ISE

Industrial and
Systems Engineering

Mixed Integer Bilevel Optimization with k-optimal Follower: A Hierarchy of Bounds

Xueyu Shi and Oleg Prokopyev

Department of Industrial and System Engineering, University of Pittsburgh, Pittsburgh, PA

Ted K. Ralphs

Department of Industrial and System Engineering, Lehigh University, Bethlehem, PA

COR@L Technical Report 20T-012-R2

LEHIGH UNIVERSITY

COR@L
COMPUTATIONAL OPTIMIZATION RESEARCH AT LEHIGH

Mixed Integer Bilevel Optimization with k-optimal Follower: A Hierarchy of Bounds

Xueyu Shi*[1], Oleg Prokopyev†[1], and Ted K. Ralphs‡[2]

[1]Department of Industrial and System Engineering, University of Pittsburgh, Pittsburgh, PA

[2]Department of Industrial and System Engineering, Lehigh University, Bethlehem, PA

Original Publication: June 29, 2020

Last Revised: August 11, 2022

Abstract

We consider mixed integer bilevel linear optimization problems in which the decision variables of the lower-level (follower's) problem are all binary. We propose a general modeling and solution framework motivated by the practical reality that in a Stackelberg game, the follower does not always solve their optimization problem to optimality. They may instead implement a locally optimal solution with respect to a given upper-level decision. Such scenarios may occur when the follower's computational capabilities are limited, or when the follower is not completely rational. Our framework relaxes the typical assumption of perfect rationality that underlies the standard modeling framework by defining a hierarchy of increasingly stringent assumptions about the behavior of the follower. Namely, at level k of this hierarchy, it is assumed that the follower produces a k-optimal solution. Associated with this hierarchy is a hierarchy of upper and lower bounds that are in fact valid for the classical case in which complete rationality of the follower is assumed. Two mixed integer linear programming (MILP) formulations are derived for the resulting optimization problems. Extensive computational results are provided to demonstrate the effectiveness of the proposed MILP formulations and the quality of the bounds produced. The latter are shown to dominate the standard approach based on a single-level relaxation at a reasonable computational cost. Finally, we also explore a class of bilevel problems for which 2-optimal lower-level solutions imply global optimality, and hence we can solve these bilevel problems exactly using the developed MILP formulations.

1 Introduction

In bilevel optimization problems [13, 15], two independent decision-makers with their own distinct objective functions are involved in a hierarchical decision-making process. First, the upper-level decision-maker (referred to as the leader) makes the decision. Then, the lower-level decision-maker (referred to as the follower) determines the response in terms of their own optimization model, whose feasible region and objective function are parameterized on the leader's decision. Importantly, the follower's response also affects the leader's objective function. This is the reason the

*xus6@pitt.edu

†droleg@pitt.edu

‡ted@lehigh.edu

leader must take the follower's possible reactions into account when maximizing/minimizing their own benefits/costs.

1.1 Formal Setting

In this paper, we focus on a broad class of mixed integer bilevel linear optimization problems (MIBLP) in which the follower's decision variables are all binary. In particular, the considered class of problems is formally stated as:

$$\begin{aligned} \eta^* = \max_{x,y}\ & \alpha^1 x + \alpha^2 y \\ \text{s.t.}\ \ & x \in \mathcal{X}, \\ & y \in \arg\max_{\bar{y}} \{\beta \bar{y} :\ Ax + G\bar{y} \le d, \bar{y} \in \{0,1\}^n\}, \end{aligned} \tag{BP}$$

where $\mathcal{X} = \mathcal{P} \cap (\mathbb{Q}_+^{p_1} \times \mathbb{Z}_+^{p_2})$ for a given polyhedron $\mathcal{P}$; $A \in \mathbb{Z}^{m\times(p_1+p_2)}, G \in \mathbb{Z}^{m\times n}$, $d \in \mathbb{Z}^m$; and $\alpha^1 \in \mathbb{Q}^{(p_1+p_2)}$, $\alpha^2 \in \mathbb{Q}^n$, and $\beta \in \mathbb{Q}_+^n$ are given row vectors. Let

$$\mathcal{S} = \{(x, y) \in \mathcal{X} \times \{0,1\}^n :\ Ax + Gy \le d\}$$

be the feasible region of the relaxation obtained by dropping the optimality conditions on the follower's decision. We refer to x and y as the upper-level (leader's) variables and the lower-level (follower's) variables, respectively. Given $x \in \mathcal{X}$, we refer to

$$\mathcal{S}(x) = \{y \in \{0,1\}^n :\ (x, y) \in \mathcal{S}\} \quad \text{and} \quad \mathcal{R}(x) = \{y \in \mathcal{S}(x) :\ \beta y \ge \beta\bar{y} \quad \forall \bar{y} \in \mathcal{S}(x)\},$$

as the follower's feasible region and the follower's (rational) reaction set, respectively. The feasible region $\mathcal{S}(x)$ represents the options available to the follower, given a fixed solution x chosen by the leader, while the reaction set $\mathcal{R}(x)$ is the subset of the feasible region containing the feasible solutions that maximize the follower's objection function.

Note that (BP) specifies the optimistic case of the bilevel problem (BP). That is, there is an implicit assumption that the follower selects the most favorable solution for the leader from their reaction set $\mathcal{R}(x)$, i.e., if (x, y) is optimal for (BP), then we must have $y \in \mathcal{R}(x)$ and $\alpha^2 y \ge \alpha^2 \bar{y}$ for all $\bar{y} \in \mathcal{R}(x)$.

If we relax the optimality requirement on the follower's solution, the resulting *single-level relaxation* of (BP) is the mixed integer linear optimization problem (MILP)

$$\eta^{\text{SLR}} = \max_{(x,y)\in\mathcal{S}} \{\alpha^1 x + \alpha^2 y\}, \tag{SLR}$$

with feasible region $\mathcal{S}$. Problem (SLR) is also known as the *high-point problem* in the related literature; see, e.g., [43].

Observe that a feasible solution for (BP) can be obtained by fixing the leader's decision to an optimal solution obtained from (SLR). Specifically, assume (x^0, y^0) is optimal for (SLR). Then if $\hat{y}^0 \in \mathcal{R}(x^0)$ is an optimal (optimistic) follower's decision for the lower-level problem corresponding to x^0, then clearly, $(x^0, \hat{y}^0)$ is a feasible solution for (BP). Denote by $\hat{\eta}^{\text{SLR}} := \alpha^1 x_0 + \alpha^2 \hat{y}_0$ the leader's objective function value corresponding to $(x_0, \hat{y}_0)$. Thus, we have that:

$$\hat{\eta}^{\text{SLR}} \le \eta^* \le \eta^{\text{SLR}}.$$

The majority of branch-and-bound and branch-and-cut approaches for MIBLPs in the literature, see, e.g., [38, 43, 52, 55], solve either (SLR) or the linear programming relaxation of (SLR) to obtain the initial lower and upper bounds. We refer to the bounds obtained by this single-level relaxation-based approach as the SLR-based bounds.

Throughout the paper, we make the following assumptions, which are standard in the bilevel optimization literature:

Assumption 1. *All entries in A, G and d are integers; $\beta \geq 0$.*

Assumption 2. *$\mathcal{P}$ is compact.*

As for Assumption 1, the components of A, G, and d can always be scaled to be integral as long as they are rational; this assumption is often exploited in the literature on MILPs, see, e.g., [18]. As for the second part of Assumption 1, this is without loss of generality, since whenever $\beta_j < 0$ for some $j \in [n] := \{1, 2, \ldots, n\}$, we can simply replace $\bar{y}_j$ by $1 - \bar{y}_j$, and then reduce the problem into an equivalent one with this assumption satisfied. As for Assumptions 2, it is common in the bilevel optimization literature to assume the compactness of the leader's feasible region, see, e.g., [17, 49, 52] and the references therein.

We also make a technical assumption involving the set of upper-level variables that we refer to as the *linking variables*, those with at least one non-zero coefficient in the lower-level constraint matrix. The linking variables are formally defined as those with indices in the set

$$L = \left\{ i \in \{1, \ldots, p_1 + p_2\} : A^i \neq 0 \right\},$$

where A^i denotes the i^{th} column of A. We make the following assumption to ensure that optimal solutions of (BP) exist [54]:

Assumption 3. *All linking variables are integer variables.*

Additional notation. For any positive integer n, we use $[n]$ to denote the set $\{1, \ldots, n\}$. We denote by $e^j \in \mathbb{R}^n$ the j^{th} unit vector; by $\mathbf{0}$ the vector with all components equal to 0 and $\mathbf{1}$ the vector with all components equal to 1; and finally by M^i and m^i the i^{th} column and row of matrix M, respectively. All vectors are column vectors by default, with the exception of vectors used exclusively as objective function vectors (and the rows of matrices), which are taken to be row vectors for notational simplicity.

1.2 Related Work

Due to a broad range of important applications including interdiction [7, 16, 27], price setting [9, 33], and network design [6, 20, 58], among others, bilevel optimization problems have been extensively studied in the past two decades. Most of the research efforts have been focused on bilevel problems, where the lower-level problem is a linear optimization problem. The latter assumption allows for application of the necessary and sufficient optimality conditions (strong duality and complementary slackness) which, in turn, can be used to reduce the original bilevel problem into a single-level MILP [4, 59]. However, in recent years there has been increased interest in the exact solution methods for general MIBLPs, where the lower-level problem may involve integer decision variables, see, e.g., [17, 38, 55, 61]. The presence of integer variables at the lower level require application of somewhat more sophisticated methods. In particular, one of the primary examples is MibS [16, 52], an open-source solver, that exploits advanced cutting-plane-based approaches [51] within a branch-and-bound framework.

In general, solving MIBLPs is quite challenging. The initial bounds used within branch-and-bound and branch-and-cut frameworks are of critical importance for the overall performance of this type of exact methods. Unfortunately, for general classes of bilevel problems there are no theoretical guarantees on the quality of the SLR-based bounds. Moreover, the computational experiments available in the literature also indicate that (SLR) typically yields relatively poor bounds; see, e.g., [11]. Consequently, exact solution of large-scale bilevel optimization problems remains an intractable task, in particular, if one compares available bilevel solvers to the state-of-the-art commercial solvers for MILPs, e.g., CPLEX [26]. The latter is capable of handling millions of variables and constraints for many broad classes of MILPs. On the other hand, the most recent version of MibS can typically solve medium-sized problems with only up to several hundred integer decision variables at the lower level.

There are only a few studies, mostly focused on special classes of bilevel problem, that use bounding methods that are not SLR-based. In particular, the studies in [11, 53] describe rather effective continuous relaxation based bounds that provide a lower bound for the mixed integer max-min optimization problems, where the objective functions of the decision-makers have the same functional form but have opposite signs. Specifically, these approaches relax the integrality constraints of the follower and then reformulate the resulting bilevel linear optimization problem as a MILP through optimality conditions. In [14], the knapsack interdiction problem is studied and the concept of the critical items in the follower's knapsack problem is explored to further improve the bounds of the continuous relaxation. However, the aforementioned types of bounding methods exploit specific structure of the considered classes of bilevel problems and are not applicable for more general MIBLPs.

1.3 Overview

The remainder of the paper is organized as follows. Section 2 provides a formal description of our proposed framework and summarizes our contributions. Section 3 develops the hierarchy of lower and upper bounds for the considered class of MIBLPs. Then, in Section 4, we describe two single-level MILP reformulations for our bilevel problems. Section 5 considers a class of bilevel problems for which the follower's local optimality implies global optimality and hence, the follower can be viewed as rational, despite using a solution methodology typically employed as a heuristic. Finally, extensively computational experiments are conducted to illustrate the effectiveness of our proposed framework in Section 6.

2 Bilevel Optimization With k-optimal Follower

The main idea in the remainder of the paper is to consider an optimality-based relaxation of (BP) in which the follower's response is required only to be a locally optimal solution to the parametric follower's problem. To formally define what we mean by locally optimal, we utilize the Hamming distance [24] between two binary vectors y and $\bar{y}$, which is the number of positions at which these vectors are different, i.e., $\|\bar{y} - y\|_1$. Then given a positive integer k, in response to the leader's decision x, the follower chooses a solution in the k-optimal reaction set, defined as follows.

Definition 1. *The* k-swap neighborhood *of* $y \in \{0,1\}^n$ *is the set*

$$\mathcal{N}_k(y) = \{\bar{y} \in \{0,1\}^n : \|\bar{y} - y\|_1 \leq k\} \qquad (\mathcal{N}_k)$$

of all vectors within Hamming distance k *of* y.

Definition 2. *The* k-optimal reaction set *with respect to* $x \in \mathcal{X}$ *is the set*

$$\mathcal{R}_k(x) = \{y \in \mathcal{S}(x) : \beta y \geq \beta \bar{y} \quad \forall\, \bar{y} \in \mathcal{N}_k(y) \cap \mathcal{S}(x)\} \qquad (\mathcal{R}_k)$$

of all k-optimal solutions to the lower-level problem with respect to $x \in \mathcal{X}$.

The concept of a k-optimal solution in Definition 2 is commonly used in the literature on combinatorial optimization in methods that exploit local optimality as a way of generating heuristic solutions; see examples in the context of the traveling salesman [1, 12, 23], routing and scheduling [8, 45, 46], as well as many other combinatorial optimization problems [29, 30, 41, 47].

The mixed integer bilevel linear optimization problem with k-optimal follower is then formally stated as follows:

$$\begin{aligned} \eta_k^* = \max_{x,y} \ & \alpha^1 x + \alpha^2 y \\ \text{s.t.} \ & x \in \mathcal{X}, \\ & y \in \mathcal{R}_k(x). \end{aligned} \tag{BP$_k$}$$

We say that (x, y) is feasible for (BP$_k$) if $x \in \mathcal{X}$ and $y \in \mathcal{R}_k(x)$. Note that, as with (BP), the formulation of (BP$_k$) implicitly assumes the optimistic case, although the approach can also be generalized to the pessimistic case (there is a brief discussion on this issue in Section 7).

Observe that (x, y) is a feasible solution for (BP$_0$) if and only if $(x, y) \in \mathcal{S}$. Hence, $\mathcal{R}_0(x) = \mathcal{S}(x)$ and (BP$_0$) is equivalent to (SLR). Furthermore, $\mathcal{R}_n(x) = \mathcal{R}(x)$, so that (BP$_n$) is equivalent to (BP). We show in Section 3 that for any other $k \in \{0, \ldots, n\}$, the optimal objective function value of (BP$_k$) provides an upper bound on the optimal objective function value of (BP); furthermore, the optimal objective function value of (BP$_k$) is monotonic in k. Optimal solutions to (BP$_k$) can also be used to derive a hierarchy of monotonically increasing lower bounds for (BP).

To solve (BP$_k$), in Section 4 we propose two single-level MILP formulations solvable by standard MILP solvers. The first one follows a disjunctive-based approach with additional logical variables. The second formulation extends the previous idea and exploits the inherent structure of (BP$_k$) through the lens of mixing-set inequalities [3, 22, 62]. Our extensive numerical study with these MILP formulations indicates that the developed bounds are substantially better than the SLR-based bounds. Furthermore, the bounds converge to the optimal objective function value of (BP) for rather small values of k, and the required computational efforts is small. This observation suggests that the bounds provided by (BP$_k$) have tremendous potential for boosting the performance of exact solvers, especially for bilevel problems with low quality single-level relaxation bounds, which are common in practice.

Aside from the obvious usefulness of the bounds that can be derived from (BP$_k$), another important reason for studying (BP$_k$) is that it provides an exact reformulation for classes of MIBLPs for which k-optimal solutions (ideally, for some small fixed k) are also globally optimal for the lower-level problem. In Section 5, we exploit this idea by showing that 2-optimality of the lower-level problem implies global optimality for a general class of bilevel matroid problems. Specialized approaches for solving the bilevel minimum spanning tree problem (BMST) are then developed and tested in Section 6.3.

Finally, the proposed modeling framework (BP$_k$) provides a natural connection between theoretical exact formulations for hierarchical decision-making problems and the practical considerations arising in many real-life applications. In standard exact formulations of bilevel optimization problems, it is typically assumed that the follower is completely rational and their computational resources are sufficient to solve the lower-level problem to global optimality for any leader's decision. In many practical settings, it is clear that this is an unrealistic assumption. The follower often faces a situation in which either their computational resources are limited or they simply lack the knowledge to develop an efficient approach to obtaining the exact solution of their lower-level problem (this may be the case, in particular, when the follower's problem is NP-hard). Furthermore, in practice it is often the case that the follower only seeks a high-quality sub-optimal solution within reasonable time. To address such "inexact" followers, Smith et al. [50] and Zare et al. [60] study mxied integer bilevel optimization problems, where the reaction solutions by the follower are computed using a finite set of heuristic algorithms. In our framework, we do not specify the follower's strategies and algorithms, but instead quantify their possible reaction by exploiting the concept of locally optimal solutions. Thus, the proposed (BP_k) problems can be viewed as a more general modeling framework for hierarchical decision-making problems than the classical MIBLPs.

2.1 Characterization of k-optimal Reaction Set

Although the definition of the k-optimal reaction set given is already straightforward, we would ideally like a characterization that can be used to formulate (BP_k) as a mathematical optimization problem (preferably an MILP). To develop such a characterization, we first define the notion of an *improving k-swap.*

Definition 3. *A vector $w \in \{-1, 0, 1\}^n$ represents an* improving k-swap *if $\|w\|_1 = k$ and $\beta w > 0$. The set of all improving k-swaps is denoted by $\mathcal{T}_k$.*

In the above definition, the members of $\mathcal{T}_k$ represent ways of flipping the values of k variables in a given solution to get a new solution with improved objective function value. Note that membership in $\mathcal{T}_k$ only considers the number of flips and their effect on the objective function value, not the effect on feasibility of the lower-level problem, since the effect of the k-swap on feasibility would vary depending on the solution.

To illustrate, let $x \in \mathcal{X}$, $y \in \mathcal{S}(x)$, and $w \in \mathcal{T}_k$ be given. Applying the k-swap to y, we get $y + w$, which is an improved solution if and only if $y + w \in \mathcal{S}(x)$. Note that $y + w$ may be infeasible either because $G(y + w) \not\leq d - Ax$ or because $y + w \notin \{0, 1\}^n$, e.g., $y_i = w_i = 1$.

The following necessary and sufficient conditions characterize membership in $\mathcal{R}_k(x)$ for $x \in \mathcal{X}$. Informally, the result says that y is k-optimal if and only if no improving j-swap is feasible for $j \in [k]$.

Proposition 1. *Let $(x, y) \in \mathcal{S}$. Then $y \in \mathcal{R}_k(x)$ if and only if $y + w \notin \mathcal{S}(x)$ for all $w \in \mathcal{T}^k$, where $\mathcal{T}^k = \cup_{j\in[k]}\mathcal{T}_j$.*

Proof. Let $k \in [n]$ and $x \in \mathcal{X}$ be given. There are two parts to the proof.

"$\Leftarrow$" We prove the contrapositive. Let $y \notin \mathcal{R}_k(x)$ given. Then there exists $\bar{y} \in \mathcal{N}_k(y) \cap \mathcal{S}(x)$ such that $\beta\bar{y} > \beta y$. Let $\bar{w} = \bar{y} - y$ and $j = \|w\|_1$. We have $j \in [k]$, $w \in \mathcal{T}_j$, and $y + w \in \mathcal{S}(x)$, so the contrapositive is proven.

"$\Rightarrow$" We again prove the contrapositive. We therefore have $j \in [k]$ and $w \in \mathcal{T}_j$ such that $\bar{y} = y + w \in \mathcal{S}(x)$. Then $\bar{y} \in \mathcal{N}_k(y) \cap \mathcal{S}(x)$ and $\beta\bar{y} > \beta y$, so the contrapositive is proven.

■

We next illustrate Definition 3 and Proposition 1 as follows.

Example 1. *Consider the bilevel knapsack problem*

$$\begin{aligned}
\max_{x,y} \ & -\beta y \\
s.t. \ & \sum_{j=1}^{n} x_j \leq b, x \in \{0,1\}^n, \\
& y \in \arg \max_{\bar{y}\in\{0,1\}^n} \{\beta\bar{y} : \ a\bar{y} \leq C, x_j + \bar{y}_j \leq 1 \ \forall j \in [n]\},
\end{aligned}$$

where $n = 6$*, vector* $\beta = (70, 40, 39, 37, 17, 15)$*, weight vector* $a = (28, 25, 20, 18, 13, 10)$*, the leader's and the follower's knapsack capacities are given by* $b = 1$ *and* $C = 30$*, respectively. We note that*

$$\min_x \max_y \ \beta y = -\max_x \Big(-\max_y \ \beta y \Big)$$

and thus, the considered example is essentially an instance of the knapsack interdiction problem [16].

