Solving Bilevel Mixed Integer Program by Reformulations and Decomposition

June, 2014

Abstract

In this paper, we study bilevel mixed integer programming (MIP) problem and present a novel computing scheme based on reformulations and decomposition strategy. By converting bilevel MIP into a constrained mathematical program, we present its single-level reformulations that are friendly to perform analysis and build insights. Then, we develop a decomposition algorithm based on column-and-constraint generation method, which converges to the optimal value within finite operations. A preliminary computational study on randomly generated instances is presented, which demonstrates that the developed computing scheme has a superior capacity over existing methods. As it is generally applicable, easy-to-use and computationally strong, we believe that this solution method makes an important progress in solving challenging bilevel MIP problem.

Key words: bilevel optimization, mixed integer programming, reformulation, decomposition algorithm

1 Introduction

Bilevel optimization is an optimization scheme to model a non-centralized system that has two decision makers (DMs) at different levels driven by their own interests. Decisions made by the upper-level DM affect the feasible decision set of the lower-level DM, while an equilibrium response, i.e., an optimal decision, from the lower-level constitutes a part of the performance evaluation of the upper-level. Indeed, because of such sequential interaction between them, this decision making structure is also called Stackelberg leader-follower game, where the upper-level and lower-level

DMs are treated as the leader and follower respectively. Mathematically, by defining two optimization problems for those DMs respectively, the whole decision making structure is formulated as the following bilevel optimization model.

$$\mathbf{BiMIP}: \quad \Theta^* = \min \ \mathbf{fx} + \mathbf{gy} + \mathbf{hz} \tag{1}$$

$$\text{s.t.} \quad \mathbf{Ax} \le \mathbf{b}, \ \mathbf{x} \in \mathbb{R}_+^{m_c} \times \mathbb{Z}_+^{m_d}, \tag{2}$$

$$(\mathbf{y}, \mathbf{z}) \in \mathcal{F}(\mathbf{x}) \equiv \arg\max\{\mathbf{wy} + \mathbf{vz} : \mathbf{Py} + \mathbf{Nz} \le \mathbf{R} - \mathbf{Kx}, \mathbf{y} \in \mathbb{R}_+^{n_c}, \mathbf{z} \in \mathbb{Z}_+^{n_d}\} \tag{3}$$

where $\mathbf{x}$ represents the upper-level decision variables, $\mathbf{y}$ represents the lower-level continuous decision variables, and $\mathbf{z}$ represents the lower-level discrete decision variables. The differentiation between $\mathbf{y}$ and $\mathbf{z}$ is to highlight discrete variables and to streamline our exposition in the remainder of this paper. Clearly, the membership requirement in (3) ensures that an optimal solution of the lower-level DM is fed back to the upper-level DM. Note that formulation **BiMIP** is often referred to as *optimistic bilevel formulation* as the upper-level DM is able to select $(\mathbf{y}, \mathbf{z})$, if $\mathcal{F}(\mathbf{x})$ is not a singleton, for her own favor. A more conservative model, which is less studied, is *pessimistic bilevel formulation* [34] where the lower-level DM is assumed to against her by choosing the least favorable one. In this paper, we follow the majority of existing research and focus our study on optimistic bilevel formulation.

Since its introduction in 1970s [11, 16], bilevel optimization has received enormous research attention and has been widely applied to study and support many practical hierarchical decision making problems. Such situations happen very often in transportation planning and network capacity expansion [12, 19, 35], government policy making [8, 18], revenue management [13, 20], and computational biology [41, 15]. With deregulations of power systems and organizations of electricity markets, bilevel and general multilevel optimization have been applied to deal with various management challenges between market administrators and participants, including power generation, market bidding, and capacity expansion [43, 27, 31]. Moreover, the interdiction model, a special class of bilevel optimization model, where the upper-level and lower-level DMs have completely opposite interests, has been intensively utilized in military and homeland security applications for strategic planning and system vulnerability analysis [1, 44, 10, 14, 33].

Although it is widely applied in modeling and analyzing practical problems, computing bilevel mixed integer programming (MIP) problem, i.e., **BiMIP** in (1-3), is not easy. Even for the simplest bilevel linear programming (LP) problem, whose both upper and lower-level problems are linear programs, it is theoretically NP-hard [25, 3]. Yet, using results of Karush-Kuhn-Tucker (KKT) optimality conditions or strong duality from linear programming theory, bilevel LP is often converted into an LP with complementarity constraints or an MIP, which yields a computationally feasible approach to solve practical issues. Indeed, because those crucial structural properties are only applicable to LP, the majority of existing research efforts does not consider general **BiMIP** with mixed integer lower-level problems. Up to know, only a few algorithms have been developed [22, 37, 49, 50] that are able to compute bilevel problem whose lower-level problem has discrete variables. Nevertheless, we note that those algorithms either (i) heavily depend on enumerative Branch-and-Bound strategies based on a rather weak relaxation, or (ii) involve complicated operations that are problem specific and challenging for most researchers and practitioners. Hence, existing methods are of very limited computational capability. As a consequence, there is no commonly accepted approach and little support is available to transform **BiMIP** into a decision making tool for real system practice. Given such situation, some researchers consider general **BiMIP** as an open problem in operations research [21].

In this paper, to improve our solution capacity on **BiMIP** and to change its application status in practical systems, we present a new solution scheme that has a clear mathematical foundation for analysis and a simple algorithmic structure for implementation. In theoretical aspect, through reformulations, it provides strong and computationally friendly relaxations, and ensures validity and convergence of the whole procedure. In algorithmic aspect, it employs an easy-to-use decomposition approach, which minimizes unnecessary operations and computational expenses to derive exact solutions. In computational aspect, on a set of random instances, it often produces optimal solutions within a very small number of algorithmic operations and demonstrates a superior performance over existing ones. Given those advantages, we believe that this new solution method to **BiMIP** is an effective tool that is of a great significance in practice.

We organize the rest of this paper as follows. In Section 2, we briefly review existing solution methods for bilevel optimization problem. In Section 3, we present an equivalent formulation different from **BiMIP** formulation in (1-3) and analyze its advantages. In Section 4, we introduce a few single-level reformulations and discuss their mathematical implications. In Section 5, we describe a decomposition method that converges to the optimal value within a finite number of iterations. In Section 6, we present numerical results obtained on a set of randomly generated instances. Section 7 concludes this paper with a discussion on future research directions.

