uncertainty. Hence, the lower-level problem is an x-parameterized problem in which we can embed any of the concepts known for single-level optimization under uncertainty. For instance, if only the lower-level objective function is uncertain and the follower is assumed to behave in an optimistic way, we are solving Problem (3) with

$$S(x) := \operatorname*{arg\,min}_{y' \in Y} \left\{ \max_{u \in \mathcal{U}} f(x, u, y') \colon g(x, y') \geq 0 \right\}.$$

2.1.2. *Wait-and-See Follower.* In this setting, the leader first takes a here-and-now decision, i.e., without knowing the realization of uncertainty. Then, the uncertainty is revealed and, finally, the follower decides in a wait-and-see fashion, taking the leader's decision as well as the realization of the uncertainty into account. Hence, one considers the timing

$$\text{leader } x \quad \curvearrowright \quad \text{uncertainty } u \quad \curvearrowright \quad \text{follower } y = y(x,u). \tag{8}$$

This means that the leader does not have full knowledge about the lower-level problem. Thus, she wants to hedge against the worst-case reaction of the follower. Here, "worst-case" may not only imply the robustness of the leader w.r.t. lower-level data uncertainty but also her conservatism regarding the cooperation of the follower. For instance, to protect against the worst-case realization of the uncertainties w.r.t. the leader's objective function, we consider the problem

$$\text{“}\min_{x \in \bar{X}} \max_{u \in \mathcal{U}}\text{”} \quad F(x,y) \quad \text{s.t.} \quad y \in S(x,u), \tag{9}$$

where $S(x,u)$ is the set of optimal solutions of the (x,u)-parameterized problem

$$\min_{y \in Y} \quad f(x,u,y) \quad \text{s.t.} \quad g(x,u,y) \geq 0.$$

The quotation marks in (9) express the ill-posedness of the bilevel problem in the case that the set $S(x,u)$ is not a singleton. Hence, one also needs to distinguish between the optimistic and the pessimistic case in the robust setting. Indeed, both situations can be motivated by practical applications. For instance, the pessimistic robust bilevel problem appears when the leader wants to hedge against the worst-case both w.r.t. lower-level data uncertainty as well as w.r.t. the potentially unknown level of cooperation of the follower. On the other hand, there may also be situations in which the follower still hedges against his uncertainties in a robust way but, in the case of ambiguous optimal solutions, acts in an optimistic way. This might be the case in energy markets with sufficiently regulated agents, where a strong level of regulation might lead to an optimistic robust bilevel problem.

3. AN ACADEMIC EXAMPLE

Let us consider the linear bilevel problem taken from Beck, Ljubić, et al. (2022a) that is given by

$$\begin{aligned} \text{“}\min_{x \in \mathbb{R}}\text{”} \quad & F(x,y) = x + y && \text{(10a)}\\ \text{s.t.} \quad & x - y \geq -1, && \text{(10b)}\\ & 3x + y \geq 3, && \text{(10c)}\\ & y \in S(x), && \text{(10d)} \end{aligned}$$

where $S(x)$ denotes the set of optimal solutions of the x-parameterized lower level

$$\begin{aligned} \min_{y \in \mathbb{R}} \quad & f(x,y) = -0.1y && \text{(11a)}\\ \text{s.t.} \quad & -2x + y \geq -7, && \text{(11b)}\\ & -3x - 2y \geq -14, && \text{(11c)}\\ & 0 \leq y \leq 2.5. && \text{(11d)} \end{aligned}$$

The problem is depicted in Figure 1 (left). The upper- and lower-level constraints are represented with dashed and solid lines, respectively. The optimal solution $(x^*, y^*) = (1.5, 2.5)$ is the same for both the optimistic and the pessimistic setting and it is illustrated by the thick dot. Suppose now that the lower-level objective function is uncertain. To this end, we consider $\tilde{f}(x,u,y) = (-0.1 + u)y$ and assume that u only takes values in the uncertainty set $\mathcal{U} = \{u \in \mathbb{R} \colon |u| \leq 0.5\}$. In what follows,

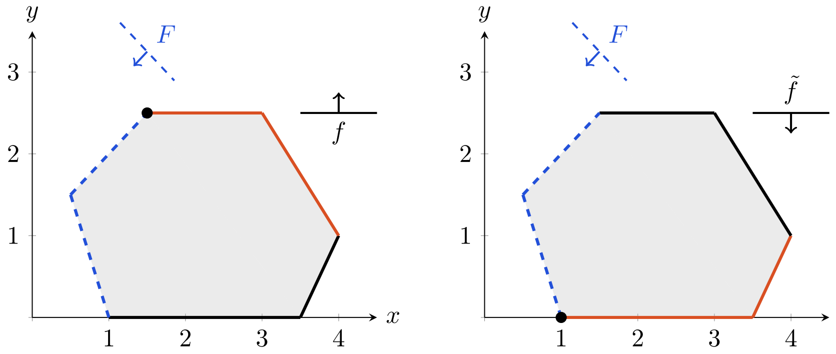

Figure 1. Both figures show the upper-level constraints (dashed blue lines), the lower-level constraints (solid black and orange lines), the shared constraint set (gray area), and the bilevel feasible set (solid orange lines) of the bilevel problem (10) and (11). The deterministic variant of the problem is depicted on the left and the variant with a here-and-now follower is given on the right.

we distinguish between a follower taking a here-and-now or a wait-and-see decision to illustrate how the considered timing may affect the solution of the problem.

3.1. **Here-and-Now Follower.** We first consider the timing in (7). The robustified lower-level problem is thus given by

$$\min_{y\in\mathbb{R}} \max_{u\in\mathcal{U}} \quad \tilde{f}(x,u,y) = (-0.1+u)y \quad \text{s.t.} \quad \text{(11b)–(11d)}.$$

Using classic techniques from robust optimization, we obtain a modified gradient of the lower-level objective function, which is shown in Figure 1 (right). The optimal solution $(x^*, y^*) = (1, 0)$ of this problem is represented by the thick dot. In particular, there is a unique lower-level response for every feasible x, which is why we do not need to distinguish between the optimistic and the pessimistic case.

3.2. **Wait-and-See Follower.** We now consider the timing in (8), i.e., the overall robustified bilevel problem reads

$$\text{“}\min_{x\in\mathbb{R}} \max_{u\in\mathcal{U}}\text{”} \quad F(x,y) \quad \text{s.t.} \quad \text{(10b)–(10c)}, \; y \in S(x,u),$$

where $S(x,u)$ is the set of optimal solutions of the (x,u)-parameterized lower level

$$\min_{y\in\mathbb{R}} \quad \tilde{f}(x,u,y) = (-0.1+u)y \quad \text{s.t.} \quad \text{(11b)–(11d)}.$$

To solve this problem, we need to distinguish the following three cases.

(i) $-0.5 \le u < 0.1$: This case corresponds to the setting that is depicted in Figure 1 (left). The optimal follower's reaction is thus given by

$$y(x,u) = \begin{cases} 2.5, & x \le 3, \\ -1.5x + 7, & 3 \le x \le 4. \end{cases} \tag{12}$$

Note, however, that the bilevel problem is infeasible for $x < 1.5$. In particular, this means that the robust optimal leader's decision $x^* = 1$ for the case with a here-and-now follower is no longer bilevel feasible if the follower decides in a wait-and-see fashion.

(ii) $u = 0.1$: Any feasible decision of the follower, i.e., any $y \in \mathbb{R}$ that satisfies (11b)–(11d), is optimal for the x-parameterized lower level. Hence, the distinction between an optimistic and a pessimistic follower is necessary. In the optimistic setting, the follower would react with

$$y(x,u) = \begin{cases} 0, & x \leq 3.5, \\ 2x - 7, & 3.5 \leq x \leq 4. \end{cases} \tag{13}$$

This corresponds to the setting that is depicted in Figure 1 (right). A pessimistic follower, however, would select (12). Note that the bilevel problem with an optimistic follower turns out to be infeasible for $x < 1$ and, again, the problem is infeasible for $x < 1.5$ if a pessimistic follower is considered.

(iii) $0.1 < u \leq 0.5$: The optimal follower's reaction is given by (13). Again, the overall bilevel problem turns out to be infeasible for $x < 1$.

To determine an optimal solution of the bilevel problem (10) and (11) with a wait-and-see follower, we thus consider the worst-case realization of each of the previous three cases w.r.t. the leader's decision x. Hence, we need to solve

$$\min_{x} \quad \hat{F}(x) \quad \text{s.t.} \quad 1.5 \leq x \leq 4 \tag{14}$$

with the piecewise-linear function

$$\hat{F}(x) = \begin{cases} x + 2.5, & 1.5 \leq x \leq 3, \\ -0.5x + 7, & 3 \leq x \leq 4. \end{cases}$$

In particular, the solution $x^* = 1.5$ of Problem (14) is an optimal decision of the leader in both the optimistic and the pessimistic setting. After observing the realization of the uncertainty, the corresponding response of the follower is then given by

$$y^*_{\text{o}}(x^*, u) = \begin{cases} 2.5, & -0.5 \leq u < 0.1, \\ 0, & 0.1 \leq u \leq 0.5 \end{cases}$$

in the optimistic setting, whereas, for the pessimistic case, we have

$$y^*_{\text{p}}(x^*, u) = \begin{cases} 2.5, & -0.5 \leq u \leq 0.1, \\ 0, & 0.1 < u \leq 0.5. \end{cases}$$

Note that, if $u \in [-0.5, 0.1)$ realizes, at the point $x^* = 1.5$, the deterministic solution $(x^*, y(x^*))$ and the robust bilevel solutions $(x^*, y(x^*, u))$ coincide. However, the optimal follower's response $y(x^*, u)$ in the robust setting may change significantly for $u \geq 0.1$.

4. SELECTED RESULTS FROM THE LITERATURE

The field of robust bilevel optimization is still in its infancy. For a detailed discussion of existing modeling and solution approaches, we refer to our recent survey (Beck, Ljubić, et al. 2022a). In deterministic bilevel optimization, a standard solution approach is to reformulate the problem as a classic, i.e., single-level, problem. This can be done, e.g., by replacing the lower level with its Karush–Kuhn–Tucker (KKT) conditions (Fortuny-Amat and McCarl 1981). The same holds true for robust bilevel problems whenever the robust counterpart of the lower-level problem can be reformulated as a deterministic problem for which the KKT conditions are necessary and sufficient. However, these reformulation techniques cannot be applied anymore if discrete variables are introduced in the lower level. Due to their intrinsic hardness, approaches for discrete robust bilevel problems have not been investigated a lot up to now. In single-level optimization, the knapsack problem is one of the most thoroughly studied discrete optimization problem due to its relevance both in theory and practice; see, e.g., Pisinger and Toth (1998). Bilevel knapsack problems naturally extend their single-level counterparts such as to capture hierarchical and, in particular, competitive settings (Caprara et al. 2013; Della Croce and Scatamacchia 2020; DeNegre 2011; Fischetti, Ljubić, et al. 2019; Fischetti, Monaci, et al. 2018). Moreover, the bilevel knapsack interdiction problem is commonly used as a benchmark for testing bilevel optimization solvers; see, e.g., DeNegre and Ralphs (2009) and Tang et al. (2016). It is thus not surprising that bilevel knapsack problems are also among the first discrete bilevel problems studied under uncertainty—both in terms of complexity questions and solution approaches. The remainder of this section is thus dedicated to a brief overview of recent results from the literature for robust bilevel knapsack problems.

4.1. **Complexity Results for Robust Continuous Bilevel Knapsack Problems with a Wait-and-See Follower.** We start by considering the robust continuous bilevel knapsack problem with an uncertain lower-level objective, i.e., we consider the problem

$$\max_{x\in[x^-,x^+]} \quad \min_{c\in\mathcal{U},y\in\mathbb{R}^n} \quad d^\top y \tag{15a}$$

$$\text{s.t.} \quad y \in \arg\max_{y'} \left\{c^\top y' \colon a^\top y' \leq x,\, 0 \leq y' \leq 1\right\} \tag{15b}$$

with x^-, $x^+ \in \mathbb{R}$, $x^- \leq x^+$, a, c, $d \in \mathbb{R}^n_{\geq 0}$, and an uncertainty set $\mathcal{U} \subseteq \mathbb{R}^n$. In this setting, the leader first decides on the knapsack's capacity x. Then, the uncertainties regarding the lower-level objective function coefficients realize. Finally, the follower solves a knapsack problem according to the realization of his own profits, which may differ from those of the leader. Hence, the follower decides in a wait-and-see fashion, i.e., the timing in (8) is considered. The leader's aim is to choose the capacity of the knapsack in such a way that her own profit of the items packed by the follower is maximized. Whenever the follower's choice of items is not unique, the pessimistic approach is considered. The deterministic variant of Problem (15) can be solved in polynomial time, which makes it a good starting point to address the question of how uncertainties may affect the hardness of the underlying bilevel problem.

Driven by this question, Buchheim and Henke (2020, 2022) show that the complexity of Problem (15) strongly depends on the considered type of the uncertainty set. For discrete uncertainty sets as well as for interval uncertainty under the independence assumption, i.e., for the case in which the follower's objective function coefficients independently take values in given intervals, Problem (15) remains solvable in polynomial time. However, the problem becomes NP-hard if the uncertainty set is the Cartesian product of discrete sets. In particular, this shows that replacing the uncertainty set by its convex hull may significantly change the problem, which is very much in contrast to the situation in single-level robust optimization. NP-hardness is also shown for the variants of the problem with polytopal uncertainty sets and uncertainty sets that are defined by a p-norm with $p \in [1, \infty)$. In particular, for all NP-hard variants of the problem, even the evaluation of the leader's objective function is NP-hard.

As a generalization of the aforementioned works, Buchheim, Henke, and Hommelsheim (2021) are concerned with complexity questions for robust bilevel combinatorial problems of the form

$$\text{“}\max_{x\in X} \; \min_{c\in\mathcal{U}}\text{”} \quad d^\top y \tag{16a}$$

$$\text{s.t.} \quad y \in \arg\max_{y'\in\mathbb{R}^{n_y}} \left\{c^\top y' \colon By' \leq Ax + b\right\}. \tag{16b}$$

with $X \subseteq \{0,1\}^{n_x}$, $A \in \mathbb{R}^{m \times n_x}$, $B \in \mathbb{R}^{m \times n_y}$, c, $d \in \mathbb{R}^{n_y}$, and $b \in \mathbb{R}^m$. Again, it is assumed that the lower-level objective function coefficients are uncertain, that the uncertainties take values in a given uncertainty set $\mathcal{U} \subseteq \mathbb{R}^{n_y}$, and that the follower decides in a wait-and-see fashion. As before, the quotation marks in (16a) express the ambiguity in the case that the lower level does not have a unique solution. The deterministic variant of Problem (16) is known to be NP-easy.[2] However, it is shown that interval uncertainty renders Problem (16) significantly harder than the consideration of discrete uncertainty sets. More precisely, the robust counterpart can be Σ_2^P-hard[3] for interval uncertainty under the independence assumption, whereas it can be NP-hard for uncertainty sets $\mathcal{U}$ with $|\mathcal{U}| = 2$ and strongly NP-hard for general discrete uncertainty sets. In particular, it is shown that replacing the discrete uncertainty set by its convex hull may increase the complexity of the problem at hand, which is in line with the results in Buchheim and Henke (2020, 2022).

4.2. **Solution Approaches for the Bilevel Knapsack Interdiction Problem with a Here-and-Now Follower.** Beck, Ljubić, et al. (2022b) study discrete linear min-max problems with uncertainties regarding the lower-level objective function coefficients. In contrast to the aforementioned works, which all follow the notion of strict robustness, the authors consider a Γ-robust approach (Bertsimas and Sim 2003, 2004). The problem under consideration thus reads

$$\min_{x} \quad c^\top x + d^\top y \tag{17a}$$

$$\text{s.t.} \quad Ax \geq a, \; x \in X \subseteq \mathbb{Z}^{n_x}, \tag{17b}$$

$$y \in \operatorname*{arg\,max}_{y' \in Y(x)} \left\{ d^\top y' - \max_{\{S \subseteq [n_y] \colon |S| \leq \Gamma\}} \sum_{i \in S} \Delta d_i y'_i \right\}, \tag{17c}$$

where $\Gamma \in [n_y] := \{1, \dots, n_y\}$ and $Y(x) \subseteq \mathbb{Z}_+^{n_y}$ denotes the lower-level feasible set. Here, the timing in (7) is considered, i.e., both the leader and the follower decide before the uncertainty realizes. The authors present two approaches to reformulate Problem (17). The first approach is based on an extended formulation, whereas the second one exploits the fact that Problem (17) can be interpreted as a single-leader multi-follower problem with independent followers. Based on these reformulations, the authors propose generic branch-and-cut frameworks to solve the problem. Moreover, it is shown that the same techniques can also be used for the case in which uncertainties only arise in a single packing-type constraint on the lower level. To assess the applicability of the proposed branch-and-cut methods, the authors focus on the Γ-robust knapsack interdiction problem (Caprara et al. 2016). In this setting, both players share a common set of items and the leader has the ability to influence the follower's decision by prohibiting the usage of certain items by the follower. The authors derive problem-tailored cuts and perform a computational study on 200 robustified knapsack interdiction instances with up to 55 items, i.e., with up to 55 variables on both the upper and the lower level.

5. A First Glimpse at Decision Uncertainty

Although being subject to data uncertainty, both decision makers in the bilevel problem are assumed to take perfectly rational decisions in the sense that they can perfectly anticipate or observe the other's decision and that they can solve their problem to global optimality. In decision making theory, however, it is well known

---

[2] A decision problem is NP-easy if it can be polynomially reduced to an NP-complete decision problem (Buchheim, Henke, and Hommelsheim 2021).

[3] This class contains those problems that can be solved in nondeterministic polynomial time, provided that there exists an oracle that solves problems that are in NP in constant time.

that these assumptions regarding perfect information and rationality are rarely satisfied in a real-world context. Luckily, bilevel optimization under uncertainty allows to relax these assumptions in multiple ways. Throughout this article, we assumed that the major source of uncertainty stems from unknown or noisy input data. However, bilevel optimization involves (at least) two decision makers and, hence, other uncertainties in the decision making process are also possible. Another possible one is *decision uncertainty* in which, e.g., the leader is not sure about the reaction of the follower (for instance if the follower does not necessarily choose an optimal solution) or in which the follower is not sure about the observed leader's decision. We are not going into the details here but want to give a few pointers to the relevant literature that covers such aspects. If the leader is uncertain about her anticipation of the follower's optimal reaction and, thus, may want to hedge against sub-optimal follower reactions, the resulting setup can be modeled using so-called near-optimal robust bilevel models; see, e.g., Besançon et al. (2019). As an extreme case of the former aspect it may be the case that the upper-level player knows that the follower will play against her. This is the setting of a pessimistic bilevel optimization problem, which is also rather naturally connected to the field of robust optimization; see, e.g., Wiesemann et al. (2013). However, if the level of cooperation or confrontation of the follower is not known, this leads to intermediate cases in between of the optimistic and the pessimistic case; see, e.g., Aboussoror and Loridan (1995) and Mallozzi and Morgan (1996). Moreover, in many situations it is not possible for the follower to perfectly observe the optimal decision of the leader and the follower thus may want to hedge against all possible leader decisions in some uncertainty set around the observation. Such settings are tackled in, e.g., Bagwell (1995), Beck and Schmidt (2021), and van Damme and Hurkens (1997). Finally, even if all data and the rational reaction of the follower is known and even if the leader can, in principle, fully anticipate the (globally) optimal reaction of the follower, it might still be the case that limited intellectual or computational resources render it impossible for the follower to take a globally optimal decision. In such situations, a follower might resort to heuristic approaches and the leader may be uncertain w.r.t. which heuristic is used. For a good primer in this context, we refer to the recent paper by Zare et al. (2020).

The above list is by far not comprehensive. A much more detailed discussion of these and other aspects can be found in our recent survey (Beck, Ljubić, et al. 2022a). However, it is hopefully clear now how much more diverse the sources of uncertainty can be in bilevel optimization as compared to single-level optimization. Hence, we expect a lot of research in this area in future years.

References

Aboussoror, A. and P. Loridan (1995). "Strong-weak Stackelberg Problems in Finite Dimensional Spaces." In: *Serdica. Mathematical Journal* 21, pp. 151–170. URL: `https://www.semanticscholar.org/paper/Strong-weak-Stackelberg-Problems-in-Finite-Spaces-Aboussoror-Loridan/c875814905ed516f7e693d9040683c1115f31c27`.

Bagwell, K. (1995). "Commitment and observability in games." In: *Games and Economic Behavior* 8.2, pp. 271–280. DOI: `10.1016/S0899-8256(05)80001-6`.

Beck, Y., I. Ljubić, and M. Schmidt (2022a). *A Survey on Bilevel Optimization Under Uncertainty*. Tech. rep. URL: `https://optimization-online.org/2022/06/8963/`.

Beck, Y., I. Ljubić, and M. Schmidt (2022b). *Exact Methods for Discrete Γ-Robust Interdiction Problems*. Tech. rep. URL: `http://www.optimization-online.org/DB_FILE/2021/11/8678.pdf`.

Beck, Y. and M. Schmidt (2021). “A robust approach for modeling limited observability in bilevel optimization.” In: *Operations Research Letters* 49.5, pp. 752–758. DOI: `10.1016/j.orl.2021.07.010`.

Ben-Tal, A., L. El Ghaoui, and A. Nemirovski (2009). *Robust Optimization*. Princeton University Press.

Ben-Tal, A., A. Goryashko, E. Guslitzer, and A. Nemirovski (2004). “Adjustable robust solutions of uncertain linear programs.” In: *Mathematical Programming* 99.2, pp. 351–376. DOI: `10.1007/s10107-003-0454-y`.

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`.

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`.

Besançon, M., M. F. Anjos, and L. Brotcorne (2019). *Near-optimal Robust Bilevel Optimization*. URL: `https://arxiv.org/pdf/1908.04040.pdf`.

Birge, J. R. and F. Louveaux (2011). *Introduction to Stochastic Programming*. Springer Science & Business Media. DOI: `10.1007/978-1-4614-0237-4`.

Bracken, J. and J. T. McGill (1973). “Mathematical programs with optimization problems in the constraints.” In: *Operations Research* 21.1, pp. 37–44. DOI: `10.1287/opre.21.1.37`.

Buchheim, C. and D. Henke (2020). *The bilevel continuous knapsack problem with uncertain follower’s objective*. URL: `https://arxiv.org/abs/1903.02810`.

Buchheim, C. and D. Henke (2022). “The robust bilevel continuous knapsack problem with uncertain coefficients in the follower’s objective.” In: *Journal of Global Optimization* 83, pp. 803–824. DOI: `10.1007/s10898-021-01117-9`.

Buchheim, C., D. Henke, and F. Hommelsheim (2021). “On the Complexity of Robust Bilevel Optimization With Uncertain Follower’s Objective.” In: *Operations Research Letters* 49.5, pp. 703–707. DOI: `10.1016/j.orl.2021.07.009`.

Candler, W. and R. Norton (1977). *Multi-level Programming*. Discussion Papers, Development Research Center, International Bank for Reconstruction and Development. World Bank. URL: `https://books.google.de/books?id=TicmAQAAMAAJ`.

Caprara, A., M. Carvalho, A. Lodi, and G. J. Woeginger (2013). “A Complexity and Approximability Study of the Bilevel Knapsack Problem.” In: *Integer Programming and Combinatorial Optimization*. Ed. by M. Goemans and J. Correa. Vol. 7801. IPCO 2013. Springer, Berlin, Heidelberg, pp. 98–109. DOI: `10.1007/978-3-642-36694-9_9`.

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`.

Della Croce, F. and R. Scatamacchia (2020). “An exact approach for the bilevel knapsack problem with interdiction constraints and extensions.” In: *Mathematical Programming* 183, pp. 249–281. DOI: `10.1007/s10107-020-01482-5`.

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.

DeNegre, S. T. and T. K. Ralphs (2009). "A Branch-and-cut Algorithm for Integer Bilevel Linear Programs." In: *Operations research and cyber-infrastructure*. Springer, pp. 65–78. DOI: 10.1007/978-0-387-88843-9_4.

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.

Fortuny-Amat, J. and B. McCarl (1981). "A Representation and Economic Interpretation of a Two-Level Programming Problem." In: *The Journal of the Operational Research Society* 32.9, pp. 783–792. DOI: 10.1057/jors.1981.156.

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.

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 of Computational Optimization* 9, p. 100007. DOI: 10.1016/j.ejco.2021.100007.

Mallozzi, L. and J. Morgan (1996). "Hierarchical Systems with Weighted Reaction Set." In: *Nonlinear Optimization and Applications*. Ed. by G. D. Pillo and F. Giannessi. Springer, Boston, MA, pp. 271–282. DOI: 10.1007/978-1-4899-0289-4_19.

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.

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.

Tang, Y., J. P. Richard, and J. C. Smith (2016). "A class of algorithms for mixed-integer bilevel min–max optimization." In: *Journal of Global Optimization* 66, pp. 225–262. DOI: 10.1007/s10898-015-0274-7.

van Damme, E. and S. Hurkens (1997). "Games with Imperfectly Observable Commitment." In: *Games and Economic Behavior* 21.1–2, pp. 282–308. DOI: 10.1006/game.1997.0524.

von Stackelberg, H. (1934). *Marktform und Gleichgewicht*. Springer.

von Stackelberg, H. (1952). *Theory of the market economy*. Oxford University Press.

Wiesemann, W., A. Tsoukalas, P. Kleniati, and B. Rustem (2013). "Pessimistic Bilevel Optimization." In: *SIAM Journal on Optimization* 23.1, pp. 353–380. DOI: 10.1137/120864015.

Zare, M. H., O. A. Prokopyev, and D. Sauré (2020). "On Bilevel Optimization with Inexact Follower." In: *Decision Analysis* 17.1. DOI: 10.1287/deca.2019.0392.

(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`

Article 2

A survey on bilevel optimization under uncertainty

Yasmine Beck, Ivana Ljubić, and Martin Schmidt
European Journal of Operational Research (2023)
https://doi.org/10.1016/j.ejor.2023.01.008

A SURVEY ON BILEVEL OPTIMIZATION UNDER UNCERTAINTY

YASMINE BECK, IVANA LJUBIĆ, AND MARTIN SCHMIDT

ABSTRACT. Bilevel optimization is a very active field of applied mathematics. The main reason is that bilevel optimization problems can serve as a powerful tool for modeling hierarchical decision making processes. This ability, however, also makes the resulting problems challenging to solve—both in theory and practice. Fortunately, there have been significant algorithmic advances in the field of bilevel optimization so that we can solve much larger and also more complicated problems today compared to what was possible to solve two decades ago. This results in more and more challenging bilevel problems that researchers try to solve today. This survey gives a detailed overview of one of these more challenging classes of bilevel problems: bilevel optimization under uncertainty. We review the classic ways of addressing uncertainties in bilevel optimization using stochastic or robust techniques. Moreover, we highlight that the sources of uncertainty in bilevel optimization are much richer than for usual, i.e., single-level, problems since not only the problem's data can be uncertain but also the (observation of the) decisions of the two players can be subject to uncertainty. We thus also review the field of bilevel optimization under limited observability, the area of problems considering only near-optimal decisions, and discuss intermediate solution concepts between the optimistic and pessimistic cases. Finally, we also review the rich literature on applications studied using uncertain bilevel problems such as in energy, for interdiction games and security applications, in management sciences, and networks.

1. INTRODUCTION

Bilevel optimization is a rather young field of research that dates back to the early publications by Bracken and McGill (1973) as well as Candler and Norton (1977), having its game-theoretic foundations dating back to the seminal works by von Stackelberg (1934, 1952). While being very powerful modeling tools that allow to consider hierarchical decision making processes, bilevel optimization models are also very hard to solve—both in theory and practice. For instance, NP-hardness (Jeroslow 1985) and strong NP-hardness (Hansen et al. 1992) have been shown in the 1980s and early 1990s. The intrinsic hardness of bilevel optimization leads to the fact that, on the one hand, the field has been propelled theoretically first (see, e.g., the seminal textbook by Dempe (2002) and the more recent book by Dempe, Kalashnikov, et al. (2015)) but that computational bilevel optimization still has been in its infancy until the late 2000s. Since then, several innovative works pushed the computational study of these problems so that we can solve relevant practical instances of realistic size today; see Kleinert et al. (2021) for a very recent survey on this and related topics as well as the annotated bibliography on bilevel optimization by Dempe (2020).

Rather naturally, the operations research, mathematics, engineering, and economics communities, which all use bilevel optimization to model and solve real-world problems in their respective fields, started to study more and more complicated

---

Date: November 19, 2022.

2020 *Mathematics Subject Classification.* 90-02, 90C70, 90C90, 91A26, 91A65.

Key words and phrases. (R) Uncertainty modelling, Bilevel optimization, Optimization under uncertainty, Bounded rationality, Survey.

bilevel problems. These more complicated problems are even harder than "usual", e.g., continuous and maybe linear, bilevel problems since they introduce different aspects that make the resulting problems both more challenging in theory and practice—among others: (i) mixed-integer aspects, in particular in the lower-level problem, (ii) nonlinearities and nonconvexities both in the upper- and the lower-level problem, or (iii) large-scale instances that could not have been solved a few decades ago.

Additionally, and this leads us to the main topic of this survey, more and more researchers from all of the above mentioned communities started to study bilevel optimization problems under uncertainty. In classic, i.e., single-level, optimization, there have been mainly two paths to address uncertainty in optimization models: stochastic optimization (Birge and Louveaux 2011; Kall and Wallace 1994) and robust optimization (Ben-Tal, El Ghaoui, et al. 2009; Ben-Tal and Nemirovski 1998; Bertsimas, Brown, et al. 2011; Soyster 1973). The same two paths have been followed as well in bilevel optimization starting from the 1990s on.

However, the sources of uncertainty are much richer in bilevel optimization compared to usual, i.e., single-level, optimization. To make this more concrete, a linear optimization problem $\min\{c^\top x \colon Ax \geq b\}$ can only be subject to uncertainty due to uncertainties in the problem's data c, A, and b. Throughout this survey, we will denote this setting as *data uncertainty*. Moreover, bilevel optimization may be subject to an additional source of uncertainty, which is due to its nature that combines two different decision makers in one model. Hence, there can be further uncertainty involved either if the leader is not sure about the reaction of the follower or if the follower is not certain about the observed leader's decision. We will denote this additional type of uncertainty as *decision uncertainty*. Obviously, decision uncertainty does not play any role in single-level optimization since only one decision maker is involved.

Both data as well as decision uncertainty can—and maybe should—be considered under the wider umbrella of what economists call *bounded rationality*; see, e.g., Rubinstein (1998) and Simon (1972). In economics, *rationality* usually means that the different agents of a system (e.g., the leader and the follower in bilevel optimization or in a Stackelberg game) act as optimizers, meaning that they all implement a fully rational decision process that consists in solving an optimization problem to global optimality while knowing all of the data required to parameterize the given instance of the problem to be solved. Moreover, in a game-theoretical context (as bilevel optimization naturally is), a fully rational decision process also needs to include that the decision of the other players can either be observed or anticipated perfectly.

This point of view leads to a wider scope for bilevel optimization under uncertainty than it can be the case for single-level optimization. Besides the classic topic of data uncertainty, bilevel optimization under uncertainty may cover the following aspects:

(i) The leader may be uncertain about her[1] anticipation of the follower's rational reaction and, thus, may want to hedge against this uncertainty; see, e.g., Besançon et al. (2019).

(ii) As an extreme case of the former aspect it may be the case that the upper-level player knows that the follower will play against her. This is the setting of a pessimistic bilevel optimization problem, which is rather naturally connected to the field of robust optimization; see, e.g., Wiesemann et al. (2013). However, if the level of cooperation or confrontation of the follower

---

[1] According to the experimental results collected by the male author of this survey while assigning work to co-authors during the writing process, we decided to use "her" for the leader and "his" for the follower throughout the paper.

is not known, this leads to intermediate cases in between the optimistic and the pessimistic case; see, e.g., Aboussoror and Loridan (1995) and Mallozzi and Morgan (1996). Rather obviously, this is another realization of decision uncertainty.

(iii) It can also be the case that the leader can anticipate the rational reaction of the follower but that the follower is not able to perfectly observe the leader's decision. In this case, the follower—if aware of this aspect—usually tries to hedge against this uncertainty (Bagwell 1995; Beck and Schmidt 2021; van Damme and Hurkens 1997).

(iv) Even if all data and the rational reaction of the follower is known and even if the leader can, in principle, fully anticipate the optimal reaction of the follower, it might still be the case that limited intellectual or computational resources make it impossible that globally optimal decisions are taken. In contrast, only approximately optimal or heuristic answers of, e.g., the follower need to be considered, imposing the challenge that the leader does not know which heuristic or which approximation is applied by the follower. As a good primer in this context, we refer to the recent paper by Zare, Prokopyev, et al. (2020).

This list is not comprehensive but should make clear how much more diverse the sources of uncertainty can be in bilevel optimization as compared to single-level optimization.

Throughout this survey, we will highlight different aspects of bounded rationality as it has been roughly discussed above. Most of the papers, but not all of them, that we will review are concerned with the typical setting of data uncertainty. However, the interest of the mathematical optimization as well operations research community in decision uncertainty is growing. The survey thus has two main goals. First, to almost comprehensively describe the state-of-the-art of bilevel optimization under uncertainty. Second, to also view the existing research under a bounded-rationality lens to put the existing literature into a broader (game-theoretic or economic) context as well as to open doors to future research that maybe would stay locked—or not even seen—while using a different lens.

The remainder of this survey is structured as follows. In Section 2, we define the overall problem statement and discuss both data as well as decision uncertainty using illustrating examples. Afterward, in Section 3, we then discuss the existing (and mostly theoretical) literature on bilevel optimization under uncertainty along the lines on how uncertainty is modeled. Thus, we explicitly consider stochastic bilevel problems, bilevel problems with robust modeling of data uncertainty, bilevel problems with near-optimal lower-level decisions, limited observability of decisions of the leader, and the field of intermediate solution concepts between the optimistic and pessimistic cases. In Section 4, we then review papers on bilevel optimization under uncertainty that are devoted to specific applications. Here, we focus on applications from the field of energy, security, management science, and networks. We close the paper with some concluding words in Section 5, where we also mention open questions and other possible directions for future research. Throughout the survey, we assume that the reader is familiar with standard concepts of robust and stochastic optimization.

2. General Problem Statement

We study bilevel problems of the general form

$$\text{“}\min_{x\in X}\text{”} \quad F(x,y) \tag{1a}$$

$$\text{s.t.} \quad G(x,y) \geq 0, \tag{1b}$$

$$y \in S(x), \tag{1c}$$

where $S(x)$ denotes the set of optimal solutions of the x-parameterized problem

$$\min_{y\in Y} \quad f(x,y) \tag{2a}$$

$$\text{s.t.} \quad g(x,y) \geq 0. \tag{2b}$$

We refer to Problem (1) as the upper-level (or the leader's) problem and to Problem (2) as the lower-level (or the follower's) problem. Moreover, the variables $x \in X$ and $y \in Y$ are called the leader's and the follower's variables, respectively. The sets $X \subseteq \mathbb{R}^{n_x}$ and $Y \subseteq \mathbb{R}^{n_y}$ can be used to denote integrality constraints. The objective functions are given by $F, f : \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \to \mathbb{R}$ and the constraint functions by $G : \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \to \mathbb{R}^m$ as well as $g : \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \to \mathbb{R}^\ell$. A summary of important notation used throughout this paper can also be found in Table 1. The quotation marks in (1a) express the ill-posedness of the bilevel problem in the case that the lower-level problem does not have a unique solution. To deal with this ambiguity, it is common to pursue either an optimistic or a pessimistic approach to bilevel optimization; see, e.g., Dempe (2002). For the ease of presentation, we focus on the optimistic setting at this point, i.e., we study

$$\min_{x,y} \quad F(x,y) \quad \text{s.t.} \quad G(x,y) \geq 0,\ x \in X,\ y \in S(x).$$

We consider bilevel problems of the above form, which are, however, affected by various kinds of uncertainty. This setting is relevant for many practical applications since uncertainty is an important aspect of bounded rationality; see, e.g., Simon (1972). In this survey article, we mainly distinguish between two types of uncertainty: data uncertainty and decision uncertainty.

2.1. **Data Uncertainty.**

Data uncertainty arises if, e.g., the lower-level player only has access to inaccurate or incomplete data. To illustrate this aspect, let us assume that the right-hand sides of the lower-level constraints are uncertain. For a feasible leader's decision x and a specific realization of the uncertainty u, the set of optimal follower's decisions is then given by

$$S(x,u) := \operatorname*{arg\,min}_{y\in Y} \left\{ f(x,y) \colon g(x,y) \geq z(u) \right\},$$

where $z(u) \in \mathbb{R}^\ell$ represents the lower-level right-hand side vector for the given uncertainty realization u. In mathematical optimization, it is common to use one of the following two variants to deal with data uncertainty.

(i) Uncertainties are assumed to take values in a given uncertainty set $\mathcal{U}$. Pursuing a robust approach, we hedge against the worst-case realization of the uncertainties w.r.t. the leader's optimal objective function value. For the ease of presentation, we assume that there are no coupling constraints, i.e., there are no upper-level constraints that explicitly depend on y. To this end, we define $\bar{X} := \{x \in X \colon G(x) \geq 0\}$ with $G : \mathbb{R}^{n_x} \to \mathbb{R}^m$. We then solve

$$\min_{x\in\bar{X}} \max_{u\in\mathcal{U}} \min_{y\in S(x,u)} \quad F(x,y).$$

Here, we consider the robust and optimistic case in which the leader may influence the follower's decision in her favor. However, the consideration of

Table 1. Central Notation.

Sets	
$X \subseteq \mathbb{R}^{n_x}$	Set of all admissible upper-level decisions
$\mathcal{X}(x) \subseteq \mathbb{R}^{n_x}$	(Imperfectly) perceived upper-level feasible region
$\bar{X} \subseteq \mathbb{R}^{n_x}$	Upper-level feasible region without coupling constraints
$Y \subseteq \mathbb{R}^{n_y}$	Set of all admissible lower-level decisions
$\mathcal{U}$	Uncertainty set
Ω	Finite set of possible scenarios
Variables	
$x \in X$	Upper-level decision
$\bar{x} \in \mathcal{X}(x)$	(Imperfectly) observed upper-level decision
$u \in \mathcal{U}$, $\omega \in \Omega$, z	Uncertainty realization
$y = y(x) \in Y$	Lower-level response for given x
$y(x,u)$, $y(x,\omega) \in Y$	Lower-level response for given x and uncertainty realization u or ω
Functions	
$F, f : \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \to \mathbb{R}$	Upper and lower-level objective functions
$G : \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \to \mathbb{R}^m$	Upper-level constraints
$g : \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \to \mathbb{R}^\ell$	Lower-level constraints
$\varphi : X \to \mathbb{R}$	Lower-level optimal-value function
Point-to-Set Mappings	
$Y(x)$	Lower-level feasible set for given x
$S(x)$	Set of optimal lower-level solutions for given x
$S(x,u)$, $S(x,z)$	Set of optimal lower-level solutions for given x and uncertainty realization u or z
$S(x,\varepsilon)$	Set of ε-optimal lower-level solutions for given x with $\varepsilon > 0$

other solution concepts is also possible for uncertain bilevel problems. For instance, the most conservative situation in which the leader anticipates a pessimistic follower is given by

$$\min_{x \in \bar{X}} \max_{u \in \mathcal{U}} \max_{y \in S(x,u)} F(x,y).$$

Eventually, the application at hand dictates which model is appropriate.

(ii) Adopting a stochastic approach, it is assumed that the uncertainties can be described by a given probability distribution. Here, we hedge against uncertainties in a probabilistic sense by optimizing, e.g., the expected value. In line with all the existing literature on stochastic bilevel optimization (see also Section 3.1), we again focus on the setting without coupling constraints. Hence, we solve

$$\min_{x \in \bar{X}} \mathbb{E}_u \left[\Phi_u(x)\right] \quad \text{with} \quad \Phi_u(x) := \min_{y \in S(x,u)} F(x,y).$$

In both cases, we consider the timing

$$\text{leader } x \quad \curvearrowright \quad \text{uncertainty } u \quad \curvearrowright \quad \text{follower } y = y(x,u).$$

This means that the leader first takes a here-and-now decision, i.e., without knowing the realization of uncertainty. Then, the uncertainty realizes and, finally, the follower decides in a wait-and-see manner, taking the leader's decision as well as the realization of the uncertainty into account. However, other timings are possible as well, for instance, if we consider problems of the form

$$\begin{aligned} \min_{x,y} \quad & F(x,y) \\ \text{s.t.} \quad & G(x,y) \geq 0,\ x \in X, \\ & y \in \operatorname*{arg\,min}_{\bar{y} \in Y} \left\{ f(x,\bar{y}) \colon g(x,\bar{y}) \geq z(u) \text{ for all } u \in \mathcal{U} \right\}. \end{aligned}$$

This is another robust bilevel optimization problem but this time, the leader takes a here-and-now decision and the follower also decides before the uncertainty realizes. The decision of the follower is robust in the sense that it is required to remain feasible for all possible realizations of the uncertainty. Hence, one considers the timing

$$\text{leader } x \quad \curvearrowright \quad \text{follower } y = y(x) \quad \curvearrowright \quad \text{uncertainty } u. \tag{3}$$

So far, we have discussed the case of lower-level right-hand side uncertainty. Of course, data uncertainty may also occur at other locations of the problem such as, e.g., in the upper-level problem's data or in the objective function of the follower. The following examples illustrate robust and stochastic approaches to deal with uncertain data at various locations of the lower-level problem.

Example 1. *Let us consider the linear bilevel problem*

$$\begin{aligned} \min_{x,y \in \mathbb{R}} \quad & F(x,y) = x - 4y && \text{(4a)} \\ s.t. \quad & x - y \geq -1, && \text{(4b)} \\ & 3x + y \geq 3, && \text{(4c)} \\ & y \in S(x), && \text{(4d)} \end{aligned}$$

where $S(x)$ denotes the set of optimal solutions of the x-parameterized lower-level problem

$$\begin{aligned} \min_{y \in \mathbb{R}} \quad & f(x,y) = -0.1y && \text{(5a)} \\ s.t. \quad & -2x + y \geq -7, && \text{(5b)} \\ & -3x - 2y \geq -14, && \text{(5c)} \\ & 0 \leq y \leq 2.5. && \text{(5d)} \end{aligned}$$

The problem is depicted in Figure 1 (left). The upper- and lower-level constraints are represented with dashed and solid lines, respectively. The optimal solution $(x,y) = (1.5, 2.5)$ of the deterministic bilevel problem (4) and (5) is illustrated by the thick dot. Suppose now that the lower-level objective function is uncertain. We follow a robust approach and assume that the follower decides in a here-and-now fashion, i.e., we consider the timing in (3). The uncertain objective function coefficient is assumed to take values in the uncertainty set $\mathcal{U} = \{-0.1 + \zeta \colon |\zeta| \leq 0.5\} = [-0.6, 0.4]$, which leads to a modified gradient of the lower-level objective function. This effect is shown in Figure 1 (right). The optimal solution of the uncertain bilevel problem is represented by the thick square. In particular, we obtain a completely different solution than in the deterministic case if we take data uncertainty into account. Let us further point out that, in this example, we do not have to distinguish between the optimistic and the pessimistic case since the solution of the (robust) lower-level problem is unique for every feasible x.

Example 2. *Consider the linear bilevel problem*

$$\min_{x,y \in \mathbb{R}} \quad F(x,y) = -y \quad s.t. \quad x \geq 0, \quad y \in S(x), \tag{6}$$

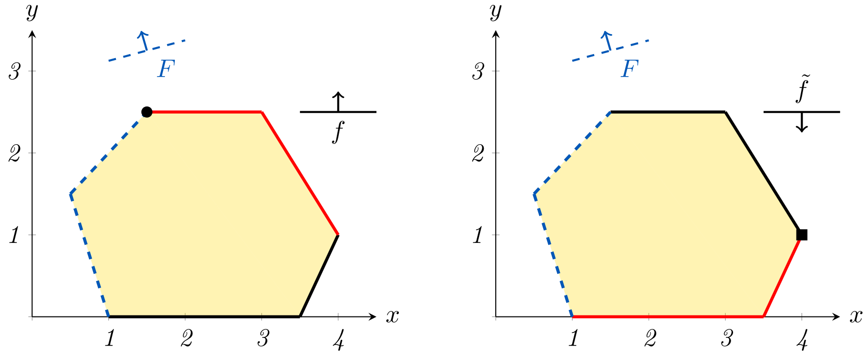

FIGURE 1. Both figures show the upper-level constraints (dashed blue lines), the lower-level constraints (solid black and red lines), the shared constraint set (yellow area), and the bilevel feasible set (solid red lines) of the bilevel problem (4) and (5). The deterministic variant of the problem is depicted on the left and the variant of the problem that accounts for a robust modeling of an uncertain lower-level objective function $\tilde{f}$ is given on the right.

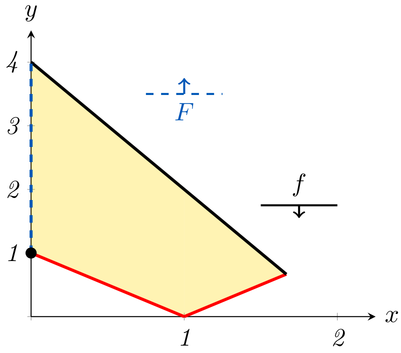

FIGURE 2. The upper-level constraint (dashed blue line), the lower-level constraints (solid black and red lines), the shared constraint set (yellow area), and the bilevel feasible set (solid red lines) of the bilevel problem (6) and (7).

where $S(x)$ denotes the set of optimal solutions of the x-parameterized lower-level problem

$$\min_{y \geq 0} \quad f(x,y) = y \tag{7a}$$

$$\text{s.t.} \quad x + y \geq 1, \tag{7b}$$

$$-x + y \geq -1, \tag{7c}$$

$$-2x - y \geq -4. \tag{7d}$$

The problem is depicted in Figure 2. The upper- and lower-level constraints are represented with dashed and solid lines, respectively. The unique optimal solution $(x,y) = (0,1)$ of the deterministic bilevel problem (6) and (7) is illustrated by the thick dot. Suppose now that the right-hand side of Constraint (7c) is uncertain. We pursue a stochastic approach and assume that the right-hand side $b(\omega) \in \mathbb{R}$

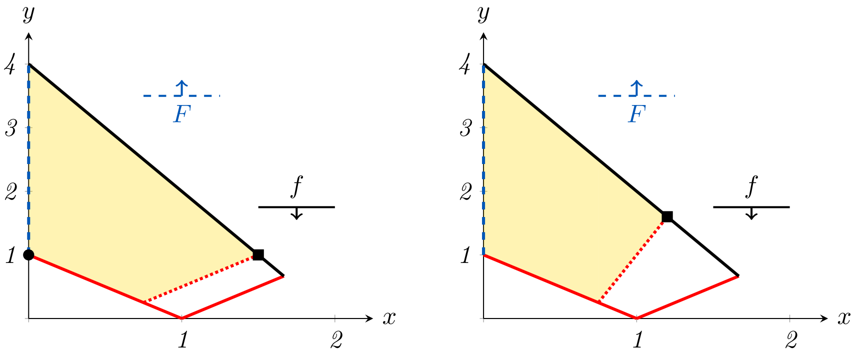

FIGURE 3. Variants of the bilevel problem (6) and (7) with lower-level right-hand side uncertainty (left) and uncertain lower-level constraint coefficients (right).

depends on the scenario $\omega \in \Omega = \{\omega^1, \omega^2\}$ *with* $b(\omega^1) = -1$ *and* $b(\omega^2) = -1/2$. *We further assume that both scenarios have probability* $p^1 = p^2 = 1/2$. *We start by considering each scenario individually. Note that the realization of* ω^1 *corresponds to the deterministic setting. Hence, the unique optimal solution for scenario* ω^1 *is given by* $(x, y^1) = (0, 1)$. *Here and in what follows, we set* $y^i = y(\omega^i)$ *for* $i = 1, 2$. *The realization of scenario* ω^2 *leads to a parallel shift of the uncertain lower-level constraint. This effect is shown in Figure 3 (left). It can also be seen that the solution is not unique anymore if scenario* ω^2 *is considered. Both* $(0, 1)$ *and* $(3/2, 1)$*—which are illustrated by the thick dot and the thick square, respectively—yield an optimal objective function value of* -1. *To hedge against lower-level right-hand side uncertainty, we optimize the expected value of the upper-level objective function, i.e., we solve*

$$\min_{x, y^1, y^2 \in \mathbb{R}} \quad -p^1 y^1 - p^2 y^2 \quad \text{s.t.} \quad x \geq 0,\ y^1 \in S(x, \omega^1),\ y^2 \in S(x, \omega^2). \tag{8}$$

The unique solution for the variant of Problem (6) *and* (7) *with lower-level right-hand side uncertainty is given by* $(x, y^1, y^2) = (0, 1, 1)$. *Despite the consideration of data uncertainty, the overall bilevel solution does not change significantly compared to the deterministic setting. However, the following shows that this may not always be the case.*

To this end, we focus on the stochastic modeling of uncertain constraint coefficients in (7c). *The constraint coefficients* $a(\omega) \in \mathbb{R}^2$ *are assumed to depend on the scenario* $\omega \in \Omega = \{\omega^1, \omega^2\}$ *with* $a(\omega^1) = (-1, 1)$ *and* $a(\omega^2) = (-3/2, 1/2)$. *We further assume that scenario* ω^1 *has probability* $p^1 = 1/3$, *whereas* ω^2 *has probability* $p^2 = 2/3$. *Again, the realization of scenario* ω^1 *corresponds to the deterministic setting. Thus, the unique optimal solution for scenario* ω^1 *is given by* $(x, y^1) = (0, 1)$. *The setting in which scenario* ω^2 *realizes is shown in Figure 3 (right). The unique optimal solution* $(x, y^2) = (6/5, 8/5)$ *is illustrated by the thick square. Hedging against data uncertainty by optimizing over the expected value yields the unique overall stochastic bilevel solution* $(x, y^1, y^2) = (6/5, 1/5, 8/5)$, *which can be obtained by solving the corresponding scenario-expanded formulation* (8). *In particular, the solution is attained at a completely different vertex of the bilevel feasible set than in the deterministic case.*

2.2. **Decision Uncertainty.** Decision uncertainty refers to the case in which the players may face uncertainties regarding the decision of the other player. For instance, the follower may lack the ability or the resources to obtain an optimal solution and, thus, takes a "satisfactory" solution instead of an optimal one. For a given leader's decision x, this can be modeled using the set of ε-optimal reactions of the follower, which is given by

$$S(x, \varepsilon) = \{y \in Y \colon g(x, y) \geq 0,\ f(x, y) \leq \varphi(x) + \varepsilon\}, \quad \varepsilon > 0,$$

where

$$\varphi(x) := \min_{y \in Y} \{f(x, y) \colon g(x, y) \geq 0\}$$

denotes the lower-level's optimal-value function. The parameter ε quantifies the follower's willingness to deviate from his optimal objective function value. As a consequence of the follower's ε-optimality, the leader is uncertain about the actual response of the follower. The aim of the leader may thus be to hedge against the worst-case ε-optimal reaction of the follower, i.e., one studies the problem

$$\min_{x \in X} \max_{y \in S(x, \varepsilon)} \quad F(x, y) \quad \text{s.t.} \quad G(x, \hat{y}) \geq 0 \text{ for all } \hat{y} \in S(x, \varepsilon). \tag{9}$$

In particular, Problem (9) can be reformulated as a specific instance of the pessimistic bilevel problem considered by Wiesemann et al. (2013).

Another reason for decision uncertainty may be the follower's limited capability to observe the decision of the leader. A reasonable assumption in this context, however, may be that the follower has an insight into the leader's scope of action. One way to account for this type of decision uncertainty is the following modeling. Let $\mathcal{X}(x) \subset \mathbb{R}^{n_x}$ denote the x-dependent set containing all possible decisions of the leader, which is assumed to be known by the follower. Clearly, the actual decision x of the leader belongs to $\mathcal{X}(x)$. The follower then takes his decision to hedge against all possible leader decisions $\bar{x} \in \mathcal{X}(x)$, e.g., by pursuing a robust approach that leads to the bilevel model

$$\begin{aligned} \min_{x, y} \quad & F(x, y) \\ \text{s.t.} \quad & G(x, y) \geq 0,\ x \in X, \\ & y \in \operatorname*{arg\,min}_{\bar{y} \in Y} \left\{ \max_{\bar{x} \in \mathcal{X}(x)} f(\bar{x}, \bar{y}) \colon g(\bar{x}, \bar{y}) \geq 0 \text{ for all } \bar{x} \in \mathcal{X}(x) \right\} \end{aligned}$$

with a robustified follower's objective function and a robustified feasible set of the lower-level problem.

Example 3. *To illustrate some modeling approaches for decision uncertainty discussed so far, we consider the following example taken from Beck and Schmidt (2021) and Besançon et al. (2019):*

$$\begin{aligned} \min_{x, y \in \mathbb{R}} \quad & F(x, y) = x - 10y && \text{(10a)} \\ \text{s.t.} \quad & x - 4y \geq -11, && \text{(10b)} \\ & -x - 2y \geq -13, && \text{(10c)} \\ & x \geq 0,\ y \in S(x). && \text{(10d)} \end{aligned}$$

Here, $S(x)$ again denotes the set of optimal solutions of the x-parameterized lower-level problem, which is given by

$$\begin{aligned} \min_{y \in \mathbb{R}} \quad & f(x, y) = y && \text{(11a)} \\ \text{s.t.} \quad & 2x + y \geq 5, && \text{(11b)} \\ & -5x + 4y \geq -30, && \text{(11c)} \\ & y \geq 0. && \text{(11d)} \end{aligned}$$

The deterministic bilevel problem (10) and (11) is depicted in Figure 4. Again, the upper- and lower-level constraints are represented with dashed and solid lines,

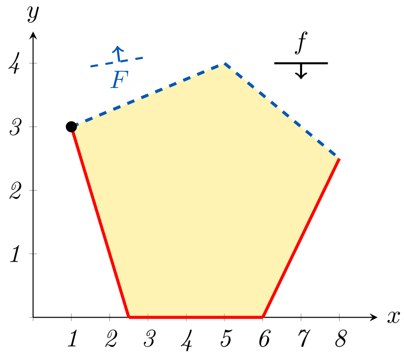

FIGURE 4. The upper-level constraints (dashed blue lines), the lower-level constraints as well as the bilevel feasible set (solid red lines), and the shared constraint set (yellow area) of the bilevel problem (10) and (11).

respectively, and the optimal solution $(x, y) = (1, 3)$ *of the problem is illustrated by the thick dot. Note that, in the deterministic case, the solid lines also correspond to the bilevel feasible set. The modeling of an* ε*-optimal follower leads to a parallel shift of the upper-level constraints since the leader needs to make sure that her decision remains feasible for every* ε*-optimal decision of the follower. Hence, the bilevel feasible set is reduced as it can be seen in Figure 5 (left) for* $\varepsilon = 0.5$*. The optimal solution of the bilevel problem (10) and (11) with an* ε*-optimal follower is given by the thick dot.*

To illustrate the effect of a follower with limited capability to perfectly observe the actual decision x *of the leader, let us assume that the perceived leader's decision* $\bar{x}$ *belongs to the uncertainty set* $\mathcal{X}(x) = \{x + \zeta \colon |\zeta| \leq 0.5\}$*. Similar to the setting with an* ε*-optimal follower, pursuing a robust approach to account for limited observability reduces the bilevel feasible set as it can be seen in Figure 5 (right). Here, however, the consideration of limited observability leads to a parallel shift of the lower-level constraints that explicitly depend on the variable of the leader. The reason is that, in this setting, the follower makes a decision, which must be feasible for all possible realizations of the leader's decision. The optimal solution of the bilevel problem (10) and (11) with a follower who cannot perfectly observe the actual decision of the leader is illustrated by the thick square.*

To sum up, the consideration of uncertainties in bilevel optimization may impact the solution of the problem significantly. Moreover, the obtained solution depends on the specific modeling of uncertainty that is taken into account. Hence, uncertainties are important to be considered—especially if decision makers are involved who are subject to bounded rationality.

Apart from the aforementioned aspects, there are many other modeling approaches to account for uncertainties in bilevel optimization for which we provide an in-depth discussion in the following sections.

3. DIFFERENT APPROACHES TO ACCOUNT FOR UNCERTAINTY

3.1. **Stochastic Bilevel Problems.**

To the best of our knowledge, the first series of papers on stochastic bilevel problems are the ones by Patriksson and Wynter (1999, 1997); see also Christiansen et al. (2001) for a follow-up paper. Before we review the stochastic setups discussed in these papers, let us first state the deterministic

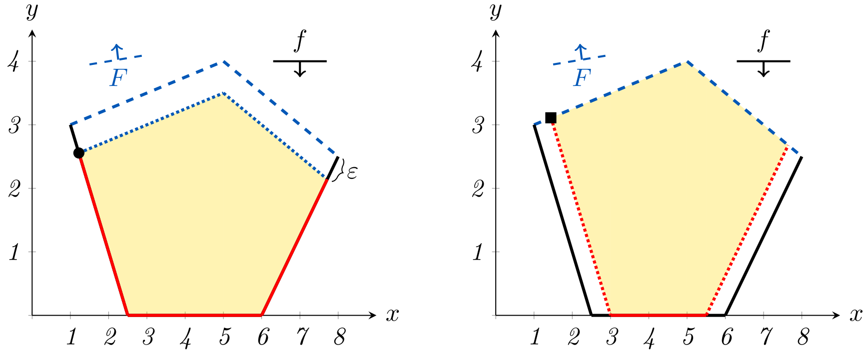

FIGURE 5. Both figures show the shared constraint set (yellow area) and the bilevel feasible set (red lines) of the linear bilevel problem (10) and (11) with an ε-optimal follower with $\varepsilon = 0.5$ (left) and the variant of the problem with a follower facing limited observability (right).

problem, which the authors call a "generalized" bilevel optimization problem since the lower-level is given by a variational inequality problem. To define this, we consider a bilevel setting in which no coupling constraints are present, i.e., there are no upper-level constraints that explicitly depend on the follower's variables. Further, we consider the function $T : X \times \mathbb{R}^{n_y} \to \mathbb{R}^{n_y}$ and the lower-level feasible set $Y(x)$. Then, the variational inequality problem is to find a point $y^* \in Y(x)$ such that

$$T(x, y^*)^\top (y - y^*) \geq 0 \quad \text{for all} \quad y \in Y(x) \tag{12}$$

holds for a given leader's decision $x \in X$. In the cited papers, the authors state this problem in a geometric way by using normal cones. To this end, let $Y(x)$ be convex for all possible leader decisions $x \in X$. We call a vector v a normal of the convex set $Y(x)$ at a point $\bar{y} \in Y(x)$ if

$$v^\top (y - \bar{y}) \leq 0 \quad \text{for all} \quad y \in Y(x)$$

holds. The set of such vectors is then denoted by $N_{Y(x)}(\bar{y})$ and is called the normal cone to $Y(x)$ at $\bar{y}$. Having this notation at hand, we can re-write the variational inequality (12) as

$$-T(x, y^*) \in N_{Y(x)}(y^*). \tag{13}$$

For more background on variational analysis we refer the interested reader to Rockafellar and Wets (1998). Hence, the overall deterministic, generalized bilevel problem is given by

$$\min_{x \in X} \quad F(x, y) \quad \text{s.t.} \quad -T(x, y) \in N_{Y(x)}(y).$$

The classic, i.e., non-generalized, bilevel problem is covered since for continuously differentiable and pseudo-convex lower-level objective functions f as well as non-empty, closed, and convex lower-level feasible sets $Y(x)$, the minimum principle ensures that the parameterized variational inequality (13) characterizes the global optimal solutions of the parameterized lower-level problem

$$\min_{y} \quad f(x, y) \quad \text{s.t.} \quad y \in Y(x)$$

if one identifies $T(x, y) = \nabla_y f(x, y)$ and $Y(x) = \{y \in \mathbb{R}^{n_y} : g(x, y) \geq 0\}$.

3.1.1. *Risk-Neutral Models.* In Patriksson and Wynter (1997), the authors generalize traditional two-stage[2] stochastic problems to generalized bilevel optimization problems under uncertainty. They illustrate this novel class of problems using an example from the field of traffic equilibrium modeling that includes uncertain travel demands as stochastic parameters. As a solution approach, the authors propose a descent method that uses sensitivity analysis to obtain derivatives for setting up search directions, a line-search method (of Armijo-type) to get step sizes, and a projection onto the feasible set of the upper-level player to obtain a new iterate.

In Patriksson and Wynter (1999), the same authors consider stochastic mathematical programs with equilibrium constraints (SMPECs)—a class of problems that comprises stochastic bilevel problems if compact optimality conditions such as the Karush–Kuhn–Tucker (KKT) conditions or the above minimum principle are both necessary and sufficient for lower-level optimality. We refer the reader to Lin and Fukushima (2010) for a survey on SMPECs. Given the deterministic setup above, the stochastic and risk-neutral counterpart reads

$$\min_{x \in X} \quad \mathbb{E}_\omega \left[F(x, y(\omega)) \right], \tag{14}$$

where for all $\omega \in \Omega$,

$$y(\omega) \in \left\{ y \in \mathbb{R}^{n_y} : \; -T(x, \omega, y) \in N_{Y(x,\omega)}(y) \right\} \tag{15}$$

denotes the solutions of the lower-level variational inequality problem, which is parameterized by the upper-level decision x and the random variable ω. We recover the stochastic bilevel problem by identifying $T(x, \omega, y) = \nabla_y f(x, \omega, y)$ and $Y(x, \omega) = \{y \in \mathbb{R}^{n_y} : g(x, \omega, y) \geq 0\}$. As usual, the random variable ω is defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The timing covered by the problem given in (14) and (15) is the following. The leader first takes her decision in a here-and-now manner, then the uncertainty realizes and, finally, the follower decides based on the realization of uncertainty and the leader's decision. Hence, we consider the timing

$$\text{leader } x \quad \curvearrowright \quad \text{uncertainty } \omega \quad \curvearrowright \quad \text{follower } y = y(x, \omega) \tag{16}$$

and the considered stochastic setting is risk-neutral since the expected value is minimized in the upper-level objective function. The presented theoretical results in Patriksson and Wynter (1999) are concerned with existence of solutions as well as convexity and directional differentiability of the (implicitly defined) upper-level objective function. Finally, a subgradient descent method is sketched that follows the main ideas already discussed for the special case in Patriksson and Wynter (1997). Going further, Patriksson (2008a,b) again considers SMPECs (in the context of equilibrium problems from structural optimization and traffic assignment problems) and shows that, under some additional assumptions, the SMPEC solutions continuously depend on the probability distribution used to model the uncertainty in the lower-level problem. Note that the author calls this continuity property "robustness", which has nothing to do with robustness in the sense of robust optimization. The papers discussed so far do not contain any numerical results.

In Christiansen et al. (2001), the authors study stochastic bilevel problems for truss topology optimization problems in which the external load applied to the truss is uncertain and the random variables are assumed to be discrete and finite. Hence, a finite set of scenarios is considered. Using classic inf-compactness assumptions, i.e., lower semicontinuity and bounded level sets of the upper-level objective function, existence of solutions is shown. This is extended by an existence result without requiring inf-compactness. However, additional and problem-specific assumptions

---

[2] Here and in what follows, we try to clearly distinguish between "stages" and "levels", where "stages" always refer to stages as considered in, e.g., two-stage stochastic optimization, whereas "levels" are always used in the sense of multilevel optimization, e.g., as in bilevel optimization.

are used such as that the upper-level objective function is quadratic, which then allows to invoke the classic existence theorem by Frank and Wolfe (1956). As in the other papers discussed so far, the authors derive further results on the directional differentiability of the implicitly given upper-level objective function to design a subgradient method that further exploits parallelization across the scenarios. In contrast to the papers discussed before, Christiansen et al. (2001) also present a small numerical case study for a stochastic truss topology optimization problem.

These early results on stochastic bilevel optimization are wrapped up in the handbook chapter by Wynter (2008). Again, existence of optimal solutions is discussed for the case of a discrete set of scenarios. Moreover, sufficient conditions for the convexity of stochastic bilevel problems are given as well. Convexity, however, is only possible if the upper-level problem does not contain any coupling constraints and if the objective function of the upper level only depends on the optimal value of the lower level and not on the follower's decision itself. As already done in the original works discussed above, the author also collects sufficient conditions that ensure that the upper-level objective function is Lipschitz continuous and directionally differentiable. The handbook chapter closes with a sketch of the subgradient method from the previously discussed original works.

The above mentioned parallelization techniques are a special case of decomposition methods that are very prominent in the literature on stochastic optimization; cf., e.g., the L-shaped and other Benders-inspired decomposition methods for two- or multistage stochastic optimization; see, e.g., Birge and Louveaux (2011), Fischetti, Ljubić, and Sinnl (2016, 2017), Rahmaniani et al. (2018), and Van Slyke and Wets (1969). These decomposition techniques have also been carried over to the case of stochastic MPECs. For this, see, e.g., Shapiro and Xu (2008), where scenario generation techniques are discussed to obtain a finite-dimensional deterministic equivalent (which is a deterministic MPEC or bilevel optimization problem) with separate blocks of follower decisions for each realization of the uncertainty. Such block structures are then exploited by decomposition methods. For an overview of these methods, we refer the reader to the PhD thesis by Henkel (2014), where a broad literature review is given w.r.t. what has been published up to 2014. The PhD thesis itself studies both the classic KKT as well as the optimal-value reformulation for stochastic bilevel problems with discrete and finite probability distributions. Based on the KKT approach, an integer-programming based method is designed and evaluated that also uses problem-tailored decomposition techniques. The numerical results contain instances with up to 20 scenarios and with up to 20 lower-level variables.

Rather recently, Bolusani et al. (2020) consider the similarities between multilevel mixed-integer linear optimization and multistage stochastic mixed-integer linear optimization with recourse. For the bilevel stochastic setting mentioned above (with a discrete set of scenarios, a wait-and-see follower, and a risk-neutral leader), they exploit the block-angular structure of the bilevel deterministic equivalent and propose both a Benders-like decomposition as well as a cutting-plane method. These methods are also implemented in MibS; see Tahernejad (2019).

Note that for a risk-neutral here-and-now follower and with a given set of discrete scenarios, one can easily turn bilevel two-stage stochastic optimization problems into their bilevel deterministic counterparts as it is done in two-stage stochastic optimization. Hence, any general purpose solver for deterministic bilevel problems (see, e.g., Fischetti, Ljubić, Monaci, et al. (2017) or MibS by Tahernejad et al. (2020), see `https://coin-or.github.io/MibS/`) can be used for that purpose.

3.1.2. *Bilevel Models with a Quantile Criterion.* Another branch of research does not consider the optimization of an expected value in the leader's objective function but studies a quantile criterion that ensures that a certain upper-level objective function value is not exceeded with a given probability. In different application contexts, these settings have been considered first in Chen et al. (2007) and Katagiri et al. (2014) and the first theoretical analysis has been carried out, to the best of our knowledge, by Ivanov (2014). The mathematical setup is given as follows. The lower-level solution set is given by

$$S(x,z) = \arg\min_{y} \left\{ d^\top y \colon Dy \geq z - Cx,\ y \geq 0 \right\},$$

where $z \in Z$ is a realization of a random vector. For the ease of presentation, we omit to state the dimensions of all vectors and matrices. With this at hand, the so-called loss function of the leader is defined as

$$\Phi(x,z) = \begin{cases} \min_{y \in S(x,z)} c_y^\top y, & \text{if } S(x,z) \neq \emptyset, \\ +\infty, & \text{if } S(x,z) = \emptyset, \end{cases}$$

and the corresponding α-quantile function reads

$$\Phi_\alpha(x) = \min \left\{ \varphi \colon \mathbb{P}(\Phi(x,Z) \leq \varphi) \geq \alpha \right\},$$

where $\mathbb{P}(\cdot)$ is the probability measure induced by the distribution of the random vector Z. Finally, the bilevel optimization problem with a quantile criterion is given by

$$\min_{x \in X} \quad c_x^\top x + \Phi_\alpha(x).$$

The paper first presents theoretical results regarding Lipschitz continuity of the leader's loss function. After also proving the continuity of the corresponding quantile function, an existence result is obtained. Finally, the paper shows that the studied problem can be reformulated as a single-level mixed-integer linear optimization problem in the case of a discrete and finite distribution of the random variables. To this end, the classic linearization technique by Fortuny-Amat and McCarl (1981) is used, including the choice of sufficiently large big-M values and by introducing additional binary variables to linearize the KKT complementarity conditions. Finally, a numerical case study is presented with 2 upper-level and 3 lower-level variables as well as 4 upper-level and 2 lower-level constraints. The number of considered scenarios is 25.

The paper by Dempe, Ivanov, et al. (2017) builds on Ivanov (2014) and generalizes the setting. Again, a stochastic bilevel problem with a quantile criterion is considered in which the lower-level problem depends on the realization of a random vector and on the leader's decision. In contrast to Ivanov (2014), the authors now consider the so-called "a priori statement" of the problem, meaning that the follower's decision variables are chosen from a set of functions depending on random parameters, whereas the so-called "a posteriori statement" is studied in Ivanov (2014). Hence, Dempe, Ivanov, et al. (2017) study the timing

$$\text{leader } x \quad \curvearrowright \quad \text{follower } y = y(x) \quad \curvearrowright \quad \text{uncertainty } \omega, \tag{17}$$

which differs significantly from the one in (16) since the follower now also takes his decision before the uncertainty is realized. In Dempe, Ivanov, et al. (2017), the leader's problem can be nonlinear and the follower's problem is linear in the follower's variables. The authors establish a mixed-integer nonlinear single-level reformulation for the case that the random vector has a finite set of realizations and propose assumptions under which an optimal solution of the original problem exists. The numerical case study considers a single academic instance of roughly the same size as in Ivanov (2014).

In contrast to Ivanov (2014), where only the lower-level feasible set is affected by randomness, in Ivanov (2018), the setting is studied in which the randomness affects the objective function of the lower-level problem in a way such that its objective function is linear for a given realization of uncertainty and a given leader's decision. The lower level's feasible set, however, is both independent of the leader's decision as well as of the realization of uncertainty. The leader again minimizes a quantile function of her loss function depending on the leader's and the follower's decision. It is shown that the lower-level problem has a unique solution (with probability 1) if the distribution is absolutely continuous and if the lower-level's feasible set is non-empty and bounded. The loss function is proven to be lower semicontinuous, which implies that the overall bilevel problem with a quantile criterion has an optimal solution if the upper-level's feasible set is non-empty and bounded as well. The quantile function itself, however, is not continuous. Finally, sample average approximation is applied and the convergence of its limit points is shown before a small-scale optimal tax rate problem is considered in a case study.