For a fixed leader's decision $x_0 = (1, 0, 0, 0, 0, 0)^\top$*, the lower-level problem becomes a knapsack problem given by*

$$\begin{aligned} \max_{y \in \{0,1\}^6} \quad & 70y_1 + 40y_2 + 39y_3 + 37y_4 + 17y_5 + 15y_6 \\ s.t. \quad & \qquad 25y_2 + 20y_3 + 18y_4 + 13y_5 + 10y_6 \le 30, \\ & y_1 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad = 0. \end{aligned}$$

First, consider the case of $k = 1$*. From Definition 3, we have* $\mathcal{T}_1 = \{w \in \{-1, 0, 1\}^n : \ \beta w > 0, \|w\|_1 = 1\}$*. Since* $\beta_j > 0$ *for each* j*, it follows that*

$$\mathcal{T}_1 = \{e^j : \ j \in [n]\}.$$

Furthermore, we have y *is a 1-optimal solution for the follower if and only if it is maximal (i.e.,* $ay \le C$ *and* $ay + a_j > C$ *for any* j *such that* $y_j = 0$*). Therefore, it can be verified that in this example we have that*

$$\mathcal{R}_1(x_0) = \{\{2\}, \{3, 6\}, \{4, 6\}, \{5, 6\}\}.$$

In the above, for simplicity we use subsets of the selected items in the follower's knapsack to describe $\mathcal{R}_1(x_0)$ *instead of the corresponding binary vector* y*, e.g., set* $\{2\}$ *denotes the follower's solution* $y = (0, 1, 0, 0, 0, 0)^\top$*.*

Next, it is also easy to verify that

$$\mathcal{T}_2 = \{e^i - e^j : \ i, j \in [n], \beta_i > \beta_j\} \cup \{e^i + e^j : i, j \in [n], i \ne j\}$$

and the 2-optimal reaction set of the follower is given by $\mathcal{R}_2(x_0) = \{\{2\}, \{3, 6\}\} \subseteq \mathcal{R}_1(x_0)$*. Then we can compute*

$$\begin{aligned} \mathcal{T}_3 = & \{e^i + e^j + e^k : i, j, k \in [n], i \ne j \ne k\} \cup \\ & \{e^i + e^j - e^k : i, j, k \in [n], i \ne j \ne k, \beta_i + \beta_j > \beta_k\} \cup \\ & \{e^i - e^j - e^k : i, j, k \in [n], i \ne j \ne k, \beta_i > \beta_j + \beta_k\}. \end{aligned}$$

Finally, we have that $\mathcal{R}_3(x_0) = \cdots = \mathcal{R}_n(x_0) = \{\{3, 6\}\}$*. It implies that in this instance the* k*-optimal reaction sets of follower are monotone decreasing and converge to* $\mathcal{R}(x_0)$ *for* $k = 3$*.* ■

2.2 Complexity of (BP$_k$)

The next question of interest is to explore the theoretical computational complexity of computing (BP$_k$) for a fixed value of k. It is easy to guess that since the problem of simply finding a feasible solution to a 0-1 integer optimization problem is already NP-hard, then (BP$_k$) should also be NP-hard. In fact, for any pure 0-1 integer optimization problem, one can easily construct a bilevel optimization problem with a lower-level problem in which a solution is optimal if and only if it is 1-optimal.

Theorem 1. (BP$_k$) *is* NP*-hard for any fixed integer* $k \ge 1$*.*

Proof. We show that pure binary integer optimization can be reduced to an instance of (BP$_k$). Let $\mathcal{X} \subseteq \{0,1\}^n$ be the feasible region of a pure binary integer optimization problem with objective function vector $\alpha \in \mathbb{R}^n$. Let an instance of (BP$_k$) be defined as follows. The set $\mathcal{X}$ is as given. We let $A = I_n$, $G = -I_n$, $d = 0$, $\beta_i = -1$ for $i \in [n]$, $\alpha^1 = \alpha$ and $\alpha^2 = 0$. Then the second-level problem is trivially solvable, since $\mathcal{R}_k(x) = \mathcal{R}(x) = \{x\}$ for all $x \in \mathcal{X} \subseteq \{0,1\}^n$. The solution to this instance of (BP$_k$) also solves

$$\max_{x \in \mathcal{X}} \alpha x$$

■

Remark 1. *It is evident that the proof did not depend at all on the lower-level problem. As long as* $\alpha^2 = 0$*, the solution to* (BP$_k$) *will be the same as that of the binary integer optimization problem. A stronger result also holds—that the decision version of* (BP$_k$) *is* NP*-complete for fixed* k*—as we show later.*

Remark 2. *The MIBLP is known to be hard for complexity class* Σ_2^p *in general [28, 37]. However, in Section 4, we demonstrate a polynomial-time procedure to reduce* (BP$_k$) *to a single-level linear MILP of polynomial size for a fixed value of* k*. Thus, for any fixed* k*, the decision version of* (BP$_k$) *is in class* NP*.*

3 Hierarchy of Bounds

In this section, we formally describe the hierarchy of upper and lower bounds associated with (BP$_k$). Some basic properties of these bounds are also established.

3.1 Upper Bounds

We first show that (BP$_k$) provides a natural hierarchy of upper bounds for (BP). In particular, these upper bounds can be shown to be progressively tighter with increasing k. Formally:

Theorem 2. $\eta^{SLR} = \eta_0^* \geq \eta_1^* \geq \eta_2^* \geq \cdots \geq \eta_n^* = \eta^*$.

Proof. Recall that η_k^* is computed as

$$\eta_k^* = \max \{\alpha^1 x + \alpha^2 y : x \in \mathcal{X}, y \in \mathcal{R}_k(x)\}.$$

To prove $\eta_k^* \geq \eta_{k+1}^*$ for $k = 0, 1, \ldots, n-1$, it is sufficient to observe from Proposition 1 that $\mathcal{R}_k(x) \supseteq \mathcal{R}_{k+1}(x)$. If $k = 0$, then $\mathcal{N}_0(y) = \{y\}$ and $\mathcal{R}_0(x) = \mathcal{S}(x)$, which implies that $\eta^{\text{SLR}} = \eta_0^*$. If $k = n$, then $\mathcal{N}_n(y) = \{0,1\}^n$ and $\mathcal{R}_n(x) = \mathcal{R}(x)$, which implies that $\eta_n^* = \eta^*$. ■

Our computational study in Section 6 indicates that the upper bound η_k^* is substantially tighter than the SLR-based bound, η_0^*, even for $k = 1$, and the optimal objective function value of (BP$_k$) converges to that of (BP) rather fast. The following example is provided as an illustration.

Example 1 (continued). *Observe that in Example 1,* $\eta_0^* = 0$ *with optimal solution* $(x^0, y^0) = (\mathbf{0}, \mathbf{0})$*. For* $k = 1$*, we can verify that the optimal leader's decision for* (BP$_1$) *is* $\{6\}$ *and the follower's* 1*-optimal solution is* $\{5\}$*, resulting in* $\eta_1^* = -17$*. For* $k = 2$*, the leader selects* $\{1\}$*, with the follower's* 2*-optimum given by* $\{2\}$*; thus,* $\eta_2^* = -40$*. For* $k = 3$*, the leader's optimal decision is* $\{1\}$ *and the follower chooses* $\{3, 6\}$ *with* $\eta_3^* = -54$*. We can further verify that* $\eta_4^* = \cdots = \eta_6^* = -54$ *with the leader's and the follower's optimal decisions given by* $\{1\}$ *and* $\{3, 6\}$*, respectively.* ■

In the above example in order to obtain an optimal solution for (BP) it is sufficient to solve (BP_3). Thus, a natural question is to establish when an optimal solution for (BP_k) is also optimal for (BP). The following results provide possible answers for this question.

Proposition 2. *Given an optimal solution (x^k, y^k) for* (BP_k)*, if $y^k \in \mathcal{R}(x^k)$ (i.e., y^k is also a globally optimal solution for the lower-level problem with respect to the leader's decision x^k), then (x^k, y^k) is also optimal for* (BP) *and $\eta_k^* = \eta_{k+1}^* = \cdots = \eta_n^* = \eta^*$.*

Proof. Since $y^k \in \mathcal{R}(x^k)$, we have that (x^k, y^k) is a bilevel feasible solution. It follows that $\eta^* \geq \alpha x^k + \beta y^k = \eta_k^*$ as (x^k, y^k) is optimal for (BP_k). Therefore, based on Theorem 2, we have $\eta_k^* = \eta_{k+1}^* = \cdots = \eta_n^*$. ■

Note that Proposition 2 provides a practical approach to verify whether $\eta_k^* = \eta^*$. Next, we discuss a more general condition.

Proposition 3. *If any k-optimal solution of the lower-level problem is also a globally optimal solution for the lower-level problem, then $\eta_k^* = \eta_{k+1}^* = \cdots = \eta_n^*$.*

Proof. It follows directly from the observation that $\mathcal{R}_k(x) = \mathcal{R}_{k+1}(x) = \cdots = \mathcal{R}_n(x)$. ■

For some single-level combinatorial optimization problems, locally optimal solutions are also globally optimal for reasonably small values of k, e.g., the minimum spanning tree problem with $k = 2$. Hence, Proposition 3 provides us with one possible approach for treating bilevel generalizations of such problems. In Section 5, we use Proposition 3 in the context of a general class of bilevel matroid problems.

On the other hand, it also may occur that the upper bounds provided by (BP_k), $k \geq 1$, do not improve the SLR-based bound, η_0^*. The next example illustrates this situation.

Example 2. *Consider an instance of the bilevel knapsack problem:*

$$\max_{x,y} \alpha(x+y)$$
$$s.t. \sum_{j=1}^{n} x_j \leq 1, x_1 = 0, x \in \{0,1\}^n,$$
$$y \in \arg\max_{\bar{y} \in \{0,1\}^n} \{\sum_{j=1}^{n} \beta \bar{y} : ax + a\bar{y} \leq C, x_j + \bar{y}_j \leq 1 \; \forall j \in [n]\},$$

where $n = 4$, vector $\alpha = (15, 7, 6, 5)$, vector $\beta = (10, 9, 7, 4)$, weight vector $a = (10, 6, 4, 3)$, and $C = 10$. We can verify that (SLR), (BP_1) *and* (BP_2) *lead to the same optimal objective function value of* 15 *with the leader's decision $\emptyset$ and the follower's decision $\{1\}$. That is, $\eta_0^* = \eta_1^* = \eta_2^* = 15$. On the other hand, $\eta^* = \eta_3^* = 13$ for* (BP_3) *with the leader's decision $\emptyset$ and the follower's decision $\{2, 3\}$.* ■

Finally, this example can be generalized via the following analytical result.

Proposition 4. *If $\alpha^2 w > 0$ for any $w \in \mathcal{T}^k$ and $1 \leq k \leq n$, then $\eta_0^* = \eta_1^* = \cdots = \eta_k^*$.*

Proof. From Theorem 2, we have $\eta_0^* \geq \cdots \geq \eta_{k-1}^* \geq \eta_k^*$. Thus, we only need to prove that $\eta_0^* \leq \eta_k^*$. Suppose (x^0, y^0) is an optimal solution of (BP_0). Then it suffices to show that (x^0, y^0) is also a feasible solution for (BP_k), that is $y^0 \in \mathcal{R}_k(x^0)$. We first note that $y^0 \in \mathcal{S}(x^0)$ and $\alpha^2 y^0 \geq \alpha^2 y$ for any $y \in \mathcal{S}(x^0)$.

Suppose $y^0 \notin \mathcal{R}_k(x^0)$, then based on Proposition 1, there exists $w \in \mathcal{T}^k$ such that $y^0 + w \in \{0,1\}^n$ and $y^0 + w \in \mathcal{S}(x^0)$. Since $\alpha^2 w > 0$, then $\alpha^2(y + w) > \alpha^2 y$, which contradicts with the assumption that (x^0, y^0) is optimal for (BP_0). Hence, $y \in \mathcal{R}_k(x^0)$ and the result follows. ■

Corollary 1. *If $\alpha_j^2 > 0$ for all $j \in [n]$ such that $\beta_j > 0$, then $\eta_0^* = \eta_1^*$.*

3.2 Lower Bounds

As outlined in Section 2, (BP_k) can also be exploited to construct lower bounds for (BP). Formally, let (x^k, y^k) be an optimal solution for (BP_k). Denote by $\hat{y}^k$ the follower's optimal solution that corresponds to the leader's decision x^k, i.e., $\hat{y}^k \in \mathcal{R}(x^k)$. Clearly, a pair $(x^k, \hat{y}^k)$ forms a bilevel feasible solution for (BP). Then we define:

$$\hat{\eta}_k = \alpha^1 x^k + \alpha^2 \hat{y}^k,$$

which provides a valid lower bound for η^*.

However, it can be verified (see an example below) that $\hat{\eta}_k$ is not necessarily monotonic in k. To present a hierarchy of monotonically increasing lower bounds for (BP), we need to slightly modify the definition of $\hat{\eta}_k$ as follows:

Definition 4. *Given an integer $k, 1 \leq k \leq n$, the modified $\hat{\eta}_k$, denoted as $\hat{\eta}'_k$, is given by*

$$\hat{\eta}'_k = \max_{t=0,1,\ldots,k} \{\alpha^1 x^t + \alpha^2 \hat{y}^t\}.$$

Then we obtain the following hierarchy of lower bounds:

Theorem 3. *$\hat{\eta}_k \leq \hat{\eta}'_k$ for $1 \leq k \leq n$, and $\hat{\eta}^{SLR} = \hat{\eta}_0 \leq \hat{\eta}'_1 \leq \hat{\eta}'_2 \leq \cdots \leq \hat{\eta}'_n = \eta^*$.*

Proof. Note that $\hat{\eta}_k = \alpha^1 x^k + \alpha^2 \hat{y}^k \leq \eta^*$. Thus, based on Definition 4 we have $\hat{\eta}_k \leq \hat{\eta}'_k = \max_{t=0,1\ldots,k}\{\hat{\eta}_k\} \leq \eta^*$. Also, it directly follows from Definition 4 that $\hat{\eta}'_k \leq \hat{\eta}'_{k+1}$ for $0 \leq k \leq n-1$. If $k = n$, then $\mathcal{R}_n(x) = \mathcal{R}(x)$, which yields that $\eta^* = \eta^*_n = \hat{\eta}_n \leq \hat{\eta}'_n \leq \eta^*$, and the result follows. ■

We next illustrate the considered lower bounds in the following example.

Example 3. *Consider an instance of the knapsack interdiction problem:*

$$\max_{x \in \{0,1\}^6} \Big(- \max_{y \in \{0,1\}^6} 11y_1 + 2y_2 + 7y_3 + 8y_4 + 3y_5 + 10y_6 \Big)$$

$$s.t. \sum_{j=1}^{6} x_j \leq 1, \; x_j + y_j \leq 1 \; \forall j \in [6],$$

$$14y_1 + 12y_2 + 6y_3 + 5y_4 + 3y_5 + 2y_6 \leq 14.$$

Then we compute that $\eta^*_0 = 0$ with the leader's decision $\mathbf{0}$, which leads to $\hat{\eta}_0 = \hat{\eta}'_0 = -25$ with follower's decision $\{4, 5, 6\}$. For $k = 1$, the leader's optimal decision set is $\{6\}$ with $\eta^*_1 = -2$. We can compute $\hat{\eta}_1 = \hat{\eta}'_1 = -18$ with follower's decision $\{3, 4, 5\}$. For $k = 2$, we see that $\mathbf{0}$ is optimal for the leader with $\eta^*_2 = -11 < \eta^*_1$, while its corresponding lower bound $\hat{\eta}_2 = -25 < \hat{\eta}_1$ with the follower's decision $\{3, 4, 6\}$. On the contrary, our modified lower bound $\hat{\eta}'_2 = \max_{k=0,1,2}\{\hat{\eta}_k\} = -18 \geq \hat{\eta}'_1$. We can further verify that $\hat{\eta}'_3 = \cdots = \hat{\eta}'_6 = \eta^* = -18$. ■

4 Extended Formulations

In this section, we present two extended formulations based on the concepts presented so far. In Section 4.1, we begin by describing a formulation of (BP_1), i.e., the case when $k = 1$, using a disjunctive-based approach. We then generalize this formulation to general k in Section 4.2. In doing so, we show that (BP_k) is polynomially reducible to a single-level MILP for any fixed k. Additionally, several preprocessing procedures are considered to reduce the number of variables and constraints in the proposed MILP formulations. Finally, by looking carefully into the structure of (BP_k), we identify, somewhat surprisingly, a mixing-set substructure within (BP_k), which is exploited to develop a tighter extended MILP formulation in Section 4.3.

4.1 Formulation for (BP_1)

For $k = 1$, Definitions 1 and 2 imply that the the follower's 1-optimal reaction set is defined as:

$$\mathcal{R}_1(x) = \{y \in \mathcal{S}(x) : \beta y \geq \beta \bar{y} \quad \forall\, \bar{y} \in \mathcal{N}_1(y) \cap \mathcal{S}(x)\},$$

where $\mathcal{N}_1(y) = \{\bar{y} \in \{0,1\}^n : ||y - \bar{y}||_1 \leq 1\}$. We have $\mathcal{T}_1 = \{e^j :\ \beta_j > 0\}$. Then Proposition 1 can be used to provide a simplified condition to determine whether $y \in \mathcal{R}_1(x)$ as follows:

Proposition 5. *Let $(x, y) \in \mathcal{S}$. Then $y \in \mathcal{R}_1(x)$ if and only if either $y_j = 1$, or $y_j = 0$ and $y + e^j \notin \mathcal{S}(x)$ for $j \in J^+_\beta$, where $J^+_\beta = \{j \in [n] : \beta_j > 0\}$.*

Observe that for any $y \in \{0,1\}^n$, if $y \notin \mathcal{S}(x)$, then based on Assumption 1, there must exist some row $i \in [m]$ such that

$$a^i x + g^i y \geq d_i + 1,$$

which (recalling that a^i and g^i are the i^{th} rows of matrices A and G, respectively) yields that

$$\Big\{y \in \{0,1\}^n : y \notin \mathcal{S}(x)\Big\} = \bigcup_{i=1}^{m} \Big\{y \in \{0,1\}^n : g^i y \geq d_i + 1 - a^i x\Big\}.$$

To develop a formulation that reflects the conditions in Proposition 5, we introduce binary variables z_{ij} for $i \in [m]$, $j \in J^+_\beta$, such that when $z_{ij} = 1$, we must have $a^i x + g^i(y + e^j) \geq d_i + 1$, i.e., $z_{ij} = 1 \Rightarrow y + e^j \notin \mathcal{S}(x)$. Using this set of auxiliary binary variables, we reformulate (BP_1) as a linear MILP in the following form:

$$\eta_1^* = \max_{x,y,z}\ \alpha^1 x + \alpha^2 y \tag{BP_1-DF-a}$$

$$\text{s.t.}\ \ (x, y) \in \mathcal{S}, \tag{BP_1-DF-b}$$

$$a^i x + g^i y + z_{ij}(\mu_i - h_{ij}) \geq \mu_i \quad \forall i \in [m], j \in J^+_\beta, \tag{BP_1-DF-c}$$

$$\sum_{i=1}^{m} z_{ij} + y_j \geq 1 \quad \forall j \in J^+_\beta, \tag{BP_1-DF-d}$$

$$z_{ij} \in \{0,1\} \quad \forall i \in [m], j \in J^+_\beta, \tag{BP_1-DF-e}$$

where $h_{ij} = d_i + 1 - g^i_j$, and μ_i is a sufficiently small constant parameter (below we provide some additional discussion on its appropriate values).

To demonstrate the correctness of the obtained formulation, we make the following observations. First, note that constraint (BP_1-DF-c) ensures that if $z_{ij} = 1$, then we must have $a^i x + g^i(y + e^j) \geq d_i + 1$; if $z_{ij} = 0$, then constraint (BP_1-DF-c) is always satisfied for appropriate values of μ_i. Constraint (BP_1-DF-d) guarantees that if $y_j = 0$ for some $j \in J^+_\beta$, then there exists $i \in [m]$ such that $z_{ij} = 1$. This, in turn, ensures that $y + e^j \notin \mathcal{S}(x)$. On the other hand, since $y_j = 1$ already implies $y + e^j \notin \mathcal{S}(x)$, we do not need to ensure violation of any constraints in this case.