2 Existing Solution Methods of Bilevel MIP

Although many efficient algorithms and software packages have been designed and developed for general or structured single-level MIP, computing bilevel MIP remains challenging. One fundamental reason is that the bilevel formulation itself in (1-3) is defined in a rather implicit way. Let 3-tuple $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ represent one solution of **BiMIP**. Its feasible set, which is often called the *inducible region*, is defined in the following parametric fashion.

$$\Omega^I = \{(\mathbf{x}, \mathbf{y}, \mathbf{z}) : \mathbf{A}\mathbf{x} \leq \mathbf{b}, (\mathbf{y}, \mathbf{z}) \in \mathcal{F}(\mathbf{x}), \mathbf{x} \in \mathbb{R}_+^{m_c} \times \mathbb{Z}_+^{m_d}\}.$$

Analyzing such parametric representation and developing structural insights are difficult. However, under a special situation where the lower-level problem is a pure LP, i.e., no $\mathbf{z}$ variables and $n_d = 0$, Ω^I's mathematical representation can be drastically simplified. Recall that for an LP problem with a finite optimal value, its Karush-Kuhn-Tucker (KKT) conditions, which include primal feasibility, dual feasibility, and complementary slackness conditions, are necessary and sufficient to characterize an optimal solution. In other words, those KKT conditions can be used to represent optimal solutions. Hence, we can replace $\mathcal{F}(\mathbf{x})$ by the corresponding KKT conditions [24]. As a result, Ω^I is reformulated into the following set with linear and disjunctive constraints (from complementary slackness conditions in (6)), where π is the vector of dual variables with appropriate dimension n_1. For the easiness of exposition, we employ $\perp$ signs to compactly represent complementarity conditions.

$$\Omega^I = \{\mathbf{x} \in \mathbb{R}_+^{m_c} \times \mathbb{Z}_+^{m_d}, \mathbf{y} \in \mathbb{R}_+^{n_c}, \pi \in \mathbb{R}_+^{n_1} : \quad \mathbf{A}\mathbf{x} \leq \mathbf{b}, \tag{4}$$

$$\mathbf{P}\mathbf{y} \leq \mathbf{R} - \mathbf{K}\mathbf{x}, \ \mathbf{P}^t\pi \geq \mathbf{w}^t, \tag{5}$$

$$\mathbf{y} \perp (\mathbf{P}^t\pi - \mathbf{w}^t), \ \pi \perp (\mathbf{R} - \mathbf{K}\mathbf{x} - \mathbf{P}\mathbf{y})\}. \tag{6}$$

Therefore, **BiMIP** is equivalently converted into the following single-level optimization model.

$$\min\{\mathbf{fx} + \mathbf{gy} : (4) - (6)\} \tag{7}$$

Such single-level reformulation provides a great convenience to design computing algorithms. For example, the disjunctive structure displayed in (6) motivates many Branch-and-Bound methods, including [24, 6, 7, 26, 48]. Actually, **BiMIP** with Ω^I in (4-6) is a mathematical program with linear complementarity constraints (MPCC) [46], which is often solved by nonlinear programming algorithms [42, 30, 40]. Moveover, constraints of complementary slackness conditions can be linearized using additional binary variables [3], along with nonnegativity properties of decision variables. Specifically, consider a complementary slackness condition $\pi_i(\mathbf{R} - \mathbf{Kx} - \mathbf{Py})_i = 0$, where subscript i is to denote i^{th} component of the associated vector. It is equivalent to

$$\pi_i \leq M\delta_i, \ (\mathbf{R} - \mathbf{Kx} - \mathbf{Py})_i \leq M(1 - \delta_i), \ \delta_i \in \{0, 1\} \tag{8}$$

where M is a sufficiently large number. By applying this conversion to every complementary slackness condition, the whole **BiMIP** formulation then can be reformulated into a regular MIP problem. Indeed, in concert with powerful professional MIP solvers, this MIP reformulation strategy has been widely employed as it allows researchers and practitioners to compute **BiMIP** instances with little effort on designing sophisticated algorithms.

A natural connection between bilevel MIP and single-level MIP can be established if we simply require $(\mathbf{y}, \mathbf{z})$ to be feasible, instead of optimal, for a given $\mathbf{x}$. Then, bilevel MIP reduces to the following single-level MIP, which is called *high point problem* [37].

$$\phi^* = \min_{(\mathbf{x},\mathbf{y},\mathbf{z}) \in \Omega} (\mathbf{fx} + \mathbf{gy} + \mathbf{hz})$$

where

$$\Omega = \{\mathbf{x} \in \mathbb{R}^{m_c}_+ \times \mathbb{Z}^{m_d}_+, \mathbf{y} \in \mathbb{R}^{n_c}_+, \mathbf{z} \in \mathbb{Z}^{n_d}_+ : \mathbf{Ax} \leq \mathbf{b}, \mathbf{Py} + \mathbf{Nz} + \mathbf{Kx} \leq \mathbf{R}\}.$$

It is easy to see that $\Omega^I \subseteq \Omega$ and high point problem is a relaxation to **BiMIP**. Indeed, it has been proven in [17, 9, 5] that when both upper and lower-levels are LP problems, i.e.,

$m_d = n_d = 0$, an optimal solution to **BiMIP** will be an extreme point of Ω. Hence, a few algorithms are developed to identify an optimal solution by evaluating extreme points of Ω, which is often referred to as *vertex enumeration*, in an efficient way [17, 39, 4, 9]. Nevertheless, we point out that those vertex enumeration methods only work for instances with LP upper and lower-level problems. On the contrary, the KKT conditions based reformulation strategy is more general as it does not restrict the upper-level to be an LP.

When the lower-level problem has discrete variables, which renders KKT conditions invalid, algorithm development becomes scarce. Among a couple of algorithms [37, 22, 49, 45] developed over the last 20 years, almost all of them are directly implemented with Branch-and-Bound techniques [37, 22, 49, 45, 50], which are more suitable for those with pure IP lower-level problems [22, 49]. Actually, high point problem is generally adopted as the fundamental relaxation within those Branch-and-Bound schemes. However, as demonstrated in [37], this relaxation is very weak, leading to very large Branch-and-Bound trees with long computation time. To improve performance of Branch-and-Bound methods, fast heuristic variants are designed [37], cutting planes are generated for strengthening [22], and a sophisticated master-sub problem decomposition method is designed to guide the branching process and to create nodes [49]. Different from Branch-and-Bound methods, a parametric integer programming approach is developed in [32], which, however, is rather a conceptual method with no numerical evaluation.