3.1.3. *Convex Risk Measures.* The aforementioned models based on a quantile criterion just discussed can be seen as models that incorporate some kind of risk measure. In the case of quantiles, these risk measures (such as the value-at-risk; VaR) are not convex. The key idea of incorporating risk measures in the upper-level's objective function of stochastic and linear bilevel problems (such as the quantile function in the previous section) is also studied in Burtscheidt, Claus, and Dempe (2020), where the setting of law-invariant, convex, and coherent risk measures is considered for uncertainties in the right-hand side of the follower's problem. More formally, the considered problem is an optimistic parametric bilevel linear problem of the form

$$\min_x \left\{ c_x^\top x + \min_y \left\{ c_y^\top y \colon y \in S(x, z) \right\} : x \in X \right\}$$

with a parameter z and the lower-level solution set mapping

$$S(x, z) = \arg\min_y \left\{ d^\top y \colon Dy \geq z - Cx \right\}.$$

The stochastic bilevel problem is then obtained by assuming that the parameter $z = Z(\omega)$ is the realization of a random vector Z defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Define now

$$G = \{(x, z) \colon \exists y \text{ with } Dy \geq z - Cx\}$$

to be the set of all upper-level decisions and parameters so that there is a feasible lower-level solution. Moreover, set

$$F(x, z) = c_x^\top x + \min_y \left\{ c_y^\top y \colon y \in S(x, z) \right\}$$

as an abbreviation for the upper-level objective function in dependence of the leader's decision x and the parameter z. An additional nonanticipativity constraint then leads us to the timing in (16), i.e., the leader takes a here-and-now decision while the follower takes a wait-and-see decision. Further, denote the Borel probability measure induced by the random vector Z by $\mu_Z = \mathbb{P} \circ Z^{-1}$. With this, it is assumed that

$$X \subseteq \{x \colon (x, z) \in G \text{ for all } z \in \operatorname{supp}(\mu_Z)\},$$

i.e., there exists a lower-level feasible point for all possible leader decisions and all realizations of uncertainty. After fixing properly chosen function spaces and a suitable risk measure $\mathcal{R}$, the stochastic bilevel problem is given by

$$\min_x \quad \mathcal{R}[F(x, Z(\cdot))] \quad \text{s.t.} \quad x \in X.$$

The class of considered risk measures includes, among others, the expected value, the conditional value-at-risk (CVaR), or a worst-case risk measure. Note that the latter includes the case of robust and linear bilevel problems. The key results of the paper are about (local) Lipschitz continuity of the upper-level objective function. Moreover, differentiability results are given for those risk measures that are based on expectations. For the case of discrete and finite distributions, equivalent deterministic bilevel problems are derived for the expectation-based risk measures. Finally, the MPCC regularization strategy proposed by Scholtes (2001) is applied to the KKT reformulation of the deterministic equivalent and a convergence result for the limit points of this regularization approach is proven.

Most of the content presented in Burtscheidt, Claus, and Dempe (2020) can also be found in the book chapter by Burtscheidt and Claus (2020), where the same class of problems is considered. The new content given there is about stochastic dominance constraints. Moreover, equivalent deterministic counterparts are proven for some of the cases that have not been included in Burtscheidt, Claus, and Dempe (2020).

In a follow-up paper, Claus (2021a) considers the same risk-neutral setting as in Burtscheidt, Claus, and Dempe (2020) and the first-order necessary optimality conditions for the risk-neutral case from the latter paper are accompanied by second-order sufficient conditions. It was known from Burtscheidt, Claus, and Dempe (2020) that the expectation functional, i.e., the upper-level objective function in the risk-neutral case, is continuously differentiable if the underlying probability measure is absolutely continuous w.r.t. the Lebesgue measure. However, (even local) Lipschitz continuity of the gradient may fail to hold under the assumptions used in Burtscheidt, Claus, and Dempe (2020). The main novel assumptions now are the boundedness of the support and the uniform boundedness of the Lebesgue density of the probability measure, which ensure (as the main result of the paper) the Lipschitz continuity of the gradient.

In the paper by Claus (2021c), the overall timing with a leader taking a here-and-now decision, while the follower reacts in a wait-and-see manner, is kept but the way how randomness enters the lower-level problem is different. Instead of considering a stochastic right-hand side, the lower-level's feasible set is now a fixed polyhedron but the coefficients of the linear objective function of the follower are affected by randomness. This still leads to the case that the upper-level objective function value is a random variable itself that is parameterized by the leader's decision. The main result is the development of sufficient conditions for the existence of optimal solutions for a wide class of risk measures, including the expected value, CVaR, or the worst-case risk measure that leads to a robust and linear bilevel problem. The main stepping stone towards this result is the proof of the continuity of the risk functionals. Surprisingly, these sufficient conditions are the same both for the optimistic and the pessimistic bilevel problem, which is not the case in the deterministic setting.

The, up to now, last paper in this row of research is Claus (2021b) in which the two settings considered so far are combined: randomness in the lower-level's right-hand side as well as in its objective function coefficients. The results of Claus (2021c) cannot be applied anymore since the extreme points of the lower-level's feasible set now depend on the realization of the randomness as well. However, lower semicontinuity can be achieved for a class of functions derived from convex risk measures. Further, a continuity result is derived that can also be applied to the pessimistic case as well. These continuity properties can then, as usual, be used to obtain existence results under further but classic compactness assumptions. As it is the case for all other papers discussed so far in this subsection, the paper does not contain any numerical results.

In Burtscheidt, Claus, Conti, et al. (2021), the same analytical spirit is followed but is now brought to the field of pessimistic and stochastic bilevel optimization.

Again, continuity results are derived in order to finally obtain an existence result. After these theoretical developments, the authors study an applied problem from the field of mechanical shape optimization that is modeled as a pessimistic and stochastic bilevel problem and also present some numerical results.

3.1.4. *Chance Constraints.* Up to now, all papers except for Dempe, Ivanov, et al. (2017) studied the timing in (16). In the case of the other timing (17), fulfilling the random lower-level constraints cannot be guaranteed almost surely in general but only with a certain probability. In a natural way, this leads to the use of probabilistic or chance constraints.

To the best of our knowledge, the first combination of bilevel optimization and chance constraints is studied in Kosuch et al. (2012). The authors study a standard linear bilevel problem that is extended by a probabilistic knapsack constraint in the upper-level problem. The setting is motivated by pricing applications in networks. For finite probability distributions, the authors first derive a deterministic equivalent formulation using additional binary variables and big-M constraints. This all takes place in the upper-level problem and is completely independent of the lower level. Afterward, the authors apply the classic mixed-integer linear reformulation by Fortuny-Amat and McCarl (1981), leading to a mixed-integer linear single-level problem. Based on this reformulation, a so-called min-max scheme is designed with bounds obtained from suitably chosen Lagrangian relaxations. The authors present numerical results for up to 1000 lower- and upper-level variables and for up to 100 scenarios.

From a more application-driven point of view, other chance-constrained bilevel problems have been studied in Pramanik and Banerjee (2012) and Yang, Zhang, et al. (2009). However, the only paper that we are aware of in which a chance constraint is considered in the lower-level problem and in which, thus, the timing in (17) is considered, is Heitsch et al. (2022). There, the chance constraint is shown to be convex and oracles for function as well as gradient evaluations are provided while the chance constraint itself cannot be stated in closed form. This setting is considered as having a convex black-box function in the lower level which is then tackled by an outer-approximation based cutting-plane method. However, since the lower-level chance constraint can thus only by satisfied by a prescribed tolerance $\varepsilon > 0$, it does not seem to be possible to prove anything about the upper-level objective function value obtained by this method—an issue that is also considered in the recent paper by Beck, Schmidt, et al. (2022) on continuous but nonconvex lower-level problems.

3.1.5. *Knapsack Problems.* The first paper on bilevel knapsack problems is Dempe and Richter (2000), where the authors consider the problem in a purely deterministic setting. The first paper combining the bilevel knapsack problem with random data appeared ten years later (Özaltın et al. 2010). The setting is as follows. In the lower-level problem, the follower solves the knapsack problem

$$\max_{y} \quad d^\top y \quad \text{s.t.} \quad a^\top y \leq b(x,\omega), \; y \in \{0,1\}^n, \tag{18}$$

in which the knapsack's capacity depends on the scenario $\omega \in \Omega$ and on the scalar leader's decision x, where b is a non-decreasing function in x. Moreover, the leader solves the problem

$$\begin{aligned} \max_{x \in \mathbb{R}} \quad & \mathbb{E}_\omega \left[c_y^\top y(x,\omega) - c_x x \right] \\ \text{s.t.} \quad & x \in [x^-, x^+] \subset \mathbb{R}, \quad y(x,\omega) \in S(x,\omega) \text{ for all } \omega \in \Omega, \end{aligned}$$

where $S(x,\omega)$ is the set of optimal solutions of the (x,ω)-parameterized problem (18). Thus, the leader maximizes the value of the items in the knapsack (packed by the follower) but has different values c_y for the separate items, compared to the values d

of the follower. Consequently, we are again in the setting of (16), consider the risk-neutral case, and the uncertainty is modeled by a finite set of scenarios. The authors develop necessary and sufficient conditions for the existence of an optimal solution. Under the additional assumption that the leader's decision can only take integer values, the problem is reformulated as a two-stage stochastic program with binary first- and second-stage decisions that uses the optimal-value function of the lower-level problem. For evaluating the subproblem of this two-stage stochastic program, the authors design a so-called branch-and-backtrack algorithm and then, using this sub-routine, develop a branch-and-cut method to solve the overall problem. Computational results are reported on 16 randomly generated test instances with up to 100 items and 200 scenarios.

Very recently, Buchheim, Henke, and Irmai (2022) considered the continuous bilevel knapsack problem. Here, the lower-level problem is the continuous knapsack problem

$$\max_{y} \quad d(\omega)^\top y \quad \text{s.t.} \quad a^\top y \leq x, \; y \in [0,1]^n,$$

which depends on a random variable ω so that the lower-level's objective function is uncertain. The upper-level player then solves

$$\max_{x \in \mathbb{R}} \quad \mathbb{E}_\omega \left[c_y^\top y(x,\omega) - c_x x \right] \quad \text{s.t.} \quad x \in [x^-, x^+],$$

where $y(x,\omega)$ is a solution of the lower-level problem given above, which is parameterized by the uncertainty ω and the upper-level knapsack capacity decision x. Hence, the leader has different values of the items to be packed and the follower's values of the items are uncertain for the leader. This models the risk-neutral case—for a robust consideration of this setup, we refer the reader to Buchheim and Henke (2022). The authors purely focus on complexity questions. The deterministic problem is known to be solvable in polynomial time; see, e.g., Dempe, Kalashnikov, et al. (2015). First, it is shown that the problem stays polynomial-time solvable if the random variable has finite support, which needs to belong to the input of the problem together with the corresponding probabilities. If, however, the random variable has a finite and componentwise uniform distribution, the problem becomes #P-hard, which is shown by a reduction from #Knapsack.[3] The same hardness result holds true for continuous instead of finite, componentwise uniform distributions. Moreover, even the evaluation of the upper-level objective function is #P-hard for these two cases. Since these results all are hardness results in the weak sense, the authors also derive tailored pseudo-polynomial time algorithms based on dynamic programming. Finally, an additive approximation scheme is derived for arbitrary continuous distributions with independent components. Open questions in this field are the consideration of risk measures other than the expected value as well as the study of cases in which the uncertain lower-level coefficients are correlated.

3.1.6. *General Mixed-Integer Stochastic Bilevel Problems.* Except for Özaltın et al. (2010), where the special situation of a bilevel knapsack problem is studied, no other general mixed-integer stochastic bilevel problems have been considered until the paper by Yanıkoğlu and Kuhn (2018) in which a stochastic bilevel problem is studied with the following properties. The leader decides here-and-now and takes a binary decision without knowing the follower's objective function coefficients and right-hand side. Moreover, the objective function coefficients for the follower's variables in the leader's objective are unknown as well. The uncertainty then reveals

---

[3] The #-symbol indicates that the corresponding counting version of the given decision problem is considered.

and the follower takes his continuous decision having full information. The main focus is on the pessimistic setting, i.e., the bilevel problem at hand is given by

$$\begin{aligned} \min_{x \in X} \quad \sup_{y \in L_n^2} \quad & c_x^\top x + \mathbb{E}_\omega \left[c_y(\omega)^\top y(\omega) \right] \\ \text{s.t.} \quad & y(\omega) \in \operatorname*{arg\,min}_{y' \in \mathbb{R}^n} \left\{ d(\omega)^\top y' \colon Ay' \le b(x, \omega) \right\}. \end{aligned}$$

Here, $X \subseteq \{0,1\}^{n_x}$ and L_n^2 is the set of all n-dimensional square-integrable functions of ω. Hence, the risk-neutral setting with the classic timing (16) is considered but in a rather general setting with randomness both in the upper- as well as in the lower-level problem. The key results from this contribution are that the authors develop both primal and dual decision rules to restrict the search space L_n^2 for the lower-level answers from the point of view of the leader. Both bounding problems are equivalent to MILPs that are not significantly harder than the nominal variant of the original problem. The developed techniques are then applied to a facility location problem with up to 35 locations. To the best of our knowledge, this paper is the first one that specifically focuses on stochastic and pessimistic bilevel optimization problems.

Even more recently, Zhang and Özaltın (2021) consider an even more complicated setting w.r.t. the decision variables since they consider stochastic integer bilevel problems, i.e., problems with purely integer variables both at the upper and the lower level. However, they focus on the optimistic setting while claiming that the developed techniques also work for the pessimistic case. Moreover, randomness only appears in the lower-level's right-hand side and has a discrete distribution with finite support. Their solution approach is based on the value-function reformulation and uses an integer complementarity slackness theorem that is extended to bilevel integer programs. Using these approaches, they finally solve the given class of problems using a tailored branch-and-bound method, which is tested on bilevel facility interdiction problems with up to 200 constraints and 400 variables in the lower level and with up to 50 000 scenarios. It turns out that the approach is rather insensitive w.r.t. the number of scenarios but is sensitive w.r.t. the number of random right-hand sides.

3.2. Robust Approaches to Model Data Uncertainty.

To the best of our knowledge, robust approaches to model uncertainties in the context of bilevel optimization have been much less investigated compared to stochastic approaches. For an overview of robust optimization techniques for single-level optimization, we refer the reader to Ben-Tal, El Ghaoui, et al. (2009), Ben-Tal, Goryashko, et al. (2004), Ben-Tal and Nemirovski (1998), Bertsimas, Brown, et al. (2011), and Soyster (1973). In robust setups, it is assumed that uncertainties take values in a given uncertainty set $\mathcal{U}$. In single-level optimization, uncertainty sets are typically modeled using geometries such as boxes, polyhedra, ellipsoids, or cones. When considered in a bilevel setting, however, the literature so far focuses mainly on either box or polyhedral uncertainty sets.

Chuong and Jeyakumar (2017) study optimistic bilevel problems with linear constraints and a linear lower-level objective function, while the objective of the leader is a polynomial. They allow for interval uncertainties in the upper- and lower-level constraints. Hence, the authors consider the problem

$$\begin{aligned} \min_{x,y} \quad & F(x,y) \\ \text{s.t.} \quad & \tilde{A}_{i\cdot} x + \tilde{B}_{i\cdot} y \ge \tilde{a}_i \quad \text{for all } \left(\tilde{A}_{i\cdot}, \tilde{B}_{i\cdot}, \tilde{a}_i \right) \in \mathcal{U}_i^{\mathrm{L}},\ i \in [m], \\ & x \in \mathbb{R}^{n_x},\ y \in S(x) \end{aligned}$$

with $[m] := \{1, \dots, m\}$. Here, uncertain parameters are emphasized by a tilde, $\tilde{A}_{i\cdot}$ denotes the ith row of the matrix $\tilde{A}$, and $S(x)$ is the set of optimal solutions of the robust lower-level problem

$$\begin{aligned} \min_{y} \quad & d^\top y \\ \text{s.t.} \quad & C_{j\cdot}x + \tilde{D}_{j\cdot}y \geq \tilde{b}_j \quad \text{for all } \left(\tilde{D}_{j\cdot}, \tilde{b}_j\right) \in \mathcal{U}_j^{\mathrm{F}},\ j \in [\ell]. \end{aligned}$$

The uncertainty sets on the upper and the lower level are given by $\mathcal{U}_i^{\mathrm{L}}$ and $\mathcal{U}_j^{\mathrm{F}}$, respectively. The matrix C is assumed to be certain. To solve the robust counterpart of the problem, a sequence of single-level nonconvex polynomial relaxations of the uncertain bilevel problem is introduced, which in turn can be reformulated as a sequence of semidefinite linear problems. The authors show that the optimal values of the relaxed problem converge to the robust global optimal value given that the leader's objective function is coercive and that the lower-level problem satisfies Slater's constraint qualification for every feasible decision of the leader and for every possible realization of the uncertainty in the lower-level constraints.

Uncertainties regarding the lower-level objective function are considered in Borrero et al. (2022) in the context of sequential optimistic linear bilevel problems in which the leader and the follower interact over multiple time periods. The follower's time-invariant objective function coefficients are initially unknown to the leader but they are assumed to take values in a given uncertainty set. In each time period, the leader may refine her perception of the uncertainty set by observing the follower's optimal response, which in turn is based on the leader's decision given her current knowledge of the problem data. Various mechanisms to update the uncertainty set are proposed, which differ in the amount of information from the follower's feedback that is taken into account. The updating process may then characterize the leader's policy, which is a sequence of functions that map the information from previous time periods, i.e., previous upper-level decisions and the observed feedback of the follower, to a feasible leader's decision in the current time period. The authors discuss different policies of the leader that can be interpreted as a robust modeling of a follower who is also uncertain about his objective function coefficients. Results on the convergence of the leader's policies to a full-information solution are provided for different update mechanisms. Moreover, an upper bound on the number of time periods necessary for said convergence is established, which is referred to as "time stability". In the worst-case, the upper bound may be exponential but the presented computational results suggest that this bound is rather loose. To illustrate the performance of the proposed policies, the authors provide numerical experiments on 20 randomly generated road network instances with a layered topology with 4 layers and 4 nodes per layer. Further, experiments are performed on 3 variants of the "infiltration network" near the Arizona-Mexico border with a network of 38 nodes and 109 arcs, which has been described in Unsal (2010).

Recently, Zhang, Liu, et al. (2022) are concerned with the existence of solutions of uncertain multi-leader-multi-follower problems that are modeled as Nash–Stackelberg–Nash games. Uncertainties arise in the objective functions of both the leaders' and the followers' problems as well as in the strategy sets of the followers. In the considered setting, the leaders first take a here-and-now decision that particularly influences the description of the uncertainty set. Thus, the authors consider a special type of decision-dependent uncertainty; see also Lappas and Gounaris (2018) and Nohadani and Sharma (2018) for general works on decision-dependent uncertainty. Then, the uncertainty realizes and, finally, the followers decide on their actions taking the leaders' decision as well as the realization of uncertainty into account. To hedge against decision-dependent uncertainties, a worst-case approach is considered.

Buchheim and Henke (2020, 2022) are concerned with complexity questions for the bilevel continuous knapsack problem in which the leader controls the capacity of the knapsack and the follower faces uncertainties regarding the profits of the items. The authors' main focus is on the pessimistic setting and a worst-case oriented approach is pursued to account for data uncertainty. While the deterministic problem can be solved in polynomial time (Dempe, Kalashnikov, et al. 2015), the complexity of the robust variant of the problem strongly depends on the considered type of the uncertainty set. First, the authors show that the robust counterpart remains solvable in polynomial time for discrete uncertainty sets as well as for interval uncertainty under the independence assumption, i.e., if the follower's objective function coefficients independently take values in given intervals. However, the problem is NP-hard if the uncertainty set is the Cartesian product of discrete sets. In particular, this means that replacing the uncertainty set by its convex hull can significantly change the problem in the bilevel context, which is in contrast to the situation in single-level optimization. NP-hardness is also shown for the variants of the problem with polytopal uncertainty sets and uncertainty sets that are defined by a p-norm with $p \in [1, \infty)$. Moreover, even the evaluation of the leader's objective function is NP-hard for the latter three cases.

The complexity of robust bilevel combinatorial problems with a linear follower facing an uncertain objective function is addressed in Buchheim, Henke, and Hommelsheim (2021). The deterministic variant of the problem is known to be NP-easy.[4] It is shown that interval uncertainty renders the bilevel problem significantly harder than the consideration of discrete uncertainty sets. To be more precise, the robust counterpart can be Σ_2^P-hard for interval uncertainty under the independence assumption, whereas it can be NP-hard for uncertainty sets $\mathcal{U}$ with $|\mathcal{U}| = 2$ and strongly NP-hard for general discrete uncertainty sets. In particular, it is shown that replacing the discrete uncertainty set by its convex hull may increase the complexity of the problem at hand, which is in line with the results in Buchheim and Henke (2020, 2022).

All papers discussed so far deal with data uncertainties in the sense of strict robustness. Beck, Ljubić, et al. (2022) propose a Γ-robust approach for mixed-integer linear min-max problems with lower-level data uncertainty. They follow the notion of Γ-robustness introduced by Bertsimas and Sim (2003, 2004) to account for an uncertain follower's objective, i.e., they study the problem

$$\begin{aligned}
\min_x \quad & c^\top x + d^\top y \\
\text{s.t.} \quad & Ax \geq a,\ x \in X \subseteq \mathbb{Z}^{n_x}, \\
& y \in \operatorname*{arg\,max}_{\bar{y} \in Y(x)} \left\{ d^\top \bar{y} - \max_{\{S \subseteq [n_y]\colon |S| \leq \Gamma\}} \sum_{i \in S} \Delta d_i \bar{y}_i \right\}
\end{aligned}$$

with $\Gamma \in [n_y]$ and the lower-level feasible set $Y(x) \subseteq \mathbb{Z}_+^{n_y}$. Furthermore, uncertainties in a single packing-type constraint are considered. Two approaches to model this situation are presented—an extended formulation and a multi-scenario formulation. In particular, the authors establish that the Γ-robust bilevel problem can be interpreted as a single-leader-multi-follower problem with independent followers in the case that all of the follower's variables are binary. A branch-and-cut framework to solve the robustified bilevel problem is proposed. As an application, the authors consider the knapsack interdiction problem for which problem-tailored cuts are provided; see, e.g., Section 4.2. The authors conduct a computational study on

---

[4] A decision problem is NP-easy if it can be polynomially reduced to an NP-complete decision problem (Buchheim, Henke, and Hommelsheim 2021).

200 robustified knapsack interdiction instances with up to 55 items, which are based on the nominal instances described in Caprara et al. (2016).

3.3. Lower-Level Near-Optimality.

In the classic setting of bilevel optimization, it is assumed that the leader anticipates an optimal reaction of the follower. In many practical applications, however, exact solutions of the lower-level problem cannot be expected. Possible reasons might be that an exact solution cannot be obtained in a reasonable amount of time or that there is simply no exact solution method to solve the problem at hand. Hence, the follower will take any "satisfactory" decision instead of an optimal one, i.e., one considers the set of ε-optimal follower's decisions

$$S(x, \varepsilon) = \{y \in Y : g(x, y) \geq 0, \ f(x, y) \leq \varphi(x) + \varepsilon\}, \quad \varepsilon > 0,$$

for a feasible decision x of the leader. Here, the parameter ε specifies the follower's willingness to deviate from his optimal objective function value. Since the leader faces response uncertainty due to the follower's near-optimality, this modeling approach accounts for decision uncertainty w.r.t. the lower-level player. This is in contrast to the concepts presented in the previous sections that cover data uncertainty.

Different approaches to deal with an ε-optimal follower have been studied in the literature. Maybe the first appearance of an ε-optimal follower can be found in Loridan and Morgan (1989), where the ε is used as a regularization parameter in order to prove existence results for the bilevel problem at hand.

Besançon et al. (2019) consider the effect of near-optimality on the upper-level constraints by exploiting a robust approach to hedge against deviations from the optimal reaction of the follower. In this setting, an optimal solution of the overall bilevel problem is required to remain feasible for all ε-optimal follower's decisions. To this end, the leader hedges against the worst-possible reaction of the follower by considering the problem

$$\begin{aligned}
\min_{x,z} \quad & F(x, z) \\
\text{s.t.} \quad & f(x, z) \leq \varphi(x), \ g(x, z) \geq 0, \ x \in X, \ z \in Y, \\
& \min_{y} \{G(x, y) : y \in S(x, \varepsilon)\} \geq 0.
\end{aligned}$$

Here, the leader controls the variables z that model the follower's optimal response, i.e., $z \in S(x, 0) \subseteq S(x, \varepsilon)$, and the upper-level constraints are protected against ε-optimal follower's decisions in a robust way. Based on the Karush–Kuhn–Tucker (KKT) conditions of the lower-level problem, a single-level reformulation of the bilevel problem with an ε-optimal follower is provided if the lower level is a convex problem. Finally, the authors propose a solution method for purely linear near-optimal robust bilevel problems. The applicability of the proposed method is assessed in a computational study on 1200 randomly generated linear near-optimal robust bilevel instances with up to 20 variables and 20 constraints each on the upper and the lower level.

The complexity of near-optimal robustness concepts is analyzed in Besançon et al. (2021). The authors not only consider near-optimal robust bilevel problems but also investigate general multilevel optimization problems with an ε-optimal decision maker at an arbitrary lower level of the problem. Under suitable assumptions, they show that the robust modeling of near-optimality at a lower level remains in the same complexity class as the problem without uncertainty.

Motivated by military and law-enforcement applications, Zare, Özaltın, et al. (2018) consider bilevel problems with a follower that willingly deviates from his optimal objective function value to adversely affect the leader. For a feasible decision x of the leader, the authors define the set of near-optimal follower's decisions as

$$S_\alpha(x) = \{y \in Y \colon f(x,y) \le \alpha\varphi(x) + (1-\alpha)U,\, g(x,y) \ge 0\}\,.$$

Here, U is an upper bound for the lower-level objective function value and $\alpha \in [0,1]$ denotes the follower's willingness to deviate from his optimal objective function value. Note that for $\alpha = 1$, the set of (exact) optimal decisions of the follower is considered, whereas any feasible follower's decision can be chosen by the follower for $\alpha = 0$. This notion of near-optimality is considered in the context of pessimistic bilevel optimization (see, e.g., Liu, Fan, et al. (2020), Tsoukalas et al. (2009), and Wiesemann et al. (2013)) by introducing the so-called α-pessimistic bilevel problem

$$\min_{x \in X} \max_{y \in S_\alpha(x)} F(x,y).$$

The authors focus on the case in which (i) there are no coupling constraints and (ii) the functions F, f, and g are affine. Furthermore, they allow for integer variables on the upper level. The proposed model accounts for different levels of conservatism regarding the uncertainty of the follower's commitment. In particular, the α-pessimistic bilevel problem includes the standard pessimistic bilevel problem as well as the min-max problem by either setting $\alpha = 1$ or $\alpha = 0$, respectively. As an extension of the proposed model, the authors further embed an α-pessimistic follower into the context of strong-weak bilevel problems; see Section 3.5.

Pita, Jain, Ordóñez, et al. (2009), Pita, Jain, Tambe, et al. (2010), and Pita, Portway, et al. (2008) address ε-optimal follower's decisions in the context of Bayesian Stackelberg games (Conitzer and Sandholm 2006). They consider the setting in which there is only one leader type, e.g., a security entity, and multiple follower types, e.g., multiple attacker types. Each follower type may select an ε-optimal strategy and the actual follower type is unknown to the leader. To hedge against near-optimality of each follower type, the leader pursues a worst-case oriented approach. Given an a priori probability distribution over the follower types, the leader then optimizes over expected values under the worst-case assumption. The authors further combine the modeling of near-optimal follower types with the concept of limited observability regarding the leader's strategy, which is discussed in more detail in the following section.

In many practical applications, the lower-level problem cannot be solved to global optimality either because there is no exact solution method available or due to tractability reasons. Beck, Schmidt, et al. (2022) consider an illustrative example of a bilevel problem with a continuous but nonconvex lower-level problem for which only ε-feasible solutions—at least for the nonlinear constraints of the lower-level problem—can be anticipated. The authors show that such ε-feasible bilevel solutions can be arbitrarily far away from the overall exact optimal solution of the bilevel problem.

Zare, Prokopyev, et al. (2020) study the setting in which the follower can choose any solution method out of a discrete set $\mathcal{H}$ of possible options that may include exact methods, heuristics, or approximation algorithms. The leader knows this set of potential solution methods but she is uncertain regarding the actual choice of the follower. To hedge against sub-optimal follower's decisions that may stem from the follower's use of an inexact solution method, the authors propose three modeling approaches. First, they follow a robust approach by hedging against the worst-possible choice of the follower's solution algorithm. The leader then solves the problem

$$\min_{x \in X} \max_{h \in \mathcal{H}} F(x, y^h),$$

where y^h denotes the "solution" of the lower-level problem using method $h \in \mathcal{H}$. Let us point out that the proposed model does not contain coupling constraints but allows for integer upper- and lower-level variables. The second approach follows to some extent the notion of Γ-robustness, which has been proposed in Bertsimas and Sim (2003) for single-level optimization. Here, the previous model is adapted such that the leader only hedges against the Γth least damaging choices of the solution algorithm for the lower-level problem. The parameter $\Gamma \in \{1, \dots, |\mathcal{H}|\}$ is used to control the leader's level of conservatism. In contrast to the aforementioned modeling approaches, the third model relies on the leader's prior knowledge about the probability p_h that the follower will use method $h \in \mathcal{H}$. Hence, the leader hedges against lower-level algorithmic uncertainty in a probabilistic sense by optimizing over the expected value, i.e.,

$$\min_{x \in X} \ \mathbb{E}_h \left[F(x, y^h) \right] := \sum_{h \in \mathcal{H}} p_h F(x, y^h).$$

The authors provide a detailed discussion of the proposed models for bilevel knapsack problems with a follower that may choose from an exact cutting plane method or several greedy approaches. In their computational study, the authors focus on defender-attacker problems that can be formulated as bilevel knapsack models. Numerical results are presented for 32 randomly generated instances with 15 items in which the follower can choose from an exact method and two greedy heuristics.

Shi et al. (2020) pursue the same idea as Zare, Prokopyev, et al. (2020) and study bilevel problems with an inexact follower. However, instead of specifying the available solution methods, the authors exploit the notion of k-optimality to capture local optimal solutions of the follower. The modeling of a k-optimal follower is then used to derive valid lower and upper bounds for the original mixed-integer linear bilevel problem in which the lower level is solved to global optimality. An extensive computational study on randomly generated instances of the knapsack interdiction problem, the bilevel vertex cover problem, and the bilevel minimum spanning tree problem affirms the quality of the proposed bounds.

3.4. Limited Observability.

Another approach to account for decision uncertainty in bilevel optimization considered in the literature is known under the notion of limited observability. In this setting, the follower may, e.g., not be able to perfectly observe the leader's decision due to cognitive limitations and the leader thus faces follower's response uncertainty. This aspect plays a significant role for many practical applications—especially in defender-attacker scenarios; see Section 4.

To the best of our knowledge, uncertainties regarding the observability of the upper-level decision are first addressed in Bagwell (1995) and van Damme and Hurkens (1997). The authors consider Stackelberg games involving noise in the follower's observation of the leader's strategy. It is shown that the leader's first-mover advantage is completely lost for pure-strategy equilibria of the "noisy Stackelberg game". Nevertheless, a mixed-strategy equilibrium may exist for which the outcome converges to the outcome of the Stackelberg game under perfect observability.

Following a different approach to account for limited observability of the leader's decision, Pita, Portway, et al. (2008) are concerned with Bayesian Stackelberg games in which the leader's decision $\bar{x}$ that is perceived by the follower may deviate from the actual leader's strategy x by a bounded observation error δ. A discrete set of observation errors—which is known in advance—specifies the strategies of the leader the follower might observe. As the follower assumes that $\bar{x}$ is the strategy the leader actually plays, the follower's response is thus based on $\bar{x}$ instead of x. The leader then hedges against the worst-possible reaction of the follower due to his erroneous observation. In particular, the authors model limited observability for the setting in which there is only one leader type, e.g., a security entity, and multiple follower types, e.g., multiple types of attackers. The leader does not know the actual follower type but has prior knowledge on the probability distribution over the follower types and thus optimizes over expected values. Pita, Jain, Ordóñez, et al. (2009) and Pita, Jain, Tambe, et al. (2010) provide an extension of the previous approach by using anchoring theory (Kelly et al. 2006; Tversky and Koehler 1994) to model the follower's limited capability of observing the actual decision of the leader. Given that—due to limited observability—the follower does not have any information on the actual decision of the leader, the follower tends to assign an equal probability, i.e., a uniform distribution, to each feasible leader's strategy. The more information is revealed to the follower, e.g., via further observations, the more the follower relies on his perception of the leader's strategy. Hence, the perceived decision of the leader $\bar{x}$ is represented as

$$\bar{x}_i = \alpha \frac{1}{N} + (1-\alpha)x_i, \quad i \in [n_x], \tag{19}$$

where N denotes the number of pure strategies available to the leader and $\alpha \in [0,1]$ indicates how much weight the follower leaves on the uniform distribution. To be comprehensive, let us further mention that the discussed articles provide an additional extension by combining the concept of limited observability of the leader's decision with near-optimal decisions of the follower.

A similar setting as in Pita, Jain, Ordóñez, et al. (2009), Pita, Jain, Tambe, et al. (2010), and Pita, Portway, et al. (2008) is considered in Yin and Tambe (2012). In contrast to the literature discussed so far, however, they model the perceived leader's decision $\bar{x}$ as a linear perturbation of the actual leader's decision x, i.e., $\bar{x} = F^\top x + f$, where F and f have appropriate dimensions and are assumed to follow some known continuous distribution. Moreover, the authors allow for data uncertainty and a noisy execution of the leader's strategy. A branch-and-cut method to solve this type of Bayesian Stackelberg game is proposed that incorporates Benders decomposition and heuristic branching rules. The method is assessed using 30 Stackelberg games with randomly generated utility matrices, 5 pure strategies per player, and up to 200 follower types.

Due to the vast amount of publications on limited observability in the context of Bayesian Stackelberg games or, more generally, Stackelberg security games, the literature discussed so far is far from being comprehensive. Hence, we refer the interested reader to Jain et al. (2010), Kar et al. (2018), Kiekintveld et al. (2011), Paruchuri, Pearce, Marecki, et al. (2008), Paruchuri, Pearce, Tambe, et al. (2007), Pita, Jain, Ordóñez, et al. (2009), Pita, Jain, Tambe, et al. (2010), Pita, Portway, et al. (2008), Sinha et al. (2018), Yin, Jain, et al. (2011), and Yin and Tambe (2012), as well as to the references therein.

Beck and Schmidt (2021) consider limited observability regarding the upper-level decision for bilinear bilevel problems. The authors assume that the perceived leader's decision $\bar{x}$ only takes values in a given polyhedral uncertainty set

$$\mathcal{X}(x) = \{x + P\zeta \colon H\zeta \geq h\},$$

where x denotes the actual decision of the leader and P, H, h as well as the perturbation vector ζ have appropriate dimensions. The leader hedges against follower's response uncertainty due to limited observability by pursuing a worst-case oriented approach. A single-level reformulation of the robustified bilevel problem is obtained by exploiting the KKT conditions of the lower-level problem. In particular, it is shown that the robustified bilevel problem remains in the same problem class as the problem without taking limited observability into account. Moreover, an ex-post relation between the modeling of limited observability and robust bilevel problems with an uncertain right-hand side of the lower level is established.

Korzhyk et al. (2011) consider a game in extensive form to model limited observability. To this end, nature is introduced as an additional player in the game. The game then proceeds as follows. Based on a given probability distribution, nature first determines whether the follower can perfectly observe the leader's decision or not. Second, the leader selects a probability distribution over her set of pure strategies without knowing the decision made by nature. Finally, the follower responds optimally after possibly observing the leader's decision depending on nature's choice. The authors propose a method that alternates between solving Nash and Stackelberg games in each iteration. Upper bounds on the number of iterations to obtain an equilibrium are provided.

In Karwowski et al. (2020), limited observability of the leader's strategy is again modeled using anchoring theory. The authors extend the approach proposed in Pita, Jain, Ordóñez, et al. (2009) and Pita, Jain, Tambe, et al. (2010) to account for multiple time steps by considering multi-step extensive-form games. In this setting, a fixed number of leader-follower games is played successively and, in each round of the game, the perceived leader's strategy takes the form in (19). This leads to nonlinearities which cannot be tackled by classic solution methods. Hence, the authors propose to simplify this modeling of limited observability by assuming that only the leader's decision in the last round of the game cannot be perfectly observed such as to avoid nonlinear terms. This modeling of limited observability is embedded in three state-of-the-art MILP methods as well as two other methods (Monte-Carlo sampling and an evolutionary algorithm) for solving sequential Stackelberg games, whose performance is assessed in a computational study on 25 warehouse games with up to 7 time steps.

Lu et al. (2015) introduce a novel notion of robustness for bilevel problems with upper- and lower-level decision uncertainty. In their modeling, the perceived leader's and follower's decisions are assumed to take values in a small neighborhood around the actual strategies. They pursue a robust approach in the sense that an optimal solution of the bilevel problem must be feasible for all possible realizations of the uncertainty. In contrast to traditional robustness concepts (Ben-Tal, El Ghaoui, et al. 2009; Bertsimas and Sim 2003), however, they do not follow a worst-case oriented philosophy. Instead, they hedge against uncertainties by either optimizing mean objective function values or by optimizing the relative deviations from the (exact) optimal objective function values w.r.t. a predefined tolerance.

Finally, Molan and Schmidt (2022) consider one of the extreme cases of limited observability and assume that the leader does not know the optimization problem of the follower at all but tries to learn the best-response function based on past data regarding the outcomes of the same bilevel game. The authors use a neural network to learn the best-response function and apply a tailored Lipschitz optimization approach to solve the resulting optimization problem that contains the input-output mapping of the trained neural network as a constraint.

3.5. Intermediate Solution Concepts in Between the Optimistic and the Pessimistic Case.

In general, the bilevel problem as defined in (1) and (2) is ill-posed if the lower-level problem does not have a unique solution. To overcome this issue, it is common to pursue either the optimistic or the pessimistic approach to bilevel optimization; see, e.g., Dempe (2002). The optimistic (or strong) approach corresponds to the setting in which the leader and the follower fully cooperate. Thus, the follower chooses his solution such as to favor the leader w.r.t. the leader's objective function value, i.e., the leader considers the problem

$$\min_{x \in \bar{X}} \min_{y \in S(x)} \quad F(x, y)$$

with $\bar{X} := \{x \in X : G(x) \geq 0\}$ and $G : \mathbb{R}^{n_x} \to \mathbb{R}^m$. Here, and in the remainder of this section, we focus on the setting without coupling constraints. In the pessimistic (or weak) setting, the follower aims to harm the leader by selecting the worst-possible reaction in terms of the leader's objective function value. The leader thus considers the problem

$$\min_{x \in \bar{X}} \max_{y \in S(x)} \quad F(x, y).$$

In particular, the pessimistic bilevel problem is a special case of a robust problem in which the set $S(x)$ is interpreted as the uncertainty set. In the literature, the optimistic approach is predominantly used. However, the general pessimistic setting has gained increasing attention recently; see, e.g., Aboussoror and Mansouri (2005), Aussel and Svensson (2019), Tsoukalas et al. (2009), Wiesemann et al. (2013), and Zheng, Wan, Sun, et al. (2013) or the recent surveys in Liu, Fan, et al. (2020, 2018) and the references therein. Nevertheless, the optimistic and the pessimistic approach represent two extreme situations regarding the follower's level of cooperation, which may be inappropriate for certain applications, e.g., if the leader is uncertain regarding the level of cooperation of the follower. The resulting necessity for intermediate solution concepts is, e.g., tackled by so-called strong-weak bilevel problems. This modeling approach allows for a partial cooperation of the follower by considering a convex combination of the leader's objective functions in the optimistic and the pessimistic setting. So far and to the best of our knowledge, intermediate solution concepts have only been considered in the literature for bilevel problems without coupling constraints, i.e., one considers the problem

$$\min_{x \in \bar{X}} \quad \left\{ \beta \min_{y \in S(x)} F(x, y) + (1 - \beta) \max_{y \in S(x)} F(x, y) \right\}.$$

Here, the parameter $\beta \in [0, 1]$ can be interpreted as the follower's probability of cooperation and is used to adjust the leader's level of conservatism. Note that for $\beta = 1$, the optimistic setting is considered, whereas for $\beta = 0$, this modeling corresponds to the standard pessimistic bilevel problem.

Among the first works that address intermediate solution concepts in between the optimistic and the pessimistic case are Aboussoror and Loridan (1995) and Mallozzi and Morgan (1996). In Mallozzi and Morgan (1996) the following two cases are distinguished. First, the authors consider the case in which the set of optimal follower's decisions is discrete, i.e., $S(x) = \{y^1(x), \dots, y^k(x)\}$. It is assumed that the leader has prior knowledge on the likelihood that the follower will choose a certain solution and model this setting in a stochastic sense by optimizing over expected values. Hence, the authors consider the problem

$$\min_{x \in X} \quad \sum_{j=1}^{k} p_j(x) F(x, y^j(x)),$$

where $p_j(x)$ denotes the probability that the follower chooses $y^j(x)$. Second, more general reaction sets $S(x)$ of the follower are considered which are assumed to be Lebesgue measurable with non-zero measure. In this case, the leader assigns a probability measure $\mu_x(y)$ on $S(x)$ for every feasible upper-level decision $x \in X$. The authors then consider the problem

$$\min_{x \in X} \quad \int_{S(x)} F(x, y) \, \mathrm{d}\mu_x(y).$$

For both cases, illustrative examples and comparisons of the obtained results with the pessimistic formulation are provided.

Sufficient conditions for the existence of solutions of strong-weak bilevel problems in finite-dimensional spaces are established in Aboussoror and Loridan (1995).

The authors show that a sequence of ε-optimal solutions converges to an optimal solution of the original problem under sufficient conditions. Similarly, existence results for the variant of the problem in infinite-dimensional spaces are provided in Aboussoror, Adly, et al. (2017). In particular, and in contrast to the techniques used in Aboussoror and Loridan (1995), no convexity assumptions are required to establish these existence results.

It is well known that bilevel problems are intrinsically hard to solve. Even linear (optimistic) bilevel problems are strongly NP-hard; see, e.g., Hansen et al. (1992). Since the consideration of a pessimistic follower adds another level to the problem formulation, pessimistic bilevel optimization—and thus strong-weak bilevel optimization—is expected to be even more challenging. When it comes to developing solution methods, the literature so far focuses on the easiest instantiations of intermediate bilevel formulations, namely linear strong-weak bilevel problems. Cao and Leung (2002) propose a penalty-based approach for this type of problem. They reformulate the strong-weak problem as a classic linear bilevel problem which can be tackled using standard solution approaches, e.g., by solving the KKT reformulation. To obtain an optimal intermediate solution, however, the resulting problem needs to be solved for all possible values of the exogenous parameter β, which models the follower's level of cooperation. This issue is addressed in Zheng, Wan, Jia, et al. (2015). The authors provide a method that avoids the "enumeration" over all $\beta \in [0, 1]$ to determine the critical points of the optimal-value function of the leader.

A relaxation-and-correction scheme to solve the linear strong-weak bilevel problem is presented in Zeng (2020). The author proposes a relaxation of the original problem by including two sets of variables and constraints of the follower—one for the optimistic and one for the pessimistic case—and relaxing the optimality of the pessimistic follower. The original problem is thus reduced to a single-leader-multi-follower problem with two independent followers. Using the KKT conditions of each of the lower-level problems, a single-level reformulation is obtained, which can be solved by state-of-the-art solvers. Afterward, the optimality of the pessimistic follower's solution is ensured via a correction step.

All the mentioned approaches so far take the follower's level of cooperation β as a parameter that is specified by the leader in advance. Jia, Wan, et al. (2011) introduce the concept of considering the follower's level of cooperation as a decision-dependent random function $\beta(x)$ of the leader's variables x. Salas and Svensson (2020) follow this approach in the context of multi-leader-multi-follower problems. In this setting, the classic optimistic and pessimistic approaches are ill-defined since the cooperation of the followers with one of the leaders may result in non-cooperation or partial cooperation with another leader. To overcome this issue, the authors assume that each leader i has a "belief" β^i regarding the followers' choice. The belief assigns a probability measure to an optimal solution of the followers for each feasible leaders' decision $x \in X$. Hence, the followers' response is a random variable that follows the decision-dependent probability distribution $\beta^i(x)$. Each leader then hedges against uncertainties regarding the followers' response by optimizing over expected values. The authors provide results on the existence of optimal solutions for this setting.

Jia and Zheng (2013) consider another intermediate solution concept that allows the leader to make side-payments to the follower. In this setting, the leader willingly gives up a portion of her optimal objective function value to offer an incentive for the follower to cooperate, i.e., one considers the problem

$$\begin{aligned} \max_{x,y,\beta} \quad & \beta F(x,y) \\ \text{s.t.} \quad & f(x,y) + (1-\beta)F(x,y) \geq \bar{\varphi}(x), \\ & x \in \bar{X},\ y \in S(x),\ \beta \in [\alpha, 1] \end{aligned}$$

with the lower-level optimal-value function

$$\bar{\varphi}(x) := \max_{y \in Y} \left\{ f(x,y) \colon g(x,y) \geq 0 \right\}.$$

Here, $\beta \in [\alpha, 1]$ determines how much of the leader's optimal objective function value is given to the follower as a side-payment. The parameter $\alpha \in [0,1]$ can be interpreted as a minimum allocation proportion. In particular, β is understood as a variable of the problem, which is optimized in the proposed procedure. For the linear formulation of the problem, the authors present a solution method that relies on a penalty-based reformulation and that exploits strong duality.

4. Applications

In this section, we review different areas of application in which bilevel optimization under uncertainty is used. We start with energy applications in Section 4.1, discuss the large field of interdiction problems in Section 4.2, and end in Section 4.3 with a discussion of applications from management science.

4.1. **Energy.** There is a rather large amount of research papers in which uncertain bilevel optimization is applied to the field of energy research. Before we review the separate contributions in detail, let us first briefly summarize the main commonalities of the research in this area.

(i) In the majority of papers, a linear or bilinear setting is considered. If a bilinearity is present, it usually consists of the mix of an upper- and a lower-level variable so that the lower-level problem stays a parameterized linear problem for which compact optimality certificates such as the strong duality or the KKT theorem are available.

(ii) Except for Heitsch et al. (2022), see also Section 3.1.4, all papers consider the timing in (16), i.e., the leader takes her decision here-and-now, then the uncertainty realizes, and the follower afterward takes a wait-and-see decision.

(iii) The vast majority of papers considers a stochastic setup and only a few papers also include robustness aspects. The considered stochastic setting is usually given by a discrete and finite probability distribution that allows for stating the deterministic equivalent.

4.1.1. *Power Markets, Contracting, and Networks.* An important application of bilevel optimization in the field of electricity is the optimization of retailer problems. Maybe the earliest application of uncertain bilevel optimization in this area is Carrión et al. (2009), where the authors consider stochastic bilevel optimization for determining optimal retailer trading strategies in future markets. Here, the retailer acts as the leader and takes her here-and-now decisions under uncertainty regarding future pool prices, the clients' demands, and the prices of the rivaling retailers. The clients act as the follower in a wait-and-see manner. Risk aversion in the upper-level is modeled using a weighted sum of the expected value and the CVaR. Moreover, stochasticity is covered by using a finite set of scenarios, which can additionally be reduced in size by exploiting further scenario reduction techniques as, e.g., discussed in Growe-Kuska et al. (2003). This allows to consider the deterministic equivalent, which has a linear lower-level problem (in the follower's variables) so that the classic

KKT approach with big-M reformulations à la Fortuny-Amat and McCarl (1981) can be applied. The remaining bilinearities that are all products of prices and quantities are then linearized using duality theory.

In Askeland et al. (2020), the authors also consider a retailer's problem but try to figure out how to use stochastic bilevel optimization to design electricity network tariffs to incentivize flexible end-users (so-called prosumers) to adapt their consumption patterns. To this end, the upper level determines the tariff at the planning stage and the curtailment of load at the operational stage while the lower level models end-users of electricity, which can either be consumers or prosumers. Thus, the lower-level problem models the operational decisions. Uncertainty is again modeled via a finite scenario set and the classic KKT reformulation is used to obtain a single-level optimization problem in which the nonlinear KKT complementarity constraints are tackled via SOS-1 techniques.

A related setting is considered in Fanzeres, Street, et al. (2015), where the authors study a hybrid approach by mixing stochastic and robust optimization to determine optimal energy supply contracting strategies. A single-level reformulation is obtained via strong duality and, as before, stochasticity is modeled using a finite number of scenarios.

Another important application of bilevel optimization in the field of power markets is to determine the optimal bidding strategy of a strategic generator in wholesale electricity markets; see, e.g., Fampa et al. (2008). There, the authors consider the classic setup in the sense that the strategic generator is the upper-level player. However, the lower-level problem does not cover a Nash equilibrium problem to model the rival's behavior at the market but models the (unknown) actions of the competitors using a finite set of scenarios with exogenous probabilities. In this setup, the upper level maximizes the expected profit of the strategic generator while in the lower-level problem, a market clearing is determined that minimizes the overall system costs. After some reformulations, the model at hand has a large similarity to bilevel models considered for optimal taxation of goods and services such as discussed in, e.g., Brotcorne et al. (2000) and Labbé, Marcotte, et al. (1998). A tailored heuristic is proposed that is motivated by the latter papers and to solve the problem to global optimality, the KKT reformulation is used including binary decompositions of integer variables to linearize the occurring nonlinearities.

In Haghighat (2014), the same overall question is tackled but with a transmission-constrained economic dispatch model in the lower level. Again, uncertainty enters the model in order to capture the unknown offers of the rivals and the market demand. In contrast to Fampa et al. (2008), however, uncertainty is handled using the Γ-robust approach (Bertsimas and Sim 2004). After deriving the robust counterpart using duality-based techniques of continuous robust optimization, a single-level reformulation is obtained by using the classic KKT approach and further nonlinearities are handled via big-M constraints and relaxation-linearization techniques (Sherali and Adams 1990). Finally, a hybrid model is presented that also takes stochasticity into account by a finite number of scenarios in order to model uncertain quantity offers of rival generators.

Ambrosius, Egerer, Grimm, and Weijde (2020) and Ambrosius, Egerer, Grimm, and van der Weijde (2022) consider multilevel optimization models for electricity markets that tackle the situation in which investment decisions have to be taken subject to uncertainty w.r.t. the future network congestion management regime such as nodal or zonal pricing.

All papers discussed so far in this section model stochasticity by considering finite scenario sets and by writing down the deterministic equivalent. For large scenario sets, this leads to large-scale single-level reformulations that might be hard to solve even for state-of-the-art commercial solvers. This is especially the case for modeling strategic investment decisions that need to incorporate a large planning horizon and, thus, very large scenario sets. This issue is considered in Kallabis et al. (2020), where the authors tackle this situation using a rolling-horizon approach. Here, the strategic investment by the generator is part of the upper level, whereas the lower level includes the decisions of the market operator that clears the market in order to maximize welfare for given consumer demands, given installed generation capacities, and given price bids of the producers. The investment process is split into multiple stages so that wait-and-see decisions can be modified over time. The rolling-horizon approach is then applied to an MPEC that is obtained by the deterministic equivalent for given scenario trees of electricity demand and by using KKT conditions for the lower level as well as SOS-1 techniques to tackle KKT complementarity conditions.

In another very recent contribution (Zeng, Dong, et al. 2020), the optimal configuration of electricity vehicle charging stations and the corresponding pricing schemes are studied. These two decisions are modeled in the upper level and the lower-level problem comprises the actual charging decisions of plug-in electric vehicle owners. The lower-level objective function is of min-max type and, thus, models a robust setup. The classic KKT reformulation leads to a single-level problem with a max-min-max structure in the objective, which is then solved using the column-and-constraint generation method developed by Zeng and Zhao (2013).

Furthermore, Kovacevic and Pflug (2014) survey bilevel modeling of electricity swing option pricing. The authors carefully develop their model, which leads to a stochastic and dynamic multistage bilevel problem. The survey also discusses solution techniques—in particular for bilinear swing option problems.

4.1.2. *Gas Market and Further Energy Applications.* Besides applications in the power sector, there are also other energy sectors that have been modeled using uncertain bilevel optimization problems. For instance, in U-tapao et al. (2016), the authors set up a bilevel model for optimizing wastewater treatment plans by deciding on the size of compressed natural gas (CNG) filling stations and their locations. The lower-level problem consists of many downstream markets including agriculture, CNG transportation, residential natural gas, and electricity markets. Each downstream market is modeled by its own KKT conditions plus suitably chosen market clearing conditions. Uncertainties stem from, e.g., fuel or electricity prices and are modeled using a finite set of scenarios. It is assumed that there is no correlation between these uncertain aspects—which is, what the authors admit, a simplification. The overall setting leads to an SMPEC, which is reformulated as a mixed-integer linear problem via SOS-1 techniques.

Another branch of research on the European natural gas market started rather recently with the modeling paper by Grimm et al. (2019). There, a four-level model of the so-called European entry-exit gas market system is developed and it is shown that this system can be reduced to a bilevel problem under suitable economic assumptions such as perfect competition. The upper level then consists of the decisions of the transmission system operator (TSO) whereas the lower-level problem models the long- to short-term market behavior of gas buyers and sellers. Due to the EU regulation, the upper-level problem contains a robust constraint so that the overall problem of the TSO can be seen as a special case of an adjustable robust problem (Ben-Tal, El Ghaoui, et al. 2009; Ben-Tal, Goryashko, et al. 2004); see, e.g., Labbé, Plein, Schmidt, and Thürauf (2021) for an in-depth discussion of this relationship. The overall model is rather challenging; see Labbé, Plein, and Schmidt (2020), Labbé, Plein, Schmidt, and Thürauf (2021), Schewe et al. (2020), and Thürauf (2022) for complexity studies and Böttger et al. (2021), Plein et al.

(2021), and Schewe et al. (2022) for solution techniques and numerical results. Let us finally remark that the paper by Heitsch et al. (2022), which we already discussed in Section 3.1.4, also considers this bilevel problem with additional uncertainties in the lower-level problem that are modeled using a chance constraint.

4.2. **Interdiction, Defender-Attacker, and Security Applications.** Interdiction problems are a special class of bilevel optimization problems in which the leader (who acts as a defender) aims at preventing adversary activities of the follower (who acts as an attacker). They are typically used for identifying vulnerabilities of a system when it comes to potential disruptions (being accidental or intentional) to its infrastructure. Interdiction problems follow the common structure of bilevel problems without coupling constraints, namely

$$\max_{x \in \bar{X}} \left\{ \min_{y \in Y} f(x,y) \colon g(x,y) \geq 0 \right\},$$

where the set $\bar{X} = \{x \in X \colon G(x) \geq 0\}$ describes feasible interdiction policies. Let N_y denote the assets that can be interdicted by the leader. Decisions of the leader and the follower are linked through constraints $g(x,y) \geq 0$, which are typically given as $y_i \leq U_i(1 - x_i)$ for all $i \in N_y$. Here, U_i is the default available capacity of asset i (if not interdicted). Interdiction actions (modeled by variables x_i) can be discrete (in which case the assets are made unavailable for the follower) or continuous (if the capacity of the asset is modified, based on the intensity of interdiction). If the interdiction affects only the objective function of the follower, then their nominal objective function value determined as $\sum_{i \in N_y} d_i y_i$ is modified by adding a bilinear term $\sum_{i \in N_y} \delta_i y_i x_i$ to it, where δ_i represents the cost increase for each asset interdicted by the leader. For a comprehensive survey on interdiction problems, we refer to Smith and Song (2020) and Section 6 in Kleinert et al. (2021). Interdiction problems on networks are the most frequently studied problem variants in which the leader controls the network resources (nodes, edges) by eliminating them, reducing their capacities, or increasing the costs of their usage.

In more realistic settings of defender-attacker games, either party may not have complete information about their opponent's strategy or about the underlying conditions such as the network topology, arc or node costs, or their capacities. Hence, interdiction problems under uncertainty are gaining increasing attention of the bilevel optimization community. Also, for these problems, we can distinguish between wait-and-see followers (who observe the leader's decision and the realization of random variables), which is determined by the timing as in (16), and here-and-now followers (who—as customary—observe the leader's decision but need to deal with parameter uncertainty in the second stage of the lower-level problem), which is the timing given in (17).

Early examples of interdiction problems under uncertainty include the stochastic shortest path interdiction introduced by Israeli (1999) or the stochastic maximum-flow interdiction studied by Cormican et al. (1998).

4.2.1. *Stochastic Interdiction Problems (SIPs).* In SIPs, the uncertainty can be in costs or capacities as well as in the effect of interdiction, i.e., an interdicted resource can be partially or completely destroyed only with a certain probability. It is commonly assumed that the underlying probability distribution of the random variables is known to the leader, who is a risk-neutral decision maker and is thus optimizing the expected value of the opposite of the follower's objective function.

For the example given above, the SIP variant is given by

$$\max_{x\in\bar{X}} \mathbb{E}_\omega \left[\min_{y\in Y} \sum_{i\in N_y} (d_i + \delta_i(\omega)x_i)y_i\right], \tag{20}$$

where the actual value of $\delta_i(\omega)$ is revealed to the follower before he makes his decision. If the uncertainty is in the asset cost after the interdiction, then $\delta_i(\omega)$ typically represents the increase of cost in scenario ω. On the other hand, if the uncertainty is in the effect of interdiction, a binary random vector $\tilde{s}_i$ is associated to each asset $i \in N_y$. The value of $\tilde{s}_i$ is equal to one with probability p_i (indicating that the interdiction attempt of this asset is successful) or zero with probability $1 - p_i$ (otherwise). It is typically assumed that separate interdiction attempts are independent and that each asset can be interdicted at most once. In this case, for a given δ_i representing the increase of the asset cost after interdiction, we have $\delta_i(\omega) = \delta_i\tilde{s}_i(\omega)$, where ω represents a possible scenario realization.

Two main approaches to model the uncertainty are adopted in the literature: (i) sample average approximation (SAA) or (ii) sequential approximation (SA). SAA allows to transform the original two-stage (or multi-stage) stochastic problem into its deterministic equivalent while guaranteeing a certain quality of the obtained solution for a sufficiently large number of scenarios; see Kleywegt et al. (2002) for further details. On the other hand, SA starts with a small set of discrete scenarios for which valid lower and upper bounds for the original problem are derived. These scenarios are then iteratively refined by partitioning the uncertainty space until a sufficiently small gap between lower and upper bounds is obtained.

Both SAA and SA allow to transform the original problem into its deterministic interdiction counterpart with a potentially large number of discrete scenarios $\omega \in \Omega$. For problem (20), its deterministic equivalent is given as

$$\max_{x\in\bar{X}} \sum_{\omega\in\Omega} p_\omega \left\{\min_{y^\omega\in Y} \sum_{i\in N_y} (d_i + \delta_i(\omega)x_i)y_i^\omega\right\}, \tag{21}$$

where y^ω refers to lower-level variables representing the optimal response of the follower in scenario ω. The latter bilevel problem can then be reformulated as a single-level MI(N)LP using common techniques for interdiction problems such as dualization or a strong-duality-based reformulation (if the lower level problem is convex for a given choice of x and ω) as well as penalization (otherwise). More details on these reformulation techniques can be found in recent surveys by Kleinert et al. (2021) and Smith and Song (2020). Due to the large number of scenarios involved, the employment of sophisticated decomposition techniques is indispensable in order to develop computationally effective methods. Bailey et al. (2006) consider a generalization of SIPs in which the leader acts as an interdictor, and the follower's decision making process is modeled using a Markov decision process (MDP). As in SIPs, we are given a finite set of discrete scenarios, however, for each (discrete) choice of the leader and for each scenario realization, the follower is solving an MDP. The authors propose a Benders decomposition approach under the assumption that transition probabilities in the MDP are not affected by the decisions of the leader.

In the following, we review some of the most studied applications of stochastic interdiction problems in networks.

Stochastic Shortest Path Interdiction: In this setting, the leader wishes to interdict arcs of a given network by increasing their cost within a limited interdiction budget so that the shortest path between two distinct nodes s and t in the resulting network is maximized. In his PhD thesis, Israeli (1999) assumes that the success of an interdiction attempt is uncertain and, hence, the original deterministic max-min objective function is replaced by maximizing the expected length of the shortest path over all possible interdiction decisions. This corresponds to Model (21) in which binary interdiction variables are associated to arcs of the network, the set $\bar{X}$ models all feasible interdictions under a given knapsack-like budget constraint, and the set Y models all s-t paths in the given network. The author points out that single-level reformulations can be derived following reformulation methods proposed by Israeli and Wood (2002) for the deterministic shortest path interdiction. However, the major difficulty arises from the exponential number of possible scenarios and the author proposes several SA-based methods to deal with them.

Nguyen and Smith (2022) study a variant of this problem in which the base arc cost $d_i \geq 0$ introduced above is not known to the leader, whereas the cost increase $\delta_i \geq 0$ caused by interdiction is certain. The leader assumes that the base cost values are uniformly distributed within given (arc-specific) intervals, whereas the follower acts in a wait-and-see manner. As customary, the leader maximizes the expected shortest path cost attainable by the follower. Nguyen and Smith (2022) develop an SA approach inspired by the work of Cormican et al. (1998) with bounds derived using Jensen's inequality. The authors also provide several algorithmic strategies for accelerating the convergence of their exact approach. Computational results are provided for randomly generated networks with up to 20 nodes.

Held et al. (2005) consider a class of network interdiction problems introduced by Hemmecke et al. (2003) in which the network topology is uncertain and probabilities of each possible network configuration scenario $\omega \in \Omega$ are provided as data: If an arc is not available in a given scenario, its cost is given by $d_i(\omega) = \infty$ and, as customary, $\delta_i(\omega) \geq 0$ denotes the cost increase in case of arc interdiction in scenario ω. In this setting, the leader wishes to maximize the probability of *sufficient disruption*, e.g., to maximize the probability that the shortest path length in the interdicted network is above a given threshold value $1 > \varphi > 0$. The problem is stated as

$$\max_{x \in \bar{X}} \mathbb{P}\left(\min_{y \in Y(\omega)} \sum_{i \in N_y} (d_i(\omega) + \delta_i(\omega) x_i) y(\omega)_i \geq \varphi \right).$$

The authors show that the latter problem can be decomposed by scenarios and solved in a cutting-plane fashion using the method of Riis and Schultz (2003). A computational study is conducted on networks with up to 110 nodes and considering up to 100 scenarios. In a similar setting studied by Song and Shen (2016), dubed *risk-averse* shortest path interdiction, the sufficient disruption is imposed in form of a chance constraint over a discrete set of scenarios. The leader controls interdiction variables x and minimizes the interdiction cost subject to $\sum_{\omega \in \Omega} p_\omega z_\omega \geq 1 - \varepsilon$. Here, the binary variable z_ω is set to one if and only if the follower's shortest path in scenario ω is above the threshold φ, i.e.,

$$\min_{y \in Y(\omega)} \sum_{i \in N_y} (d_i(\omega) + \delta_i(\omega) x_i) y(\omega)_i \geq \varphi z_\omega.$$

The authors propose several families of valid inequalities and develop a branch-and-cut algorithm based on scenario decomposition. In this work, computational results are provided for small grid graphs with up to 64 nodes and for two transportation networks with up to 44 nodes. In both cases, up to 1000 scenarios are considered.

Stochastic Maximum Flow Interdiction: Cormican et al. (1998) investigate the case of minimizing the expected maximum s-t flow in a given network by removing some arcs (or reducing their capacities), assuming that the effect of interdiction is uncertain. More precisely, the interdiction success of each arc is assumed to be an independent binary random variable such that a successful interdiction (which can be performed with some known probability) leaves the arc with no capacity. They also introduce other variants of the problem in which both interdiction success and arc capacities are random or in which multiple interdictions per arc are allowed. The authors apply SA based on classic bounding techniques in stochastic optimization and gradually refine the set of discrete scenarios until obtaining a sufficiently small optimality gap. Janjarassuk and Linderoth (2008) apply a duality-based MILP reformulation, combined with SAA and Benders decomposition, and implement a distributed algorithm. The authors obtain significant speed-ups compared to previous techniques, which is due to a successful combination of decomposition, sampling, parallel computing, and heuristics. They conduct a computational study on grid graphs with up to 400 nodes and consider between 50 and 5000 scenarios.

Lei et al. (2018) consider a variant of the problem with uncertain interdiction effects in which the follower, after observing the leader's interdiction action and before the uncertainty is revealed, can add additional arc capacities to mitigate flow losses. The authors study risk-neutral and risk-averse approaches to model the leader's behavior, i.e., they incorporate the expectation, left-tail, and right-tail CVaR for evaluating maximum flows under uncertainty in the leader's and follower's objectives. The resulting bilevel or trilevel mathematical models are reformulated into single-level MILP formulations and, using an SAA approach, are applied to real-world network instances. Their instances are derived from SNDlib (see `sndlib.zib.de`) and contain 100 scenarios. Finally, Atamtürk et al. (2020) assume that arc capacities are uncertain and that their mean values along with the covariance matrix are known to the leader. The leader removes a subset of arcs from the network while minimizing the VaR of the maximum flow on the resulting network for a given confidence level. The authors propose a heuristic procedure based on successive quadratic optimization embedded in a bisection search. Computational results are reported for a set of grid networks with $q \times q$ nodes (created in a similar way as in Janjarassuk and Linderoth (2008)) with q ranging between 20 and 100.

Maximum Reliability Path Interdiction: In this family of problems, the origin and the destination node chosen by the follower may be uncertain. The leader installs sensors at some arcs of the network and the follower seeks to find a path in the resulting network that maximizes the probability of remaining undetected. The probabilities of being detected with and without sensors installed are known to both players. The problem is, e.g., relevant for preventing nuclear smuggling activities between two countries by placing sensors at their borders (Morton et al. 2007; Pan, Charlton, et al. 2003; Pan and Morton 2008). The leader installs sensors within a limited budget so as to minimize the expected value of the maximum reliability path over all possible source-target choices of the follower. When the probabilities of traversing interdicted arcs are all strictly positive and the follower's source-target choice is known, one can obtain a deterministic shortest-path interdiction problem with a logarithmic transformation; see Morton (2011) and Towle and Luedtke (2018). Morton et al. (2007) and Pan and Morton (2008) propose step inequalities to exploit the relationships between evasion probabilities associated with different paths. Sullivan et al. (2014) use the problem transformation in a bipartite network, provide polyhedral results for a single-scenario, i.e., the deterministic, case, and show how these can be exploited in a multi-scenario setting. The problem can also be reformulated as a two-stage stochastic problem and Bodur et al. (2017) use this reformulation for assessing novel generic ideas to strengthen Benders cuts by exploiting integrality of the first-stage variables. Towle and Luedtke (2018) provide an alternative path-based reformulation and develop a branch-and-cut algorithm that is based on supermodular cuts from Nemhauser et al. (1978) and Ahmed and Atamtürk (2011). The latter approach is shown to be the new state-of-the-art w.r.t. the set of instances considered in the previous literature. These instances are built from a network with 783 nodes as well as 2586 arcs and 456 scenarios are considered. A generalization of this problem with applications in cyber security is studied by Ertem and Bier (2021). Lunday and Sherali (2012) study the problem setting in which the source-target pair is known to the leader, however, there are multiple interdiction resources that can be employed and the impact of their combination can be nonlinear.

Michalopoulos et al. (2015) assume that, in addition to the above uncertainties, the leader is uncertain about the interdiction budget as well. Hence, a three-stage stochastic interdiction approach is proposed in which the leader starts by forming a priority list and assigning an appropriate number of interdiction locations to each priority level. After the budget is revealed, the leader installs sensors at the highest priority locations until the budget is exhausted. Finally, in the third stage, the follower solves the maximum-reliability path problem. A tabu search heuristic is used to solve this challenging problem.

Contrary to the above assumption that the follower's path is deterministic, once the source-target pair and the interdicted arcs are revealed, Collado et al. (2017) assume that the path chosen by the follower remains uncertain to the leader. This setting corresponds to a decision uncertainty caused, e.g., by the uncertainty regarding the follower's criterion for choosing his path through the network or by random influences encountered by the follower while traversing the network. Hence, the follower's path choice is known to the leader only in terms of a probability distribution reflecting her beliefs. The authors propose a risk-neutral and a risk-averse modeling approach while assuming that the leader deploys her resources (modeled using continuous interdiction variables) before and after discovering the source chosen by the follower. The authors employ a mean-upper semideviation risk measure for the risk-aversion approach. The risk-averse problem is approximated and reformulated as a single-level LP model.

4.2.2. *Robust Interdiction Problems.* There is much less literature available for interdiction problems in which there is no assumption on the distribution of the uncertain parameters. Chauhan (2020) studies maximum-flow interdiction with interval uncertainty w.r.t. arc capacities and consumption of resources required to interdict an arc. Following the Γ-robust modeling approach (Bertsimas and Sim 2004), the leader assumes that only a limited number of arcs can be subject to uncertainty, both w.r.t. interdiction budget and arc capacities. It is a wait-and-see setting in which the follower observes arcs removed by the leader and a realization of the uncertain arc capacities for the remaining ones. The leader seeks for an interdiction strategy that protects her from the worst-case outcome in which the capacity of Γ arcs is the largest possible. An MILP formulation is proposed along with three heuristics based on Lagrangian relaxation, Benders decomposition, and a combination of the two.

Nikoofal and Zhuang (2011) consider defender-attacker problems with applications in counter-terrorism. The leader and the follower have different perceptions regarding the valuation of the damage caused by attacking given targets. The leader searches to minimize the damage caused by the follower, however, the follower's valuation of targets is unknown and it is assumed to belong to bounded intervals. Using the Γ-robust approach, the authors assume that the total scaled deviation of the uncertainty parameters cannot exceed a given threshold Γ and the corresponding robustness constraint is added at the upper level. In a follow-up article, Nikoofal and Zhuang (2015) study the trade-off between disclosure of the defense strategy by the leader (which corresponds to a Stackelberg game) versus secrecy (which corresponds to a simultaneous game), assuming in both cases that the leader only knows intervals to which the attacker's valuation of targets belongs. Their results show that the leader's benefit by making the first move, i.e., by playing the Stackelberg game, is only considerable if the follower and the leader share a similar valuation of the targets. Thus, the optimal defense allocation in a simultaneous game provides a better protection against uncertainty in the follower's valuation of targets.