For a constraint (BP_1-DF-c) to be redundant when $z_{ij} = 0$, it is sufficient to have

$$\mu_i \leq \min_{(x,y)\in\mathcal{S}} a^i x + g^i y.$$

Based on Assumption 2, there exist vectors l and u such that for any $x \in \mathcal{X}$ we have $l \leq x \leq u$, which leads to a straightforward choice of μ_i as

$$\mu_i = \sum_{j:a^i_j>0} a^i_j l_j + \sum_{j:a^i_j<0} a^i_j u_j + \sum_{j:g^i_j<0} g^i_j.$$

If all entries in A and G are non-negative, then μ_i can be trivially set to 0. Note that the chosen value of μ_i may influence the quality of the above MILP reformulation (BP_1-DF). A tighter μ_i can be achieved for bilevel problems with identifiable structures; this issue is further addressed in Section 6.

Next, we discuss how to reduce the number of variables and constraints in the formulation (BP_1-DF) through some simple preprocessing steps.

Proposition 6. *(i) If $g^i_j \leq 0$ for some $i \in [m]$ and $j \in J^+_\beta$, then $z_{ij} = 0$ for any feasible solution of* (BP_1-DF).

(ii) If $\alpha^2_j > 0$ and $g^i_j \geq 0$ for some $j \in J^+_\beta$ and all $i \in [m]$, then removing variables z_{ij} and the associated constraints for all $i \in [m]$ in (BP_1-DF) *does not change its optimal objective function value.*

Proof. (*i*) Suppose (x, y, z) is a feasible solution of (BP_1-DF). Since $(x, y) \in \mathcal{S}$, then $a^i x + g^i y \leq d_i$ for all $i \in [m]$. If $z_{ij} = 1$ for some $i \in [m]$ and $j \in J^+_\beta$ such that $g^i_j \leq 0$, then from constraint (BP_1-DF-c), we have

$$a^i x + g^i y \geq d_i + 1 - g^i_j \geq d_i + 1,$$

where the second inequality follows from the assumption that $g^i_j \leq 0$. Thus, we obtain a contradiction and it immediately follows that z_{ij} should be equal to 0 in any feasible solution.

(*ii*) We refer to (BP'_1-DF) as the problem obtained by removing z_{ij_0} for some $j_0 \in J^+_\beta$ and all $i \in [m]$ such that $\alpha^2_{j_0} > 0$ and $g^i_{j_0} \geq 0$. Denote by η'_1 the resulting optimal objective function value of (BP'_1-DF). It is clear that $\eta'_1 \geq \eta^*_1$.

Assume (x', y', z') is an optimal solution of (BP'_1-DF). Next, it is sufficient to show that $y' \in \mathcal{R}_1(x')$ as the latter implies that $\eta'_1 \leq \eta^*_1$. By Proposition 5 (recall that z_{ij} is not removed from (BP_1-DF) for any $j \neq j_0$) we have that $y' + e^j \notin \mathcal{S}(x')$ for any $j \in J^+_\beta$ and $j \neq j_0$. Therefore, we only need to show $y' + e^{j_0} \notin \mathcal{S}(x')$.

If $y'_{j_0} = 1$, then the statement holds trivially. If $y'_{j_0} = 0$, then suppose $y' + e^{j_0} \in \mathcal{S}(x)$. Consider $j \in J^+_\beta$ such that $j \neq j_0$ and $y'_j = 0$. Since $y' + e^j \notin \mathcal{S}(x')$, assume that $a^i x + g^i(y' + e^j) \geq d_i + 1$ for some $i \in [m]$ (i.e., $z_{ij} = 1$ in constraint (BP_1-DF-c)). It immediately follows that $a^i x + g^i(y' + e^j + e^{j_0}) \geq d_i + 1$ as $g^i_{j_0} \geq 0$. Therefore, we have $y' + e^{j_0} + e^j \notin \mathcal{S}(x')$ for any $j \in J^+_\beta$. Hence, based on Proposition 5, $y' + e^{j_0}$ is a feasible solution for (BP_1-DF). However, $\eta^*_1 \geq \alpha^1 x' + \alpha^2(y' + e^{j_0}) = \alpha^1 x' + \alpha^2 y' + \alpha^2_{j_0} > \alpha^1 x' + \alpha^2 y' = \eta'_1$, which contradicts with the fact that $\eta^*_1 \leq \eta'_1$. Thus, $y' + e^{j_0} \notin \mathcal{S}(x')$ and the result follows.

Note that (*i*) simply follows from the fact that given any $y \in \mathcal{S}(x)$, if $y_j = 0$, then $a^i x + g^i(y + e^j) \leq d$, as $g^i_j \leq 0$.

■

Remark 3. *Note that if z_{ij} can be fixed to zero, then the corresponding constraint in* (BP_1-DF-c) *can also be removed.* ■

Remark 4. *Observe that the statement in Proposition 6 (ii) is consistent with Corollary 1. In particular, if $\alpha^2_j > 0$ for all $j \in J^+_\beta$ (as in Corollary 1), then in fact, we do not need to consider the signs of g^i_j and all z_{ij} variables can be removed, as* (BP_1-DF) *coincides with* (SLR). ■

Remark 5. *The proposed MILP formulation can be further strengthened as follows. Assume there exist $j_1, j_2 \in [n]$ such that $g^i_{j_1} \leq g^i_{j_2}$ for all $i \in [m]$. Clearly, if $y + e^{j_1} \notin \mathcal{S}(x)$ for some $y \in \mathcal{S}(x)$*

and $y_{j_1} = 0$*, then* $y + e^{j_2} \notin \mathcal{S}(x)$*. Thus, we can replace constraint* (BP_1-DF-d) *for* j_2 *with*

$$\sum_{i=1}^{m} z_{ij_2} + y_{j_1} y_{j_2} \geq y_{j_1}. \tag{1}$$

Constraint (1) *takes the value of* y_{j_1} *into consideration: if* $y_{j_1} = 0$*, then constraints* (BP_1-DF-c) *and* (BP_1-DF-d) *for* j_1 *ensure* $y + e^{j_1} \notin \mathcal{S}(x)$*. Hence, the values of* z_{ij_2} *are not required to be considered, as* $y + e^{j_2} \notin \mathcal{S}(x)$ *is already implied by* $y + e^{j_1} \notin \mathcal{S}(x)$*; otherwise, if* $y_{j_1} = 1$*, constraint* (1) *reduces to* (BP_1-DF-d)*. We can linearize the nonlinear item* $y_{j_1} y_{j_2}$ *through McCormick envelopes [25] by introducing additional binary variables.* ■

Next, we discuss how to generalize the MILP formulation (BP_1-DF) for (BP_1) to (BP_k).

4.2 Formulation for (BP_k)

Similar to the MILP formulation (BP_1-DF), we introduce binary variables z_{iw} for $i \in [m]$ and $w \in \mathcal{T}^k$ to verify the condition that $y + w \notin \mathcal{S}(x)$ (recall Proposition 1). We then reformulate (BP_k) as:

$$\eta_k^* = \max_{x,y,z} \alpha^1 x + \alpha^2 y \tag{BP_k-DF-a}$$

$$\text{s.t. } (x, y) \in \mathcal{S}, \tag{BP_k-DF-b}$$

$$a^i x + g^i y + z_{iw}(\mu_i - h_{iw}) \geq \mu_i \quad \forall i \in [m], w \in \mathcal{T}^k, \tag{BP_k-DF-c}$$

$$\sum_{i=1}^{m} z_{iw} + w^\top y + \|w^-\|_1 \geq 1 \quad \forall w \in \mathcal{T}^k, \tag{BP_k-DF-d}$$

$$z_{iw} \in \{0, 1\} \quad \forall i \in [m],\ w \in \mathcal{T}^k, \tag{BP_k-DF-e}$$

where μ_i is a sufficiently small constant chosen as in Section 4.1; $h_{iw} = d_i + 1 - g^i w$ for all $i \in [m]$ and $w \in \mathcal{T}^k$; and $\|w^-\|_1$ is the number of entires of w with negative values.

If $z_{iw} = 1$, then constraint (BP_k-DF-c) implies that $a^i x + g^i(y + w) \geq d_i + 1$; on the contrary, if $z_{iw} = 0$, then the associated constraint (BP_k-DF-c) is redundant. As for constraints (BP_k-DF-d), first note that $w^\top y + \|w^-\|_1 \geq 0$ for all $y \in \{0, 1\}^n$ and $w^\top y + \|w^-\|_1 = 0$ if and only if $y + w \in \{0, 1\}^n$. Hence, $w^\top y + \|w^-\|_1 > 0$ implies $y + w \notin \mathcal{S}(x)$. When $w^\top y + \|w^-\|_1 = 0$, we ensure $y + w \notin \mathcal{S}(x)$ by again employing Proposition 1, ensuring that there exists at least one $i \in [m]$ such that $z_{iw} = 1$, which further implies that $y + w \notin \mathcal{S}(x)$ by constraint (BP_k-DF-c). Hence, the overall combination of constraints (BP_k-DF-c) and (BP_k-DF-d) ensure $y + w \notin \mathcal{S}(x)$.

For a fixed k, the cardinality of $\mathcal{T}^k$ is $O(n^k)$ and the number of variables and constraints in (BP_k-DF) is $O(mn^k)$. Therefore, the above reformulation is of polynomial size for any fixed k. This is an interesting observation from the theoretical perspective in two respects. First, for a fixed value of k, we can say that (BP_k) is polynomially reducible to an MILP. It immediately follows that (BP_k) is in class NP; recall our discussions in Remark 2. Furthermore, if for some fixed k, any follower's k-optimal solution is globally optimal, then based on Proposition 3, we have (BP) $\equiv$ (BP_k). Thus, we conclude that the decision version of such a class of bilevel problems is NP-complete and the problem itself is NP-hard. For example, in Section 5 we show that (BP) $\equiv$ (BP_2) for a general class of bilevel matroid problems.

From the practical implementation perspective, the magnitude of $O(n^k)$ is still substantial, even for small values of k, when n is large. Thus, we now consider two essential questions: how to reduce the size of (BP_k-DF) through some simple preprocessing to remove irrelevant members of $\mathcal{T}^k$ and how to efficiently enumerate a relevant subset of $\mathcal{T}^k$ in an efficient manner.

We first discuss procedures for fixing some variable values and removing redundant variables and constraints from (BP$_k$-DF).

Proposition 7. *Let $k \geq 1$ be given.*

(i) If $g^i w \leq 0$ for some $w \in \mathcal{T}^k$ and $i \in [m]$, then $z_{iw} = 0$ for any feasible solution of (BP$_k$-DF)*.*

(ii) If $h_{iw^0} = h_{iw^1}$ for some $i \in [m]$ and $w^0, w^1 \in \mathcal{T}^k$, then there exists an optimal solution to (BP$_k$-DF) *in which $z_{iw^1} = z_{iw^0}$.*

Proof. (i) Suppose (x, y, z) is a feasible solution of (BP$_k$-DF). Since $(x, y) \in \mathcal{S}$, then $a^i x + g^i y \leq d_i$. If $z_{iw} = 1$, then from constraint (BP$_k$-DF-c), we have

$$a^i x + g^i y \geq d_i + 1 - g^i w \geq d_i + 1,$$

where the second inequality follows from the assumption that $g^i w \leq 0$. Thus, we obtain a contradiction and it immediately follows that z_{iw} should be equal to 0 in any feasible solution.

(ii) Suppose (x^*, y^*, z^*) is an optimal solution of (BP$_k$-DF), and $z^*_{iw^0} \neq z^*_{iw^1}$. Without loss of generality, assume $z^*_{iw^0} = 1$ and $z^*_{iw^1} = 0$. Define z' as follows: $z'_{iw^1} = 1$ and $z'_{iw} = z^*_{iw}$ for all $i \in [m]$ and $w \in \mathcal{T}^k \setminus \{w^1\}$. Observe that constraints (BP$_k$-DF-d) are trivially satisfied for (x^*, y^*, z'). Since $z^*_{iw^0} = 1$ and $h_{iw^0} = h_{iw^1}$, then $a^i x^* + g^i y^* \geq h_{iw^1}$. Thus, we can verify that constraints (BP$_k$-DF-c) also hold for (x^*, y^*, z'). Consequently, it follows that (x^*, y^*, z') is also optimal for (BP$_k$) and $z'_{iw^0} = z'_{iw^1}$.

■

Note that when we either fix $z_{iw^1} = 0$ or set it equal to z_{iw^0} for some $i \in [m]$ and $w^0, w^1 \in \mathcal{T}^k$, the corresponding constraint (BP$_k$-DF-c) can also be removed. Our computational study in Section 6 indicates that the preprocessing procedures in Proposition 7 can significantly decrease the number of variables and constraints (BP$_k$-DF-c). Also, observe that the computational efforts required to evaluate conditions (i) and (ii) in Proposition 7 are of order $O(nm|\mathcal{T}^k|)$, and $O(m|\mathcal{T}^k| \log(|\mathcal{T}^k|))$, respectively. Thus, the cardinality of $\mathcal{T}^k$ is a primary driver of the efficiency of the proposed approach. In view of this, we next discuss how to prune the components in $\mathcal{T}^k$ in order to effectively reduce the formulation size.

Proposition 8. *Let $k \geq 1$ be given.*

(i) Given $w^0 \in \mathcal{T}^{k-1}$ and $\ell \in [n]$ such that $w^0_\ell = 0$, let $w^1 = w^0 + e^\ell \in \mathcal{T}^k$. If $g^i_\ell \geq 0$ for all $i \in [m]$, then removing variables z_{iw^1} and the associated constraints for all $i \in [m]$ in (BP$_k$-DF) *does not change its optimal objective function value.*

(ii) Given $w^0 \in \mathcal{T}^k$ and $\ell \in [n]$ such that $w^0_\ell = 1$, let $w^1 = w^0 - e^\ell \in \mathcal{T}^{k-1}$. If $g^i_\ell \leq 0$ for all $i \in [m]$, then removing variables z_{iw^1} and the associated constraints for all $i \in [m]$ in (BP$_k$-DF) *does not change its optimal objective function value.*

(iii) If $\alpha^2 w > 0$ for some $w \in \mathcal{T}^k$, and $g^i w \geq 0$ for all $i \in [m]$, then removing variables z_{iw} and the associated constraints for all $i \in [m]$ in (BP$_k$-DF) *does not change its optimal objective function value.*

Proof. (i) Denote by (BP$'_k$-DF) and η'_k the problem and its optimal objective function value, respectively, where z_{iw^1} and associated constraints are removed from (BP$_k$-DF). It is clear that $\eta'_k \geq \eta^*_k$. It suffices to show that given any optimal solution (x', y', z') in (BP$'_k$-DF), there exists z^* such that (x', y', z^*) is a feasible solution of (BP$_k$-DF), which leads to $\eta'_k \leq \eta^*_k$

If $y' + w^1 \notin \{0,1\}^n$, then let $z^*_{iw^1} = 0$ for all $i \in [m]$, and $z^*_{iw} = z'_{iw}$ for all $w \in \mathcal{T}^k \setminus \{w^1\}$. Clearly, (x', y', z^*) is feasible for (BP$_k$-DF). Otherwise, if $y' + w^1 \in \{0,1\}^n$, then based on constraint (BP$'_k$-DF-d) for w^0, we have $\sum_{i=1}^m z'_{iw^0} \geq 1$. Since $g^i_\ell \geq 0$, we have $h_{iw^1} = d_i + 1 - g^i w^1 \leq h_{iw^0}$ for all $i \in [m]$. Let $z^*_{iw^1} = z'_{iw^0}$, and $z^*_{iw} = z'_{iw}$ for all $w \in \mathcal{T}^k \setminus \{w^1\}$, then we can verify that (x', y', z^*) is feasible for (BP$_k$-DF). This observation completes the proof.

(*ii*) The proof is similar to (*i*) above, and omitted for brevity.

(*iii*) Denote by (BP$'_k$-DF) and η'_k the corresponding problem and its optimal objective function value, respectively where z_{iw} and associated constraints are removed from $\mathcal{T}^k$ in (BP$_k$-DF). It is clear that $\eta'_k \geq \eta^*_k$. Suppose (x', y', z') is an optimal solution of (BP$'_k$). We next focus on proving that $y' \in \mathcal{R}_k(x)$, which implies that there exists z^* such that (x', y', z^*) is feasible for (BP$_k$-DF) and $\eta'_k \leq \eta^*_k$. Note that to verify $y' \in \mathcal{R}_k(x)$, we only need to show that $y' + w \notin \mathcal{S}(x)$.

If $y' + w \notin \{0,1\}^n$, then $y' + w \notin \mathcal{S}(x)$ trivially. Otherwise, we proceed by contradiction, so suppose $y' + w \in \mathcal{S}(x)$. Since $g^i w \geq 0$, then we can verify that $(x', y' + w)$ is also feasible for (BP$'_k$-DF). Moreover, $\alpha^1 x' + \alpha^2 (y' + w) > \alpha^1 x' + \alpha^2 y'$, which contradicts our initial assumption that (x', y', z') is an optimal solution for (BP$'_k$-DF). Thus, $y' + w \notin \mathcal{S}(x)$ and the result follows. ■

We outline the pseudocode of the procedure to construct $\mathcal{T}^k$ in Algorithm 1. In lines 8-11, we first find the component $\ell \in [n]$, for which the corresponding value of β is smallest among $j \in J^+_G$ for which $w_j = 1$, where $J^+_G := \{j \in [n] : g^i_j \geq 0 \ \forall i \in [m]\}$. Then w is discarded based on Proposition 8(*i*). Following Proposition 8(*ii*), we discard w in lines 12-14 if there exists $\ell \in J^-_G = \{j \in [n] : g^i_j \leq 0 \ \forall i \in [m]\}$ such that $w_\ell = -1$. We finally evaluate the conditions of Proposition 8(*iii*) in lines 15-17 to determine whether to add w into $\mathcal{T}^k$.

4.3 Strengthened Formulation for (BP$_k$)

Next, we provide a strengthened formulation based on the inherent structural properties of (BP$_k$). Though the results here are derived from first principles, we note that the properties we exploit and the resulting formulations are closely related to the mixing-set inequalities, which have been studied in a number of contexts [3, 22, 62].

To derive the stronger formulation, we first re-write constraints (BP$_k$-DF-c) as

$$a^i x + g^i y \geq \mu_i - (\mu_i - h_{iw}) z_{iw} \ \forall i \in [m], w \in \mathcal{T}^k. \tag{2}$$

This rewriting is to emphasize that the combined effect of the additional binary variables introduced is to essentially set the right-hand side of a single constraint to one of a number of different possible values. This is illustrated even more clearly by combining the inequalities (2) associated with $i \in [m]$ into the single inequality

$$a^i x + g^i y \geq \max_{w \in \mathcal{T}^k} \{\mu_i - (\mu_i - h_{iw}) z_{iw}\} \ \forall i \in [m]. \tag{3}$$

Of course, this inequality involves a non-linear function, but there is a way to combine the inequalities in a different way that yields a strong linear formulation that we describe next.

The next property that we use is that after the preprocessing procedure described in Proposition 7, the values of h_{iw} are distinct in the remaining constraints (BP$_k$-DF-c) for each $i \in [m]$. Denote the number of distinct values of h_{iw} for $i \in [m]$ as ℓ_i. We assume without loss of generality that $h_{iw^1} < h_{iw^2} < \cdots < h_{iw^{\ell_i}}$ for $w^1, w^2, \ldots, w^{\ell_i}$ in $\mathcal{T}^k$.

Algorithm 1 Algorithm for constructing $\mathcal{T}^k$

```
    Input A, G, d, α^1, α^2, β
 1: T^k ← ∅
 2: J_G^+ ← {j ∈ [n] : g_j^i ≥ 0 ∀i ∈ [m]}
 3: J_G^- ← {j ∈ [n] : g_j^i ≤ 0 ∀i ∈ [m]}
 4: for all w ∈ {−1, 0, 1}^n such that ‖w‖_1 ≤ k do
 5:     if βw ≤ 0 then
 6:         Discard w and go to Line 4
 7:     end if
 8:     β_0 ← min{β_j : w_j = 1 and j ∈ J_G^+}
 9:     if βw ≥ β_0 then
10:         Discard w and go to Line 4          // based on Proposition 8(i)
11:     end if
12:     if ∃j ∈ J_G^- such that w_j = −1 then
13:         Discard w and go to Line 4          // based on Proposition 8(ii)
14:     end if
15:     if α^2 w > 0 and g^i w > 0 for all i ∈ [m] then
16:         Discard w and go to Line 4          // based on Proposition 8(iii)
17:     end if
18:     T^k ← T^k ∪ {w}
19: end for
    Return T^k
```

The strengthened formulation for (BP_k) is then given by:

$$\tilde{\eta}_k^* = \max_{x,y,z} \alpha^1 x + \alpha^2 y \tag{BP$_k$-Mix-a}$$

$$\text{s.t.} \quad (x, y) \in \mathcal{S}, \tag{BP$_k$-Mix-b}$$

$$a^i x + g^i y + \sum_{j=1}^{\ell_i} (h_{iw^{j-1}} - h_{iw^j}) z_{iw^j} \geq \mu_i \quad \forall i \in [m], \tag{BP$_k$-Mix-c}$$

$$z_{iw^j} \geq z_{iw^{j+1}} \quad \forall j \in [\ell_i], i \in [m], \tag{BP$_k$-Mix-d}$$

$$\sum_{i=1}^{m} z_{iw} + w^\top y + \|w^-\|_1 \geq 1 \quad \forall w \in \mathcal{T}^k, \tag{BP$_k$-Mix-e}$$

$$z_{iw^j} \in \{0, 1\} \quad \forall j \in [\ell_i], i \in [m], \tag{BP$_k$-Mix-f}$$

where $h_{iw^0} = \mu_i$ for all $i \in [m]$. The underlying concept is the one illustrated earlier, that the value of the (variable) right-hand side is, in fact, controlled by the largest h_{iw} whose corresponding variable z_{iw} is equal to one (as illustrated in (3)), so that the whole set of constraints involving the original constraint $i \in [m]$ can be collapsed into a single constraint that dominates the set of original ones.