The interdiction model, as a structured bilevel optimization problem, is formulated as the following.

$$\Theta^* = \min_{\mathbf{x}} \max_{(\mathbf{y},\mathbf{z})\in\mathcal{F}^0(\mathbf{x})} \mathbf{wy} + \mathbf{vz} \tag{9}$$

where $\mathcal{F}^0(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}_+^{n_c}, \mathbf{z} \in \mathbb{Z}_+^{n_d} : \mathbf{Py} + \mathbf{Nz} \leq \mathbf{R} - \mathbf{Kx}\}$. Because the gain of one level is exactly the loss of the other one, it represents a zero-sum game and is widely employed in decision making and analysis of security and defense applications. When the lower-level is an LP, the interdiction model is often reformulated into a single-level problem with a bilinear objective function as in (10), based on the strong duality (with π satisfying all constraints in the dual problem).

$$\min_{\mathbf{x}} \max_{\mathbf{y}\in\mathcal{F}^0(\mathbf{x})} \mathbf{wy} \Leftrightarrow \min_{\mathbf{x},\pi} \pi(\mathbf{R} - \mathbf{Kx}). \tag{10}$$

In many cases [1, 2, 38], especially for network interdiction applications, $\mathbf{x}$ are binary variables. So, the nonlinear products between $\mathbf{x}$ and π can easily be linearized and the whole formulation is converted into an MIP. We mention that, compared with the reformulation based on KKT conditions, this formulation is with less variables and constraints and typically demands much less computational expenses [1]. Certainly, as one type of **BiMIP**, neither reformulation strategy is applicable when discrete variables appear in the lower-level. Nevertheless, because network interdiction problems often demonstrate clear structural properties, some specialized computing methods are developed for those with mixed integer lower-level problems [45, 21].

Overall, we note that, for **BiMIP** with an LP lower-level problem, the KKT conditions based single-level reformulation strategy is arguably the most popular solution method. In particular, the MIP representation of that reformulation, in concert with existing powerful MIP solvers, provides a convenient and efficient computing approach. As a result, this type of **BiMIP** has been employed as an analytical tool in decision making of real problems. Nevertheless, for **BiMIP** with an MIP lower-level problem, the situation is very different. Although several computing methods are proposed or designed, they are either designed for those with special structures or their implementations involve sophisticated analysis and complicated operations. Moreover, the popular high point problem relaxation is weak, which indicates the associated Branch-and-Bound schemes are less effective. Hence, up to now, there is no commonly accepted algorithm to compute this general type of **BiMIP**. Without support of effective solution methods or computing tools, some researchers consider general **BiMIP** "*still unsolved by the operations research community*" [21]. Consequently, its application in addressing real problems is very restricted.

In the following sections, different from traditional methods that mainly depend on Branch-and-Bound techniques, we develop a new solution scheme based on reformulation and decomposition methods. Specifically, in Section 3 and 4, we introduce a couple of reformulations, and discuss their connections and advantages to the traditional **BiMIP** formulation in (1-3). Then, based on those reformulations, we present a decomposition method and prove its convergence and complexity in Section 5.

3 A Revisit of Bilevel MIP Problem

First, we consider a couple of situations where **BiMIP** can be simplified.

Proposition 1. *1. If high point problem is infeasible,* **BiMIP** *is infeasible.*

2. If $(\mathbf{g}, \mathbf{h}) = \alpha(\mathbf{w}, \mathbf{v})$ *with* $\alpha \leq 0$, **BiMIP** *reduces to high point problem.*

3. If $(\mathbf{g}, \mathbf{h}) = \alpha(\mathbf{w}, \mathbf{v})$ *with* $\alpha > 0$, **BiMIP** *reduces to*

$$\Theta^* = \min_{\mathbf{x}} \mathbf{fx} + \alpha \max_{(\mathbf{y},\mathbf{z}) \in \mathcal{F}^0(\mathbf{x})} \mathbf{wy} + \mathbf{vz}. \tag{11}$$

□

In the remainder of this paper, without loss of generality we assume that high point problem is feasible and Ω is not empty. Nevertheless, it is observed that, even that Ω is not empty, provided that there are discrete decision variables in the lower-level, i.e., n_d is non-zero, **BiMIP** may not have any optimal solution [32]. Under such situation, min should be replaced by inf in (1). To concentrate on designing computing algorithms for **BiMIP**, we assume that it has a finite optimal solution $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ in this paper. We mention that, when inf cannot be reduced to min, the presented research in this paper could be applied to derive an ϵ-optimal solutions, which is often sufficient in solving practical instances.

Most of existing research on solving bilevel MIP directly studies **BiMIP** formulation in (1-3) and seeks to derive critical properties for deeper insights and more support for computational improvements. Instead of following that convention, we duplicate decision variables and constraints of the lower-level problem and provide the following equivalent formulation, to which we denote as $\mathbf{BiMIP_d}$.

$$\begin{aligned}
\mathbf{BiMIP_d}: \ \Theta^* \ = \ & \min \ \mathbf{fx} + \mathbf{gy}^0 + \mathbf{hz}^0 && (12)\\
\text{s.t.} \ \ & \mathbf{Ax} \leq \mathbf{b}, \ \mathbf{x} \in \mathbb{R}_+^{m_c} \times \mathbb{Z}_+^{m_d}, && (13)\\
& \mathbf{Py}^0 + \mathbf{Nz}^0 \leq \mathbf{R} - \mathbf{Kx}, \ \mathbf{y}^0 \in \mathbb{R}_+^{n_c}, \ \mathbf{z}^0 \in \mathbb{Z}_+^{n_d}, && (14)\\
& \mathbf{wy}^0 + \mathbf{vz}^0 \geq \max\{\mathbf{wy} + \mathbf{vz} : \mathbf{Py} + \mathbf{Nz} \leq \mathbf{R} - \mathbf{Kx}, \mathbf{y} \in \mathbb{R}_+^{n_c}, \mathbf{z} \in \mathbb{Z}_+^{n_d}\}. && (15)
\end{aligned}$$

Note that (14-15) ensure that $(\mathbf{y}^0, \mathbf{z}^0)$ is an optimal solution to the lower-level problem for any $\mathbf{x}$. Hence, the equivalence between $\mathbf{BiMIP_d}$ and **BiMIP** is straightforward. Similar to the conventional solution concept of **BiMIP**, we let 3-tuple $(\mathbf{x}, \mathbf{y}^0, \mathbf{z}^0)$ to represent a solution of

$\mathbf{BiMIP_d}$ and point out that an optimal solution to the lower-level problem, with respect to $\mathbf{x}$, can be obtained by setting $(\mathbf{y}, \mathbf{z}) = (\mathbf{y}^0, \mathbf{z}^0)$.