Gillen et al. (2021) study the spread of cascading behavior in a social network using a linear threshold (LT) model with a given set of activated nodes. In the cascading LT model, new nodes are getting activated if the sum of arc weights of their already activated neighbors is above a given threshold value. In the defender-attacker problem studied by Gillen et al. (2021), the leader tries to fortify some nodes by increasing their influence threshold within a limited budget in order to reduce the total number of activated nodes at the end of the propagation process. The follower's problem models the propagation process and the authors assume that the arc weights (modeling the influence of a node to its neighbors) are subject to interval uncertainty. The authors use a Γ-robust approach to deal with uncertainty and develop an iterative procedure in which the problem is solved on a subgraph using a cutting-plane procedure. The subgraph is then expanded and the procedure is repeated until the convergence criteria are met. A computational study is conducted on a selection of real-world networks from the SNAP library (see `https://snap.stanford.edu`) with some of them containing more than 200 000 nodes.

Beck, Ljubić, et al. (2022) propose a generic branch-and-cut framework for solving min-max mixed-integer optimization problems with a Γ-robust uncertainty modeling in the lower level. The follower takes decisions after observing the action of the leader and before facing the uncertainty, i.e., the timing (17) is assumed. The follower aims to hedge against a subset of deviations of uncertain parameters and the lower-level problem contains discrete variables. Two cases of interval uncertainty are studied: in the coefficients of the lower-level's objective function and in the coefficients of a single packing-type constraint. Two generic reformulations as a single-level MILP are proposed and problem-tailored cuts have been derived for the Γ-robust variants of the knapsack interdiction problem. These cuts assume that Γ-robust follower sub-problems satisfy a downward monotonicity property, which arises in many packing-type applications. In this context, these cuts are a generalization of what has been proposed in Fischetti, Ljubić, Monaci, et al. (2019) for the deterministic knapsack interdiction problem.

4.2.3. *Other Interdiction Problems with Incomplete or Asymmetric Information.* Bayrak and Bailey (2008) study a shortest path network interdiction problem with asymmetric information in which the follower has inaccurate or incomplete information while the leader has complete knowledge of the network. Hence, the perceived arc costs and their increase caused by interdiction are different for the two players. Therefore, the objective functions of the leader and the follower do not coincide, and the problem assumes a structure of a more general bilevel optimization problem with a convex lower-level model. The authors propose a reformulation as a single-level MILP and demonstrate in their computational study that asymmetric information allows to obtain improved interdiction policies compared to those using symmetric information.

Pay et al. (2018) consider a stochastic shortest path interdiction problem with incomplete risk preferences of the leader. As customary, there is uncertainty in the arc costs and in the interdiction effect on each arc and, after the interdiction, the follower reacts in a wait-and-see fashion. A finite set of scenarios is used to model uncertainty realizations. Contrary to previous studies, the authors assume that the leader is risk-averse and that there exists a utility function that summarizes her risk preferences but that her knowledge about this function is incomplete. To this end, the authors propose two ways to deal with this ambiguity: (i) to use historical data and pairwise comparison of lotteries to fit a piecewise concave utility function and run the stochastic interdiction model afterward, or (ii) to integrate utility estimation within the optimization model. The latter leads to a robust approach (originally proposed in Armbruster and Delage (2015) and Hu and Mehrotra (2015)) in which the leader searches for an interdiction strategy that maximizes her utility in a robust fashion, i.e., by considering an infimum over the function space of all possible utility functions $u \in U$ consistent with leader's preferences:

$$\max_{x \in X} \inf_{u \in U} \mathbb{E}\left[u\left(\min_{y \in Y} \sum_{\omega \in \Omega} p_\omega \sum_{i \in N_y} (d_i + \delta_i(\omega) x_i) y_i\right)\right].$$

A single-level MILP reformulation for this problem is proposed and solved using a branch-and-cut procedure. Computational results are reported for small grid networks with up to 49 nodes and up to 1000 scenarios.

Sequential shortest path interdiction games with asymmetric information have been studied in a series of papers by Borrero et al. (2019, 2016) and Yang, Borrero, et al. (2021). In the sequential decision making setting proposed by Borrero et al. (2016), there is a repeated interaction between the leader and the follower. At each period, the leader attacks the follower by interdicting a subset of assets with the goal of maximizing the cumulative follower's costs over the given time period. The leader has incomplete knowledge concerning the structure and arc costs in the network, with only lower and upper bounds on the arc costs being available. The leader learns about the network and arc costs trough sequential interdiction actions (thanks to the optimal responses of the follower) and dynamically adapts her interdiction strategy. The leader observes the chosen path along with its cost in order to possibly update her strategy in the following iteration. The authors introduce the concept of "policy time-stability", representing the number of learning iterations needed for the leader to reach the same interdiction strategy as if she would have complete information. Two policies (a greedy and a pessimistic one) are proposed and studied from the theoretical and computational perspective. A computational study is carried out on randomly generated uniform graphs with 40 and 50 nodes. The results have been later generalized in Borrero et al. (2019), where general interdiction problems are considered such that the leader does not know all of the follower's resources and constraints and the follower's cost coefficients are assumed to belong to a polyhedral uncertainty set. Three types of follower's feedback are studied: standard (the leader observes the value of the follower's objective only), value-perfect (cost coefficients of the follower's activity are revealed, too), and response-perfect (the full decision vector of the follower is revealed to the leader). For their computational study, the authors use knapsack interdiction instances with up to 15 items.

In a follow-up article, Yang, Borrero, et al. (2021) also study the sequential shortest path interdiction problem with incomplete information. Contrary to Borrero et al. (2016), where the feedback includes the shortest path chosen by the follower, in Yang, Borrero, et al. (2021), this feedback is limited to the length of the chosen path. In terms of policy time stability, the authors show that, in the worst case, the number of iterations of the proposed greedy interdiction policies is exponential and that convergence in polynomial time is possible if more information is provided through the feedback. Computational experiments are conducted on layered graph networks with between 3 and 10 layers as well as 7 nodes in each layer. Finally, more general two-player sequential games are studied in Borrero et al. (2022), see also Section 3.2.

4.3. **Management Science.** In this last part of this section, we review applications of uncertain bilevel optimization in the fields of networks, supply chains, facility location, and finance.

Networks: Toll pricing in networks is a bilevel optimization problem in which the leader sets the tolls for road segments of a transportation network so that the revenue raised from tariffs is maximized. The leader anticipates that user flows are assigned to cheapest paths in the resulting network; see Brotcorne et al. (2000), Labbé, Marcotte, et al. (1998), and Labbé and Violin (2016). Alizadeh et al. (2013) consider a two-stage stochastic toll pricing problem in which the leader faces uncertainties regarding travel demand and travel costs. These uncertainties are modeled through a discrete set of scenarios. The first-stage decision of the leader is to set the tariffs while maximizing the expected revenues. The tariffs set in the first stage can be modified within a pre-defined interval in the second stage once the uncertainties are revealed. The authors show how to reduce the problem to its deterministic bilevel equivalent and conduct a sensitivity analysis w.r.t. the constraints linking the tariffs at the first and the second stage. Gilbert et al. (2015) consider another stochastic variant in which the users choose their paths according to a discrete choice model. The followers namely choose a path that minimizes their disutility, which, besides the arc costs and tolls, contains an additional additive component unknown to the leader. This unobserved term is assumed to have a logistic distribution. The authors provide two heuristics derived from approximations of the logit revenue function. In her PhD thesis, Violin (2014) studies the toll pricing problem with interval uncertainty considering travel demand or the cost of toll-free paths. She applies Γ-robust models to some cases whose deterministic counterparts can be solved in polynomial time, e.g., pricing on a single arc, the single commodity case, or the application of a unit toll. Dokka et al. (2016), see also Dokka et al. (2017), study the toll pricing problem under uncertainty on non-toll costs. For the leader, the distribution of non-toll costs is unknown but, based on historical information, she assumes that the distribution is fixed and belongs to a set of non-negative distributions. The followers observe the toll rates and use the full knowledge of non-toll data (in a wait-and-see fashion) to choose the shortest paths in the resulting network. The authors consider a single follower case and model the uncertainty using the concept of distributional robustness; see, e.g., Goh and Sim (2010). For this highly complex problem, they provide mathematical formulations and heuristics for networks with multiple parallel source-target arcs and networks with a polynomial number of source-target paths.

In the literature, robust optimization is frequently applied to the hazmat network design problem. Given an existing road network, the deterministic hazmat network design problem asks for selecting the road segments that should be closed for hazmat transport so as to minimize the total risk. Here, each commodity has its own arc risk, which is taken into consideration if the arc is used in the resulting transportation network. Hence, the problem can be seen as a bilevel problem in which the leader selects the road segments to be closed while the followers, i.e., the hazmat carriers, solve shortest path problems with different source-target pairs. Moreover, it is assumed that there are no congestion effects in the resulting network. Longsheng et al. (2017) propose a robust approach to model *generalized bounded rationality* in route choice behavior modeling. They test their concept on the robust hazmat network design problem in which robust optimization models the uncertainty in the cost of shortest paths of the lower-level problems. The uncertainty models the bounded rationality of lower-level decision makers caused by their perception error. Among others, polyhedral and ellipsoidal uncertainty sets are considered and an exact method based on cutting planes is proposed. A similar problem is studied by

Xin et al. (2015), however, it is assumed that the arc risks are subject to interval uncertainty instead, whereas the arc lengths are deterministic and known to both players. The leader searches for a subset of road segments to block so that the maximum regret w.r.t. path risks over all commodities is minimized. The authors propose a heuristic approach and test it using a case study of the road network of the Guangdong province in China. Arguing that the minimax regret approach of Xin et al. (2015) is too conservative, Sun et al. (2015) use Γ-robustness to deal with arc risk uncertainties. Two cases are considered in which Γ corresponds to (i) the overall number of arcs that are subject to uncertainty or (ii) the number of arcs which are subject to uncertainty for all shipments. The authors provide single-level reformulations and a Lagrangian heuristic. In related works, Berglund and Kwon (2014) and Liu and Kwon (2020) consider a combined facility location and hazmat network design problem. They assume interval uncertainty w.r.t. transportation demands and arc risks and adopt the Γ-robust approach. The leader minimizes a linear combination of fixed facility opening cost and the worst-case arc risk exposure. The followers choose the routes that minimize the transportation cost to the nearest hazmat facility. A single-level MILP formulation and a genetic algorithm are proposed in Berglund and Kwon (2014). Liu and Kwon (2020) develop an exact method based on cutting planes combined with Benders decomposition. In this latest computational study, the authors consider a road network of the city of Ravenna with 110 nodes and 143 arcs.

Attack graphs, represented by trees or directed acyclic graphs, are frequently used to model vulnerabilities of a system. In such a network, nodes represent attack states and arcs correspond to the transition of states fulfilled by attack activities; see, e.g., Bhuiyan (2018) and further references therein. Zheng and Albert (2019) use an attack graph with completion times on its arcs to model applications for infrastructure protection planning, e.g., to mitigate cyber-security or supply chain attacks. The leader, who acts as the defender, implements policies and invests in cyber-infrastructure security, whereas multiple adversaries try to exploit vulnerabilities of the system to carry out attacks as soon as possible. An attack corresponds to a "project" whose fastest completion time is associated with a critical path in the attack graph. In their setting, arc delay times (imposed by interdiction activities) are uncertain, the followers act in a wait-and-see manner, and the leader deploys limited resources to impose delays on the arcs so that the total weighted expected completion time of all adversarial attacks is maximized. After considering a finite set of discrete scenarios and dualizing the lower-level problem, the resulting max-max problem is decomposed using a Lagrangian heuristic approach based on subgradient optimization. Another more general Stackelberg game on attack graphs is studied by Letchford and Vorobeychik (2013) and a risk-averse approach based on a CVaR model is studied by Bhuiyan (2018).

In telecommunication industry, bilevel optimization can be used to model hierarchical decision making between a network operator (the leader) and virtual operators (the followers). Audestad et al. (2006) study the problem in which the network operator solves the pricing problem so as to maximize her profit or the market share. The leader decides on the capacity leased to the virtual operators and sets the prices for these capacities as well as for the service to the customers. The followers maximize their own profit function after observing the decisions of the leader. The authors describe a two-stage stochastic bilevel optimization model in which the uncertainty of customer demands is modeled with a discrete set of scenarios. Once the customer demand is revealed, the leader has a possibility to extend the service and to change the pricing decisions. Additional results, including a Lagrangian optimization method for finding locally optimal solutions, are provided in the PhD thesis of Werner (2004). DeMiguel and Xu (2009) consider stochastic multi-leader-follower games with demand uncertainty and use it to model competition in the telecommunication industry. The leaders only know the distribution of the demand, whereas the followers have information on the exact realization of uncertainty. The leaders compete in a Cournot setting and each leader searches for a Stackelberg solution w.r.t. her followers. The followers compete in a Cournot fashion with all leaders and the other followers. The authors show the existence and uniqueness of an equilibrium for the considered stochastic model and propose a solution approach based on sample average approximation. We refer to Hu and Fukushima (2015) for further studies on multi-leader-follower Stackelberg games.

Supply Chains and Facility Location: Multi-period facility location interdiction with stochastic resource constraints is studied by Zhang and Özaltın (2021). The problem is modeled as a stochastic bilevel problem with integer variables at both levels and a branch-and-bound procedure is proposed, see Section 3.1.6 for further details. Ryu et al. (2004) consider supply chain optimization in which the upper-level problem is a plant planning problem and the lower-level problem is a distribution network problem with stochastic demands. A similar problem but in a multi-objective setting is investigated by Roghanian et al. (2007). Yeh et al. (2015) use bilevel optimization to model timber supply chains in which the harvesters decide first on the quantity to be harvested and the manufacturers decide later on how much to utilize. The authors study a problem of investing in biofuel production under uncertainty and model it as a two-stage stochastic problem in which, for each realization of uncertain parameters, the second stage is the aforementioned bilevel problem. In Su and Geunes (2013), a Stackelberg game is considered between a single-supplier and a multi-retailer supply chain under asymmetric demand information. The supplier sets price discounts first (in a multi-period setting) under uncertainty concerning the retail-store demand. After the actual demand of each store is revealed, individual store order quantities are determined in each period. The goal of the supplier is to determine the pricing strategy that maximizes her expected payoff while anticipating the store order quantities.

Finance: Yan et al. (2014) consider a stochastic approach to supply chain financing under demand uncertainty. Fanzeres, Ahmed, et al. (2019) investigate a revenue-maximizing strategic bidding problem with uncertainty concerning the competitors' bidding strategy. A two-stage robust optimization model with equilibrium constraints is proposed and reformulated as a bilevel problem with equilibrium constraints. After deriving a single-level reformulation, the authors implement a column-and-constraint generation approach.

5. Possible Directions for Future Research

Although the study of bilevel optimization problems under uncertainty started rather recently in the 1990s, there already has been substantial work in this field; see, e.g., Figure 6, which shows the number of papers per year cited in this survey.

Nevertheless, there are very many topics still open for future research. In this last section of the survey, we thus sketch a few of these potential future research topics.

(i) For the stochastic approach to address data uncertainty, only a few works exist (see, e.g., Carrión et al. (2009) and Kovacevic and Pflug (2014)) that go beyond the risk-neutral case for nonlinear bilevel models. This leaves a wide open space for future research combining risk-averse modeling (such as using the CVaR as in the papers mentioned above) for nonlinear models. Except for some interdiction problems, nonlinear bilevel models

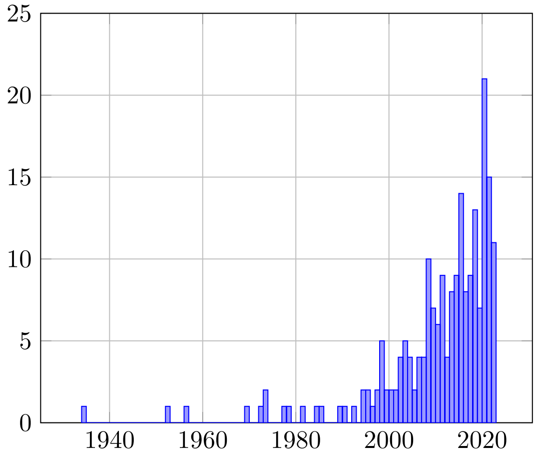

FIGURE 6. Papers per year as cited in this survey

under uncertainty are only rarely discussed in the literature and most of the research on stochastic bilevel optimization so far focuses on the linear case. The reason, most likely, is that even nonlinear bilevel optimization is an extremely challenging field and the combination of such nonlinearities with further uncertainty considerations makes such problems even more hard to study and solve.

(ii) For stochastic setups, there have been some works that algorithmically exploit (quasi-)block structures. However, apart from Held et al. (2005) and Song and Shen (2016) who study special interdiction problems, most of these works consider the risk-neutral case. Hence, the field of structure-exploiting algorithms for risk-averse bilevel models contains many possible directions for future research as well.

(iii) It seems that there is only little research (see, e.g., Adasme et al. (2013) and Dokka et al. (2016)) that combines distributional robustness (Goh and Sim 2010) and bilevel optimization although both fields, standalone, are very active fields of research today. However, since both fields on their own are already very difficult, their combination will be even more challenging.

(iv) The literature on stochastic bilevel optimization is rather theoretical. There is (again, to the best of our knowledge) no general computational paper in this field besides those that consider specific applications and that almost all use the deterministic equivalent as the main algorithmic workhorse.

(v) Robust approaches to account for uncertainties in the context of bilevel optimization are still in their infancy. So far, most works focus on the strictly or Γ-robust case. Hence, the consideration of other robustness concepts such as light robustness (Fischetti and Monaci 2009) or adjustable robustness (Ben-Tal, Goryashko, et al. 2004) may be reasonable directions for future research.

(vi) In robust setups, uncertainties are predominantly modeled using either interval or polyhedral uncertainty sets. This leaves room to study other uncertainty set geometries and, in particular, models with discrete uncertainty sets.

(vii) We are not aware of any work that considers intermediate solution concepts between the optimistic and the pessimistic approach for bilevel problems with coupling constraints. Moreover, and to the best of our knowledge, solution methods for strong-weak bilevel problems have only been considered for the linear case.

(viii) Almost all papers on uncertain bilevel optimization consider the timing in which the uncertainty realizes after the decision of the leader has been taken and before the follower decides. Although some first works (Beck, Ljubić, et al. 2022; Heitsch et al. 2022) have been published on the alternative timing given in (17), this setup still is a rather open field—in particular in the case of chance constraints as part of the lower-level problem.

(ix) There are only a few papers (Salas and Svensson 2020; Zhang, Liu, et al. 2022) that started to pave the way for the study of decision-dependent uncertainties. This is, however, another completely open field of research.

(x) Finally, there are no well-curated collections of instances in the field of bilevel optimization under uncertainty, which would, of course, help the community a lot when it comes to developing and testing novel algorithmic ideas.

Acknowledgments

The third author thanks the DFG for their support within projects A05 and B08 in CRC TRR 154. Moreover, we are very grateful for the many comments and the fruitful discussions on the topics of this survey during the Dagstuhl seminar "Optimization at the Second Level (22441)".

References

Aboussoror, A., S. Adly, and F. E. Saissi (2017). "Strong-Weak Nonlinear Bilevel Problems: Existence of Solutions in a Sequential Setting." In: *Set-Valued and Variational Analysis* 25, pp. 113–132. DOI: `10.1007/s11228-016-0369-4`.

Aboussoror, A. and P. Loridan (1995). "Strong-weak Stackelberg Problems in Finite Dimensional Spaces." In: *Serdica. Mathematical Journal* 21, pp. 151–170. URL: `https://www.semanticscholar.org/paper/Strong-weak-Stackelberg-Problems-in-Finite-Spaces-Aboussoror-Loridan/c875814905ed516f7e693d9040683c1115f31c27`.

Aboussoror, A. and A. Mansouri (2005). "Weak linear bilevel programming problems: existence of solutions via a penalty method." In: *Journal of Mathematical Analysis and Applications* 304.1, pp. 399–408. DOI: `10.1016/j.jmaa.2004.09.033`.

Adasme, P., A. Lisser, and C. Wang (2013). "A Distributionally Robust Formulation for Stochastic Quadratic Bi-level Programming." In: *ICORES 2013 - Proceedings of the 2nd International Conference on Operations Research and Enterprise Systems, Barcelona, Spain, 16-18 February, 2013*. Ed. by B. Vitoriano and F. Valente. SciTePress, pp. 24–31. DOI: `10.5220/0004207100240031`.

Ahmed, S. and A. Atamtürk (2011). "Maximizing a class of submodular utility functions." In: *Mathematical programming* 128.1-2, pp. 149–169. DOI: `10.1007/s10107-009-0298-1`.

Alizadeh, S., P. Marcotte, and G. Savard (2013). "Two-stage stochastic bilevel programming over a transportation network." In: *Transportation Research Part B: Methodological* 58, pp. 92–105. DOI: `10.1016/j.trb.2013.10.002`.

Ambrosius, M., J. Egerer, V. Grimm, and A. van der Weijde (2020). "Uncertain bidding zone configurations: The role of expectations for transmission and generation capacity expansion." In: *European Journal of Operational Research* 285.1, pp. 343–359. DOI: `10.1016/j.ejor.2020.01.024`.

Ambrosius, M., J. Egerer, V. Grimm, and A. H. van der Weijde (2022). "Risk aversion in multilevel electricity market models with different congestion pricing regimes." In: *Energy Economics* 105, p. 105701. DOI: `10.1016/j.eneco.2021.105701`.

Armbruster, B. and E. Delage (2015). “Decision Making Under Uncertainty When Preference Information Is Incomplete.” In: *Management Science* 61.1, pp. 111–128. DOI: `10.1287/mnsc.2014.2059`.

Askeland, M., T. Burandt, and S. A. Gabriel (2020). “A stochastic MPEC approach for grid tariff design with demand-side flexibility.” In: *Energy Systems*. DOI: `10.1007/s12667-020-00407-7`.

Atamtürk, A., C. Deck, and H. Jeon (2020). “Successive Quadratic Upper-Bounding for Discrete Mean-Risk Minimization and Network Interdiction.” In: *INFORMS Journal of Computing* 32.2, pp. 346–355. DOI: `10.1287/ijoc.2018.0870`.

Audestad, J.-A., A. A. Gaivoronski, and A. Werner (2006). “Extending the stochastic programming frameworkfor the modeling of several decision makers: pricing and competition in the telecommunication sector.” In: *Annals of Operations Research* 142, pp. 19–39. DOI: `10.1007/s10479-006-6159-x`.

Aussel, D. and A. Svensson (2019). “Is Pessimistic Bilevel Programming a Special Case of a Mathematical Program with Complementarity Constraints?” In: *Journal of Optimization Theory and Applications* 181, pp. 504–520. DOI: `10.1007/s10957-018-01467-7`.

Bagwell, K. (1995). “Commitment and observability in games.” In: *Games and Economic Behavior* 8.2, pp. 271–280. DOI: `10.1016/S0899-8256(05)80001-6`.

Bailey, M. D., S. M. Shechter, and A. J. Schaefer (2006). “SPAR: stochastic programming with adversarial recourse.” In: *Operations Research Letters* 34.3, pp. 307–315. DOI: `https://doi.org/10.1016/j.orl.2005.05.002`.

Bayrak, H. and M. D. Bailey (2008). “Shortest path network interdiction with asymmetric information.” In: *Networks* 52.3, pp. 133–140. DOI: `10.1002/net.20236`.

Beck, Y., I. Ljubić, and M. Schmidt (2022). *Exact Methods for Discrete* Γ*-Robust Interdiction Problems.* Tech. rep. URL: `http://www.optimization-online.org/DB_FILE/2021/11/8678.pdf`.

Beck, Y. and M. Schmidt (2021). “A robust approach for modeling limited observability in bilevel optimization.” In: *Operations Research Letters* 49.5, pp. 752–758. DOI: `10.1016/j.orl.2021.07.010`.

Beck, Y., M. Schmidt, J. Thürauf, and D. Bienstock (2022). *On a Computationally Ill-Behaved Bilevel Problem with a Continuous and Nonconvex Lower Level.* Tech. rep. URL: `https://arxiv.org/abs/2202.01033`.

Ben-Tal, A., L. El Ghaoui, and A. Nemirovski (2009). *Robust optimization.* Princeton University Press.

Ben-Tal, A., A. Goryashko, E. Guslitzer, and A. Nemirovski (2004). “Adjustable robust solutions of uncertain linear programs.” In: *Mathematical Programming* 99.2, pp. 351–376. DOI: `10.1007/s10107-003-0454-y`.

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`.

Berglund, P. G. and C. Kwon (2014). “Robust facility location problem for hazardous waste transportation.” In: *Networks and Spatial Economics* 14.1, pp. 91–116. DOI: `10.1007/s11067-013-9208-4`.

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`.

Besançon, M., M. F. Anjos, and L. Brotcorne (2021). “Complexity of near-optimal robust versions of multilevel optimization problems.” In: *Optimization Letters* 15, pp. 2597–2610. DOI: 10.1007/s11590-021-01754-9.

Besançon, M., M. F. Anjos, and L. Brotcorne (2019). *Near-optimal Robust Bilevel Optimization*. URL: https://arxiv.org/pdf/1908.04040.pdf.

Bhuiyan, T. H. (2018). “Risk-averse bi-level stochastic network interdiction model for cyber-security risk management.” Mississippi State University.

Birge, J. R. and F. Louveaux (2011). *Introduction to stochastic programming*. Springer Science & Business Media. DOI: 10.1007/978-1-4614-0237-4.

Bodur, M., S. Dash, O. Günlük, and J. Luedtke (2017). “Strengthened Benders Cuts for Stochastic Integer Programs with Continuous Recourse.” In: *INFORMS Journal on Computing* 29.1, pp. 77–91. DOI: 10.1287/ijoc.2016.0717.

Bolusani, S., S. Coniglio, T. K. Ralphs, and S. Tahernejad (2020). “A Unified Framework for Multistage Mixed Integer Linear Optimization.” In: *Bilevel Optimization: Advances and Next Challenges*. Ed. by S. Dempe and A. Zemkoho. Cham: Springer International Publishing, pp. 513–560. DOI: 10.1007/978-3-030-52119-6_18.

Borrero, J. S., O. Prokopyev, and D. Sauré (2022). “Learning in Sequential Bilevel Linear Programming.” In: *INFORMS Journal on Computing*. Online first. DOI: 10.1287/ijoo.2021.0063.

Borrero, J. S., O. A. Prokopyev, and D. Sauré (2019). “Sequential Interdiction with Incomplete Information and Learning.” In: *Operations Research* 67.1, pp. 72–89. DOI: 10.1287/opre.2018.1773.

Borrero, J. S., O. A. Prokopyev, and D. Sauré (2016). “Sequential Shortest Path Interdiction with Incomplete Information.” In: *Decision Analysis* 13.1, pp. 68–98. DOI: 10.1287/deca.2015.0325.

Böttger, T., V. Grimm, T. Kleinert, and M. Schmidt (2021). “The Cost of Decoupling Trade and Transport in the European Entry-Exit Gas Market with Linear Physics Modeling.” In: *European Journal on Operational Research*. Online first. DOI: 10.1016/j.ejor.2021.06.034.

Bracken, J. and J. T. McGill (1973). “Mathematical programs with optimization problems in the constraints.” In: *Operations Research* 21.1, pp. 37–44. DOI: 10.1287/opre.21.1.37.

Brotcorne, L., M. Labbé, P. Marcotte, and G. Savard (2000). “A bilevel model and solution algorithm for a freight tariff-setting problem.” In: *Transportation Science* 34.3, pp. 289–302. DOI: 10.1287/trsc.34.3.289.12299.

Buchheim, C. and D. Henke (2020). *The bilevel continuous knapsack problem with uncertain follower’s objective*. URL: https://arxiv.org/abs/1903.02810.

Buchheim, C. and D. Henke (2022). “The robust bilevel continuous knapsack problem with uncertain coefficients in the follower’s objective.” In: *Journal of Global Optimization* 83, pp. 803–824. DOI: 10.1007/s10898-021-01117-9.

Buchheim, C., D. Henke, and F. Hommelsheim (2021). “On the Complexity of Robust Bilevel Optimization With Uncertain Follower’s Objective.” In: *Operations Research Letters* 49.5, pp. 703–707. DOI: 10.1016/j.orl.2021.07.009.

Buchheim, C., D. Henke, and J. Irmai (2022). “The Stochastic Bilevel Continuous Knapsack Problem with Uncertain Follower’s Objective.” In: *Journal of Optimization Theory and Applications* 194.2, pp. 521–542. DOI: 10.1007/s10957-022-02037-8.

Burtscheidt, J. and M. Claus (2020). “Bilevel Linear Optimization Under Uncertainty.” In: *Bilevel Optimization: Advances and Next Challenges*. Ed. by S. Dempe and A. Zemkoho, pp. 485–511. DOI: 10.1007/978-3-030-52119-6_17.

Burtscheidt, J., M. Claus, S. Conti, M. Rumpf, J. Sassen, and R. Schultz (2021). *A Pessimistic Bilevel Stochastic Problem for Elastic Shape Optimization.* URL: https://arxiv.org/abs/2103.02281.

Burtscheidt, J., M. Claus, and S. Dempe (2020). "Risk-Averse Models in Bilevel Stochastic Linear Programming." In: *SIAM Journal on Optimization* 30.1, pp. 377–406. DOI: 10.1137/19M1242240.

Candler, W. and R. Norton (1977). *Multi-level Programming.* Discussion Papers, Development Research Center, International Bank for Reconstruction and Development. World Bank. URL: https://books.google.de/books?id=TicmAQAAMAAJ.

Cao, D. and L. C. Leung (2002). "A partial cooperation model for non-unique linear two-level decision problems." In: *European Journal of Operational Research* 140.1, pp. 134–141. DOI: 10.1016/S0377-2217(01)00225-9.

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.

Carrión, M., J. M. Arroyo, and A. J. Conejo (2009). "A Bilevel Stochastic Programming Approachfor Retailer Futures Market Trading." In: *IEEE Transactions on Power Systems* 24.3, pp. 1446–1456. DOI: 10.1109/TPWRS.2009.2019777.

Chauhan, D. R. (2020). "Robust Maximum Flow Network Interdiction Problem." PhD thesis. Portland State University. DOI: 10.15760/etd.7315.

Chen, A., J. Kim, Z. Zhou, and P. Chootinan (2007). "Alpha Reliable Network Design Problem." In: *Transportation Research Record* 2029.1, pp. 49–57. DOI: 10.3141/2029-06.

Christiansen, S., M. Patriksson, and L. Wynter (2001). "Stochastic bilevel programming in structural optimization." In: *Structural and Multidisciplinary Optimization* 21, pp. 361–371. DOI: 10.1007/s001580100115.

Chuong, T. D. and V. Jeyakumar (2017). "Finding Robust Global Optimal Values of Bilevel Polynomial Programs with Uncertain Linear Constraints." In: *Journal of Optimization Theory and Applications* 173.2, pp. 683–703. DOI: 10.1007/s10957-017-1069-4.

Claus, M. (2021a). "A Second-Order Sufficient Optimality Condition for Risk-Neutral Bi-level Stochastic Linear Programs." In: *Journal of Optimization Theory and Applications* 188, pp. 243–259. DOI: 10.1007/s10957-020-01775-x.

Claus, M. (2021b). "Existence of Solutions for a Class of Bilevel Stochastic Linear Programs." In: *European Journal of Operational Research.* DOI: 10.1016/j.ejor.2021.12.004.

Claus, M. (2021c). "On continuity in risk-averse bilevel stochastic linear programming with random lower level objective function." In: *Operations Research Letters* 49.3, pp. 412–417. DOI: 10.1016/j.orl.2021.04.007.

Collado, R., S. Meisel, and L. Priekule (2017). "Risk-averse stochastic path detection." In: *European Journal of Operational Research* 260.1, pp. 195–211. DOI: 10.1016/j.ejor.2016.12.002.

Conitzer, V. and T. Sandholm (2006). "Computing the Optimal Strategy to Commit To." In: *Proceedings of the 7th ACM Conference on Electronic Commerce.* New York, NY, USA: Association for Computing Machinery, pp. 82–90. DOI: 10.1145/1134707.1134717.

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.

DeMiguel, V. and H. Xu (2009). "A Stochastic Multiple-Leader Stackelberg Model: Analysis, Computation, and Application." In: *Operations Research* 57.5, pp. 1220–1235. DOI: 10.1287/opre.1080.0686.

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

Dempe, S. and K. Richter (2000). “Bilevel programming with knapsack constraints.” In: *Central European Journal of Operations Research* 8.2, pp. 93–107.

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.

Dempe, S., S. Ivanov, and A. Naumov (2017). “Reduction of the bilevel stochastic optimization problem with quantile objective function to a mixed-integer problem.” In: *Applied Stochastic Models in Business and Industry* 33.5, pp. 544–554. DOI: 10.1002/asmb.2254.

Dempe, S., V. Kalashnikov, G. A. Pérez-Valdés, and N. Kalashnykova (2015). *Bilevel Programming Problems*. Springer. DOI: 10.1007/978-3-662-45827-3.

Dokka, T., A. B. Zemkoho, S. S. Gupta, and F. T. Nobibon (2017). *Robust toll pricing: A novel approach*. DOI: 10.48550/ARXIV.1712.01570.

Dokka, T., A. B. Zemkoho, S. S. Gupta, and F. T. Nobibon (2016). “Pricing Toll Roads under Uncertainty.” In: *16th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems, ATMOS 2016, August 25, 2016, Aarhus, Denmark*. Ed. by M. Goerigk and R. F. Werneck. Vol. 54. OASIcs. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 4:1–4:14. DOI: 10.4230/OASIcs.ATMOS.2016.4.

Ertem, M. and V. M. Bier (2021). “A Stochastic Network-Interdiction Model For Cyber Security.” In: *2021 5th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, pp. 171–176. DOI: 10.1109/ISMSIT52890.2021.9604681.

Fampa, M., L. A. Barroso, D. Candal, and L. Simonetti (2008). “Bilevel optimization applied to strategic pricing in competitive electricity markets.” In: *Computational Optimization and Applications* 39, pp. 121–142. DOI: 10.1007/s10589-007-9066-4.

Fanzeres, B., S. Ahmed, and A. Street (2019). “Robust strategic bidding in auction-based markets.” In: *European Journal of Operational Research* 272.3, pp. 1158–1172. DOI: 10.1016/j.ejor.2018.07.027.

Fanzeres, B., A. Street, and L. A. Barroso (2015). “Contracting Strategies for Renewable Generators: A Hybrid Stochastic and Robust Optimization Approach.” In: *IEEE Transactions on Power Systems* 30.4, pp. 1825–1837. DOI: 10.1109/TPWRS.2014.2346988.

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., I. Ljubić, and M. Sinnl (2016). “Benders decomposition without separability: A computational study for capacitated facility location problems.” In: *European Journal of Operational Research* 253.3, pp. 557–569. DOI: 10.1016/j.ejor.2016.03.002.

Fischetti, M., I. Ljubić, and M. Sinnl (2017). “Redesigning Benders Decomposition for Large-Scale Facility Location.” In: *Management Science* 63.7, pp. 2146–2162. DOI: 10.1287/mnsc.2016.2461.

Fischetti, M. and M. Monaci (2009). “Light Robustness.” In: *Robust and Online Large-Scale Optimization: Models and Techniques for Transportation Systems*.

Ed. by R. K. Ahuja, R. H. Möhring, and C. D. Zaroliagis. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 61–84. DOI: 10.1007/978-3-642-05465-5_3.

Fortuny-Amat, J. and B. McCarl (1981). "A Representation and Economic Interpretation of a Two-Level Programming Problem." In: *The Journal of the Operational Research Society* 32.9, pp. 783–792. DOI: 10.1057/jors.1981.156.

Frank, M. and P. Wolfe (1956). "An algorithm for quadratic programming." In: *Naval Research Logistics Quarterly* 3.1-2, pp. 95–110. DOI: 10.1002/nav.3800030109.

Gilbert, F., P. Marcotte, and G. Savard (2015). "A Numerical Study of the Logit Network Pricing Problem." In: *Transportation Science* 49.3, pp. 706–719. DOI: 10.1287/trsc.2014.0560.

Gillen, C. P., A. Veremyev, O. A. Prokopyev, and E. L. Pasiliao (2021). "Fortification Against Cascade Propagation Under Uncertainty." In: *INFORMS Journal on Computing* 33.4, pp. 1481–1499. DOI: 10.1287/ijoc.2020.0992.

Goh, J. and M. Sim (2010). "Distributionally Robust Optimization and Its Tractable Approximations." In: *Operations Research* 58.4-part-1, pp. 902–917. DOI: 10.1287/opre.1090.0795.

Grimm, V., L. Schewe, M. Schmidt, and G. Zöttl (2019). "A Multilevel Model of the European Entry-Exit Gas Market." In: *Mathematical Methods of Operations Research* 89.2, pp. 223–255. DOI: 10.1007/s00186-018-0647-z.

Growe-Kuska, N., H. Heitsch, and W. Römisch (2003). "Scenario reduction and scenario tree construction for power management problems." In: *2003 IEEE Bologna Power Tech Conference Proceedings,* vol. 3. DOI: 10.1109/PTC.2003.1304379.

Haghighat, H. (2014). "Strategic offering under uncertainty in power markets." In: *International Journal of Electrical Power & Energy Systems* 63, pp. 1070–1077. DOI: 10.1016/j.ijepes.2014.05.049.

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.

Heitsch, H., R. Henrion, T. Kleinert, and M. Schmidt (2022). "On Convex Lower-Level Black-Box Constraints in Bilevel Optimization with an Application to Gas Market Models with Chance Constraints." In: *Journal of Global Optimization*. DOI: 10.1007/s10898-022-01161-z.

Held, H., R. Hemmecke, and D. L. Woodruff (2005). "A decomposition algorithm applied to planning the interdiction of stochastic networks." In: *Naval Research Logistics* 52.4, pp. 321–328. DOI: https://doi.org/10.1002/nav.20079.

Hemmecke, R., R. Schultz, and D. L. Woodruff (2003). "Interdicting Stochastic Networks with Binary Interdiction Effort." In: *Network Interdiction and Stochastic Integer Programming*. Ed. by D. L. Woodruff. Boston, MA: Springer US, pp. 69–84. DOI: 10.1007/0-306-48109-X_4.

Henkel, C. (2014). "An Algorithm for the Global Resolution of Linear Stochastic Bilevel Programs." PhD thesis. Universität Duisburg-Essen. URL: https://duepublico2.uni-due.de/receive/duepublico_mods_00035553.

Hu, J. and S. Mehrotra (2015). "Robust decision making over a set of random targets or risk-averse utilities with an application to portfolio optimization." In: *IIE Transactions* 47.4, pp. 358–372. DOI: 10.1080/0740817X.2014.919045.

Hu, M. and M. Fukushima (2015). "Multi-Leader-Follower Games: Models, Methods and Applications." In: *Journal of the Operations Research Society of Japan* 58.1, pp. 1–23. DOI: 10.15807/jorsj.58.1.

Israeli, E. (1999). "System Interdiction and Defense." PhD thesis. Naval Postgraduate School Monterey, California. URL: https://apps.dtic.mil/sti/pdfs/ADA361997.pdf.

Israeli, E. and R. K. Wood (2002). “Shortest-path network interdiction.” In: *Networks* 40.2, pp. 97–111. DOI: `10.1002/net.10039`.

Ivanov, S. V. (2018). “A Bilevel Stochastic Programming Problem with Random Parameters in the Follower’s Objective Function.” In: *Journal of Applied and Industrial Mathematics* 12.4, pp. 658–667. DOI: `10.1134/S1990478918040063`.

Ivanov, S. V. (2014). “Bilevel Stochastic Linear Programming Problems with Quantile Criterion.” In: *Automation and Remote Control* 75, pp. 107–118. DOI: `10.1134/S0005117914010081`.

Jain, M., J. Tsai, J. Pita, C. Kiekintveld, S. Rathi, M. Tambe, and F. Ordóñez (2010). “Software Assistants for Randomized Patrol Planning for the LAX Airport Police and the Federal Air Marshal Service.” In: *INFORMS Journal on Applied Analytics* 40.4, pp. 253–338. DOI: `10.1287/inte.1100.0505`.

Janjarassuk, U. and J. Linderoth (2008). “Reformulation and sampling to solve a stochastic network interdiction problem.” In: *Networks* 52.3, pp. 120–132. DOI: `https://doi.org/10.1002/net.20237`.

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`.

Jia, S., Z. Wan, Y. Feng, and G. Wang (2011). “New partial cooperation model for bilevel programming problems.” In: *Journal of Systems Engineering and Electronics* 22.2, pp. 263–266. URL: `https://ieeexplore.ieee.org/abstract/document/6077633`.

Jia, S. and Y. Zheng (2013). “A New Coordination Model for Ill-Posed Bilevel Programming Problem.” In: *Applied Mechanics and Materials* 411–414, pp. 1943–1947. DOI: `10.4028/www.scientific.net/AMM.411-414.1943`.

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

Kallabis, T., S. A. Gabriel, and C. Weber (2020). “Strategic generation investment using a stochastic rolling-horizon MPEC approach.” In: *Energy Systems*. DOI: `10.1007/s12667-020-00413-9`.

Kar, D., T. H. Nguyen, F. Fang, M. Brown, A. Sinha, M. Tambe, and A. X. Jiang (2018). “Trends and Applications in Stackelberg Security Games.” In: *Handbook of Dynamic Game Theory*. Ed. by T. Başar and G. Zaccour. Springer, Cham., pp. 1223–1269. DOI: `10.1007/978-3-319-44374-4_27`.

Karwowski, J., J. Mańdziuk, and A. Zychowski (2020). *Anchoring Theory in Sequential Stackelberg Games*. URL: `https://arxiv.org/abs/1912.03564`.

Katagiri, H., T. Uno, K. Kato, H. Tsuda, and H. Tsubaki (2014). “Random fuzzy bilevel linear programming through possibility-based value at risk model.” In: *International Journal of Machine Learning and Cybernetics* 5.2, pp. 211–224. DOI: `10.1007/s13042-012-0126-4`.

Kelly, E. S., C. R. Fox, and Y. S. Rottenstreich (2006). “Between ignorance and truth: Partition dependence and learning in judgment under uncertainty.” In: *Journal of Experimental Psychology. Learning, Memory, and Cognition* 32.6, pp. 1385–1402. DOI: `10.1037/0278-7393.32.6.1385`.

Kiekintveld, C., J. Marecki, and M. Tambe (2011). “Approximation Methods for Infinite Bayesian Stackelberg Games: Modeling Distributional Payoff Uncertainty.” In: *Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems*. AAMAS ’11, pp. 1005–1012.

Kleinert, T., M. Labbé, I. Ljubić, and M. Schmidt (2021). “A Survey on Mixed-Integer Programming Techniques in Bilevel Optimization.” In: *EURO Journal of Computational Optimization* 9, p. 100007. DOI: `10.1016/j.ejco.2021.100007`.

Kleywegt, A. J., A. Shapiro, and T. Homem-de-Mello (2002). "The sample average approximation method for stochastic discrete optimization." In: *SIAM Journal on Optimization* 12.2, pp. 479–502. DOI: `10.1137/S1052623499363220`.

Korzhyk, D., V. Conitzer, and R. Parr (2011). "Solving Stackelberg Games with Uncertain Observability." In: *The 10th International Conference on Autonomous Agents and Multiagent Systems*. Vol. 3. AAMAS 2011. International Foundation for Autonomous Agents and Multiagent Systems, pp. 1013–1020. URL: `http://www.ifaamas.org/Proceedings/aamas2011/papers/C7_G76.pdf`.

Kosuch, S., P. L. Bodic, J. Leung, and A. Lisser (2012). "On a Stochastic Bilevel Programming Problem." In: *Networks* 59.1, pp. 107–116. DOI: `10.1002/net.20482`.

Kovacevic, R. M. and G. C. Pflug (2014). "Electricity swing option pricing by stochastic bilevel optimization: A survey and new approaches." In: *European Journal of Operational Research* 237.2, pp. 389–403. DOI: `10.1016/j.ejor.2013.12.029`.

Labbé, M., P. Marcotte, and G. Savard (1998). "A bilevel model of taxation and its application to optimal highway pricing." In: *Management Science* 44.12-part-1, pp. 1608–1622. DOI: `10.1287/mnsc.44.12.1608`.

Labbé, M., F. Plein, and M. Schmidt (2020). "Bookings in the European Gas Market: Characterisation of Feasibility and Computational Complexity Results." In: *Optimization and Engineering* 21.1, pp. 305–334. DOI: `10.1007/s11081-019-09447-0`.

Labbé, M., F. Plein, M. Schmidt, and J. Thürauf (2021). "Deciding feasibility of a booking in the European gas market on a cycle is in P for the case of passive networks." In: *Networks*. DOI: `10.1002/net.22003`.

Labbé, M. and A. Violin (2016). "Bilevel programming and price setting problems." In: *Annals of Operations Research* 240.1, pp. 141–169. DOI: `10.1007/s10479-015-2016-0`.

Lappas, N. H. and C. E. Gounaris (2018). "Robust optimization for decision-making under endogenous uncertainty." In: *Computers & Chemical Engineering* 111, pp. 252–266. DOI: `10.1016/j.compchemeng.2018.01.006`.

Lei, X., S. Shen, and Y. Song (2018). "Stochastic maximum flow interdiction problems under heterogeneous risk preferences." In: *Computers & Operations Research* 90, pp. 97–109. DOI: `10.1016/j.cor.2017.09.004`.

Letchford, J. and Y. Vorobeychik (2013). "Optimal Interdiction of Attack Plans." In: *Proceedings of the 2013 International Conference on Autonomous Agents and Multi-Agent Systems*. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, pp. 199–206.

Lin, G.-H. and M. Fukushima (2010). "Stochastic equilibrium problems and stochastic mathematical programs with equilibrium constraints: A survey." In: *Pacific Journal of Optimization* 6.3, pp. 455–482.

Liu, J., Y. Fan, Z. Chen, and Y. Zheng (2020). "Methods for Pessimistic Bilevel Optimization." In: *Bilevel Optimization - Advances and Next Challenges*. Springer, Cham, pp. 403–420. DOI: `10.1007/978-3-030-52119-6_14`.

Liu, J., Y. Fan, Z. Chen, and Y. Zheng (2018). "Pessimistic Bilevel Optimization: A Survey." In: *International Journal of Computational Intelligence Systems* 11.1, pp. 725–736. DOI: `10.2991/ijcis.11.1.56`.

Liu, X. and C. Kwon (2020). "Exact robust solutions for the combined facility location and network design problem in hazardous materials transportation." In: *IISE Transactions* 52.10, pp. 1156–1172. DOI: `10.1080/24725854.2019.1697017`.

Longsheng, S., M. H. Karwan, and K. Changhyun (2017). “Generalized Bounded Rationality and Robust Multicommodity Network Design.” In: *Operations Research* 66.1. DOI: 10.1287/opre.2017.1621.

Loridan, P. and J. Morgan (1989). “ϵ-Regularized two-level optimization problems: Approximation and existence results.” In: *Optimization*. Ed. by S. Dolecki. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 99–113. DOI: 10.1007/BFb0083589.

Lu, Z., K. Deb, and A. Sinha (2015). “Handling decision variable uncertainty in bilevel optimization problems.” In: pp. 1683–1690. DOI: 10.1109/CEC.2015.7257089.

Lunday, B. J. and H. D. Sherali (2012). “Network interdiction to minimize the maximum probability of evasion with synergy between applied resources.” In: *Annals of Operations Research* 196.1, pp. 411–442. DOI: 10.1007/s10479-012-1135-0.

Mallozzi, L. and J. Morgan (1996). “Hierarchical Systems with Weighted Reaction Set.” In: *Nonlinear Optimization and Applications*. Ed. by G. D. Pillo and F. Giannessi. Springer, Boston, MA, pp. 271–282. DOI: 10.1007/978-1-4899-0289-4_19.

Michalopoulos, D. P., J. W. Barnes, and D. P. Morton (2015). “Prioritized interdiction of nuclear smuggling via tabu search.” In: *Optimization Letters* 9.8, pp. 1477–1494. DOI: 10.1007/s11590-014-0829-4.

Molan, I. and M. Schmidt (2022). *Using Neural Networks to Solve Linear Bilevel Problems with Unknown Lower Level*. Tech. rep. URL: http://www.optimization-online.org/DB_HTML/2022/07/8972.html.

Morton, D. P. (2011). “Stochastic network interdiction.” In: *Wiley Encyclopedia of Operations Research and Management Science*. Ed. by J. J. Cochran, L. A. C. Jr., P. Keskinocak, J. P. Kharoufeh, and J. C. Smith. Hoboken: Wiley. DOI: 10.1002/9780470400531.

Morton, D. P., F. Pan, and K. J. Saeger (2007). “Models for nuclear smuggling interdiction.” In: *IIE Transactions* 39.1, pp. 3–14. DOI: 10.1080/07408170500488956.

Nemhauser, G. L., L. A. Wolsey, and M. L. Fisher (1978). “An analysis of approximations for maximizing submodular set functions—I.” In: *Mathematical Programming* 14.1, pp. 265–294. DOI: 10.1007/BF01588971.

Nguyen, D. H. and J. C. Smith (2022). “Network interdiction with asymmetric cost uncertainty.” In: *European Journal of Operational Research* 297.1, pp. 239–251. DOI: 10.1016/j.ejor.2021.04.055.

Nikoofal, M. E. and J. Zhuang (2015). “On the value of exposure and secrecy of defense system: First-mover advantage vs. robustness.” In: *European Journal of Operational Research* 246.1, pp. 320–330. DOI: 10.1016/j.ejor.2015.04.043.

Nikoofal, M. E. and J. Zhuang (2011). “Robust Allocation of a Defensive Budget Considering an Attacker’s Private Information.” In: *Risk Analysis* 32.5. DOI: 10.1111/j.1539-6924.2011.01702.x.

Nohadani, O. and K. Sharma (2018). “Optimization under Decision-Dependent Uncertainty.” In: *SIAM Journal on Optimization* 28.2, pp. 1773–1795. DOI: 10.1137/17M1110560.

Özaltın, O. Y., O. A. Prokopyev, and A. J. Schaefer (2010). “The bilevel knapsack problem with stochastic right-hand sides.” In: *Operations Research Letters* 38, pp. 328–333. DOI: 10.1016/j.orl.2010.04.005.

Pan, F., W. S. Charlton, and D. P. Morton (2003). “A Stochastic Program for Interdicting Smuggled Nuclear Material.” In: *Network Interdiction and Stochastic Integer Programming*. Ed. by D. L. Woodruff. Boston, MA: Springer US, pp. 1–19. DOI: 10.1007/0-306-48109-X_1.

Pan, F. and D. P. Morton (2008). "Minimizing a stochastic maximum-reliability path." In: *Networks* 52.3, pp. 111–119. DOI: https://doi.org/10.1002/net.20238.

Paruchuri, P., J. P. Pearce, J. Marecki, M. Tambe, F. Ordóñez, and S. Kraus (2008). "Playing Games for Security: An Efficient Exact Algorithm for Solving Bayesian Stackelberg Games." In: *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems.* AAMAS '08. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, pp. 895–902.

Paruchuri, P., J. P. Pearce, M. Tambe, F. Ordóñez, and S. Kraus (2007). "An Efficient Heuristic Approach for Security against Multiple Adversaries." In: *Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems.* AAMAS '07. New York, NY, USA: Association for Computing Machinery. DOI: 10.1145/1329125.1329344.

Patriksson, M. (2008a). "On the applicability and solution of bilevel optimization models in transportation science: A study on the existence, stability and computation of optimal solutions to stochastic mathematical programs with equilibrium constraints." In: *Transportation Research Part B: Methodological* 42.10, pp. 843–860. DOI: 10.1016/j.trb.2008.05.001.

Patriksson, M. (2008b). "Robust bi-level optimization models in transportation science." In: *Philosophical Transactions of the Royal Society A* 366.1872. DOI: 10.1098/rsta.2008.0007.

Patriksson, M. and L. Wynter (1999). "Stochastic mathematical programs with equilibrium constraints." In: *Operations Research Letters* 25.4, pp. 159–167. DOI: 10.1016/S0167-6377(99)00052-8.

Patriksson, M. and L. Wynter (1997). *Stochastic Nonlinear Bilevel Programming.* URL: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.16.5155.

Pay, B. S., J. R. W. Merrick, and Y. Song (2018). "Stochastic network interdiction with incomplete preference." In: *Networks* 73.1, pp. 3–22. DOI: 10.1002/net.21831.

Pita, J., M. Jain, F. Ordóñez, M. Tambe, S. Kraus, and R. Magori-Cohen (2009). "Effective solutions for real-world Stackelberg games: When agents must deal with human uncertainties." In: *AAMAS 2009—Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems.* Vol. 1, pp. 369–376. DOI: 10.1017/CBO9780511973031.010.

Pita, J., M. Jain, M. Tambe, F. Ordóñez, and S. Kraus (2010). "Robust solutions to Stackelberg games: Addressing bounded rationality and limited observations in human cognition." In: *Artificial Intelligence* 174.15, pp. 1142–1171. DOI: 10.1016/j.artint.2010.07.002.

Pita, J., C. Portway, M. Tambe, P. Paruchuri, and S. Kraus (2008). "Robust Solutions in Stackelberg Games: Addressing Boundedly Rational Human Preference Models." In: *AAAI 2008—4th Multidisciplinary Workshop on Advances in Preference Handling.* URL: https://www.aaai.org/Papers/Workshops/2008/WS-08-09/WS08-09-010.pdf.

Plein, F., J. Thürauf, M. Labbé, and M. Schmidt (2021). "Bilevel Optimization Approaches to Decide the Feasibility of Bookings in the European Gas Market." In: *Mathematical Methods of Operations Research.* S.I.: Energy Networks. DOI: 10.1007/s00186-021-00752-y.

Pramanik, S. and D. Banerjee (2012). "Chance Constrained Quadratic Bi-level Programming Problem." In: *International Journal of Modern Engineering Research* 2.4, pp. 2417–2424.

Rahmaniani, R., T. G. Crainic, M. Gendreau, and W. Rei (2018). "Accelerating the Benders Decomposition Method: Application to Stochastic Network Design Problems." In: *SIAM Journal on Optimization* 28.1, pp. 875–903. DOI: `10.1137/17M1128204`.

Riis, M. and R. Schultz (2003). "Applying the Minimum Risk Criterion in Stochastic Recourse Programs." In: *Computational Optimization and Applications* 24.2–3, pp. 267–287. DOI: `10.1023/A:1021862109131`.

Rockafellar, R. T. and R. J.-B. Wets (1998). *Variational Analysis*. Springer Berlin Heidelberg 1998. DOI: `10.1007/978-3-642-02431-3`.

Roghanian, E., S. Sadjadi, and M. Aryanezhad (2007). "A probabilistic bi-level linear multi-objective programming problem to supply chain planning." In: *Applied Mathematics and Computation* 188.1, pp. 786–800. DOI: `https://doi.org/10.1016/j.amc.2006.10.032`.

Rubinstein, A. (1998). *Modeling Bounded Rationality*. The MIT Press. URL: `https://mitpress.mit.edu/books/modeling-bounded-rationality`.

Ryu, J.-H., V. Dua, and E. N. Pistikopoulos (2004). "A bilevel programming framework for enterprise-wide process networks under uncertainty." In: *Computers & Chemical Engineering* 28.6-–7, pp. 1121–1129. DOI: `10.1016/j.compchemeng.2003.09.021`.

Salas, D. and A. Svensson (2020). *Existence of solutions for deterministic bilevel games under a general Bayesian approach*. URL: `https://arxiv.org/pdf/2010.05368v4.pdf`.

Schewe, L., M. Schmidt, and J. Thürauf (2020). "Computing Technical Capacities in the European Entry-Exit Gas Market is NP-Hard." In: *Annals of Operations Research* 295, pp. 337–362. DOI: `10.1007/s10479-020-03725-2`.

Schewe, L., M. Schmidt, and J. Thürauf (2022). "Global Optimization for the Multilevel European Gas Market System with Nonlinear Flow Models on Trees." In: *Journal of Gloal Optimization*. DOI: `10.1007/s10898-021-01099-8`.

Scholtes, S. (2001). "Convergence Properties of a Regularization Scheme for Mathematical Programs with Complementarity Constraints." In: *SIAM Journal on Optimization* 11.4, pp. 918–936. DOI: `10.1137/S1052623499361233`.

Shapiro, A. and H. Xu (2008). "Stochastic mathematical programs with equilibrium constraints, modelling and sample average approximation." In: *Optimization* 57.3, pp. 395–418. DOI: `10.1080/02331930801954177`.

Sherali, H. D. and W. P. Adams (1990). "A Hierarchy of Relaxations between the Continuous and Convex Hull Representations for Zero-One Programming Problems." In: *SIAM Journal on Discrete Mathematics* 3.3, pp. 411–430. DOI: `10.1137/0403036`.

Shi, X., O. Prokopyev, and T. K. Ralphs (2020). *Mixed Integer Bilevel Optimization with k-optimal Follower: A Hierarchy of Bounds*. Tech. rep. URL: `https://coral.ise.lehigh.edu/~ted/files/papers/HierarchyOfBounds20.pdf`.

Simon, H. (1972). "Theories of Bounded Rationality." In: *McGuire, C.B. and Radner, R., Eds., Decision and Organization*. North Holland Publishing Company, pp. 161–176. URL: `http://innovbfa.viabloga.com/files/Herbert_Simon___theories_of_bounded_rationality___1972.pdf`.

Sinha, A., F. Fang, B. An, C. Kiekintveld, and M. Tambe (2018). "Stackelberg Security Games: Looking Beyond a Decade of Success." In: *IJCAI*. URL: `https://ink.library.smu.edu.sg/sis_research/4792/`.

Smith, J. C. and Y. Song (2020). "A survey of network interdiction models and algorithms." In: *European Journal of Operational Research* 283.3, pp. 797–811. DOI: `10.1016/j.ejor.2019.06.024`.

Song, Y. and S. Shen (2016). “Risk-Averse Shortest Path Interdiction.” In: *INFORMS Journal on Computing* 28.3, pp. 527–539. DOI: 10.1287/ijoc.2016.0699.

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.

Su, Y. and J. Geunes (2013). “Multi-period price promotions in a single-supplier, multi-retailer supply chain under asymmetric demand information.” In: *Annals of Operations Research* 211.1, pp. 447–472. DOI: 10.1007/s10479-013-1485-2.

Sullivan, K. M., J. C. Smith, and D. P. Morton (2014). “Convex hull representation of the deterministic bipartite network interdiction problem.” In: *Mathematical Programming* 145.1–2, pp. 349–376. DOI: 10.1007/s10107-013-0650-3.

Sun, L., M. H. Karwan, and C. Kwon (2015). “Robust Hazmat Network Design Problems Considering Risk Uncertainty.” In: *Transportation Science* 50.4, pp. 1188–1203. DOI: 10.1287/trsc.2015.0645.

Tahernejad, S. (2019). “Two-Stage Mixed Integer Stochastic Bilevel Linear Optimization.” PhD thesis. Lehigh University.

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.

Thürauf, J. (2022). “Deciding the Feasibility of a Booking in the European Gas Market is coNP-hard.” In: *Annals of Operations Research*. To appear. URL: http://www.optimization-online.org/DB_HTML/2020/05/7803.html.

Towle, E. and J. R. Luedtke (2018). “New solution approaches for the maximum-reliability stochastic network interdiction problem.” In: *Computational Management Science* 15.3-4, pp. 455–477. DOI: 10.1007/s10287-018-0321-1.

Tsoukalas, A., W. Wiesemann, and B. Rustem (2009). *Global Optimisation of Pessimistic Bi-Level Problems*. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.154.9021&rep=rep1&type=pdf.

Tversky, A. and D. J. Koehler (1994). “Support theory: A nonextensional representation of subjective probability.” In: *Psychological Review* 101.4, pp. 547–567. DOI: 10.1037/0033-295X.101.4.547.

U-tapao, C., S. Moryadee, S. A. Gabriel, C. Peot, and M. Ramirez (2016). “A stochastic, two-level optimization model for compressed natural gas infrastructure investments in wastewater management.” In: *Journal of Natural Gas Science and Engineering* 28, pp. 226–240. DOI: 10.1016/j.jngse.2015.11.039.

Unsal, O. (2010). “Two-Person Zero-Sum Network-Interdiction Game with Multiple Inspector Types.” MA thesis. Naval Postgraduate School Montery,CA. URL: https://apps.dtic.mil/sti/citations/ADA524779.

van Damme, E. and S. Hurkens (1997). “Games with Imperfectly Observable Commitment.” In: *Games and Economic Behavior* 21.1–2, pp. 282–308. DOI: 10.1006/game.1997.0524.

Van Slyke, R. M. and R. Wets (1969). “L-shaped linear programs with applications to optimal control and stochastic programming.” In: *SIAM Journal on Applied Mathematics* 17.4, pp. 638–663. DOI: 10.1137/0117061.

Violin, A. (2014). “Mathematical programming approaches to pricing problems.” Université Libre de Bruxelles.

von Stackelberg, H. (1934). *Marktform und Gleichgewicht*. Springer.

von Stackelberg, H. (1952). *Theory of the market economy*. Oxford University Press.

Werner, A. S. (2004). “Bilevel stochastic programming problems: analysis and application to telecommunications.” PhD thesis. Norwegian University of Science and Technology. URL: https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/264997.

Wiesemann, W., A. Tsoukalas, P.-M. Kleniati, and B. Rustem (2013). “Pessimistic Bilevel Optimization.” In: *SIAM Journal on Optimization* 23.1, pp. 353–380. DOI: `10.1137/120864015`.

Wynter, L. (2008). “Stochastic Bilevel Programs.” In: *Encyclopedia of Optimization*. Ed. by C. Floudas and P. Pardalos. Springer, Boston, MA. DOI: `10.1007/978-0-387-74759-0_646`.

Xin, C., L. Qingge, J. Wang, and B. Zhu (2015). “Robust optimization for the hazardous materials transportation network design problem.” In: *Journal of Combinatorial Optimization* 30, pp. 320–334. DOI: `10.1007/s10878-014-9751-z`.

Yan, N., H. Dai, and B. Sun (2014). “Optimal bi-level Stackelberg strategies for supply chain financing with both capital-constrained buyers and sellers.” In: *Applied Stochastic Models in Business and Industry* 30.6, pp. 783–796. DOI: `10.1002/asmb.2021`.

Yang, J., J. S. Borrero, O. A. Prokopyev, and D. Sauré (2021). “Sequential Shortest Path Interdiction with Incomplete Information and Limited Feedback.” In: *Decision Analysis* 18.3, pp. 218–244. DOI: `10.1287/deca.2021.0426`.

Yang, J., M. Zhang, B. He, and C. Yang (2009). “Bi-level programming model and hybrid genetic algorithm for flow interception problem with customer choice.” In: *Computers & Mathematics with Applications* 57.11, pp. 1985–1994. DOI: `10.1016/j.camwa.2008.10.035`.

Yanıkoğlu, İ. and D. Kuhn (2018). “Decision Rule Bounds for Two-Stage Stochastic Bilevel Programs.” In: *SIAM Journal on Optimization* 28.1, pp. 198–222. DOI: `10.1137/16M1098486`.

Yeh, K., C. Whittaker, M. J. Realff, and J. H. Lee (2015). “Two stage stochastic bilevel programming model of a pre-established timberlands supply chain with biorefinery investment interests.” In: *Computers & Chemical Engineering* 73, pp. 141–153. DOI: `https://doi.org/10.1016/j.compchemeng.2014.11.005`.

Yin, Z., M. Jain, M. Tambe, and F. Ordóñez (2011). “Risk-Averse Strategies for Security Games with Execution and Observational Uncertainty.” In: *Twenty-Fifth AAAI Conference on Artificial Intelligence*.

Yin, Z. and M. Tambe (2012). “A Unified Method for Handling Discrete and Continuous Uncertainty in Bayesian Stackelberg Games.” In: *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems*. Vol. 2. AAMAS 2012. International Foundation for Autonomous Agents and Multiagent Systems, pp. 855–862. URL: `https://dl.acm.org/doi/10.5555/2343776.2343819`.

Zare, M. H., O. Y. Özaltın, and O. A. Prokopyev (2018). “On a class of bilevel linear mixed-integer programs in adversarial settings.” In: *Journal of Global Optimization* 71, pp. 91–113. DOI: `10.1007/s10898-017-0549-2`.

Zare, M. H., O. A. Prokopyev, and D. Sauré (2020). “On Bilevel Optimization with Inexact Follower.” In: *Decision Analysis* 17.1. DOI: `10.1287/deca.2019.0392`.

Zeng, B., H. Dong, R. Sioshansi, F. Xu, and M. Zeng (2020). “Bilevel Robust Optimization of Electric Vehicle Charging Stations With Distributed Energy Resources.” In: *IEEE Transactions on Industry Applications* 56.5, pp. 5836–.5847. DOI: `10.1109/TIA.2020.2984741`.

Zeng, B. (2020). “A Practical Scheme to Compute the Pessimistic Bilevel Optimization Problem.” In: *INFORMS Journal on Computing* 32.4. DOI: `10.1287/ijoc.2019.0927`.

Zeng, B. and L. Zhao (2013). “Solving two-stage robust optimization problems using a column-and-constraint generation method.” In: *Operations Research Letters* 41.5, pp. 457–461. DOI: `10.1016/j.orl.2013.05.003`.

Zhang, J. and O. Y. Özaltın (2021). “Bilevel Integer Programs with Stochastic Right-Hand Sides.” In: *INFORMS Journal on Computing* 33.4, pp. 1644–1660. DOI: 10.1287/ijoc.2020.1055.

Zhang, Y., F. Liu, Z. Wang, Y. Chen, S. Feng, Q. Wu, and Y. Hou (2022). *On Nash-Stackelberg-Nash Games under Decision-Dependent Uncertainties: Model and Equilibrium*. URL: https://arxiv.org/pdf/2202.11880.pdf.

Zheng, K. and L. A. Albert (2019). “Interdiction models for delaying adversarial attacks against critical information technology infrastructure.” In: *Naval Research Logistics* 66.5, pp. 411–429. DOI: 10.1002/nav.21859.

Zheng, Y., Z. Wan, K. Sun, and T. Zhang (2013). “An exact penalty method for weak linear bilevel programming problem.” In: *Journal of Applied Mathematics and Computing* 42, pp. 41–49. DOI: 10.1007/s12190-012-0620-6.

Zheng, Y., Z. Wan, S. Jia, and G. Wang (2015). “A new method for strong-weak linear bilevel programming problem.” In: *Journal of Industrial & Management Optimization* 11.2, pp. 529–547. DOI: 10.3934/jimo.2015.11.529.

(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

Article 3

Exact methods for discrete Γ-robust interdiction problems with an application to the bilevel knapsack problem

Yasmine Beck, Ivana Ljubić, and Martin Schmidt

Mathematical Programming Computation (2023)
https://doi.org/10.1007/s12532-023-00244-6

EXACT METHODS FOR DISCRETE Γ-ROBUST INTERDICTION PROBLEMS WITH AN APPLICATION TO THE BILEVEL KNAPSACK PROBLEM

YASMINE BECK, IVANA LJUBIĆ, AND MARTIN SCHMIDT

ABSTRACT. Developing solution methods for discrete bilevel problems is known to be a challenging task—even if all parameters of the problem are exactly known. Many real-world applications of bilevel optimization, however, involve data uncertainty. We study discrete min-max problems with a follower who faces uncertainties regarding the parameters of the lower-level problem. Adopting a Γ-robust approach, we present an extended formulation and a multi-follower formulation to model this type of problem. For both settings, we provide a generic branch-and-cut framework. Specifically, we investigate interdiction problems with a monotone Γ-robust follower and we derive problem-tailored cuts, which extend existing techniques that have been proposed for the deterministic case. For the Γ-robust knapsack interdiction problem, we computationally evaluate and compare the performance of the proposed algorithms for both modeling approaches.

1. INTRODUCTION

In the last years and decades, bilevel optimization problems have gained increasing attention due to their ability to model hierarchical decision making processes that occur in various applications such as transportation (Ben-Ayed et al. 1992; Migdalas 1995), energy markets (Arroyo 2010; Grimm et al. 2019), or pricing (Dempe and Zemkoho 2012; Labbé et al. 1998). In bilevel problems, the decision maker on the upper level (the leader) makes a decision anticipating the reaction of the lower-level player (the follower). In this paper, we consider discrete linear bilevel problems of the form

$$\begin{aligned} \min_{x} \quad & c^\top x + d^\top y && \text{(1a)}\\ \text{s.t.} \quad & Ax \geq a, && \text{(1b)}\\ & x \in X \subseteq \mathbb{Z}^{n_x}, && \text{(1c)}\\ & y \in \arg\max_{y'} \left\{ d^\top y' \colon y' \in Y(x) \subseteq \mathbb{Z}_+^{n_y} \right\}, && \text{(1d)} \end{aligned}$$

where $Y(x)$ denotes the lower-level feasible set that is parameterized by the leader's variables x. The set X is used to denote integrality constraints. Moreover, we have $c \in \mathbb{R}^{n_x}$, $d \in \mathbb{R}^{n_y}$, $A \in \mathbb{R}^{k \times n_x}$, and $a \in \mathbb{R}^k$. We refer to (1a)–(1c) as the upper-level and to (1d) as the lower-level problem. Note that Problem (1) is a min-max problem. Hence, the follower's response yields the worst-possible outcome for the leader, which is why there is no need to distinguish between the optimistic and the pessimistic approach; see, e.g., Dempe (2002). Let us further point out that we do not consider coupling constraints in the upper level, which is crucial for the validity of the branch-and-cut methods we propose in the following sections. In particular, this type of problem covers the important classes of interdiction (G. Brown et al. 2006; Cormican et al. 1998; DeNegre 2011; Fischetti, Ljubić, et al. 2019; Furini, Ljubić, Malaguti, et al. 2021; Israeli and Wood 2002; Wood 2011) and blocking problems (Bazgan et al. 2013; Furini, Ljubić, Malaguti, et al. 2020; Golden 1978; Pajouh 2020; Pajouh, Boginski, et al. 2014; Pajouh, Walteros, et al. 2015; Zenklusen et al. 2009) that arise in various real-world applications such as in critical infrastructure defense, network

---

Date: April 21, 2023.