We refer the readers for more technical discussion about the extended formulation of the mixing-set inequality in [39]. Note that the z variables here ultimately play the same role as in the original formulation, but we enforce more structure on them with the addition of the precedence constraints (BP_k-Mix-d) for reasons that will become clear in the proof of the next theorem.

Theorem 4. (BP_k-Mix) *and* (BP_k-DF) *have the same optimal objective function values, that is,* $\eta_k^* = \tilde{\eta}_k^*$.

Proof. Let (x, y, z) be a feasible solution for (BP$_k$-DF). We show that there exists a feasible solution to (BP$_k$-Mix) with the same objective function value. For $i \in [m]$, let $j_0 = \max\{j \in [\ell_i] : z_{iw^j} = 1\}$. Let $z'_{iw^j} = 1$ for all $j = 1, \ldots, j_0$ and $z'_{iw^j} = 0$ for $j = j_0 + 1, \ldots, \ell_i$. Recalling (3), we have $a^i x + g^i y - h_{iw^{j_0}} \geq 0$. Then

$$
\begin{aligned}
a^i x + g^i y + \sum_{j=1}^{\ell_i} (h_{iw^{j-1}} - h_{iw^j}) z'_{iw^j} &= a^i x + g^i y + \sum_{j=1}^{j_0} (h_{iw^{j-1}} - h_{iw^j}) \\
&= a^i x + g^i y + h_{iw^0} - h_{iw^{j_0}} \\
&= a^i x + g^i y + \mu_i - h_{iw^{j_0}} \\
&\geq \mu_i.
\end{aligned}
$$

Therefore, (x, y, z') satisfies constraints (BP$_k$-Mix-c) and is feasible for (BP$_k$-Mix). It also has the same objective function value (BP$_k$-Mix) as (x, y, z) has in (BP$_k$-DF). It immediately yields that $\eta^*_k \leq \tilde{\eta}^*_k$.

On the other hand, let (x', y', z') be a feasible solution of (BP$_k$-Mix). For $i \in [m]$, let $j_0 = \max\{j \in [\ell_i] : z'_{iw^j} = 1\}$. Based on constraint (BP$_k$-Mix-c), we have

$$
\begin{aligned}
a^i x' + g^i y' + \sum_{j=1}^{\ell_i} (h_{iw^{j-1}} - h_{iw^j}) z'_{iw^j} &= a^i x' + g^i y' + h_{iw^0} - h_{iw^{j_0}} \\
&= a^i x' + g^i y' + \mu_i - h_{iw^{j_0}} \\
&\geq \mu_i
\end{aligned}
$$

which results in $a^i x' + g^i y' \geq h_{iw^{j_0}}$. Observe that constraints (BP$_k$-DF-c) are trivially satisfied for i and $j \geq j_0$ because $z'_{iw^j} = 0$ for $j \geq j_0$. If $j < j_0$, then $z'_{iw^j} = 1$, and we have $a^i x' + g^i y' \geq h_{iw^{j_0}} > h_{iw^j}$, which implies that (x', y', z') satisfies constraints (BP$_k$-DF-c). Therefore, (x', y', z') is itself feasible for (BP$_k$-DF) and $\tilde{\eta}^*_k \leq \eta^*_k$. ∎

When we remove the integrality constraints of decision variables in a MILP, the resulting linear optimization problem (LP) is referred to as the LP relaxation of the original MILP. It is known that the tightness of the LP relaxations for MILPs is a critical factor affecting the overall performance of the solver. We then show that the MILP formulation (BP$_k$-Mix) is stronger than (BP$_k$-DF).

Proposition 9. *The LP relaxation of* (BP$_k$-Mix) *is at least as strong as that of* (BP$_k$-DF)*.*

Proof. Let (x, y, z) be any feasible solution for the LP relaxation of (BP$_k$-Mix). It suffices to show that (x, y, z) is also feasible for the LP relaxation of (BP$_k$-DF). Since constraints (BP$_k$-DF-d) are also included in (BP$_k$-Mix), we need to show that (x, y, z) satisfies constraints (BP$_k$-DF-c). Based on constraints (BP$_k$-Mix-c) and (BP$_k$-Mix-d), we have, for any $i \in [m]$, and $j_0 \in [\ell_i]$,

$$
\begin{aligned}
a^i x + g^i y + (\mu_i - h_{iw^{j_0}}) z_{iw^{j_0}} &= a^i x + g^i y + \sum_{j=1}^{j_0} (h_{iw^{j-1}} - h_{iw^j}) z_{iw^{j_0}} \\
&\geq a^i x + g^i y + \sum_{j=1}^{j_0} (h_{iw^{j-1}} - h_{iw^j}) z_{iw^j} \\
&\geq \mu_i,
\end{aligned}
$$

where the first inequality follows from $z_{iw^j} \geq z_{iw^{j+1}}$ (recall constraint (BP$_k$-Mix-d)) and $h_{iw^{j-1}} - h_{iw^j} < 0$; the second inequality follows from our initial assumption that (x, y, z) satisfies constraint (BP$_k$-Mix-c). ∎

We next illustrate Theorem 4 and Proposition 9 with the following example.

Example 4. *Consider an instance of the bilevel problem:*

$$\min_{x \in \mathbb{R}} \max_{y} \{y_1 + y_2 : \ 2y_1 + 3y_2 \leq 4, \ y \in \{0,1\}^2\}.$$

If $k = 1$*, then* $\mathcal{T}_1 = \{e^1, e^2\}$*. Let* z_1, z_2 *be the binary variables in* (BP$_k$-DF) *that correspond to* e^1 *and* e^2*, respectively. We set* $\mu = 0$ *for constraint* (BP$_k$-DF-c)*. Then the feasible region for the LP relaxation of* (BP$_k$-DF) *with* $k = 1$ *is given as:*

$$\mathcal{Q}^1_{DF} = \left\{ (x, y, z) \in \mathbb{R} \times [0,1]^4 : \begin{array}{l} 2y_1 + 3y_2 \leq 4, \\ 2y_1 + 3y_2 \geq 3z_1, \\ 2y_1 + 3y_2 \geq 2z_2, \\ y_1 + z_1 \geq 1, \\ y_2 + z_2 \geq 1 \end{array} \right\}.$$

The feasible region for the LP relaxation of (BP$_k$-Mix) *with* $k = 1$ *is given as:*

$$\mathcal{Q}^1_{Mix} = \left\{ (x, y, z) \in \mathbb{R} \times [0,1]^4 : \begin{array}{l} 2y_1 + 3y_2 \leq 4, \\ 2y_1 + 3y_2 \geq 2z_2 + z_1, \\ y_1 + z_1 \geq 1, \\ y_2 + z_2 \geq 1, \\ z_2 \geq z_1 \end{array} \right\}.$$

It is easy to verify that $\mathcal{Q}^1_{Mix} \subseteq \mathcal{Q}^1_{DF}$*. Moreover, we can find a solution* $(0, 1, \frac{2}{3}, 1, \frac{1}{3})^\top \in \mathcal{Q}^1_{DF}$*, and this solution is not feasible in* $\mathcal{Q}^1_{Mix}$*. It immediately follows that* $\mathcal{Q}^1_{Mix} \subset \mathcal{Q}^1_{DF}$*.* ■

Remark 6. *Based on the results of the mixing-set in [3, 22], we can show that for any* $i \in [m]$*, the inequalities*

$$a^i x + g^i y + \sum_{j=1}^{\ell} (h_{iw^{t_{j-1}}} - h_{iw^{t_j}}) z_{iw^{t_j}} \geq \mu_i \quad \forall \{w^{t_1}, w^{t_2}, \ldots, w^{t_\ell}\} \subseteq \mathcal{T}^k,$$

where $h_{iw^{t_1}} < h_{iw^{t_2}} < \cdots < h_{iw^{t_\ell}}$ *are valid for* (BP$_k$-DF)*. The above inequalities are called star inequalities, which can be further shown to describe the convex hull of a mixing-set polytope; see [3, 22]. In our strengthened formulation* (BP$_k$-Mix)*, we note that constraint* (BP$_k$-Mix-c) *is a special class of star inequalities. We exploit the inherent structure in* (BP$_k$) *to simplify the star inequalities and provide the stronger formulation* (BP$_k$-Mix)*.* ■

5 Bilevel Matroid Optimization

Next, we focus on exploring in more detail a special case of (BP$_k$). Specifically, by exploiting Proposition 3 we show that (BP) ≡ (BP$_k$) for a general class of bilevel matroid problems whenever $k \geq 2$.

Definition 5 ([35])**.** *Let* $[n] = \{1, \ldots, n\}$ *be a finite set, and let* $\mathcal{F}$ *be a set of subsets of* $[n]$*. We say that* $M = ([n], \mathcal{F})$ *is a matroid if the following conditions are satisfied:*

(*i*) $\emptyset \in \mathcal{F}$*;*

(*ii*) *$S \in \mathcal{F}$ and $S' \subseteq S$ implies $S' \subseteq \mathcal{F}$;*

(*iii*) *for any $S, S' \in \mathcal{F}$, if $|S| > |S'|$, then there exists $j \in S \setminus S'$ such that $S' \cup \{j\} \in \mathcal{F}$.*

Elements of $\mathcal{F}$ are called independent sets, and the remaining sets of $[n]$ are called dependent sets.

Denote the set that contains the characteristic vectors of all independent sets of a matroid $M = ([n], \mathcal{F})$ as:

$$\mathcal{I} = \{y^S \in \{0,1\}^n : \ S \in \mathcal{F}\},$$

where y^S is the characteristic vector of set S such that $y^S_j = 1$ for $j \in S$, and $y^S_j = 0$, otherwise. The basic properties of $\mathcal{I}$ are shown as follows.

Lemma 1. *If $\mathcal{I}$ is the set of characteristic vectors of the independent sets of a matroid, then $\mathcal{I}$ satisfies the following statements:*

(*i*) *$0 \in \mathcal{I}$;*

(*ii*) *given $y, y' \in \{0,1\}^n$ and $y \leq y'$, if $y' \in \mathcal{I}$, then $y \in \mathcal{I}$;*

(*iii*) *if $y, y' \in \mathcal{I}$, and $\|y\|_1 < \|y'\|_1$, then there exists $j \in \{i \in [n] : \ y_i = 0, y'_i = 1\}$ such that $y + e^j \in \mathcal{I}$.*

Proof. It directly follows from Definition 5. ■

Denote by $\mathcal{F}_B$ and $\mathcal{I}_B$ the set of all maximal independent sets in $\mathcal{F}$ and the set of its corresponding characteristic vectors, respectively. The most fundamental matroid optimization problems are maximum-weight independent set problem $\max\{\beta y : \ y \in \mathcal{I}\}$ and minimum-weight basis problem $\min\{\beta y : \ y \in \mathcal{I}_B\}$.

It is well known that these two matroid optimization problems are polynomially solvable by a greedy algorithm [34, 35], which iteratively selects an element with the largest/smallest weight among the remaining elements. In particular, observe that the obtained greedy solutions are 2-optimal. We next provide a self-contained proof to show that if the lower-level feasible region contains the matroid structure, (BP) is equivalent to (BP$_k$) whenever $k \geq 2$.

Lemma 2. *Suppose $\mathcal{S}(x)$ is the characteristic vector set of all independent sets of a matroid for any $x \in \mathcal{X}$. Let $y, y' \in \mathcal{S}(x)$ and $y' \leq y$. If $y' \in \mathcal{R}_k(x)$ for some $k \geq 1$, then we have*

(*i*) *$\beta y' = \beta y$;*

(*ii*) *$y \in \mathcal{R}_k(x)$.*

Proof. (*i*) Since $\beta \geq 0$ and $y' \leq y$, we have $\beta y' \leq \beta y$. Suppose $\beta y' < \beta y$. Then there exists at least one j such that $y'_j = 0, y_j = 1$ and $\beta_j > 0$. By Lemma 1(*ii*), we have $y' + e^j \in \mathcal{S}(x)$. Then we have a contradiction by observing that $y' + e^j \in \mathcal{N}_k(y)$ for $k \geq 1$ and $\beta(y' + e^j) > \beta y'$.

(*ii*) According to the definition of $\mathcal{R}_k(x)$, we need to show that $\beta y \geq \beta \bar{y}$ for any $\bar{y} \in \mathcal{N}_k(y) \cap \mathcal{S}(x)$. Let $w = y - \bar{y}$, then $w \in \{-1, 0, 1\}^n$ and $\|w\|_1 \leq k$.

Construct $w' \in \{-1, 0, 1\}^n$ such that $w'_j = 0$ if $w_j = 1$ and $y'_j = 0$; otherwise $w'_j = w_j$. Thus, $\|w'\|_1 \leq \|w\|_1 \leq k$ and $y' - w' \in \{0,1\}^n$. Observe that $y' - w' \leq y - w = \bar{y}$. Thus, based on Lemma 1(*ii*), we have $y' - w' \in \mathcal{S}(x)$. Following our condition that $y' \in \mathcal{R}_k(x)$, we have $\beta y' \geq \beta(y' - w')$ as $y' - w' \in \mathcal{N}_k(y') \cap \mathcal{S}(x)$, which implies that $\beta y \geq \beta(y - w') \geq \beta(y - w)$ and the result follows.

■

It is worth noting that Lemma 2 implies that if $y' \in \mathcal{R}_k(x)$ and y' is not a maximal independent characteristic vector in $\mathcal{I}_B$, then there exists a maximal independent characteristic vector $y \in \mathcal{I}_B$ and $y \geq y'$ such that $y \in \mathcal{R}_k(x)$ and $\beta y = \beta y'$.

Theorem 5. *If $\mathcal{S}(x)$ is the characteristic vector set of all independent sets of a matroid for any $x \in \mathcal{X}$, then* (BP) $\equiv$ (BP$_k$) *for any integer $k \geq 2$.*

Proof. By Proposition 3, it is sufficient to show that for any leader's decision x the corresponding 2-optimal follower's response is also a globally optimal solution for the lower-level problem. That is, $\mathcal{R}_2(x) = \mathcal{R}(x)$. Since $\mathcal{R}(x) \subseteq \mathcal{R}_2(x)$, we only need to verify that $\mathcal{R}_2(x) \subseteq \mathcal{R}(x)$ for any $x \in \mathcal{X}$.

Based on Lemmas 2(i) and 2(ii), it suffices to focus on the maximal independent set in $\mathcal{R}_2(x)$ and $\mathcal{R}(x)$. Let y' and y^* be the characteristic vector of any maximal independent set in $\mathcal{R}_2(x)$ and $\mathcal{R}(x)$, respectively. It is clear that $\|y'\|_1 = \|y^*\|_1$ based on Lemma 1(iii). We next show that $\beta y' = \beta y^*$, which implies that $y' \in \mathcal{R}(x)$.

Without loss of generality, assume $\beta_1 \geq \beta_2 \geq \cdots \geq \beta_n$; also, suppose $y' = \sum_{j=1}^{t} e^{i'_j}$ and $y^* = \sum_{j=1}^{t} e^{i^*_j}$ such that $i'_1 < i'_2 < \cdots < i'_t$ and $i^*_1 < i^*_2 < \cdots < i^*_t$. Suppose there exists $j \in [t]$ such that $\beta_{i'_j} \neq \beta_{i^*_j}$, let $\ell = \min\{j \in [t] : \beta_{i'_j} \neq \beta_{i^*_j}\}$. Then we need to consider two cases: $\beta_{i'_\ell} < \beta_{i^*_\ell}$, and $\beta_{i'_\ell} > \beta_{i^*_\ell}$.

We first discuss the case that $\beta_{i'_\ell} < \beta_{i^*_\ell}$. Since $\sum_{j=1}^{\ell} e^{i^*_j} \in \mathcal{I}, \sum_{j=1}^{\ell-1} e^{i'_j} \in \mathcal{I}$ and $\| \sum_{j=1}^{\ell} e^{i^*_j} \|_1 > \| \sum_{j=1}^{\ell-1} e^{i'_j} \|_1$, based on Lemma 1($iii$), there exists $j^* \in \{1, \ldots, \ell\}$ such that $\sum_{j=1}^{\ell-1} e^{i'_j} + e^{i^*_{j^*}} \in \mathcal{I}$. Observe that $\| \sum_{j=1}^{\ell-1} e^{i'_j} + e^{i^*_{j'}} \|_1 < \|y'\|_1$. Based on Lemma 1($iii$), there exists j' such that $y'_{i'_{j'}} = 1$, $i'_{j'} \geq i'_\ell$, and $\bar{y} := y' - e^{i'_{j'}} + e^{i^*_{j^*}} \in \mathcal{I}$. Note that $\bar{y} \in \mathcal{N}_2(y')$ and $\beta \bar{y} > \beta y'$ as $\beta_{i^*_{j^*}} \geq \beta_{i^*_\ell} > \beta_{i'_\ell} \geq \beta_{i'_{j'}}$, which contradicts with the assumption that $y' \in \mathcal{R}_2(x)$. Thus, the first case (i.e., $\beta_{i'_\ell} < \beta_{i^*_\ell}$) is considered.

The proof for the case of $\beta_{i'_\ell} > \beta_{i^*_\ell}$ is similar and omitted for brevity. Therefore, we have that $\beta_{i'_j} = \beta_{i^*_j}$ for all $j \in [t]$ and the result immediately follows.

■

Remark 7. *We can further establish another related result that is similar to Theorem 5 for bilevel problems in which the follower solves a minimization problem of the form $\min_{y \in \{0,1\}^n} \{\beta y : y \in \mathcal{S}(x)\}$. Specifically, if $\mathcal{S}(x)$ is the characteristic vector set of all maximal independent sets of a matroid, then we have that* (BP) $\equiv$ (BP$_k$) *for any integer $k \geq 2$.*

To illustrate Theorem 5, we next outline several bilevel problems such that their lower-level problems are reducible to a matroid optimization problem. Specifically, we focus on problems with

$$\mathcal{S} = \{(x, y) \in \mathcal{X} \times \{0,1\}^n : \ y \in \mathcal{I}(x)\},$$

where $\mathcal{I}(x)$ is some characteristic vector set of a matroid for any $x \in \mathcal{X}$. We also consider two special cases of the above set given by:

- $\mathcal{S} = \{(x, y) \in \mathcal{X} \times \{0,1\}^n : \ y \in \mathcal{I}, \ x + y \leq 1\}$, and $\mathcal{I}$ is the characteristic vector set of a matroid; and
- $\mathcal{S} = \{(x, y) \in \mathcal{X} \times \{0,1\}^n : \ x + y \in \mathcal{I}\}$, and $\mathcal{I}$ is the characteristic vector set of a matroid.

There are a number of single-level combinatorial optimization problems that contain the matroid structure; we refer the reader to [35, 44, 57] and the references therein. In the literature, there are several bilevel generalizations of these problems with the following feasible sets at the lower level:

(i) The knapsack set with a cardinality constraint: Given a positive budget C, let

$$\mathcal{I} = \{y \in \{0,1\}^n : \sum_{j=1}^{n} y_j \leq C\},$$

which is the characteristic vector set of the uniform matroid; see, e.g., [57]. The bilevel knapsack problem has been extensively studied in recent years; see [10, 11, 14, 16].

(ii) The knapsack problem with multiple disjoint cardinality constraints: Given a partition of set $[n]$, $\{N_i\}_{i=1}^r$ and budgets C_i for each class i, let

$$\mathcal{I} = \{y \in \{0,1\}^n : \sum_{j \in N_i} y_j \leq C_i \ \forall i \in [r]\},$$

which is the characteristic vector set of the partition matroid; see, e.g., [2]. Some interesting results for the bilevel multidimensional knapsack problem are developed in [18].