Besides our independent work on this reformulation, we note that this type of formulation was presented in the early 1990s [47] for bilevel linear programming problems to design global optimization algorithms. As more variables and constraints are involved in, it does not appear to be an effective formulation to solve bilevel linear optimization problems. Hence, over the last 20 years, this formulation has received very limited attention. We next present some observations and insights on $\mathbf{BiMIP_d}$, from which we believe that this formulation actually is an informative and convenient representation for analysis and algorithm design for general bilevel MIP.

Remarks:

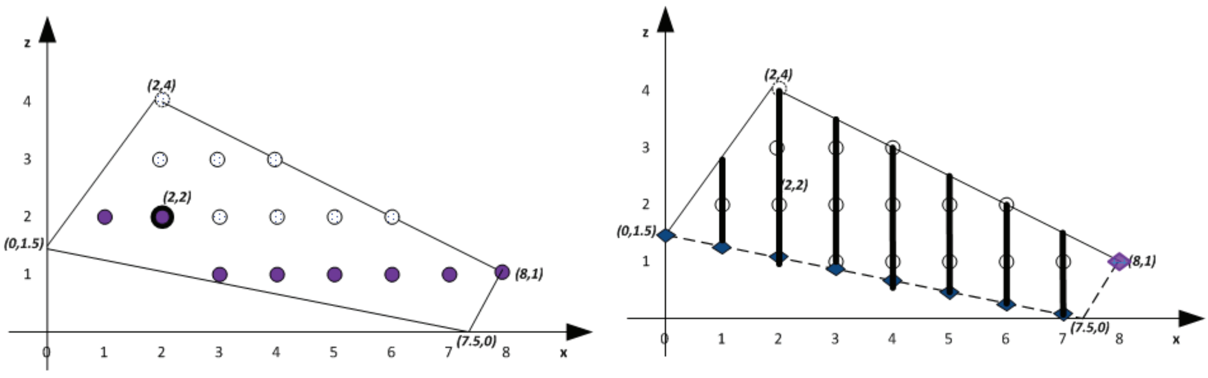
(i) By replicating variables (and constraints), it provides a complete variable set, including the original upper-level variables $\mathbf{x}$ and the lower-level variables, which are now represented by $(\mathbf{y}^0, \mathbf{z}^0)$, at control of the upper-level DM. Mathematically, due to constraint (15), it is clear that the feasible set of $(\mathbf{x}, \mathbf{y}^0, \mathbf{z}^0)$ is Ω^I, the inducible region of the complete bilevel MIP problem. Conceptually, the upper-level DM will be able to use $(\mathbf{y}^0, \mathbf{z}^0)$ to *simulate* the response of the lower-level DM, and to evaluate the impact of that response in her decision scheme. Indeed, they naturally support us to capture further interactions and restrictions between the upper and lower-level solutions. For example, as mentioned in [36], the upper-level constraint (2) may depend on the lower-level solution, which is of the following form where $(\mathbf{y}, \mathbf{z}) \in \mathcal{F}(\mathbf{x})$.

$$\mathbf{Ax} + \mathbf{By} + \mathbf{Cz} \leq \mathbf{b}.$$

For such situation, because of (15), it can be easily reformulated as

$$\mathbf{Ax} + \mathbf{By}^0 + \mathbf{Cz}^0 \leq \mathbf{b},$$

which replaces the corresponding part in (13). We mention that for such type of bilevel MIP problems, our reformulation strategy and decomposition algorithm developed in the following sections work well without modifications. For the structured interdiction model where the upper-level problem has an opposite objective function and does not have any constraint involving lower-level solution, we also mention in Section 4.3 that such variable replication is not necessary.



(a) Ω and Ω^I (b) Inducible regions if $z \in \mathbf{R}_+$

Figure 1: Feasible Sets of Various Problems

(ii) Unlike **BiMIP** that imposes a membership restriction on $(\mathbf{y}, \mathbf{z})$, an inequality constraint (15) on $(\mathbf{y}^0, \mathbf{z}^0)$ would be more friendly to general mathematical programming tools. Indeed, if we ignore (15), or equivalently replace the right-hand-side of (15) by $-\infty$, it is straightforward to see that the resulting formulation is high point problem, a weak relaxation of $\mathbf{BiMIP_d}$ or **BiMIP**. Naturally, if we can strengthen the right-hand-side of (15), a stronger relaxation will be obtained. (iii) Note that bilevel MIP with a lower-level LP can be solved by using its KKT condition based reformulation. Nevertheless, it has been recognized that relaxing the lower-level MIP into an LP does not yield a valid method to solve the original bilevel MIP. Through $\mathbf{BiMIP_d}$ and (15), it is rather easy to identify the actual reason: given that the LP relaxation has an optimal value larger than that of original MIP problem, replacing the right-hand-side of (15) by its LP relaxation will lead to a very different constraint that may cut off a large portion of the bilevel feasible set Ω^I.

Next, we employ an instance presented in Moore and Bard [37] to illustrate our understanding.

Example 1. *(adapted from [37])*
Consider the following bilevel MIP problem represented in the popular **BiMIP** *form.*

$$\min_{x \in \mathbb{Z}_+} -x - 10z \tag{16}$$
$$z \in \arg\max\{-z : -25x + 20z \leq 30,\ x + 2z \leq 10,\ 2x - z \leq 15,\ 2x + 10z \geq 15,\ z \in \mathbb{Z}_+\}.$$

Its corresponding $\mathbf{BiMIP_d}$ *reformulation is*

$$\begin{aligned}
&\min_{(x,z^0)\in\mathbb{Z}^2_+} -x-10z^0 \\
&-25x+20z^0\leq 30,\ x+2z^0\leq 10,\ 2x-z^0\leq 15,\ 2x+10z^0\geq 15, \qquad (17)\\
&-z^0\geq \max\{-z:-25x+20z\leq 30,\ x+2z\leq 10,\ 2x-z\leq 15,\ 2x+10z\geq 15,\ z\in\mathbb{Z}_+\}.
\end{aligned}$$

For this instance, its optimal value is -22 *and its unique optimal solution is* $(x,z)=(2,2)$ *for* **BiMIP** *form (or* $(x,z^0)=(2,2)$ *for* $\mathbf{BiMIP_d}$ *form) .*

In Figure 1, we provide feasible sets of related problems of this instance. The collection of solid dots in Figure 1(a) is the inducible region (also the feasible set of (17)*) of this instance. The collection of all dots in the convex polytope in Figure 1(a), including both solid and empty ones, represents the feasible set of high point problem, i.e., the set* Ω*. Note that high point problem actually is equivalent to* $\mathbf{BiMIP_d}$ *formulation with the last constraint replaced by* $-z^0\geq -\infty$*. Clearly, this high point problem is a relaxation to the bilevel MIP instance. Nevertheless, comparing its optimal value, which is* -42 *from* $(x,z)=(2,4)$*, to that of the bilevel MIP, we observe that this relaxation is weak.*