2010 *Mathematics Subject Classification.* 90C11, 90C27, 90C57, 90C70.

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

disruption, or marketing. A recent survey on network interdiction models and algorithms can be found in Smith and Song (2020).

Due to their nested structure, even linear bilevel problems are strongly NP-hard in general; see, e.g., Hansen et al. (1992). Moreover, merely checking feasibility for mixed-integer bilevel problems is an NP-hard problem. Thus, it is a difficult task to develop solution methods—especially for bilevel problems that involve discrete variables. In the seminal work by Moore and Bard (1990), the first branch-and-bound method for solving mixed-integer linear bilevel problems is discussed. The idea is extended by DeNegre and Ralphs (2009) who provide a branch-and-cut approach that is based on techniques of standard integer linear programming. In particular, this work can be considered as a turning point regarding computational mixed-integer bilevel optimization that is followed by many influential works on solution methods for bilevel problems; see, e.g., Fischetti, Ljubić, et al. (2017, 2018), Tahernejad, Ralphs, and DeNegre (2020), and Xu and Wang (2014). For a detailed discussion on further techniques for mixed-integer bilevel optimization, we refer to the recent survey in Kleinert et al. (2021).

Throughout this paper, we say that an upper-level decision x is feasible if $x \in X$ and $Ax \geq a$ are satisfied. Moreover, we will hold on to the following.

Assumption 1. *For every feasible decision x of the leader, the lower-level feasible set $Y(x)$ is non-empty.*

Assumption 2. *The shared constraint set $\{(x, y) \colon Ax \geq a,\ x \in X,\ y \in Y(x)\}$ is non-empty and compact.*

Assumption 3. *All linking variables, i.e., all variables of the leader that appear in the lower-level constraints, are bounded integers.*

Assumptions 1–3 are necessary to ensure that Problem (1) has an optimal solution. For a feasible upper-level decision x, we further define the lower-level optimal-value function

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

to re-write Problem (1) as the single-level problem

$$\begin{aligned} \min_{x \in X, \eta \in \mathbb{R}} \quad & c^\top x + \eta && \text{(2a)} \\ \text{s.t.} \quad & Ax \geq a, && \text{(2b)} \\ & \eta \geq \Phi(x). && \text{(2c)} \end{aligned}$$

Up to this point, we implicitly made the assumption that the input data of both players is certain. However, in many practical situations, this is not the case and the players are forced to make decisions under uncertainty. In mathematical optimization, there are two main approaches to deal with data uncertainty—stochastic optimization (Birge and Louveaux 2011) and robust optimization (Ben-Tal et al. 2009; Bertsimas, D. Brown, et al. 2010; Soyster 1973). In stochastic optimization, it is assumed that the uncertainties can be described by probability distributions that are known in advance. In this setting, the decision maker hedges against uncertainties in a probabilistic sense, e.g., by optimizing over expected values, by considering chance constraints, or some risk-averse models. In between stochastic and robust optimization there is the further approach of distributional robustness; see, e.g., Goh and Sim (2010). In this paper, however, we focus on a robust approach. In robust optimization, the decision maker is interested in a solution that is feasible for all possible realizations of the uncertain data that are assumed to take values in a given uncertainty set. Thus, one pursues a worst-case-oriented philosophy. However, a major point of criticism regarding this approach is the possible over-conservatism of solutions in the sense that ensuring robustness can be very expensive. Addressing this matter, Bertsimas and Sim (2003, 2004) and Sim (2004) propose a more flexible robust approach—the so-called Γ-robust approach—which allows to control the level of conservatism of the solution. In this setting, it is assumed that the decision maker hedges against the cases in which only a subset of the uncertain parameters will change as to adversely affect the solution of the problem at hand.

In the context of bilevel optimization, problems involving data uncertainties have been investigated using both stochastic as well as robust approaches. Cormican et al. (1998) and Israeli (1999) address stochastic network interdiction problems with uncertainties regarding the interdiction success and uncertain arc capacities, respectively. A stochastic approach for interdiction problems under uncertainty is also considered in the survey by Smith and Song (2020). Further works that pursue a stochastic approach for more general bilevel problems under uncertainty can be found, e.g., in Burtscheidt and Claus (2020), Burtscheidt, Claus, and Dempe (2020), Dempe, Ivanov, et al. (2017), Ivanov (2018), and Yanıkoğlu and Kuhn (2018). To the best of our knowledge, robust approaches to address data uncertainty in bilevel optimization have been much less investigated. In the context of power markets, a Γ-robust approach to deal with uncertain lower-level data is considered in Haghighat (2014). Chuong and Jeyakumar (2017) consider problems with uncertain upper- as well as lower-level constraints and solve the robust counterpart via a sequence of semidefinite programming relaxations. In Zeng et al. (2020), a worst-case oriented approach for bilevel problems with lower-level data uncertainty is addressed. Buchheim and Henke (2020) and Buchheim, Henke, and Hommelsheim (2021) present complexity results for robust bilevel problems with uncertainties regarding the lower-level objective function coefficients. For a brief introduction to robust bilevel optimization, we refer to Beck, Ljubić, et al. (2022) and for a general discussion on bilevel optimization under uncertainty, we refer to the recent survey by Beck, Ljubić, et al. (2023).

Let us further mention that, in bilevel optimization, the sources for uncertainty are much richer compared to classic, i.e., single-level, optimization. In bilevel optimization, uncertainties may not only arise in the problem's data but there may also be uncertainty regarding the (observation of the) decisions of the two players. Despite being still in its infancy, there are a few works that consider robust approaches to deal with decision uncertainty. Besançon et al. (2019) propose a robust approach to hedge against near-optimal lower-level decisions. In this context, complexity results are discussed in Besançon et al. (2021). A similar setting in which the leader anticipates sub-optimal follower's decisions due to lower-level algorithmic uncertainty is considered in Zare et al. (2020). The authors consider the setting in which the leader hedges against the Γth least damaging choices of the solution algorithm for the lower-level problem, which is to some extent related to the notion of Γ-robustness proposed in Bertsimas and Sim (2003). Lastly, robust optimization techniques are used in Beck and Schmidt (2021) to model follower's response uncertainty due to limited observability regarding the leader's decision.

The contributions of this paper are the following. We study discrete linear bilevel problems involving a follower who faces uncertainties regarding the parameters of the lower-level problem. In this context, we pursue the same idea as in Bertsimas and Sim (2003, 2004) and Sim (2004) so that the follower aims to only hedge against a subset of deviations of uncertain parameters. In contrast to the aforementioned literature, we consider bilevel problems that involve discrete variables on the lower level. Therefore, standard reformulation techniques like replacing the lower-level problem by its Karush–Kuhn–Tucker conditions (see, e.g., Fortuny-Amat and McCarl (1981)) cannot be applied anymore.

With regard to the uncertainties, we distinguish the following two cases. On the one hand, we assume that the lower-level's objective function coefficients are uncertain. Instead of d_i, we consider the uncertain coefficients $\bar{d}_i$, where $\bar{d}_i \in [d_i - \Delta d_i, d_i]$ for all $i \in [n_y] := \{1, \dots, n_y\}$. We denote d_i as the nominal value of the ith coefficient of the lower-level's objective function and Δd_i as its maximum deviation from the nominal value. For a feasible upper-level decision x, the robust counterpart of the lower-level problem (1d) in which the follower hedges against at most $\Gamma_d \in \{0, \dots, n_y\}$ deviations in the objective function coefficients is given by

$$\Phi_d(x) = \max_{y \in Y(x)} \left\{ d^\top y - \max_{\{S \subseteq [n_y] \colon |S| \le \Gamma_d\}} \sum_{i \in S} \Delta d_i y_i \right\}. \tag{3}$$

The overall bilevel problem with a Γ_d-robust follower facing uncertain objective function coefficients can thus be written as

$$\min_{x \in X, \eta \in \mathbb{R}} \quad c^\top x + \eta \quad \text{s.t.} \quad Ax \ge a, \, \eta \ge \Phi_d(x). \tag{4}$$

On the other hand, we deal with uncertainties regarding the lower-level constraints. Here, we focus on the specific case of a single packing-type constraint of the form $\bar{w}^\top y + v^\top x \le C$ with $v \in \mathbb{R}^{n_x}$, $C \in \mathbb{R}$, and $\bar{w} \in \mathbb{R}^{n_y}$, where $\bar{w}_i \in [w_i, w_i + \Delta w_i]$ for all $i \in [n_y]$ with $w, \Delta w \in \mathbb{R}^{n_y}$. For a feasible upper-level decision x, the robust counterpart in which the follower hedges against at most $\Gamma_w \in \{0, \dots, n_y\}$ deviations in the constraint coefficients is given by

$$\Phi_w(x) = \max_{y \in Y(x)} \left\{ d^\top y \colon w^\top y + v^\top x + \max_{\{S \subseteq [n_y] \colon |S| \le \Gamma_w\}} \sum_{i \in S} \Delta w_i y_i \le C \right\} \tag{5}$$

such that the Γ_w-robust counterpart of Problem (2) can be written as

$$\min_{x \in X, \eta \in \mathbb{R}} \quad c^\top x + \eta \quad \text{s.t.} \quad Ax \ge a, \, \eta \ge \Phi_w(x). \tag{6}$$

Let us mention that we focus on this specific case to ensure that both of the two approaches we present in this paper to model discrete bilevel problems with a robust follower can deal with uncertain lower-level constraints. Nevertheless, possible extensions are sketched in Section 5.

Without loss of generality, we further impose the following.

Assumption 4. *The deviations are non-negative, i.e.,* $\Delta d_i, \Delta w_i \ge 0$ *for all* $i \in [n_y]$.

To model the two types of situations, we consider an approach using an *extended formulation.* Additionally, we present an approach using a *multi-follower formulation* for the special case in which all lower-level variables are binary, i.e., $Y(x) \subseteq \{0,1\}^{n_y}$. However, the multi-follower-based approach can be extended naturally to allow for additional non-binary lower-level variables as long as the coefficients of the objective function or the constraints corresponding to the non-binary variables are not subject to uncertainty. We present a generic branch-and-cut framework to solve the value-function reformulation of the robustified bilevel problem. Moreover, we derive problem-tailored cuts for interdiction problems that can be used in the proposed branch-and-cut procedure. These cuts assume that Γ-robust follower sub-problems satisfy a downward monotonicity property, which arises in many packing-type applications. In this context, it is our aim to provide a natural extension of the results that have been proposed in Fischetti, Ljubić, et al. (2019) for the deterministic case. The main results of this paper are stated in Theorems 2–5 and form the core for the implementation of the proposed solution methods.

The remainder of the paper is organized as follows. In Section 2, we provide an extended formulation and a multi-follower formulation to model discrete linear bilevel problems with a Γ-robust lower-level problem. We present a generic branch-and-cut method to solve these problems. In Section 3, we focus on interdiction problems with a follower problem that satisfies a downward monotonicity property. In Section 4, we evaluate the effectiveness of the proposed approaches in a numerical study using the bilevel knapsack interdiction problem, which is a prominent example of an interdiction problem that satisfies the monotonicity property. Finally, we conclude in Section 5.

2. Generic Branch-and-Cut Frameworks

The aim of this section is to present generic branch-and-cut frameworks that can be used to solve the Γ-robustification of Problem (2) as stated around (4) and (6). The methods are similar to a procedure proposed by Wood (2011), which resembles (generalized) Benders decomposition (Benders 1962; Geoffrion 1972). To initialize the methods, we start by solving the problem in which the integrality constraints on the variables x as well as $\eta \geq \Phi(x)$ are omitted. This means that we first consider the linear problem

$$\begin{aligned} \min_{x,\eta} \quad & c^\top x + \eta \\ \text{s.t.} \quad & (x,\eta) \in \Omega_0 := \{(x,\eta) \in \mathbb{R}^{n_x} \times \mathbb{R} \colon Ax \geq a,\ x \in \mathring{X},\ \eta \geq \eta^-\}. \end{aligned} \tag{P$_0$}$$

Here, $\mathring{X}$ is a continuous relaxation of X, i.e., the integer points contained in $\mathring{X}$ coincide with X. Furthermore, we use an a priori lower bound $\eta^- \in \mathbb{R}$ on $\Phi(x)$ for all feasible leader's decisions x. In the following sections, we elaborate on how to obtain a valid lower bound in our setting. After considering Problem (P$_0$), we iteratively add valid inequalities or branch to cut off integer-infeasible points and also add valid inequalities to cut off bilevel infeasible points. Let

$$\begin{aligned} \min_{x,\eta} \quad & c^\top x + \eta \\ \text{s.t.} \quad & (x,\eta) \in \Omega_j \subseteq \mathbb{R}^{n_x} \times \mathbb{R} \end{aligned} \tag{P$_j$}$$

be the problem of node j of the branch-and-cut search tree. Here, the set Ω_j contains all valid inequalities that have been added previously to cut off integer-infeasible and bilevel-infeasible points as well as all branching decisions. If either Problem (P$_j$) is infeasible or, if the objective function value corresponding to an optimal solution (x^j, η^j) exceeds the current upper bound U, we can fathom node j. Otherwise, we do the following. First, we check if the upper-level variables x^j satisfy the integrality constraints, i.e., we check if $x^j \in X$ holds. If this is not the case, we separate a fractional solution by either exploiting standard cutting planes from mixed-integer linear optimization as elaborated in, e.g., Cornuéjols (2008), or by branching. Otherwise, we proceed by checking for bilevel feasibility, i.e., we check if $\eta^j \geq \Phi(x^j)$ is satisfied. For this purpose, we solve a reformulation of the robust counterpart of the lower-level problem that is parameterized by the current leader's decision $x^j \in X$. In the following sections, we will elaborate on how to obtain these reformulations. In particular, we present two approaches—an extended formulation and a multi-follower formulation—that are derived from Theorem 1 and 3, respectively, in Bertsimas and Sim (2003). Based on the latter two types of formulations, valid cuts to separate bilevel-infeasible points can be obtained. Nevertheless, the development of such cuts strongly depends on the specific problem considered at the lower level. Hence, the branch-and-cut frameworks presented in the remainder of this section remain fairly general and need to be adapted accordingly to capture the application problem at hand. We will show such adaptations for the bilevel knapsack interdiction problem in the following sections.

2.1. Extended Formulation.

One possibility to reformulate the robust counterparts (3) and (5) is to allow for an extended variable space of the follower that involves additional continuous variables.

Lemma 1. *For a feasible upper-level decision x, the robust counterpart of the lower-level problem (3) can be solved as the mixed-integer linear problem*

$$\begin{aligned} \Phi_d(x) = \max_{y,z,\theta} \quad & \sum_{i=1}^{n_y} d_i y_i - \Gamma_d \theta - \sum_{i=1}^{n_y} z_i && && \text{(7a)} \\ s.t. \quad & z_i + \theta \geq \Delta d_i y_i, && i \in [n_y], && \text{(7b)} \\ & (z,\theta) \geq 0, && && \text{(7c)} \\ & y \in Y(x). && && \text{(7d)} \end{aligned}$$

The extended formulation for the case of uncertainties in the lower-level constraint is stated in the following lemma.

Lemma 2. *For a feasible upper-level decision x, the robust counterpart of the lower-level problem (5) can be solved as the mixed-integer linear problem*

$$\begin{aligned}
\Phi_w(x) = \max_{y,z,\theta} \quad & \sum_{i=1}^{n_y} d_i y_i && \text{(8a)}\\
\text{s.t.} \quad & \sum_{i=1}^{n_y} w_i y_i + \Gamma_w \theta + \sum_{i=1}^{n_y} z_i \le C - \sum_{i=1}^{n_x} v_i x_i, && \text{(8b)}\\
& z_i + \theta \ge \Delta w_i y_i, \quad i \in [n_y], && \text{(8c)}\\
& (z,\theta) \ge 0, && \text{(8d)}\\
& y \in Y(x). && \text{(8e)}
\end{aligned}$$

Both lemmas can be shown in analogy to the proof of Theorem 1 in Bertsimas and Sim (2003). Before we comment on how the previous reformulations can be embedded in a branch-and-cut framework to separate bilevel-infeasible points, let us now provide valid lower bounds that can be used in Problem (P_0) to initialize the method.

Proposition 1. *Let x be a feasible upper-level decision. Then, by Assumption 2, there is a vector of finite upper bounds $u \in \mathbb{R}^{n_y}_+$ for the follower's variables such that $Y(x) \subseteq [0, u_1] \times \cdots \times [0, u_{n_y}]$ and such that*

$$\eta^- := \sum_{i=1}^{n_y} \min\{d_i, 0\} u_i - \Gamma_d \max_{i \in [n_y]} \{\Delta d_i u_i\} - \sum_{i=1}^{n_y} \Delta d_i u_i$$

is a valid lower bound for the optimal objective function value of the x-parameterized problem (7).

All the proofs that we omit here can be found in Appendix A. In the following, we provide a lower bound for the variant of the problem with uncertainties regarding the follower's packing-type constraint.

Proposition 2. *Let x be a feasible upper-level decision. Then, by Assumption 2, there is a vector of finite upper bounds $u \in \mathbb{R}^{n_y}_+$ for the follower's variables such that $Y(x) \subseteq [0, u_1] \times \cdots \times [0, u_{n_y}]$ and such that*

$$\eta^- := \sum_{i=1}^{n_y} \min\{d_i, 0\} u_i$$

is a valid lower bound for the optimal objective function value of the x-parameterized problem (8).

The method to process node j of the branch-and-cut search tree that exploits an extended formulation (as stated in the last two lemmas) is formally stated in Algorithm 1. To determine $\Phi(x^j)$ for the current leader's decision x^j in Step 9, we need to solve the x^j-parameterized robust lower-level problem. Depending on the considered uncertainty model, this can either be Problem (7) or Problem (8) for which we set $\Phi(x^j) = \Phi_d(x^j)$ or $\Phi(x^j) = \Phi_w(x^j)$ accordingly. In Step 11, we generate a valid cut to exclude the bilevel-infeasible point (x^j, η^j). To this end, one can use generic cuts like (generalized) no-good cuts; see, e.g., Tahernejad and Ralphs (2020).

2.2. **Multi-Follower Approach.**

An alternative reformulation of the robust counterparts (3) and (5) can be obtained under the following additional assumptions.

Assumption 5. *All lower-level variables y are binary, i.e., $Y(x) \subseteq \{0,1\}^{n_y}$.*

Assumption 6. *The indices are ordered such that the deviations are given in non-increasing order, i.e., $\Delta d_i \ge \Delta d_{i+1}$ or $\Delta w_i \ge \Delta w_{i+1}$ for all $i \in [n_y]$ with $\Delta d_{n_y+1} = 0$ and $\Delta w_{n_y+1} = 0$.*

Algorithm 1 Processing node j using the extended formulation

```
1:  Solve Problem (P_j).
2:  if Problem (P_j) is infeasible then
3:      Fathom the current node.
4:  Let $(x^j, \eta^j)$ denote the solution of Problem (P_j).
5:  if $c^\top x^j + \eta^j \geq U$ then
6:      Fathom the current node.
7:  if $x^j \notin X$ then
8:      Either generate cuts valid for $\Omega_j \cap (X \times \mathbb{R})$, augment $\Omega_j$, and go to Step 1 or branch.
9:  Determine $\Phi(x^j)$ and set $U \leftarrow \min\{U, c^\top x^j + \Phi(x^j)\}$.
10: if $\eta^j < \Phi(x^j)$ then
11:     Generate a valid cut that excludes $(x^j, \eta^j)$ from $\Omega_j$, augment $\Omega_j$, and go to Step 1.
```

Note that Assumption 6 is w.l.o.g. as long as we do not have uncertainties in both the objective function coefficients and the constraint coefficients on the lower level. Assumptions 5 and 6 are necessary to exploit Theorem 3 in Bertsimas and Sim (2003), which is what we do in the following.

Lemma 3. *Let x be a feasible upper-level decision. Under Assumptions 5 and 6, solving the robust counterpart of the lower-level problem* (3) *is equivalent to solving $n_y + 1$ problems of the nominal type, i.e.,*

$$\Phi_d(x) = \max_{\ell \in \{1,\dots,n_y+1\}} \Phi_d^\ell(x) \tag{9}$$

holds, where for all $\ell \in \{1, \dots, n_y + 1\}$, we have

$$\Phi_d^\ell(x) = -\Gamma_d \Delta d_\ell + \max_{y \in Y(x)} \left\{ \tilde{d}(\ell)^\top y \right\} \tag{10}$$

with

$$\tilde{d}(\ell)_i = \begin{cases} d_i - (\Delta d_i - \Delta d_\ell), & 1 \leq i \leq \ell, \\ d_i, & \ell + 1 \leq i \leq n_y. \end{cases}$$

Note that the lower-level optimal-value function (9) is defined as the maximum of $n_y + 1$ value functions. Thus, the robustification of Problem (1) can be interpreted as a single-leader-multi-follower problem with $n_y + 1$ many followers, which is why we refer to (9) as *multi-follower formulation*. Moreover, we refer to (10) as the ℓth *follower sub-problem* throughout this paper.

In the following, we provide a valid lower bound for the optimal objective function value of (9) that can be used in Problem (P$_0$).

Proposition 3. *Let x be a feasible upper-level decision and let $\ell \in \{1, \dots, n_y+1\}$ be arbitrary but fixed. Under Assumption 5, a valid lower bound for the optimal objective function value of the x-parameterized problem* (9) *is given by*

$$\eta^- := -\Gamma_d \Delta d_\ell + \sum_{i=1}^{n_y} \min \left\{ \tilde{d}(\ell)_i, 0 \right\}.$$

The multi-follower reformulation for the case of uncertainties in the lower-level constraint is stated in the following lemma.

Lemma 4. *Let x be a feasible upper-level decision. Under Assumptions 5 and 6, solving the robust counterpart of the lower-level problem* (5) *is equivalent to solving $n_y + 1$ problems of the nominal type, i.e.,*

$$\Phi_w(x) = \max_{\ell \in \{1,\dots,n_y+1\}} \Phi_w^\ell(x)$$

holds, where for all $\ell \in \{1, \dots, n_y + 1\}$*, we have*

$$\Phi_w^\ell(x) = \max_{y \in Y(x)} \left\{ \sum_{i=1}^{n_y} d_i y_i : \Gamma_w \Delta w_\ell + \sum_{i=1}^{n_y} \tilde{w}(\ell)_i y_i + \sum_{i=1}^{n_x} v_i x_i \leq C \right\} \tag{11}$$

with

$$\tilde{w}(\ell)_i = \begin{cases} w_i + (\Delta w_i - \Delta w_\ell), & 1 \leq i \leq \ell, \\ w_i, & \ell + 1 \leq i \leq n_y. \end{cases}$$

Lemmas 3 and 4 can be shown in analogy to the proof of Theorem 3 in Bertsimas and Sim (2003).

Note that, in the case of uncertainties regarding the follower's inequality constraint, we have $\Delta d_\ell = 0$ such that $\tilde{d}(\ell) = d$ holds for all $\ell \in \{1, \dots, n_y + 1\}$. Using the latter, Proposition 3 also provides a valid lower bound for the setting considered in Lemma 4. Let us further note that we consider the same deterministic lower-level objective function in the extended formulation as well as in the multi-follower formulation. Moreover, we would like to point out that the fact that there are only *binary* variables corresponding to uncertain coefficients on the lower level is a crucial point for the validity of the multi-follower formulation.

In what follows, we omit the subscripts d and w that are used to denote the considered uncertainty modeling for notational convenience. Further, we will hold on to an improvement of the previous results that has been established in Álvarez-Miranda, Ljubić, et al. (2013) by reducing the number of nominal problems to be considered to $n_y - \Gamma + 2$, i.e.,

$$\Phi(x) = \max_{\ell \in \{\Gamma, \dots, n_y + 1\}} \Phi^\ell(x). \tag{12}$$

Note that a further reduction of the number of nominal problems to be solved has been published by Lee and Kwon (2014). Since this is, however, not the core focus of this paper, we stay with the version as published by Álvarez-Miranda, Ljubić, et al. (2013).

The method to process node j that exploits the multi-follower formulation for the robust counterpart of the lower-level problem is formally stated in Algorithm 2. In contrast to the approach using the extended formulation, in which a single cut is added at each node of the branch-and-cut search tree in case of bilevel-infeasibility, a cut for each follower sub-problem $\ell \in \{\Gamma, \dots, n_y + 1\}$ that satisfies $\eta^j < \Phi^\ell(x^j)$ is added in Step 12 of Algorithm 2. This means that up to $n_y - \Gamma + 2$ cuts could be added at each node. However, it would also be valid to consider, e.g., adding only the most violated cut for the given leader's decision x^j. We will address this aspect in detail when we discuss various cut separation strategies in Section 4.

Algorithm 2 Processing node j using the multi-follower approach

```
1: Solve Problem (P_j).
2: if Problem (P_j) is infeasible then
3:     Fathom the current node.
4: Let $(x^j, \eta^j)$ denote the solution of Problem (P_j).
5: if $c^\top x^j + \eta^j \geq U$ then
6:     Fathom the current node.
7: if $x^j \notin X$ then
8:     Either generate cuts valid for $\Omega_j \cap (X \times \mathbb{R})$, augment $\Omega_j$, and go to Step 1 or branch.
9: for all $\ell \in \{\Gamma, \dots, n_y + 1\}$ do
10:    Solve the $\ell$th lower-level sub-problem to obtain $\Phi^\ell(x^j)$.
11:    if $\eta^j < \Phi^\ell(x^j)$ then
12:        Generate a valid cut that excludes $(x^j, \eta^j)$ from $\Omega_j$, augment $\Omega_j$.
13: Set $\Phi(x^j) \leftarrow \max_{\ell \in \{\Gamma, \dots, n_y + 1\}} \Phi^\ell(x^j)$ and $U \leftarrow \min\{U, c^\top x^j + \Phi(x^j)\}$.
14: If at least one cut was added in Step 12, then go to Step 1.
```

Theorem 1. *If we embed either Algorithm 1 or 2 into a usual branch-and-bound framework, we obtain a correct method that terminates with an optimal solution* (x^*, η^*) *after a finite number of nodes and after adding an overall finite number of cuts.*

Proof. We first recall that all linking variables are bounded integers. We then observe that, w.l.o.g., non-linking variables can be moved to the lower level; see also Bolusani et al. (2020). Hence, the finite termination is due to the finiteness of the number of feasible upper-level decisions, the finiteness of branch-and-cut methods to solve the lower-level sub-problems, and from the fact that a leader's decision cannot be selected twice. If (x, η) is a non-optimal leader's decision, i.e., $\eta < \Phi(x)$, adding a globally valid inequality excludes (x, η) from being feasible for all subsequent considerations. If an upper-level decision were ever to occur again, i.e., if there would exist solutions $(x^k, \eta^k) = (x^j, \eta^j)$ of Problem (P_j) with $j < k$, then $\eta^j = \eta^k \geq \Phi(x^j)$ would have to hold and the termination criterion would be satisfied. Thus, an optimal solution cannot be overlooked. In particular, the number of cuts possibly added to the problem formulation is finite and in $O(2^{n_x})$. □

Note that the sub-problems that are solved in Step 10 of Algorithm 2 are independent. This means that the objective function and the constraints of a sub-problem only include the upper-level decision x and the lower-level variables corresponding to the ℓth sub-problem with $\ell \in \{\Gamma, \dots, n_y + 1\}$. Thus, the sub-problems can be solved in parallel. Further note that we have not specified how the cuts that are added in Algorithm 1 and 2 are generated. For an overview of various cutting planes that can be used for general classes of mixed-integer linear bilevel problems, we refer to Tahernejad and Ralphs (2020). Nevertheless, stronger formulations can be obtained for certain problems; see, e.g., Fischetti, Ljubić, et al. (2019) and Furini, Ljubić, Segundo, et al. (2021). Thus, it is often essential to exploit specific properties of the application problem at hand to derive valid cuts, which is what we do in the remainder of the paper.

3. Interdiction Cuts for Monotone Γ-Robust Followers

In the following, we focus on interdiction problems with a follower problem that satisfies a downward monotonicity property as in Fischetti, Ljubić, et al. (2019). We refer to this type of problem as an interdiction problem with a monotone follower. In this setting, both players share a common set of items indexed by $i \in [n]$. The leader has the ability to influence the follower's decision by prohibiting the usage of certain items by the follower. This is established by either setting the leader's variable $x_i = 1$ to interdict item $i \in [n]$ for the follower or $x_i = 0$, otherwise. For the ease of presentation, we assume the following.

Assumption 7. *The number of variables on the upper and the lower level is the same, i.e.,* $n_x = n_y = n$ *holds.*

In particular, this means that all variables of the leader need to be binary, i.e., $x \in X = \{0, 1\}^n$. However, the following results can as well be adapted to account for non-interdicting (and thus possibly non-binary) variables of the leader. The case in which the lower-level problem also includes variables that are not subject to interdiction can be handled by partitioning the follower's variable set into interdicted and non-interdicted variables as it is done in Fischetti, Ljubić, et al. (2019).

Assumption 8. *For a feasible upper-level decision* x*, the* x*-parameterized lower-level feasible set is of the form*

$$Y(x) = \{y \in \mathcal{Y} \colon By \leq b,\ y_i \leq u_i(1 - x_i),\ i \in [n]\} \tag{13}$$

with $\mathcal{Y} \subseteq \mathbb{Z}^n_+$, $B \in \mathbb{R}^{m \times n}_+$, $b \in \mathbb{R}^m_+$, *and a vector of finite upper bounds* $u \in \mathbb{R}^n_+$.

Assumption 8 ensures that the nominal lower-level problem (1d) satisfies a downward monotonicity property, which we formally define in the following. Further note that the leader's variables x are linked to the lower-level problem only via the interdiction constraints $y_i \leq u_i(1-x_i)$. Both aspects are essential for the validity of the cuts we propose in the remainder of this section.

Proposition 4 (Monotonicity Property)**.** *Let x be a feasible decision of the leader. Further, let $y \in Y(x)$ and let $y' \in \mathcal{Y}$ be such that $y' \leq y$ holds. Then, y' is a feasible follower's decision for the given leader's decision x, i.e., $y' \in Y(x)$.*

It is noteworthy that, under Assumption 8, all items with non-positive objective function coefficients are not chosen by the follower. Consequently, the leader does not need to spend interdiction resources on these items and we could thus omit all items with non-positive objective function coefficients in the problem formulation. This leads us to, w.l.o.g., making the following assumption.

Assumption 9. *All lower-level objective function coefficients are positive, i.e., $d_i > 0$ for all $i \in [n]$.*

Finally, we will hold on to the following, which is reasonable in the interdiction setting.

Assumption 10. *There are no terms depending on the leader's decision in the upper-level objective function, i.e., $c = 0$.*

In the following sections, we show that the downward monotonicity property remains satisfied when Γ-robust followers are considered, which we exploit to derive valid cuts. In Section 3.1, we focus on the case of uncertainties in the follower's objective function coefficients. We derive two variants of *interdiction cuts* based on the two approaches discussed in Section 2. We devote Section 3.2 to strengthened formulations for the proposed interdiction cuts. Finally, the case of uncertainties in the lower-level constraint is addressed in Section 3.3.

3.1. Problems with Uncertain Objective Function Coefficients.

According to the notation considered in the previous sections, we assume that all lower-level objective function coefficients may be subject to uncertainty and that the follower hedges against at most Γ_d deviations in the objective function coefficients. The robust counterpart of the lower-level problem and of the overall bilevel problem is given in (3) and (4), respectively. The corresponding extended formulation and the multi-follower formulation have already been stated in Section 2. For the extended formulation, we only need to replace (7d) with (13), i.e., we consider the problem

$$\begin{aligned}
\Phi(x) = \max_{y,z,\theta} \quad & \sum_{i=1}^{n} d_i y_i - \Gamma_d \theta - \sum_{i=1}^{n} z_i && && \text{(14a)}\\
\text{s.t.} \quad & z_i + \theta \geq \Delta d_i y_i, && i \in [n], && \text{(14b)}\\
& By \leq b, && && \text{(14c)}\\
& y_i \leq u_i(1-x_i), && i \in [n], && \text{(14d)}\\
& (z,\theta) \geq 0,\ y \in \mathcal{Y}. && && \text{(14e)}
\end{aligned}$$

For the multi-follower approach, we need to replace the feasible set in (10) with (13). Due to Assumption 5, $u_i = 1$ is a trivially valid upper bound for all $i \in [n]$. For $\ell \in \{\Gamma_d, \dots, n_y + 1\}$, we thus consider the sub-problem

$$\Phi^{\ell}(x) = -\Gamma_d \Delta d_\ell + \max_{y} \left\{ \tilde{d}(\ell)^\top y \colon By \leq b,\ y_i \leq 1 - x_i,\ y_i \in \{0,1\},\ i \in [n] \right\}. \quad (15)$$

Proposition 5. *Let x be a feasible decision of the leader. Then, the x-parameterized problem (15) satisfies the downward monotonicity property. Moreover, let (y,z,θ) be feasible for the x-parameterized problem (14), and let $y' \in \mathcal{Y}$ be such that $y' \leq y$ holds. Then, (y',z,θ) is feasible for the x-parameterized problem (14) as well.*

Due to Proposition 5 and for the ease of presentation, we also say that the extended formulation (14) satisfies the downward monotonicity property. In what follows, we exploit the previous result to introduce penalized formulations for Problems (14) and (15) that are used to derive valid interdiction cuts. To this end, we omit the interdiction constraints $y_i \leq u_i(1-x_i)$ in the problem formulation and instead add the penalty terms $-d_i y_i x_i$ to the objective function for all $i \in [n]$.

Proposition 6. *Let x be a feasible upper-level decision. Then, Problem (14) and the mixed-integer linear problem*

$$\begin{aligned}
\Phi(x) = \max_{y,z,\theta} \quad & \sum_{i=1}^{n} d_i y_i (1-x_i) - \Gamma_d \theta - \sum_{i=1}^{n} z_i && && (16a)\\
\text{s.t.} \quad & z_i + \theta \geq \Delta d_i y_i, && i \in [n], && (16b)\\
& By \leq b, && && (16c)\\
& y_i \leq u_i, && i \in [n], && (16d)\\
& (z,\theta) \geq 0,\ y \in \mathcal{Y} && && (16e)
\end{aligned}$$

admit the same optimal value.

For the multi-follower-based approach, we obtain the following similar result.

Proposition 7. *Let x be a feasible upper-level decision and let $\ell \in \{\Gamma_d, \dots, n+1\}$ be arbitrary but fixed. Then, Problem (15) and the problem*

$$\Phi^{\ell}(x) = -\Gamma_d \Delta d_\ell + \max_{y \in Y} \left\{ \sum_{i=1}^{n} \tilde{d}(\ell)_i y_i (1-x_i) \right\} \tag{17}$$

with

$$Y = \{ y \in \{0,1\}^n : By \leq b \}$$

admit the same optimal value.

The feasible set of Problem (16) and of each sub-problem (17) is independent from the leader's decision. Moreover, the objective functions are linear for fixed x. Thus, an optimal solution is attained at a vertex of the convex hull of the respective feasible set. We set

$$\Psi = \{ (y,z,\theta) \in \mathcal{Y} \times \mathbb{R}^n \times \mathbb{R} : (y,z,\theta) \text{ satisfy } (16b)\text{–}(16e) \}.$$

In what follows, we use $\hat{\Psi}$ and $\hat{Y}$ to denote the set containing the finite number of vertices of the convex hull of Ψ and Y, respectively. Further, let (x,η) be feasible for Problem (2) with the lower-level optimal-value function (3). Due to Proposition 6, we have

$$\eta \geq \Phi(x) \geq \sum_{i=1}^{n} d_i \hat{y}_i (1-x_i) - \Gamma_d \hat{\theta} - \sum_{i=1}^{n} \hat{z}_i$$

for arbitrary but fixed $(\hat{y}, \hat{z}, \hat{\theta}) \in \hat{\Psi}$. Consequently, the interdiction cuts

$$\eta \geq \sum_{i=1}^{n} d_i \hat{y}_i (1-x_i) - \Gamma_d \hat{\theta} - \sum_{i=1}^{n} \hat{z}_i \quad \text{for all } \left(\hat{y}, \hat{z}, \hat{\theta}\right) \in \hat{\Psi} \tag{18}$$

are valid for Problem (2). To derive valid cuts for the multi-follower-based approach, we do the following. Under Assumptions 5 and 6, we have

$$\begin{aligned}
\eta \geq \Phi(x) &= \max_{\ell \in \{\Gamma_d, \dots, n+1\}} \Phi^{\ell}(x)\\
&\geq \Phi^{\ell}(x) = -\Gamma_d \Delta d_\ell + \max_{y \in \hat{Y}} \left\{ \sum_{i=1}^{n} \tilde{d}(\ell)_i y_i (1-x_i) \right\}\\
&\geq -\Gamma_d \Delta d_\ell + \sum_{i=1}^{n} \tilde{d}(\ell)_i \hat{y}_i (1-x_i)
\end{aligned}$$

for arbitrary but fixed $\hat{y} \in \hat{Y}$ and for all $\ell \in \{\Gamma_d, \dots, n+1\}$. The first equality follows from (12) and the second one holds due to Proposition 7. As a result, the cuts

$$\eta \geq -\Gamma_d \Delta d_\ell + \sum_{i=1}^{n} \tilde{d}(\ell)_i y_i (1 - x_i) \quad \text{for all } \hat{y} \in \hat{Y},\ \ell \in \{\Gamma_d, \dots, n+1\} \tag{19}$$

are valid for Problem (2). In particular, we may obtain different valid cuts for each $\ell \in \{\Gamma_d, \dots, n+1\}$. Finally, we exploit the previous results to equivalently reformulate the interdiction problem with a Γ_d-robust follower facing uncertain objective function coefficients.

Theorem 2. *Problem (2) with the lower-level optimal-value function (3) can be equivalently reformulated by replacing Constraint (2c) with (18). Under Assumptions 5 and 6, an equivalent reformulation can be obtained by replacing Constraint (2c) with (19).*

3.2. Cut Strengthening and Enhanced Formulations.

In this section, we provide enhancements and techniques to strengthen the cuts proposed in the previous section. First, we introduce the notion of maximal packings.

Definition 1. *A follower's decision $(\hat{y}, \hat{z}, \hat{\theta}) \in \hat{\Psi}$ is a* maximal packing *w.r.t. the extended formulation (16) if there is no $(y', \hat{z}, \hat{\theta}) \in \hat{\Psi} \setminus \{\hat{y}\}$ such that $\hat{y} \leq y'$ holds.*

We exploit this notion to avoid the generation of unnecessary cuts.

Proposition 8. *Let $(\hat{y}, \hat{z}, \hat{\theta}) \in \hat{\Psi}$ be a non-maximal packing for Problem (16) and let $(y', \hat{z}, \hat{\theta}) \in \hat{\Psi} \setminus \{\hat{y}\}$ be chosen such that $\hat{y} \leq y'$ holds. Then, the interdiction cut (18) associated with $(\hat{y}, \hat{z}, \hat{\theta})$ is dominated by the interdiction cut associated with $(y', \hat{z}, \hat{\theta})$.*

Here and in what follows, *domination* between two cuts is understood in the sense that the feasible set induced by the dominating cut is contained in the feasible set induced by the dominated one. Due to the previous result, it is sufficient to only consider the interdiction cuts that correspond to maximal packings of the follower.

Definition 2. *A follower's decision $\hat{y} \in \hat{Y}$ is a* maximal packing *w.r.t. the multi-follower formulation (17) if there is no $y' \in \hat{Y} \setminus \{\hat{y}\}$ such that $\hat{y} \leq y'$ holds.*

Note that there is no need to specify $\ell \in \{\Gamma_d, \dots, n+1\}$ in the previous definition since, in each sub-problem, we consider the same set $\hat{Y}$, which is independent from ℓ. To obtain a dominance result for interdiction cuts associated with maximal packings in the multi-follower setting, we need to further study the properties of the modified objective function coefficients $\tilde{d}(\ell)$ for each $\ell \in \{\Gamma_d, \dots, n+1\}$. Note that the modified objective function coefficients can be non-positive for certain items in some follower sub-problems. If $\tilde{d}(\ell)_i \leq 0$ holds for all $\ell \in \{\Gamma_d, \dots, n+1\}$, the ith item will not be chosen in any of the follower sub-problems. Thus, the leader does not need to spend interdiction resources on the ith item and $x_i = y_i = 0$ can be fixed. This is equivalent to completely omitting the ith item in the problem formulation. However, if there is an item $i \in [n]$ with non-positive modified objective function coefficients only for some follower sub-problems, i.e., $\tilde{d}(k)_i \leq 0$ for all $k \in \mathcal{S} \subset \{\Gamma_d, \dots, n+1\}$ and $\tilde{d}(l)_i > 0$ for all $l \in \{\Gamma_d, \dots, n+1\} \setminus \mathcal{S}$, the ith item might be part of an optimal solution. Therefore, we introduce the following notation. For each $\ell \in \{\Gamma_d, \dots, n+1\}$, we define the set

$$D_+^\ell := \{i \in [n] \colon \tilde{d}(\ell)_i > 0\}.$$

Proposition 9. *The interdiction cuts (19) can be replaced with*

$$\eta \geq -\Gamma_d \Delta d_\ell + \sum_{i \in D_+^\ell} \tilde{d}(\ell)_i \hat{y}_i (1 - x_i) \quad \textit{for all } \hat{y} \in \hat{Y},\ \ell \in \{\Gamma_d, \dots, n+1\}. \tag{20}$$

In particular, the cuts (20) dominate the basic interdiction cuts (19).

With the previous considerations, we can finally state a dominance result for interdiction cuts associated with maximal packings in the multi-follower setting.

Proposition 10. *Let $\hat{y} \in \hat{Y}$ be a non-maximal packing for Problem (17) and let $y' \in \hat{Y} \setminus \{\hat{y}\}$ be such that $\hat{y} \leq y'$ holds. Then, the interdiction cuts (20) associated with $\hat{y}$ are dominated by the interdiction cuts associated with y'.*

Note that the previous result is not valid for the basic interdiction cuts as stated in (19) since, in general, $\tilde{d}(\ell)_i \hat{y}_i (1 - x_i) \leq \tilde{d}(\ell)_i y'_i (1 - x_i)$ does not hold for all $i \notin D^\ell_+$. Moreover, we would like to mention that we also considered maximal packings for the leader. However, preliminary computational tests revealed that this does not improve the performance of the overall solution method. Thus, we decided to refrain from using this ingredient in our computational study in Section 4. However, we exploit dominance properties among items to obtain further enhancements. First, we introduce additional inequalities regarding the leader's decision x. In what follows, $A_{\cdot i}$ denotes the ith column of A.

Theorem 3. *For all $s, t \in [n]$, $s \neq t$, chosen such that $A_{\cdot s} > A_{\cdot t}$, $B_{\cdot s} \leq B_{\cdot t}$, $u_s \geq u_t$, $d_s \geq d_t$, and $d_s - \Delta d_s \geq d_t - \Delta d_t$, the dominance inequality*

$$x_t \leq x_s \tag{21}$$

is satisfied in at least one optimal solution of Problem (2) with the lower-level optimal-value function (3).

Proof. Let (x^*, η^*) be an optimal solution of Problem (2) with the lower-level optimal-value function (3). Without loss of generality, suppose that there are exactly two distinct items s, $t \in [n]$ that satisfy the requirements of the theorem but for which the dominance inequality (21) does not hold, i.e., $x^*_s = 0$ and $x^*_t = 1$. Otherwise, we repeat the following procedure as long as there are still items left that satisfy the requirements but violate the corresponding dominance inequality. The idea is to construct an optimal leader's decision that satisfies the dominance inequality. To this end, we set

$$x'_i = \begin{cases} x^*_i, & i \in [n] \setminus \{s, t\}, \\ 1, & i = s, \\ 0, & i = t. \end{cases}$$

By construction, x' is feasible for Problem (2) with the lower-level optimal-value function (3) and satisfies the dominance inequality associated with s and t. Without loss of generality, we show that x' is also an optimal solution of Problem (2) using the extended formulation. Let (y', z', θ') be an optimal solution of Problem (14) for x', i.e., we have $y'_s = 0$. Moreover, $z'_i = \max\{\Delta d_i y'_i - \theta', 0\}$ holds for all $i \in [n]$ due to Constraints (14b), (14e), and the objective function. If $y'_t = 0$ holds, (y', z', θ') is also a feasible follower's decision for x^* and we obtain

$$\Phi(x') = \sum_{i=1}^{n} d_i y'_i - \Gamma_d \theta' - \sum_{i=1}^{n} z'_i \leq \Phi(x^*).$$

If $y'_t \geq 1$ holds, we consider the alternative follower's decision $(\hat{y}, \hat{z}, \theta')$ with

$$\hat{y}_i = \begin{cases} y'_i, & i \in [n] \setminus \{s, t\}, \\ y'_t, & i = s, \\ 0, & i = t, \end{cases}$$

and

$$\hat{z}_i = \begin{cases} z'_i, & i \in [n] \setminus \{s, t\}, \\ \max\{\Delta d_s y'_t - \theta', 0\}, & i = s, \\ 0, & i = t. \end{cases}$$

By construction, $(\hat{y}, \hat{z}, \theta')$ is feasible for Problem (14) given the leader's decision x^* and we obtain

$$\begin{aligned}
\Phi(x') &= \sum_{i=1}^{n} d_i y_i' - \Gamma_d \theta' - \sum_{i=1}^{n} z_i' \\
&= d_s \underbrace{y_s'}_{=0} + d_t y_t' + \sum_{i\in[n]\setminus\{s,t\}} d_i \underbrace{y_i'}_{=\hat{y}_i} - \Gamma_d \theta' - \underbrace{z_s'}_{=0} - z_t' - \sum_{i\in[n]\setminus\{s,t\}} \underbrace{z_i'}_{=\hat{z}_i} \\
&= d_t y_t' + \sum_{i\in[n]\setminus\{s,t\}} d_i \hat{y}_i - \Gamma_d \theta' - \max\{\Delta d_t y_t' - \theta', 0\} - \sum_{i\in[n]\setminus\{s,t\}} \hat{z}_i \\
&= \sum_{i\in[n]\setminus\{s,t\}} d_i \hat{y}_i - \Gamma_d \theta' + \min\{(d_t - \Delta d_t) y_t' + \theta', d_t y_t'\} - \sum_{i\in[n]\setminus\{s,t\}} \hat{z}_i \\
&\le \sum_{i\in[n]\setminus\{s,t\}} d_i \hat{y}_i - \Gamma_d \theta' + \min\{(d_s - \Delta d_s) \hat{y}_s + \theta', d_s \hat{y}_s\} - \sum_{i\in[n]\setminus\{s,t\}} \hat{z}_i \\
&= d_s \hat{y}_s + d_t \underbrace{\hat{y}_t}_{=0} + \sum_{i\in[n]\setminus\{s,t\}} d_i \hat{y}_i - \Gamma_d \theta - \max\{\Delta d_s \hat{y}_s - \theta', 0\} - \underbrace{\hat{z}_t}_{=0} - \sum_{i\in[n]\setminus\{s,t\}} \hat{z}_i \\
&= \sum_{i=1}^{n} d_i \hat{y}_i - \Gamma_d \theta' - \sum_{i=1}^{n} \hat{z}_i \\
&\le \Phi(x^*).
\end{aligned}$$

Hence, the alternative leader's decision x' is optimal for Problem (2) with the lower-level optimal-value function (3) and satisfies the dominance inequality associated with s and t. This concludes the proof. □

Note that, in the case of only binary follower's variables, the requirement $u_s \geq u_t$ in the previous theorem is trivially satisfied since $u_i = 1$ for all $i \in [n]$ is a valid upper bound. Further note that the requirement $A_{\cdot s} > A_{\cdot t}$ can be relaxed to $A_{\cdot s} \geq A_{\cdot t}$ if the matrix A has only non-negative entries.

In the remainder of this section, we provide lifted cuts that dominate their respective basic counterparts stated in (18) and (20). We start by lifting the basic interdiction cuts corresponding to the extended formulation.

Theorem 4. *For any arbitrary but fixed $(\hat{y}, \hat{z}, \hat{\theta}) \in \hat{\Psi}$, let $K \in [n]$, $S_a = \{a_1, \dots, a_K\} \subset [n]$, and $S_b = \{b_1, \dots, b_K\} \subset [n]$ be such that $S_a \cap S_b = \emptyset$, $\hat{y}_{a_k} \geq 1$, $\hat{y}_{b_k} = 0$, $B_{\cdot a_k} \geq B_{\cdot b_k}$, $u_{a_k} \leq u_{b_k}$, $\Delta d_{a_k} < \Delta d_{b_k}$, and $d_{a_k} - \Delta d_{a_k} < d_{b_k} - \Delta d_{b_k}$ for all $k \in [K]$. Then, the following lifted interdiction cut is valid for Problem (2) with the lower-level optimal-value function (3):*

$$\eta \geq \sum_{i=1}^{n} d_i \hat{y}_i (1 - x_i) + \sum_{k=1}^{K} \left((d_{b_k} - \Delta d_{b_k}) - (d_{a_k} - \Delta d_{a_k}) \right) \hat{y}_{a_k} (1 - x_{b_k}) - \Gamma_d \hat{\theta} - \sum_{i=1}^{n} \hat{z}_i. \quad (22)$$

Proof. Let (x, η) be a feasible leader's decision for Problem (2) with the lower-level optimal-value function (3). If $x_{b_k} = 1$ holds for all $k \in [K]$, we obtain the basic interdiction cut as stated in (18), which is satisfied by x. Otherwise, we define $\mathcal{K} := \{k \in [K] \colon x_{b_k} = 0\}$ and consider an alternative follower's decision $(y', z', \hat{\theta})$, which is obtained as follows. For all $k \in \mathcal{K}$, we set $y_{a_k}' = 0$, $y_{b_k}' = \hat{y}_{a_k}$ as well as $z_{a_k}' = 0$ and $z_{b_k}' = \max\{\Delta d_{b_k} \hat{y}_{a_k} - \hat{\theta}, 0\}$. We further define $\bar{\mathcal{K}} := [n] \setminus \{a_k, b_k \colon k \in \mathcal{K}\}$ and set $y_i' = \hat{y}_i$ as well as $z_i' = \hat{z}_i$ for all $i \in \bar{\mathcal{K}}$. By construction, we have $(y', z', \hat{\theta}) \in \Psi$. Since the interdiction cuts (18) are not only valid for vertices contained in the set $\hat{\Psi}$ but also for any lower-level feasible point, the leader's decision x also satisfies the basic interdiction cut associated with $(y', z', \hat{\theta})$, i.e.,

$$\begin{aligned}
\eta &\geq \sum_{i=1}^{n} d_i y'_i(1-x_i) - \Gamma_d\hat{\theta} - \sum_{i=1}^{n} z'_i \\
&= \sum_{i\in\bar{\mathcal{K}}} d_i \underbrace{y'_i}_{=\hat{y}_i}(1-x_i) + \sum_{k\in\mathcal{K}}(d_{a_k}\underbrace{y'_{a_k}}_{=0}(1-x_{a_k}) + d_{b_k}\underbrace{y'_{b_k}}_{=\hat{y}_{a_k}}\underbrace{(1-x_{b_k})}_{=1}) \\
&\quad - \Gamma_d\hat{\theta} - \sum_{i\in\bar{\mathcal{K}}}\underbrace{z'_i}_{=\hat{z}_i} - \sum_{k\in\mathcal{K}}(\underbrace{z'_{a_k}}_{=0} + z'_{b_k}) \\
&= \sum_{i\in\bar{\mathcal{K}}} d_i\hat{y}_i(1-x_i) - \Gamma_d\hat{\theta} - \sum_{i\in\bar{\mathcal{K}}}\hat{z}_i + \sum_{k\in\mathcal{K}}\left(d_{b_k}\hat{y}_{a_k} - \max\{\Delta d_{b_k}\hat{y}_{a_k} - \hat{\theta}, 0\}\right).
\end{aligned} \tag{23}$$

In particular, we have $\hat{z}_{a_k} = \max\{\Delta d_{a_k}\hat{y}_{a_k} - \hat{\theta}, 0\}$ and $\hat{z}_{b_k} = 0$ for all $k \in [K]$. Hence, we can re-write Inequality (22) as

$$\begin{aligned}
\eta &\geq \sum_{i=1}^{n} d_i\hat{y}_i(1-x_i) + \sum_{k=1}^{K}\left((d_{b_k} - \Delta d_{b_k}) - (d_{a_k} - \Delta d_{a_k})\right)\hat{y}_{a_k}(1-x_{b_k}) - \Gamma_d\hat{\theta} - \sum_{i=1}^{n}\hat{z}_i \\
&= \sum_{i\in\bar{\mathcal{K}}} d_i\hat{y}_i(1-x_i) + \sum_{k\in\mathcal{K}}(d_{a_k}\hat{y}_{a_k}(1-x_{a_k}) + d_{b_k}\underbrace{\hat{y}_{b_k}}_{=0}(1-x_{b_k})) \\
&\quad + \sum_{k\in\mathcal{K}}((d_{b_k} - \Delta d_{b_k}) - (d_{a_k} - \Delta d_{a_k}))\hat{y}_{a_k}(1-\underbrace{x_{b_k}}_{=0}) \\
&\quad + \sum_{k\in[K]\setminus\mathcal{K}}((d_{b_k} - \Delta d_{b_k}) - (d_{a_k} - \Delta d_{a_k}))\hat{y}_{a_k}(1-\underbrace{x_{b_k}}_{=1}) - \Gamma_d\hat{\theta} \\
&\quad - \sum_{i\in\bar{\mathcal{K}}}\hat{z}_i - \sum_{k\in\mathcal{K}}(\hat{z}_{a_k} + \underbrace{\hat{z}_{b_k}}_{=0}) \\
&= \sum_{i\in\bar{\mathcal{K}}} d_i\hat{y}_i(1-x_i) - \Gamma_d\hat{\theta} - \sum_{i\in\bar{\mathcal{K}}}\hat{z}_i + \sum_{k\in\mathcal{K}}\Big(d_{a_k}\hat{y}_{a_k}(1-x_{a_k}) \\
&\qquad + \left((d_{b_k} - \Delta d_{b_k}) - (d_{a_k} - \Delta d_{a_k})\right)\hat{y}_{a_k} - \max\{\Delta d_{a_k}\hat{y}_{a_k} - \hat{\theta}, 0\}\Big) \\
&= \sum_{i\in\bar{\mathcal{K}}} d_i\hat{y}_i(1-x_i) - \Gamma_d\hat{\theta} + \sum_{k\in\mathcal{K}}\left(d_{b_k}\hat{y}_{a_k} - \max\{\Delta d_{b_k}\hat{y}_{a_k} - \hat{\theta}, 0\}\right) \\
&\qquad - \sum_{k\in\mathcal{K}}\Big(d_{a_k}\hat{y}_{a_k}x_{a_k} + (\Delta d_{b_k} - \Delta d_{a_k})\hat{y}_{a_k} \\
&\qquad\qquad + \max\{\Delta d_{a_k}\hat{y}_{a_k} - \hat{\theta}, 0\} - \max\{\Delta d_{b_k}\hat{y}_{a_k} - \hat{\theta}, 0\}\Big)
\end{aligned} \tag{24}$$

Note that the right-hand side of (24) corresponds to the right-hand side of (23) with the additional term

$$\begin{aligned}
&\sum_{k\in\mathcal{K}}\Big(d_{a_k}\hat{y}_{a_k}x_{a_k} + (\Delta d_{b_k} - \Delta d_{a_k})\hat{y}_{a_k} \\
&\qquad + \max\{\Delta d_{a_k}\hat{y}_{a_k} - \hat{\theta}, 0\} - \max\{\Delta d_{b_k}\hat{y}_{a_k} - \hat{\theta}, 0\}\Big)
\end{aligned} \tag{25}$$

subtracted. We now show that the latter term is non-negative. For all $k \in \mathcal{K}$, we have

$$\begin{aligned}
&\max\{\Delta d_{a_k}\hat{y}_{a_k} - \hat{\theta}, 0\} - \max\{\Delta d_{b_k}\hat{y}_{a_k} - \hat{\theta}, 0\} \\
&= \begin{cases} (\Delta d_{a_k} - \Delta d_{b_k})\hat{y}_{a_k}, & \Delta d_{b_k}\hat{y}_{a_k} \geq \Delta d_{a_k}\hat{y}_{a_k} \geq \hat{\theta}, \\ -\Delta d_{b_k}\hat{y}_{a_k} + \hat{\theta}, & \Delta d_{b_k}\hat{y}_{a_k} \geq \hat{\theta} \geq \Delta d_{a_k}\hat{y}_{a_k}, \\ 0, & \hat{\theta} \geq \Delta d_{b_k}\hat{y}_{a_k} \geq \Delta d_{a_k}\hat{y}_{a_k}, \end{cases}
\end{aligned}$$

and we consequently obtain

$$\begin{aligned}
&d_{a_k}\hat{y}_{a_k}x_{a_k} + (\Delta d_{b_k} - \Delta d_{a_k})\,\hat{y}_{a_k} + \max\{\Delta d_{a_k}\hat{y}_{a_k} - \hat{\theta}, 0\} - \max\{\Delta d_{b_k}\hat{y}_{a_k} - \hat{\theta}, 0\} \\
&= \begin{cases} d_{a_k}\hat{y}_{a_k}x_{a_k}, & \Delta d_{b_k}\hat{y}_{a_k} \geq \Delta d_{a_k}\hat{y}_{a_k} \geq \hat{\theta}, \\ d_{a_k}\hat{y}_{a_k}x_{a_k} - \Delta d_{a_k}\hat{y}_{a_k} + \hat{\theta}, & \Delta d_{b_k}\hat{y}_{a_k} \geq \hat{\theta} \geq \Delta d_{a_k}\hat{y}_{a_k}, \\ d_{a_k}\hat{y}_{a_k}x_{a_k} + (\Delta d_{b_k} - \Delta d_{a_k})\hat{y}_{a_k}, & \hat{\theta} \geq \Delta d_{b_k}\hat{y}_{a_k} \geq \Delta d_{a_k}\hat{y}_{a_k}. \end{cases}
\end{aligned}$$

The latter is greater or equal to $d_{a_k}\hat{y}_{a_k}x_{a_k}$ for all $k \in \mathcal{K}$. In particular, we have $d_{a_k}\hat{y}_{a_k}x_{a_k} = 0$ if $x_{a_k} = 0$ and $d_{a_k}\hat{y}_{a_k} > 0$, otherwise. To sum up, the term in (25) is non-negative. This means that feasibility w.r.t. Inequality (23) implies feasibility w.r.t. Inequality (24). Since (23) is a valid inequality, this concludes the proof. $\square$

In the next theorem, we consider lifted versions for the enhanced interdiction cuts stated in (20).

Theorem 5. *For any arbitrary but fixed* $\hat{y} \in \hat{Y}$ *and* $\ell \in \{\Gamma_d, \dots, n+1\}$, *let* $K \in [n]$, $S_a^\ell = \{a_1, \dots, a_K\} \subset D_+^\ell$, *and* $S_b^\ell = \{b_1, \dots, b_K\} \subset D_+^\ell$ *be chosen such that* $S_a^\ell \cap S_b^\ell = \emptyset$, $\hat{y}_{a_k} = 1$, $\hat{y}_{b_k} = 0$, $B_{\cdot a_k} \geq B_{\cdot b_k}$, *and* $\tilde{d}(\ell)_{a_k} < \tilde{d}(\ell)_{b_k}$ *hold for all* $k \in [K]$. *Under Assumptions 5 and 6, the lifted interdiction cut*

$$\eta \geq -\Gamma_d \Delta d_\ell + \sum_{i \in D_+^\ell} \tilde{d}(\ell)_i \hat{y}_i (1 - x_i) + \sum_{k=1}^{K} (\tilde{d}(\ell)_{b_k} - \tilde{d}(\ell)_{a_k})(1 - x_{b_k}) \tag{26}$$

is valid for Problem (2) *with the lower-level optimal-value function* (3).

Proof. Let (x, η) be a feasible leader's decision for Problem (2) with the lower-level optimal-value function (3). If $x_{b_k} = 1$ holds for all $k \in [K]$, we obtain the enhanced interdiction cut as stated in (20), which is satisfied by x. Otherwise, let us define $\mathcal{K} := \{k \in [K] \colon x_{b_k} = 0\}$ as well as $\bar{\mathcal{K}} := D_+^\ell \setminus \{a_k, b_k \colon k \in \mathcal{K}\}$ and consider the alternative decision of the follower given by

$$y'_i = \begin{cases} \hat{y}_i, & i \in \bar{\mathcal{K}}, \\ 1, & i \in \{b_k \colon k \in \mathcal{K}\}, \\ 0, & i \in \{a_k \colon k \in \mathcal{K}\}. \end{cases}$$

By construction, we have $y' \in Y$. Since the interdiction cuts (20) are not only valid for vertices contained in the set $\hat{Y}$ but also for any lower-level feasible point, the leader's decision x satisfies the basic interdiction cut associated with y' and ℓ, i.e.,

$$\begin{aligned}
\eta \geq & -\Gamma_d \Delta d_\ell + \sum_{i \in D_+^\ell} \tilde{d}(\ell)_i y'_i (1 - x_i) \\
= & -\Gamma_d \Delta d_\ell + \sum_{i \in \bar{\mathcal{K}}} \tilde{d}(\ell)_i \underbrace{y'_i}_{=\hat{y}_i} (1 - x_i) \\
& + \sum_{k \in \mathcal{K}} \Big(\tilde{d}(\ell)_{a_k} \underbrace{y'_{a_k}}_{=0} (1 - x_{a_k}) + \tilde{d}(\ell)_{b_k} \underbrace{y'_{b_k}(1 - x_{b_k})}_{=1} \Big) \\
= & -\Gamma_d \Delta d_\ell + \sum_{i \in \bar{\mathcal{K}}} \tilde{d}(\ell)_i \hat{y}_i (1 - x_i) + \sum_{k \in \mathcal{K}} \tilde{d}(\ell)_{b_k}.
\end{aligned} \tag{27}$$

Further, we can re-write Inequality (26) as

$$
\begin{aligned}
\eta \geq & -\Gamma_d \Delta d_\ell + \sum_{i \in D_+^\ell} \tilde{d}(\ell)_i \hat{y}_i (1 - x_i) + \sum_{k=1}^{K} (\tilde{d}(\ell)_{b_k} - \tilde{d}(\ell)_{a_k})(1 - x_{b_k}) \\
= & -\Gamma_d \Delta d_\ell + \sum_{i \in \bar{\mathcal{K}}} \tilde{d}(\ell)_i \hat{y}_i (1 - x_i) + \sum_{k \in \mathcal{K}} \tilde{d}(\ell)_{a_k} \underbrace{\hat{y}_{a_k}}_{=1} (1 - x_{a_k}) \\
& + \sum_{k \in \mathcal{K}} \tilde{d}(\ell)_{b_k} \underbrace{\hat{y}_{b_k}}_{=0} (1 - x_{b_k}) + \sum_{k=1}^{K} (\tilde{d}(\ell)_{b_k} - \tilde{d}(\ell)_{a_k})(1 - x_{b_k}) \\
= & -\Gamma_d \Delta d_\ell + \sum_{i \in \bar{\mathcal{K}}} \tilde{d}(\ell)_i \hat{y}_i (1 - x_i) + \sum_{k \in \mathcal{K}} \tilde{d}(\ell)_{a_k} (1 - x_{a_k}) \\
& + \sum_{k \in \mathcal{K}} (\tilde{d}(\ell)_{b_k} - \tilde{d}(\ell)_{a_k}) \underbrace{(1 - x_{b_k})}_{=1} + \sum_{k \in [K] \setminus \mathcal{K}} (\tilde{d}(\ell)_{b_k} - \tilde{d}(\ell)_{a_k}) \underbrace{(1 - x_{b_k})}_{=0} \\
= & -\Gamma_d \Delta d_\ell + \sum_{i \in \bar{\mathcal{K}}} \tilde{d}(\ell)_i \hat{y}_i (1 - x_i) + \sum_{k \in \mathcal{K}} \tilde{d}(\ell)_{a_k} (1 - x_{a_k}) \\
& + \sum_{k \in \mathcal{K}} \left(\tilde{d}(\ell)_{b_k} - \tilde{d}(\ell)_{a_k} \right) \\
= & -\Gamma_d \Delta d_\ell + \sum_{i \in \bar{\mathcal{K}}} \tilde{d}(\ell)_i \hat{y}_i (1 - x_i) + \sum_{k \in \mathcal{K}} \tilde{d}(\ell)_{b_k} - \sum_{k \in \mathcal{K}} \tilde{d}(\ell)_{a_k} x_{a_k}.
\end{aligned}
\tag{28}
$$

Note that the right-hand side of (28) corresponds to the right-hand side of (27) with the additional term

$$\sum_{k \in \mathcal{K}} \tilde{d}(\ell)_{a_k} x_{a_k} \geq 0$$

subtracted. Here, we have $\tilde{d}(\ell)_{a_k} x_{a_k} = 0$ if $x_{a_k} = 0$ and $\tilde{d}(\ell)_{a_k} x_{a_k} > 0$, otherwise. Hence, feasibility w.r.t. Inequality (27) implies feasibility w.r.t. Inequality (28). Since (27) is a valid inequality, this concludes the proof. □

Let us point out that the items in the sets S_a and S_b (or S_a^ℓ and S_b^ℓ in the multi-follower case) can be paired in different ways. This might yield different lifted cuts. To this end, we consider the following separation procedure. For an item a that is a candidate to enter the set S_a or S_a^ℓ, i.e., $\hat{y}_a \geq 1$ holds, we select its counterpart b among all items that satisfy the requirements with maximum value of $((d_b - \Delta d_b) - (d_a - \Delta d_a))\hat{y}_{a_k}(1 - x_b)$ and $(\tilde{d}(\ell)_b - \tilde{d}(\ell)_a)(1 - x_b)$ for the extended formulation and the multi-follower-based approach, respectively. If such a pair (a, b) exists, items a and b are inserted into the sets S_a and S_b (or S_a^ℓ and S_b^ℓ in the multi-follower case) and then removed from any further consideration.

3.3. **Problems with an Uncertain Lower-Level Constraint.**

To conclude this section, we briefly address uncertainties that only arise in a single packing-type constraint of the follower as stated around (5) and (6). For the validity of the proposed cuts, we need to impose the following.

Assumption 11. *The uncertain lower-level constraint does not contain terms depending on the leader's decision, i.e.,* $v = 0$.

Assumption 12. *All constraint coefficients of the uncertain lower-level constraint are non-negative, i.e.,* $w_i \geq 0$ *holds for all* $i \in [n]$.

Assumption 11 is an analog of Assumption 8 stating that the leader's variables x are linked to the lower-level problem only via the interdiction constraints $y_i \leq u_i(1 - x_i)$. Assumption 12 ensures that the Γ_w-robust lower-level problems (8) and (11) satisfy the downward monotonicity property. Both assumptions are necessary to exploit a penalized formulation of the Γ_w-robust follower's problem to derive valid cuts as it is done in the case of uncertain objective function coefficients. Again, we remove interdiction constraints and add penalty terms $-d_i y_i x_i$ to the objective function for all $i \in [n]$. Hence, we consider the same deterministic objective function

$$\sum_{i=1}^{n} d_i y_i (1 - x_i)$$

in the extended formulation as well as in the multi-follower formulation. In particular, the objective function is linear for a fixed leader's decision x. However, the description of the resulting feasible set differs for both approaches. For the extended formulation, we maximize over the feasible set of the penalized follower's problem projected onto the y-space, which is given by

$$\Theta = \{y \in \mathcal{Y} \colon \exists\, z, \theta \geq 0 \text{ such that (8b) and (8c) are satisfied}\}\,.$$

The feasible set Θ is independent from the leader's decision. Hence, an optimal solution of the Γ_w-robust follower's problem is attained at a vertex of the convex hull of Θ. We denote $\hat{\Theta}$ as the set containing the finite number of vertices of the convex hull of Θ. Then, the interdiction cuts

$$\eta \geq \sum_{i=1}^{n} d_i \hat{y}_i (1 - x_i) \quad \text{for all } \hat{y} \in \hat{\Theta} \tag{29}$$

are valid for Problem (2) and can equivalently replace Constraint (2c). Under Assumptions 5 and 6, when considering the multi-follower approach, we maximize over the ℓ-dependent feasible sets

$$\left\{ y \in \{0,1\}^n \colon \Gamma_w \Delta w_\ell + \sum_{i=1}^{n} \tilde{w}(\ell)_i y_i \leq C \right\},$$

which may differ in each follower sub-problem. However, it is easy to see that an optimal solution needs to be contained in the union of all ℓ-dependent sets. Let Y' be the set containing the finite number of vertices of the convex hull of the union of all ℓ-dependent sets. Then, we obtain the interdiction cuts for the multi-follower case by replacing $\hat{y} \in \hat{\Theta}$ with $\hat{y} \in Y'$ in (29), which can equivalently replace Constraint (2c) in Problem (2).

4. COMPUTATIONAL RESULTS

We now provide detailed numerical results for the proposed methods to solve interdiction problems with a monotone Γ-robust follower. Our solution approaches are implemented in Python 3.6.9 and Gurobi 9.1.2 is used to solve all arising optimization problems. [1] To add the interdiction cuts described in the previous sections, we use Gurobi's lazy constraint callbacks, which requires to set the parameter LazyConstraints to 1. All other parameters have been left at their default settings. The tests have been realized on an Intel XEON SP 6126 at 2.6 GHz (6 cores) with 32 GB RAM, which is part of the high performance cluster "Elwetritsch" at TU Kaiserslautern within the "Alliance of High Performance Computing Rheinland-Pfalz" (AHRP).[2] For each test run, we set the time limit to 1 h. We refer to the branch-and-cut method that exploits the multi-follower approach as MF and to the one that is based on the extended formulation as Ext. In Section 3.2, we have discussed several enhancements to improve the performance of MF and Ext, which we assess in the following. The aim is to determine a "winner setting" for MF and Ext and to compare both approaches. For the ease of presentation, we focus on the bilevel knapsack interdiction problem as a typical example of an interdiction problem with a monotone follower. Moreover, we only discuss the results for problems with uncertainties regarding the objective function coefficients.

Before we start, let us also comment on that there is no other tailored method in the literature that solves bilevel knapsack interdiction problems using a Γ-robust treatment of uncertainties. Hence, there are no alternative methods that we could compare with. However,

---

[1] The code for the approaches presented in this paper is publicly available at https://github.com/YasmineBeck/gamma-robust-knapsack-interdiction-solver.