(iii) The spanning tree set: Given an undirected graph $G = (N, E)$, let

$$\mathcal{I}_B = \{y \in \{0,1\}^n : \ G[y] \text{ is a spanning tree of graph } G\},$$

where $G[y] := G[E_y] = (N, E_y)$ is the subgraph induced by edges in $E_y = \{(i,j) \in E : y_{ij} = 1\}$. The spanning tree set contains all maximal independent sets of a tree matroid [2], and its bilevel versions are considered in [19, 48, 56].

(iv) Unit-time task scheduling problem: Given a set of unit-time tasks $N \in \{1, \ldots, n\}$ and their deadlines d_i for each task $i \in N$, let

$$\mathcal{I} = \{y \in \{0,1\}^n\} : \text{ there exists a schedule for tasks } N_y \text{ without delay}\},$$

where $N_y = \{i \in N : \ y_i = 1\}$. Several scheduling problems are shown to have matroid structures; see, e.g., [34, 35, 57]. Examples of bilevel scheduling problems can be found in [31, 38].

In Section 6.3, we provide a case study illustrating the use of (BP$_k$) to solve the bilevel minimum spanning tree problem (BMST) [19, 48, 56], in which the follower's optimization problem involves constructing a minimum spanning tree (MST) in a graph.

6 Computational Experiments

In this section, we report the results of our computational experiments with several classes of bilevel problems. We would like to point out that our main goal is not to solve general bilevel problems to optimality, but rather to evaluate (i) the quality of the proposed lower and upper bounds provided by (BP$_k$) as well as (ii) the performance of our MILP formulations. Therefore, we do not compare our approaches against specialized algorithms designed for solving particular classes of bilevel problems. Instead, the generic mixed integer bilevel solver MibS [52] and the SLR-based bounds are used as the main benchmarks.

This section is organized as follows. We first consider the knapsack interdiction problem (KIP) in Section 6.1. In Section 6.2, we consider the bilevel vertex cover problem. Section 6.3 illustrates our results developed in Section 5 by running the experiments on the bilevel minimum spanning tree problem. Finally, we note that our numerical experiments are performed using CPLEX 12.8 [26] on an Ubuntu 16.04 system with a 3.2GHz CPU and 19 GB of RAM.

6.1 Knapsack Interdiction Problem (KIP)

We consider the knapsack interdiction problem [16, 18] given as:

$$\min_{x \in \{0,1\}^n} \max_{y \in \{0,1\}^n} \sum_{j=1}^{n} \beta_j y_j \tag{KIP-a}$$

$$\text{s.t.} \quad \sum_{j=1}^{n} a_j^i x_j \le h_i \quad \forall i \in [m_\ell], \tag{KIP-b}$$

$$\sum_{j=1}^{n} g_j^i y_j \le d_i \quad \forall i \in [m_f], \tag{KIP-c}$$

$$x_j + y_j \le 1 \quad \forall j \in [n], \tag{KIP-d}$$

where $\beta_j, a_j^i, g_j^i, h_i$ and d_i are positive integers for all i and j; parameters m_ℓ and m_f denote the number of knapsack constraints at the upper and lower levels, respectively. We refer to the knapsack interdiction problem with k-optimal follower as (KIP$_k$). Since the leader aims to minimize their objective function, we note that the optimal objective function values of (KIP$_k$), η_k^*, provide a hierarchy of lower bounds for η^*. The objective function values $\hat{\eta}_k^*$ of bilevel feasible solutions that are constructed from (KIP$_k$), on the other hand, are natural upper bounds for η^*.

Experimental setup. To construct the test instances, we follow an approach similar to the one used in [11, 42]. The costs β_j as well as the weights a_j^i and g_j^i are generated randomly and independently using the discrete uniform distribution in interval $[0, 100]$. For each $n \in \{10, 20, 30, 40, 50\}$ and $r \in \{1, 2, \ldots, 10\}$, parameter d_i, $i \in [m_f]$, is set to $\lceil \frac{r}{11} \sum_{j=1}^{n} g_j^i \rceil$; parameter h_i, $i \in [m_\ell]$, is generated using the discrete uniform distribution in interval $[\frac{\sum_{i=1}^{m_f} d_i}{m_f} - 10, \frac{\sum_{i=1}^{m_f} d_i}{m_f} + 10]$. We construct 10 instances for each pair of n and r, and report the corresponding average performance.

In our computational experiments, we set the time limit for MibS and (KIP$_k$) to 600 seconds (10 minutes); all results are reported in seconds. Denote by η_M^* and $\hat{\eta}_M^*$ the best lower and upper bounds obtained by MibS at termination, respectively. Whenever CPLEX cannot solve (KIP$_k$) to optimality within the time limit, the best lower bound reported by CPLEX is referred to as η_k^*. Then the leader's feasible solution reported by CPLEX is used to derive the upper bound $\hat{\eta}_k^*$.

Next, we first discuss the sizes of the formulation (BP$_k$-DF) for (KIP$_k$) after the preprocessing steps based on Propositions 7 and 8 are applied. Then we first conduct the experiments on (KIP) instances with a single knapsack constraint at both levels, i.e., $m_\ell = m_f = 1$. Finally, we also explore the quality of our bounds for the instances of (KIP) with multiple knapsack constraints at the lower level.

Formulation size for (KIP$_k$). The average cardinality of $\mathcal{T}^k$ and the average number of constraints (BP$_k$-DF-c) for each n and m_f are shown in Table 1. Observe that despite the fact that the cardinality of $\mathcal{T}^k$ grows exponentially with respect to k, the number of constraints (BP$_k$-DF-c) after preprocessing has a rather moderate increase and is roughly a concave function with respect to m_f, n and k. Since the numbers of constraints (BP$_k$-DF-c) and variables z are equal, these results also indicate that we introduce a reasonably small number of additional binary variables z in our MILP reformulations for (KIP$_k$).

Results for (KIP) **with a single knapsack constraint at both levels ($m_\ell = m_f = 1$).** To evaluate MILP formulations (BP$_k$-DF) and (BP$_k$-Mix) for (KIP$_k$), we need first to select a tight

Table 1. The average cardinality of $\mathcal{T}^k$ and the average number of constraints (BP$_k$-DF-c) for the formulation (BP$_k$-DF) of (KIP$_k$) after the preprocessing procedure.

		$k=1$		$k=2$		$k=3$	
	n	$\lvert\mathcal{T}^k\rvert$	#(BP$_k$-DF-c)	$\lvert\mathcal{T}^k\rvert$	#(BP$_k$-DF-c)	$\lvert\mathcal{T}^k\rvert$	#(BP$_k$-DF-c)
$m_f = 1$	10	9	9	53	25	163	57
	20	20	18	208	61	1,324	137
	30	30	26	461	81	4,503	163
	40	40	33	812	88	10,694	174
	50	50	40	1,262	93	20,889	182
$m_f = 3$	10	10	28	54	137	174	357
	20	20	54	208	347	1,344	837
	30	30	78	460	467	4,520	986
	40	40	98	812	524	10,704	1,056
	50	50	118	1,262	549	20,856	1,087
$m_f = 5$	10	10	47	54	232	174	599
	20	20	90	208	576	1,348	1,404
	30	30	129	460	778	4,515	1,640
	40	40	165	811	868	10,702	1,750
	50	50	196	1,261	911	20,897	1,809

value of μ for constraint (BP$_k$-DF-c), which corresponds to (KIP-c). Recall from our discussion in Section 4 that μ can be set to some lower bound for the term $\sum_{j=1}^n a_j^2 y_j$. For the knapsack interdiction problem, we separate the leader's decisions into two possible groups:

(*i*) if $\sum_{j=1}^n a_j^2(1-x_j) \leq C_F$ for some leader's feasible decision x, then the lower-level problem has a k-optimal solution $y_j = 1 - x_j$ for all $j \in [n]$ and any $k \geq 1$. Therefore, the value of μ for such leader's decision x can be trivially achieved by solving:

$$\mu^1 = \min_{x \in [0,1]^n} \{\sum_{j=1}^n a_j^2(1-x_j) : \sum_{j=1}^n a_j^1 x_j \leq C_L\}.$$

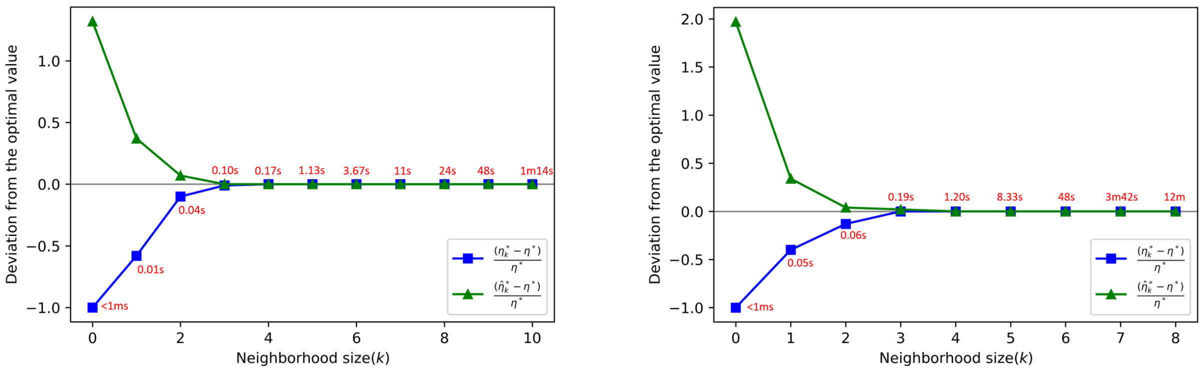
(*ii*) if $\sum_{j=1}^n a_j^2(1-x_j) > C_F$ for some leader's feasible decision x, then any follower's k-optimal solution y is maximal for $k \geq 1$, that is $\sum_{j=1}^n a_j^2 y_j \leq C_F$ and $\sum_{j=1}^n a_j^2 y_j + a_\ell^2 > C_F$ for any ℓ such that $y_\ell = 0$ and $x_\ell = 0$. Therefore, the value of μ for such leader's decision x can be set to:

$$\mu^2 = C_F - \max_{j \in [n]}\{a_j^2\}.$$

Therefore, we set $\mu = \min\{\lceil \mu^1 \rceil, \mu^2\}$ in our experiments.

Next, we evaluate the quality of our bounds, which are also depicted in Figure 1 for $n = 15$ and $n = 20$. The horizontal axis shows the value of k and the vertical axis indicates the deviation of the bounds from the true optimal objective function value η^* of the bilevel optimization problem. In Figure 1 we also indicate the solution time of the formulation (BP$_k$-DF) for each k, please see the labels in red.

One immediate observation is that the bounds provided by (KIP$_k$) for $k \geq 1$ are substantially better than the SLR-based bounds (i.e., $k = 0$). For example, in Figure 1(a) for $n = 15$ and $r = 4$,



(a) $n = 15, r = 4$.

(b) $n = 20, r = 4$.

Figure 1. The average deviations from the optimal value of (KIP$_k$) for different k. The solution time of formulation (BP$_k$-DF) is shown in red.

the gaps between the optimal objective function value and the SLR-based bounds are 197% and 100%, respectively. On the contrast, the bounds provided by (KIP$_1$) are only 40% and 34% away from the optimal values, respectively, and are computed within 0.02 seconds in total.

Consistent with Theorem 2, we can see from Figure 1 that the obtained lower bounds η_k^* improves monotonically with respect to k, but, of course, at increased computational expense. Moreover, these lower bounds converge rapidly and the optimal value η^* is achieved at $k = 3$ for both cases, with the corresponding (KIP$_3$) solved to optimality within a second. As for $\hat{\eta}_k^*$, although these upper bounds are not guaranteed to be monotone with respect to k (recall our discussion in Section 3.2), it is interesting to observe that for almost all of the tested instances, $\hat{\eta}_k^*$ provides better upper bound with the increase of k.

For both cases in Figure 1, we observe that the required computational time is small for sufficiently small values of k. For $n = 15$ and $r = 4$, the solver can efficiently tackle the formulation (BP$_k$-DF) for (KIP$_k$) with $k \geq 10$. However, we observed memory limitations in some cases for $n = 20$ and $r = 4$, when $k \geq 9$. We attribute it to the fact that the formulation size grows considerably with the increase of k (recall our discussion in Section 4.2). Therefore, we next focus on examining the performance of (KIP$_k$) for $k \leq 3$.

Specifically, in our next set of experiments, we compare the performances of the formulations for (KIP$_k$) against (SLR) and the general bilevel solver, MibS [52]. The average performances for the considered solution approaches are presented in Tables 2 and 3. In particular, for MibS and (SLR) (i.e., $k = 0$), we report the average runtime in seconds (column "Time"). For (KIP$_k$), $k \in \{1, 2, 3\}$, the average runtime in seconds for formulations (BP$_k$-DF) and (BP$_k$-Mix) are shown in columns "Time" and "ExtTime", respectively. For each solution approach, the ratios between the achieved bounds and $\hat{\eta}_M^*$ (i.e., $\frac{\eta_k^*}{\hat{\eta}_M^*}$ and $\frac{\hat{\eta}_k^*}{\hat{\eta}_M^*}, k \in \{0, 1, 2, 3\}$) are reported in columns "ObjL" and "ObjU", respectively.

MibS succeeds in solving the small instances to optimality within the time limit up to $n = 30$. We can also observe that the instances with either sufficiently small r or large r were easier to solve, e.g., when $r \geq 6$ in Table 3. This is because when the leader has a large interdiction budget (i.e., r is large), then the feasible region of the follower is typically small, which leads to fewer bilevel feasible solutions. Similarly, a scarce budget (i.e., small r) results in a small number of feasible decisions for the leader, and also makes the overall problem easier.

From Tables 2 and 3, we observe that (SLR) provides rather poor bounds for all considered instances. On the contrast, it is usually possible to obtain a bound equal to the optimal value by solving (KIP$_k$) for some small k. For the easy instances in Table 3, $k = 1$ suffices. For the hard instances in Table 2, the improvements provided by (KIP$_1$) are less significant; nevertheless, the optimal values are attained by η_3^* and $\hat{\eta}_3^*$ for more than half of the instances.

In Tables 2 and 3, we also highlight that the MILP formulations (BP$_k$-DF) and (BP$_k$-Mix) for all tested instances can be solved to optimality in under 20 seconds. We can also observe that there is no significant difference between these two formulations in terms of their running time performance in Tables 2 and 3.

Results for (KIP) **with multiple knapsack constraints at the lower level** ($m_f > 1$)**.** Since the formulation sizes of (BP$_k$-DF) and (BP$_k$-Mix) depend only on the size of the lower-level problem, we use $m_\ell = 1$ in our experiments, i.e., there is a single constraint at the upper-level problem. For simplicity, we set $\mu_i = 0$ for all $i \in [m_f]$ in (BP$_k$-DF) and (BP$_k$-Mix).

We report the average performance of (KIP) instances with $m_f \in \{3, 5\}$ in Table 4. In particular, similar to the discussions for (KIP) with a single knapsack constraint, larger values of r correspond to easier instances. Similar to the previous set of experiments, the bounds provided by (KIP$_k$) dramatically improve with the increase of k at the expense of more computational efforts. We observe that the optimal value of (KIP) can be achieved by (KIP$_3$) in most of our test instances.

For $k = 1$, the quality of achieved bounds notably outperforms those provided by (SLR). The formulations (BP$_k$-DF) and (BP$_k$-Mix) for (KIP$_1$) have fairly fast and stable solution times across all the instances, which implies that (KIP$_1$) is very scalable. Therefore, using (KIP$_1$) instead of (SLR) as the initial relaxation problem could be a promising approach for speeding up the performance of general branch-and-cut solvers.

Furthermore, Table 4 shows that (BP$_k$-Mix) significantly outperforms (BP$_k$-DF) (in contrast to our previous set of experiments for problems with a single constraint). This observation can be justified by a more complex structure of the lower-level problem for test instances with multiple constraints and highlights our theoretical results in Proposition 9. In particular, (BP$_k$-Mix) provides a speedup of at least one order of magnitude for $k \geq 2$ in Table 4; see e.g., the results for $m_f = 3$, $n = 40$ and $r = 2$. On the other hand, we recognize that our proposed formulations (BP$_k$-DF) and (BP$_k$-Mix) for (KIP$_k$) with $k \geq 2$ are sensitive to the size of instances and the value of parameter r. We attribute it to the fact the cardinality of $\mathcal{T}^k$ remains in order of $O(n^k)$ despite our preprocessing; see Table 1. Hence, we have the exponential growth of the number of constraints in (BP$_k$-DF-d) with the increase of k, which leads to the deterioration of the overall performance.

6.2 Bilevel Vertex Cover (BVC)

Given a graph $G = (N, E)$, the vertex cover problem is to find a subset of vertices whose total weight is as small as possible such that each vertex in the graph is either in this subset or connected to at least one vertex in this subset [21]. We consider its bilevel extension with interdiction constraints, referred to as the bilevel vertex cover (BVC) problem [5]. In BVC, the leader first removes vertices from N subject to some budgetary constraint, and then the follower solves the vertex cover problem.

		Mibs [52]		$k=0$			$k=1$				$k=2$				$k=3$			
n	r	ObjL	Time	ObjL	ObjU	Time	ObjL	ObjU	Time	ExtTime	ObjL	ObjU	Time	ExtTime	ObjL	ObjU	Time	ExtTime
10	1	1	< 0.01	0.02	1.75	< 0.01	0.35	1.75	< 0.01	< 0.01	0.99	1.01	< 0.01	< 0.01	1	1	< 0.01	< 0.01
10	2	1	< 0.01	0.01	1.79	< 0.01	0.23	1.4	0.01	0.01	0.78	1.17	0.01	0.01	0.99	1	0.01	0.01
10	3	1	< 0.01	0.01	1.92	< 0.01	0.42	1.3	0.01	0.01	0.94	1.11	0.03	0.04	1	1	0.05	0.08
10	4	1	< 0.01	0.01	2.71	< 0.01	0.34	1.27	0.01	0.01	0.94	1.04	0.03	0.02	1	1.01	0.04	0.07
10	5	1	< 0.01	0.03	8.32	< 0.01	0.49	1.44	0.01	0.01	0.95	1.06	0.02	0.02	1	1	0.04	0.04
20	1	1	< 0.01	0	1.58	< 0.01	0.14	1.39	0.01	0.01	0.75	1.1	0.03	0.04	0.97	1	0.06	0.06
20	2	1	1.5	0	1.8	< 0.01	0.14	1.28	0.01	0.01	0.71	1.16	0.06	0.09	0.99	1.03	0.15	0.15
20	3	1	9.6	0	1.98	< 0.01	0.19	1.27	0.02	0.01	0.78	1.11	0.05	0.08	0.98	1.02	0.14	0.21
20	4	1	15	0	2.6	< 0.01	0.34	1.46	0.02	0.01	0.84	1.13	0.05	0.09	0.99	1	0.12	0.14
20	5	1	7.7	0	3.43	< 0.01	0.63	1.35	0.03	0.02	0.89	1.03	0.05	0.07	1	1.01	0.12	0.15
30	1	1	2.6	0	1.53	< 0.01	0.06	1.38	0.01	0.01	0.7	1.09	0.09	0.11	0.99	1	0.33	0.29
30	2	0.97	445.3	0	1.82	< 0.01	0.1	1.35	0.02	0.01	0.73	1.12	0.1	0.11	0.97	1.01	0.58	0.75
30	3	0.85	529.9	0	1.93	< 0.01	0.19	1.3	0.02	0.01	0.73	1.1	0.1	0.1	0.97	1.02	0.55	0.77
30	4	0.81	600	0	2.54	< 0.01	0.31	1.5	0.03	0.01	0.8	1.16	0.1	0.14	0.99	1.01	0.45	0.56
30	5	0.91	402.4	0	3.71	< 0.01	0.66	1.49	0.05	0.02	0.99	1.02	0.13	0.1	1	1	0.38	0.25
40	1	0.97	178	0	1.59	< 0.01	0.07	1.33	0.02	0.01	0.55	1.19	0.15	0.13	0.96	1.01	1.57	1.53
40	2	0.7	600	0	1.8	< 0.01	0.11	1.44	0.02	0.01	0.65	1.12	0.16	0.17	0.96	1.03	1.69	2.53
40	3	0.57	600	0	2	< 0.01	0.2	1.35	0.03	0.02	0.69	1.11	0.21	0.21	0.97	1.02	2.2	2.79
40	4	0.54	600	0	2.49	< 0.01	0.34	1.44	0.04	0.02	0.82	1.17	0.21	0.25	0.99	1	1.51	2.27
40	5	0.6	600	0	3.21	< 0.01	0.62	1.51	0.07	0.04	0.92	1.09	0.22	0.17	0.99	1	1.64	1.44
50	1	0.9	426.5	0	1.56	< 0.01	0.06	1.38	0.03	0.02	0.57	1.24	0.25	0.23	0.98	1.01	6.82	7.75
50	2	0.69	600	0	1.79	< 0.01	0.09	1.37	0.03	0.02	0.69	1.08	0.3	0.33	0.98	1.01	8.84	11.38
50	3	0.49	600	0	2.05	< 0.01	0.18	1.3	0.03	0.02	0.64	1.07	0.3	0.38	0.98	1.01	8.09	14.57
50	4	0.41	600	0	2.53	< 0.01	0.36	1.4	0.07	0.04	0.77	1.12	0.32	0.37	0.97	0.98	6.26	8.2
50	5	0.46	600	0	3.31	< 0.01	0.64	1.49	0.15	0.05	0.92	1.05	0.35	0.26	0.99	0.99	4.27	4.48

Table 2. Results for the instances of (KIP) with $r \in \{1, 2, \ldots, 5\}$. For MibS and (SLR) (i.e., $k = 0$), we report the average runtime in seconds (column "Time"). For (KIP_k), $k \in \{1, 2, 3\}$, the average runtime in seconds for formulations (BP_k-DF) and (BP_k-Mix) are shown in columns "Time" and "ExtTime," respectively. For each solution approach, the ratios between the achieved bounds and the best upper bound, $\hat{\eta}^*_M$, reported by MibS (i.e., $\frac{\eta^*_k}{\hat{\eta}^*_M}$ and $\frac{\hat{\eta}^*_k}{\hat{\eta}^*_M}, k \in \{M, 0, 1, 2, 3\}$) are shown in the columns "ObjL" and "ObjU," respectively.