If we relax the lower-level variable z *in* (16) *to be continuous, which equivalently sets both* z^0 *and* z *in* (17) *to be continuous, the resulting bilevel MIP has its inducible region be the collection of diamond points in Figure 1(b), and its optimal value* -18 *from* $(x,z)=(8,1)$*. Clearly, this inducible region does not have any strong connection with respect to* Ω^I *as their intersection is the single point* $(8,1)$*. A deeper insight can be developed by analyzing* (17) *and Figure 1(b). Note that the collection of bold lines (including all diamond points), which is the feasible set of* (x,z^0) *without considering the optimality constraint between* z^0 *and* z *in* (17)*, contains* Ω^I *as its subset. Nevertheless, if that optimality constraint between* z^0 *and* z *is imposed, because a larger optimal value serves as its right-hand-side, it cuts off most part of that collection, including* Ω^I*, and just leaves diamond points as feasible points.*

Actually, even that we keep z^0 *as discrete and simply relax* z *to be continuous in* (17)*, we cannot derive a good solution either. Under such situation, that optimality constraint between* z^0 *and* z *cuts off almost all integer points of* Ω *and just leaves point* $(8,1)$ *as the only feasible solution.* □

4 Reformulations and Strong Relaxations of Bilevel MIP

4.1 A Single-Level Reformulation and Strong Relaxations

In this subsection, we present a single-level equivalent reformulation of $\mathbf{BiMIP_d}$, which is the basis of our solution scheme. The main idea of this reformulation is to expand (15) through enumeration. We make one assumption that for any possible $(\mathbf{x}, \mathbf{z})$, the remaining lower-level problem has a finite optimal value. Such assumption is similar to the *relatively complete recourse* property, a frequently used concept in stochastic programming literature to ensure the feasibility of the recourse problem under any feasible choice of the first stage decision. Hence, we refer to this assumption as *relatively complete response.*

First, we separate discrete and continuous variables in the lower-level to restructure the right-hand-side of (15) as follows:

$$\mathbf{w}\mathbf{y^0} + \mathbf{v}\mathbf{z^0} \geq \max_{\mathbf{z}\in\mathbf{Z}} \mathbf{v}\mathbf{z} + \max\{\mathbf{w}\mathbf{y} : \mathbf{P}\mathbf{y} \leq \mathbf{R} - \mathbf{K}\mathbf{x} - \mathbf{N}\mathbf{z}, \mathbf{y} \in \mathbb{R}^{n_c}_{+}\} \tag{18}$$

where $\mathbf{Z}$ represents the collection of all possible $\mathbf{z}$. One may think that it is pointless to replace (15) by (18), given that $\mathbf{BiMIP_d}$ becomes even harder. However, as the second maximization problem in (18) is a pure LP, the classical reformulation method using KKT conditions can be applied. Hence, we have its equivalent form

$$\begin{aligned}
&\mathbf{w}\mathbf{y_0} + \mathbf{v}\mathbf{z_0} \geq \max_{\mathbf{z}\in\mathbf{Z}} \mathbf{v}\mathbf{z} + \mathbf{w}\mathbf{y} \\
&\text{s.t. } (\mathbf{y}, \pi) \in \{\mathbf{P}\mathbf{y} \leq \mathbf{R} - \mathbf{K}\mathbf{x} - \mathbf{N}\mathbf{z},\ \mathbf{P}^t\pi \geq \mathbf{w}^t,\ \mathbf{y} \perp (\mathbf{P}^t\pi - \mathbf{w}^t),\ \pi \perp (\mathbf{R} - \mathbf{K}\mathbf{x} - \mathbf{N}\mathbf{z} - \mathbf{P}\mathbf{y})\}.
\end{aligned}$$

Unless explicitly mentioned, we assume that $\mathbf{Z}$ is a finite set such that $\mathbf{Z} = \{\mathbf{z}^1, \ldots, \mathbf{z}^k\}$. We believe that such assumption is very mild in practice. Note that if some components of $\mathbf{z}$ can take very large values, it may not be necessary to treat them as discrete variables and we can consider them as continuous ones. Then, by enumerating $\mathbf{z}^j$ and introducing corresponding variables $(\mathbf{y}^j, \pi^j)$ and their related KKT conditions, we have the following result.

Theorem 2. *The formulation* $\mathbf{BiMIP_d}$ *in (12-15) is equivalent to its expanded single-level for-*

mulation

$$\Sigma_{\mathbf{Z}}: \quad \min \quad \mathbf{fx} + \mathbf{gy}^0 + \mathbf{hz}^0 \tag{19}$$

$$s.t. \quad (13-14)$$

$$\mathbf{wy}^0 + \mathbf{vz}^0 \geq \mathbf{vz}^j + \mathbf{wy}^j, \quad 1 \leq j \leq k \tag{20}$$

$$\mathbf{Py}^j \leq \mathbf{R} - \mathbf{Kx} - \mathbf{Nz}^j, \ \mathbf{P}^t\pi^j \geq \mathbf{w}^t, \quad 1 \leq j \leq k \tag{21}$$

$$\mathbf{y}^j \perp (\mathbf{P}^t\pi^j - \mathbf{w}^t), \ \pi^j \perp (\mathbf{R} - \mathbf{Kx} - \mathbf{Nz}^j - \mathbf{Py}^j), \quad 1 \leq j \leq k \tag{22}$$

$$\mathbf{y}^j \in \mathbb{R}^{n_c}_+, \pi^j \in \mathbb{R}^{n_1}_+, \ 1 \leq j \leq k. \tag{23}$$

Remarks:

(*i*) Note that the aforementioned equivalent reformulation $\Sigma_{\mathbf{Z}}$ is a mathematical program with complementarity constraints. It can be easily converted into a regular MIP using the linearization technique presented in (8), which, therefore, enables us to readily solve bilevel MIP using popular MIP solvers. Or, it can be directly computed by using Branching-and-Bound techniques on complementarity constraints. Various methods [46, 30, 40, 28] for mathematical program with complementarity constraints may be applicable, most of which, however, only deal with continuous problems.

(*ii*) For some bilevel MIP, if its upper-level problem is an MIP and its lower-level problem includes discrete variables, we may not be able to achieve its infimum, for which we need to replace min with inf in (1) and (12) [32]. Although our focus is to develop algorithms for those with optimal solutions, it is interesting to note that the relatively complete response property provides a sufficient condition to ensure the existence of an optimal solution.