[2] We kindly acknowledge the support of RHRK (https://rz.rptu.de/en/).

using the extended formulation, one could transform the Γ-robust interdiction problem into a standard mixed-integer linear bilevel problem. The latter can, in general, be solved with the MibS solver (Tahernejad, Ralphs, and DeNegre 2020) and the general branch-and-cut solver presented in Fischetti, Ljubić, et al. (2017). We tested both solvers for 40 robustified knapsack instances with 35 items; see below for more details on our test set. MibS is not able to solve any of these instances—although 35 items belong to the smallest class of instances in our test set. The smallest optimality gap we get is 105 %. The other solver solved 2 out of 40 instances. The runtimes are 1797.27 s and 3188.13 s, which is more than a factor of 360 or 635 longer than our tailored methods take. Consequently, we omit a more detailed comparison with these general-purpose bilevel solvers.

4.1. Generation of Knapsack Test Instances.

To test our solution approaches, we consider the bilevel knapsack interdiction problem that has been considered in Caprara et al. (2016) and which is formally stated as

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

with $B, C \in \mathbb{Z}_+$, and $p, v, w \in \mathbb{Z}_+^n$. In particular, the bilevel knapsack interdiction problem is a prominent example for an interdiction problem with a monotone follower. The motivation for us to focus on this type of problem is the following. Classic knapsack problems belong to the most intensively studied discrete optimization problems, which is due to their relevance in many real-world applications, e.g., in the field of economics. In particular, the bilevel knapsack interdiction problem naturally extends the classic knapsack problem such as to capture competitive situations; see, e.g., DeNegre (2011) for a specific application in corporate strategy. Moreover, the knapsack interdiction problem is commonly used as a benchmark for testing bilevel optimization solvers; see, e.g., DeNegre and Ralphs (2009) and Tang et al. (2016). In its deterministic variant, the knapsack interdiction problem has been studied, e.g., in Caprara et al. (2013, 2016), Della Croce and Scatamacchia (2020), DeNegre (2011), Fischetti, Ljubić, et al. (2019), Fischetti, Monaci, et al. (2018), Shi et al. (2020), and Tang et al. (2016). For our computational study, we generate deterministic knapsack interdiction instances according to Martello et al. (1999), which we adapt to account for a Γ-robust follower. Before we comment on the uncertainty parameterizations that we consider, let us briefly describe the generation of the knapsack instances. The profits p_i and the follower's weights w_i take uncorrelated integer values from the interval $[0, 100]$. For each instance size $n \in \{35, 40, 45, 50, 55, \dots, 100\}$, 10 instances have been generated. The follower's knapsack capacity C is set to $\lceil (N/11) \sum_{i=1}^n w_i \rceil$, where $N \in \{1, \dots, 10\}$ is used to identify the instance number. The leader's weights v_i and the interdiction budget B are uniformly random integers from the intervals $[0, 100]$ and $[C - 10, C + 10]$. Let us point out that the deterministic knapsack instances with size $n \in \{35, 40, \dots, 55\}$ are taken from Caprara et al. (2016). All other instances have not yet been studied in the literature and are newly generated. To study the effects of a Γ-robust follower, we consider four different uncertainty parameterizations. We assume that the deviations take either 10 % or 25 % of the nominal value. The parameter Γ is set to either 10 % or 50 % of the instance size n. In the case of a fractional value for Γ, we then consider the closest integer. Hence, our test set contains 560 robustified knapsack instances. Finally, let us mention that we do not consider any instances with more than 100 items because even the most advanced variants of the presented approaches cannot solve all of the 560 robustified knapsack instances described above within a reasonable amount of time.

4.2. Lifted Cuts and Dominance Inequalities.

We now assess the influence of lifted cuts and dominance inequalities on the overall performance of the solution method. To this end, we consider the following four settings.

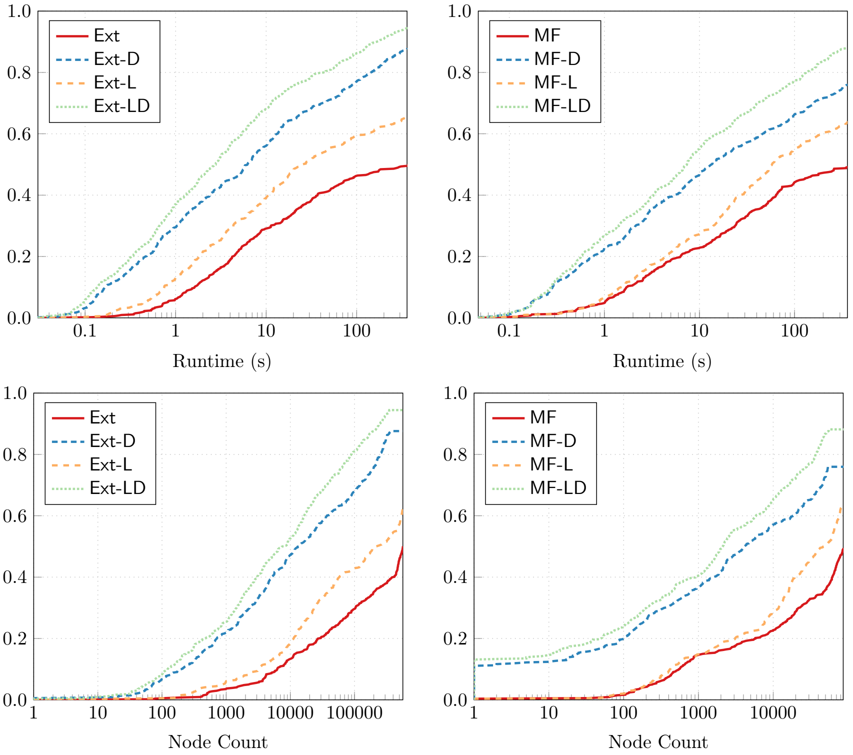

FIGURE 1. Log-scaled ECDF plots of the runtimes (in s) (top) and the number of branch-and-bound nodes (bottom) for Ext (left) and MF (right) using lifted cuts and/or dominance inequalities.

- **MF/Ext:** The basic setting in which only basic interdiction cuts are added without any further enhancements.
- **MF-D/Ext-D:** The basic setting with the addition of dominance inequalities (21) regarding the leader's decision.
- **MF-L/Ext-L:** The basic setting but instead of considering basic interdiction cuts, we add lifted interdiction cuts.
- **MF-LD/Ext-LD:** Like MF-L or Ext-L but with the addition of dominance inequalities (21).

As a default for MF, we consider the cut separation strategy in which all violated (lifted) interdiction cuts are added to the problem formulation. Figure 1 shows the empirical cumulative distribution functions (ECDFs) w.r.t. the runtimes and the number of branch-and-bound nodes of the four settings. The ECDFs can be interpreted as the percentage of instances (y-axis) that can be solved within a certain amount of time and/or after investigating a certain number of branch-and-bound nodes (log-scaled x-axis). Note that, to have a fair comparison, we exclude 29 instances that none of the considered variants can solve within the time limit. Moreover, we exclude 19 instances that every variant can solve within 5 s to avoid drawing erroneous conclusions because of low runtimes. Hence, we consider a total of 512 instances at this point. While lifted interdiction cuts only slightly improve the performance of MF and Ext, the use of dominance inequalities significantly enhances the performance of both approaches. The combination of lifted cuts and dominance inequalities yields only minor further improvement compared to MF-D and Ext-D. Nevertheless, MF-LD and Ext-LD dominate all other settings of the respective solution approach. This observation is also underlined by the results in Table 1. In what follows, we thus hold on to the variants

Table 1. Mean and median runtimes (in s), mean and median number of branch-and-bound nodes as well as the number of solved instances for the variants with lifted cuts and/or dominance inequalities.

	runtimes		node count		
	mean	median	mean	median	solved
Ext	1863.94	1507.90	689196.96	574953.50	268
Ext-D	583.12	53.52	108588.91	10254.50	473
Ext-L	1337.09	188.62	435544.75	168709.50	353
Ext-LD	260.25	21.86	46412.61	5593.50	510
MF	1933.78	2147.53	106190.09	87702.00	265
MF-D	1018.79	100.61	26374.73	3197.50	410
MF-L	1482.54	510.86	69689.89	36435.00	344
MF-LD	573.54	62.09	15212.53	1771.00	476

with additional dominance inequalities and lifted interdiction cuts as our "winner setting" for both approaches.

4.3. **The Benefits of Maximal Packings.** We consider maximal packings of the follower to avoid the generation of unnecessary interdiction cuts. This is achieved in the following manner. We determine a feasible follower's decision y for the current x by either solving the follower's extended formulation (14) or the sub-problems (15). In the next step, we complete the follower's decision to a maximal packing in a greedy-like fashion. To this end, we order the indices of the follower's variables according to non-increasing profit-to-weight ratio and then gradually add items that still fit into the follower's knapsack. When considering the extended formulation, however, we need to further check if the follower's decision (y, z, θ) satisfies $\theta \geq \Delta p_i$. This requirement is necessary to ensure the feasibility of the maximal packing w.r.t. the constraints $z_i + \theta \geq \Delta p_i y_i$ for all $i \in [n]$. Finally, we generate and add only the interdiction cut that corresponds to the follower's maximal packing.

To assess the influence of maximal packings, we adopt the parameterizations of our previous "winner settings" MF-LD and Ext-LD with the difference that we now only add cuts corresponding to maximal packings of the follower. We refer to these settings as MF-LD-Max and Ext-LD-Max. Again, for a fair comparison, we exclude 21 instances that none of the variants can solve within the time limit and 90 instances that every variant can solve within 5 s so that we consider a total of 449 instances here. Based on Figure 2, it can be seen that adding only the interdiction cuts corresponding to maximal packings of the follower improves the performance of the overall solution method for both approaches. In particular, this holds true for the easier instances, which is also underlined by the number of solved problems presented in Table 2. Table 2 further shows considerably smaller mean and median runtimes for MF-LD-Max and Ext-LD-Max compared to their counterparts without maximal packings. Also the mean and median number of branch-and-bound nodes visited is significantly smaller in both settings when maximal packings of the follower are considered. To sum up, the observations drawn from Figure 2 and Table 2 clearly suggest that MF-LD-Max and Ext-LD-Max are the "winner settings" among the considered variants.

As mentioned in Section 3.2, we also considered leader's maximal packings. However, preliminary computational results revealed that maximal packings for the leader interfere with Gurobi's integrated branching rules and node selection, which is why we decided to refrain from this ingredient.

4.4. **The Impact of Warmstarting.** In our computational study, we also investigated how warmstarting the proposed methods may affect their performance. To this end, we considered two options—a heuristic similar to the one presented in Fischetti, Ljubić, et al. (2019) and solving the nominal knapsack interdiction problem—to determine an initial feasible (and

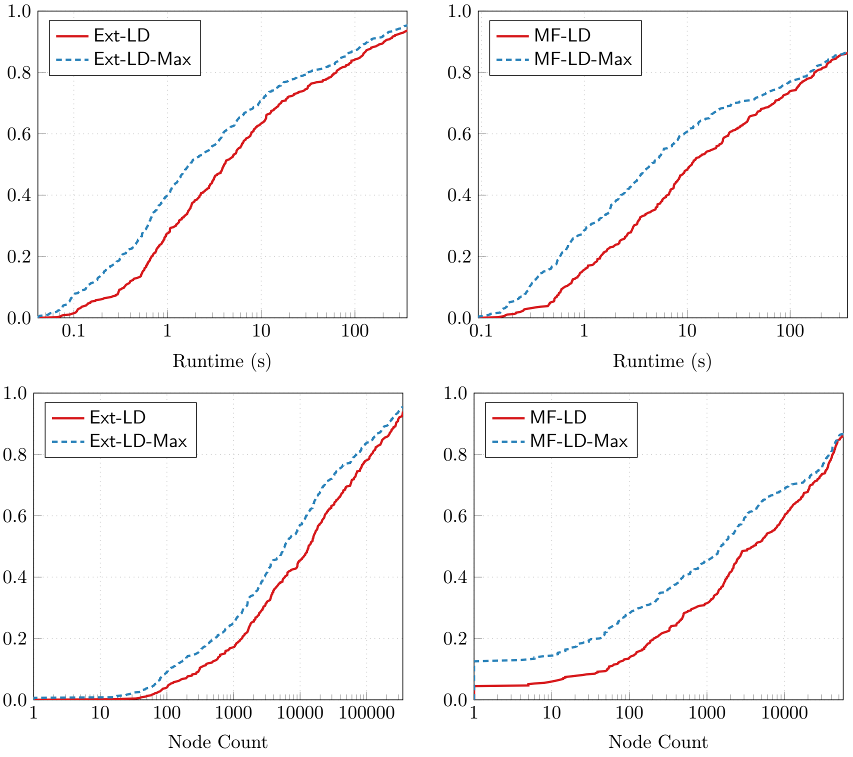

Figure 2. Log-scaled ECDF plots of the runtimes (in s) (top) and the number of branch-and-bound nodes (bottom) for Ext (left) and MF (right) considering maximal packings of the follower.

Table 2. Mean and median runtimes (in s), mean and median number of branch-and-bound nodes as well as the number of solved instances for the variants with and without maximal packings of the follower.

	runtimes		node count		
	mean	median	mean	median	solved
Ext-LD	360.70	37.71	61901.68	12293.00	439
Ext-LD-Max	252.71	15.32	41166.66	5253.00	448
MF-LD	718.13	94.62	18089.67	2789.00	405
MF-LD-Max	616.47	37.65	15079.47	1373.00	406

potentially good) decision of the leader for Problem (P_0). The pair (x, η) that we obtain using any of the two options is then provided to Gurobi as MIP start.

The computational results obtained for both settings revealed that warmstarting the methods has no significant impact on their performance, neither for MF nor for Ext. This is why we omit the details on the runtimes and the number of branch-and-bound nodes here. As per default, both methods are warmstarted using the heuristic in all subsequent considerations.

4.5. **Comparison of Different Cut Separation Strategies and the Potential of Parallelization.** We further evaluate enhancement techniques that exploit the special structure of the multi-follower formulation. Due to the fact that the overall problem can be considered as a single-leader-multi-follower problem with independent followers, we can make use of parallelization as briefly mentioned in Section 2. In this context, we can further consider different cut separation strategies instead of adding all violated interdiction cuts in each iteration of the algorithm. This way, the number of cuts added to the problem formulation can be reduced, which might speed up the overall solution method. We adopt the parameterizations of the previous "winner setting", i.e., we consider the multi-follower formulation with lifted cuts corresponding to maximal packings of the follower, dominance inequalities regarding the leader's decision, and heuristic warmstarts. For notational convenience, we omit MF-LD-Max as a prefix for the considered variants when we focus on the following cut separation strategies.

All-In: The default setting in which all interdiction cuts that are violated by the current leader's decision are added to the problem formulation.

Most-Violated: A single cut is added corresponding to the interdiction inequality (20) that is maximally violated by the current leader's decision.

Sorting: In Álvarez-Miranda, Fernández, et al. (2015), the authors propose a learning mechanism to identify the sub-problems that produce potentially good cuts. This is done by taking the information of previous iterations into account. We adapt the proposed strategy to our setting such that a single potentially good cut is added in each iteration.

First-In: We iterate over $\ell \in \{\Gamma, \dots, n+1\}$, add the first interdiction cut that is violated by the current leader's decision, and then break the loop.

Random: Among the violated interdiction cuts, we randomly choose a single cut and add it to the problem formulation.

To assess the potential of parallelization, we consider so-called *idealized runtimes*, which reflect the overall runtime of the solution method provided that there are sufficient capacities available to solve all sub-problems in parallel. For each instance, the idealized runtimes are computed after solving the problems sequentially by taking the maximum over the runtime of each sub-problem.

In Figure 3, we compare the aforementioned cut separation strategies w.r.t. sequential and idealized runtimes. Note that we exclude 21 instances that none of the considered variants can solve within the time limit and 43 instances that every variant can solve within 5 s. Hence, we consider a total of 496 instances here. It can be seen that the first insertion strategy First-In harms the performance of the solution method for both sequential and idealized runtimes as a benchmark. The last observation is also reflected by the results on the number of branch-and-bound nodes presented in Figure 4 (left). A possible reason for this might be the following. By adding the first violated interdiction cut, later follower sub-problems that might produce stronger cuts are neglected. In particular, it is possible that the cut corresponding to the same follower sub-problem is added in each iteration of the algorithm. To overcome this situation, we consider Random as a modification of First-In. Based on Figure 3, it can be observed that this variant performs significantly better. In particular, Random outperforms all other cut separation strategies w.r.t. sequential runtimes. The potential reasons for the effectiveness of the randomized first insertion strategy are twofold. On the one hand, it seems beneficial to add a single cut in each iteration of the algorithm; otherwise, as for the strategy All-In, the leader's problem can get extremely large w.r.t. the number of constraints. On the other hand, Random has comparatively low computational costs. For Sorting, it can be seen that having information from previous iterations does not lead to a better choice than randomly adding a violated cut. When considering Most-Violated, we need to solve all sub-problems to determine the most violated interdiction cut, which seems to be rather expensive. The previous observations are underlined by the results in Table 3, since the sequential mean and median runtimes for Random are considerably smaller compared to all other cut separation strategies. Focusing on idealized runtimes, however, it is noteworthy that Random is no longer the "winning" cut separation strategy. Based on Figure 3 and Table 3, Most-Violated dominates the randomized first insertion strategy w.r.t. idealized runtimes.

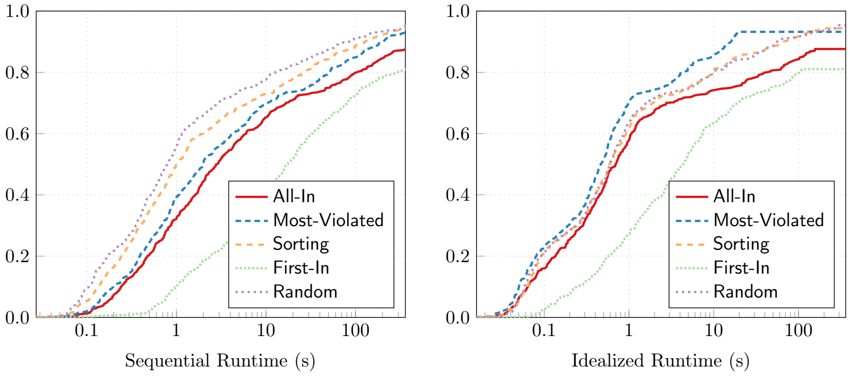

Figure 3. Log-scaled ECDF plots of the sequential runtimes (left) and the idealized runtimes (right) (in s) for different cut separation strategies.

Table 3. Sequential and idealized mean and median runtimes (in s) as well as the number of solved instances for the variants with different cut separation strategies.

	sequential runtimes		idealized runtimes		
	mean	median	mean	median	solved
All-In	543.17	24.85	267.35	6.03	452
Most-Violated	346.88	17.98	37.53	4.35	481
Sorting	245.16	9.19	138.89	5.53	487
First-In	855.10	157.92	466.59	37.68	418
Random	183.52	7.23	156.57	5.61	492

This is to be expected since we can benefit the most from parallelization for Most-Violated, where we indeed have to solve all of the sub-problems. In particular, the previous observations suggest that the higher computational costs for solving all sub-problems—especially if this is done in parallel—can be compensated to some extent by the strength of the added cuts. This strength is further visualized in Figure 4 (right) and Table 4. Here, to assess the quality of the added cuts, we consider the runtimes independently of the times spent for the cut generation. However, it is noteworthy that, apart from the strength of Most-Violated, the results also suggest that Sorting and Random yield cuts of good quality for the easier instances. Finally, we further highlight the benefit of adding the most-violated cut by considering the mean and median number of branch-and-bound nodes presented in Table 4. It can be observed that the mean and median number of nodes for Most-Violated are significantly smaller than the ones for almost all of the other cut separation strategies. Only All-In has a considerably smaller mean number of investigated nodes compared to Most-Violated. However, as stated earlier, it seems beneficial to add a single cut in each iteration of the algorithm instead of adding all violated cuts, which is underlined by both the runtime results as well as the number of solved instances.

To sum up, Most-Violated yields the strongest cuts among the considered variants. Moreover, provided that the necessary capacities are available to solve all sub-problems in parallel, Most-Violated particularly outperforms all other variants w.r.t. idealized runtimes. Hence, we consider Most-Violated as our “winning” strategy in the idealized setting. When considering sequential runtimes, however, the cost of solving all sub-problems cannot be completely compensated by the strength of the added cuts. Since the overall runtime is our decisive criterion, we prefer Random in the sequential setting.

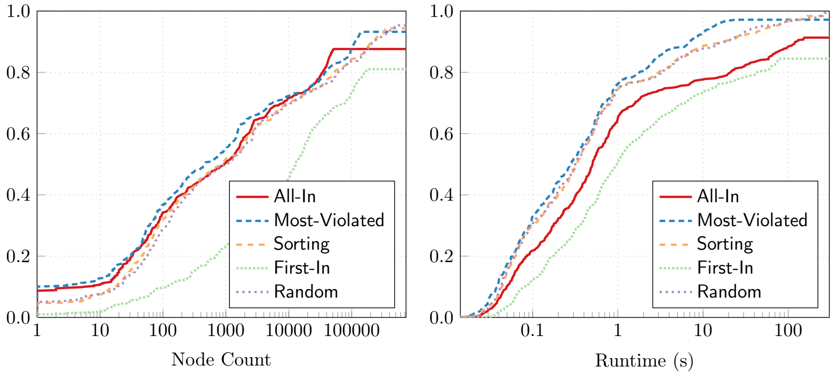

FIGURE 4. Log-scaled ECDF plots of the number of branch-and-bound nodes (left) and the runtimes (in s) independent of the times spent for the cut generation (right) for different cut separation strategies.

TABLE 4. Mean and median runtimes (in s) independent of the times spent for the cut generation as well as the mean and median number of branch-and-bound nodes for the variants with different cut separation strategies.

	runtimes		node count	
	mean	median	mean	median
All-In	258.36	5.05	14468.14	670.00
Most-Violated	27.67	2.83	24159.59	351.00
Sorting	104.84	3.26	39044.19	587.50
First-In	342.75	9.72	72248.80	11379.00
Random	113.29	3.16	45086.26	690.50

4.6. **Comparison of the Solution Approaches.** We now compare the "winning" parameterizations of the extended formulation and the multi-follower formulation. For the multi-follower-based approach, we particularly distinguish between the sequential and the idealized setting. Hence, we consider MF-LD-Max-Random, MF-LD-Max-Most-Violated, and Ext-LD-Max, which we abbreviate with MF-seq, MF-ideal, and Ext in the following. Again, we exclude 16 instances that none of the considered variants can solve within the time limit and 137 instances that every variant can solve within 5 s. In total, we thus consider 407 instances. Figure 5 (left) shows the ECDF plots w.r.t. the runtimes of the three considered variants. Note that we consider sequential and idealized runtimes for MF-seq and MF-ideal, respectively. We observe that MF-ideal clearly outperforms the remaining two approaches. This particularly affirms that the strength of the multi-follower-based approach lies in the possibility to parallelize the solution of the follower sub-problems. The previous observation is also underlined by the mean and median runtimes in Table 5. It can be seen that MF-ideal has significantly smaller mean and median runtimes compared to MF-seq and Ext. Despite being not as significant as when considering the runtimes, the same qualitative behavior can also be observed for the results on the number of nodes in Figure 5 (right) and Table 5. If, however, the capacity is not available to have an idealized parallelization, the multi-follower-based approach MF-seq still performs slightly better than Ext. Based on Figure 5, MF-seq seems to have an advantage over Ext on the easier instances. The last observation is also supported by the results on the mean and median runtimes presented in Table 5. Yet the number of solved instances suggests that Ext performs slightly better than MF-seq on the harder instances. Let us emphasize, however, that the amount of instances that can be solved by Ext but not

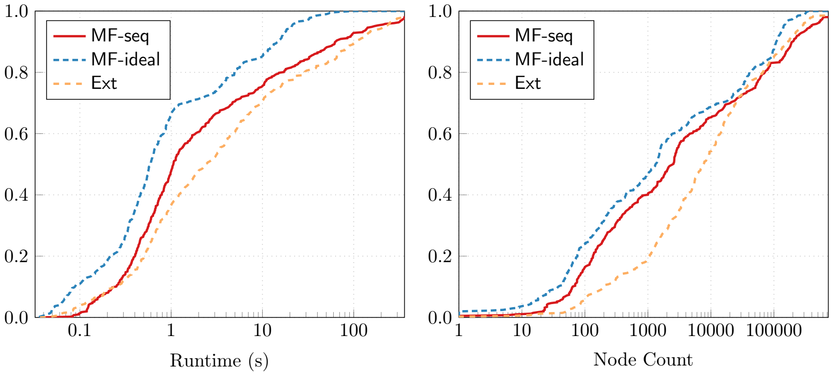

FIGURE 5. Log-scaled ECDF plots of the runtimes (in s) (left) and the number of branch-and-bound nodes (right) for the "winner settings" of Ext and MF.

TABLE 5. Mean and median runtimes (in s), mean and median number of branch-and-bound nodes as well as the number of solved instances for the "winner settings" of Ext and MF.

	runtimes		node count		
	mean	median	mean	median	solved
Ext	324.42	24.16	49818.23	7680.00	401
MF-seq	267.57	10.59	62861.15	2546.00	398
MF-ideal	43.64	5.99	29502.58	1333.00	387

by MF-seq is comparatively small, which is why we consider MF-seq as the overall better method in the sequential setting.

4.7. **The Computational Cost of Robustness.** To conclude our computational study, we address the computational cost of robustness. This expression captures the effect of robustification, e.g., on the overall runtimes of the solution methods. To evaluate the computational cost of robustness, we compare the three "winning" approaches MF-seq, MF-ideal, and Ext for the considered uncertainty parameterizations, which are referred to as (Δ, Γ). Here, $\Delta \in \{10, 25\}$ is used to specify the considered percentage deviations in the objective function coefficients and $\Gamma \in \{10, 50\}$ denotes the percentage that the parameter Γ takes of the instance size. Based on Table 6, it can be seen that the mean and median runtimes to solve the Γ-robust knapsack interdiction problem increase with increasing values of Δ and Γ for Ext. For MF-seq and MF-ideal, however, this does not seem to be the case in principle. Detailed runtime results for each knapsack instance can further be found in Table 8. For both the multi-follower approach and the extended formulation, we compare the nominal runtimes to the mean runtimes obtained from all considered uncertainty parameterizations. Note that we label the sequential and idealized mean runtimes with the superscripts seq and ideal, respectively.

To further assess the cost of robustness w.r.t. the runtimes, we measure the relative performance of the method in the Γ-robust and in the nominal case. Note that we restrict ourselves to the instances that have been considered in Section 4.6 for a fair comparison, i.e., we exclude 16 instances that none of the "winner" settings can solve within the time limit and 137 instances that every variant can solve within 5 s. To measure the relative performance, we determine the coefficient of runtimes $q_i = t_{i,\text{rob}}/t_{i,\text{nom}}$ for each knapsack instance i. Here, $t_{i,\text{rob}}$ and $t_{i,\text{nom}}$ denote the runtimes of the considered solution method for instance i in

TABLE 6. Mean and median runtimes (in s) for the "winner" settings of Ext and MF for the uncertainty parameterizations given by (Δ, Γ).

		uncertainty parameterization			
	$(\Delta, \Gamma) =$	(10,10)	(10,50)	(25,10)	(25,50)
mean	Ext	201.93	331.11	234.78	540.96
	MF-seq	301.37	234.55	334.64	192.01
	MF-ideal	43.63	52.38	43.93	34.94
median	Ext	6.34	41.50	10.70	84.54
	MF-seq	10.35	10.63	10.44	11.06
	MF-ideal	7.66	5.40	7.14	4.86

the Γ-robust and in the nominal case, respectively. In Figure 6, we show box-plots for the coefficients of runtimes corresponding to MF-seq, MF-ideal, and Ext. Each box in Figure 6 represents the distribution of the determined coefficients q over the 407 considered instances for the three "winning" solution methods. It can be seen that the box for Ext is considerably larger compared to the other two approaches. This shows that the coefficients for MF-seq and MF-ideal are less dispersed, which suggests that the performance of these methods is more stable compared to the one of Ext. In particular, the boxes for MF-seq and MF-ideal are of similar size, which reflects the similarity between the multi-follower-based solution approaches. Note that MF-seq and MF-ideal rely on the same solution procedure. The only difference between these approaches is that MF-seq is a sequential method and MF-ideal exploits parallelization. Let us point out that, in Figure 6, we use a logarithmic scaling of the y-axis for runtime coefficients greater than 40 such as to capture the spread of the outliers in a detailed and comprehensive way. For Ext, the outliers are widely scattered, which shows that the performance of the method is rather volatile. In contrast to that, smaller ranges of the outliers can be observed for MF-seq and MF-ideal. All previous observations are also affirmed by the results in Table 7. Based on Figure 6 and Table 7, it thus seems reasonable to prefer the multi-follower approach over the extended formulation since it seems to be the more stable method. Interestingly, we also observe coefficients of runtimes strictly smaller than 1 for some of the instances. This means that the robust counterpart can be solved faster than the nominal problem, i.e., robustification does not necessarily always lead to increased computational costs. However, the mean and median coefficients of runtimes shown in Table 7 emphasize that the robustification is not "for free". However, the computational cost of robustness for the multi-follower-based approaches—MF-ideal in particular—is comparatively small.

Table 7. Minimum, mean, median, and maximum values of the coefficients of runtimes for the "winner" settings of Ext and MF.

	min	mean	median	max
Ext	0.79	128.59	14.68	4365.61
MF-seq	0.76	7.71	4.95	74.66
MF-ideal	0.04	4.60	1.66	65.37

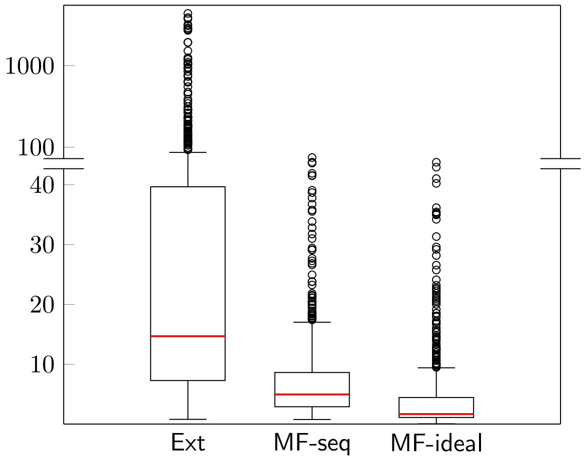

Figure 6. Box-plots of the coefficients of runtimes $q_i = t_{i,\mathrm{rob}}/t_{i,\mathrm{nom}}$ for the "winner" settings of Ext and MF.

To sum up, Γ-robust solutions are obtained at the expense of increased computational difficulty of the problem but, provided that there are sufficient capacities available to solve all of the follower sub-problems in parallel, the price of robustness w.r.t. runtimes is comparatively small. Thus, we prefer MF-ideal over MF-seq and Ext. In the sequential setting, the results in Table 7 clearly suggests that MF-seq outperforms Ext w.r.t. the stability of the method. Based on the results in Table 5, however, the advantage of MF-seq over Ext is not as significant as when the relative performance measure is considered. In particular, as elaborated in Section 4.6, Ext seems to have an advantage on harder instances, which also justifies the use of the extended formulation.

Table 8: Runtime results (in s) for the nominal knapsack instances and mean runtimes for their robust counterparts.

		Ext		MF		
size	instance	nominal	robust	nominal	robust$^{\mathrm{seq}}$	robust$^{\mathrm{ideal}}$
35	1	0.22	1.71	0.37	0.74	0.27
	2	0.49	6.00	0.49	1.13	0.45
	3	0.56	13.70	0.77	4.68	1.92
	4	0.22	1.75	1.34	6.08	3.23
	5	0.28	2.74	2.01	5.80	4.16
	6	0.14	0.71	0.38	0.69	0.49
	7	0.18	1.20	0.37	3.15	2.83
	8	0.16	0.53	0.14	0.39	0.18
	9	0.04	0.65	0.13	0.37	0.28

	10	0.05	0.31	0.04	0.32	0.14
40	1	0.33	2.28	0.36	1.43	0.50
	2	0.44	10.22	0.56	2.44	0.93
	3	1.47	42.99	1.25	11.75	3.41
	4	0.29	2.90	1.18	6.85	2.64
	5	0.20	2.51	2.09	6.11	4.53
	6	0.25	0.94	0.34	0.80	0.70
	7	0.15	1.18	0.35	0.86	0.52
	8	0.18	0.56	0.22	0.57	0.27
	9	0.18	4.85	0.31	3.81	3.45
	10	0.02	0.07	0.02	0.19	0.13
45	1	0.38	4.70	0.38	1.51	0.51
	2	0.38	10.91	0.46	1.61	0.61
	3	0.77	33.53	3.44	14.58	4.25
	4	0.43	34.35	7.34	22.32	6.71
	5	0.35	16.26	4.69	10.93	5.99
	6	0.21	1.39	0.46	3.41	2.98
	7	0.06	1.70	0.38	0.67	0.39
	8	0.13	0.86	0.27	1.02	0.67
	9	0.14	1.08	0.15	0.71	0.47
	10	0.20	1.01	0.22	0.70	0.45
50	1	0.36	5.11	0.40	1.57	0.77
	2	4.70	121.56	7.57	26.17	6.96
	3	5.30	76.00	10.08	19.37	7.45
	4	0.21	13.50	3.93	12.48	2.86
	5	0.42	45.12	13.53	34.86	9.56
	6	0.22	1.29	0.43	0.98	0.58
	7	0.17	2.46	0.40	1.46	0.97
	8	0.05	0.75	0.35	0.67	0.39
	9	0.16	9.69	0.30	4.36	4.08
	10	0.06	0.39	0.06	0.57	0.21
55	1	1.08	14.10	0.79	5.53	1.27
	2	2.55	86.23	4.31	19.58	4.45
	3	299.82	2191.23	261.03	870.73	196.44
	4	30.68	338.33	40.64	192.79	48.64
	5	0.67	38.71	15.82	20.45	12.10
	6	0.44	21.98	0.91	3.65	3.41
	7	0.23	8.73	0.57	4.25	3.59
	8	0.24	32.00	0.44	5.18	4.72
	9	0.14	3.21	0.16	1.18	0.74
	10	0.21	2.47	0.21	2.73	2.53
60	1	0.49	3.73	0.32	1.88	0.44
	2	3.28	58.43	1.49	14.08	3.30
	3	28.27	424.31	22.57	114.68	31.02
	4	32.84	358.24	35.87	182.28	54.59
	5	1.06	25.08	3.70	17.98	10.79
	6	0.66	6.44	0.43	5.37	5.16
	7	0.22	5.77	0.41	1.15	0.56
	8	0.24	4.99	0.44	3.47	2.88
	9	0.23	3.74	0.31	3.84	3.54
	10	0.20	1.53	0.27	0.74	0.46

65	1	0.39	4.97	0.51	1.93	0.45
	2	5.71	99.99	5.05	28.94	6.22
	3	66.36	705.40	61.75	172.25	56.87
	4	119.97	773.86	106.79	548.47	128.77
	5	0.59	78.25	9.81	60.65	28.67
	6	0.34	24.47	0.79	7.99	6.72
	7	0.28	19.71	0.79	5.67	5.27
	8	0.28	13.19	0.27	1.29	0.67
	9	0.28	7.37	0.18	5.67	5.28
	10	0.23	1.32	0.29	0.86	0.47
70	1	0.37	5.77	0.55	3.20	0.79
	2	6.95	173.91	10.91	48.16	10.97
	3	171.28	1477.50	197.35	594.18	166.50
	4	52.55	1168.15	143.76	566.51	160.45
	5	0.82	200.63	35.73	148.16	53.30
	6	0.42	32.51	0.74	6.44	4.02
	7	0.31	32.62	0.70	1.38	0.70
	8	0.18	20.74	0.21	2.16	1.42
	9	0.26	5.06	0.23	4.85	3.92
	10	0.23	2.67	0.15	1.56	1.01
75	1	0.65	10.65	0.76	4.53	0.97
	2	30.46	302.07	32.78	82.94	18.50
	3	185.97	1669.87	148.04	594.80	159.71
	4	2.37	1420.08	282.86	1476.37	349.61
	5	0.56	165.93	27.51	87.01	34.18
	6	0.40	40.84	1.29	4.12	3.31
	7	0.21	21.62	0.41	1.08	0.91
	8	0.21	15.62	0.30	6.81	6.63
	9	0.24	3.41	0.31	1.43	0.82
	10	0.11	0.91	0.28	0.85	0.48
80	1	0.66	15.50	0.65	9.21	1.94
	2	58.70	584.49	31.46	194.11	41.11
	3	513.98	2410.95	302.54	1554.35	399.95
	4	0.93	–	496.44	–	–
	5	0.52	187.66	27.88	109.24	38.73
	6	0.30	75.02	1.28	11.47	8.50
	7	0.35	42.64	0.65	7.76	6.42
	8	0.30	16.52	0.31	5.23	4.66
	9	0.10	4.79	0.08	1.85	1.13
	10	0.25	0.90	0.30	1.35	0.58
85	1	1.07	43.68	2.47	26.83	3.19
	2	84.63	875.17	89.83	353.61	73.44
	3	425.13	3112.44	333.26	2305.22	376.08
	4	0.97	–	635.59	–	–
	5	0.77	218.35	30.64	91.05	34.17
	6	0.36	55.90	1.31	4.23	2.17
	7	0.28	53.94	1.25	8.13	7.25
	8	0.21	22.15	0.24	7.73	6.85
	9	0.17	6.58	0.15	9.70	7.37
	10	0.11	0.71	0.09	1.37	0.52
90	1	2.60	43.73	2.73	19.40	3.87

	2	130.91	1184.05	110.48	506.53	106.59
	3	1324.50	–	1102.64	–	–
	4	1.25	–	1788.24	–	–
	5	0.85	298.27	29.03	74.70	28.46
	6	0.48	90.69	1.22	10.88	6.64
	7	0.29	53.34	0.92	9.32	7.33
	8	0.24	19.87	0.36	8.75	7.54
	9	0.21	7.75	0.21	8.54	6.76
	10	0.18	2.63	0.22	1.82	0.75
95	1	1.54	64.46	2.55	33.70	4.15
	2	140.75	1203.87	101.87	702.22	104.00
	3	1935.09	–	1266.83	–	–
	4	1.87	–	–	–	–
	5	1.87	683.32	54.15	206.94	47.86
	6	0.54	155.16	2.57	12.58	3.53
	7	0.35	143.62	1.08	11.34	8.30
	8	0.31	39.91	0.46	8.81	7.65
	9	0.63	10.83	0.79	4.93	2.32
	10	0.14	1.79	0.14	1.56	0.53
100	1	3.09	69.32	1.61	37.42	3.71
	2	203.20	1597.40	136.96	694.00	175.65
	3	–	–	–	–	–
	4	3.47	–	–	–	–
	5	0.82	1122.00	71.78	391.87	105.00
	6	0.63	491.01	3.35	23.15	10.84
	7	0.45	461.73	1.21	10.86	8.76
	8	0.71	58.45	0.70	2.37	1.62
	9	0.13	15.88	0.33	2.92	1.01
	10	0.32	3.56	0.33	9.13	8.12

5. Conclusion

In this paper, we consider discrete min-max problems with a follower facing uncertain lower-level data. We exploit a Γ-robust approach so that the follower only hedges against a subset of deviations in the uncertain parameters as to adversely affect the solution of the problem. We present two approaches—an extended formulation and a multi-follower formulation—to model this type of situation. For both frameworks, we present a fairly generic branch-and-cut method. Nevertheless, we can obtain stronger formulations for certain types of problems. As an example, we consider interdiction problems with a monotone Γ-robust follower to derive problem-tailored cuts that generalize existing interdiction cuts from the literature. Finally, we conduct a computational study to assess the performance of the two proposed solution approaches. To this end, we focus on the bilevel knapsack interdiction problem, which is one of the most prominent examples of monotone interdiction problems.

The computational results suggest that the extended formulation (`Ext`) performs slightly better on harder knapsack instances. However, smaller overall mean and median runtimes and a more stable performance of the method compared to `Ext` can be observed for the multi-follower formulation (`MF`). In particular, we can exploit parallelization for the multi-follower formulation, which is a major strength of this solution approach. Nevertheless, the study justifies the use of both the extended formulation as well as the multi-follower approach.

Despite the contribution of this paper, there are still several interesting research questions that require further investigation. We briefly sketch three of them.

(1) Throughout this paper, we assume that there are no coupling constraints, i.e., there are no upper-level constraints explicitly depending on the variables of the follower. This is a crucial assumption for the validity of the proposed methods. Otherwise, we would not be able to project the follower's variables out of the problem using the optimal-value function as it is done in Section 1. Nevertheless, developing solution methods for Γ-robust bilevel problems with coupling constraints is a reasonable aspect of future work.

(2) In this paper, we focus on interdiction problems with a monotone Γ-robust follower to obtain problem-tailored cuts. An interesting direction for future research would be to investigate if generic cuts such as, e.g., intersection cuts (Fischetti, Ljubić, et al. 2016, 2018) can be adapted to the setting described in Section 2.

(3) Finally, we would like to emphasize that we only consider uncertainties in a single packing-type constraint in the lower level. The extended formulation can easily be adapted to allow for deviations in multiple lower-level constraints. For the multi-follower formulation, however, this situation significantly increases the difficulty of the problem. This is due to the assumption regarding the ordering of the indices, which is required to exploit the results in Bertsimas and Sim (2003). In general, it is not possible to order the indices such that the deviations in the constraint coefficients are non-increasing if there are multiple uncertain constraints. Thus, this aspect is left for future research.

Acknowledgements

The third author thanks the DFG for their support within projects A05 and B08 in CRC TRR 154.

References

Álvarez-Miranda, E., E. Fernández, and I. Ljubić (2015). “The Recoverable Robust Facility Location Problem.” In: *Transportation Research Part B: Methodological* 79, pp. 93–120. DOI: `10.1016/j.trb.2015.06.001`.

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

Arroyo, J. M. (2010). “Bilevel programming applied to power system vulnerability analysis under multiple contingencies.” In: *IET Generation, Transmission & Distribution* 4.2, pp. 178–190. DOI: `10.1049/iet-gtd.2009.0098`.

Bazgan, C., S. Toubaline, and D. Vanderpooten (2013). “Critical edges/nodes for the minimum spanning tree problem: complexity and approximation.” In: *Journal of Combinatorial Optimization* 26, pp. 178–189. DOI: `10.1007/s10878-011-9449-4`.

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`.

– (2023). “A survey on bilevel optimization under uncertainty.” In: *European Journal of Operational Research*. DOI: `10.1016/j.ejor.2023.01.008`.

Beck, Y. and M. Schmidt (2021). “A robust approach for modeling limited observability in bilevel optimization.” In: *Operations Research Letters* 49.5, pp. 752–758. DOI: `10.1016/j.orl.2021.07.010`.

Ben-Ayed, O., C. Blair, D. Boyce, and L. LeBlanc (1992). “Construction of a real-world bilevel linear programming model of the highway network design problem.” In: *Annals of Operations Research* 34, pp. 219–254. DOI: `10.1007/BF02098181`.

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

Benders, J. F. (1962). “Partitioning procedures for solving mixed-variables programming problems.” In: *Numerische Mathematik* 4, pp. 238–252. DOI: `10.1007/BF01386316`.

Bertsimas, D., D. Brown, and C. Caramanis (2010). “Theory and Applications of Robust Optimization.” In: *SIAM Review* 53. 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.

– (2004). “The Price of Robustness.” In: *Operations Research* 52.1, pp. 35–53. DOI: 10.1287/opre.1030.0065.

Besançon, M., M. F. Anjos, and L. Brotcorne (2019). *Near-optimal Robust Bilevel Optimization.* URL: https://arxiv.org/pdf/1908.04040.pdf.

Besançon, M., M. F. Anjos, and L. Brotcorne (2021). “Complexity of near-optimal robust versions of multilevel optimization problems.” In: *Optimization Letters* 15, pp. 2597–2610. DOI: 10.1007/s11590-021-01754-9.

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

Bolusani, S., S. Coniglio, T. K. Ralphs, and S. Tahernejad (2020). “A Unified Framework for Multistage Mixed Integer Linear Optimization.” In: *Bilevel Optimization: Advances and Next Challenges.* Ed. by S. Dempe and A. Zemkoho. Cham: Springer International Publishing, pp. 513–560. DOI: 10.1007/978-3-030-52119-6_18.

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.

Buchheim, C. and D. Henke (2020). *The bilevel continuous knapsack problem with uncertain follower’s objective.* URL: https://arxiv.org/abs/1903.02810.

Buchheim, C., D. Henke, and F. Hommelsheim (2021). *On the Complexity of Robust Bilevel Optimization With Uncertain Follower’s Objective.* URL: https://arxiv.org/abs/2105.08378.

Burtscheidt, J. and M. Claus (2020). “Bilevel Linear Optimization Under Uncertainty.” In: *Bilevel Optimization: Advances and Next Challenges.* Ed. by S. Dempe and A. Zemkoho. Springer International Publishing, pp. 485–511. DOI: 10.1007/978-3-030-52119-6_17.

Burtscheidt, J., M. Claus, and S. Dempe (2020). “Risk-Averse Models in Bilevel Stochastic Linear Programming.” In: *SIAM Journal on Optimization* 30.1, pp. 377–406. DOI: 10.1137/19M1242240.

Caprara, A., M. Carvalho, A. Lodi, and G. J. Woeginger (2013). “A Complexity and Approximability Study of the Bilevel Knapsack Problem.” In: *Integer Programming and Combinatorial Optimization.* Ed. by M. Goemans and J. Correa. Vol. 7801. IPCO 2013. Springer, Berlin, Heidelberg, pp. 98–109. DOI: 10.1007/978-3-642-36694-9_9.

– (2016). “Bilevel Knapsack with Interdiction Constraints.” In: *INFORMS Journal on Computing* 28.2, pp. 319–333. DOI: 10.1287/ijoc.2015.0676.

Chuong, T. D. and V. Jeyakumar (2017). “Finding Robust Global Optimal Values of Bilevel Polynomial Programs with Uncertain Linear Constraints.” In: *Journal of Optimization Theory and Applications* 173.2, pp. 683–703. DOI: 10.1007/s10957-017-1069-4.

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.

Della Croce, F. and R. Scatamacchia (2020). “An exact approach for the bilevel knapsack problem with interdiction constraints and extensions.” In: *Mathematical Programming* 183, pp. 249–281. DOI: 10.1007/s10107-020-01482-5.

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

Dempe, S. and A. B. Zemkoho (2012). “Bilevel road pricing: theoretical analysis and optimality conditions.” In: *Annals of Operations Research* 196, pp. 223–240. DOI: 10.1007/s10479-011-1023-z.

Dempe, S., S. Ivanov, and A. Naumov (2017). “Reduction of the bilevel stochastic optimization problem with quantile objective function to a mixed-integer problem.” In: *Applied Stochastic Models in Business and Industry* 33.5, pp. 544–554. DOI: 10.1002/asmb.2254.

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`.

DeNegre, S. T. and T. K. Ralphs (2009). “A Branch-and-cut Algorithm for Integer Bilevel Linear Programs.” In: *Operations research and cyber-infrastructure*. Springer, pp. 65–78. DOI: `10.1007/978-0-387-88843-9_4`.

Fischetti, M., I. Ljubić, M. Monaci, and M. Sinnl (2016). “Intersection Cuts for Bilevel Optimization.” In: *Integer Programming and Combinatorial Optimization*. Ed. by Q. Louveaux and M. Skutella. IPCO 2016. Springer, pp. 77–88. DOI: `10.1007/978-3-319-33461-5_7`.

– (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`.

– (2018). “On the use of intersection cuts for bilevel optimization.” In: *Mathematical Programming* 172, pp. 77–103. DOI: `10.1007/s10107-017-1189-5`.

– (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`.

Fortuny-Amat, J. and B. McCarl (1981). “A Representation and Economic Interpretation of a Two-Level Programming Problem.” In: *The Journal of the Operational Research Society* 32.9, pp. 783–792. DOI: `10.1057/jors.1981.156`.

Furini, F., I. Ljubić, E. Malaguti, and P. Paronuzzi (2020). “On Integer and Bilevel Formulations for the k-Vertex Cut Problem.” In: *Mathematical Programming Computation* 12, pp. 133–164. DOI: `10.1007/s12532-019-00167-1`.

– (2021). “Casting light on the hidden bilevel combinatorial structure of the capacitated vertex separator problem.” In: *Operations Research*. DOI: `10.1287/opre.2021.2110`.

Furini, F., I. Ljubić, P. S. Segundo, and Y. Zhao (2021). “A branch-and-cut algorithm for the Edge Interdiction Clique Problem.” In: *European Journal of Operational Research* 294.1, pp. 54–69. DOI: `10.1016/j.ejor.2021.01.030`.

Geoffrion, A. M. (1972). “Generalized Benders decomposition.” In: *Journal of Optimization Theory and Applications* 10, pp. 237–260. DOI: `10.1007/BF00934810`.

Goh, J. and M. Sim (2010). “Distributionally Robust Optimization and Its Tractable Approximations.” In: *Operations Research* 58.4-part-1, pp. 902–917. DOI: `10.1287/opre.1090.0795`.

Golden, B. (1978). “A problem in network interdiction.” In: *Naval Research Logistics Quarterly* 25.4. DOI: `10.1002/nav.3800250412`.

Grimm, V., L. Schewe, M. Schmidt, and G. Zöttl (2019). “A Multilevel Model of the European Entry-Exit Gas Market.” In: *Mathematical Methods of Operations Research* 89.2, pp. 223–255. DOI: `10.1007/s00186-018-0647-z`.

Haghighat, H. (2014). “Strategic offering under uncertainty in power markets.” In: *International Journal of Electrical Power & Energy Systems* 63, pp. 1070–1077. DOI: `10.1016/j.ijepes.2014.05.049`.

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. (1999). “System Interdiction and Defense.” PhD thesis. Naval Postgraduate School Monterey, California. URL: `https://apps.dtic.mil/sti/pdfs/ADA361997.pdf`.

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

Ivanov, S. (2018). “A Bilevel Stochastic Programming Problem with Random Parameters in the Follower’s Objective Function.” In: *Journal of Applied and Industrial Mathematics* 12.4, pp. 658–667. DOI: `10.1134/S1990478918040063`.

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*. DOI: 10.1016/j.ejco.2021.100007. Forthcoming.

Labbé, M., P. Marcotte, and G. Savard (1998). “A Bilevel Model of Taxation and Its Application to Optimal Highway Pricing.” In: *Management Science* 44.12-part 1. DOI: 10.1287/mnsc.44.12.1608.

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. DOI: 10.1287/mnsc.45.3.414.

Migdalas, A. (1995). “Bilevel programming in traffic planning: Models, methods and challenge.” In: *Journal of Global Optimization* 7, pp. 381–405. DOI: 10.1007/BF01099649.

Moore, J. T. and J. F. Bard (1990). “The mixed integer linear bilevel programming problem.” In: *Operations Research* 38.5, pp. 911–921. DOI: 10.1287/opre.38.5.911.

Pajouh, F. M. (2020). “Minimum cost edge blocker clique problem.” In: *Annals of Operations Research* 294, pp. 345–376. DOI: 10.1007/s10479-019-03315-x.

Pajouh, F. M., V. Boginski, and E. L. Pasiliao (2014). “Minimum vertex blocker clique problem.” In: *Networks* 64.1. DOI: 10.1002/net.21556.

Pajouh, F. M., J. L. Walteros, V. Boginski, and E. L. Pasiliao (2015). “Minimum edge blocker dominating set problem.” In: *European Journal of Operational Research* 247.1, pp. 16–26. DOI: 10.1016/j.ejor.2015.05.037.

Shi, X., O. Prokopyev, and T. K. Ralphs (2020). *Mixed Integer Bilevel Optimization with k-optimal Follower: A Hierarchy of Bounds*. Tech. rep. URL: http://www.optimization-online.org/DB_FILE/2020/06/7874.pdf.

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

Smith, J. C. and Y. Song (2020). “A survey of network interdiction models and algorithms.” In: *European Journal of Operational Research* 283.3, pp. 797–811. DOI: 10.1016/j.ejor.2019.06.024.

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. and T. K. Ralphs (2020). *Valid Inequalities for Mixed Integer Bilevel Linear Optimization Problems*. Tech. rep. URL: https://engineering.lehigh.edu/sites/engineering.lehigh.edu/files/_DEPARTMENTS/ise/pdf/tech-papers/20/20T_013.pdf.

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.

Tang, Y., J. P. Richard, and J. C. Smith (2016). “A class of algorithms for mixed-integer bilevel min–max optimization.” In: *Journal of Global Optimization* 66, pp. 225–262. DOI: 10.1007/s10898-015-0274-7.

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.

Xu, P. and L. Wang (2014). “An exact algorithm for the bilevel mixed integer linear programming problem under three simplifying assumptions.” In: *Computers & Operations Research* 41, pp. 309–318. DOI: 10.1016/j.cor.2013.07.016.

Yanıkoğlu, İ. and D. Kuhn (2018). “Decision Rule Bounds for Two-Stage Stochastic Bilevel Programs.” In: *SIAM Journal on Optimization* 28.1, pp. 198–222. DOI: 10.1137/16M1098486.

Zare, M. H., O. A. Prokopyev, and D. Sauré (2020). “On Bilevel Optimization with Inexact Follower.” In: *Decision Analysis* 17.1. DOI: 10.1287/deca.2019.0392.

Zeng, B., H. Dong, R. Sioshansi, F. Xu, and M. Zeng (2020). “Bilevel Robust Optimization of Electric Vehicle Charging Stations With Distributed Energy Resources.” In: *IEEE Transactions on Industry Applications* 56.5, pp. 5836–.5847. DOI: `10.1109/TIA.2020.2984741`.

Zenklusen, R., B. Ries, C. Picouleau, D. de Werra, M.-C. Costa, and C. Bentz (2009). “Blockers and transversals.” In: *Discrete Mathematics* 309.13, pp. 4306–4314. DOI: `10.1016/j.disc.2009.01.006`.

Appendix A. Omitted Proofs

Proof of Proposition 1. Let x be a feasible upper-level decision and let (y^*, z^*, θ^*) be an optimal solution of the x-parameterized lower-level problem (7). By Assumption 2, there is a vector $u \in \mathbb{R}^{n_y}_+$ such that $Y(x) \subseteq [0, u_1] \times \cdots \times [0, u_{n_y}]$ and, in particular, $0 \leq y^*_i \leq u_i$ holds for all $i \in [n_y]$. Consequently, we obtain

$$d^\top y^* \geq \sum_{i=1}^{n_y} \min\{d_i, 0\} u_i.$$

By (7b), $\theta^* \geq 0$, and $y^* \leq u$, an optimal solution of the follower further satisfies

$$z^*_i = \max\{\Delta d_i y^*_i - \theta^*, 0\} \leq \Delta d_i y^*_i \leq \Delta d_i u_i$$

for all $i \in [n_y]$. Finally, we have $\theta^* \leq \max\{\Delta d_i u_i \colon i \in [n_y]\}$ due to (7b) and the optimality of θ^*, which concludes the proof. □

Proof of Proposition 2. This follows immediately from the first part of the proof of Proposition 1. □

Proof of Proposition 3. Let x be a feasible upper-level decision and let $\ell \in \{1, \dots, n_y + 1\}$ be arbitrary but fixed. Further, let y^ℓ be an optimal solution of the ℓth sub-problem (10) that is parameterized in x. By (9) and since $y^\ell \in \{0, 1\}^{n_y}$, we obtain

$$\Phi_d(x) \geq -\Gamma_d \Delta d_\ell + \sum_{i=1}^{n_y} \tilde{d}(\ell)_i y^\ell_i \geq -\Gamma_d \Delta d_\ell + \sum_{i=1}^{n_y} \min\left\{\tilde{d}(\ell)_i, 0\right\},$$

which concludes the proof. □

Proof of Proposition 4. Let x be a feasible upper-level decision. Further, let $y \in Y(x)$ and let $y' \in \mathcal{Y}$ be such that $y' \leq y$ holds. Due to $B \in \mathbb{R}^{m \times n}_+$, we obtain

$$\begin{aligned} By' &\leq By \leq b, \\ y'_i &\leq y_i \leq u_i(1 - x_i), \qquad i \in [n], \end{aligned}$$

Thus, y' is a feasible follower’s decision for the given x, i.e., $y' \in Y(x)$. □

Proof of Proposition 5. Let x be a feasible upper-level decision. First, we show that the extended formulation (14) satisfies the monotonicity property. To this end, let (y, z, θ) be a feasible follower’s decision for Problem (14) for the given x. Further, let $y' \in \mathcal{Y}$ be such that $y' \leq y$ holds. Due to $B \in \mathbb{R}^{m \times n}_+$ and $\Delta d_i \geq 0$ for all $i \in [n]$, we obtain

$$\begin{aligned} z_i + \theta &\geq \Delta d_i y_i \geq \Delta d_i y'_i, \quad i \in [n], \\ By' &\leq By \leq b, \\ y'_i &\leq y_i \leq u_i(1 - x_i), \quad i \in [n], \end{aligned}$$

i.e., the follower’s decision (y', z, θ) is feasible for Problem (14) for the given x. Second, we show that each sub-problem (15) satisfies the monotonicity property. Note that there is no need to specify $\ell \in \{\Gamma_d, \dots, n + 1\}$ since the feasible set of (15) does not depend on ℓ. For the given x, let y be a feasible follower’s decision for sub-problem (15). Further, let $y' \in \{0, 1\}^n$ be such that $y' \leq y$. Since we restrict ourselves to binary follower’s variables in the multi-follower case, we have valid upper bounds $u_i = 1$ for all $i \in [n]$. Applying the same arguments as before, y' is feasible for sub-problem (15) for the given x. Consequently, the Γ_d-robust lower-level problems (14) and (15) satisfy the monotonicity property. □

Proof of Proposition 6. Let x be a feasible upper-level decision. For notational convenience, let $\Psi(x)$ and Ψ denote the feasible set of Problem (14) and (16), respectively. Further, let (y^*, z^*, θ^*) be an optimal solution of Problem (14) for the given leader's decision x. Then, (y^*, z^*, θ^*) is also feasible for Problem (16), i.e., $(y^*, z^*, \theta^*) \in \Psi$. In particular, $y_i^* x_i = 0$ holds for all $i \in [n]$, i.e., both problems have the same objective function value for (y^*, z^*, θ^*). Thus, we obtain

$$\begin{aligned}&\max_{y,\theta,z} \left\{ \sum_{i=1}^{n} d_i y_i - \Gamma_d \theta - \sum_{i=1}^{n} z_i \colon (y, z, \theta) \in \Psi(x) \right\} \\ \leq\ &\max_{y,\theta,z} \left\{ \sum_{i=1}^{n} d_i y_i (1 - x_i) - \Gamma_d \theta - \sum_{i=1}^{n} z_i \colon (y, z, \theta) \in \Psi \right\}.\end{aligned} \tag{30}$$

Let $(\hat{y}, \hat{z}, \hat{\theta})$ be an optimal solution of Problem (16) for the given leader's decision x. Without loss of generality, we assume that there is exactly one item $k \in [n]$ for which the interdiction constraint $\hat{y}_k \leq u_k(1 - x_k)$ is not satisfied, i.e., $\hat{y}_k \geq 1$ and $x_k = 1$. Otherwise, we repeat the following as long as there are no more items left that violate the interdiction constraint. We consider the alternative follower's decision $(y', z', \hat{\theta})$ with

$$y_i' = \begin{cases} \hat{y}_i, & i \in [n] \setminus \{k\}, \\ 0, & i = k, \end{cases}$$

and

$$z_i' = \begin{cases} \hat{z}_i, & i \in [n] \setminus \{k\}, \\ 0, & i = k. \end{cases}$$

By construction, $(y', z', \hat{\theta})$ is feasible for Problem (16) and satisfies all interdiction constraints. Moreover, we obtain

$$\begin{aligned}&\max_{y,\theta,z} \left\{ \sum_{i=1}^{n} d_i y_i (1 - x_i) - \Gamma_d \theta - \sum_{i=1}^{n} z_i \colon (y, z, \theta) \in \Psi \right\} \\ &= d_k \hat{y}_k \underbrace{(1 - x_k)}_{=0} + \sum_{i \in [n] \setminus \{k\}} d_i \underbrace{\hat{y}_i}_{=y_i'} (1 - x_i) - \Gamma_d \hat{\theta} - \hat{z}_k - \sum_{i \in [n] \setminus \{k\}} \underbrace{\hat{z}_i}_{=z_i'} \\ &= d_k \underbrace{y_k'}_{=0} (1 - x_k) + \sum_{i \in [n] \setminus \{k\}} d_i y_i' (1 - x_i) - \Gamma_d \hat{\theta} - \underbrace{z_k'}_{=0} - \hat{z}_k - \sum_{i \in [n] \setminus \{k\}} z_i' \\ &\leq \sum_{i=1}^{n} d_i y_i' (1 - x_i) - \Gamma_d \hat{\theta} - \sum_{i=1}^{n} z_i',\end{aligned}$$

i.e., the alternative follower's decision is optimal for Problem (16). In particular, it is also feasible for Problem (14), i.e., $(y', z', \hat{\theta}) \in \Psi(x)$, and we have $y_i' x_i = 0$ for all $i \in [n]$. Hence, we obtain

$$\begin{aligned}
&\max_{y,\theta,z}\left\{\sum_{i=1}^{n} d_i y_i(1-x_i) - \Gamma_d\theta - \sum_{i=1}^{n} z_i \colon (y,z,\theta)\in\Psi\right\}\\
&= \sum_{i=1}^{n} d_i y'_i(1-x_i) - \Gamma_d\hat{\theta} - \sum_{i=1}^{n} z'_i\\
&= \sum_{i=1}^{n} d_i y'_i - \Gamma_d\hat{\theta} - \sum_{i=1}^{n} z'_i\\
&\le \max_{y,\theta,z}\left\{\sum_{i=1}^{n} d_i y_i - \Gamma_d\theta - \sum_{i=1}^{n} z_i \colon (y,z,\theta)\in\Psi(x)\right\}.
\end{aligned} \tag{31}$$

Due to (30) and (31), Problem (14) and (16) admit the same optimal value. □

Proof of Proposition 7. Let x be a feasible upper-level decision and let $\ell \in \{\Gamma_d,\dots,n+1\}$ be arbitrary but fixed. Further, let y^* be an optimal solution of the ℓth sub-problem (15) for the given leader's decision x. Then, y^* is also feasible for the ℓth sub-problem (17), i.e., $y^*\in Y$. In particular, $y^*_i x_i = 0$ holds for all $i\in[n]$, i.e., both sub-problems have the same objective function value for y^*. Thus, we obtain

$$\max_y\left\{\sum_{i=1}^{n}\tilde{d}(\ell)_i y_i \colon y\in Y(x)\right\} \le \max_y\left\{\sum_{i=1}^{n}\tilde{d}(\ell)_i y_i(1-x_i)\colon y\in Y\right\}. \tag{32}$$

Let $\hat{y}$ be an optimal solution of the ℓth sub-problem (17) for the given leader's decision x. Without loss of generality, suppose there is exactly one item $k\in[n]$ for which the interdiction constraint $\hat{y}_k \le 1 - x_k$ is not satisfied, i.e., $\hat{y}_k = 1 = x_k$. Then, we consider the alternative follower's decision

$$y'_i = \begin{cases}\hat{y}_i, & i\in[n]\setminus\{k\},\\ 0, & i = k.\end{cases}$$

By construction, y' is feasible for the ℓth sub-problem (17) and satisfies all interdiction constraints. Moreover, we obtain

$$\begin{aligned}
\max_y\left\{\sum_{i=1}^{n}\tilde{d}(\ell)_i y_i(1-x_i)\colon y\in Y\right\} &= \tilde{d}(\ell)_k\hat{y}_k\underbrace{(1-x_k)}_{=0} + \sum_{i\in[n]\setminus\{k\}}\tilde{d}(\ell)_i\underbrace{\hat{y}_i}_{=y'_i}(1-x_i)\\
&= \tilde{d}(\ell)_k\underbrace{y'_k}_{=0}(1-x_k) + \sum_{i\in[n]\setminus\{k\}}\tilde{d}(\ell)_i y'_i(1-x_i)\\
&= \sum_{i=1}^{n}\tilde{d}(\ell)_i y'_i(1-x_i),
\end{aligned}$$

i.e., the alternative follower's decision is optimal for Problem (17). In particular, it is also feasible for Problem (15), i.e., $y'\in Y(x)$, and we have $y'_i x_i = 0$ for all $i\in[n]$. Hence, we obtain

$$\begin{aligned}
\max_y\left\{\sum_{i=1}^{n}\tilde{d}(\ell)_i y_i(1-x_i)\colon y\in Y\right\} &= \sum_{i=1}^{n}\tilde{d}(\ell)_i y'_i(1-x_i)\\
&= \sum_{i=1}^{n}\tilde{d}(\ell)_i y'_i\\
&\le \max_y\left\{\sum_{i=1}^{n}\tilde{d}(\ell)_i y_i\colon y\in Y(x)\right\}.
\end{aligned} \tag{33}$$

Due to (32) and (33), Problem (15) and (17) admit the same optimal value. □

Proof of Theorem 2. Let $(x, \eta) \in X \times \mathbb{R}$ be a given leader's decision. Due to the validity of the proposed cuts, it suffices to show that the feasibility of (x, η) w.r.t. either the interdiction cuts (18) or (19) implies $\eta \geq \Phi(x)$. To this end, suppose that (x, η) satisfies $Ax \geq a$ and either the interdiction cuts

$$\eta \geq \sum_{i=1}^{n} d_i \hat{y}_i (1 - x_i) - \Gamma_d \hat{\theta} - \sum_{i=1}^{n} \hat{z}_i \quad \text{for all } (\hat{y}, \hat{z}, \hat{\theta}) \in \hat{\Psi}$$

or

$$\eta \geq -\Gamma_d \Delta d_\ell + \sum_{i=1}^{n} \tilde{d}(\ell)_i \hat{y}_i (1 - x_i) \quad \text{for all } \hat{y} \in \hat{Y},\ \ell \in \{\Gamma_d, \dots, n+1\}.$$

By Propositions 6 and 7, this is equivalent to

$$\eta \geq \max_{\hat{y}, \hat{\theta}, \hat{z}} \left\{ \sum_{i=1}^{n} d_i \hat{y}_i (1 - x_i) - \Gamma_d \hat{\theta} - \sum_{i=1}^{n} \hat{z}_i \colon (\hat{y}, \hat{z}, \hat{\theta}) \in \hat{\Psi} \right\} = \Phi(x)$$

and

$$\eta \geq \max_{\ell \in \{\Gamma_d, \dots, n+1\}} \left\{ -\Gamma_d \Delta d_\ell + \max_{\hat{y} \in \hat{Y}} \left\{ \sum_{i=1}^{n} \tilde{d}(\ell)_i \hat{y}_i (1 - x_i) \right\} \right\} = \Phi(x),$$

which concludes the proof. □

Proof of Proposition 8. This follows immediately from $d_i > 0$ and from the fact that $x_i \in \{0, 1\}$ implies $\hat{y}_i (1 - x_i) \leq y'_i (1 - x_i)$ for all $i \in [n]$. □

Proof of Proposition 9. Let (x, η) be feasible for Problem (2) with the lower-level optimal-value function (3). Further, let $\ell \in \{\Gamma_d, \dots, n+1\}$ be arbitrary but fixed. Due to $x_i \in \{0, 1\}$ for all $i \in [n]$, we have $\tilde{d}(\ell)_i (1 - x_i) \leq 0$ for all $i \notin D_+^\ell$. Hence, all follower's variables y_i with $i \notin D_+^\ell$ could be omitted in this sub-problem, i.e., we obtain

$$\Phi^\ell(x) = \max_{y} \left\{ \sum_{i \in D_+^\ell} \tilde{d}(\ell)_i y_i (1 - x_i) \colon y \in \hat{Y} \right\}.$$

The validity of the new interdiction cuts (20) can be shown in a similar way as it is done in Section 3. In particular,

$$\begin{aligned}
& -\Gamma_d \Delta d_\ell + \sum_{i=1}^{n} \tilde{d}(\ell)_i \hat{y}_i (1 - x_i) \\
= & -\Gamma_d \Delta d_\ell + \sum_{i \in D_+^\ell} \tilde{d}(\ell)_i \hat{y}_i (1 - x_i) + \sum_{i \notin D_+^\ell} \underbrace{\tilde{d}(\ell)_i \hat{y}_i (1 - x_i)}_{\leq 0} \\
\leq & -\Gamma_d \Delta d_\ell + \sum_{i \in D_+^\ell} \tilde{d}(\ell)_i \hat{y}_i (1 - x_i)
\end{aligned}$$

holds for all $\hat{y} \in \hat{Y}$ and $\ell \in \{\Gamma_d, \dots, n+1\}$. Thus, the cuts (20) dominate the basic interdiction cuts (19). □

Proof of Proposition 10. This can be shown in analogy to the proof of Proposition 8.