		Mibs [52]		$k=0$			$k=1$				$k=2$				$k=3$			
n	r	ObjL	Time	ObjL	ObjU	Time	ObjL	ObjU	Time	ExtTime	ObjL	ObjU	Time	ExtTime	ObjL	ObjU	Time	ExtTime
10	6	1	0.02	0.12	54.65	< 0.01	0.75	1.31	0.01	0.01	0.99	1	0.01	0.01	1	1	0.01	0.01
10	7	1	0.02	0.22	81.65	< 0.01	0.95	1.04	0.01	0.01	0.99	1	0.01	0.01	1	1	0.02	0.02
10	8	1	0.02	0.02	9.35	< 0.01	1	1	0	< 0.01	1	1	< 0.01	< 0.01	1	1	0.01	0.01
10	9	1	0.01	0.23	97.48	< 0.01	1	1	0	< 0.01	1	1	< 0.01	< 0.01	1	1	0.01	0.01
10	10	1	0.01	0.26	126.21	< 0.01	1	1	0	< 0.01	1	1	< 0.01	< 0.01	1	1	0.01	0.01
20	6	1	2.02	0.01	5.82	< 0.01	0.97	1.15	0.03	0.02	1	1	0.05	0.05	1	1	0.08	0.1
20	7	1	0.98	0.01	10.1	< 0.01	0.99	1.06	0.02	0.01	1	1	0.03	0.03	1	1	0.05	0.07
20	8	1	0.32	0.03	28	< 0.01	1	1	0.02	0.01	1	1	0.02	0.01	1	1	0.04	0.07
20	9	1	0.04	0.15	157.31	< 0.01	1	1	0.01	0.01	1	1	0.01	0.01	1	1	0.02	0.05
20	10	1	0.02	0.47	435.3	< 0.01	1	1	0	< 0.01	1	1	0.01	0.01	1	1	0.01	0.03
30	6	0.99	282.74	0.01	6.4	< 0.01	0.99	1.23	0.06	0.04	1	1	0.1	0.1	1	1	0.25	0.18
30	7	1	19.62	0.01	9.03	< 0.01	0.99	1.08	0.06	0.02	1	1	0.07	0.05	1	1	0.15	0.16
30	8	1	3.19	0.02	33.22	< 0.01	1	1	0.03	0.01	1	1	0.03	0.03	1	1	0.08	0.1
30	9	1	0.24	0.15	224.45	< 0.01	1	1	0.01	0.01	1	1	0.02	0.02	1	1	0.06	0.1
30	10	1	0.02	0.18	263.71	< 0.01	1	1	0.01	0.01	1	1	0.01	0.01	1	1	0.05	0.07
40	6	0.76	542.45	0	5.25	0.01	0.97	1.17	0.17	0.04	0.99	0.99	0.17	0.12	0.99	0.99	0.77	0.74
40	7	0.98	116.57	0.01	14.58	< 0.01	1	1	0.06	0.03	1	1	0.09	0.08	1	1	0.34	0.54
40	8	0.98	129.03	0.01	20.78	0.01	1	1	0.06	0.02	1	1	0.05	0.06	1	1	0.21	0.49
40	9	1	2.33	0.12	261.7	< 0.01	1	0.96	0.04	0.01	1	1	0.02	0.03	1	1	0.13	0.37
40	10	1	0.2	0.39	804.49	< 0.01	1	1	0.01	0.01	1	1	0.01	0.02	1	1	0.1	0.26
50	6	0.57	600	0	5.37	0.01	0.96	1.03	0.29	0.07	0.99	0.99	0.29	0.12	0.99	0.99	3.26	1.89
50	7	0.78	545.39	0.01	11.74	< 0.01	0.99	0.99	0.18	0.04	0.99	0.99	0.2	0.1	0.99	0.99	1.99	1.52
50	8	0.97	382.01	0.01	16.94	0.01	1	1	0.1	0.02	1	1	0.06	0.09	1	1	0.6	1.23
50	9	1	5.96	0.03	78.9	< 0.01	1	1	0.02	0.01	1	1	0.03	0.05	1	1	0.32	0.8
50	10	1	0.07	0.33	804.69	< 0.01	1	1	0.01	0.01	1	1	0.02	0.03	1	1	0.21	0.5

Table 3. Results for the instances of (KIP) with $r \in \{6, 7, \ldots, 10\}$. For MibS and (SLR) (i.e., $k = 0$), we report the average runtime in seconds (column "Time"). For (KIP$_k$), $k \in \{1, 2, 3\}$, the average runtime in seconds for formulations (BP$_k$-DF) and (BP$_k$-Mix) are shown in columns "Time" and "ExtTime," respectively. For each solution approach, the ratios between the achieved bounds and the best upper bound, $\hat{\eta}^*_M$, reported by MibS (i.e., $\frac{\eta^*_k}{\hat{\eta}^*_M}$ and $\frac{\hat{\eta}^*_k}{\hat{\eta}^*_M}, k \in \{M, 0, 1, 2, 3\}$) are shown in the columns "ObjL" and "ObjU," respectively.

			Mibs [52]		$k=0$			$k=1$				$k=2$				$k=3$			
	n	r	ObjL	Time	ObjL	ObjU	Time	ObjL	ObjU	Time	ExtTime	ObjL	ObjU	Time	ExtTime	ObjL	ObjU	Time	ExtTime
	20	3	1	146.59	0.01	2.27	< 0.01	0.21	1.54	0.08	0.04	0.62	1.3	1.12	0.79	0.95	1.02	8.23	5.79
	20	4	1	162.38	0	2.49	< 0.01	0.37	1.6	0.13	0.05	0.79	1.22	1.64	0.66	0.99	1	16.81	3.76
	20	5	1	41.21	0	2.62	0.01	0.41	1.37	0.12	0.04	0.79	1.21	2.34	0.58	1	1	15.31	4.36
	20	6	1	20.52	0	3.45	< 0.01	0.75	1.43	0.13	0.05	0.98	1.06	2.03	0.35	1	1	9.48	1.83
	20	7	1	0.62	0.01	9.31	< 0.01	1	1.08	0.05	0.02	1	1	0.15	0.09	1	1	0.42	0.32
	30	3	0.61	600	0	1.85	0.01	0.19	1.43	0.12	0.07	0.52	1.2	36.65	4.33	0.92	1.05	501.1	71.14
	30	4	0.71	600	0	2.51	0.01	0.3	1.66	0.14	0.08	0.64	1.31	68.35	3.98	0.99	1.01	338.76	50.08
$m_f = 3$	30	5	0.78	564.6	0	3.41	0.01	0.59	1.7	0.19	0.09	0.91	1.13	60.18	1.67	1	1	239.47	19.13
	30	6	0.94	461.67	0	5.32	0.01	0.88	1.25	0.28	0.08	1	1	19.48	0.42	1	1	72.01	4.51
	30	7	1	138.96	0.02	22.67	0.01	1	1	0.14	0.07	1	1	1.15	0.1	1	1	7.63	0.7
	40	3	0.39	600	0	1.89	0.01	0.18	1.38	0.16	0.1	0.5	1.24	241.89	13.47	0.91	1.01	600	327.48
	40	4	0.45	600	0	2.4	0.01	0.31	1.51	0.28	0.12	0.64	1.28	365.95	12.06	0.97	1.01	600	240.46
	40	5	0.53	600	0	2.82	0.01	0.47	1.49	0.32	0.12	0.78	1.17	436.04	7.09	0.99	1	600	116.86
	40	6	0.71	560.78	0	5.79	0.01	0.93	1.21	0.6	0.11	1	1	86.17	1.45	1	1	231.03	17.82
	40	7	0.83	544.32	0	8.33	0.01	1	1	0.36	0.1	1	1	7.95	0.19	1	1	83.18	2.7
	20	3	1	213.32	0	1.83	0.01	0.19	1.48	0.1	0.07	0.53	1.32	2.03	2.51	0.93	1.08	20.75	32.92
	20	4	1	144.56	0	2.31	0.01	0.31	1.46	0.11	0.07	0.68	1.32	7.13	4.65	0.99	1	80.87	39.33
	20	5	1	60.71	0	2.97	< 0.01	0.45	1.51	0.12	0.08	0.84	1.14	11.56	2.2	1	1.01	53.88	16.48
	20	6	1	7.62	0	4.71	< 0.01	0.85	1.46	0.14	0.08	1	1	1.65	0.4	1	1	6.42	2.72
	20	7	1	0.75	0	24.21	< 0.01	0.96	1.11	0.06	0.04	1	1	0.41	0.12	1	1	1.07	1.21
	30	3	0.55	600	0	1.72	0.01	0.19	1.49	0.13	0.1	0.48	1.27	88.75	22.5	0.91	1.02	600	287.14
	30	4	0.63	600	0	2.39	0.01	0.31	1.53	0.19	0.13	0.62	1.26	323.6	14.51	0.97	1.01	574.77	238.13
$m_f = 5$	30	5	0.72	600	0	2.96	0.01	0.51	1.52	0.36	0.12	0.86	1.2	471.29	13.9	0.99	1	583.49	95.97
	30	6	0.89	496.44	0	4.37	0.01	0.87	1.28	0.49	0.12	0.98	1.03	169.73	1.75	1	1	248.91	20.62
	30	7	0.99	210.67	0	8.88	0.01	0.97	1.02	0.32	0.11	1	1	11.59	0.37	1	1	70.4	5.06
	40	3	0.33	600	0	1.78	0.02	0.17	1.4	0.4	0.13	0.38	1.22	574.95	85.51	0.61	1.02	600	600
	40	4	0.33	600	0	2.24	0.01	0.34	1.65	0.81	0.13	0.61	1.27	600	83.92	0.85	0.99	600	541.18
	40	5	0.45	600	0	2.77	0.01	0.53	1.49	1.27	0.13	0.79	1.14	569.92	25.57	0.96	0.97	600	421.23
	40	6	0.6	600	0	4.44	0.01	0.91	1.34	1.27	0.12	0.99	0.99	173.04	2.26	0.99	0.99	580.04	65.21
	40	7	0.8	581.89	0	7.12	0.01	1	1	1.17	0.12	1	1	37.96	0.48	1	1	380.96	19.53

Table 4. Results for the instances of (KIP) for $m_f \in \{3, 5\}$. For MibS and (SLR) (i.e., $k = 0$), we report the average runtime in seconds (column "Time"). For (KIP$_k$), $k \in \{1, 2, 3\}$, the average runtime in seconds for formulations (BP$_k$-DF) and (BP$_k$-Mix) are shown in columns "Time" and "ExtTime," respectively. For each solution approach, the ratios between the achieved bounds and the best upper bound, $\hat{\eta}^*_M$, reported by MibS (i.e., $\frac{\eta^*_k}{\hat{\eta}^*_M}$ and $\frac{\hat{\eta}^*_k}{\hat{\eta}^*_M}, k \in \{M, 0, 1, 2, 3\}$) are shown in the columns "ObjL" and "ObjU," respectively.

Formally, the BVC problem is stated as:

$$\begin{aligned} \max_{x,y}\ & \sum_{j=1}^{n} \alpha_j y_j \\ \text{s.t.}\ & \sum_{j=1}^{n} x_j \le b, x \in \{0,1\}^n, \\ & y \in \arg\min_{\bar{y}} \{\sum_{j=1}^{n} \beta_j \bar{y}_j : \ \bar{y} \in \mathcal{S}_{\mathrm{VC}}(x)\}, \end{aligned} \tag{BVC}$$

where $n = |N|$, and

$$\mathcal{S}_{\mathrm{VC}}(x) = \Big\{ y \in \{0,1\}^n : \ x_j + y_j \le 1 \ \forall j \in N, \sum_{i \in N_j} y_i \ge 1 \ \forall j \in N \Big\},$$

where $N_j := \{i \in N : (i,j) \in E\} \cup \{j\}$ is the extended neighborhood of $j \in N$, that includes j itself. If $\alpha_j = \beta_j$ for all $j \in N$, then we say that the BVC problem is symmetric, and can be referred to as the vertex cover interdiction problem; otherwise, the BVC problem is asymmetric. Note that in the formulations (BP_k-DF) and (BP_k-Mix), a valid μ_j is the upper bound for the term $\sum_{i \in N_j} y_j$ for each $j \in N$. Thus, we trivially set $\mu_j = |N_j|$ in our experiments.

Experimental setup. Let (BVC_k) be the bilevel vertex cover problem with k-optimal follower. In the BVC problem, the leader maximizes her objective function, thus η_k^* and $\hat{\eta}_k^*$ provide valid upper and lower bounds for η^*, respectively. Denote by η_M^* and $\hat{\eta}_M^*$ the best upper and lower bounds reported by MibS at termination, respectively. In our experiments, we use $\hat{\eta}_M^*$ as the benchmark.

We randomly construct graphs with SNAP [36], in which the degree of each vertex is no less than $\lceil \frac{n}{2} \rceil$ with respect to the vertex size $n \in \{10, 15, 20, \ldots, 45, 50\}$. We refer to the minimum vertex degree in the graph as "deg". For each pair of the considered classes (i.e., the number of vertices n along with the specific value of deg and b), we report the average performance over 10 randomly constructed instances. The time limit for MibS is set to 30 minutes in order to avoid the out-of-memory error. The time limit for (BVC_k) is also set to 30 minutes.

Results and discussions. In Figure 2, we first depict the deviation of our proposed bounds from the true optimal objective function value η^* of (BVC) for different values of k. We evaluate the average performance of 10 randomly constructed symmetric (BVC) instances for each $n \in \{15, 20\}$. The solution time is indicated as the labels in red. Similar to Figure 1 for (KIP), we observe that the bounds provided by (BVC_k), $k \ge 1$, significantly outperform the quality of SLR-based bounds. Furthermore, our bounds converge rapidly to the optimal value η^* for relatively small values of k.

The numerical results for MibS and $\mathrm{BVC}_k, k \in \{0, 1, 2, 3\}$, are reported for both symmetric and asymmetric objectives in Table 5. In particular, for MibS and (SLR) (i.e., $k = 0$), we report the average runtime in seconds (column "Time"). For $\mathrm{BVC}_k, k \in \{1, 2, 3\}$, the average runtime in seconds for formulations (BP_k-DF) and (BP_k-Mix) are shown in columns "Time" and "ExtTime", respectively. For each solution approach, the ratios between the achieved bounds and $\hat{\eta}_M^*$ (i.e., $\frac{\hat{\eta}_k^*}{\hat{\eta}_M^*}$ and $\frac{\eta_k^*}{\hat{\eta}_M^*}, k \in \{M, 0, 1, 2, 3\}$) are reported in columns "ObjL" and "ObjU", respectively.

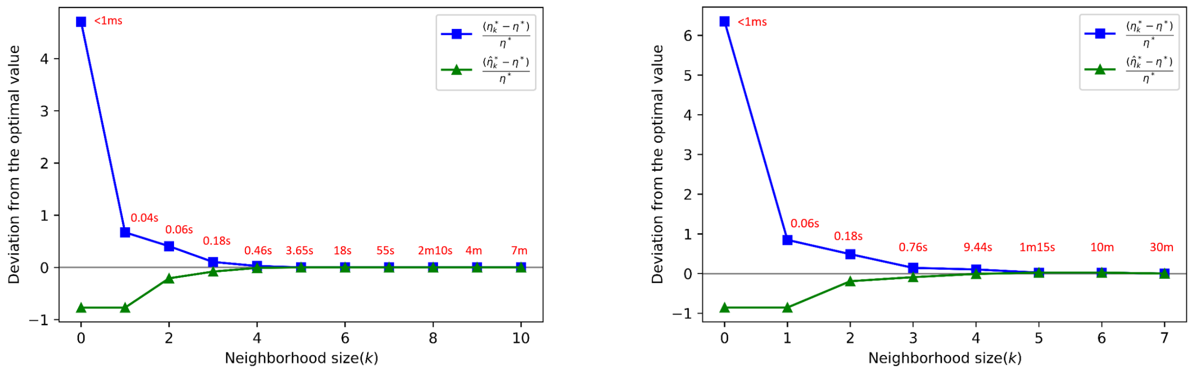
In Table 5, we observe that MibS can only handle the smallest instances within the computational limits. For instances with $n = 50$, deg$= b = 25$ and symmetric objectives, the average ratio between the best upper and lower bounds (i.e., $\frac{\eta_M^*}{\hat{\eta}_M^*}$) obtained by MibS at termination is 32.19, and

			Mibs [52]		$k=0$			$k=1$				$k=2$				$k=3$			
n	deg	b	ObjU	Time	ObjL	ObjU	Time	ObjL	ObjU	Time	ExtTime	ObjL	ObjU	Time	ExtTime	ObjL	ObjU	Time	ExtTime
Symmetric Objective																			
10	5	5	1	0.57	0.27	4.00	< 0.01	0.27	1.38	0.03	0.03	0.89	1.06	0.04	0.04	0.95	1.03	0.06	0.06
15	8	8	1	25.66	0.23	5.71	< 0.01	0.23	1.67	0.04	0.04	0.79	1.41	0.06	0.06	0.92	1.10	0.18	0.24
20	10	10	1	984.85	0.14	7.19	< 0.01	0.14	1.81	0.06	0.06	0.79	1.46	0.18	0.18	0.89	1.12	0.76	1.32
25	13	13	6.96	1800	0.14	9.47	< 0.01	0.14	2.30	0.06	0.06	0.79	1.66	0.35	0.38	1.08	1.37	2.35	4.52
30	15	15	10.75	1800	0.17	12.90	< 0.01	0.17	2.92	0.10	0.10	0.73	2.05	1.02	0.86	1.15	1.76	9.10	18.67
35	18	18	16.40	1800	0.16	18.49	< 0.01	0.16	3.77	0.25	0.17	0.97	2.56	2.32	2.20	1.17	2.03	43.49	71.31
40	20	20	20.33	1800	0.25	22.26	< 0.01	0.25	4.31	0.57	0.38	0.87	2.81	6.56	6.68	1.28	2.37	197.25	479.93
45	23	23	27.28	1800	0.18	29.08	< 0.01	0.18	5.31	0.75	0.57	0.98	3.39	12.91	14.10	1.70	2.83	737.55	1277.52
50	25	25	32.19	1800	0.17	33.86	< 0.01	0.17	5.86	1.13	0.98	0.83	3.94	31.58	36.76	1.63	2.94	1800	1800
Asymmetric Objective																			
10	5	5	1	0.54	0.50	3.12	< 0.01	0.50	1.07	0.03	0.03	0.78	1.04	0.04	0.04	0.98	1	0.07	0.06
15	8	8	1	24.75	0.37	3.59	< 0.01	0.37	1.11	0.04	0.04	0.53	1.06	0.07	0.07	0.69	1	0.13	0.16
20	10	10	1	925.88	0.46	4.35	< 0.01	0.46	1.19	0.05	0.05	0.62	1.06	0.13	0.13	0.66	1.02	0.61	0.80
25	13	13	4.30	1800	0.43	5.94	< 0.01	0.43	1.37	0.07	0.07	0.70	1.28	0.29	0.29	0.79	1.24	1.38	2.35
30	15	15	5.66	1800	0.54	6.78	< 0.01	0.54	1.53	0.1	0.09	0.79	1.40	0.68	0.58	0.89	1.29	4.98	10.27
35	18	18	7.27	1800	0.44	8.25	< 0.01	0.48	1.68	0.27	0.17	0.72	1.54	1.16	0.99	0.75	1.44	12.76	25.78
40	20	20	8.58	1800	0.55	9.39	< 0.01	0.55	1.84	0.55	0.34	0.68	1.63	2.98	2.80	0.81	1.47	72.02	112.00
45	23	23	9.79	1800	0.66	10.51	< 0.01	0.66	1.85	0.82	0.67	0.68	1.71	4.59	4.42	0.89	1.56	204.57	228.72
50	25	25	10.66	1800	0.61	11.26	< 0.01	0.61	1.97	1.07	0.93	0.62	1.86	8.18	8.10	0.85	1.70	684.02	656.31

Table 5. Results for the instances of (BVC). For MibS and (SLR) (i.e., $k = 0$), we report the average runtime in seconds (column "Time"). For (BVC_k), $k \in \{1, 2, 3\}$, the average runtime in seconds for formulations (BP_k-DF) and (BP_k-Mix) are shown in columns "Time" and "ExtTime," respectively. For each solution approach, the ratios between the achieved bounds and $\hat{\eta}^*_M$ (i.e., $\frac{\hat{\eta}^*_k}{\hat{\eta}^*_M}$ and $\frac{\eta^*_k}{\hat{\eta}^*_M}, k \in \{M, 0, 1, 2, 3\}$) are reported in columns "ObjL" and "ObjU," respectively.