Corollary 3. *If a bilevel MIP problem has the relatively complete response property, its optimal solutions exist, which can be obtained by branching on complementarity constraints of* $\Sigma_{\mathbf{Z}}$. □

(*iii*) Similar to well-known Benders Reformulation and Dantzig-Wolfe Reformulation, this equivalent reformulation, however, could be extremely large, which is more of a theoretical value rather than of a practical significance. One idea is to consider some subset of those constraints.

Corollary 4. *Let* $\underline{\mathbf{Z}}$ *be a subset of* $\mathbf{Z}$. *A partial reformulation constructed with respect to* $\underline{\mathbf{Z}}$, *which is denoted by* $\Sigma_{\underline{\mathbf{Z}}}$, *is a relaxation and provides a lower bound to* $\mathbf{BiMIP_d}$. *In particular,* *it is stronger than high point problem by strengthening the right-hand-side of* (15) *through using KKT conditions.* □

It is easy to see that we can always obtain a stronger relaxation and better lower bound by considering a larger $\underline{\mathbf{Z}}$ and computing the corresponding $\Sigma_{\underline{\mathbf{Z}}}$. Hence, it would be beneficial to develop a procedure to dynamically expand $\underline{\mathbf{Z}}$.

4.2 Relatively Complete Response Property and Extended Formulation

For a bilevel MIP problem that does not have the relatively complete response property, there exists some $(\mathbf{x}, \mathbf{z})$ tuple such that the remaining lower-level linear program is infeasible. For such a situation, we can introduce additional variables $\tilde{\mathbf{y}} = (\tilde{y}_1, \ldots, \tilde{y}_{n_1})$ with big-M penalty coefficients for constraint violations. Specifically, we replace (15) in $\mathbf{BiMIP_d}$ with the following constraint where $\mathbf{I}$ is the identity matrix.

$$\mathbf{wy}^0 + \mathbf{vz}^0 \geq \max\{\mathbf{wy} + \mathbf{vz} - M \sum_i \tilde{y}_i : \mathbf{Py} + \mathbf{Nz} \leq \mathbf{R} - \mathbf{Kx} + \mathbf{I}\tilde{\mathbf{y}},\ (\mathbf{y}, \tilde{\mathbf{y}}) \in \mathbb{R}_+^{n_c+n_1}, \mathbf{z} \in \mathbb{Z}_+^{n_d}\} \quad (24)$$

We refer to the formulation $\mathbf{BiMIP_d}$ with constraint (15) replaced by constraint (24) as the extended formulation of the original one.

Proposition 5. (i) *The extended formulation has the relatively complete response property.* (ii) *Assume M is sufficiently large. The extended formulation is a relaxation of the original formulation in (12-15). Moreover, there exists an optimal solution to the extended formulation that is also feasible and optimal to the original one.*

Proof. The first statement is obvious. We provide a proof for the second one. To show that the extended $\mathbf{BiMIP_d}$ formulation with (15) replaced by (24) is a relaxation to the original one, it is sufficient to show that if $(\mathbf{x}, \mathbf{y}^0, \mathbf{z}^0)$ is in the inducible region, i.e., being feasible to the original formulation, it is also feasible to the extended one.

For this $\mathbf{x}$, let $((\mathbf{y}^*, \tilde{\mathbf{y}}^*), \mathbf{z}^*)$ denote an optimal solution to the lower-level problem of the extended formulation as in the right-hand-side of (24). Also, it is easy to see that $((\mathbf{y}^0, \mathbf{0}), \mathbf{z}^0)$ is feasible to the same problem. By contradiction, we assume that

$$\mathbf{wy}^* + \mathbf{vz}^* - M \sum_i \tilde{y}_i^* > \mathbf{wy}^0 + \mathbf{vz}^0 - M \sum_i 0 = \mathbf{wy}^0 + \mathbf{vz}^0. \quad (25)$$

Nevertheless, because M is sufficiently large, unless $\tilde{\mathbf{y}}^* = \mathbf{0}$, (25) will not be valid. When $\tilde{\mathbf{y}}^* = \mathbf{0}$, $(\mathbf{y}^*, \mathbf{z}^*)$ is feasible to the lower-level problem of the original formulation, which, according to (25), is better than $(\mathbf{y}^0, \mathbf{z}^0)$. Hence, in either cases, we have a contradiction. Therefore, we conclude that $((\mathbf{y}^0, \mathbf{0}), \mathbf{z}^0)$ is an optimal solution to the lower-level problem of the extended formulation. Because (24) is satisfied, $(\mathbf{x}, \mathbf{y}^0, \mathbf{z}^0)$ is feasible to the extended formulation.

Furthermore, for any given $\mathbf{x}$, it is valid for any M that

$$\max\{\mathbf{wy} + \mathbf{vz} - M\sum_i \tilde{y}_i : \mathbf{Py} + \mathbf{Nz} \leq \mathbf{R} - \mathbf{Kx} + \mathbf{I}\tilde{\mathbf{y}}, (\mathbf{y}, \tilde{\mathbf{y}}) \in \mathbb{R}_+^{n_c+n_1}, \mathbf{z} \in \mathbb{Z}_+^{n_d}\} \geq$$
$$\max\{\mathbf{wy} + \mathbf{vz} : \mathbf{Py} + \mathbf{Nz} \leq \mathbf{R} - \mathbf{Kx},\ \mathbf{y} \in \mathbb{R}_+^{n_c}, \mathbf{z} \in \mathbb{Z}_+^{n_d}\}. \tag{26}$$

So, for an optimal solution of the extended formulation, it must satisfy (15) of the original formulation, which ensures its feasibility. Because the extended formulation is a relaxation to the original one, its optimality follows. □

Clearly, the extended formulation provides a practical strategy to handle instances without the relatively complete response property. Nevertheless, according to Corollary 3, if bilevel MIP does not have any optimal solution, it can be inferred that there is no finite big-M to ensure validity of the second statement of Proposition 5. Next, we employ one example from [32] to provide an illustration.