(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`

Article 4

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

Yasmine Beck, Ivana Ljubić, and Martin Schmidt
Preprint
https://optimization-online.org/?p=26186

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

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

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

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

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

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

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 $x^\ell$ to the deterministic min-max problem
          $v_\ell \leftarrow \min_{x \in X} \{c^\top x + \Phi_\ell(x)\}.$        ($\ell$-Min-Max)
 4:   Set $L \leftarrow \max\{L, v_\ell\}$.
 5:   if $U \leq L$ then
 6:     return $x^*$, $L$, $U$
 7:   Solve the $x^\ell$-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*

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

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

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

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 $x^* \leftarrow$ None, $L \leftarrow -\infty$, and $U \leftarrow \infty$.
 2: for all $\ell \in \mathcal{L}$ do
 3:     Compute a solution $x^\ell$ to the deterministic min-max problem
            $v_\ell \leftarrow \min_{x \in X} \left\{ c^\top x + \Phi_\ell(x) \right\}.$     ($\ell$-Min-Max)
 4: Sort the indices such that $v_{\ell_1} \le v_{\ell_2} \le \cdots \le v_{\ell_{|\mathcal{L}|}}$ holds and set $L \leftarrow v_{\ell_{|\mathcal{L}|}}$.
 5: Set $i \leftarrow 1$.
 6: while $i \le |\mathcal{L}|$ and $L < U$ do
 7:     Solve the $x^{\ell_i}$-parameterized $\Gamma$-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*

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

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

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

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 $c^\top x^* + d^\top y^* \ge c^\top x^\ell + d^\top y^\ell$ 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 sub-problems (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 $(x^\ell, y^\ell)$ to the bilevel problem
            $\min_{x,y} \quad c^\top x + d^\top y$
            $\text{s.t.} \quad x \in X,$                                        ($\ell$-BMIP)
            $\quad y \in \arg\max_{y' \in Y(x)} \left\{ -\Gamma \Delta f_\ell + \tilde{f}(\ell)^\top y' \right\}.$
 4: Sort the indices such that
        $c^\top x^{\ell_1} + d^\top y^{\ell_1} \le c^\top x^{\ell_2} + d^\top y^{\ell_2} \le \dots \le c^\top x^{\ell_{|\mathcal{L}|}} + d^\top y^{\ell_{|\mathcal{L}|}}$
    holds and set $L \leftarrow c^\top x^{\ell_1} + d^\top y^{\ell_1}$.
 5: Set $i \leftarrow 1$.
 6: while $i \le |\mathcal{L}|$ and $L < U$ do
 7:     Solve the $x^{\ell_i}$-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 $(x^{\ell_i}, \hat{y})$, $i \in \{1, \dots, |\mathcal{L}|\}$, obtained from Line 7 of the algorithm satisfies $x^{\ell_i} \in X$, $\hat{y} \in Y(x^{\ell_i})$, and

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

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 $c^\top x^* + d^\top y^* \leq c^\top x^\ell + d^\top y^\ell$ *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 $L \leq U$ holds. By assumption, we further have $c^\top x^* + d^\top y^* \leq c^\top x^\ell + d^\top y^\ell$ for all $\ell \in \mathcal{L}$, which is equivalent to

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

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* $x^k = x^\ell$, $d^\top y^k \leq d^\top y^\ell$ *and* $\tilde{f}(\ell)^\top y^k \geq \tilde{f}(\ell)^\top y^\ell$ *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

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

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 $c^\top x^k + d^\top y^k \leq c^\top x^\ell + d^\top y^\ell$ 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*

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

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

$$k \leftarrow \operatorname*{arg\,max}_{\ell \in \mathcal{L}} \{\Phi_\ell(\hat{x})\}$$

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 ($\ell$-BMIP), an index set $\mathcal{L}$ as in (5), an index $\ell_i \in \mathcal{L}$
Output: A solution $\hat{y}$ to the $x^{\ell_i}$-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 $\ell$th 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

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

in which all parameters are assumed to be non-negative integers, i.e., $B, C \in \mathbb{Z}_{\ge 0}$, and $f, v, w \in \mathbb{Z}^n_{\ge 0}$. 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

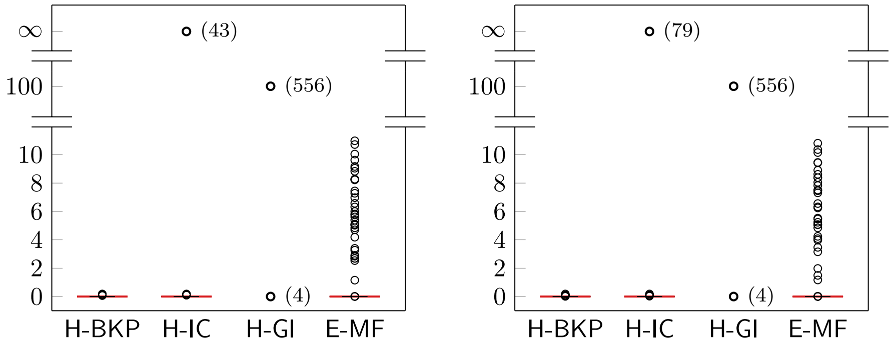

Figure 1. Box-plots of the optimality gaps (in %) for the approaches H-BKP, H-IC, H-GI, and E-MF in the min-max setting with integer (left) and continuous deviations (right).

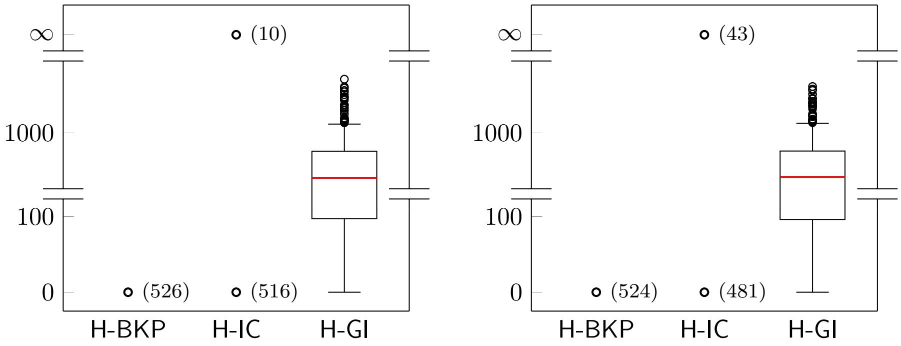

Figure 2. Box-plots of the ex-post optimality gaps (in %) for the approaches H-BKP, H-IC, and H-GI for the 526 instances with integer deviations (left) and the 524 instances with continuous deviations (right) that have been solved to global optimality by E-MF. Values above 100 % are shown on a log-scaled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

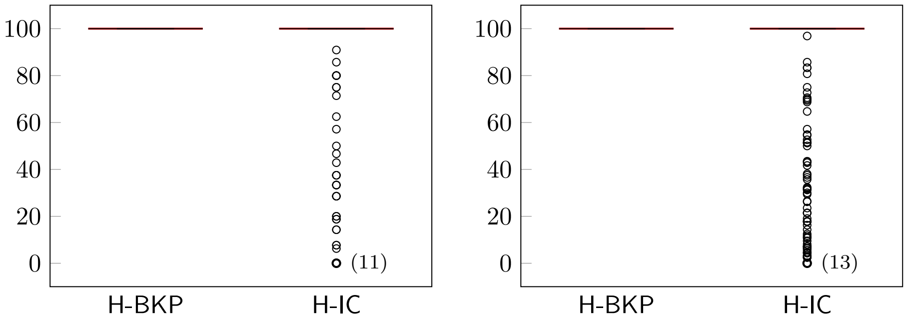

Figure 3. Box-plots of the percentages of solved sub-problems (out of the total number of sub-problems to be solved) for the approaches H-BKP and H-IC in the min-max setting with integer deviations (left) and continuous deviations (right).

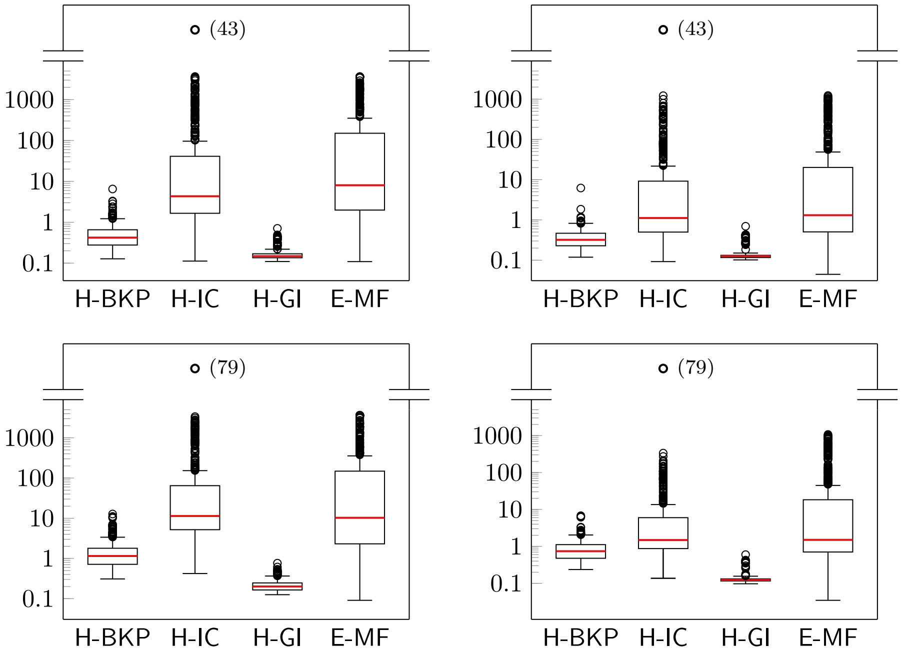

Figure 4. Box-plots of the sequential (left) and the idealized parallel runtimes (right) for the approaches H-BKP, H-IC, H-GI, and E-MF in the min-max setting with integer deviations (top) and continuous deviations (bottom). Sequential and idealized parallel runtimes (in s) are depicted on a log-scaled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

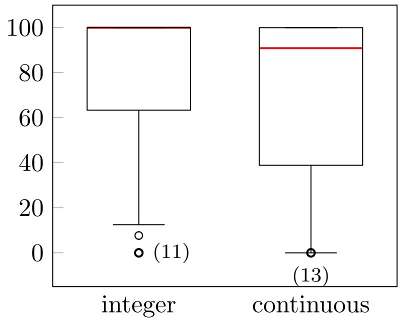

Figure 5. Box-plots of the percentages of solved sub-problems (out of the total number of sub-problems to be solved) for H in the general bilevel setting with integer and continuous deviations.

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

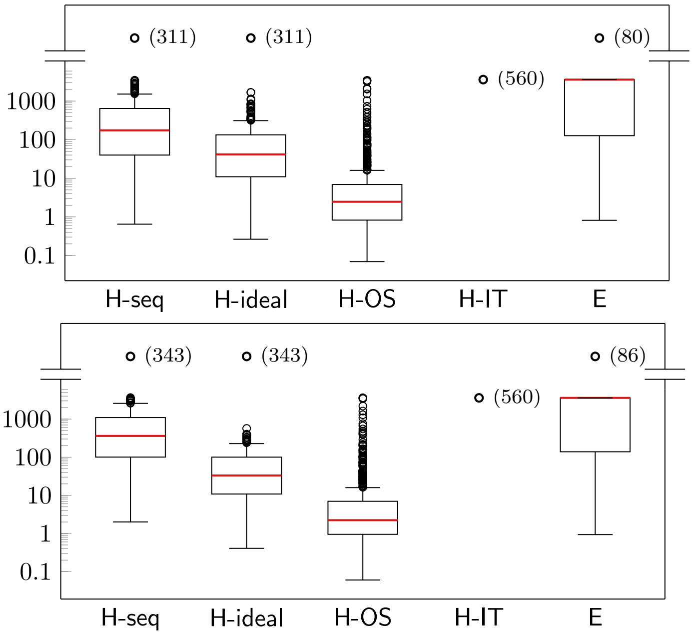

FIGURE 6. Box-plots of the runtimes for the approaches in the general bilevel setting with integer (top) and continuous deviations (bottom). Runtimes (in s) are depicted on a log-scaled y-axis.

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

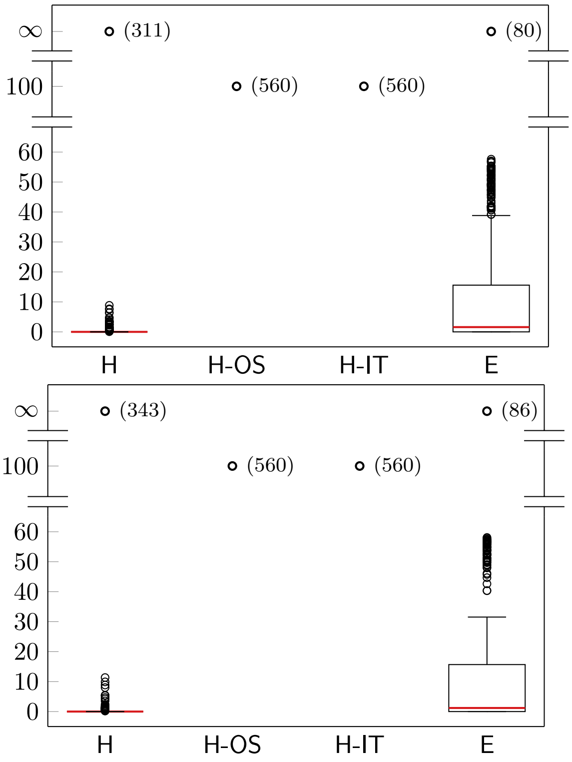

FIGURE 7. Box-plots of the optimality gaps (in %) for the approaches in the general bilevel setting with integer (top) and continuous deviations (bottom).

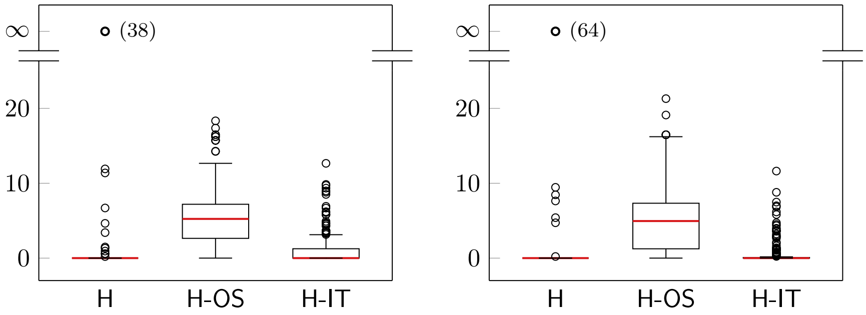

FIGURE 8. Box-plots of the ex-post optimality gaps (in %) for the approaches H, H-OS, and H-IT for the 236 instances with integer (left) and the 230 instances with continuous deviations (right) that have been solved to global optimality by E.

(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

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

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,

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

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 $x \geq x^\ell$ 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 $y_i \leq 1 - x_i = 1 - x_i^\ell$ 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

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

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

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

i.e., x solves (ℓ-Min-Max) as well.

Proof of Lemma 4. By Lemma 1, we have

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

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

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

Moreover, we have

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

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`

Article 5

A Toll-Setting Problem with Robust Wardrop Equilibrium Conditions Under Budgeted Uncertainty

Yasmine Beck, Martine Labbé, and Martin Schmidt
Preprint
https://optimization-online.org/?p=26949

A Toll-Setting Problem with Robust Wardrop Equilibrium Conditions Under Budgeted Uncertainty

Yasmine Beck, Martine Labbé, Martin Schmidt

Abstract. We consider the problem of determining optimal tolls in a traffic network in which a toll-setting authority aims to maximize revenues and the users of the network act in the sense of Wardrop's user equilibrium. The setting is modeled as a mathematical problem with equilibrium constraints and a mixed-integer, nonlinear, and nonconvex reformulation is presented that exploits binary variables and big-M constants. We prove existence of optimal solutions to this problem, derive correct big-Ms, and provide valid inequalities. Moreover, we consider the setting in which the network users hedge against uncertainties regarding their travel costs. We model this setting using robust Wardrop equilibria under budgeted uncertainty and prove existence of robust solutions. Finally, we present preliminary computational results to illustrate the impact of considering robust travel decisions on the revenues realized by the toll-setting authority.

1. Introduction

In traffic networks, collecting tolls is a powerful tool for network management and for influencing travel behavior. For instance, revenues generated by imposing tolls may support the maintenance of existing infrastructure or fund the construction of new roads. In addition, tolls may be used to manage traffic flow by alleviating congestion and encouraging the more efficient use of road capacity. Thus, it is evident that determining optimal tolls in a traffic network is an important aspect of transportation science. In this context, the toll-setting authority has to decide on the tolls while anticipating the reaction of the users of the traffic network, who usually try to minimize costs and time spent on travel. The overall toll-setting problem can thus be seen as a single-leader multi-follower game in which the toll-setting authority acts as the leader and the users of the traffic network act as the followers. Influential works in this context include, e.g., Brotcorne et al. (2001), Dempe and Zemkoho (2012), Dewez et al. (2008), Kalashnikov et al. (2020), and Labbé et al. (1998, 2000).

In this paper, we consider a multi-commodity traffic network in which a toll-setting authority decides on the tolls of (some of) the arcs of the network. While we consider the setting in which the toll-setting authority aims to maximize revenues by imposing tolls, we emphasize that other objective functions may be possible as well. Regarding the users of the traffic network, we assume that they act according to Wardrop's user equilibrium (Wardrop 1952; Wardrop and Whitehead 1952), minimizing their individual travel costs that are parameterized by the imposed tolls. In this paper, we do not make any assumptions about the separability of the travel costs but they are assumed to be affine-linear in the traffic flows. We model the overall toll-setting problem as a mathematical problem with equilibrium

---

Date: July 9, 2024.

2020 *Mathematics Subject Classification.* 90C26, 90C33, 90C35, 90C70.

Key words and phrases. Transportation networks, Pricing, Wardrop equilibrium, Mixed-integer nonlinear optimization, Robust optimization.

constraints (MPEC); see, e.g., Luo et al. (1996) for a general overview. In contrast to Dempe and Zemkoho (2012), who study a similar setting for separable cost functions from a theoretical point of view, we consider the toll-setting problem more from a computational perspective. To this end, we reformulate the problem as a mixed-integer, nonlinear, and nonconvex problem (nonconvex MINLP) that exploits binary variables and big-M constants. The latter can be tackled using state-of-the-art general-purpose solvers. We provide results on the existence of optimal solutions to this problem as well as valid inequalities to enhance the problem formulation. Moreover, we derive valid big-M constants.

In addition, we study the toll-setting problem in which the users of the traffic network face uncertainties regarding their travel costs, which we tackle using techniques from robust optimization; see, e.g., Ben-Tal et al. (2009), Bertsimas et al. (2011), and Soyster (1973). To this end, we pursue similar ideas compared to those in Ito (2011) and Ordóñez and Stier-Moses (2007, 2010), who also consider so-called robust Wardrop equilibria. In Ito (2011), a strictly robust setup is considered to hedge against uncertainties regarding the travel costs. The author makes the necessary continuity assumptions on the robustified travel cost functions to ensure that robust Wardrop equilibria exist. In particular, the author considers ellipsoidal uncertainty sets. In Ordóñez and Stier-Moses (2007), the authors pursue a Γ-robust approach (Bertsimas and Sim 2003; Sim 2004) to hedge against uncertain travel costs. The authors provide existence results for robust Wardrop equilibria and present a column-generation algorithm to compute them. In a follow-up paper, Ordóñez and Stier-Moses (2010) provide more extensive theoretical and computational details on this approach, along with further equilibrium concepts to hedge against uncertain travel costs. All aforementioned works have in common that the authors focus on the robust traffic assignment problem in which a path-based formulation is used to model the travelers' behavior. The modeling framework considered in this paper differs from those in Ito (2011) and Ordóñez and Stier-Moses (2007, 2010) in the following two aspects. First, we consider the problem of determining optimal tolls in a traffic network that incorporates robust Wardrop equilibria in the constraints of the problem. Second, we study robust Wardrop equilibria under budgeted uncertainty, which necessitates the use of a node-arc formulation to model the travelers' behavior. To the best of our knowledge, there are no other works in the literature that consider such a network pricing under robust Wardrop equilibria. We illustrate the impact of considering robust travel decisions on the revenues realized by the toll-setting authority through a case study on a subnetwork of the well-known Sioux Falls network (LeBlanc et al. 1975). Here, we observe that addressing uncertainties in the travel costs may significantly impact the travel behavior and, in particular, lead to increased revenues realized by imposing tolls.

The remainder of this paper is organized as follows. In Section 2, we present the overall toll-setting problem, which we model as an MPEC. In Section 3, we present an MINLP reformulation of the toll-setting problem, prove existence of optimal solutions, derive valid big-Ms, and provide valid inequalities. In Section 4, we present a robustified variant of the toll-setting problem under budgeted uncertainty. We provide an MINLP reformulation of this problem and prove existence of robust solutions. In Section 5, we present preliminary computational results to illustrate the impact of considering robust travel decisions on the revenues realized by the toll-setting authority. Finally, we conclude in Section 6.

2. The MPEC Model

We consider a traffic network that is modeled using a directed graph $\mathcal{G} = (\mathcal{N}, \mathcal{A})$ with node set $\mathcal{N}$ and arc set $\mathcal{A} \subseteq \mathcal{N} \times \mathcal{N}$. We elaborate on the graph's connectivity later in this section. In what follows, we denote $f = (f_a)_{a\in\mathcal{A}}$ as the vector of all arc flows and $\tau = (\tau_a)_{a\in\mathcal{A}}$ as the tolls imposed on the arcs of the network. The aim of the toll-setting authority is to maximize the revenues

$$\sum_{a\in\mathcal{A}} \tau_a f_a$$

that are realized by charging tolls on certain arcs of the network. Throughout this paper, we make the following assumption.

Assumption 1. *The tolls τ are subject to constraints described by a polytope $\mathcal{T} = \{\tau \in \mathbb{R}^{|\mathcal{A}|} \colon B\tau \leq b\} \neq \emptyset$ for some matrix B and a vector b of appropriate dimension.*

The constraints $B\tau \leq b$ are used to model, e.g., lower and upper bounds on the tolls or toll-free arcs. Here and in what follows, we assume that the set $\mathcal{T}$ induces a finite upper bound τ_a^+ as well as a lower bound of zero for the toll τ_a on each arc $a \in \mathcal{A}$. Arcs $a \in \mathcal{A}$ for which the set $\mathcal{T}$ imposes the upper bound $\tau_a^+ = 0$ are called toll-free arcs. All remaining arcs are called toll arcs. The overall toll-setting problem can now be stated as

$$\max_{\tau,f,x} \quad \sum_{a\in\mathcal{A}} \tau_a f_a \quad \text{s.t.} \quad \tau \in \mathcal{T}, \; (f,x) \in S(\tau). \tag{1}$$

Here, the set $S(\tau)$ is used to denote the Wardrop equilibria that are parameterized by the imposed tolls τ, which we discuss in detail in the following section. In particular, Problem (1) can be interpreted as a single-leader multi-follower problem in which the toll-setting authority acts as the leader and the users of the traffic network act as the followers. By optimizing over the tolls τ and the variables f and x, we consider the so-called optimistic approach as it is known in bilevel optimization; see, e.g., Dempe (2002). This means that, whenever there are multiple optimal route choices for the users of the network, they choose the ones that favor the leader the most w.r.t. the associated revenues. The latter is a common assumption in the literature; see, e.g., Brotcorne et al. (2001) and Labbé et al. (1998).

2.1. **Wardrop Equilibrium Conditions.** For node subsets $\mathcal{O}, \mathcal{D} \subseteq \mathcal{N}$, we denote the set of all origin-destination (OD) pairs of the network as $\mathcal{K} \subseteq \mathcal{O} \times \mathcal{D}$. For the ease of presentation, we consider a single commodity for each OD pair $k \in \mathcal{K}$. Let $x^k = (x_a^k)_{a\in\mathcal{A}} \in \mathbb{R}^{|\mathcal{A}|}$ denote the flow vector of commodity $k \in \mathcal{K}$. The vector of arc flows is then given by

$$f = \sum_{k\in\mathcal{K}} x^k \in \mathbb{R}^{|\mathcal{A}|}. \tag{2}$$

Throughout this paper, we make the following assumptions.

Assumption 2. *For every node $i \in \mathcal{N}$, there is at least one path that connects node i to each destination node $j \in \mathcal{D}$.*

Assumption 3. *For every commodity $k \in \mathcal{K}$, the travel demand $d_k \in \mathbb{R}$ is nonnegative and fixed.*

We emphasize that Assumption 2 is a standard assumption in the literature; cf., e.g., Assumption 2.A in Patriksson (2015). Let us further mention that Assumption 3 is w.l.o.g. since any elastic-demand problem can equivalently be reformulated as a fixed-demand problem; see, e.g., Dantzig et al. (1976) and Gartner (1980). For each commodity $k = (\alpha_k, \omega_k) \in \mathcal{K}$, flow conservation can now be modeled via

$$\sum_{a\in\delta^{\text{in}}(i)} x_a^k - \sum_{a\in\delta^{\text{out}}(i)} x_a^k = d_i^k, \quad i \in \mathcal{N}, \tag{3}$$

with

$$d_i^k = \begin{cases} +d_k, & i = \omega_k, \\ 0, & i \in \mathcal{N} \setminus \{\alpha_k, \omega_k\}, \\ -d_k, & i = \alpha_k. \end{cases}$$

Here, $\delta^{\text{in}}(i)$ and $\delta^{\text{out}}(i)$ denote the sets of in- and outgoing arcs of node $i \in \mathcal{N}$, respectively. Next, we elaborate on Wardrop's second[1] principle to model user-optimized behavior. It is assumed that the users of the traffic network seek to minimize their individual travel costs such that no user can reduce costs by unilaterally changing routes. This behavior can be modeled as

$$0 \leq c_a^k(f; \tau_a) + t_j^k - t_i^k \perp x_a^k \geq 0, \quad a = (i,j) \in \mathcal{A}, \; k \in \mathcal{K}. \tag{4}$$

A similar setting is, e.g., considered in Section 3.6.2 in Ferris and Pang (1997). In (4), the cost for commodity $k \in \mathcal{K}$ to travel along an arc $a \in \mathcal{A}$ is given by the function $c_a^k(f; \tau_a)$ that depends on the overall flows f and that is parameterized by the imposed toll τ_a. Moreover, t_i^k denotes the minimum cost to reach the destination of commodity $k \in \mathcal{K}$ from node $i \in \mathcal{N}$. We abbreviate $t = (t^k)_{k \in \mathcal{K}}$ with $t^k = (t_i^k)_{i \in \mathcal{N}} \in \mathbb{R}^{|\mathcal{N}|}$. To sum up, the τ-parameterized set of Wardrop equilibria is given by

$$S(\tau) := \{(f, x) \colon \exists t \text{ such that } (f, x, t) \text{ solves } (2)\text{–}(4)\}.$$

3. An MINLP Reformulation

We introduce additional binary variables $z \in \{0,1\}^{|\mathcal{A}| \cdot |\mathcal{K}|}$ to obtain a reformulation of Problem (1) that is given by

$$\max_{\tau, f, x, t, z} \quad \sum_{a \in \mathcal{A}} \tau_a f_a \tag{5a}$$

$$\text{s.t.} \quad \tau \in \mathcal{T}, \quad f = \sum_{k \in \mathcal{K}} x^k, \tag{5b}$$

$$\sum_{a \in \delta^{\text{in}}(i)} x_a^k - \sum_{a \in \delta^{\text{out}}(i)} x_a^k = d_i^k, \qquad i \in \mathcal{N}, \; k \in \mathcal{K}, \tag{5c}$$

$$x_a^k \geq 0, \quad c_a^k(f; \tau_a) + t_j^k - t_i^k \geq 0, \qquad a = (i,j) \in \mathcal{A}, \; k \in \mathcal{K}, \tag{5d}$$

$$c_a^k(f; \tau_a) + t_j^k - t_i^k \leq M_a^k (1 - z_a^k), \qquad a = (i,j) \in \mathcal{A}, \; k \in \mathcal{K}, \tag{5e}$$

$$x_a^k \leq M_a^k z_a^k, \qquad a \in \mathcal{A}, \; k \in \mathcal{K}, \tag{5f}$$

$$z_a^k \in \{0, 1\}, \qquad a \in \mathcal{A}, \; k \in \mathcal{K}. \tag{5g}$$

Problem (5) is as a mixed-integer nonlinear problem (MINLP) due to bilinearities in the objective function as well as possible nonlinearities in the travel cost functions $c_a^k(f; \tau_a)$, $a \in \mathcal{A}$, $k \in \mathcal{K}$. By construction, Problem (5) is equivalent to the toll-setting problem (1) if the big-M constants M_a^k, $a \in \mathcal{A}$, $k \in \mathcal{K}$, are chosen sufficiently large. To obtain such constants, however, we need further knowledge about the travel cost functions $c_a^k(f; \tau_a)$. In this paper, we assume that the travel cost functions are affine-linear in the flows f so that, in Problem (5), we consider a bilinear objective that is optimized over mixed-integer and linear constraints. We acknowledge that this is a strong assumption. However, we illustrate in Section 5 that even under this simplifying assumption, solving the toll-setting problem is a highly challenging task.

---

[1] In the literature, the user optimum is commonly referred to as Wardrop's second principle, despite it being introduced first in Wardrop and Whitehead (1952); see, e.g., the respective discussion by Ferris and Pang (1997).

Assumption 4. *For every commodity $k \in \mathcal{K}$, the travel cost functions $c^k(f;\tau) = \left(c_a^k(f;\tau_a)\right)_{a\in\mathcal{A}}$ are affine-linear in the flows, i.e., there exists a matrix $C^k \in \mathbb{R}_{\geq 0}^{|\mathcal{A}|\times|\mathcal{A}|}$ and a vector $c^{\text{fix},k} \in \mathbb{R}_{>0}^{|\mathcal{A}|}$ with $c^k(f;\tau) = C^k f + c^{\text{fix},k} + \tau$.*

We emphasize that we do not make any assumptions about the separability of the travel cost functions in Assumption 4. In traffic assignment problems, it is often interesting to consider travel costs that are non-separable. This means that, for $k \in \mathcal{K}$, the costs $c_a^k(f;\tau_a)$ may not only depend on the flow f_a on arc $a \in \mathcal{A}$ itself but also on the flows $f_{a'}$, $a' \neq a \in \mathcal{A}$, on the other arcs. For motivating examples as well as further discussions on non-separable travel costs, we refer to Dafermos (1971). In Assumption 4, non-separability implies that the travel cost matrix C^k is not a diagonal matrix.

The remainder of this section is organized as follows. In Section 3.1, we elaborate on how to obtain sufficiently large big-M constants that can be used in Problem (5). Afterward, in Section 3.2, we prove the existence of an optimal solution to the toll-setting problem (5). In Section 3.3, we provide valid inequalities to strengthen the formulation in (5).

3.1. Computing Big-Ms.

In what follows, we provide bounds for the flow variables f and x as well as for the minimum travel costs t, which are essential for obtaining sufficiently large big-Ms that can be used in the MINLP reformulation (5) of the toll-setting problem. For this purpose, we first establish the existence of a Wardrop equilibrium for any given toll-setting policy $\tau \in \mathcal{T}$.

Lemma 1. *Let $\tau \in \mathcal{T}$ be given arbitrarily. Then, under Assumptions 1–4, $S(\tau) \neq \emptyset$ holds, i.e., there exists a Wardrop equilibrium for the given tolls τ.*

Proof. For every commodity $k = (\alpha_k, \omega_k) \in \mathcal{K}$, there exists at least one path that connects α_k and ω_k due to Assumption 2. Moreover, Assumption 1 implies $\tau_a \geq 0$ for all $a \in \mathcal{A}$. Hence, by Assumptions 1 and 4, the travel cost functions $c_a^k(f;\tau_a)$ are positive and continuous for all $a \in \mathcal{A}$ and $k \in \mathcal{K}$. Moreover, by Assumption 3, the travel demand is fixed, positive, and bounded from above. Under Assumptions 1–4, we can thus apply Theorem 5.5 in Aashtiani and Magnanti (1981), which yields the existence of a Wardrop equilibrium in the path formulation. As a consequence, there also exists a Wardrop equilibrium in the node-arc formulation; see, e.g., the discussion in Section 2.2.2 in Patriksson (2015) for further details. ☐

Next, we provide bounds for the commodity flow variables x in a Wardrop equilibrium.

Proposition 1. *Let $\tau \in \mathcal{T}$ be given arbitrarily. Then, under Assumptions 1–4, there exists $(f, x) \in S(\tau)$ that satisfies*

$$0 \leq x_a^k \leq d_k, \quad a \in \mathcal{A},\ k \in \mathcal{K}.$$

In particular, $x_a^k = 0$ holds for all $a \in \delta^{in}(\alpha_k) \cup \delta^{out}(\omega_k)$ with $k = (\alpha_k, \omega_k) \in \mathcal{K}$.

Proof. By Assumptions 1–4, we can apply Lemma 1, i.e., there exists $(f, x) \in S(\tau)$. The non-negativity of the commodity flows x immediately follows from (4). We now prove the upper bound. To this end, let $k = (\alpha_k, \omega_k) \in \mathcal{K}$ be given arbitrarily. By the flow decomposition theorem, we obtain

$$x_a^k = \sum_{\{p\in\mathcal{P}^k \colon a\in p\}} h_p + \sum_{\{\ell\in\mathcal{C} \colon a\in\ell\}} g_\ell^k, \quad a \in \mathcal{A};$$

see, e.g., Theorem 3.5 in Ahuja et al. (1993). Here, $\mathcal{P}^k$ denotes the set of all simple paths between the origin α_k and the destination ω_k of commodity k and $\mathcal{C}$ denotes the set of all cycles in the traffic network. The vectors $h = (h_p)_{p\in\mathcal{P}^k}$

and $g^k = (g^k_\ell)_{\ell \in \mathcal{C}}$ are used for the path and cycle flows, respectively. Suppose that there is a cycle $\ell \in \mathcal{C}$ with positive flow, i.e., $g^k_\ell > 0$ holds. This implies that $x^k_{a'} > 0$ holds for all $a' \in \ell$. Wardrop's second principle (4) thus yields

$$\sum_{a' \in \ell} c^k_{a'}(f; \tau_{a'}) = 0,$$

which is a contradiction to Assumptions 1 and 4. Hence, there cannot be a cycle with positive flow. Consequently, $x^k_a = 0$ holds for all $a \in \delta^{\text{in}}(\alpha_k) \cup \delta^{\text{out}}(\omega_k)$ with $k = (\alpha_k, \omega_k) \in \mathcal{K}$. From flow conservation (3), we thus obtain

$$\sum_{a \in \delta^{\text{in}}(\omega_k)} x^k_a = \sum_{a \in \delta^{\text{out}}(\alpha_k)} x^k_a = d_k. \tag{6}$$

Moreover, again due to flow conservation (3), we have

$$d_k = \sum_{a \in \delta^{\text{in}}(\omega_k)} x^k_a \geq \sum_{a \in \delta^{\text{out}}(i)} x^k_a = \sum_{a \in \delta^{\text{in}}(i)} x^k_a, \quad i \in \mathcal{N} \setminus \{\alpha_k, \omega_k\}. \tag{7}$$

The non-negativity of the commodity flows as well as (6) and (7) finally yield $x^k_a \leq d_k$ for all $a \in \mathcal{A}$, which concludes the proof. □

Since the overall arc flows f and the commodity flows x^k, $k \in \mathcal{K}$, are linearly coupled, Proposition 1 also yields valid bounds for the arc flows.

Proposition 2. *Let $\tau \in \mathcal{T}$ be given arbitrarily. Then, under Assumptions 1–4, there exists $(f, x) \in S(\tau)$ that satisfies*

$$0 \leq f_a \leq \sum_{k \in \mathcal{K}} d_k, \quad a \in \mathcal{A}.$$

Proof. By Assumptions 1–4, we can apply Lemma 1, i.e., there exists $(f, x) \in S(\tau)$. For all $a \in \mathcal{A}$, we have

$$0 \leq f_a = \sum_{k \in \mathcal{K}} x^k_a \leq \sum_{k \in \mathcal{K}} d_k.$$

Here, the first inequality follows from (2) and the non-negativity of the commodity flows given by (4), the equality follows from (2), and the last inequality is due to Proposition 1. □

In the following proposition, we provide bounds for the minimum travel costs t. The key idea is that, whenever a vector (f, x, t) solves (2)–(4), we can shift all values of t by the same amount while still satisfying the conditions.

Proposition 3. *Let $\tau \in \mathcal{T}$ be given arbitrarily and suppose that Assumptions 1–4 hold. Then, for all $(f, x) \in S(\tau)$, there exists t such that (f, x, t) solves (2)–(4) and t has the following properties:*

(i) *For all $k = (\alpha_k, \omega_k) \in \mathcal{K}$, it holds $t^k_{\omega_k} = 0$.*
(ii) *For all $k = (\alpha_k, \omega_k) \in \mathcal{K}$ and $i \in \mathcal{N} \setminus \{\omega_k\}$, it holds*

$$0 \leq t^k_i \leq \min_{p \in \mathcal{P}^k_i} \left\{ \sum_{a \in p} \left(\sum_{a' \in \mathcal{A}} C^k_{aa'} \sum_{q \in \mathcal{K}} d_q + c^{\text{fix},k}_a + \tau^+_a \right) \right\}$$

with $\mathcal{P}^k_i$ being the set of all simple paths between nodes i and ω_k.

Proof. Under Assumptions 1–4, there exists t such that (f, x, t) solves (2)–(4) due to Lemma 1. For all $k \in \mathcal{K}$, let now $\Delta t^k = t^k_{\omega_k}$ and consider $t^k_i - \Delta t^k$ instead of t^k_i for all $i \in \mathcal{N}$. Then, by construction, $t^k_{\omega_k} = 0$ holds for all $k \in \mathcal{K}$. Moreover, we have

$$0 \leq c^k_a(f; \tau_a) + (t^k_j - \Delta t^k) - (t^k_i - \Delta t^k) = c^k_a(f; \tau_a) + t^k_j - t^k_i \perp x^k_a \geq 0,$$

for all $a = (i, j) \in \mathcal{A}$ and all $k \in \mathcal{K}$, i.e., Wardrop's second principle (4) remains satisfied. Hence, and since Conditions (2) and (3) do not depend on t, there exists (f, x, t) with $t^k_{\omega_k} = 0$ for all $k \in \mathcal{K}$ that solves (2)–(4) for the given tolls τ. This proves (i). Let now $k \in \mathcal{K}$ be given arbitrarily. Summing over Conditions (4) and applying (i) yields

$$t_i^k \leq \sum_{a \in p} c_a^k(f; \tau_a) + t^k_{\omega_k} = \sum_{a \in p} c_a^k(f; \tau_a), \quad i \in \mathcal{N}, \ p \in \mathcal{P}_i^k,$$

which is equivalent to

$$t_i^k \leq \min_{p' \in \mathcal{P}_i^k} \left\{ \sum_{a \in p'} c_a^k(f; \tau_a) \right\}, \quad i \in \mathcal{N}. \tag{8}$$

We now show that for every node $i \in \mathcal{N}$ that is traversed in a path with positive commodity flow, the corresponding inequality in (8) is satisfied with equality. To this end, let $p \in \mathcal{P}_i^k$, $i \in \mathcal{N}$, be a path with positive commodity flow, i.e., $x_a^k > 0$ for all $a \in p$. Due to the complementarity in (4), we thus have $t_n^k = c_a^k(f; \tau_a) + t_m^k$ for all $a = (n, m) \in p$, which yields

$$t_i^k = \sum_{a \in p} c_a^k(f; \tau_a).$$

In particular, this means that p is a minimum-cost path from node i to ω_k. Next, we show that at least one equilibrium is preserved by setting

$$t_i^k = \min_{p' \in \mathcal{P}_i^k} \left\{ \sum_{a \in p'} c_a^k(f; \tau_a) \right\} \tag{9}$$

for all nodes $i \in \mathcal{N} \setminus \{\omega_k\}$. By our previous considerations, it suffices to consider arcs with zero flow, i.e., $a = (n, m) \in \mathcal{A}$ with $x_a^k = 0$. In this case, the complementarity in (4) is trivially satisfied. Hence, we only need to show that $t_n^k \leq c_a^k(f; \tau_a) + t_m^k$ holds for t_n^k and t_m^k as defined in (9). To this end, let

$$p_m = \operatorname*{arg\,min}_{p' \in \mathcal{P}_m^k} \left\{ \sum_{a' \in p'} c_{a'}^k(f; \tau_{a'}) \right\}.$$

Since $p_m \cup \{a\} \in \mathcal{P}_n^k$, we have

$$t_n^k = \min_{p' \in \mathcal{P}_n^k} \left\{ \sum_{a' \in p'} c_{a'}^k(f; \tau_{a'}) \right\} \leq c_a^k(f; \tau_a) + \sum_{a' \in p_m} c_{a'}^k(f; \tau_{a'}) = c_a^k(f; \tau_a) + t_m^k,$$

where both equalities follow from (9).

Since $k \in \mathcal{K}$ was chosen arbitrarily, we conclude that (f, x, t) with $t = (t^k)_{k \in \mathcal{K}}$, $t^k = (t_i^k)_{i \in \mathcal{N}}$ as in (9), and $t^k_{\omega_k} = 0$ solves (2), (3) and (4) for the given tolls τ. The non-negativity of t now follows from Assumptions 1 and 4. For all $i \in \mathcal{N} \setminus \{\omega_k\}$ and $k \in \mathcal{K}$, applying Assumptions 1 and 4 as well as Propositions 1 and 2 to (9) finally yields

$$\begin{aligned} t_i^k &\leq \min_{p' \in \mathcal{P}_i^k} \left\{ \sum_{a \in p'} \left(\sum_{a' \in \mathcal{A}} C_{aa'}^k f_{a'} + c_a^{\text{fix},k} + \tau_a \right) \right\} \\ &\leq \min_{p' \in \mathcal{P}_i^k} \left\{ \sum_{a \in p'} \left(\sum_{a' \in \mathcal{A}} C_{aa'}^k \sum_{q \in \mathcal{K}} d_q + c_a^{\text{fix},k} + \tau_a^+ \right) \right\}. \end{aligned} \qquad \square$$

Finally, we note that sufficiently large big-M constants for Problem (5) can be obtained by exploiting Assumptions 1 and 4 as well as Propositions 2 and 3.

3.2. **Existence of Solutions.** We now show the existence of an optimal solution to the overall toll-setting problem (1). To this end, we start with the following result.

Corollary 1. *Let $\tau \in \mathcal{T}$ be given arbitrarily. Then, under Assumptions 1–4, it holds*

$$S(\tau) = \left\{(f,x)\colon \exists t \text{ so that } (f,x,t) \text{ solves } (2)\text{–}(4) \text{ with } 0 \le t_i^k \le u_i^k,\ i \in \mathcal{N},\ k \in \mathcal{K}\right\},$$

where, for all $k = (\alpha_k, \omega_k) \in \mathcal{K}$, we have $u_{\omega_k}^k = 0$ and

$$u_i^k := \min_{p \in \mathcal{P}_i^k} \left\{ \sum_{a \in p} \left(\sum_{a' \in \mathcal{A}} C_{aa'}^k \sum_{q \in \mathcal{K}} d_q + c_a^{\text{fix},k} + \tau_a^+ \right) \right\}, \quad i \in \mathcal{N} \setminus \{\omega_k\}.$$

Corollary 1 immediately follows from Proposition 3 in which we state that imposing $t_i^k \le u_i^k$, $i \in \mathcal{N}$, $k \in \mathcal{K}$, does not affect the flows (f,x) in a Wardrop equilibrium.

Theorem 1. *Under Assumptions 1–4, the toll-setting problem (1) has an optimal solution (τ, f, x).*

Proof. By Assumption 1 and Lemma 1, the toll-setting problem (1) is feasible, i.e.,

$$\mathcal{F} := \{(\tau, f, x)\colon \tau \in \mathcal{T},\ (f,x) \in S(\tau)\} \neq \emptyset.$$

From Corollary 1 and Propositions 1–3, we further obtain that the feasible set $\mathcal{F}$ of Problem (1) is bounded. Moreover, the set $\mathcal{F}$ is described by a finite number of continuous functions, which implies that $\mathcal{F}$ is closed. Since the function $(\tau, f) \mapsto \sum_{a \in \mathcal{A}} \tau_a f_a$ is continuous as well, the Weierstraß theorem thus ensures that the toll-setting problem (1) has an optimal solution. ☐

3.3. **Valid Inequalities.** We now conclude this section by providing valid inequalities for the feasible set of Problem (5) as well as valid inequalities for optimal solutions to the problem.

Proposition 4. *Let $\tau \in \mathcal{T}$ be given arbitrarily. Further, let $i, j \in \mathcal{N}$ be such that $(i,j), (j,i) \in \mathcal{A}$ holds. Then, under Assumptions 1 and 4, the inequalities*

$$z_{(i,j)}^k + z_{(j,i)}^k \le 1, \quad k \in \mathcal{K},$$

are valid for the feasible set of Problem (5).

Proof. We prove the claim by contradiction. To this end, let (τ, f, x, t, z) be feasible for Problem (5) and let $k \in \mathcal{K}$ be given arbitrarily. Suppose now that $z_{(i,j)}^k = 1 = z_{(j,i)}^k$ holds. Then, Constraints (5e) yield

$$c_{(i,j)}^k(f; \tau_{(i,j)}) + t_j^k = t_i^k \quad \text{and} \quad c_{(j,i)}^k(f; \tau_{(j,i)}) + t_i^k = t_j^k.$$

From the latter, we obtain

$$t_i^k = c_{(i,j)}^k(f; \tau_{(i,j)}) + \left(c_{(j,i)}^k(f; \tau_{(j,i)}) + t_i^k\right) \iff 0 = c_{(i,j)}^k(f; \tau_{(i,j)}) + c_{(j,i)}^k(f; \tau_{(j,i)}),$$

which is a contradiction to Assumptions 1 and 4. Hence, a feasible point for Problem (5) satisfies $z_{(i,j)}^k + z_{(j,i)}^k \le 1$. ☐

From Proposition 4, we particularly obtain $0 \le x_{(i,j)}^k \perp x_{(j,i)}^k \ge 0$ for all nodes $i, j \in \mathcal{N}$ with $(i,j), (j,i) \in \mathcal{A}$ and all commodities $k \in \mathcal{K}$. This means that, under the assumption of positive travel costs, there cannot be positive commodity flow on both an arc and its reversed arc. Finally, we provide valid inequalities for the tolls τ in an optimal solution to Problem (5).

Proposition 5. *Under Assumptions 1–4, there exists an optimal solution* (τ, f, x, t, z) *to Problem* (5) *that satisfies*

$$\tau_a \geq \tau_a^+ \left(1 - \sum_{k \in \mathcal{K}} z_a^k\right), \quad a \in \mathcal{A}. \tag{10}$$

Proof. Under Assumptions 1–4, there exists an optimal solution (τ, f, x, t, z) to Problem (5) due to Theorem 1. By Assumption 1, an optimal solution particularly satisfies $\tau_a \geq 0$ for all $a \in \mathcal{A}$. For all arcs $a \in \mathcal{A}$ for which the set $\mathcal{T}$ imposes the upper bound $\tau_a^+ = 0$, Inequality (10) is trivially satisfied. Hence, we only need to consider arcs $a \in \mathcal{A}$ with $\tau_a^+ > 0$ in the following. Suppose that there is an arc $a \in \mathcal{A}$ with $\tau_a^+ > 0$ for which Inequality (10) is violated. If there exists $k \in \mathcal{K}$ with $z_a^k = 1$, this implies $\tau_a < 0$, which is a contradiction to the feasibility of τ. Hence, if a feasible point violates Inequality (10), $\tau_a < \tau_a^+$ and $z_a^k = 0$ for all $k \in \mathcal{K}$ needs to hold. In particular, the latter implies $x_a^k = 0$ for all $k \in \mathcal{K}$ and, thus, $f_a = 0$. We now set

$$\hat{\tau}_{a'} = \begin{cases} \tau_{a'}, & a' \in \mathcal{A} \setminus \{a\}, \\ \tau_{a'}^+, & a' = a. \end{cases}$$

By construction and due to Assumption 1, $(\hat{\tau}, f, x, t, z)$ satisfies Constraints (5b), (5c), (5d), (5f), and (5g). Moreover, (5e) is satisfied for a sufficiently large big-M constant M_a^k for all $k \in \mathcal{K}$. Hence, $(\hat{\tau}, f, x, t, z)$ is feasible for Problem (5). By construction, we further have $\tau_{a'} f_{a'} = \hat{\tau}_{a'} f_{a'}$ for all $a' \in \mathcal{A}$, i.e., $(\hat{\tau}, f, x, t, z)$ solves Problem (5) as well. In particular, $\hat{\tau}_a \geq \tau_a^+(1 - \sum_{k \in \mathcal{K}} z_a^k)$ holds. Repeating the previous procedure until there are no arcs left that violate Inequality (10) concludes the proof. ☐

4. ROBUSTIFICATION

Up to now, we have considered the setting in which the users of the traffic network act under perfect information. In real-world applications, however, travelers often face uncertainties when making their decisions. For instance, the travel costs may be subject to uncertainty due to unforeseen events such as accidents, maintenance work, or changing weather conditions. Hence, the assumption of perfect information seems to be rather strong. In this section, we consider the toll-setting problem (1) under uncertainties regarding the travel costs, which we tackle using techniques from robust optimization. In Section 4.1, we present a robustified variant of the toll-setting problem in which the network users hedge against uncertain travel costs within a predefined and user-specific uncertainty set. We model this setting as a mathematical problem with robustified Wardrop equilibrium conditions, for which we present an MINLP reformulation that exploits binary variables and big-M constants in Section 4.2. Section 4.3 is devoted to deriving valid big-Ms. We conclude by proving the existence of robust solutions in Section 4.4.

4.1. A Robust Toll-Setting Problem.

We start from the nominal Wardrop equilibrium model given by Conditions (2)–(4), for which we now assume that the travel costs of each arc $a \in \mathcal{A}$ and each commodity $k \in \mathcal{K}$ are not known exactly. More formally, we impose the following.

Assumption 5. *For all* $a \in \mathcal{A}$ *and* $k \in \mathcal{K}$, *the travel costs* $c_a^k(f; \tau_a)$ *are subject to additive deviations* $Y_a^k \Delta c_a^k$ *with* Y_a^k *being a random variable with support in* $[0, 1]$ *and* $\Delta c_a^k \geq 0$.

The parameters $\Delta c_a^k \geq 0$ denote upper bounds on the possible deviation from the nominal travel costs. Since it is unlikely that the costs realize in a worst-case sense on every arc of the network and, hence, to avoid being overly conservative, we assume that each commodity $k \in \mathcal{K}$ hedges against deviations of up to $\Gamma^k \in \{0, \dots, |\mathcal{A}|\}$. The robustified version of Wardrop's second principle (4) then reads

$$0 \le c_a^k(f;\tau_a) + y_a^k \Delta c_a^k + t_j^k - t_i^k \perp x_a^k \ge 0, \quad a=(i,j) \in \mathcal{A},\ k \in \mathcal{K}. \tag{11}$$

Here, for a commodity $k \in \mathcal{K}$ and a given flow vector x^k, the vector y^k solves

$$\max_{y^k} \quad \sum_{a \in \mathcal{A}} (\Delta c_a^k x_a^k) y_a^k \tag{12a}$$

$$\text{s.t.} \quad \sum_{a \in \mathcal{A}} y_a^k \le \Gamma^k, \tag{12b}$$

$$0 \le y_a^k \le 1, \quad a \in \mathcal{A}. \tag{12c}$$

Problem (12) is a linear problem for fixed x^k, $k \in \mathcal{K}$. Hence, the KKT conditions are necessary and sufficient optimality conditions, i.e., replacing Problem (12) by its KKT conditions yields an equivalent reformulation of (11) and (12) that is given by

$$0 \le c_a^k(f;\tau_a) + y_a^k \Delta c_a^k + t_j^k - t_i^k \perp x_a^k \ge 0, \quad a=(i,j) \in \mathcal{A},\ k \in \mathcal{K}, \tag{13a}$$

$$0 \le \xi^k + \zeta_a^k - \Delta c_a^k x_a^k \perp y_a^k \ge 0, \quad a \in \mathcal{A},\ k \in \mathcal{K}, \tag{13b}$$

$$0 \le 1 - y_a^k \perp \zeta_a^k \ge 0, \quad a \in \mathcal{A},\ k \in \mathcal{K}, \tag{13c}$$

$$0 \le \Gamma^k - \sum_{a \in \mathcal{A}} y_a^k \perp \xi^k \ge 0, \quad k \in \mathcal{K}. \tag{13d}$$

For notational convenience, we use $\tilde{c}_a^k(f;\tau_a) := c_a^k(f;\tau_a) + y_a^k \Delta c_a^k$ to denote the robustified travel costs for commodity $k \in \mathcal{K}$ on arc $a \in \mathcal{A}$ in the following. We further emphasize that the equilibrium conditions (2) and (3) do not explicitly depend on the travel costs. Hence, the set of robust Wardrop equilibria for given tolls τ and fixed $\Gamma = (\Gamma^k)_{k \in \mathcal{K}}$ can be stated as

$$S_{\text{rob}}(\tau) = \{(f,x) \colon \exists (t,y,\xi,\zeta) \text{ such that } (f,x,t,y,\xi,\zeta) \text{ solves } (2), (3), \text{and } (13)\}.$$

The overall robustified toll-setting problem is then given by

$$\max_{\tau,f,x} \quad \sum_{a \in \mathcal{A}} \tau_a f_a \quad \text{s.t.} \quad \tau \in \mathcal{T},\ (f,x) \in S_{\text{rob}}(\tau). \tag{14}$$

4.2. **An MINLP Reformulation.** Similar as it is done in Section 3, we exploit sufficiently large big-M constants and additional binary variables to linearize the complementarity constraints in (13). An MINLP reformulation of the robustified toll-setting problem (14) then reads

$$\max_{\tau,f,x,t,r} \quad \sum_{a \in \mathcal{A}} \tau_a f_a \tag{15a}$$

$$\text{s.t.} \quad \tau \in \mathcal{T}, \quad f = \sum_{k \in \mathcal{K}} x^k, \tag{15b}$$

$$\sum_{a \in \delta^{\text{in}}(i)} x_a^k - \sum_{a \in \delta^{\text{out}}(i)} x_a^k = d_i^k, \quad i \in \mathcal{N},\ k \in \mathcal{K}, \tag{15c}$$

$$\tilde{c}_a^k(f;\tau_a) + t_j^k - t_i^k \ge 0, \quad a=(i,j) \in \mathcal{A},\ k \in \mathcal{K}, \tag{15d}$$

$$\tilde{c}_a^k(f;\tau_a) + t_j^k - t_i^k \le M_a^k(1 - z_a^k), \quad a=(i,j) \in \mathcal{A},\ k \in \mathcal{K}, \tag{15e}$$

$$x_a^k \ge 0, \quad x_a^k \le M_a^k z_a^k, \quad a \in \mathcal{A},\ k \in \mathcal{K}, \tag{15f}$$

$$\xi^k + \zeta_a^k - \Delta c_a^k x_a^k \ge 0, \quad a \in \mathcal{A},\ k \in \mathcal{K}, \tag{15g}$$

$$\xi^k + \zeta_a^k - \Delta c_a^k x_a^k \le N_a^k w_a^k, \quad a \in \mathcal{A},\ k \in \mathcal{K}, \tag{15h}$$

$$y_a^k \ge 0, \quad y_a^k \le 1 - w_a^k, \quad a \in \mathcal{A},\ k \in \mathcal{K}, \tag{15i}$$

$$y_a^k \leq 1, \quad y_a^k \geq v_a^k, \qquad a \in \mathcal{A},\ k \in \mathcal{K}, \tag{15j}$$
$$\zeta_a^k \geq 0, \quad \zeta_a^k \leq L_a^k v_a^k, \qquad a \in \mathcal{A},\ k \in \mathcal{K}, \tag{15k}$$
$$\xi^k \geq 0, \quad \xi^k \leq R^k q_k, \qquad k \in \mathcal{K}, \tag{15l}$$
$$\sum_{a \in \mathcal{A}} y_a^k \leq \Gamma^k, \qquad k \in \mathcal{K}, \tag{15m}$$
$$\Gamma^k - \sum_{a \in \mathcal{A}} y_a^k \leq R^k(1 - q_k), \qquad k \in \mathcal{K}, \tag{15n}$$
$$q_k,\ v_a^k,\ w_a^k,\ z_a^k \in \{0,1\}, \qquad a \in \mathcal{A},\ k \in \mathcal{K}. \tag{15o}$$

Here, r contains all variables that are used for the robustification of the travel costs as well as the variables that are introduced for the linearization of the complementarity constraints, i.e., $r := (y, \xi, \zeta, q, v, w, z)$. By construction, Problem (15) is equivalent to the robustified toll-setting problem (14) for sufficiently large constants L_a^k, M_a^k, N_a^k, and R^k for all $a \in \mathcal{A}$ and $k \in \mathcal{K}$. Before we elaborate on how to obtain such constants in Section 4.3, we provide enhanced formulations for Problem (15) in the remainder of this section. For this purpose, we need the following auxiliary lemma.

Lemma 2. *Let $k \in \mathcal{K}$, $x^k \in \mathbb{R}_{\geq 0}^{|\mathcal{A}|}$, and $\Gamma^k \in \{0, \dots, |\mathcal{A}|\}$ be given arbitrarily. Then, Problem (12) has an optimal solution y^k that satisfies*

$$\sum_{a \in \mathcal{A}} y_a^k = \Gamma^k.$$

Proof. The objective function of Problem (12) is linear for fixed x^k, the zero vector is feasible for Problem (12), and the feasible set of Problem (12) is compact. By the Weierstraß theorem, Problem (12) thus has an optimal solution y^k. Moreover, the system $My^k \leq v$ that describes the feasible set of Problem (12) consists of a totally unimodular matrix $M \in \{0,1\}^{(|\mathcal{A}|+1) \times |\mathcal{A}|}$ and the right-hand side vector $v = \left(\Gamma^k, 1, \dots, 1\right)^\top \in \mathbb{Z}^{|\mathcal{A}|+1}$. Hence, by Proposition 3.3 in Wolsey (2020), Problem (12) has an integer solution. For $\Gamma^k = 0$, Constraint (12b) is satisfied with equality since $y^k = 0$ is the only feasible point for Problem (12). If $\sum_{a \in \mathcal{A}} y_a^k < \Gamma^k$ holds for some $\Gamma^k \geq 1$, the integrality of y^k implies that there exists at least one arc $a \in \mathcal{A}$ with $y_a^k = 0$. If $\Delta c_a^k x_a^k > 0$ holds, we have a contradiction to the optimality of y^k. Hence, there exists $a \in \mathcal{A}$ with $y_a^k = 0$ and $\Delta c_a^k x_a^k = 0$. We set

$$\hat{y}_{a'}^k = \begin{cases} y_{a'}^k, & a' \in \mathcal{A} \setminus \{a\}, \\ 1, & a' = a. \end{cases}$$

By construction, $\hat{y}$ solves Problem (12) as well. Repeating the previous procedure until Constraint (12b) is satisfied with equality concludes the proof. □

By Lemma 2, we can reduce the size of Problem (15) by eliminating the auxiliary binary variables q used for the linearization of the complementarity in (13d). To this end, we can replace Constraints (15l), (15m), and (15n) by

$$\xi^k \geq 0, \quad \sum_{a \in \mathcal{A}} y_a^k = \Gamma^k, \quad k \in \mathcal{K}.$$

While the complementarity in (13d) is satisfied for any $\xi \geq 0$, we emphasize that finite upper bounds on the variables ξ are required to obtain valid big-M constants N_a^k, $a \in \mathcal{A}$, $k \in \mathcal{K}$, for Constraint (15h). Hence, we additionally impose $\xi^k \leq R^k$ for all $k \in \mathcal{K}$.

Proposition 6. *The inequalities*

$$v_a^k + w_a^k \leq 1, \quad a \in \mathcal{A},\ k \in \mathcal{K},$$

are valid for the feasible set of Problem (15).

Proof. We prove the claim by contradiction. To this end, let $(\tau, f, x, t, y, \xi, \zeta, q, v, w, z)$ be feasible for Problem (15) and suppose that there exists an arc $a \in \mathcal{A}$ and a commodity $k \in \mathcal{K}$ for which the inequality $v_a^k + w_a^k \leq 1$ is violated. Hence, we have $v_a^k = 1 = w_a^k$. Then, we obtain $y_a^k = 1$ by Constraint (15j), which contradicts $y_a^k = 0$ that is obtained from Constraint (15i). □

Finally, we note that the robustified travel costs $\tilde{c}_a^k(f; \tau_a)$ are positive for all $a \in \mathcal{A}$ and $k \in \mathcal{K}$ under Assumptions 1, 4, and 5. Thus, the valid inequalities derived in Proposition 4 are also valid for Problem (15). Moreover, since we do not use any information about the travel costs to prove the validity of Inequalities (10), the latter are valid for optimal solutions to the robustified toll-setting problem as well.

4.3. **Computing Big-Ms.** We now derive bounds for the variables of Problem (15), which we exploit to obtain sufficiently large big-M constants L_a^k, M_a^k, N_a^k, and R^k for all $a \in \mathcal{A}$ and $k \in \mathcal{K}$. For this purpose, we first prove the existence of a robust Wardrop equilibrium for given tolls $\tau \in \mathcal{T}$.

Theorem 2. *Let $\tau \in \mathcal{T}$ as well as $\Gamma = (\Gamma^k)_{k \in \mathcal{K}}$ with $\Gamma^k \in \{0, \dots, |\mathcal{A}|\}$ for all $k \in \mathcal{K}$ be given arbitrarily. Then, under Assumptions 1–5, there exists $(f, x) \in S_{rob}(\tau)$.*

Proof. By Lemma 2, Problem (12) has an optimal solution y^k for arbitrarily given $x^k \in \mathbb{R}_{\geq 0}^{|\mathcal{A}|}$, $k \in \mathcal{K}$. Since the KKT conditions are necessary and sufficient for Problem (12), there exist ξ^k and $\zeta^k = (\zeta_a^k)_{a \in \mathcal{A}}$ such that (y^k, ξ^k, ζ^k) solves (13b)–(13d) for all $k \in \mathcal{K}$. Hence, Conditions (13b)–(13d) cannot induce any infeasibility. Similar as it is done in the proof of Lemma 1, we now apply Theorem 5.5 in Aashtiani and Magnanti (1981) to prove the existence of a robust Wardrop equilibrium. For the application of this theorem, it only remains to show that the robustified travel cost functions $\tilde{c}_a^k(f; \tau_a)$ are positive and continuous for all $a \in \mathcal{A}$ and $k \in \mathcal{K}$. The positivity of the travel cost functions immediately follows from Assumptions 1, 4, and 5. Moreover, Problem (12) is a linear problem for given commodity flows x^k so that we can use classic sensitivity results as, e.g., Proposition 4.3.3 in Bertsekas (2016). As a consequence, the function $x^k \mapsto y_a^k \Delta c_a^k$ with y^k being an optimal solution to the x^k-parameterized linear problem (12), is continuous for all $a \in \mathcal{A}$ and $k \in \mathcal{K}$. Thus, under Assumption 4, the robustified travel cost functions $\tilde{c}_a^k(f; \tau_a) := c_a^k(f; \tau_a) + y_a^k \Delta c_a^k$ are continuous. This concludes the proof. □

Remark 1. *The budgeted uncertainty modeling used in Problem (12) is closely related to the so-called Γ-robust approach presented in Bertsimas and Sim (2003) and Sim (2004). Pursuing a Γ-robust approach in our setting, however, would imply that the users of the traffic network hedge against uncertain travel costs on at most Γ^k many arcs of the network. This requires imposing integrality on the variables y^k in Problem (12), which leads to robustified travel cost functions $\tilde{c}_a^k(f; \tau_a) := c_a^k(f; \tau_a) + y_a^k \Delta c_a^k$ that are no longer continuous. Continuity is needed to prove the existence of robust Wardrop equilibria using Theorem 5.5 in Aashtiani and Magnanti (1981). Hence, proving existence of Γ-robust Wardrop equilibria in the sense of Bertsimas and Sim (2003) and Sim (2004) most likely requires different techniques compared to those used in the last proof.*

In the remainder of this section, we provide bounds for the flow variables f and x as well as for the variables t, ξ, and ζ in a robust Wardrop equilibrium.

Corollary 2. *Let $\tau \in \mathcal{T}$ as well as $\Gamma = (\Gamma^k)_{k \in \mathcal{K}}$ with $\Gamma^k \in \{0, \dots, |\mathcal{A}|\}$ for all $k \in \mathcal{K}$ be given arbitrarily. Then, under Assumptions 1–5, there exists $(f, x) \in S_{rob}(\tau)$ that satisfies*

$$0 \leq x_a^k \leq d_k, \quad a \in \mathcal{A}, \ k \in \mathcal{K},$$

as well as

$$0 \le f_a \le \sum_{k \in \mathcal{K}} d_k, \quad a \in \mathcal{A}.$$

Under Assumptions 1, 4, and 5, the robustified travel cost functions $\tilde{c}_a^k(f;\tau_a)$ are positive for all $a \in \mathcal{A}$ and $k \in \mathcal{K}$. Hence, Corollary 2 can be shown in analogy to the proofs of Propositions 1 and 2 by replacing the nominal travel cost functions $c_a^k(f;\tau_a)$ with the robustified travel cost functions $\tilde{c}_a^k(f;\tau_a)$ for all $a \in \mathcal{A}$ and $k \in \mathcal{K}$.

Corollary 3. *Let $\tau \in \mathcal{T}$ as well as $\Gamma = (\Gamma^k)_{k \in \mathcal{K}}$ with $\Gamma^k \in \{0, \dots, |\mathcal{A}|\}$ for all $k \in \mathcal{K}$ be given arbitrarily and suppose that Assumptions 1–5 hold. Then, for all $(f,x) \in S_{rob}(\tau)$, there exists (t,y,ξ,ζ) such that (f,x,t,y,ξ,ζ) solves (2), (3) and (13), and t has the following properties:*

(i) *For all $k = (\alpha_k, \omega_k) \in \mathcal{K}$, it holds $t_{\omega_k}^k = 0$.*

(ii) *For all $k = (\alpha_k, \omega_k) \in \mathcal{K}$ and $i \in \mathcal{N} \setminus \{\omega_k\}$, it holds*

$$0 \le t_i^k \le \min_{p \in \mathcal{P}_i^k} \left\{ \sum_{a \in p} \left(\sum_{a' \in \mathcal{A}} C_{aa'}^k \sum_{q \in \mathcal{K}} d_q + c_a^{\mathit{fix},k} + \tau_a^+ + \Delta c_a^k \right) \right\} =: u_i^k$$

with $\mathcal{P}_i^k$ being the set of all simple paths between nodes i and ω_k. In particular, it holds

$$S_{rob}(\tau) = \{(f,x) : \exists (t,y,\xi,\zeta) \text{ such that } (f,x,t,y,\xi,\zeta) \text{ solves } (2), (3), \text{and } (13)$$
$$\text{with } 0 \le t_i^k \le u_i^k,\ i \in \mathcal{N},\ k \in \mathcal{K}\}.$$

Corollary 3 can be shown in analogy to the proof of Proposition 3 by replacing the nominal travel cost functions $c_a^k(f;\tau_a)$ with the robustified travel cost functions $\tilde{c}_a^k(f;\tau_a)$ for all $a \in \mathcal{A}$ and $k \in \mathcal{K}$.

Proposition 7. *Let $\tau \in \mathcal{T}$ as well as $\Gamma = (\Gamma^k)_{k \in \mathcal{K}}$ with $\Gamma^k \in \{0, \dots, |\mathcal{A}|\}$ for all $k \in \mathcal{K}$ be given arbitrarily. Then, under Assumptions 1–5, there exists (f,x,t,y,ξ,ζ) that solves (2), (3), and (13) with*

$$0 \le \xi^k \le \max_{a \in \mathcal{A}} \left\{ \Delta c_a^k d_k \right\}, \quad k \in \mathcal{K}, \tag{16}$$

and

$$0 \le \zeta_a^k \le \Delta c_a^k d_k, \quad a \in \mathcal{A},\ k \in \mathcal{K}. \tag{17}$$

In particular, it holds

$$S_{rob}(\tau) = \{(f,x) : \exists (t,y,\xi,\zeta) \text{ such that } (f,x,t,y,\xi,\zeta)$$
$$\text{solves } (2), (3), (13), (16) \text{ and } (17)\}.$$

Proof. By Theorem 2, there exists (f,x,t,y,ξ,ζ) that solves (2), (3), and (13). Since the non-negativity of ξ and ζ immediately follows from the feasibility w.r.t. Conditions (13), we only need to prove the upper bounds. To this end, let $k \in \mathcal{K}$ be given arbitrarily. If $\Gamma^k = 0$ holds, commodity k does not hedge against any uncertainties regarding the travel costs, i.e., no additional variables ξ^k and ζ^k are introduced for the robustification of commodity k. Consequently, it suffices to consider the case $\Gamma^k \ge 1$. Due to Lemma 2, we can assume w.l.o.g. that $\sum_{a \in \mathcal{A}} y_a^k = \Gamma^k$ holds in a robust Wardrop equilibrium. In particular, this implies that there exists at least one arc $a \in \mathcal{A}$ with $y_a^k > 0$. Condition (13b) then yields $\xi^k + \zeta_a^k = \Delta c_a^k x_a^k$. From the non-negativity of ξ^k and ζ_a^k, we thus obtain

$$0 \le \zeta_a^k \le \Delta c_a^k x_a^k \quad \text{and} \quad 0 \le \xi^k \le \Delta c_a^k x_a^k$$

for all $a \in \mathcal{A}$ with $y_a^k > 0$. Moreover, for all $a \in \mathcal{A}$ with $y_a^k = 0$, we obtain $\zeta_a^k = 0$ from (13c). Taking all previous considerations into account, we obtain valid bounds

$$\begin{aligned} & 0 \leq \xi^k \leq \max_{a \in \mathcal{A}} \left\{ \Delta c_a^k x_a^k \right\} \leq \max_{a \in \mathcal{A}} \left\{ \Delta c_a^k d_k \right\}, && k \in \mathcal{K}, \\ & 0 \leq \zeta_a^k \leq \Delta c_a^k d_k, && a \in \mathcal{A},\ k \in \mathcal{K}, \end{aligned}$$

by exploiting Corollary 2 as well as Assumptions 3 and 5. We further note that imposing these bounds does not affect the flows (f, x) in a robust Wardrop equilibrium for the given tolls τ. This concludes the proof. □

Finally, sufficiently large big-M constants for Problem (15) can be obtained by exploiting Assumptions 1, 4, and 5, Corollaries 2 and 3, as well as Proposition 7.

4.4. **Existence of Solutions.**

We conclude this section by showing the existence of solutions to the robustified toll-setting problem (14).