			Mibs [52]	$k=0$			$k=1$				$k=2$				$k=3$			
n	deg	b	Time	ObjL	ObjU	Time	ObjL	ObjU	Time	Ext Time	ObjL	ObjU	Time	Ext Time	ObjL	ObjU	Time	Ext Time
Symmetric Objective																		
20	10	10	984.85	0.14	7.19	< 0.01	0.14	1.81	0.06	0.06	0.79	1.46	0.18	0.18	0.89	1.12	0.76	1.33
20	11	10	1245.49	0.15	7.09	< 0.01	0.15	1.72	0.04	0.04	0.70	1.32	0.16	0.17	0.87	1.11	0.66	1.09
20	12	10	1430.29	0.11	7.16	< 0.01	0.11	1.70	0.05	0.05	0.81	1.28	0.15	0.15	0.80	1.07	0.56	0.98
20	13	10	1543.51	0.12	7.57	< 0.01	0.12	1.67	0.05	0.05	0.72	1.32	0.11	0.11	0.94	1.10	0.48	0.89
20	14	10	1534.97	0.09	7.28	< 0.01	0.09	1.46	0.05	0.05	0.68	1.21	0.10	0.10	0.89	1.02	0.37	0.65
20	15	10	1529.98	0.08	7.53	< 0.01	0.08	1.35	0.04	0.04	0.74	1.18	0.09	0.09	0.98	1.01	0.26	0.46
Asymmetric Objective																		
20	10	10	925.88	0.46	4.35	< 0.01	0.46	1.19	0.05	0.05	0.62	1.06	0.13	0.13	0.66	1.02	0.60	0.79
20	11	10	1200.43	0.42	4.77	< 0.01	0.42	1.16	0.05	0.05	0.55	1.09	0.11	0.11	0.90	1.01	0.52	0.71
20	12	10	1383.42	0.43	5.18	< 0.01	0.43	1.23	0.05	0.05	0.70	1.19	0.10	0.10	0.95	1.04	0.40	0.62
20	13	10	1489.77	0.42	5.23	< 0.01	0.42	1.10	0.05	0.05	0.54	1.03	0.09	0.09	0.96	1	0.32	0.51
20	14	10	1463.03	0.38	5.81	< 0.01	0.38	1.15	0.05	0.05	0.55	1.06	0.1	0.10	0.92	1	0.27	0.42
20	15	10	1445.94	0.27	6.25	< 0.01	0.27	1.04	0.04	0.04	0.65	1.02	0.09	0.10	0.94	1.01	0.23	0.29

Table 6. Results for the instances of (BVC_k) with different minimum vertex degrees for fixed $n = 20$ and $b = 10$. For MibS and (SLR) (i.e., $k = 0$), we report the average runtime in seconds (column "Time"). For (BVC_k), $k \in \{1, 2, 3\}$, the average runtime in seconds for formulations (BP_k-DF) and (BP_k-Mix) are shown in columns "Time" and "ExtTime," respectively. For each solution approach, the ratios between the achieved bounds and $\hat{\eta}^*_M$ (i.e., $\frac{\hat{\eta}^*_k}{\hat{\eta}^*_M}$ and $\frac{\eta^*_k}{\hat{\eta}^*_M}, k \in \{M, 0, 1, 2, 3\}$) are reported in columns "ObjL" and "ObjU," respectively.



(a) $n = 15$, deg$= 8, b = 8$.

(b) $n = 20$, deg$= 10, b = 10$.

Figure 2. The average deviations from the optimal value of (BVC_k) for different k. The solution time of formulation (BP_k-DF) is shown in red.

the average ratio between the upper bound obtained by (SLR) and $\hat{\eta}_M^*$ (i.e., $\frac{\eta_0^*}{\hat{\eta}_M^*}$) is 33.86. The difference between these two ratios is less than 2, which suggests that there is no substantial progress achieved by MibS to reduce the optimality gap after extensive branching and cut generation.

On the other hand, all instances (except for $n = 50$, deg$= b = 25$ and symmetric objectives) are solved to optimality when using formulations (BP_k-DF) and (BP_k-Mix) for (BVC_k) with $k \in \{1, 2, 3\}$. The average ratio between the obtained upper bound and $\hat{\eta}_M^*$ is reduced from 33.86 to 5.86 for $k = 1$, and further to 2.94 for $k = 3$. As for the lower bound, the average ratio between the obtained lower bound of (BVC_1) and $\hat{\eta}_M^*$ does not improve over that of (SLR). It is not a surprising result as we can easily verify that (SLR) and (BVC_1) have identical optimal values of $x = 0$. However, when $k = 2$, this ratio increases from 0.17 to 0.83. Moreover, the ratio is further improved to 1.63 when $k = 3$, which implies that the lower bound obtained by (BVC_3) is better than the best lower bound reported by MibS. For the instances with asymmetric objectives, similar improvements can be observed for (BVC_k). Hence, we conclude that the bounds by (BVC_k) are superior to the SLR-based bounds, and the overall performances of (BVC_k) is better than MibS for larger instances.

With respect to the solution times for (BVC_k), one immediate observation is that the instances with asymmetric objectives are much easier for the solver. We also observe that the average computational times are very similar for the formulations (BP_k-DF) and (BP_k-Mix), for $k \in \{1, 2\}$. Also, the runtime of formulation (BP_k-DF) is relatively better than that of formulation (BP_k-Mix) for $k = 3$. Recall Proposition 9 that the LP relaxation of (BP_k-Mix) is stronger than that of (BP_k-DF). However, the strengthened constraints in (BP_k-Mix) are denser, which may result in more computational efforts required for solving its LP relaxation.

In our next set of experiments, we explore how the quality of bounds obtained by (BVC_k) depends on the minimum vertex degree in the graph. The corresponding results are presented in Table 6. Observe that the performance of MibS deteriorates when the minimum vertex degree increases. It is intuitive given that the graph density increases for larger values of the minimum vertex degree. On the other hand, we note that the quality of bounds obtained by $\text{BVC}_k, k \geq 1$, improves for larger minimum vertex degrees and with smaller computational times. These comparisons illustrate that our proposed bounding approach is capable of exploiting the problem inherent structure, which further supports our earlier results in Section 6.1.

6.3 Bilevel Minimum Spanning Tree (BMST)

In this section, we study the bilevel minimum spanning tree problem (BMST) to illustrate our results in Section 5. Two single-level formulations for BMST are developed based on (BP_k) for $k = 2$ in Section 6.3.1. The computational experiments are then conducted in Section 6.3.2.

In particular, we focus on the variant of the BMST problem considered in [48], which is described as follows: given an undirected graph $G = (N, E)$, the leader and the follower construct a spanning tree of graph G in a hierarchical manner. The leader first selects a subset of edges from among those in $E_L \subseteq E$. The follower then selects a set of edges from E that complete a spanning tree, according to their own objective function. Formally, the BMST problem can be stated as:

$$\begin{aligned} \eta^* = & \min_{x,y} \alpha(x+y) \\ & \text{s.t. } x_{ij} \in \{0,1\} \quad \forall (i,j) \in E_L, \\ & \quad y \in \arg\min_{\bar{y}} \{\beta \bar{y} : \bar{y} \in \mathcal{S}_{\mathrm{MST}}(x)\}, \end{aligned} \tag{BMST}$$

where we let $m_0 = |E_L|$, $m = |E|$, and

$$\mathcal{S}_{\mathrm{MST}}(x) = \begin{Bmatrix} y \in \{0,1\}^m : x_{ij} + y_{ij} \leq 1 \quad \forall (i,j) \in E_L, \\ \qquad G[x+y] \text{ is a spanning tree of graph } G \end{Bmatrix},$$

where $G[x] := G[E_x] = (N, E_x)$ is the subgraph induced by edges in $E_x = \{(i,j) \in E : \ x_{ij} = 1\}$ for $x \in \{0,1\}^m$. We also define directed graph $\widetilde{G}[x] = (N, \mathcal{A}_{[x]})$, where $\mathcal{A}_{[x]} = \{(i,j), (j,i) : x_{ij} = 1\}$ for any $x \in \{0,1\}^m$.

We refer to the bilevel minimum spanning tree problem with a k-optimal follower as (BMST_k). Following our discussion in Section 5, the follower's feasible region $\mathcal{S}_{\mathrm{MST}}(x)$ is the set of characteristic vectors of all maximal independent sets of a matroid. Therefore, by Theorem 5 we have (BMST) $\equiv$ (BMST_k) for $k \geq 2$. Next, we derive single-level MILP formulations for (BMST) based on (BMST_2) and its particular structure.

6.3.1 MILP Formulations

Based on Proposition 1, we have $\mathcal{T}^2 = \{w = (e_{i_0 j_0} - e_{i_1 j_1}) : \ \beta_{i_0 j_0} < \beta_{i_1 j_1}\}$. We then explore the optimality conditions for the lower-level problem of (BMST_2).

Proposition 10. *Let x be a given leader's decision. Then y is a follower's 2-optimal solution, i.e., $y \in \mathcal{R}_2(x)$ if and only if the following two conditions hold:*

- *(i) $y \in \mathcal{S}_{\mathrm{MST}}(x)$;*
- *(ii) for any $w \in \mathcal{T}^2$, then $G[x+y+w]$ is not a spanning tree (i.e., either $x+y+w \notin \{0,1\}^m$ or $G[x+y+w]$ contains a cycle).*

Proof. The result follows directly from Proposition 1. ■

Thus, we can reformulate (BMST) as:

$$\begin{aligned} \eta^* = & \min_{x,y} \alpha(x+y) \\ & \text{s.t. } x_{ij} + y_{ij} \leq 1 \quad \forall (i,j) \in E_L, \\ & \quad G[x+y] \text{ is a spanning tree of graph } G, \\ & \quad G[x+y+w] \text{ is not a spanning tree} \quad \forall w = (e_{i_0 j_0} - e_{i_1 j_1}) \in \mathcal{T}^2, \\ & \quad x \in \{0,1\}^{m_0}, y \in \{0,1\}^m. \end{aligned}$$

To formulate the condition that $G[x+y]$ is a spanning tree via a set of linear constraints, we apply the multi-commodity flow model [40]. Let $\mathcal{A} = \{(i,j),(j,i) : (i,j) \in E\}$ be the directed arcs that are constructed from E, i.e., each edge in E is cloned into two arcs with opposite directions. Let vertex u_0 in N be an arbitrary source node. Then we impose the following constraints:

$$\sum_{(i,j)\in E_L} x_{ij} + \sum_{(i,j)\in E} y_{ij} = n-1, \tag{4a}$$

$$Af^v = \begin{cases} 1, & \text{for vertex } u_0 \\ -1, & \text{for vertex } v \\ 0, & \text{otherwise} \end{cases} \quad \forall v \in N \setminus \{u_0\}, \tag{4b}$$

$$f^v_{ij} + f^v_{ji} \le x_{ij} + y_{ij} \quad \forall (i,j) \in E, v \in N \setminus \{u_0\}, \tag{4c}$$

$$f^v_{ij} \ge 0 \quad \forall v \in N \setminus \{u_0\}, (i,j) \in \mathcal{A}, \tag{4d}$$

where A is the node-arc matrix of the directed graph $\widetilde{G} := (N, \mathcal{A})$.

To formulate the condition (ii) in Proposition 10, we first observe that if $x+y+w \notin \{0,1\}^n$, then it is clear that the condition holds; if $x+y+w \in \{0,1\}^n$ and $w = (e_{i_0j_0} - e_{i_1j_1}) \in \mathcal{T}^2$, then it implies that $w^\top y + ||w^-||_1 + |w|^\top x = 0$. Next, we use linear constraints to ensure that $G[x+y+w]$ is not a spanning tree.

Let $(i_0, j_0) \in E$ be such that $w = (e_{i_0j_0} - e_{i_1j_1}) \in \mathcal{T}^2$. Consider a shortest path problem from i_0 to j_0 in graph $\widetilde{G}[x+y+w]$, where the edge weight for (i_0, j_0) is set to n, and the weight for all other edges is set to 1. Observe that $G[x+y+w]$ is not a spanning tree if and only if the length of the shortest path from i_0 to j_0 in $\widetilde{G}[x+y+w]$ is strictly less than n. Therefore, to ensure condition (ii) in Proposition 10, we restrict the objective function of the shortest path problem to take values less than n, as follows:

$$Az^w = \begin{cases} 1, & \text{for vertex } i_0 \\ -1, & \text{for vertex } j_0 \\ 0, & \text{otherwise} \end{cases}, \tag{5a}$$

$$\sum_{(i,j)\in \mathcal{A}^w} z^w_{ij} + n z^w_{i_0j_0} \le n - 1 + w^\top y + ||w^-||_1 + |w|^\top x, \tag{5b}$$

$$0 \le z^w_{ij} \le \max\{0, x_{ij} + y_{ij} + w_{ij}\} \quad \forall (i,j) \in \mathcal{A}, \tag{5c}$$

Finally, we formalize the single-level MILP formulation for the BMST problem as:

$$\begin{aligned} \eta^* = \min_{x,y} \ & \alpha(x+y) \\ \text{s.t. } & (4), \\ & (5)\ \forall w \in \mathcal{T}^2, \\ & x_{ij} + y_{ij} \le 1 \quad \forall (i,j) \in E_L, \\ & x \in \{0,1\}^{m_0}, y \in \{0,1\}^m. \end{aligned} \tag{BMST-1}$$

Although, the mixing-set structure is not evident in the above MILP formulation, the key idea behind the extended formulation derived in Section 4.3 can be similarly applied. Recall that additional variables and precedence constraints are introduced for the extended formulation (BP_k-Mix) of (BP_k). For the BMST problem, we can thus develop another MILP formulation that also employs the precedence conditions, as described next. Note that this technique is also exploited by Shi et al. [48] (see Section 5.3), but we highlight that this idea can be generalized to other variants of the bilevel minimum spanning tree problem. For the sake of completeness and to provide a self-contained narrative, we review the concepts in our context.

Proposition 11 ([48])**.** *Assume w.l.o.g. that* $E = \{(i_k, j_k) : 1 \le k \le m\}$ *is such that* $\beta_{i_1 j_1} \le \beta_{i_2 j_2} \le \cdots \le \beta_{i_m j_m}$ *and let* $y^{<\ell}$ *be such that*

$$y^{<\ell}_{i_k j_k} = \begin{cases} y_{i_k j_k} & \text{if } k < \ell \\ 0 & \text{otherwise.} \end{cases}$$

for a given $y \in \{0,1\}^m$*. Then for a given* $x \in \mathcal{X}$*,* $y \in \mathcal{R}_2(x)$ *if and only if*

(i) $y \in \mathcal{S}_{\text{MST}}(x)$*; and*

(ii) For $1 \le \ell \le m$*,* $y_{i_\ell j_\ell} = 1$ *if and only if* i_ℓ *is disconnected from* j_ℓ *in* $\widetilde{G}[x + y^{<\ell}]$*.*

In a fashion similar to that used in deriving constraints (5), we formulate condition (*ii*) in Proposition 11 by considering a shortest path problem in graphs $G[x + y^{<\ell} + e_{i_\ell j_\ell}]$ from vertex i_ℓ to vertex j_ℓ as follows, where the formulation on the right is the LP dual of the formulation on the left.

$$\begin{aligned} \min_{z^\ell} \quad & \sum_{(i,j)\in\mathcal{A}^\ell_{[x+y^{<\ell}]}} z^\ell_{ij} + n z^\ell_{i_\ell j_\ell} \\ \text{s.t.} \quad & A_{[x+y^{<\ell}+e_{i_\ell j_\ell}]} z^\ell = \begin{cases} 1, & \text{for vertex } i_\ell \\ -1, & \text{for vertex } j_\ell \\ 0, & \text{otherwise} \end{cases}, \\ & z^\ell_{ij} \ge 0 \quad \forall\, (i,j) \in \mathcal{A}_{[x+y^{<\ell}+e_{i_\ell j_\ell}]}, \end{aligned} \qquad \begin{aligned} \max_{\pi^\ell} \quad & \pi^\ell_{i_\ell} - \pi^\ell_{j_\ell} \\ \text{s.t.} \quad & \pi^\ell_i - \pi^\ell_j \le 1 \quad \forall\, (i,j) \in \mathcal{A}_{[x+y^{<\ell}]}, \\ & \pi^\ell_{i_\ell} - \pi^\ell_{j_\ell} \le n, \end{aligned}$$

where $A_{[x+y^{<\ell}+e_{i_\ell j_\ell}]}$ is the node-arc matrix of graph $\widetilde{G}[x + y^{<\ell} + e_{i_\ell j_\ell}]$ and $\mathcal{A}_{[x+y^{<\ell}+e_{i_\ell j_\ell}]}$ is its associated set of arcs. Note that there does not exist a path from i_ℓ to j_ℓ in $\widetilde{G}[x + y^{<\ell}]$ if and only if the above shortest path problem has optimal objective function value of n. Therefore, we enforce the following constraints for condition (*ii*) in Proposition 11:

$$A^\ell z^\ell = \begin{cases} 1, & \text{for vertex } i_\ell \\ -1, & \text{for vertex } j_\ell \\ 0, & \text{otherwise} \end{cases}, \tag{6a}$$

$$z^\ell_{ij} + z^\ell_{ji} \le x_{ij} + y_{ij} \quad \forall (i,j) \in \mathcal{A}^\ell, \tag{6b}$$

$$z^\ell_{ij} \ge 0 \quad \forall (i,j) \in \mathcal{A}^\ell \cup (i_\ell, j_\ell), \tag{6c}$$

$$\pi^\ell_i - \pi^\ell_j \le 1 + \mu(1 - y_{ij}) \quad \forall\, (i,j) \in \mathcal{A}^\ell, \tag{6d}$$

$$\pi^\ell_{i_\ell} - \pi^\ell_{j_\ell} \le n - 1 + x_{i_\ell j_\ell} + y_{i_\ell j_\ell}, \tag{6e}$$

$$\pi^\ell_{i_\ell} - \pi^\ell_{j_\ell} \ge n - \mu(1 - y_{i_\ell j_\ell} + x_{i_\ell j_\ell}), \tag{6f}$$

$$z^\ell_{ij} + n z^\ell_{i_\ell j_\ell} = \pi^\ell_{i_\ell} - \pi^\ell_{j_\ell}, \tag{6g}$$

where A^ℓ is the node-arc incidence matrix of graph $\widetilde{G}^\ell = (N, \mathcal{A}^\ell)$ and $\mathcal{A}^\ell = \{(i,j),(j,i) : (i,j) \in E_L\} \cup \{(i_p, j_p), (j_p, i_p) : 1 \le p < \ell\}$, and μ is sufficiently large, e.g., n.

Based on the above, we provide another single-level MILP formulation for BMST as follows:

$$\begin{aligned}\eta^* = \min_{x,y}\ & \alpha(x+y) \\ \text{s.t. } & (4), \\ & (6)\ \forall \ell = 1,\dots,m, \\ & x_{ij} + y_{ij} \le 1 \quad \forall (i,j) \in E_L, \\ & x \in \{0,1\}^{m_0}, y \in \{0,1\}^m. \end{aligned} \tag{BMST-2}$$

All in all, the MILP formulations (BMST-1) and (BMST-2) are derived based on the local optimality conditions in Propositions 10 and 11, respectively. We note that (BMST-2) is similar to the formulation proposed in [48]. With respect to the latter, we need to point out the following observations.