Example 2. *(adapted from [32])*
Consider the following bilevel MIP problem

$$\inf \quad x - z \tag{27}$$
$$s.t. \quad 0 \leq x \leq 1,\ z \in \arg\max\Big\{ -z : -z \leq -x, z \in \{0,1\}\Big\}. \tag{28}$$

Its extended formulation is

$$\begin{aligned} \min \quad & x - z^0 \\ s.t. \quad & 0 \leq x \leq 1,\quad -z^0 \leq -x,\ z^0 \in \{0,1\} \\ & -z^0 \geq \max\Big\{ -z - M\tilde{y} : -z \leq -x + \tilde{y}, z \in \{0,1\},\ \tilde{y} \geq 0\Big\}. \end{aligned}$$

Since this extended formulation has the relatively complete response property, the original bilevel MIP in (27-28) is equivalent to the following expanded single-level formulation $\Sigma_{\mathbf{Z}}$ with $\mathbf{Z} = \{0, 1\}$.

$$\begin{aligned}
\Sigma_{\mathbf{Z}}: \quad \min \quad & x - z^0 \\
s.t. \quad & 0 \leq x \leq 1, \ -z^0 \leq -x \\
& -z^0 \geq 0 - M\tilde{y}^1, \ \tilde{y}^1 \geq x, \ \pi^1 \geq -M, \ \tilde{y}^1 \perp (\pi^1 + M), \ \pi^1 \perp (\tilde{y}^1 - x) \\
& -z^0 \geq -1 - M\tilde{y}^2, \ \tilde{y}^2 \geq x - 1, \ \pi^2 \geq -M, \ \tilde{y}^2 \perp (\pi^2 + M), \ \pi^2 \perp (\tilde{y}^2 - x + 1) \\
& z^0 \in \{0, 1\}, \ \tilde{y}^1 \geq 0, \ \tilde{y}^2 \geq 0, \ \pi^1 \leq 0, \ \pi^2 \leq 0
\end{aligned}$$

where π^1 and π^2 are dual variables for the constraint of the lower-level problem for $z = 0$ and 1 respectively.

Certainly, we can solve it numerically for any fixed M. Indeed, it can be analytically derived that there exists an optimal solution with $(x, z^0, \tilde{y}^1, \tilde{y}^2, \pi^1, \pi^2) = (\frac{1}{M}, 1, \frac{1}{M}, 0, -M, 0)$ and the optimal value is $\frac{1}{M} - 1$. $\square$

According to [32], the original bilevel MIP does not have optimal solution and the infimum of the objective function value is -1. Obviously, we can use the extended formulation and derive ϵ-optimal solutions by adjusting the value of M. In the next subsection, by using an alternative reformulation based on strong duality, we present a more straightforward illustration.

4.3 Alternative Reformulations

In addition to using KKT conditions to derive the single-level reformulation, another popular approach is to employ the strong duality theorem of linear programming. Following this line, we next present the strong duality based equivalent reformulation of $\mathbf{BiMIP_d}$. Rewriting the right-hand-side of (18) by strong duality, we have

$$\mathbf{wy^0} + \mathbf{vz^0} \geq \max_{\mathbf{z} \in \mathbf{Z}} \mathbf{vz} + \min\{(\mathbf{R} - \mathbf{Kx} - \mathbf{Nz})^t \pi : \mathbf{P^t} \pi \geq \mathbf{w}^t, \pi \in \mathbb{R}^{n_1}_+\}.$$

We can remove its min operator to obtain the next one.

$$\mathbf{wy^0} + \mathbf{vz^0} \geq \max_{\mathbf{z} \in \mathbf{Z}} \mathbf{vz} + \{(\mathbf{R} - \mathbf{Kx} - \mathbf{Nz})^t \pi : \mathbf{P^t} \pi \geq \mathbf{w}^t, \pi \in \mathbb{R}^{n_1}_+\}.$$

Then, an equivalent formulation, similar to that of Theorem 2, follows easily.

Theorem 6. *The formulation* $\mathbf{BiMIP_d}$ *in (12-15) is equivalent to its expanded single-level formulation*

$$
\begin{aligned}
\Sigma_{\mathbf{Z}}^{d}: \quad \min \quad & \mathbf{fx} + \mathbf{gy}^0 + \mathbf{hz}^0 && (29)\\
s.t. \quad & (13-14) && (30)\\
& \mathbf{wy^0} + \mathbf{vz^0} \geq \mathbf{vz}^j + (\mathbf{R} - \mathbf{Kx} - \mathbf{Nz}^j)^t \pi^j, \quad 1 \leq j \leq k && (31)\\
& \mathbf{P^t} \pi^j \geq \mathbf{w}^t, \quad 1 \leq j \leq k && (32)\\
& \pi^j \in \mathbb{R}_{+}^{n_1}, \ 1 \leq j \leq k. && (33)
\end{aligned}
$$

□

Remarks:

(i) Note that π^j variables for all j are defined by a set of same constraints: $\{\mathbf{P^t}\pi \geq \mathbf{w}^t, \pi \in \mathbb{R}_{+}^{n_1}\}$. Given that a finite optimal solution to **BiMIP** exists, which indicates a particular optimal primal and dual pair $(\mathbf{y}^j, \pi^j)$ exits, the dual feasible set defined by the aforementioned constraints is never empty. Hence, this strong duality based reformulation does not depend on any additional assumptions or property, which is less restrictive than the KKT conditions based reformulation. Actually, we think that it may reveal an essential logic implied in $\mathbf{BiMIP_d}$. Following from the non-empty dual feasible set that, for fixed $(\mathbf{x}, \mathbf{z})$, the remaining lower-level LP is either finitely optimal or infeasible. If the first case occurs, as shown in (31), a non-trivial lower bound, which is parameterized by $\mathbf{x}$, will be available. Otherwise, that lower-bound will become trivial as the right-hand-side of (31) may equal to $-\infty$. Next, we provide an illustration using Example 2.

Example 2. *(continue)*
In order to make use of strong duality for the lower-level problem, we augment the bilevel formulation in Example 2 by introducing a continuous variable y as the following.

$$
\begin{aligned}
\inf \quad & x - z\\
s.t. \quad & 0 \leq x \leq 1,\\
& z \in \arg\max \Big\{ -z + 0y : -z - y \leq -x, \ y \leq 0, \ y \geq 0, z \in \{0, 1\} \Big\}.
\end{aligned}
$$

Note that constraints on y simply force it to be 0. Given $\mathbf{Z} = \{0, 1\}$, its single-level equivalent formulation through strong duality is

$$\begin{aligned}
\inf \quad & x - z^0 && (34)\\
s.t. \quad & 0 \le x \le 1, -z^0 - y^0 \le -x^0,\ y^0 \le 0,\ y^0 \ge 0,\ z^0 \in \{0, 1\} && (35)\\
& -z^0 \ge 0 - \pi^{11} x && (36)\\
& -\pi^{11} + \pi^{12} \ge 0,\ \pi^{11}, \pi^{12} \ge 0 && (37)\\
& -z^0 \ge -1 + \pi^{21}(1 - x) && (38)\\
& -\pi^{21} + \pi^{22} \ge 0,\ \pi^{21}, \pi^{22} \ge 0 && (39)
\end{aligned}$$