Theorem 3. *Under Assumptions 1–5, the robustified toll-setting problem (14) has an optimal solution* (τ, f, x)*.*

Proof. We consider the problem

$$\begin{aligned} \max_{\tau, f, x, t, y, \xi, \zeta} \quad & \sum_{a \in \mathcal{A}} \tau_a f_a && && \text{(18a)} \\ \text{s.t.} \quad & \tau \in \mathcal{T},\ (f, x, t, y, \xi, \zeta) \text{ solves } (2), (3), \text{and } (13), && && \text{(18b)} \\ & 0 \leq t_i^k \leq u_i^k, && i \in \mathcal{N},\ k \in \mathcal{K}, && \text{(18c)} \\ & \zeta_a^k \leq \Delta c_a^k d_k, && a \in \mathcal{A},\ k \in \mathcal{K}, && \text{(18d)} \\ & \xi^k \leq \max_{a \in \mathcal{A}} \left\{ \Delta c_a^k d_k \right\}, && k \in \mathcal{K}, && \text{(18e)} \end{aligned}$$

where, for all $k \in \mathcal{K}$, we have $u_{\omega_k}^k = 0$ and

$$u_i^k := \min_{p \in \mathcal{P}_i^k} \left\{ \sum_{a \in p} \left(\sum_{a' \in \mathcal{A}} C_{aa'}^k \sum_{q \in \mathcal{K}} d_q + c_a^{\text{fix},k} + \tau_a^+ + \Delta c_a^k \right) \right\}, \quad i \in \mathcal{N} \setminus \{\omega_k\}.$$

By Assumption 1 and Theorem 2, the feasible set of Problem (18) is non-empty. Moreover, the feasible set of Problem (18) is bounded due to Assumption 1, Corollaries 2 and 3, as well as Proposition 7. In particular, the feasible set of Problem (18) is described by a finite number of continuous functions, which implies its closedness. Since the objective function of Problem (18) is continuous, the Weierstraß theorem thus ensures that Problem (18) has an optimal solution. From Part (iii) of Corollary 3 and from Proposition 7, we further obtain

$$\begin{aligned} S_{\text{rob}}(\tau) &= \{(f, x) \colon \exists t \text{ s.t. } (f, x, t, y, \xi, \zeta) \text{ solves } (2), (3), \text{and } (13)\} \\ &= \{(f, x) \colon \exists t \text{ s.t. } (f, x, t, y, \xi, \zeta) \text{ solves } (18b)\text{–}(18e)\} \end{aligned}$$

for arbitrarily given $\tau \in \mathcal{T}$. Since we consider the same objective functions in Problems (14) and (18), an optimal solution to Problem (18) solves the robust toll-setting problem (14) as well. □

5. CASE STUDY

In this section, we present a case study to illustrate how the consideration of travelers, who hedge against travel cost uncertainty in a robust way, may impact toll-setting policies. In Sections 5.1 and 5.2, we briefly discuss the computational setup and the considered test instances. In Section 5.3, we discuss the computational results of our case study.

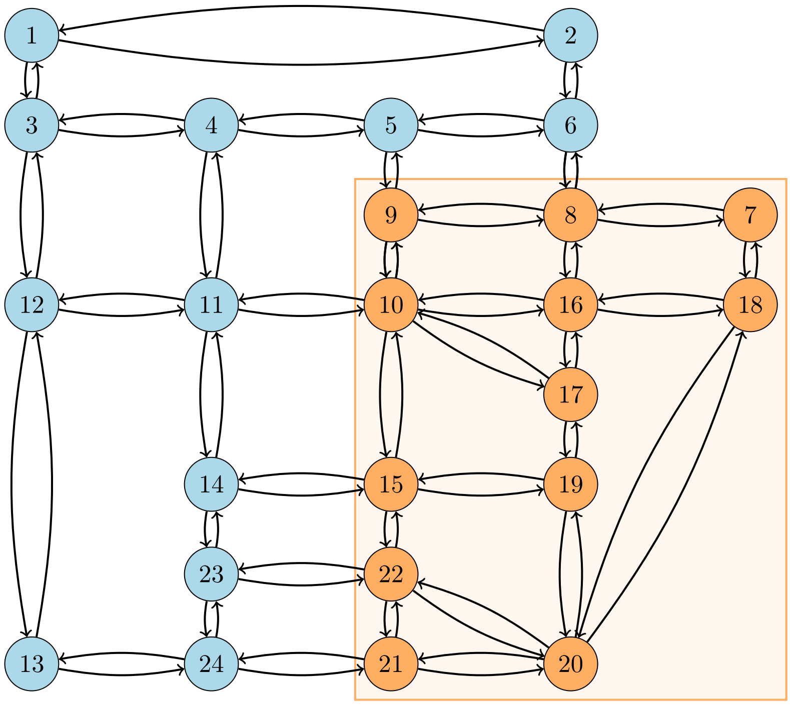

FIGURE 1. The entire Sioux Falls network (blue and orange nodes) consisting of 24 nodes and 76 arcs and the "Sioux Falls East" network (orange nodes) consisting of 12 nodes and 36 arcs.

5.1. **Computational Setup.** All tests have been realized on an Intel XEON SP 6126 at 2.6 GHz (4 cores) with 32 GB RAM, which is part of the high performance cluster "Elwetritsch" at TU Kaiserslautern within the "Alliance of High Performance Computing Rheinland-Pfalz" (AHRP).[2] The toll-setting problems (5) and (15) are implemented in Python 3.7.11 and we use Gurobi 10.0.3 to solve them. In particular, the implementation of our models includes the valid inequalities presented in Propositions 4, 5, and 6. Preliminary computational tests revealed that including these inequalities significantly enhances the solution process. Since the toll-setting models are nonconvex MINLPs, we need to set the Gurobi parameter NonConvex to 2. All other parameters have been left at their default settings. For each test run, we set a time limit (TL) of 1 h.

5.2. **Test Instances.** We consider a subnetwork of the Sioux Falls network (LeBlanc et al. 1975), which is publicly available at `https://github.com/bstabler/TransportationNetworks`. The subnetwork, which we refer to as "Sioux Falls East", consists of 12 nodes and 36 arcs. In this case study, we consider a varying number of origin-destination (OD) pairs ranging from 4 to 8. An illustration of both the entire Sioux Falls network and the "Sioux Falls East" subnetwork is given in Figure 1. The travel costs that have been provided for the Sioux Falls network are defined by the BPR function (U.S. Bureau of Public Roads 1964), which is given by

$$c_a(f_a) = c_a^{\text{fix}} \left(1 + 0.15 \left(\frac{f_a}{u_a}\right)^4\right), \quad a \in \mathcal{A}.$$

---

[2] We kindly acknowledge the support of RHRK (`https://rz.rptu.de/en/`).

Here, $c_a^{\mathrm{fix}} > 0$ denotes the fixed costs ("free-flow time") and $u_a > 0$ denotes the capacity of an arc $a \in \mathcal{A}$. In this paper, we address the problem of determining optimal tolls in a network with travel cost functions that are affine-linear in the flows f; cf. Assumption 4. Hence, we adapt the BPR functions to account for our setting. In this case study, we consider the travel cost functions

$$c_a(f_a; \tau_a) = c_a^{\mathrm{fix}} \left(1 + 1.5 \, \frac{f_a}{u_a}\right) + \tau_a, \quad a \in \mathcal{A}.$$

We emphasize that these travel costs are separable and that each commodity faces the same costs. Since the used data for the Sioux Falls network does not contain toll arcs, we have generated toll arcs using a procedure similar to the one considered in Brotcorne et al. (2000) and Minh Bui et al. (2022). The method works as follows. For a given set of OD pairs, we determine the shortest path for each commodity. For each arc of the network, we then determine the number of shortest paths that go through that arc. Afterward, we sort the arcs of the network in decreasing order w.r.t. the number of shortest paths traversing it. Following this order, we convert each arc and its reversed arc into a toll arc until $2/3$ of the desired number of toll arcs is reached. The remaining $1/3$ of the desired number of toll arcs is chosen randomly among the remaining arcs. Again, if an arc is converted to a toll arc, we also convert its reversed arc. For all arcs of the network, we impose a lower bound of 0 on the tolls and we set the upper bound $\tau_a^+ = 0$ for toll-free arcs. For toll arcs, the upper bounds τ_a^+ on the tolls are set to the fixed travel costs c_a^{fix}. As usual in the literature, we half the costs c_a^{fix} for toll arcs after their conversion. Since we define finite upper bounds on the tolls, the revenues of the toll-setting authority are bounded from above; cf. Proposition 2. In contrast to Brotcorne et al. (2000) and Minh Bui et al. (2022), we thus do not need to ensure that at least one toll-free path is preserved for each commodity when converting arcs into toll arcs. Moreover, since the used data for the Sioux Falls network does not include uncertainties, we have randomly generated the uncertainty parameters Δc_a^k and Γ^k, $a \in \mathcal{A}$, $k \in \mathcal{K}$. For all commodities, we consider the same travel cost uncertainties, i.e., we consider $\Delta c_a^k = \Delta c_a$ for all $a \in \mathcal{A}$ and $k \in \mathcal{K}$. Here, Δc_a is a uniformly distributed random integer value in the interval $[0.5 c_a^{\mathrm{fix}}, 2 c_a^{\mathrm{fix}}]$. Moreover, for each commodity $k \in \mathcal{K}$, the parameter Γ^k takes a uniformly distributed integer value in the interval $[0, 0.5|\mathcal{A}|]$. We emphasize that Γ^k may differ among commodities.

5.3. **Computational Results.** We start by considering the nominal setting, i.e., the setting without any uncertainties regarding the travel costs. Reflected by the running times and the number of investigated branch-and-bound nodes shown in Table 1, we observe that the resources required to solve Problem (5) increase with the number of OD pairs. This is to be expected as the number of OD pairs directly influences the size of the toll-setting problem. For each additional OD pair, we introduce $2|\mathcal{A}| + |\mathcal{N}| = 84$ additional variables and $3|\mathcal{A}| + |\mathcal{N}| = 120$ additional constraints in the model. Moreover, since every arc in the "Sioux Falls East" network has one reversed arc, we further add $0.5|\mathcal{A}| = 18$ valid inequalities for each OD pair. Thus, it is evident that increasing the number of OD pairs increases the amount of resources required to solve the respective toll-setting problems. In addition, the results in Table 1 indicate that increasing the number of toll arcs in the network further increases the computational burden. This is due to the fact that more toll arcs lead to more nonconvex terms in the objective function of the toll-setting problem (5), which require a special algorithmic treatment. Gurobi tackles these nonconvexities using spatial branching based on convex envelopes. Overall, even under the assumption of affine-linear travel cost functions (see Assumption 4), solving Problem (5) is a highly challenging task. The latter is particularly reflected

Table 1. The revenues realized through imposing tolls as well as the runtimes (in s) and the number of investigated branch-and-bound nodes required to solve the respective nominal toll-setting optimization problem for the "Sioux Falls East" network with varying numbers of OD pairs ("$|\mathcal{K}|$") and toll arcs ("$|\mathcal{A}^{\text{toll}}|$"). Additionally, the optimality gap is shown (in %).

$\|\mathcal{K}\|$	$\|\mathcal{A}^{\text{toll}}\|$	revenues	runtime	nodes	gap
4	4	6000.00	1.90	3544	0.00
	6	7176.11	2.76	9765	0.01
	8	7176.11	11.58	58 979	0.01
5	4	6000.00	1.28	1100	0.00
	6	7415.86	25.57	155 363	0.01
	8	7415.86	27.18	153 352	0.01
6	4	6659.88	2.33	3707	0.01
	6	7415.86	40.75	168 903	0.01
	8	8086.66	3529.97	13 370 122	0.01
7	4	8271.18	2.49	4416	0.00
	6	7415.86	162.27	845 334	0.01
	8	9697.96	TL	11 804 881	0.40
8	4	8348.09	4.04	2280	0.00
	6	7445.88	710.55	2 306 930	0.01
	8	9835.07	TL	9 180 328	11.94

Table 2. The revenues realized through imposing tolls as well as the runtimes (in s) and the number of investigated branch-and-bound nodes required to solve the respective robustified toll-setting optimization problem for the "Sioux Falls East" network with varying numbers of OD pairs ("$|\mathcal{K}|$") and toll arcs ("$|\mathcal{A}^{\text{toll}}|$"). Additionally, the optimality gap is shown (in %).

$\|\mathcal{K}\|$	$\|\mathcal{A}^{\text{toll}}\|$	revenues	runtime	nodes	gap
4	4	8757.14	286.03	250 597	0.01
	6	10 086.90	TL	3 973 530	3.49
	8	11 794.84	TL	3 491 529	4.88
5	4	7934.72	TL	2 774 570	4.24
	6	9807.33	TL	2 840 066	3.81
	8	10 624.36	TL	1 578 972	15.90

by the two instances of the "Sioux Falls East" network that cannot be solved within the time limit of 1 h.

Compared to the nominal toll-setting problem (5), the robustified toll-setting problem (15) is significantly larger w.r.t. the number of variables and constraints. For the "Sioux Falls East" network, we introduce $145|\mathcal{K}|$ additional variables and $363|\mathcal{K}|$ additional constraints for the robustification and the linearization of robustified constraints. Moreover, we add $36|\mathcal{K}|$ additional valid inequalities; cf. Proposition 6. Since, even for the nominal setting, the problem size is a limiting factor for solving the respective toll-setting problem, the computational challenges resulting from

Table 3. Nominal vs. robust travel costs for each OD pair in the "Sioux Falls East" network with 4 (left) and 5 OD pairs (right) with a varying number of toll arcs ("$|\mathcal{A}^{\text{toll}}|$").

$\lvert\mathcal{A}^{\text{toll}}\rvert$	OD pair	nominal	robust
4	(8, 20)	10.02	15.92
	(9, 21)	16.44	24.87
	(16, 21)	14.59	27.66
	(17, 22)	11.55	18.56
6	(8, 20)	10.02	15.92
	(9, 21)	15.93	24.64
	(16, 21)	13.68	28.40
	(17, 22)	11.55	18.46
8	(8, 20)	10.02	15.92
	(9, 21)	15.93	24.58
	(16, 21)	13.68	27.34
	(17, 22)	11.55	18.73

$\lvert\mathcal{A}^{\text{toll}}\rvert$	OD pair	nominal	robust
4	(8, 20)	10.02	15.90
	(9, 21)	16.44	24.93
	(21, 18)	10.20	20.23
	(16, 21)	14.59	21.47
	(17, 22)	11.55	18.95
6	(8, 20)	10.02	15.90
	(9, 21)	15.93	24.72
	(21, 18)	10.03	21.03
	(16, 21)	13.68	22.04
	(17, 22)	11.55	18.88
8	(8, 20)	10.02	15.89
	(9, 21)	15.93	24.71
	(21, 18)	10.03	21.03
	(16, 21)	13.68	22.40
	(17, 22)	11.55	18.99

larger models is thus even more pronounced in the robust setting. In Table 2, it can be seen that only the smallest of our considered instances, i.e., the one with 4 OD pairs and 4 toll arcs, can be solved within the time limit of 1 h in the robust setting. Given this limitation, we thus refrain from presenting results for larger instances with 6 or more OD pairs.

While only one of the instances considered in the robust setting can be solved, the results presented in Table 2 still provide valuable insights into the impact of robustified travel decisions on the revenues realized by the toll-setting authority. It can be seen that, for the "Sioux Falls East" network, the revenues that are realized by imposing tolls are significantly higher in the robust setting compared to the nominal one. In this context, we emphasize that it is not the toll-setting authority that hedges against uncertain travel costs in a robust way, but the users of the traffic network. In particular, the users of the traffic network decide on their route choices in a "here-and-now" fashion, i.e., before the uncertainty realizes. Viewing the overall toll-setting problem as a single-leader multi-follower game, this means that we consider multiple "here-and-now" followers. Since this problem is considered from the leader's perspective, having higher revenues in the robust setting is thus not in contrast to classic robust optimization theory. To further illustrate this, we show the nominal and the robustified travel costs faced by each commodity in the "Sioux Falls East" network in Table 3. It can be seen that, when hedging against uncertain travel costs in a robust way, users of the traffic network always face increased travel costs to reach their destination. More formally, the previous observations indicate that, while the set of feasible flows do not change in the robust compared to the nominal setting, the set of Wardrop equilibria may change. Hence, for given tolls $\tau \in \mathcal{T}$, neither $S_{\text{rob}}(\tau) \subseteq S(\tau)$ nor $S_{\text{rob}}(\tau) \supseteq S(\tau)$ holds in general.

To further illustrate the impact of robustified travel decisions on the actual route choices and the imposed tolls, we now focus on the "Sioux Falls East" instance with 4 OD pairs and 4 toll arcs, which can be solved in both the nominal and the robust setting. In Figure 2, we show the flows in a nominal Wardrop equilibrium and the tolls imposed by the toll-setting authority. It can be seen that revenues

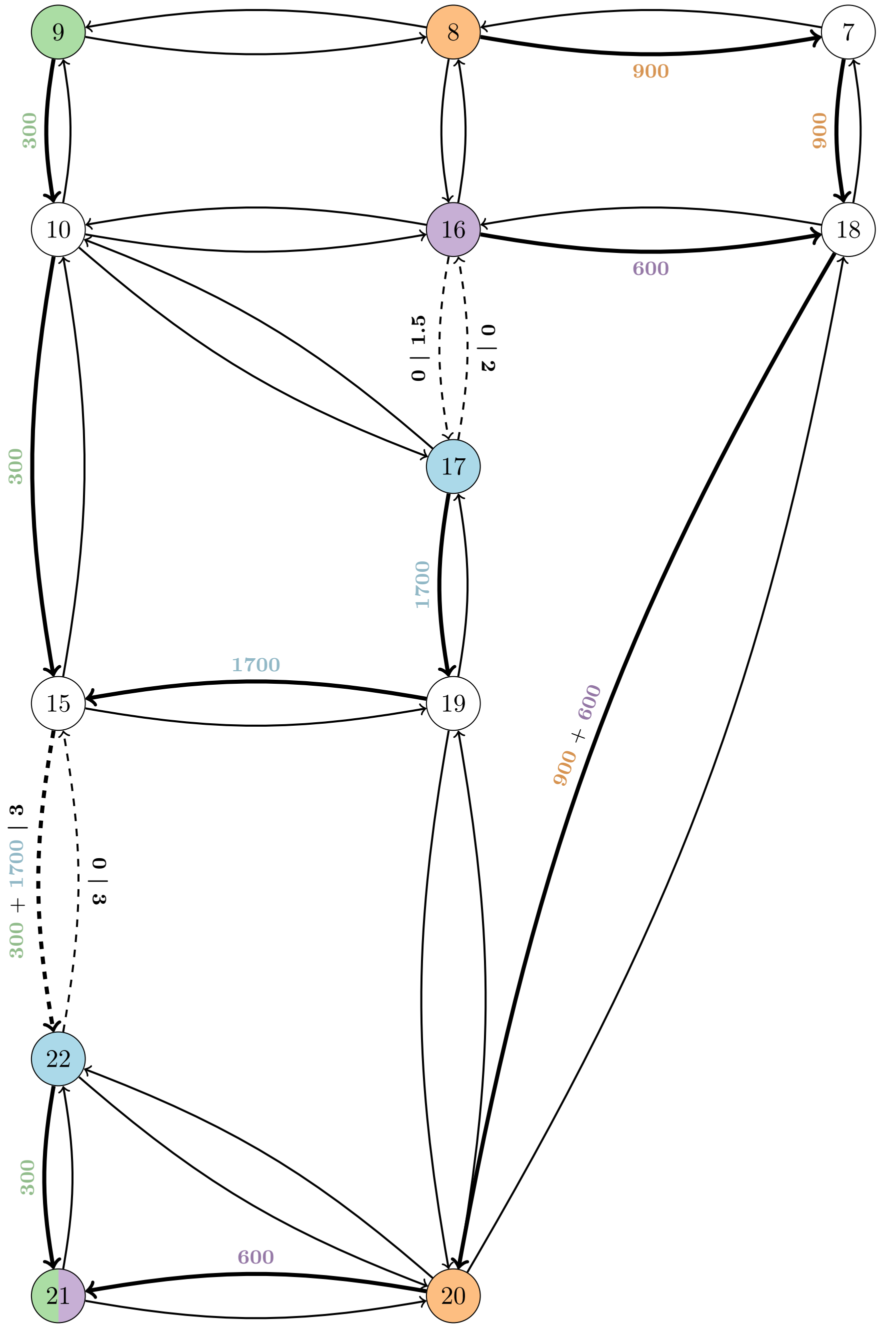

Figure 2. The “Sioux Falls East” network with 4 OD pairs and 4 toll arcs. Each OD pair is color-coded (orange, green, blue, purple). Dashed arcs represent toll arcs and solid arcs represent toll-free arcs. Edge labels correspond to commodity flows. For toll arcs, edge labels are given in the format “flow | toll”. If no label is shown, there is no flow on that edge.

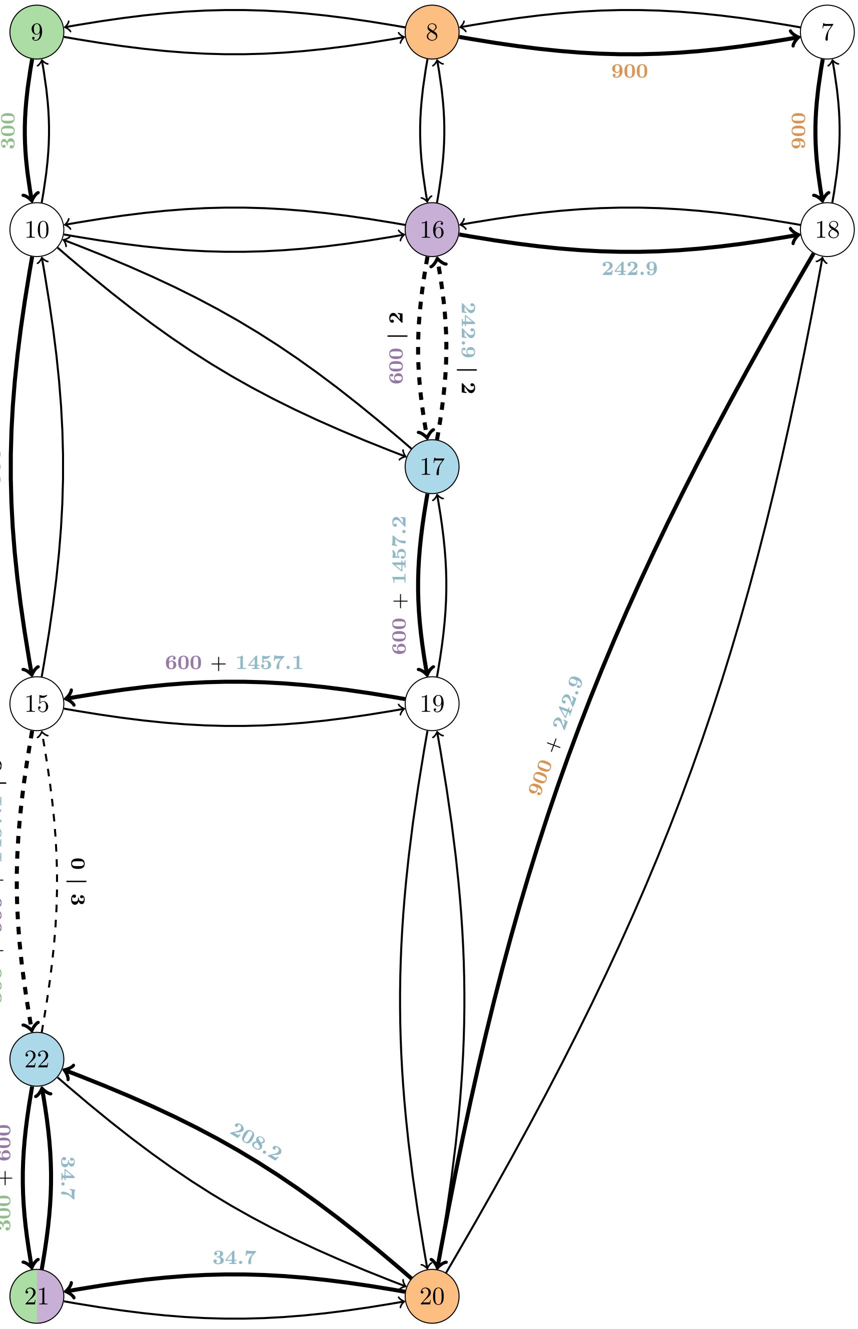

Figure 3. The robust "Sioux Falls East" network with 4 OD pairs and 4 toll arcs. Each OD pair is color-coded (orange, green, blue, purple). Dashed arcs represent toll arcs and solid arcs represent toll-free arcs. Edge labels correspond to commodity flows. For toll arcs, edge labels are given in the format "flow | toll". If no label is shown, there is no flow on that edge.

are only generated by imposing tolls on arc $(15, 22)$. The remaining toll arcs are not used by any commodity. Nevertheless, we emphasize that imposing tolls on arcs with zero flow may still be beneficial for the toll-setting authority, even if no revenues are generated. In this way, the toll-setting authority can influence the travel decisions of the users of the traffic network such as to encourage or discourage the use of specific arcs. This may lead to overall higher revenues. In particular, decreasing the imposed tolls on arcs with zero flow may affect the flows in a Wardrop equilibrium. Let us now consider the specific routes taken by each commodity. In Figure 2, we observe that the green and the blue commodities, i.e., OD pairs $(9, 21)$ and $(17, 22)$, take the most direct route to reach their destination. In doing so, they accept to pay tolls along the way. The orange and the purple commodities, i.e., OD pairs $(8, 20)$ and $(16, 21)$, do not take the most direct route and prefer to take a detour to avoid being charged tolls. However, the situation may change significantly if users of the traffic network hedge against uncertain travel costs in a robust way. In Figure 3, we show the imposed tolls and the flows in a robust Wardrop equilibrium. There are five aspects that we find particularly remarkable. First, revenues are now additionally generated by imposing tolls on the arcs that connect nodes 16 and 17. This is in contrast to the nominal setting. Second, we note that robust travel decisions do not affect the actual tolls charged on the arcs of the network for this instance. Despite the fact that the toll-setting authority imposes higher tolls on arc $(16, 17)$ in the robust setting (2 vs. 1.5), we emphasize that imposing tolls of 2 would be optimal in the nominal setting as well; cf. Proposition 5. Third, the green and the orange commodities do not change their travel decisions when hedging against travel cost uncertainties in a robust way. Also in the robust setting, the green commodity takes the most direct route, accepting the toll charges, while the orange commodity takes a detour to avoid toll arcs. Fourth, the travel decision of the purple commodity changes completely in the robust setting. Instead of taking a toll-free detour, it now takes the most direct route, which includes a toll arc. Finally, we point out that the flow of the blue commodity is split between the most direct tolled route and the toll-free detour. Moreover, the flows of the blue commodity are split between $(20, 21, 22)$ and $(20, 22)$ on the toll-free path.

To sum up, our case study illustrates that making robust travel decisions due to travel cost uncertainties may significantly impact the travel behavior and, thus, the revenues realized by imposing tolls. We have seen that users of the traffic network, who hedge against travel cost uncertainties within their user-specific uncertainty set, may be indifferent to uncertainties, change their travel decisions completely, or decide on something in between. While we have further observed that the actual toll-setting policies do not change in the robust setting for the specific “Sioux Falls East” instance considered in our case study, we note that this may not be the case in general. Nevertheless, given the significant increase in size and computational difficulty of the robustified toll-setting problem compared to the nominal one, an interesting future research question may be to identify situations in which nominal toll-setting policies also provide favorable results in the robust setting.

6. Conclusion

In this paper, we consider a multi-commodity traffic network in which a toll-setting authority aims to maximize revenues by imposing tolls on certain arcs of the network. Users of the traffic network act in the sense of Wardrop’s user equilibrium so that their individual travel costs are minimized. We model this setting as a mathematical problem with equilibrium constraints, for which we present a mixed-integer, nonlinear, and nonconvex reformulation that exploits binary variables and big-M constants. We prove existence of solutions to this problem, derive correct big-Ms, and provide valid inequalities. Moreover, we consider the setting in which the network users pursue a robust approach to hedge against uncertainties regarding their travel costs. In this paper, the uncertainties are assumed to vary within a predefined and user-specific uncertainty set, which relates to the notion of Γ-robustness. We reformulate the robustified problem as a mixed-integer, nonlinear, and nonconvex problem and prove the existence of robust solutions. To illustrate the impact of considering robust travel decisions on the revenues realized by the toll-setting authority, we further conduct a case study using a subnetwork of the well-known Sioux Falls network. We observe that addressing uncertainties in the travel costs may significantly impact the travel behavior and, in particular, may lead to increased revenues realized by imposing tolls. However, for the specific instance considered in our case study, we observe that the actual toll-setting policies do not change in the robust compared to the nominal setting. Given that solving the robustified toll-setting problem is significantly more challenging than the nominal one, a potential future research question could thus be to identify situations in which quality guarantees for nominal toll-setting policies in the robust setting are available. Another interesting research question could be to identify properties that ensure the existence of Γ-robust Wardrop equilibria in the classic sense of Bertsimas and Sim (2003) and Sim (2004). Finally, a possible direction of future research may be the development of tailored solution approaches that exploit, e.g., piecewise-linear approximations of the bilinearities in the objective function of the toll-setting problem.

Acknowledgment

Martine Labbé thanks the DFG for their support within the Research Training Group 2126: "Algorithmic Optimization".

References

Aashtiani, H. Z. and T. L. Magnanti (1981). "Equilibria on a Congested Transportation Network." In: *SIAM Journal on Algebraic Discrete Methods* 2.3. DOI: `10.1137/0602024`.

Ahuja, R. K., T. L. Magnanti, and J. B. Orlin (1993). *Network Flows: Theory, Algorithms, and Applications.* Prentice Hall.

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

Bertsekas, D. P. (2016). *Nonlinear Programming.* Athena scientific Belmont.

Bertsimas, D., D. 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`.

Brotcorne, L., M. Labbé, P. Marcotte, and G. Savard (2000). "A Bilevel Model and Solution Algorithm for a Freight Tariff-Setting Problem." In: *Transportation Science* 34.3, pp. 289–302. DOI: `10.1287/trsc.34.3.289.12299`.

Brotcorne, L., M. Labbé, P. Marcotte, and G. Savard (2001). "A Bilevel Model for Toll Optimization on a Multicommodity Transportation Network." In: *Transportation Science* 35.4. DOI: `10.1287/trsc.35.4.345.10433`.

Dafermos, S. C. (1971). "An Extended Traffic Assignment Model with Applications to Two-Way Traffic." In: *Transportation Science* 5.4, pp. 366–389. DOI: `10.1287/trsc.5.4.366`.

Dantzig, G. B., S. F. Maier, and Z. F. Lansdown (1976). *Application of Decomposition to Transportation Network Analysis*. Tech. rep. URL: `https://rosap.ntl.bts.gov/view/dot/10514`.

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

Dempe, S. and A. Zemkoho (2012). "Bilevel road pricing: theoretical analysis and optimality conditions." In: *Annals of Operations Research* 196, pp. 223–240. DOI: `10.1007/s10479-011-1023-z`.

Dewez, S., M. Labbé, P. Marcotte, and G. Savard (2008). "New formulations and valid inequalities for a bilevel pricing problem." In: *Operations Research Letters* 36.2, pp. 141–149. DOI: `10.1016/j.orl.2007.03.005`.

Ferris, M. C. and J. S. Pang (1997). "Engineering and Economic Applications of Complementarity Problems." In: *SIAM Review* 39.4, pp. 669–713. DOI: `10.1137/S0036144595285963`.

Gartner, N. H. (1980). "Optimal Traffic Assignment with Elastic Demands: A Review Part II. Algorithmic Approaches." In: *Transportation Science* 14.2, pp. 192–208. URL: `http://www.jstor.org/stable/25767978`.

Ito, Y. (2011). "Robust Wardrop Equilibria in the Traffic Assignment Problem with Uncertain Data." Master's thesis. Graduate School of Informatics, Kyoto University. URL: `http://www-optima.amp.i.kyoto-u.ac.jp/papers/master/2011_master_ito.pdf`.

Kalashnikov, V., J. G. Flores Muñiz, and N. Kalashnykova (2020). "Bilevel Optimal Tolls Problems with Nonlinear Costs: A Heuristic Solution Method." In: *Beyond Traditional Probabilistic Data Processing Techniques: Interval, Fuzzy etc. Methods and Their Applications*. Ed. by O. Kosheleva, S. P. Shary, G. Xiang, and R. Zapatrin. Cham: Springer International Publishing, pp. 481–516. DOI: `10.1007/978-3-030-31041-7_28`.

Labbé, M., P. Marcotte, and G. Savard (1998). "A Bilevel Model of Taxation and Its Application to Optimal Highway Pricing." In: *Management Science* 44.12.1. DOI: `10.1287/mnsc.44.12.1608`.

Labbé, M., P. Marcotte, and G. Savard (2000). "On a class of bilevel programs." In: *Nonlinear Optimization and Related Topics*. Ed. by G. D. Pillo and F. Giannessi. Boston, MA: Springer US, pp. 183–206. DOI: `10.1007/978-1-4757-3226-9_10`.

LeBlanc, L. J., E. K. Morlok, and W. P. Pierskalla (1975). "An efficient approach to solving the road network equilibrium traffic assignment problem." In: *Transportation Research* 9.5, pp. 309–318. DOI: `10.1016/0041-1647(75)90030-1`.

Luo, Z.-Q., J.-S. Pang, and D. Ralph (1996). *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press. DOI: `10.1017/CBO9780511983658`.

Minh Bui, Q., B. Gendron, and M. Carvalho (2022). "A Catalog of Formulations for the Network Pricing Problem." In: *INFORMS Journal on Computing* 34.5, pp. 2658–2674. DOI: `10.1287/ijoc.2022.1198`.

Ordóñez, F. and N. E. Stier-Moses (2007). "Robust Wardrop Equilibrium." In: *Network Control and Optimization*. Ed. by T. Chahed and B. Tuffin. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 247–256. DOI: `10.1007/978-3-540-72709-5_26`.

Ordóñez, F. and N. E. Stier-Moses (2010). "Wardrop Equilibria with Risk-Averse Users." In: *Transportation Science* 44.1, pp. 63–86. DOI: `10.1287/trsc.1090.0292`.

Patriksson, M. (2015). *The Traffic Assignment Problem: Models and Methods*. Courier Dover Publications.

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.

U.S. Bureau of Public Roads (1964). *Traffic assignment manual.* Washington, D.C.: U.S. Government Printing Office.

Wardrop, J. G. (1952). “Some Theoretical Aspects of Road Traffic Research.” In: *Proceedings of the Institution of Civil Engineers* 1.3, 325—362. DOI: 10.1680/ipeds.1952.11259.

Wardrop, J. G. and J. I. Whitehead (1952). “Correspondence. Some Theoretical Aspects of Road Traffic Research.” In: *ICE Proceedings: Engineering Divisions* 1.5, p. 767. DOI: 10.1680/ipeds.1952.11362.

Wolsey, L. A. (2020). *Integer Programming.* John Wiley & Sons, Ltd. DOI: 10.1002/9781119606475.

(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

(M. Labbé) (A) Université Libre de Bruxelles, Department of Computer Science, Boulevard du Triomphe, CP212, 1050 Brussels, Belgium; (B) Inria Lille - Nord Europe, Parc scientifique de la Haute Borne, 40, av. Halley - Bât A - Park Plaza, 59650 Villeneuve d’Ascq, France
Email address: martine.labbe@ulb.be

Article 6

On a Computationally Ill-Behaved Bilevel Problem with a Continuous and Nonconvex Lower Level

Yasmine Beck, Daniel Bienstock, Martin Schmidt, and Johannes Thürauf
Journal of Optimization Theory and Applications (2023)
https://doi.org/10.1007/s10957-023-02238-9

ON A COMPUTATIONALLY ILL-BEHAVED BILEVEL PROBLEM WITH A CONTINUOUS AND NONCONVEX LOWER LEVEL

YASMINE BECK, DANIEL BIENSTOCK, MARTIN SCHMIDT, JOHANNES THÜRAUF

ABSTRACT. It is well known that bilevel optimization problems are hard to solve both in theory and practice. In this paper, we highlight a further computational difficulty when it comes to solving bilevel problems with continuous but nonconvex lower levels. Even if the lower-level problem is solved to ε-feasibility regarding its nonlinear constraints for an arbitrarily small but positive ε, the obtained bilevel solution as well as its objective value may be arbitrarily far away from the actual bilevel solution and its actual objective value. This result even holds for bilevel problems for which the nonconvex lower level is uniquely solvable, for which the strict complementarity condition holds, for which the feasible set is convex, and for which Slater's constraint qualification is satisfied for all feasible upper-level decisions. Since the consideration of ε-feasibility cannot be avoided when solving nonconvex problems to global optimality, our result shows that computational bilevel optimization with continuous and nonconvex lower levels needs to be done with great care. Finally, we illustrate that the nonlinearities in the lower level are the key reason for the observed bad behavior by showing that linear bilevel problems behave much better at least on the level of feasible solutions.

1. INTRODUCTION

Bilevel optimization problems are known to be notoriously hard to solve and this holds true both in theory and in practice. In theory, bilevel problems are strongly NP-hard even if all objective functions and constraints are linear and all variables are continuous; see Hansen et al. (1992). This, of course, is also reflected in computational practice since linear bilevel problems are inherently nonsmooth and nonconvex problems. Moreover, the single-level reformulations used to solve linear bilevel problems in practice are nonconvex, complementarity-constrained problems. Their linearization requires big-M parameters that are hard to obtain in general (Kleinert et al. 2020) and that often lead to numerically badly posed problems, which are hard to tackle even for state-of-the-art commercial solvers.

In the last years, algorithmic research on bilevel optimization focused on more and more complicated lower-level problems such as mixed-integer linear models (Fischetti et al. 2017, 2018), nonlinear but still convex models (Kleinert et al. 2021a), or problems in the lower level with uncertain data (Beck et al. 2023; Buchheim and Henke 2022; Burtscheidt and Claus 2020). When it comes to the situation of a lower-level problem with continuous nonlinearities there is not too much literature—in particular in comparison to the case in which the lower-level problem is convex; see, e.g., Kleniati and Adjiman (2011, 2014a,b, 2015), Mitsos (2010), Mitsos et al. (2008), Paulavičius and Adjiman (2020), Paulavičius et al. (2020), and Paulavičius et al. (2016). Due to the brevity of this article, we do not go into the details of

---

Date: May 3, 2023.

2010 *Mathematics Subject Classification.* 90-XX, 90C26, 90C31.

Key words and phrases. Bilevel optimization, Nonconvex lower levels, Approximate feasibility, Global optimization.

the literature but refer to the seminal book by Dempe (2002) as well as the recent survey by Kleinert et al. (2021b) for further discussions of the relevant literature.

There is one important difference when crossing the border from (mixed-integer) convex to (mixed-integer) nonconvex lower-level problems. The lower-level problem can, in general, not be solved to global optimality anymore in an exact sense in finite time since we need to exploit techniques such as spatial branching to tackle nonconvexities. These techniques only lead to finite algorithms for prescribed and strictly positive feasibility tolerances; see, e.g., Locatelli and Schoen (2013) for more detailed discussions. Note that this is in clear contrast to, e.g., linear optimization. Here, the simplex method is an exact method in the sense that, if applied using exact arithmetic, the method computes a global optimal solution without any errors; see, e.g., Applegate et al. (2007). The same applies to simplex-based branch-and-bound methods for solving mixed-integer linear optimization problems. Algorithms as such are not available for continuous but nonconvex problems. This means that, just due to algorithmic necessities, we cannot expect to get exact feasible solutions of the lower-level problem anymore when doing computations for continuous but nonconvex lower-level problems. Instead, we have to deal with ε-feasible solutions—at least for the nonlinear constraints of the lower-level problem.

The aim of this paper is to present an exemplary bilevel optimization problem with continuous variables and a nonconvex lower-level problem, where the latter algorithmic aspect leads to the following severe issue:

> *Even if the feasibility tolerance for the lower level can be made extremely small, the exact bilevel solution can be arbitrarily far away from the bilevel solution that one obtains for ε-feasibility in the lower level, which in particular can be superoptimal out of proportion to ε. The same is true for the optimal objective function values.*

The main idea for the construction of this exemplary bilevel problem is based on a constraint set presented first by Bienstock et al. (2021). We explicitly note here that this construction does not make use of large constraint coefficient ranges (all coefficients are 1) or arbitrarily large degrees of polynomials (we only use quadratic or linear terms). Moreover, when considered in an exact sense, (i) the example's lower-level problem is uniquely solvable, (ii) strict complementarity holds, (iii) its convex constraint set satisfies Slater's constraint qualification for all feasible upper-level decisions, (iv) the upper level does not contain coupling constraints, and (v) the overall problem has a unique global solution as well. Thus, the bilevel program does not look like a badly-modeled problem but is shown to behave very badly in a computational sense, i.e., if only ε-feasible points for the nonlinear constraints of the lower-level problem can be considered. We also show that the observed pathological behavior arises due to the nonlinearities as we show that linear bilevel problems behave much better at least on the level of feasible points.

The example is presented in Section 2 and discussed in an exact sense in Section 3. Afterward, the example is analyzed in Section 4 for the case of ε-feasibility of nonlinear constraints. Section 5 presents an analysis of the linear bilevel case. Our final conclusions are drawn in Section 6.

2. Problem Statement

Let us consider the bilevel problem

$$\max_{x\in\mathbb{R}^2} \quad F(x,y) = x_1 - 2y_{n+1} + y_{n+2} \tag{1a}$$

$$\text{s.t.} \quad (x_1,x_2) \in [\underline{x}_1, \bar{x}_1] \times [\underline{x}_2, \bar{x}_2], \tag{1b}$$

$$y \in S(x), \tag{1c}$$

where $\underline{x}$, $\bar{x} \in \mathbb{R}^2$ with $1 \leq \underline{x}_i < \bar{x}_i$, $i \in \{1, 2\}$, denote lower and upper bounds on the variables x. Here, $S(x)$ is the set of optimal solutions of the x-parameterized problem

$$\max_{y \in \mathbb{R}^{n+2}} \quad f(x, y) = y_1 - y_n \left(x_1 + x_2 - y_{n+1} - y_{n+2}\right) \tag{2a}$$

$$\text{s.t.} \quad y_1 + y_n = \frac{1}{2}, \tag{2b}$$

$$y_i^2 \leq y_{i+1}, \quad i \in \{1, \ldots, n-1\}, \tag{2c}$$

$$y_i \geq 0, \quad i \in \{1, \ldots, n\}, \tag{2d}$$

$$y_{n+1} \in [0, x_1], \tag{2e}$$

$$y_{n+2} \in [-x_2, x_2]. \tag{2f}$$

We refer to Problem (1) as the upper-level (or the leader's) problem and to Problem (2) as the lower-level (or the follower's) problem. Let us point out that the lower-level constraints (2b) and (2c) together with $y_1 \geq 0$ have already been considered in Bienstock et al. (2021) in the context of approximately feasible solutions for single-level optimization problems. Let us further emphasize that the number of variables and constraints of the lower-level problem is linear in n.

Problem (1) is a linear problem in both the leader's and the follower's variables. The only constraints that occur in this problem are variable bounds for the leader's variables x. In particular, there are no upper-level constraints that explicitly depend on the follower's variables y, i.e., there are no coupling constraints.

Moreover, the feasible set of the lower-level problem is bounded due to the following. From (2d) and (2b), we obtain $0 \leq y_1 \leq 1/2$ as well as $0 \leq y_n \leq 1/2$ for any feasible follower's decision y. Using Constraints (2c), we further obtain $0 \leq y_i \leq 1$ for all $i \in \{1, \ldots, n\}$. Finally, we have $0 \leq y_{n+1} \leq \bar{x}_1$ as well as $-\bar{x}_2 \leq y_{n+2} \leq \bar{x}_2$ by (2e) and (2f) because the leader's variables x are bounded. Since all finitely many lower-level constraints are continuous, the feasible set of the follower's problem is compact. In addition to the compactness, the feasible set of the lower-level problem (2) is non-empty for every feasible leader's decision $(x_1, x_2) \in [\underline{x}_1, \bar{x}_1] \times [\underline{x}_2, \bar{x}_2]$. For instance, the point

$$y_i = \frac{i}{2^{2^n}},\ i \in \{1, \ldots, n-1\}, \quad y_n = \frac{1}{2} - \frac{1}{2^{2^n}}, \quad y_{n+1} = \frac{1}{2}, \quad \text{and} \quad y_{n+2} = 0$$

is strictly feasible w.r.t. the inequality constraints (2d), (2e) as well as (2f) and feasible w.r.t. the equality constraint (2b). Here, we exploit the assumption that $1 \leq \underline{x}_1, \underline{x}_2$ holds to obtain strict feasibility w.r.t. the variable bounds in (2e) and (2f). Moreover, y is also strictly feasible w.r.t. the inequality constraints (2c) due to the following. For all $i \in \{1, \ldots, n-2\}$, we have

$$y_{i+1} - y_i^2 = \frac{i+1}{2^{2^n}} - \left(\frac{i}{2^{2^n}}\right)^2 = \frac{2^{2^n}(i+1) - i^2}{\left(2^{2^n}\right)^2} = \frac{2^{2^n} + 2^{2^n} i \left(1 - \frac{i}{2^{2^n}}\right)}{\left(2^{2^n}\right)^2} > 0.$$

Furthermore, we have

$$\begin{aligned} y_n - y_{n-1}^2 &= \frac{1}{2} - \frac{1}{2^{2^n}} - \left(\frac{n-1}{2^{2^n}}\right)^2 = \frac{\left(2^{2^n}\right)^2 - 2 \cdot 2^{2^n} - 2(n-1)^2}{2 \cdot \left(2^{2^n}\right)^2} \\ &= \frac{2^{2^n}\left(2^{2^n} - 2 - \frac{2(n-1)^2}{2^{2^n}}\right)}{2 \cdot \left(2^{2^n}\right)^2} > 0. \end{aligned}$$

In particular, this means that the problem satisfies Slater's constraint qualification. Moreover, the gradient of the single equality constraint (2b) is not the null vector. Hence, the Mangasarian–Fromovitz constraint qualification (MFCQ) is also satisfied at every feasible decision of the follower. Let us further point out that all lower-level constraints are linear except for the quadratic but convex inequality constraints in (2c). Therefore, the feasible set of the lower-level problem (2) is convex. Nevertheless, the overall lower-level problem is nonconvex since the follower's objective function contains bilinear terms.

Before we solve the bilevel problem (1) and (2) in the following sections, let us briefly summarize the nice properties of the problem. The upper-level problem is linear and does not contain coupling constraints. The feasible set of the lower-level problem is convex and compact. For every feasible leader's decision, the lower-level problem further satisfies Slater's constraint qualification and the MFCQ is satisfied for every feasible follower's decision.

3. EXACT FEASIBILITY

In this section, we determine the unique exact solution of the bilevel problem (1) and (2). To this end, we start by solving the lower-level problem (2) analytically for an arbitrary but fixed feasible leader's decision $(x_1, x_2) \in [\underline{x}_1, \bar{x}_1] \times [\underline{x}_2, \bar{x}_2]$.

First, we note that any feasible follower's decision y satisfies $y_n > 0$. The reasons are as follows. Let us contrarily assume that $y_n = 0$ holds. Then, Constraint (2b) yields $y_1 = 1/2$. From $y_n = 0$ and (2c), it follows that $y_i = 0$ holds for all $i \in \{1, \dots, n\}$, which contradicts $y_1 = 1/2$. Consequently, $y_n > 0$ holds. For later reference, let us briefly summarize the previous observation.

Result 1. *For every feasible leader's decision* $(x_1, x_2) \in [\underline{x}_1, \bar{x}_1] \times [\underline{x}_2, \bar{x}_2]$, *a feasible follower's decision* y *satisfies* $y_n > 0$.

The equality constraint (2b) thus yields $y_1 < 1/2$. From (2e) and (2f), we additionally obtain

$$y_n (x_1 + x_2 - y_{n+1} - y_{n+2}) \geq 0.$$

In particular, the latter term is minimized for $(y_{n+1}, y_{n+2}) = (x_1, x_2)$. Therefore, the lower-level objective function value can be bounded from above by

$$f(x, y) = y_1 - y_n (x_1 + x_2 - y_{n+1} - y_{n+2}) \leq y_1 < \frac{1}{2}.$$

It is thus evident that an optimal follower's decision y^* satisfies $(y^*_{n+1}, y^*_{n+2}) = (x_1, x_2)$. Here, we can fix (y^*_{n+1}, y^*_{n+2}) since these variables are subject to simple bound constraints and, in particular, they are not coupled to the other variables of the follower. Hence, the follower's problem can be reduced to the convex problem

$$\max_{y} \quad y_1 \tag{3a}$$

$$\text{s.t.} \quad y_1 + y_n = \frac{1}{2}, \tag{3b}$$

$$y_i^2 \leq y_{i+1}, \quad i \in \{1, \dots, n-1\}, \tag{3c}$$

$$y_i \geq 0, \quad i \in \{1, \dots, n\}. \tag{3d}$$

As shown above, Problem (3) satisfies Slater's constraint qualification.[1] Again as shown above, the feasible set is compact. Therefore, Problem (3) has an optimal solution y^*. Because of the equality constraint (3b), the lower-level objective function value y^*_1 is maximized by minimizing y^*_n. From Constraints (3c) and the optimality of y^*, we obtain

$$y^*_i = (y^*_1)^{2^{i-1}} \qquad \text{for all } i \in \{2, \dots, n\},$$

---

[1] It can also be shown that the linear independence constraint qualification is valid at all solutions of the follower's problem (for any given leader's decision x). For the latter, see Appendix A.

where y_1^* denotes the root of the function

$$h : \left[0, \frac{1}{2}\right] \to \mathbb{R}, \quad z \mapsto z + z^{2^{n-1}} - \frac{1}{2}. \tag{4}$$

In particular, one can show that y_1^* is the unique root of (4). The function h is continuous and strictly increasing on $[0, 1/2]$. Moreover, we have $h(0) < 0$ and $h(1/2) > 0$. Consequently, there is a unique point $y_1^* \in (0, 1/2)$ such that $h(y_1^*) = 0$ holds. Furthermore, the follower's decision y^* is the unique solution of Problem (3). To see this, let us assume that there is another feasible follower's decision $\hat{y} \neq y^*$ for which the optimal objective function value y_1^* is obtained, i.e., $\hat{y}_1 = y_1^*$. Then, there must be at least one quadratic inequality constraint in (3c) that is not satisfied with equality for $\hat{y}$. Otherwise, we have $y^* = \hat{y}$. However, if there is slack in Constraints (3c), we obtain $\hat{y}_n > y_n^*$. Then, (3b) yields

$$y_1^* = \frac{1}{2} - y_n^* > \frac{1}{2} - \hat{y}_n = \hat{y}_1,$$

which is a contradiction to the optimality of $\hat{y}$.

Result 2. *For every feasible leader's decision* $(x_1, x_2) \in [\underline{x}_1, \bar{x}_1] \times [\underline{x}_2, \bar{x}_2]$, *the set of optimal solutions of the lower-level problem* (2) *is a singleton.*

In particular, Result 2 means that there is no need to distinguish between the optimistic and the pessimistic approach to bilevel optimization; see, e.g., Dempe (2002). Thus, we can finally determine an optimal leader's decision for the overall bilevel problem (1) and (2). As $(y_{n+1}^*, y_{n+2}^*) = (x_1, x_2)$ holds in the optimal follower's decision y^*, the leader actually solves the linear problem

$$\max_x \quad -x_1 + x_2 \quad \text{s.t.} \quad (x_1, x_2) \in [\underline{x}_1, \bar{x}_1] \times [\underline{x}_2, \bar{x}_2].$$

The unique optimal solution is given by $x^* = (\underline{x}_1, \bar{x}_2)$.

Result 3. *The bilevel problem* (1) *and* (2) *has a unique solution given by* $x^* = (\underline{x}_1, \bar{x}_2)$ *with an optimal objective function value of* $F^* = -\underline{x}_1 + \bar{x}_2$.

To sum up, the bilevel problem (1) and (2) not only has nice properties such as a convex and bounded lower-level feasible set as well as a lower-level problem that satisfies Slater's constraint qualification, but also has a unique optimal solution. Moreover, the strict complementarity condition holds for which we give a proof in Appendix B. Overall, the bilevel problem (1) and (2) is thus well-behaved.

4. ε-Feasibility

In what follows, we determine an optimal solution of the bilevel problem (1) and (2) under the assumption that we allow for small violations of the nonlinear lower-level constraints according to the following notion, which is motivated by the necessary special treatment of nonlinear (and, in particular, nonconvex) constraints in global optimization as we discussed it in the introduction.

Definition 1. *Let* $0 < \varepsilon \in \mathbb{R}$, $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, *and* $h : \mathbb{R}^n \to \mathbb{R}^p$ *be given. A point* $x \in \mathbb{R}^n$ *is called* ε-feasible *for the problem* $\max_{x \in \mathbb{R}^n} \{f(x) \colon g(x) \leq 0, h(x) = 0\}$ *if* $g_i(x) \leq 0$ *and* $h_j(x) = 0$ *holds for all* $i \in \{1, \dots, m\} \setminus N$ *as well as for all* $j \in \{1, \dots, p\} \setminus M$ *and if* $\max\{\max\{g_i(x) \colon i \in N\}, \max\{|h_j(x)| \colon j \in M\}\} \leq \varepsilon$ *holds, where* $N \subseteq \{1, \dots, m\}$ *and* $M \subseteq \{1, \dots, p\}$ *denote the index sets of all nonlinear inequality and equality constraints.*

A follower's decision of the form

$$y_i = 2^{-2^{i-1}}, \; i \in \{1, \dots, n-1\}, \; y_n = 0, \; y_{n+1} \in [0, x_1], \text{ and } y_{n+2} \in [-x_2, x_2] \tag{5}$$

is ε-feasible with $\varepsilon = 2^{-2^{n-1}}$ for every feasible leader's decision $(x_1, x_2) \in [\underline{x}_1, \bar{x}_1] \times [\underline{x}_2, \bar{x}_2]$ due to the following. The constraints

$$\begin{aligned} y_1 + y_n &= \frac{1}{2}, \\ y_i^2 &\le y_{i+1}, \quad i \in \{1, \dots, n-2\}, \\ y_i &\ge 0, \quad i \in \{1, \dots, n\}, \\ y_{n+1} &\in [0, x_1], \\ y_{n+2} &\in [-x_2, x_2] \end{aligned}$$

are (exactly) satisfied, whereas only the constraint $y_{n-1}^2 \le y_n$ is violated by $\varepsilon = 2^{-2^{n-1}}$. Moreover, the lower-level objective function value is 1/2.

Result 4. *If $\varepsilon \ge 2^{-2^{n-1}}$, there is an ε-feasible follower's decision y with $y_n = 0$ for every feasible leader's decision $(x_1, x_2) \in [\underline{x}_1, \bar{x}_1] \times [\underline{x}_2, \bar{x}_2]$.*

It can easily be seen that by increasing n, we can obtain arbitrarily small values for ε. In particular, there is no ε-feasible follower's decision that yields a better objective function value than 1/2. The reasons are as follows. Using the equality constraint (2b), the lower-level objective function can be re-written as

$$f(x, y) = \frac{1}{2} - y_n - y_n \left(x_1 + x_2 - y_{n+1} - y_{n+2}\right).$$

For all ε-feasible follower's decisions, we have

$$x_1 + x_2 - y_{n+1} - y_{n+2} \ge 0$$

because of the linear constraints (2e) and (2f). Consequently, a lower-level objective function value larger than 1/2 could only be obtained if $y_n < 0$. However, this is not ε-feasible w.r.t. the variable bounds (2d). In addition, a follower's decision of the form stated in (5) is thus an ε-feasible solution of the lower-level problem (2). Let us point out that, in contrast to the exact case, the follower's variables y_{n+1} and y_{n+2} do not affect the lower-level objective function value in this setting and can thus be chosen arbitrarily. Therefore, the set of ε-feasible follower's solutions is not a singleton anymore.

Result 5. *If $\varepsilon \ge 2^{-2^{n-1}}$, the set of ε-feasible follower's solutions is not a singleton for every feasible leader's decision $(x_1, x_2) \in [\underline{x}_1, \bar{x}_1] \times [\underline{x}_2, \bar{x}_2]$.*

Due to Result 5, we need to distinguish between optimistic and pessimistic solutions. Following the optimistic approach, the follower chooses $y_{n+1} = 0$ as well as $y_{n+2} = x_2$ such as to favor the leader w.r.t. the leader's objective function value. Therefore, the leader actually solves the linear problem

$$\max_x \quad x_1 + x_2 \quad \text{s.t.} \quad (x_1, x_2) \in [\underline{x}_1, \bar{x}_1] \times [\underline{x}_2, \bar{x}_2].$$

The optimistic optimal leader's decision is thus given by $x^* = (\bar{x}_1, \bar{x}_2)$. In the pessimistic case, the follower chooses $y_{n+1} = x_1$ as well as $y_{n+2} = -x_2$ such as to adversely affect the leader's decision. In this setting, the leader solves the linear problem

$$\max_x \quad -x_1 - x_2 \quad \text{s.t.} \quad (x_1, x_2) \in [\underline{x}_1, \bar{x}_1] \times [\underline{x}_2, \bar{x}_2].$$

Hence, the pessimistic optimal leader's decision is given by $x^* = (\underline{x}_1, \underline{x}_2)$. To sum up, let us state the main observations of this section.

Result 6. *Let $\varepsilon \ge 2^{-2^{n-1}}$ and suppose that we allow for ε-feasible follower's solutions. Then, the optimistic optimal solution of the bilevel problem (1) and (2) is given by $x_o^* = (\bar{x}_1, \bar{x}_2)$ with an optimal objective function value of $F_o^* = \bar{x}_1 + \bar{x}_2$.*

The pessimistic optimal solution is given by $x_p^* = (\underline{x}_1, \underline{x}_2)$ *with an optimal objective function value of* $F_p^* = -\underline{x}_1 - \underline{x}_2$.

We now finally compare the results of the exact bilevel solution with the results for the optimistic and pessimistic setting for the case of only ε-feasibility of the lower level. In the optimistic setting, the distance between the solutions is $\bar{x}_1 - \underline{x}_1$ and the difference between the corresponding objective function values is $\bar{x}_1 + \underline{x}_1$. Two aspects are remarkable. First, by enlarging the feasible interval for the variable x_1, we get an arbitrarily large error and, second, this error is independent of ε, i.e., this arbitrarily large error occurs independent of how accurate one solves the lower-level problem.

For the pessimistic setting, the distance between the solution is $\bar{x}_2 - \underline{x}_2$ and the difference between the objective function values is $\bar{x}_2 + \underline{x}_2$. Hence, we obtain the same qualitative behavior but now in dependence of the variable x_2 instead of x_1.

In summary, we obtain the following two main observations. First, we can be arbitrarily far away from the overall exact bilevel solution. Second, we also obtain arbitrarily large errors regarding the optimal objective function value of the leader. The latter is very much in contrast to the situation in single-level optimization for which sensitivity results are available; see, e.g., Proposition 4.2.2 in Bertsekas (2016). This is particularly the case for linear optimization problems, where standard sensitivity analysis results (see, e.g., Theorem 5.5 in Chvátal (1983)) apply as well and state that a small change in the right-hand side of the problem's constraints can only lead to a small change in the optimal objective function value.

Lastly, let us comment on that only very moderate values of n are required to get the wrong solution. Taking the inequality for ε from Result 6, it is easy to see that for a given tolerance ε, the parameter n needs to satisfy $n \geq \log_2(\log_2(1/\varepsilon^2))$ so that numerically computed solutions do not coincide with the exact solution for the given ε. For instance, a tolerance of $\varepsilon = 10^{-8}$ already leads to a wrong result for $n = 6$. This particularly means that the considered bilevel problem is moderate in size w.r.t. the number of constraints and variables. For $n = 6$, we only have 16 constraints and 8 variables on the lower level. We further note that the used constraint coefficients are all 1 and that the coefficients are independent from n and the given tolerance ε.

A Python code for the example considered in this paper is publicly available at `https://github.com/m-schmidt-math-opt/ill-behaved-bilevel-example` and can be used to verify the discussed results.

5. Analysis of the ε-Feasible Linear Case

In this section, we analyze the linear bilevel case, i.e., we study the problem

$$\min_{x,y} \quad c_x^\top x + c_y^\top y \tag{6a}$$

$$\text{s.t.} \quad Ax \geq a, \tag{6b}$$

$$y \in \arg\min_{\bar{y}} \left\{ d^\top \bar{y} \colon Cx + D\bar{y} \geq b \right\} \tag{6c}$$

with $c_x \in \mathbb{R}^{n_x}$, $c_y, d \in \mathbb{R}^{n_y}$, $A \in \mathbb{R}^{m \times n_x}$, $a \in \mathbb{R}^m$, $C \in \mathbb{R}^{\ell \times n_x}$, $D \in \mathbb{R}^{\ell \times n_y}$, and $b \in \mathbb{R}^\ell$. We assume that the set $\{(x, y) \in \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} : Ax \geq a,\ Cx + Dy \geq b\}$ is non-empty and compact and that for every feasible upper-level decision x, there exists a feasible lower-level decision y. This implies that the lower-level problem is bounded for every feasible upper-level decision and that the dual problem of the lower level is feasible. We also assume that the set $\{x \in \mathbb{R}^{n_x} : Ax \geq a\}$ is bounded. Moreover, we consider the setting in which the underlying linear algebra and linear optimization routines are of finite precision only.

When finite-precision procedures are used, an algorithm that solves Problem (6) will output a pair $(\hat{x}, \hat{y})$ that may be slightly infeasible. The concern, should that happen, is that the solution being output can be *superoptimal* to a degree that is not proportional to its infeasibility. As discussed in the previous section, such an outcome can be observed for general, i.e., nonlinear, bilevel problems. In this section, however, we show that linear bilevel problems behave better in some sense. To this end, we assume that our underlying solver can ensure the following properties:

- $A\hat{x} \geq a - \varepsilon e_m$ and $C\hat{x} + D\hat{y} \geq b - \varepsilon e_\ell$,
- $d^\top \hat{y} \geq \min\{d^\top y \colon C\hat{x} + Dy \geq b\} - \varepsilon$.

Here and in what follows, $0 < \varepsilon < 1$ is a given tolerance, $e_k \in \mathbb{R}^k$ is the vector of all ones, and $(\hat{x}, \hat{y})$ is used to denote a nearly feasible solution of the bilevel problem (6).

Prior to our analysis, we present a general result that will be used below. This result can be read from Theorem 3.38 (Page 112) of Conforti et al. (2014). It can also be obtained from Corollary 3.2b (Page 20) of Schrijver (1986) or from Theorem 10.2 (Page 121) of Schrijver (1986). We will use the term *size* to refer to the (bit) encoding length of a matrix, vector, or formulation, as appropriate.

Definition 2. *Let $P = \{x \in \mathbb{R}^n \colon Hx = h,\ x \geq 0\}$ with $H \in \mathbb{R}^{m \times n}$ and $h \in \mathbb{R}^m$. Given $z \in \mathbb{R}^n$, we say that z is* basic *if $Hz = h$ and, defining $B = B(z) = \{j \colon z_j \neq 0\}$, the submatrix H_B of H corresponding to the columns in B has rank $|B|$. Furthermore, if in addition $z \geq 0$, we say that z is* basic feasible.

Remark 1. *Let P be as in Definition 2. The extreme points of P are precisely the vectors z that are basic feasible.*

Theorem 1. *Let $P = \{x \in \mathbb{R}^n \colon Hx = h,\ x \geq 0\}$ with $H \in \mathbb{R}^{m \times n}$ and $h \in \mathbb{R}^m$. There is a constant $\kappa(H) > 0$ of size polynomial in the size (of the bit-encoding) of H such that, for any basic vector v, we have*

$$\|v\|_\infty \leq \kappa(H) \|h\|_\infty.$$

Proof. Let v be basic and set $J = \{j \colon v_j \neq 0\}$. Since v is basic, there is a subset of rows I of H with $|I| = |J|$ such that the following holds:

(i) The submatrix $H_{I,J}$ of H indexed by rows I and columns J is invertible.
(ii) As a consequence, it holds $v_J = H_{I,J}^{-1} h_I$, where v_J is the subvector of v indexed by J and h_I is the subvector of h indexed by I.

Using submultiplicativity of the norm, we get

$$\|v\|_\infty \leq \|H_{I,J}^{-1}\|_\infty \|h_I\|_\infty.$$

The result now follows by defining $\kappa(H)$ to be the maximum over all $\|B^{-1}\|_\infty$ for B being an invertible submatrix of H. ☐

In Theorem 1, we use what is usually termed the *standard* representation of a polyhedron. Similar statements can be derived using other representations of polyhedra, e.g., $\{x \in \mathbb{R}^n \colon Hx \leq h\}$, via well-known reformulations.

5.1. **Linear Optimization with Errors.** We start with some simple observations for classic, i.e., single-level, linear problems of the form

$$v^* := \min_{x \in \mathbb{R}^{n_x}} \left\{ v^\top x \colon Mx \geq f \right\} \tag{7}$$

with $v \in \mathbb{R}^{n_x}$, $0 \neq M \in \mathbb{R}^{m \times n_x}$, and $f \in \mathbb{R}^m$. Throughout this section, we assume that the feasible region for problem (7) is non-empty and bounded. Moreover, we denote the corresponding dual problem by

$$\max_{z \in \mathbb{R}^m} \left\{ f^\top z \colon M^\top z = v,\ z \geq 0 \right\}.$$

Next, we will derive estimates involving near-feasible and near-optimal points for Problem (7).

Lemma 1. *Suppose that there is a point* $\hat{x} \in \mathbb{R}^{n_x}$ *that is nearly feasible for Problem* (7), *i.e.,* $M\hat{x} \geq f - \varepsilon e_m$. *Then, the following holds.*

(a) *It holds*

$$v^\top \hat{x} \geq v^* - \varepsilon\kappa(M)\|v\|_\infty,$$

where $\kappa(M) > 0$ *is a constant of polynomial size.*

(b) *There exists* x^* *feasible for Problem* (7) *such that*

$$\|x^* - \hat{x}\|_\infty \leq \varepsilon\kappa_1(M)$$

holds for a certain constant $\kappa_1(M) > 0$ *of polynomial size.*

Proof. (a) Let z^* be an optimal solution of the dual problem of (7). Then,

$$v^\top \hat{x} = (z^*)^\top M\hat{x} \geq (z^*)^\top (f - \varepsilon e_m) = v^* - \varepsilon\|z^*\|_1$$

holds. In particular, this equation applies to any dual optimal z^*. Since $z \geq 0$ is a constraint of the dual problem, the dual feasible region is a pointed polyhedron, and, w.l.o.g., z^* is an extreme point. The result now follows from Theorem 1.

(b) Consider the linear optimization problem

$$\min_{x,\delta} \quad \delta \tag{8a}$$

$$\text{s.t.} \quad \|x - \hat{x}\|_\infty \leq \delta, \tag{8b}$$

$$Mx \geq f. \tag{8c}$$

That this is indeed a linear program follows by reformulating (8b) as

$$x_j - \delta \leq \hat{x}_j, \quad -x_j - \delta \leq -\hat{x}_j, \quad 1 \leq j \leq n_x. \tag{9}$$

Clearly, the resulting problem is both feasible and bounded since (7) is. Moreover, $(\hat{x}, 0)$ satisfies the constraints of this problem with additive error of at most ε. We can therefore apply (a) to this problem to obtain x^* being feasible for Problem (7) and such that

$$\|x^* - \hat{x}\|_\infty \leq \varepsilon\kappa_1(M)$$

holds, where $\kappa_1(M)$ is the κ-constant (of polynomial size in M) that applies to the matrix for Constraints (9) and (8c). ☐

Let us emphasize that the result in Lemma 1 applies for any $\varepsilon > 0$, no matter how large. In particular, it is not required that ε is "sufficiently small".

Lemma 2. *Suppose that there is a nearly primal-dual feasible and nearly primal-dual optimal pair* $(\hat{x}, \hat{z}) \in \mathbb{R}^{n_x} \times \mathbb{R}^m$ *for Problem* (7), *i.e.,* $(\hat{x}, \hat{z})$ *satisfies*

(i) $M\hat{x} \geq f - \varepsilon e_m$,
(ii) $\|M^\top \hat{z} - v\|_\infty \leq \varepsilon$, $\hat{z} \geq -\varepsilon e_m$, *and*
(iii) $v^\top \hat{x} - f^\top \hat{z} \leq \varepsilon$.

Then, there exists an optimal solution x^* *for Problem* (7) *such that*

$$\|x^* - \hat{x}\|_\infty \leq \varepsilon\kappa_3(M, v)\max\{1, \|f\|_\infty\}$$

holds for a certain constant $\kappa_3(M, v) > 0$, *whose size is polynomial in the size of the input data* M *and* v.

Proof. First, we note that Condition (ii) simply states that $\hat{z}$ is feasible for the dual of (7) up to an error of ε. We can thus apply Part (a) of Lemma 1 to obtain

$$f^\top \hat{z} \le v^* + \varepsilon\kappa_2(M)\|f\|_\infty, \tag{10}$$

where $\kappa_2(M)$ is the κ-constant for the dual of (7), which is of polynomial size in M. Together with (iii), this implies

$$v^\top \hat{x} \le v^* + \varepsilon(1 + \kappa_2(M)\|f\|_\infty). \tag{11}$$

Next, we consider the polyhedron given by

$$Mx \ge f, \tag{12a}$$

$$v^\top x \le v^*, \tag{12b}$$

which is feasible and bounded. By (i) and (11), $\hat{x}$ satisfies these inequalities with feasibility error of at most $\varepsilon(1+\kappa_2(M)\|f\|_\infty)$. By applying Part (b) of Lemma 1, we obtain that there is a feasible point x^* for (12), i.e., an optimal solution x^* for (7), such that

$$\|x^* - \hat{x}\|_\infty \le \varepsilon\kappa_1(M, v)(1 + \kappa_2(M)\|f\|_\infty)$$

holds. □

5.2. **Application to Linear Bilevel Problems.** We now return to the bilevel setup as stated in (6). To this end, note that for a given upper-level decision $x \in \mathbb{R}^{n_x}$, the dual of the lower-level problem (6c) reads

$$\max_z \quad (b - Cx)^\top z \tag{13a}$$

$$\text{s.t.} \quad D^\top z = d, \tag{13b}$$

$$z \ge 0. \tag{13c}$$

Lemma 3. *Let $x \in \mathbb{R}^{n_x}$, $\hat{y} \in \mathbb{R}^{n_y}$, and $\hat{z} \in \mathbb{R}^\ell$ be such that*

(i) $Ax \ge a$,
(ii) $D\hat{y} \ge b - Cx - \varepsilon e_\ell$,
(iii) $\|D^\top \hat{z} - d\|_\infty \le \varepsilon$, $\hat{z} \ge -\varepsilon e_\ell$, *and*
(iv) $d^\top \hat{y} - (b - Cx)^\top \hat{z} \le \varepsilon$.