(i) The only difference between (BMST-2) and the model in [48] is how to formulate condition (i) in Proposition 11 (i.e., $G[x+y]$ is a spanning tree of G). Since the considered version of (BMST) is to construct a spanning tree by joint actions of both decision-makers, the condition in Proposition 11 (i) is implied by Proposition 11 (ii) and a new condition that $G[x]$ does not contain a cycle. The formulation in [48] adopts the latter approach to reformulate (BMST), which also allows to reduce the number of constraints in comparison to the formulation in (BMST-2). However, we observe in our experiments (not reported here) that the formulation (BMST-2) is stronger than the one in [48], as the LP relaxation of (BMST-2) always provides a tighter lower bound.

(ii) The modeling approach discussed in this section can be applied to other variants of (BMST), e.g., the minimum edge blocker spanning tree problem [56], the minimum spanning tree interdiction problem [19], where the leader removes the edges in the graph to maximize the weight of follower's minimum spanning tree. Under this setting, the condition (i) in Propositions 10 and 11 is that $G[y]$ is a spanning tree of G. Thus, the constraints used in the formulation from [48] are not applicable, while the formulations (BMST-1) and (BMST-2) can be easily extended with slight modifications.

Finally, Shi et al. [48] discuss an efficient preprocessing procedure to substantially reduce the size of the formulation. We note that their procedure can also be applied to our formulations (BMST-1) and (BMST-2); we omit its discussion for brevity.

6.3.2 Computational Results for (BMST)

We now focus on computationally comparing the MILP formulations (BMST-1) and (BMST-2).

Experimental setup. We use the graphs from the test set **B** in [32]. The test set **B** contains 18 graphs with $50-100$ number of vertices. We generate our test instances similar to the procedure in [48]. Specifically, for each graph in **B**, we randomly generate the edges in E_L with a specific fractional value $\rho \in \{0.05, 0.1, 0.15\}$. Here, ρ denotes the ratio between the number of edges in E_L and in E, i.e., $\rho := \frac{|E_L|}{|E|}$.

The follower's and leader's edge weight is generated through $\beta_{ij} = w_{ij} r_{ij}$ and $\alpha_{ij} = w_{ij}(1-r_{ij})$ for each $(i,j) \in E$, respectively; where w_{ij} is the original edge weight from graphs in **B** and proportion r_{ij} is uniformly generated from interval $[0,1]$. The average performance is reported over 10 random instances for each pair of graph and ρ. We set the time limit to one hour.

			$\rho = 0.05$				$\rho = 0.1$				$\rho = 0.15$			
			(BMST-1)		(BMST-2)		(BMST-1)		(BMST-2)		(BMST-1)		(BMST-2)	
Ins.	$\lvert N\rvert$	$\lvert E\rvert$	IG (%)	Time	IG (%)	Time	IG (%)	Time	IG (%)	Time	IG (%)	Time	IG (%)	Time
b1	50	63	0	0.04	0	0.02	0	0.08	0	0.04	0	0.14	0	0.04
b2	50	63	0	0.03	0	0.02	0	0.04	0	0.03	0.52	0.23	0.52	0.06
b3	50	63	0	0.04	0	0.02	0	0.12	0	0.04	0.83	0.13	0.83	0.04
b4	50	100	0.16	0.07	0.16	0.03	0.33	0.22	0.33	0.06	1.87	0.62	1.86	0.10
b5	50	100	0	0.07	0	0.04	0.68	0.24	0.60	0.06	2.65	0.73	2.44	0.14
b6	50	100	0	0.09	0	0.04	0.72	0.28	0.53	0.08	2.52	0.67	2.42	0.12
b7	75	94	0	0.16	0	0.07	0	0.34	0	0.08	0.50	1.04	0.50	0.13
b8	75	94	0	0.07	0	0.05	0.06	0.30	0.06	0.08	0.15	0.85	0.12	0.14
b9	75	94	0	0.13	0	0.06	0	0.47	0	0.10	0.14	0.81	0.14	0.13
b10	75	150	0.16	0.30	0.16	0.10	0.73	0.96	0.70	0.18	2.43	7.38	2.39	0.45
b11	75	150	0.13	0.27	0.13	0.09	1.11	1.32	1.02	0.21	3.13	7.31	2.90	0.52
b12	75	150	0.06	0.26	0.06	0.09	0.71	1.35	0.68	0.19	2.50	6.64	2.48	0.48
b13	100	125	0.08	0.34	0.08	0.14	0.04	0.90	0.04	0.15	0.61	2.72	0.61	0.33
b14	100	125	0	0.26	0	0.11	0.13	0.91	0.13	0.16	0.25	1.78	0.25	0.20
b15	100	125	0	0.30	0	0.11	0	0.67	0	0.13	0.10	2.14	0.10	0.23
b16	100	200	0.30	1.05	0.26	0.21	1.27	4.98	1.23	0.50	3.92	201.64	3.73	1.77
b17	100	200	0.29	1.00	0.29	0.22	0.83	3.57	0.79	0.42	1.81	11.15	1.69	0.84
b18	100	200	0.05	0.43	0.05	0.14	1.39	3.11	1.32	0.38	5.12	139.41	5.05	2.22

Table 7. Results for the instances of (BMST). For each graph in **B**, the number of vertices and edges are denoted in the columns "$|N|$" and "$|E|$", respectively. For each formulation, we report the average solver's runtime in the column "Time". Column "IG (%)" reports the integrality gap, which is computed by $\text{IG}(\%) := \frac{(\eta^* - \eta_{LP})}{\eta^*} \times 100$, where η^* and η_{LP} denote the optimal objective function value of the formulation and its LP relaxation, respectively.

Results and discussion. The computational results for the MILP formulations (BMST-1) and (BMST-2) are reported in Table 7. For each graph in **B**, the number of vertices and edges are denoted in the columns "$|N|$" and "$|E|$", respectively. For each formulation, we report the average solver's runtime in the column "Time". We also report the integrality gap in the column "IG (%)", which is computed by

$$\text{IG}(\%) := \frac{(\eta^* - \eta_{LP})}{\eta^*} \times 100,$$

where η^* and η_{LP} denote the optimal objective function value of the formulation and its LP relaxation, respectively.

As expected, the BMST instances become more difficult as we increase of the number edges controlled by the leader (i.e., larger value of ρ) and the graph density. However, as we can see in Table 7, the average integrality gaps of both MILP formulations are very close and rather small, typically, under 5% for all instances even with $\rho = 0.15$. Moreover, all of the tested instances can be solved to optimality within the time limits when using both formulations. In particular, we observe that the formulation (BMST-2) based on the precedence edge orders performs better than (BMST-1). It usually requires a few seconds for the solver to handle (BMST-2), while (BMST-1) needs more than 5 minutes for large graphs and ρ. These results suggest that our formulations (BMST-1) and (BMST-2) are extremely tight, and the MILP formulation (BMST-2) might be more effective for solving the BMST problem.

7 Concluding Remarks

We considered mixed integer bilevel linear optimization problems in which the follower's decision variables are all binary. In response to the leader's decision, the proposed framework assumes that the follower does not have sufficient computational capabilities to obtain globally optimal solutions but instead implements a locally optimal solution. To capture the local optimality requirement we use the concept of k-optimality, where k is some predetermined neighborhood size of a given 0–1 vector. That is, $k = 0$ implies that the follower's objective function is completely ignored (also known as the single-level relaxation of the original bilevel problem), while $k = n$ corresponds for the fully-rational follower, who solves the lower-level problem to global optimality.

Under the assumption that the follower is optimistic, our framework naturally provides a hierarchy of upper and lower bounds for the standard bilevel optimization problem, where the follower is fully rational. To compute these bound for any fixed k, we develop single-level formulations, which can be solved by off-the-shelf solvers. In our extensive computational study the proposed bounds converge to the optimal objective function values of bilevel problems for reasonably small values of k. Moreover, the proposed framework provides lower and upper bounds of substantially better quality than those based on the widely used single-level relaxation method. Hence, our framework can be embedded into exact solvers—in particular, those that rely on single-level relaxations. Furthermore, we note that the constraints in our proposed MILP formulations are also valid inequalities for the original bilevel problem; consequently, these constraints can be applied as cutting planes in branch-and-cut solvers. Therefore, embedding the proposed methods into general purpose branch-and-cut solvers for mixed integer bilevel optimization problems provides a promising direction for future research.

Our framework can also be used for solving classes of bilevel problems, in which local optimality of a follower's decision (within some sufficiently "small" neighborhood) implies its global optimality for the lower-level problem. As an example, in our paper we exploit this idea to reformulate a general class of bilevel matroid problems as equivalent linear MILPs.

One of our framework's limitations is that the sizes of the proposed MILP formulations with large k prevent us for solving large-scale instances. Thus, improving the scalability of our approaches (e.g., by designing more advanced exact and approximate solution methods) is another important direction for future research.

Finally, if the follower is pessimistic, then our framework results in a tri-level optimization problem, which can be formalized as:

$$\begin{aligned}\max_x \min_y \ & \alpha^1 x + \alpha^2 y \\ \text{s.t. } & (x,y) \in \mathcal{S}, \\ & y \in \mathcal{R}_k(x) = \{y \in \mathcal{S}(x) : \beta y \geq \beta \hat{y} \ \forall \hat{y} \in \mathcal{N}_k(y) \cap \mathcal{S}(x)\}.\end{aligned}$$

Note that the ideas behind our MILP reformulations can be directly extended to represent $\mathcal{R}_k(x)$ via linear constraints. It implies that the above tri-level optimization problem can be reduced to a bilevel max-min problem. To solve the latter, development of more advanced solution strategies provides an interesting topic for further research.

Acknowledgements. The authors would like to thank the Review Team for their helpful and constructive comments. This research is based upon work supported by the Office of Naval Research [Grant N00014-19-1-2330].

References

[1] E. Aarts, E. H. Aarts, and J. K. Lenstra. *Local Search in Combinatorial Optimization.* Princeton University Press, Princeton, NJ, USA, 2003.

[2] R. K. Ahuja, T. L. Magnanti, J. B. Orlin, and M. Reddy. Applications of network optimization. In M. Ball, T. Magnanti, C. Monma, and G. Nemhauser, editors, *Handbooks in Operations Research and Management Science, Vol. 7*, volume 7, pages 1–83. Elsevier, Amsterdam, 1995.

[3] A. Atamtürk, G. L. Nemhauser, and M. W. Savelsbergh. The mixed vertex packing problem. *Mathematical Programming*, 89(1):35–53, 2000.

[4] C. Audet, P. Hansen, B. Jaumard, and G. Savard. Links between linear bilevel and mixed 0–1 programming problems. *Journal of Optimization Theory and Applications*, 93(2):273–300, 1997.

[5] C. Bazgan, S. Toubaline, and Z. Tuza. The most vital nodes with respect to independent set and vertex cover. *Discrete Applied Mathematics*, 159(17):1933–1946, 2011.

[6] B. Beheshti, O. A. Prokopyev, and E. L. Pasiliao. Exact solution approaches for bilevel assignment problems. *Computational Optimization and Applications*, 64(1):215–242, 2016.

[7] J. S. Borrero, O. A. Prokopyev, and D. Sauré. Sequential interdiction with incomplete information and learning. *Operations Research*, 67(1):72–89, 2019.

[8] O. Bräysy and M. Gendreau. Vehicle routing problem with time windows, Part I: Route construction and local search algorithms. *Transportation Science*, 39(1):104–118, 2005.

[9] L. Brotcorne, M. Labbé, P. Marcotte, and G. Savard. A bilevel model for toll optimization on a multicommodity transportation network. *Transportation Science*, 35(4):345–358, 2001.

[10] A. Caprara, M. Carvalho, A. Lodi, and G. J. Woeginger. A study on the computational complexity of the bilevel knapsack problem. *SIAM Journal on Optimization*, 24(2):823–838, 2014.

[11] A. Caprara, M. Carvalho, A. Lodi, and G. J. Woeginger. Bilevel knapsack with interdiction constraints. *INFORMS Journal on Computing*, 28(2):319–333, 2016.

[12] B. Chandra, H. Karloff, and C. Tovey. New results on the old k-opt algorithm for the traveling salesman problem. *SIAM Journal on Computing*, 28(6):1998–2029, 1999.

[13] B. Colson, P. Marcotte, and G. Savard. An overview of bilevel optimization. *Annals of Operations Research*, 153(1):235–256, 2007.

[14] F. Della Croce and R. Scatamacchia. An exact approach for the bilevel knapsack problem with interdiction constraints and extensions. *Mathematical Programming*, 2020. To appear.

[15] S. Dempe. *Foundations of Bilevel Programming*. Kluwer Academic Publishers, Dordrecht, 2002.

[16] S. DeNegre. *Interdiction and discrete bilevel linear programming*. PhD thesis, Lehigh University, 2011.

[17] M. Fischetti, I. Ljubić, M. Monaci, and M. Sinnl. On the use of intersection cuts for bilevel optimization. *Mathematical Programming*, 172(1-2):77–103, 2018.

[18] M. Fischetti, I. Ljubić, M. Monaci, and M. Sinnl. Interdiction games and monotonicity, with application to knapsack problems. *INFORMS Journal on Computing*, 31(2):390–410, 2019.

[19] G. N. Frederickson and R. Solis-Oba. Increasing the weight of minimum spanning trees. *Journal of Algorithms*, 33(2):244–266, 1999.

[20] Z. Gao, J. Wu, and H. Sun. Solution algorithm for the bi-level discrete network design problem. *Transportation Research Part B: Methodological*, 39(6):479–495, 2005.

[21] M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, New York, 2002.

[22] O. Günlük and Y. Pochet. Mixing mixed-integer inequalities. *Mathematical Programming*, 90(3):429–457, 2001.

[23] G. Gutin and A. P. Punnen. *The Traveling Salesman Problem and Its Variations*, volume 12. Kluwer Academic Publishers, New York, 2004.

[24] R. W. Hamming. *Coding and information theory*. Prentice-Hall, Inc., 1986.

[25] R. Horst, P. M. Pardalos, and N. Van Thoai. *Introduction to Global Optimization*. Kluwer Academic Publishers, Dordrecht, 2000.

[26] IBM. ILOG CPLEX Optimization Studio (12.8), 2017. Accessed April 9, 2020.

[27] E. Israeli and R. K. Wood. Shortest-path network interdiction. *Networks: An International Journal*, 40(2):97–111, 2002.

[28] R. G. Jeroslow. The polynomial hierarchy and a simple model for competitive analysis. *Mathematical Programming*, 32(2):146–164, 1985.

[29] K. Katayama, A. Hamamoto, and H. Narihisa. An effective local search for the maximum clique problem. *Information Processing Letters*, 95(5):503–511, 2005.

[30] B. W. Kernighan and S. Lin. An efficient heuristic procedure for partitioning graphs. *The Bell system technical journal*, 49(2):291–307, 1970.

[31] T. Kis and A. Kovács. On bilevel machine scheduling problems. *OR Spectrum*, 34(1):43–68, 2012.

[32] T. Koch, A. Martin, and S. Voß. Steinlib: An updated library on steiner tree problems in graphs. In X. Z. Cheng and D.-Z. Du, editors, *Steiner Trees in Industry*, pages 285–325. Springer, Boston, MA, 2001.

[33] M. Labbé, P. Marcotte, and G. Savard. A bilevel model of taxation and its application to optimal highway pricing. *Management Science*, 44(12-part-1):1608–1622, 1998.

[34] E. L. Lawler. *Combinatorial Optimization: Networks and Matroids*. Courier Corporation, New York, 2001.

[35] J. Lee and J. Ryan. Matroid applications and algorithms. *ORSA Journal on Computing*, 4(1):70–98, 1992.

[36] J. Leskovec and R. Sosič. Snap: A general-purpose network analysis and graph-mining library. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 8(1):1, 2016.

[37] A. Lodi, T. K. Ralphs, and G. J. Woeginger. Bilevel programming and the separation problem. *Mathematical Programming*, 146(1-2):437–458, 2014.

[38] L. Lozano and J. C. Smith. A value-function-based exact approach for the bilevel mixed-integer programming problem. *Operations Research*, 65(3):768–786, 2017.

[39] J. Luedtke, S. Ahmed, and G. L. Nemhauser. An integer programming approach for linear programs with probabilistic constraints. *Mathematical Programming*, 122(2):247–272, 2010.

[40] T. L. Magnanti and L. A. Wolsey. Optimal trees. In M. Ball, T. Magnanti, C. Monma, and G. Nemhauser, editors, *Network Models*, volume 7 of *Handbooks in Operations Research and Management Science*, pages 503 – 615. Elsevier, 1995.

[41] R. T. Maheswaran, J. P. Pearce, and M. Tambe. Distributed algorithms for dcop: A graphical-game-based approach. In *ISCA PDCS*, pages 432–439, 2004.

[42] S. Martello, D. Pisinger, and P. Toth. Dynamic programming and strong bounds for the 0-1 knapsack problem. *Management Science*, 45(3):414–424, 1999.

[43] J. T. Moore and J. F. Bard. The mixed integer linear bilevel programming problem. *Operations Research*, 38(5):911–921, 1990.

[44] J. Oxley. What is a matroid. *Cubo*, 5:179–218, 2003.

[45] H. N. Psaraftis. k-interchange procedures for local search in a precedence-constrained routing problem. *European Journal of Operational Research*, 13(4):391–402, 1983.

[46] M. W. Savelsbergh. Local search in routing problems with time windows. *Annals of Operations research*, 4(1):285–305, 1985.

[47] A. A. Schäffer. Simple local search problems that are hard to solve. *SIAM Journal on Computing*, 20(1):56–87, 1991.

[48] X. Shi, B. Zeng, and O. A. Prokopyev. On bilevel minimum and bottleneck spanning tree problems. *Networks*, 74(3):251–273, 2019.

[49] A. Sinha, P. Malo, and K. Deb. A review on bilevel optimization: From classical to evolutionary approaches and applications. *IEEE Transactions on Evolutionary Computation*, 22(2):276–295, 2017.

[50] J. C. Smith, C. Lim, and F. Sudargho. Survivable network design under optimal and heuristic interdiction scenarios. *Journal of Global Optimization*, 38(2):181–199, 2007.

[51] S. Tahernejad and T. K. Ralphs. Valid inequalities for mixed integer bilevel optimization problems. *COR@L Laboratory Report 20T-013*, 2020. Accessed June 7, 2021.

[52] S. Tahernejad, T. K. Ralphs, and S. T. DeNegre. A branch-and-cut algorithm for mixed integer bilevel linear optimization problems and its implementation. *Mathematical Programming Computation*, 12(4):529–568, 2020.

[53] Y. Tang, J.-P. P. Richard, and J. C. Smith. A class of algorithms for mixed-integer bilevel min–max optimization. *Journal of Global Optimization*, 66(2):225–262, 2016.

[54] L. Vicente, G. Savard, and J. Judice. Discrete linear bilevel programming problem. *Journal of Optimization Theory and Applications*, 89(3):597–614, 1996.

[55] L. Wang and P. Xu. The watermelon algorithm for the bilevel integer linear programming problem. *SIAM Journal on Optimization*, 27(3):1403–1430, 2017.

[56] N. Wei, J. L. Walteros, and F. M. Pajouh. Integer programming formulations for minimum spanning tree interdiction. *Optimization online*, 2019. Acessed April 9, 2020.

[57] N. White and N. M. White. *Matroid Applications.* Number 40. Cambridge University Press, 1992.

[58] H. Xu, Y. Zhang, C. G. Cassandras, L. Li, and S. Feng. A bi-level cooperative driving strategy allowing lane changes. *arXiv preprint arXiv:1912.11495*, 2019. Accessed April 9, 2020.

[59] M. H. Zare, J. S. Borrero, B. Zeng, and O. A. Prokopyev. A note on linearized reformulations for a class of bilevel linear integer problems. *Annals of Operations Research*, 272(1-2):99–117, 2019.

[60] M. H. Zare, O. A. Prokopyev, and D. Sauré. On bilevel optimization with inexact follower. *Decision Analysis*, 17(1):74–95, 2020.

[61] B. Zeng and Y. An. Solving bilevel mixed integer program by reformulations and decomposition. *Optimization online*, 2014. Accessed April 9, 2020.

[62] M. Zhao, K. Huang, and B. Zeng. A polyhedral study on chance constrained program with random right-hand side. *Mathematical Programming*, 166(1-2):19–64, 2017.