Then, there exists an optimal solution y^* for the x-parameterized lower-level problem (6c) such that

$$\|y^* - \hat{y}\|_\infty \le \varepsilon\kappa_4(A, C, D, a, b, d) \tag{14}$$

holds for a constant $\kappa_4(A, C, D, a, b, d) > 0$, whose size is polynomial in the size of the input data A, C, D, a, b, and d.

Proof. By assumption, for a given x, the lower-level problem is feasible and bounded. We can thus apply Lemma 2 since Conditions (ii)–(iv) correspond to Conditions (i)–(iii) of Lemma 2. Thus, there exists an optimal point y^* for the lower-level problem such that

$$\|y^* - \hat{y}\|_\infty \le \varepsilon\kappa_3(D, d)\max\{1, \|b - Cx\|_\infty\}$$

holds. Using the triangle inequality and the submultiplicativity of the norm, we obtain

$$\|b - Cx\|_\infty \;\le\; \|b\|_\infty + \|C\|_\infty\|x\|_\infty.$$

Since the feasible region for the upper-level problem is bounded, $\|x\|_\infty$ is upper bounded by the ∞-norm of some extreme point. We can apply Theorem 1 to obtain

$$\|x\|_\infty \le \kappa'(A)\|a\|_\infty,$$

where $\kappa'(A)$ is the κ-constant (of polynomial size in A) for the system $Ax \ge a$. The proof is now concluded by appropriately defining $\kappa_4(A, C, D, a, b, d)$. □

Now, we consider the entire bilevel problem (6) and recall a basic definition from linear optimization.

Definition 3. *Let* $z \in \mathbb{R}^\ell$ *satisfy* $D^\top z = d$ *and define* $B = B(z) = \{j : z_j \neq 0\}$. *We say that* z *is* dual basic *if the submatrix* $D_B^\top$ *of* $D^\top$ *corresponding to the columns in* B *has rank* $|B|$.

Note that z is dual basic and feasible (i.e., $z \geq 0$) if and only if z is an extreme point of the dual polyhedron to any lower-level problem.

Theorem 2. *Let* $\hat{x} \in \mathbb{R}^{n_x}$, $\hat{y} \in \mathbb{R}^{n_y}$, *and* $\hat{z} \in \mathbb{R}^\ell$ *be such that*

(i) $A\hat{x} \geq a - \varepsilon e_m$,
(ii) $D\hat{y} \geq b - C\hat{x} - \varepsilon e_\ell$,
(iii) $\|D^\top \hat{z} - d\|_\infty \leq \varepsilon$, $\hat{z} \geq -\varepsilon e_\ell$,
(iv) $d^\top \hat{y} - (b - C\hat{x})^\top \hat{z} \leq \varepsilon$,
(v) $\|\hat{z} - \tilde{z}\|_\infty \leq \varepsilon$ *for some dual basic* $\tilde{z}$.

Then, there exists a pair (x^*, y^*) *that is feasible for the bilevel problem* (6) *such that*

$$\|(x^*, y^*)^\top - (\hat{x}, \hat{y})^\top\|_\infty \leq \varepsilon\kappa_5(A, C, D, a, b, d),$$
$$|c_x^\top x^* + c_y^\top y^* - (c_x^\top \hat{x} + c_y^\top \hat{y})| \leq \varepsilon\kappa_6(A, C, D, a, b, c, d)$$

hold for certain constants $\kappa_5(A, C, D, a, b, d)$ *and* $\kappa_6(A, C, D, a, b, c, d) > 0$, *whose sizes are polynomial in the size of the input data.*

Proof. By Assumptions (i) and (ii), the pair $(\hat{x}, \hat{y}) \in \mathbb{R}^{n_x} \times \mathbb{R}^{n_y}$ is nearly feasible for the upper- and the lower-level problem of (6). Applying Part (b) of Lemma 1 to $(\hat{x}, \hat{y})$ and the system

$$\begin{bmatrix} A & 0 \\ C & D \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \geq \begin{pmatrix} a \\ b \end{pmatrix} \tag{15}$$

yields (x^*, y') with $Ax^* \geq a$, $Dy' \geq b - Cx^*$, and

$$\|(x^*, y')^\top - (\hat{x}, \hat{y})^\top\|_\infty \leq \varepsilon\kappa_7(A, C, D), \tag{16}$$

where $\kappa_7(A, C, D)$ is the κ-constant for System (15). Next, we use (iv) and obtain

$$\begin{aligned} & d^\top y' - (b - Cx^*)^\top \hat{z} \\ \leq\ & \varepsilon + d^\top (y' - \hat{y}) - (C(\hat{x} - x^*))^\top \hat{z} \\ \leq\ & \varepsilon + \left|d^\top (\hat{y} - y')\right| + \left|(C(\hat{x} - x^*))^\top \hat{z}\right|. \end{aligned} \tag{17}$$

Note that, if u, $v \in \mathbb{R}^n$, then $|u^\top v| \leq \sum_{j=1}^n |u_j v_j| \leq n\|u\|_\infty \|v\|_\infty$ holds. Moreover, if $Q \in \mathbb{R}^{m \times n}$ and $u \in \mathbb{R}^n$, then, for $1 \leq i \leq m$, $|(Qu)_i| \leq (\sum_{j=1}^n |q_{ij}|)\|u\|_\infty \leq \|Q\|_\infty \|u\|_\infty$ by definition of the infinity-norm of a matrix. Hence, using (17) we obtain

$$d^\top y' - (b - Cx^*)^\top \hat{z} \leq \ \varepsilon + n_y\|d\|_\infty \|\hat{y} - y'\|_\infty + \ell\|C\|_\infty \|\hat{x} - x^*\|_\infty \|\hat{z}\|_\infty. \tag{18}$$

Further, since $\tilde{z}$ is dual basic, $\|\tilde{z}\|_\infty$ is upper bounded by $\kappa(D)\|d\|_\infty$ due to Theorem 1. Thus, using (v) yields $\|\hat{z}\|_\infty \leq \varepsilon + \kappa(D)\|d\|_\infty$. These facts, together with (16) and (18), yield

$$d^\top y' - (b - Cx^*)^\top \hat{z} \leq \varepsilon\kappa_8(A, C, D, d),$$

with $\kappa_8(A, C, D, d) \geq 1$ being appropriately defined and of polynomial size. To sum up, x^*, y', and $\hat{z}$ satisfy

(a) $Ax^* \geq a$,
(b) $Dy' \geq b - Cx^*$,
(c) $\|D^\top \hat{z} - d\|_\infty \leq \varepsilon$, $\hat{z} \geq -\varepsilon e_\ell$,
(d) $d^\top y' - (b - Cx^*)^\top \hat{z} \leq \varepsilon\kappa_8(A, C, D, d)$.

Thus, by Lemma 3 applied to the error $\varepsilon\kappa_8(A, C, D, d) \geq \varepsilon$, there exists an optimal solution y^* for the x^*-parameterized lower-level problem (6c) such that

$$\|y^* - y'\|_\infty \leq \varepsilon\kappa_8(A, C, D, d)\kappa_4(A, C, D, a, b, d) \tag{19}$$

holds. Using this inequality and (16) concludes the proof. □

Remark 2. *Assumption (v) in Theorem 2 states that the distance between a nearly feasible and nearly optimal solution for the dual of the lower-level problem and a basic solution for the dual is small, which is a reasonable assumption in our setting.*

To summarize the statement of the theorem, the distance to feasibility and the superoptimality of a nearly feasible pair $(\hat{x}, \hat{y})$ for the bilevel problem (6) is linear in ε with coefficients κ that have polynomial size in the input data. This type of guarantee with polynomial sized coefficients is simply unavailable in the nonlinear case as we have seen in the previous sections.

6. Conclusion

In this paper, we consider an exemplary bilevel problem with continuous variables and a nonconvex lower-level problem and illustrate that numerically obtained solutions can be arbitrarily far away from an exact solution. The discrepancy between exact and numerically computed solutions is based on the fact that we cannot exactly satisfy all constraints of the nonconvex lower level when using global optimization techniques such as spatial branching. The considered problem itself is well-posed in the sense that we do not use large constraint coefficient ranges or high-degree polynomials. Moreover, we show that the constraint set of the lower-level problem is convex, compact, and that it satisfies Slater's constraint qualification. In an exact sense, we prove that the lower-level problem as well as the overall bilevel problem possess unique solutions. It is further established that LICQ holds in every follower's solution for every feasible leader's decision. While working computationally, however, we can only expect to obtain ε-feasible solutions of the nonconvex lower-level problem. Furthermore, the set of ε-feasible follower solutions is not a singleton anymore. Thus, we determine both an optimal solution for the optimistic and the pessimistic variant of the bilevel problem. By doing so, we establish that not only the obtained ε-feasible bilevel solutions can be arbitrarily far away from the overall exact bilevel solution but that there can also be an arbitrarily large error in the objective function value of the leader.

We also show that the pathological behavior observed for nonlinear lower-level problems seems to be due to the nonlinearities by showing that linear bilevel problems behave better at least on the level of feasible points. As an important question for future research, it is still open if one can prove that the bad behavior can also not appear for more general problems than linear ones, such as convex problems, in the lower level.

Finally, our results show that computational bilevel optimization with continuous but nonconvex lower levels needs to be done with great care and that ex-post checks may be needed to avoid considering arbitrarily bad points as "solutions" of the given bilevel problem.

Acknowledgements

The second author is grateful for the funding received based on an ARPA-E GO competition award. Moreover, the third author thanks the Deutsche Forschungsgemeinschaft for their support within projects A05 and B08 in the Sonderforschungsbereich/Transregio 154 "Mathematical Modelling, Simulation and Optimization using the Example of Gas Networks".

References

Applegate, D. L., W. Cook, S. Dash, and D. G. Espinoza (2007). “Exact solutions to linear programming problems.” In: *Operations Research Letters* 35.6, pp. 693–699. DOI: 10.1016/j.orl.2006.12.010.

Beck, Y., I. Ljubić, and M. Schmidt (2023). “A Survey on Bilevel Optimization Under Uncertainty.” In: *European Journal on Operational Research*. DOI: 10.1016/j.ejor.2023.01.008.

Bertsekas, D. P. (2016). *Nonlinear Programming*. Athena scientific Belmont.

Bienstock, D., A. D. Pia, and R. Hildebrand (2021). “Complexity, Exactness, and Rationality in Polynomial Optimization.” In: *Integer Programming and Combinatorial Optimization*. Ed. by M. Singh and D. P. Williamson. Vol. 12707. IPCO 2021. Springer, Cham., pp. 58–72. DOI: 10.1007/978-3-030-73879-2_5.

Buchheim, C. and D. Henke (2022). “The robust bilevel continuous knapsack problem with uncertain coefficients in the follower’s objective.” In: *Journal of Global Optimization*. DOI: 10.1007/s10898-021-01117-9.

Burtscheidt, J. and M. Claus (2020). “Bilevel Linear Optimization Under Uncertainty.” In: *Bilevel Optimization: Advances and Next Challenges*. Ed. by S. Dempe and A. Zemkoho. Springer International Publishing, pp. 485–511. DOI: 10.1007/978-3-030-52119-6_17.

Chvátal, V. (1983). *Linear Programming*. A Series of books in the mathematical sciences. New York (N. Y.): Freeman.

Conforti, M., G. Cornuéjols, and G. Zambelli (2014). *Integer Programming*. Springer International Publishing. DOI: 10.1007/978-3-319-11008-0.

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

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.

– (2018). “On the use of intersection cuts for bilevel optimization.” In: *Mathematical Programming* 172.1-2, pp. 77–103. DOI: 10.1007/s10107-017-1189-5.

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.

Kleinert, T., V. Grimm, and M. Schmidt (2021a). “Outer Approximation for Global Optimization of Mixed-Integer Quadratic Bilevel Problems.” In: *Mathematical Programming (Series B)*. DOI: 10.1007/s10107-020-01601-2.

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

Kleinert, T., M. Labbé, F. Plein, and M. Schmidt (2020). “There’s No Free Lunch: On the Hardness of Choosing a Correct Big-M in Bilevel Optimization.” In: *Operations Research* 68.6, pp. 1716–1721. DOI: 10.1287/opre.2019.1944.

Kleniati, P.-M. and C. S. Adjiman (2011). “Branch-and-Sandwich: An Algorithm for Optimistic Bi-Level Programming Problems.” In: *21st European Symposium on Computer Aided Process Engineering*. Ed. by E. Pistikopoulos, M. Georgiadis, and A. Kokossis. Vol. 29. Computer Aided Chemical Engineering. Elsevier, pp. 602–606. DOI: 10.1016/B978-0-444-53711-9.50121-8.

– (2014a). “Branch-and-Sandwich: a deterministic global optimization algorithm for optimistic bilevel programming problems. Part I: Theoretical development.” In: *Journal of Global Optimization* 60.3, pp. 425–458. DOI: 10.1007/s10898-013-0121-7.

Kleniati, P.-M. and C. S. Adjiman (2014b). “Branch-and-Sandwich: a deterministic global optimization algorithm for optimistic bilevel programming problems. Part II: Convergence analysis and numerical results.” In: *Journal of Global Optimization* 60.3, pp. 459–481. DOI: `10.1007/s10898-013-0120-8`.

– (2015). “A generalization of the Branch-and-Sandwich algorithm: From continuous to mixed-integer nonlinear bilevel problems.” In: *Computers & Chemical Engineering* 72, pp. 373–386. DOI: `10.1016/j.compchemeng.2014.06.004`.

Locatelli, M. and F. Schoen (2013). *Global Optimization.* Philadelphia, PA: Society for Industrial and Applied Mathematics. DOI: `10.1137/1.9781611972672`.

Mitsos, A. (2010). “Global solution of nonlinear mixed-integer bilevel programs.” In: *Journal of Global Optimization* 47.4, pp. 557–582. DOI: `10.1007/s10898-009-9479-y`.

Mitsos, A., P. Lemonidis, and P. I. Barton (2008). “Global solution of bilevel programs with a nonconvex inner program.” In: *Journal of Global Optimization* 42.4, pp. 475–513. DOI: `10.1007/s10898-007-9260-z`.

Paulavičius, R. and C. S. Adjiman (2020). “New bounding schemes and algorithmic options for the Branch-and-Sandwich algorithm.” In: *Journal of Global Optimization*, pp. 1–29. DOI: `10.1007/s10898-020-00874-3`.

Paulavičius, R., J Gao, P.-M. Kleniati, and C. Adjiman (2020). “BASBL: Branch-And-Sandwich BiLevel solver. Implementation and computational study with the BASBLib test set.” In: *Computers & Chemical Engineering*, p. 106609. DOI: `10.1016/j.compchemeng.2019.106609`.

Paulavičius, R., P.-M. Kleniati, and C. S. Adjiman (2016). “Global optimization of nonconvex bilevel problems: implementation and computational study of the Branch-and-Sandwich algorithm.” In: *26th European Symposium on Computer Aided Process Engineering*. Ed. by Z. Kravanja and M. Bogataj. Vol. 38. Computer Aided Chemical Engineering. Elsevier, pp. 1977–1982. DOI: `10.1016/B978-0-444-63428-3.50334-9`.

Schrijver, A. (1986). *Theory of Linear and Integer Programming.* John Wiley & Sons, Chichester.

APPENDIX A. PROOF OF THE LINEAR INDEPENDENCE CONSTRAINT QUALIFICATION

Let $(x_1, x_2) \in [\underline{x}_1, \bar{x}_1] \times [\underline{x}_2, \bar{x}_2]$ with $1 \leq \underline{x}_i < \bar{x}_i$, $i \in \{1, 2\}$, be arbitrary but fixed. Further, let y^* be the exact optimal solution of the follower for the given leader's decision x. As shown in Section 3, a follower's solution y^* satisfies $y_i^* > 0$ for all $i \in \{1, \ldots, n+2\}$. This means that the non-negativity constraints (2d) as well as the lower bound constraints in (2e) and (2f) are inactive in an optimal follower's decision. Conversely, all quadratic constraints (2c) as well as the upper bound constraints in (2e) and (2f) are active. Hence, the Jacobian matrix of the single equality constraint and the active inequality constraints in an optimal decision of the follower is given by

$$\begin{bmatrix} 1 & & & & & 1 & & \\ 2y_1^* & -1 & & & & & & \\ & 2y_2^* & -1 & & & & & \\ & & \ddots & \ddots & & & & \\ & & & 2y_{n-2}^* & -1 & & & \\ & & & & 2y_{n-1}^* & -1 & & \\ & & & & & & 1 & \\ & & & & & & & 1 \end{bmatrix}.$$

All matrix entries that are left blank here correspond to zeros. It is easy to verify that the Jacobian matrix has full rank, i.e., the linear independence constraint qualification holds.

APPENDIX B. PROOF OF THE STRICT COMPLEMENTARITY CONDITION

Let $(x_1, x_2) \in [\underline{x}_1, \bar{x}_1] \times [\underline{x}_2, \bar{x}_2]$ with $1 \leq \underline{x}_i < \bar{x}_i$, $i \in \{1, 2\}$, be arbitrary but fixed. For the x-parameterized lower-level problem (2), the Lagrangian function reads

$$\begin{aligned}\mathcal{L}(y, \alpha, \beta, \gamma, \delta^{\pm}, \pi) = &- y_1 + y_n \left(x_1 + x_2 - y_{n+1} - y_{n+2}\right) \\ &- \sum_{i=1}^{n-1} \alpha_i \left(y_{i+1} - y_i^2\right) - \sum_{i=1}^{n+1} \beta_i y_i \\ &- \gamma \left(x_1 - y_{n+1}\right) - \delta^- \left(y_{n+2} + x_2\right) \\ &- \delta^+ \left(x_2 - y_{n+2}\right) - \pi \left(y_1 + y_n - \frac{1}{2}\right)\end{aligned}$$

with the Lagrange multipliers $\alpha \in \mathbb{R}^{n-1}_{\geq 0}$, $\beta \in \mathbb{R}^{n+1}_{\geq 0}$, γ, $\delta^{\pm} \in \mathbb{R}_{\geq 0}$, and $\pi \in \mathbb{R}$. The KKT complementarity conditions of Problem (2) are given by

$$\alpha_i \left(y_{i+1} - y_i^2\right) = 0, \quad i \in \{1, \ldots, n-1\}, \tag{20a}$$

$$\beta_i y_i = 0, \quad i \in \{1, \ldots, n+1\}, \tag{20b}$$

$$\gamma \left(x_1 - y_{n+1}\right) = 0, \tag{20c}$$

$$\delta^- \left(y_{n+2} + x_2\right) = 0, \tag{20d}$$

$$\delta^+ \left(x_2 - y_{n+2}\right) = 0. \tag{20e}$$

Let y^* be the exact optimal solution of the follower for the given leader's decision x. Further, let α^*, β^*, γ^*, $(\delta^{\pm})^*$, and π^* be the corresponding Lagrange multipliers so that $(y^*, \alpha^*, \beta^*, \gamma^*, (\delta^{\pm})^*, \pi^*)$ is a KKT point of the lower-level problem. As shown in Appendix A, the linear independence constraint qualification is valid at all solutions of the follower's problem (for any given leader's decision x). Hence, the Lagrange multipliers α^*, β^*, γ^*, $(\delta^{\pm})^*$, and π^* are uniquely determined. Also, as observed in Section 3, the follower's optimal decision y^* satisfies $(y^*_{n+1}, y^*_{n+2}) = (x_1, x_2)$ and

$$y_i^* = (y_1^*)^{2^{i-1}} \qquad \text{for all } i \in \{2, \ldots, n\},$$

where y_1^* is the unique root of the function h as given in (4). We now show that the strict complementarity condition is satisfied.

Observation 1. *The point $(y^*, \alpha^*, \beta^*, \gamma^*, (\delta^{\pm})^*, \pi^*)$ satisfies $\beta_i^* = 0$ for all $i \in \{1, \ldots, n+1\}$ as well as $(\delta^-)^* = 0$.*

Proof. For all $i \in \{1, \ldots, n\}$, we have $y_i^* > 0$. By (20b), we thus obtain $\beta_i^* = 0$ for all $i \in \{1, \ldots, n\}$. From (20b), we further obtain $\beta^*_{n+1} = 0$ since $y^*_{n+1} = x_1 \geq \underline{x}_1 \geq 1 > 0$ holds. Finally, due to $y^*_{n+2} + x_2 = 2x_2 \geq 2\underline{x}_2 \geq 2 > 0$, (20d) yields $(\delta^-)^* = 0$. ☐

Observation 2. *The point $(y^*, \alpha^*, \beta^*, \gamma^*, (\delta^{\pm})^*, \pi^*)$ satisfies γ^*, $(\delta^+)^* > 0$.*

Proof. Since $(y^*, \alpha^*, \beta^*, \gamma^*, (\delta^{\pm})^*, \pi^*)$ is a KKT point,

$$\begin{aligned}\nabla_{y_{n+1}} \mathcal{L}(y^*, \alpha^*, \beta^*, \gamma^*, (\delta^{\pm})^*, \pi^*) &= -y_n^* - \beta^*_{n+1} + \gamma^* = 0, \\ \nabla_{y_{n+2}} \mathcal{L}(y^*, \alpha^*, \beta^*, \gamma^*, (\delta^{\pm})^*, \pi^*) &= -y_n^* - (\delta^-)^* + (\delta^+)^* = 0\end{aligned}$$

are satisfied. By Observation 1 and $y_n^* > 0$, we obtain γ^*, $(\delta^+)^* > 0$. ☐

Observation 3. *The point $(y^*, \alpha^*, \beta^*, \gamma^*, (\delta^{\pm})^*, \pi^*)$ satisfies $2\alpha_1^* y_1^* = 1 + \pi^*$.*

Proof. Since $(y^*, \alpha^*, \beta^*, \gamma^*, (\delta^\pm)^*, \pi^*)$ is a KKT point,

$$\nabla_{y_1} \mathcal{L}(y^*, \alpha^*, \beta^*, \gamma^*, (\delta^\pm)^*, \pi^*) = -1 + 2\alpha_1^* y_1^* - \beta_1^* - \pi^* = 0$$

is satisfied. Since, by Observation 1, we have $\beta_1^* = 0$, we obtain $2\alpha_1^* y_1^* = 1 + \pi^*$. ☐

Observation 4. *The point* $(y^*, \alpha^*, \beta^*, \gamma^*, (\delta^\pm)^*, \pi^*)$ *satisfies* $2\alpha_i^* y_i^* = \alpha_{i-1}^*$ *for all* $i \in \{2, \dots, n-1\}$.

Proof. Since $(y^*, \alpha^*, \beta^*, \gamma^*, (\delta^\pm)^*, \pi^*)$ is a KKT point,

$$\nabla_{y_i} \mathcal{L}(y^*, \alpha^*, \beta^*, \gamma^*, (\delta^\pm)^*, \pi^*) = -\alpha_{i-1}^* + 2\alpha_i^* y_i^* - \beta_i^* = 0$$

is satisfied for all $i \in \{2, \dots, n-1\}$. By Observation 1, we obtain $2\alpha_i^* y_i^* = \alpha_{i-1}^*$ for all $i \in \{2, \dots, n-1\}$. ☐

Observation 5. *The point* $(y^*, \alpha^*, \beta^*, \gamma^*, (\delta^\pm)^*, \pi^*)$ *satisfies* $\pi^* = -\alpha_{n-1}^*$.

Proof. Since $(y^*, \alpha^*, \beta^*, \gamma^*, (\delta^\pm)^*, \pi^*)$ is a KKT point,

$$\nabla_{y_n} \mathcal{L}(y^*, \alpha^*, \beta^*, \gamma^*, (\delta^\pm)^*, \pi^*) = x_1 + x_2 - y_{n+1}^* - y_{n+2}^* - \alpha_{n-1}^* - \beta_n^* - \pi^* = 0$$

is satisfied. By Observation 1, we have $\beta_n^* = 0$. Moreover, $(y_{n+1}^*, y_{n+2}^*) = (x_1, x_2)$ holds. Hence, we obtain $\pi^* = -\alpha_{n-1}^*$. ☐

Observation 6. *The point* $(y^*, \alpha^*, \beta^*, \gamma^*, (\delta^\pm)^*, \pi^*)$ *satisfies* $\alpha_i^* > 0$ *for all* $i \in \{1, \dots, n-1\}$.

Proof. We show this by contradiction. Suppose that $\alpha_1^* = 0$ holds. By Observation 4 and $y_i^* > 0$ for all $i \in \{1, \dots, n\}$, we obtain $\alpha_i^* = 0$ for all $i \in \{1, \dots, n-1\}$. Then, Observation 5 yields $\pi^* = -\alpha_{n-1}^* = 0$. By Observation 3, however, we obtain $0 = 2\alpha_1^* y_1^* = 1 + \pi^*$, i.e., $\pi^* = -1$ which is a contradiction to $\pi^* = 0$. Consequently, $\alpha_1^* > 0$ needs to holds and, thus, we have $\alpha_i^* > 0$ for all $i \in \{1, \dots, n-1\}$ by Observation 4. ☐

To sum up, we have

$$\begin{aligned}
&y_{i+1}^* = (y_i^*)^2 \quad \text{and} \quad \alpha_i^* > 0 \quad \text{for all } i \in \{1, \dots, n-1\}, \\
&y_i^* > 0 \quad \text{and} \quad \beta_i^* = 0 \quad \text{for all } i \in \{1, \dots, n+1\}, \\
&y_{n+1}^* = x_1 \quad \text{and} \quad \gamma^* > 0, \\
&y_{n+2}^* > -x_2 \quad \text{and} \quad (\delta^-)^* = 0, \\
&y_{n+2}^* = x_2 \quad \text{and} \quad (\delta^+)^* > 0,
\end{aligned}$$

i.e., strict complementarity is satisfied.

(Y. Beck, M. Schmidt, J. Thürauf) Trier University, Department of Mathematics, Universitätsring 15, 54296 Trier, Germany
Email address: `yasmine.beck@uni-trier.de`
Email address: `martin.schmidt@uni-trier.de`
Email address: `johannes.thuerauf@uni-trier.de`

(D. Bienstock) Columbia University, Department of Industrial Engineering and Operations Research, 500 W. 120th Street #315, New York, NY 10027, United States
Email address: `dano@columbia.edu`

Article 7

A robust approach for modeling limited observability in bilevel optimization

Yasmine Beck and Martin Schmidt

Operations Research Letters (2021)
https://doi.org/10.1016/j.orl.2021.07.010

A Robust Approach for Modeling Limited Observability in Bilevel Optimization

Yasmine Beck, Martin Schmidt

Abstract. Many applications of bilevel optimization contain a leader facing a follower whose reaction deviates from the one expected by the leader due to some kind of bounded rationality. We consider bilinear bilevel problems with follower's response uncertainty due to limited observability regarding the leader's decision and exploit robust optimization to model the decision making of the follower. We show that the robust counterpart of the lower level allows to tackle the problem via the lower level's KKT conditions.

1. Introduction

In bilevel optimization, some of the variables are constrained to be optimal solutions of another optimization problem, the so-called lower-level problem. The remaining variables are decided on in the so-called upper-level problem. The upper-level player (or leader) makes a decision first, anticipating the reaction of the lower-level player (or follower). The concept of this class of optimization problems dates back to the seminal publications of von Stackelberg (1932; 1954). In the last years and decades, bilevel problems have gained increasing attention due to their ability to model hierarchical decision making processes. These situations arise in various real-world applications such as in energy markets; see, e.g., Ambrosius et al. (2020), Grimm et al. (2019), and Xinmin and Ralph (2007), in transportation; see, e.g., Migdalas (1995) and Ben-Ayed et al. (1992), or in critical infrastructure defense; see, e.g., Brown et al. (2006), DeNegre (2011), Fischetti et al. (2019), Jain et al. (2010), Kiekintveld et al. (2009), Paruchuri et al. (2008), Pita, Jain, Marecki, et al. (2008), and Shieh et al. (2012). Thus, it is obvious that the capability of modeling hierarchical decision processes is important for practice. However, this ability makes bilevel problems intrinsically hard to solve. Even their easiest instantiations, namely linear bilevel problems, are strongly NP-hard; see Hansen et al. (1992).

In the classic setting of bilevel optimization, it is assumed that both players act perfectly rational. However, this assumption rarely holds in many practical applications as both players may face bounded rationality; see, e.g., Simon (1972). For instance, in Chariri (2017), it is elaborated on how decision makers are confronted with cognitive limitations preventing them from reaching a perfectly rational decision. Although Simon's theory received considerable recognition, this notion has long been abstracted from. It has been a point of controversy between economists as it has been accused of being too limited to individual psychological processes rather than that it fits the behavior of institutions and large economies; see, e.g., Dequech (2001) and Rainey (2001). For a more detailed discussion on bounded rationality, we refer to Rubinstein (1998).

Nevertheless, the consideration of bounded rationality has attained increasing attention in recent years. One possible reason for bounded rationality is data uncertainty. In the context of bilevel optimization, these types of problems have been investigated in, e.g., Haghighat (2014), Dempe et al. (2017), Ivanov (2018), Yanıkoğlu and Kuhn (2018), Burtscheidt, Claus, and Dempe (2020), and Burtscheidt and Claus (2020). Another reason

---

2020 *Mathematics Subject Classification.* 90C17, 90C46, 91A05, 91A26, 91A65.

Key words and phrases. Bilevel optimization, Robust optimization, Bounded rationality, Limited observability, Reformulations.

for bounded rationality is uncertainty about the decision of the other player. Thus, it is evident that uncertainty is an important aspect of bounded rationality. In mathematical optimization, one approach to deal with uncertainties is robust optimization (see, e.g., Soyster (1973), Ben-Tal et al. (2009), and Bertsimas et al. (2010)), which we will also exploit in this paper.

In contrast to the case of uncertain data, decision uncertainty has been much less investigated in the context of bilevel optimization. Concerning this matter, we focus on the following two papers. Recently, Besançon et al. (2019) propose a robust approach to hedge against near-optimal lower-level decisions as a generalization of the "ε-approximation" introduced in Wiesemann et al. (2013). Pita, Jain, Ordóñez, et al. (2009) consider follower's response uncertainty due to limited observability regarding the upper-level decision using anchoring biases; see, e.g., Kelly et al. (2006).

The contribution of this paper is the following. We follow the concept of Pita, Jain, Ordóñez, et al. (2009) to consider follower's response uncertainty due to limited observability regarding the upper-level decision. In contrast to the approach using anchoring biases, we pursue the same idea as in Besançon et al. (2019) and model bounded rationality using the toolbox of robust optimization. However, while these authors consider the effect of near-optimal decisions of the follower on the upper-level constraints, we focus on how limited observability regarding the upper-level decision affects the problem at the lower level.

The remainder of the paper is organized as follows. Section 2 gives a brief introduction of bilinear bilevel problems and the concept of limited observability. In Section 3, we present an illustrative example to demonstrate the importance of the proposed modeling aspect. In Section 4, the robust counterpart of the lower level is shown to be a bilinear problem as well so that we can establish an equivalent single-level reformulation by replacing the lower-level problem with its Karush–Kuhn–Tucker (KKT) conditions. We then return in Section 5 to the example to illustrate the effect of follower's decisions under bounded rationality on the problem studied in Section 3. Further, we address the relation between limited observability regarding the upper-level decision and bilevel problems with lower-level right-hand side uncertainty in Section 6. Finally, we conclude in Section 7.

2. PROBLEM STATEMENT

In this paper, we consider the bilinear bilevel problem

$$\min_{x,y} \quad c^\top x + d^\top y + x^\top R y \tag{1a}$$

$$\text{s.t.} \quad Ax + By \geq a, \tag{1b}$$

$$y \in \arg\min_{y'} \left\{ f^\top x + g^\top y' + x^\top Q y' : Cx + Dy' \geq b \right\} \tag{1c}$$

with $x, c, f \in \mathbb{R}^n$, $y, d, g \in \mathbb{R}^m$, $R, Q \in \mathbb{R}^{n \times m}$, $A \in \mathbb{R}^{k \times n}$, $B \in \mathbb{R}^{k \times m}$, $a \in \mathbb{R}^k$, $C \in \mathbb{R}^{\ell \times n}$, $D \in \mathbb{R}^{\ell \times m}$, and $b \in \mathbb{R}^\ell$. We refer to (1a)–(1b) as the upper level and to (1c) as the nominal lower level. Here, we consider the optimistic approach to bilevel optimization; see, e.g., Dempe (2002). This means that whenever the lower-level problem has multiple solutions y, the follower selects the most favorable one w.r.t. the leader's objective function.

In Problem (1), we make the strong assumption that the leader and the follower act perfectly rational. In real-world applications, however, this assumption rarely holds as both players face bounded rationality; see, e.g., Simon (1972). Here, we consider follower response uncertainty due to limited observability regarding the upper-level decision as proposed, e.g., in Pita, Portway, et al. (2008). This means that the follower cannot perfectly observe the actual leader's decision x. Nevertheless, the observed upper-level decision $\bar{x}$ provides an insight into the leader's scope of action. Given this knowledge, the follower's response is based on $\bar{x}$, which is assumed to belong to a given uncertainty set $\mathcal{U}(x)$. This leads to the robust bilevel problem

$$\begin{aligned} \min_{x,y} \quad & c^\top x + d^\top y + x^\top R y && \text{(2a)}\\ \text{s.t.} \quad & Ax + By \geq a, && \text{(2b)}\\ & y \in \arg\min_{y'} \left\{ g^\top y' + \max_{\bar{x} \in \mathcal{U}(x)} \left\{ f^\top \bar{x} + \bar{x}^\top Q y' \colon C\bar{x} + Dy' \geq b \ \forall \bar{x} \in \mathcal{U}(x) \right\} \right\} && \text{(2c)} \end{aligned}$$

with a robustified lower-level objective function as well as a robustified feasible set in the lower-level problem. Note that due to the robustification of the lower-level's objective function, the linear term in x cannot be avoided as it is usually done in (bi)linear bilevel optimization. Throughout this paper, the uncertainty set is parameterized in an affine way by

$$\mathcal{U}(x) = \{x + P\zeta \colon \zeta \in \mathcal{Z} \subseteq \mathbb{R}^q\} \tag{3}$$

with $P \in \mathbb{R}^{n \times q}$ and a polyhedral perturbation set

$$\mathcal{Z} = \{\zeta \in \mathbb{R}^q \colon H\zeta \geq h\}$$

for $H \in \mathbb{R}^{s \times q}$ and $h \in \mathbb{R}^s$. To ensure that the original leader's decision x is still part of the uncertainty set $\mathcal{U}(x)$ one can additional assume that $0 \in \mathcal{Z}$ holds.

3. Example: Bilevel Bimatrix Games

A bimatrix game is a non-cooperative two-player simultaneous-move game. This means that two competitive players, called Player 1 and Player 2, select their strategies at the same time. Each player can choose from a finite number of possible actions, which are called pure strategies. We denote $x \in \mathbb{R}^n$ as the leader's mixed strategy if

$$\sum_{i=1}^{n} x_i = 1 \quad \text{and} \quad x \geq 0$$

holds. In this case, x_i represents the probability that Player 1 plays strategy $i \in [n] := \{1, \ldots, n\}$. Analogously, we obtain the feasible set for the follower's mixed strategy $y \in \mathbb{R}^m$ via

$$\sum_{j=1}^{m} y_j = 1 \quad \text{and} \quad y \geq 0.$$

Both players attempt to minimize their objective functions $x^\top R y$ and $x^\top Q y$ with cost matrices $R, Q \in \mathbb{R}^{n \times m}$ for Player 1 and Player 2, respectively. The entries R_{ij} and Q_{ij} represent the associated costs if Player 1 chooses action $i \in [n]$ and if Player 2 selects strategy $j \in [m]$.

In what follows, we consider the sequential bimatrix game stated in Problem (4). Thus, we refer to Player 1 as the leader and to Player 2 as the follower:

$$\begin{aligned} \min_{x,y} \quad & x^\top R y && \text{(4a)}\\ \text{s.t.} \quad & \sum_{i=1}^{n} x_i = 1, \quad x \geq 0, && \text{(4b)}\\ & y \in \arg\min_{y'} \left\{ x^\top Q y' \colon \sum_{j=1}^{m} y'_j = 1, \ y' \geq 0 \right\}. && \text{(4c)} \end{aligned}$$

In Problem (4), the leader has to commit to a strategy first. Then, after observing the upper-level decision, the follower's strategy is determined. In particular, Problem (4) is a special case of Problem (1).

This type of problem has been subject to extensive research in many real-world applications in security domains such as defender-attacker scenarios and patrolling; see, e.g., Brown et al. (2006), Gatti (2008), Paruchuri et al. (2008), Pita, Portway, et al. (2008), Pita,

TABLE 1. Costs for the example in Section 3

	y_1	y_2
x_1	$(2,5)$	$(4,0)$
x_2	$(3,1)$	$(4,2)$

Jain, Marecki, et al. (2008), Kiekintveld et al. (2009), Jain et al. (2010), Shieh et al. (2012), and Yang et al. (2014) and the references therein.

Under the assumption that the follower faces limited observability regarding the upper-level decision, the reformulation of Problem (4) as a robust bilevel problem is given by

$$\begin{aligned} \min_{x,y} \quad & x^\top R y \\ \text{s.t.} \quad & \sum_{i=1}^n x_i = 1, \quad x \geq 0, \\ & y \in \arg\min_{y'} \left\{ \max_{\bar{x} \in \mathcal{U}(x)} \left\{ \bar{x}^\top Q y' : \sum_{j=1}^m y'_j = 1, \ y' \geq 0 \right\} \right\}. \end{aligned}$$

To illustrate follower's response uncertainty due to limited observability regarding the upper-level decision, we consider the example depicted in Table 1 with the respective costs for the leader and the follower. If the leader commits to a pure strategy and the follower can perfectly observe the upper-level decision, the bilevel solution is given by $x = (0,1)$ and $y = (1,0)$. If the leader commits to a mixed strategy of playing x_1 with probability $1/6$ and x_2 with probability $5/6$, the optimistic follower will select $y = (1,0)$ resulting in expected costs of $17/6$ for the leader. In Section 5, we will return to this example to illustrate the case of limited observability regarding the upper-level decision. Before, we have to consider how these robust bilevel problems can be reformulated as single-level optimization problems.

4. SINGLE-LEVEL REFORMULATION

As elaborated in Bertsimas et al. (2010), the lower-level problem (2c) can be replaced with the robust formulation

$$\begin{aligned} \min_{y,\tau} \quad & \tau && \text{(6a)} \\ \text{s.t.} \quad & g^\top y + \max_{\bar{x} \in \mathcal{U}(x)} \left\{ f^\top \bar{x} + \bar{x}^\top Q y \right\} \leq \tau, && \text{(6b)} \\ & C\bar{x} + Dy \geq b \quad \text{for all } \bar{x} \in \mathcal{U}(x), && \text{(6c)} \end{aligned}$$

in which the uncertainties only arise in the constraints of the problem. In terms of the affine parameterization of the uncertainty set, Constraint (6b) can be stated as

$$g^\top y + f^\top x + x^\top Q y + \max_{\zeta \in \mathcal{Z}} \left\{ (f + Qy)^\top P\zeta \right\} \leq \tau. \tag{7}$$

For each fixed $y \in \mathbb{R}^m$ in the follower's objective function, the inner optimization problem in (7) is equivalent to the solution of the linear problem

$$\begin{aligned} \max_{\zeta} \quad & (f + Qy)^\top P\zeta && \text{(8a)} \\ \text{s.t.} \quad & H\zeta \geq h. && \text{(8b)} \end{aligned}$$

Similarly, we can replace each constraint in (6c) with the corresponding worst-case scenario

$$C_{j\cdot}x + \min_{\zeta \in \mathcal{Z}} \left\{ (C_{j\cdot}P)\, \zeta \right\} + D_{j\cdot}y \geq b_j, \tag{9}$$

where $C_{j\cdot}$ denotes the jth row of C for $j \in [\ell]$. Thus, (9) requires the solution of

$$\min_{\zeta} \quad (C_{j\cdot}P)\,\zeta \tag{10a}$$

$$\text{s.t.} \quad H\zeta \geq h, \tag{10b}$$

for each constraint $j \in [\ell]$. We denote with $\sigma \in \mathbb{R}^s$ and $\lambda^j \in \mathbb{R}^s$, $j \in [\ell]$, the dual variables associated with Problem (8) and (10), respectively. The dual problem of (8) is given by

$$\min_{\sigma} \quad -h^\top \sigma \tag{11a}$$

$$\text{s.t.} \quad H^\top \sigma = -P^\top (f + Qy), \tag{11b}$$

$$\sigma \geq 0. \tag{11c}$$

Similarly, we obtain the dual formulation of (10) as

$$\max_{\lambda^j} \quad h^\top \lambda^j \tag{12a}$$

$$\text{s.t.} \quad H^\top \lambda^j = (C_{j\cdot}P)^\top, \tag{12b}$$

$$\lambda^j \geq 0. \tag{12c}$$

By strong duality, the objective values of the respective primal and dual problems coincide for primal-dual optimal pairs. Thus, to satisfy the uncertain constraints in (6), the following inequalities must hold for all dual variables σ and λ^j

$$f^\top x + g^\top y + x^\top Qy - h^\top \sigma \leq \tau, \tag{13a}$$

$$C_{j\cdot}x + D_{j\cdot}y + h^\top \lambda^j \geq b_j, \quad j \in [\ell], \tag{13b}$$

as well as (11b)–(11c) and (12b)–(12c). As a result, we obtain the robust counterpart

$$\min_{y,\tau,\sigma,\lambda} \quad \tau \tag{14a}$$

$$\text{s.t.} \quad f^\top x + g^\top y + x^\top Qy - h^\top \sigma \leq \tau, \tag{14b}$$

$$C_{j\cdot}x + D_{j\cdot}y + h^\top \lambda^j \geq b_j, \quad j \in [\ell], \tag{14c}$$

$$H^\top \sigma = -P^\top (f + Qy), \tag{14d}$$

$$H^\top \lambda^j = (C_{j\cdot}P)^\top, \quad j \in [\ell], \tag{14e}$$

$$\lambda^j \geq 0, \quad j \in [\ell], \tag{14f}$$

$$\sigma \geq 0, \tag{14g}$$

of the lower-level problem. In particular, (14) is a linear problem for each fixed leader's decision x. Thus, we can replace the lower level with its KKT conditions and obtain the single-level reformulation

$$
\begin{aligned}
\min_{x,y,z} \quad & c^\top x + d^\top y + x^\top R y && && \text{(15a)}\\
\text{s.t.} \quad & Ax + By \geq a, && && \text{(15b)}\\
& C_{j\cdot} x + D_{j\cdot} y + h^\top \lambda^j \geq b_j, && j \in [\ell], && \text{(15c)}\\
& H^\top \sigma = -P^\top (f + Qy), && && \text{(15d)}\\
& H^\top \lambda^j = \left(C_{j\cdot} P\right)^\top, && j \in [\ell], && \text{(15e)}\\
& g + Q^\top x - D^\top \alpha - Q^\top P \beta = 0, && && \text{(15f)}\\
& h + H\beta + \delta = 0, && && \text{(15g)}\\
& \alpha_j h + H\gamma^j + \varepsilon^j = 0, && j \in [\ell], && \text{(15h)}\\
& f^\top x + g^\top y + x^\top Q y - h^\top \sigma = \tau, && && \text{(15i)}\\
& \alpha_j \left(C_{j\cdot} x + D_{j\cdot} y + h^\top \lambda^j - b_j\right) = 0, && j \in [\ell], && \text{(15j)}\\
& \delta^\top \sigma = 0, && && \text{(15k)}\\
& \left(\varepsilon^j\right)^\top \lambda^j = 0, && j \in [\ell], && \text{(15l)}\\
& \lambda^j, \varepsilon^j \geq 0, && j \in [\ell], && \text{(15m)}\\
& \sigma, \alpha, \delta \geq 0, && && \text{(15n)}
\end{aligned}
$$

where z contains all primal variables used for modeling limited observability in the sense of robust optimization as well as all dual lower-level variables with $\sigma \in \mathbb{R}^s$, $\lambda^j \in \mathbb{R}^s$, $\tau \in \mathbb{R}$, $\alpha \in \mathbb{R}^\ell$, $\beta \in \mathbb{R}^q$, $\gamma^j \in \mathbb{R}^q$, $\delta \in \mathbb{R}^s$, and $\varepsilon^j \in \mathbb{R}^s$ for all $j \in [\ell]$. Note that τ models the robustified objective function value of the follower. Thus, we could—in principal—dispose of Constraint (15i) since τ is an auxiliary variable in Problem (14) that could be eliminated in (14) as well.

5. The Illustrative Example Revisited

We now reconsider the example in Section 3 under the assumption that the follower faces limited observability regarding the upper-level decision. We focus on the case that the perceived leader's decision $\bar{x}$ is only known to lie within a box, i.e., the uncertainty set is given by

$$\mathcal{U}(x) = \{\bar{x} = x + P\zeta : \zeta \in \mathcal{Z}\} \cap \{\bar{x} : \bar{x}_1 + \bar{x}_2 = 1\}$$

with

$$P = \begin{bmatrix} p_1 & 0 \\ 0 & p_2 \end{bmatrix} \in \mathbb{R}^{2\times 2}$$

and the perturbation set

$$\mathcal{Z} = \left\{\zeta \in \mathbb{R}^2 : -1 \leq \zeta_i \leq 1,\ i = 1, 2\right\}.$$

In particular, this is a special case of a polyhedral uncertainty set. Before we discuss the results let us briefly comment on the interpretation of the modeling in this setting—especially w.r.t. the definition of the uncertainty set $\mathcal{U}(x)$. The first set in the definition of $\mathcal{U}(x)$ corresponds to the uncertainty modeling of the leader's decision as stated around (3). In this example, the leader commits to a mixed strategy, which is a probability distribution over a finite number of possible actions. When facing limited observability, the follower anticipates the resulting simplex structure of the leader's mixed strategy, which is captured in the second set of the intersection.

The specific example is implemented in Python 3.8 and Gurobi 9.1.0 is used to solve the single-level reformulation (15). To handle the KKT complementarity constraints we exploit special ordered sets of type 1 (SOS1) to avoid big-M reformulations that can be

TABLE 2. Leader's strategies under limited observability

p_2 p_1	0	0.25	0.5	0.75	1
0	$(0.167, 0.833)$	$(0.167, 0.833)$	$(0.167, 0.833)$	$(0.167, 0.833)$	$(0.167, 0.833)$
0.25	$(0.167, 0.833)$	$(0.417, 0.583)$	$(0.417, 0.583)$	$(0.417, 0.583)$	$(0.417, 0.583)$
0.5	$(0.167, 0.833)$	$(0.417, 0.583)$	$(0.667, 0.333)$	$(0.667, 0.333)$	$(0.667, 0.333)$
0.75	$(0.167, 0.833)$	$(0.417, 0.583)$	$(0.667, 0.333)$	$(0.917, 0.083)$	$(0.917, 0.083)$
1	$(0.167, 0.833)$	$(0.417, 0.583)$	$(0.667, 0.333)$	$(0.917, 0.083)$	$(1, 0)$

troublesome; see, e.g., Pineda and Morales (2019) and Kleinert et al. (2020). Hence, we restate Constraints (15j)–(15l) as

$$\begin{aligned}
&r_j = C_{j\cdot}x + D_{j\cdot}y + h^\top \lambda^j - b_j, && j \in [\ell],\\
&\mathrm{SOS1}\left(\alpha_j, r_j\right), && j \in [\ell],\\
&\mathrm{SOS1}\left(\delta_k, \sigma_k\right), && k \in [s],\\
&\mathrm{SOS1}\left(\varepsilon_k^j, \lambda_k^j\right), && k \in [s],\ j \in [\ell].
\end{aligned}$$

Tables 2–3 summarize the nominal and robust leader's and follower's strategies for different uncertainty set parameterizations that model limited observability. The respective objective function values are given in Table 4, where the first element of the tuple denotes the value of the leader's objective and the second one gives the objective function value for the follower.

First, it can be seen that limited observability regarding the upper-level decision can lead to deviations of the follower's strategy from the nominal decision. In this example, the follower moves from playing the pure strategy $y = (1, 0)$ to committing to the mixed strategy of playing y_1 with probability $1/6$ and y_2 with probability $5/6$ regardless of the extent to which the follower is limited in the observability of the leader's decision. In particular, this means that the follower's strategy shifts entirely when facing limited observability. Due to the leader's anticipation of the follower's response uncertainty, the upper-level decision can also change significantly compared to the nominal strategy, depending on the extent of the uncertainties. Under the assumption that the follower can perfectly observe the upper-level decision, the leader tends to play x_2. However, the greater the uncertainty regarding the observability of the upper-level decision, the more the leader tends to play x_1. In particular, the leader will commit to the pure strategy of playing x_1 for the uncertainty set parameterization with $p_1 = p_2 = 1$. Therefore, not only the follower's strategy but also the leader's strategy can shift entirely if limited observability regarding the upper-level decision is taken into account. Moreover, it can be seen that the strategies of both players are symmetric w.r.t. the parameterization of the uncertainty set, which is due to the simplex structure of the leader's feasible set. Further, it is noticeable that the leader's anticipation of follower's response uncertainty due to limited observability regarding the upper-level decision always leads to significantly increased costs for the leader. In this example, the increase in the leader's costs is up to approximately 33 % of the nominal value.

To sum up, this illustrative example shows that limited observability significantly impacts the solution of the underlying bilevel problem. Thus, this modeling aspect should not be ignored if the application problem at hand contains a lower-level player that cannot perfectly observe the leader's decision.

TABLE 3. Follower's strategies under limited observability

p_2 p_1	0	0.25	0.5	0.75	1
0	$(1, 0)$	$(1, 0)$	$(1, 0)$	$(1, 0)$	$(1, 0)$
0.25	$(1, 0)$	$(0.167, 0.833)$	$(0.167, 0.833)$	$(0.167, 0.833)$	$(0.167, 0.833)$
0.5	$(1, 0)$	$(0.167, 0.833)$	$(0.167, 0.833)$	$(0.167, 0.833)$	$(0.167, 0.833)$
0.75	$(1, 0)$	$(0.167, 0.833)$	$(0.167, 0.833)$	$(0.167, 0.833)$	$(0.167, 0.833)$
1	$(1, 0)$	$(0.167, 0.833)$	$(0.167, 0.833)$	$(0.167, 0.833)$	$(0.167, 0.833)$

TABLE 4. Leader's and follower's objective function values under limited observability

p_2 p_1	0	0.25	0.5	0.75	1
0	$(2.833, 1.667)$	$(2.833, 1.667)$	$(2.833, 1.667)$	$(2.833, 1.667)$	$(2.833, 1.667)$
0.25	$(2.833, 1.667)$	$(3.764, 1.417)$	$(3.764, 1.417)$	$(3.764, 1.417)$	$(3.764, 1.417)$
0.5	$(2.833, 1.667)$	$(3.764, 1.417)$	$(3.722, 1.667)$	$(3.722, 1.667)$	$(3.722, 1.667)$
0.75	$(2.833, 1.667)$	$(3.764, 1.417)$	$(3.722, 1.667)$	$(3.681, 0.917)$	$(3.681, 0.917)$
1	$(2.833, 1.667)$	$(3.764, 1.417)$	$(3.722, 1.667)$	$(3.681, 0.917)$	$(3.667, 0.833)$

6. RELATION TO BILEVEL PROBLEMS WITH UNCERTAIN LOWER-LEVEL DATA

Compared to the problem under perfect rationality, the consideration of limited observability regarding the upper-level decision yields significantly larger optimization problems in terms of the number of variables as well as the number of constraints. Therefore, it seems reasonable to strive for a more compact formulation to model limited observability. For this purpose, we now address the relation between Problem (15) and bilevel problems with uncertain lower-level data. More specifically, we consider problems with uncertainties in the right-hand sides of the lower-level's constraints, i.e., we consider

$$\begin{aligned} \min_{x,y} \quad & c^\top x + d^\top y + x^\top R y \\ \text{s.t.} \quad & Ax + By \geq a, \\ & y \in \operatorname*{arg\,min}_{y'} \left\{ f^\top x + g^\top y' + x^\top Q y' \colon Cx + Dy' \geq \bar{b} \text{ for all } \bar{b} \in \mathcal{U}(b) \right\}. \end{aligned}$$

Here, b is the nominal right-hand side and $\mathcal{U}(b) = \mathcal{U}_1(b_1) \times \cdots \times \mathcal{U}_\ell(b_\ell)$ holds with

$$\mathcal{U}_j(b_j) = \left\{ b_j + (\tilde{p}^j)^\top \zeta^j \colon \zeta^j \in \mathcal{Z}^j \right\}$$

and

$$\mathcal{Z}^j = \left\{ \zeta^j \colon \left(\tilde{H}^j\right)^\top \zeta^j \geq \tilde{h}^j \right\},$$

for all $j \in [\ell]$. For the ease of presentation, we omit the dimensions of all vectors and matrices in this section. Similar to the derivation in Section 4, we obtain the single-level reformulation

$$\begin{aligned}
\min_{x,y,\tilde{z}} \quad & c^\top x + d^\top y + x^\top R y && && \text{(16a)}\\
\text{s.t.} \quad & Ax + By \geq a, && && \text{(16b)}\\
& C_{j\cdot} x + D_{j\cdot} y + \left(\tilde{h}^j\right)^\top \tilde{\lambda}^j \geq b_j, && j \in [\ell], && \text{(16c)}\\
& \left(\tilde{H}^j\right)^\top \tilde{\lambda}^j = -\tilde{p}^j, && j \in [\ell], && \text{(16d)}\\
& g + Q^\top x - D^\top \tilde{\alpha} = 0, && && \text{(16e)}\\
& \tilde{\alpha}_j \tilde{h}^j + \tilde{H}^j \tilde{\beta}^j + \tilde{\gamma}^j = 0, && j \in [\ell], && \text{(16f)}\\
& \tilde{\alpha}_j \left(C_{j\cdot} x + D_{j\cdot} y + \left(\tilde{h}^j\right)^\top \tilde{\lambda}^j - b_j\right) = 0, && j \in [\ell], && \text{(16g)}\\
& \left(\tilde{\gamma}^j\right)^\top \tilde{\lambda}^j = 0, && j \in [\ell], && \text{(16h)}\\
& \tilde{\lambda}^j, \tilde{\gamma}^j \geq 0, && j \in [\ell], && \text{(16i)}\\
& \tilde{\alpha} \geq 0, && && \text{(16j)}
\end{aligned}$$

where $\tilde{z}$ contains all primal variables used for the robustification of the uncertain right-hand sides as well as all resulting dual lower-level variables $\tilde{\lambda}^j$, $\tilde{\alpha}$, $\tilde{\beta}^j$, and $\tilde{\gamma}^j$. Before we show how Problem (15) relates to (16), we first consider an illustrative example taken and adapted from Besançon et al. (2019).

Example 1. *We consider the linear bilevel problem defined by* $0 \leq x, y \in \mathbb{R}$ *and the data*

$$A = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad B = \begin{bmatrix} -4 \\ -2 \end{bmatrix}, \quad a = \begin{pmatrix} -11 \\ -13 \end{pmatrix}, \quad c = 1, \quad d = -10, \quad R = 0,$$
$$C = \begin{bmatrix} 2 \\ -5 \end{bmatrix}, \quad D = \begin{bmatrix} 1 \\ 4 \end{bmatrix}, \quad b = \begin{pmatrix} 5 \\ -30 \end{pmatrix}, \quad f = 0, \quad g = 1, \quad Q = 0.$$

Further, we assume that the perceived decision of the leader as well as the uncertain lower-level data are known to lie within box constraints, i.e.,

$$\mathcal{Z} = \mathcal{Z}^j = \{\zeta \in \mathbb{R} \colon -1 \leq \zeta \leq 1\}, \quad j = 1, 2.$$

To model limited observability regarding the upper-level decision, we consider the uncertainty set parameterizations with $P \in \{0.5, 1\}$*. The example is illustrated in Figure 1. The upper- and lower-level constraints are represented with dashed and solid lines, respectively. The optimal nominal solution and the two optimal robust solutions are illustrated with thick dots. It can be seen that limited observability regarding the leader's decision leads to a parallel shift of the follower's constraints, i.e., to lower-level right-hand side uncertainty. In particular, in this example, the corresponding lower-level right-hand side uncertainty is given by* $\tilde{p}^1 = 2P$ *and* $\tilde{p}^2 = 5P$*.*

Based on the observation in Example 1, the question naturally arises on whether limited observability regarding the leader's decision can, in general, be modeled as a problem with uncertain lower-level right-hand side data. If this would be possible, it would be particularly favorable for scaling reasons, since Problem (15) can get very large. Thus, we next formally address how Problem (15) relates to Problem (16).

Theorem 2. *Let* (x, y, z) *be a solution of Problem (15) with parameters* P*,* h*, and* H *modeling the uncertainty set. Furthermore, let* $(x, y, \tilde{z})$ *be a solution of Problem (16) with the uncertainty sets modeled with the parameters* $\tilde{p}^j$*,* $\tilde{h}^j$*, and* $\tilde{H}^j$*,* $j \in [\ell]$*. Third, suppose that the lower-level's objective function does not contain bilinear terms, i.e.,* $Q = 0$*, and that*

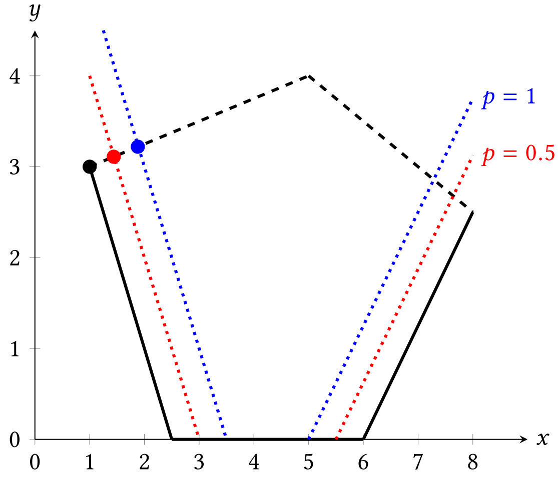

FIGURE 1. Relation between limited observability and uncertain lower-level data

$\operatorname{rank}(D^\top) = \operatorname{rank}([D^\top, g])$ *holds. Then, the uncertainty modeling parameters satisfy*

$$\left(\tilde h^j\right)^\top \tilde\lambda^j = h^\top \lambda^j, \qquad j \in [\ell], \tag{17a}$$

$$\left(\tilde\beta^j\right)^\top \tilde p^j = -\left(\gamma^j\right)^\top \left(C_{j\cdot}P\right)^\top, \qquad j \in [\ell], \tag{17b}$$

$$\left(\tilde\beta^j\right)^\top \left(\tilde H^j\right)^\top \tilde\lambda^j = -\left(\gamma^j\right)^\top H^\top \lambda^j, \qquad j \in [\ell], \tag{17c}$$

$$\left(\tilde\beta^j\right)^\top \left(\tilde H^j\right)^\top \tilde\lambda^j = \left(\gamma^j\right)^\top \left(H^\top \sigma - P^\top \left(C_{j\cdot}^\top + f\right)\right), \qquad j \in [\ell]. \tag{17d}$$

Proof. Under the imposed assumptions, we obtain $\tilde\alpha = \alpha$ by dual feasibility. Thus, (17a) follows immediately from KKT complementarity. For all $j \in [\ell]$, the multiplication of Constraint (16f) with $\tilde\lambda^j$ yields

$$\alpha_j \left(\tilde h^j\right)^\top \tilde\lambda^j + \left(\tilde\beta^j\right)^\top \left(\tilde H^j\right)^\top \tilde\lambda^j + \left(\tilde\gamma^j\right)^\top \tilde\lambda^j = 0.$$

Due to (16d) and (16h), Constraint (16f) is thus equivalent to

$$\alpha_j \left(\tilde h^j\right)^\top \tilde\lambda^j = -\left(\tilde\beta^j\right)^\top \left(\tilde H^j\right)^\top \tilde\lambda^j = \left(\tilde\beta^j\right)^\top \tilde p^j, \quad j \in [\ell]. \tag{18}$$

Similarly, multiplying (15h) with λ^j and using (15e) as well as (15l) yields

$$\alpha_j h^\top \lambda^j = -\left(\gamma^j\right)^\top H^\top \lambda^j = -\left(\gamma^j\right)^\top \left(C_{j\cdot}P\right)^\top, \quad j \in [\ell].$$

Then, plugging in (17a) and the results in (18) yields (17b)–(17c). Finally, Equation (17d) follows immediately from (17c) as well as (15d)–(15e). □

To sum up, there exists a connection between Problem (15) and a suitably chosen bilevel problem with lower-level right-hand side uncertainty. However, even though we impose rather strong assumptions (such as the rank condition), the established relation between the different uncertainty set parameterizations in (17) requires the knowledge of the lower-level primal and dual variables in advance, which means that the established result is only an ex-post relation. Thus, Problem (16) cannot be exploited to obtain a more compact formulation for our modeling of limited observability.

7. Conclusion

In this paper, we consider bilinear bilevel problems under follower's response uncertainty due to limited observability regarding the upper-level strategy. To this end, we exploit robust optimization to model decision making in the lower level under bounded rationality. An equivalent single-level reformulation is established by replacing the robust counterpart of the lower-level problem with its KKT conditions. Compared to the problem under perfect rationality, the presented modeling yields much larger optimization problems in terms of the number of variables and constraints. However, the problem remains in the same problem class as the problem without taking limited observability into account. We present an illustrative example to emphasize the importance of the proposed modeling aspect. Further, we establish an ex-post relation between the modeling of limited observability and robust bilevel problems with lower-level right-hand side uncertainty.

In this paper, polyhedral uncertainty sets have been considered. The consideration of other uncertainty set geometries might be a reasonable aspect of future work.

Acknowledgements

The second author thanks the DFG for their support within projects A05 and B08 in CRC TRR 154.

References

Ambrosius, M., V. Grimm, T. Kleinert, F. Liers, M. Schmidt, and G. Zöttl (2020). "Endogenous Price Zones and Investment Incentives in Electricity Markets: An Application of Multilevel Optimization with Graph Partitioning." In: *Energy Economics* 92. DOI: 10.1016/j.eneco.2020.104879.

Ben-Ayed, O., C. Blair, D. Boyce, and L. LeBlanc (1992). "Construction of a real-world bilevel linear programming model of the highway network design problem." In: *Annals of Operations Research* 34, pp. 219–254. DOI: 10.1007/BF02098181.

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

Bertsimas, D., D. Brown, and C. Caramanis (2010). "Theory and Applications of Robust Optimization." In: *SIAM Review* 53. DOI: 10.1137/080734510.

Besançon, M., M. F. Anjos, and L. Brotcorne (2019). *Near-optimal Robust Bilevel Optimization*. URL: https://arxiv.org/pdf/1908.04040.pdf.

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.

Burtscheidt, J. and M. Claus (2020). "Bilevel Linear Optimization Under Uncertainty." In: *Bilevel Optimization: Advances and Next Challenges*. Ed. by S. Dempe and A. Zemkoho. Springer International Publishing, pp. 485–511. DOI: 10.1007/978-3-030-52119-6_17.

Burtscheidt, J., M. Claus, and S. Dempe (2020). "Risk-Averse Models in Bilevel Stochastic Linear Programming." In: *SIAM Journal on Optimization* 30.1, pp. 377–406. DOI: 10.1137/19M1242240.

Chariri, A. (2017). "Cognitive Limitations and Decision Making." In: *JURNAL BISNIS STRATEGI* 3.2. URL: https://ejournal.undip.ac.id/index.php/jbs/article/view/14148.

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

Dempe, S., S. Ivanov, and A. Naumov (2017). "Reduction of the bilevel stochastic optimization problem with quantile objective function to a mixed-integer problem." In: *Applied Stochastic Models in Business and Industry* 33.5, pp. 544–554. DOI: 10.1002/asmb.2254.

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

Dequech, D. (2001). “Bounded Rationality, Institutions, and Uncertainty.” In: *Journal of Economic Issues* 35, pp. 911–929. DOI: 10.1080/00213624.2001.11506420.

Fischetti, M., I. Ljubic, 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.

Gatti, N. (2008). “Game Theoretical Insights in Strategic Patrolling: Model and Algorithm in Normal-Form.” In: *ECAI 2008—Proceedings of the 18th European Conference on Artificial Intelligence*, pp. 403–407. DOI: 10.3233/978-1-58603-891-5-403.

Grimm, V., L. Schewe, M. Schmidt, and G. Zöttl (2019). “A Multilevel Model of the European Entry-Exit Gas Market.” In: *Mathematical Methods of Operations Research* 89.2, pp. 223–255. DOI: 10.1007/s00186-018-0647-z.

Haghighat, H. (2014). “Strategic offering under uncertainty in power markets.” In: *International Journal of Electrical Power & Energy Systems* 63, pp. 1070–1077. DOI: 10.1016/j.ijepes.2014.05.049.

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.

Ivanov, S. V. (2018). “A Bilevel Stochastic Programming Problem with Random Parameters in the Follower’s Objective Function.” In: *Journal of Applied and Industrial Mathematics* 12.4, pp. 658–667. DOI: 10.1134/S1990478918040063.

Jain, M., J. Tsai, J. Pita, C. Kiekintveld, S. Rathi, M. Tambe, and F. Ordóñez (2010). “Software Assistants for Randomized Patrol Planning for the LAX Airport Police and the Federal Air Marshal Service.” In: *INFORMS Journal on Applied Analytics* 40.4, pp. 267–290. DOI: 10.1287/inte.1100.0505.

Kelly, E. S., C. R. Fox, and Y. S. Rottenstreich (2006). “Between ignorance and truth: Partition dependence and learning in judgment under uncertainty.” In: *Journal of experimental psychology. Learning, memory, and cognition* 32.6, pp. 1385–1402. DOI: 10.1037/0278-7393.32.6.1385.

Kiekintveld, C., M. Jain, J. Tsai, J. Pita, F. Ordóñez, and M. Tambe (2009). “Computing optimal randomized resource allocations for massive security games.” In: *AAMAS 2009—Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems*. Vol. 1, pp. 689–696. URL: https://dl.acm.org/doi/pdf/10.5555/1558013.1558108.

Kleinert, T., M. Labbé, F. Plein, and M. Schmidt (2020). “There’s No Free Lunch: On the Hardness of Choosing a Correct Big-M in Bilevel Optimization.” In: *Operations Research* 68.6, pp. 1716–1721. DOI: 10.1287/opre.2019.1944.

Migdalas, A. (1995). “Bilevel programming in traffic planning: Models, methods and challenge.” In: *Journal of Global Optimization* 7, pp. 381–405. DOI: 10.1007/BF01099649.

Paruchuri, P., J. Pearce, J. Marecki, M. Tambe, F. Ordóñez, and S. Kraus (2008). “Playing games for security: An efficient exact algorithm for solving Bayesian Stackelberg games.” In: *AAMAS 2008—Proceedings of the 7th International Conference on Autonomous Agents and Multiagent Systems*. Vol. 2, pp. 895–902. URL: https://dl.acm.org/doi/pdf/10.5555/1402298.1402348.

Pineda, S. and J. M. Morales (2019). “Solving Linear Bilevel Problems Using Big-Ms: Not All That Glitters Is Gold.” In: *IEEE Transactions on Power Systems*. DOI: 10.1109/TPWRS.2019.2892607.

Pita, J., M. Jain, J. Marecki, F. Ordóñez, C. Portway, M. Tambe, C. Western, P. Paruchuri, and S. Kraus (2008). "Deployed ARMOR protection: The application of a game theoretic model for security at the Los Angeles International Airport." In: *AAMAS 2008—Proceedings of the 7th International Conference on Autonomous Agents and Multiagent Systems*, pp. 125–132. URL: `https://dl.acm.org/doi/pdf/10.5555/1402795.1402819`.

Pita, J., M. Jain, F. Ordóñez, M. Tambe, S. Kraus, and R. Magori-Cohen (2009). "Effective solutions for real-world Stackelberg games: When agents must deal with human uncertainties." In: *AAMAS 2009—Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems*. Vol. 1, pp. 369–376. DOI: `10.1017/CBO9780511973031.010`.

Pita, J., C. Portway, M. Tambe, P. Paruchuri, and S. Kraus (2008). "Robust Solutions in Stackelberg Games: Addressing Boundedly Rational Human Preference Models." In: *AAAI 2008—4th Multidisciplinary Workshop on Advances in Preference Handling*. URL: `https://www.aaai.org/Papers/Workshops/2008/WS-08-09/WS08-09-010.pdf`.

Rainey, H. (2001). "A Reflection on Herbert Simon: A Satisficing Search for Significance." In: *Administration & Society* 33, pp. 491–507. DOI: `10.1177/00953990122019857`.

Rubinstein, A. (1998). *Modeling Bounded Rationality*. The MIT Press. URL: `https://mitpress.mit.edu/books/modeling-bounded-rationality`.

Shieh, E., B. An, R. Yang, M. Tambe, C. Baldwin, J. DiRenzo, B. Maule, and G. Meyer (2012). "PROTECT: An application of computational game theory for the security of the ports of the United States." In: *AAAI 2012—Proceedings of the 26th National Conference on Artificial Intelligence*, pp. 2173–2179. URL: `http://www.intsci.ac.cn/en/boan/papers/AAAI122.pdf`.

Simon, H. (1972). "Theories of Bounded Rationality." In: *McGuire, C.B. and Radner, R., Eds., Decision and Organization*. North Holland Publishing Company, pp. 161–176. URL: `http://innovbfa.viabloga.com/files/Herbert_Simon___theories_of_bounded_rationality___1972.pdf`.

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`.

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

– (1954). *The Theory of the Market Economy*. Oxford University Press.

Wiesemann, W., A. Tsoukalas, P. Kleniati, and B. Rustem (2013). "Pessimistic Bilevel Optimization." In: *SIAM Journal on Optimization* 23, pp. 353–380. DOI: `10.1137/120864015`.

Xinmin, H. and D. Ralph (2007). "Using EPECs to Model Bilevel Games in Restructured Electricity Markets with Locational Prices." In: *Operations Research* 55.5, pp. 809–827. DOI: `10.1287/opre.1070.0431`.

Yang, R., B. Ford, M. Tambe, and A. Lemieux (2014). "Adaptive Resource Allocation for Wildlife Protection against Illegal Poachers." In: *AAMAS 2014—Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems*, pp. 453–460. URL: `https://teamcore.seas.harvard.edu/files/teamcore/files/2014_7_teamcore_aamas2014_yang.pdf`.

Yanıkoğlu, İ. and D. Kuhn (2018). "Decision Rule Bounds for Two-Stage Stochastic Bilevel Programs." In: *SIAM Journal on Optimization* 28.1, pp. 198–222. DOI: `10.1137/16M1098486`.

(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`