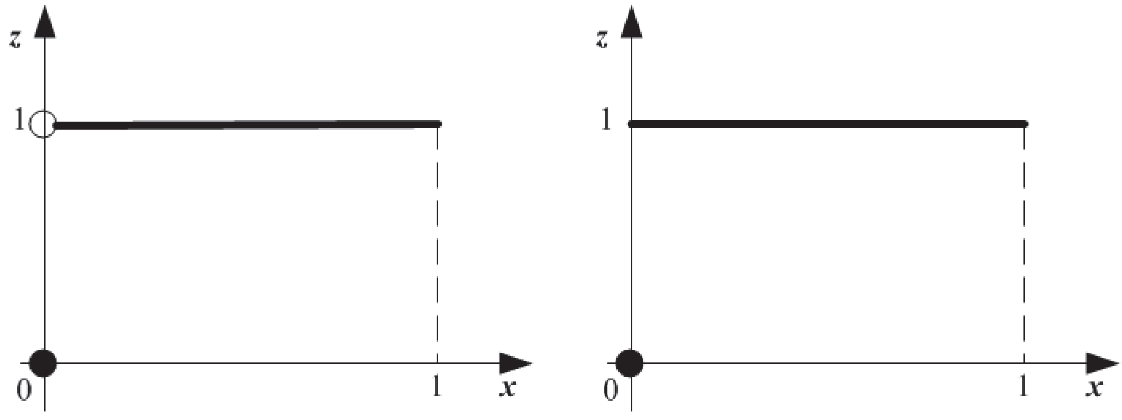
where π^{11} and π^{12} are dual variables for the first and second constraint of the lower-level problem for $z = 0$. Similarly, π^{21} and π^{22} are introduced as dual variables for $z = 1$.

From (36)*, it can be seen that when $x > 0$, the right-hand-side could be $-\infty$, given that π^{11} can be positively unbounded. So, this constraint is trivial. At the same time, from* (38)*, given $x \le 1$ and $\pi^{21} \ge 0$, it can be seen that the inequality can reduce to $-z^0 \ge -1$. Hence, both* (36) *and* (38) *can be trivial, all constraints on dual variables can be removed, and the whole formulation behaves like high point problem. Nevertheless, when $x = 0$, the situation is different. Constraint* (36) *becomes $-z^0 \ge 0$ (and* (38) *can still be trivial), which, together with nonnegativity constraint, leads to $z^0 = 0$.*

According to this discussion, we show the feasible set of (x, z^0) in Figure 2 (i.e., the inducible region Ω^I of the original bilevel MIP), as well as the feasible region Ω of the corresponding high point problem. Note that Ω^I does not include point $(0, 1)$ and it actually is a union of $\{(0, 0)\}$ and $\big\{(0, 1] \times \{1\}\big\}$. As an observation made in [32], without that point, such union is not closed. On the contrary, Ω includes point $(0, 1)$ and becomes closed and bounded, which allows high point problem to achieve its optimal value.

If we bound all dual variables by M in this single-level equivalent formulation, it can be derived that an optimal solution can be obtained by setting (x, z^0)=$(\frac{1}{M}, 1)$ with the optimal value $\frac{1}{M} - 1$, which is also same as those from the extended formulation. □

(ii) Comparing $\Sigma^d_{\mathbf{Z}}$ and the KKT conditions based $\Sigma_{\mathbf{Z}}$ in Theorem 2, it is clear that $\Sigma^d_{\mathbf{Z}}$ is of a simpler structure with less variables and constraints. Nevertheless, bilinear terms between



(a) Ω^I (b) Ω

Figure 2: Feasible Sets

$\mathbf{x}$ and π^j in constraint (31) render $\Sigma^d_{\mathbf{Z}}$ (and its relaxation defined with respect to subset $\underline{\mathbf{Z}}$) a challenging mixed integer nonlinear program. It probably is less friendly than $\Sigma_{\mathbf{Z}}$ to practitioners as the latter one can be easily linearized into an MIP model. For an instance where $\mathbf{x}$ are binary variables, products between $\mathbf{x}$ and π^j can also be linearized easily and its $\Sigma^d_{\mathbf{Z}}$ formulation can be converted into an MIP problem, which could lead to better computational performance than its KKT based one [1].

For the interdiction model presented in (9) and the formulation in (11) , we can also derive their alternative equivalent reformulations that are simpler than $\Sigma_{\mathbf{Z}}$. Next, we give a demonstration using the interdiction model. Because the upper and lower-level DM have completely opposite objective functions, we can just minimize the largest possible objective function value of the lower-level problem, without introducing $(\mathbf{y}^0, \mathbf{z}^0)$ variables and related constraints. The equivalent single-level reformulation based on KKT conditions is presented in the following and the one based on strong duality can be derived similarly.

Corollary 7. *The interdiction model in* (9) *is equivalent to its expanded single-level formulation*

$$\Sigma^I_{\mathbf{Z}} : \min\{\eta : \eta \geq \mathbf{v}\mathbf{z}^j + \mathbf{w}\mathbf{y}^j,\ 1 \leq j \leq k,\ (13),\ (21-23)\}.$$

□

As presented in Corollary 4, it is easy to see that, for any aforementioned single-level equivalent formulation, a partial reformulation constructed with respect to a subset $\underline{\mathbf{Z}}$ leads to a strong relaxation of $\mathbf{BiMIP_d}$. Next, we take advantage of this observation and develop a decomposition algorithm to solve $\mathbf{BiMIP_d}$. In our exposition, we select $\Sigma_{\mathbf{Z}}$ (and its relaxations) as the platform to describe the algorithm development while point out that the whole algorithm strategy works well for alternative reformulations $\Sigma^d_{\mathbf{Z}}$ and $\Sigma^I_{\mathbf{Z}}$.

5 Solving Bilevel MIP Problem by A Decomposition Algorithm

5.1 A Decomposition Algorithm and Computational Complexity

All the results in Section 4, including the single-level equivalent reformulations, the strong relaxations and the associated lower bounds, provide us a basis to design a dynamic solution procedure for $\mathbf{BiMIP_d}$. In particular, by expanding $\underline{\mathbf{Z}}$ and $\Sigma_{\underline{\mathbf{Z}}}$, a tighter relaxation and a stronger lower bound will be available, which also help us find a better feasible solution and a smaller upper bound for $\mathbf{BiMIP_d}$. To this end, we adopt and extend a recent *column-and-constraint generation* method, a master-subproblem computing framework initially developed for two-stage robust optimization problem [53], to solve $\mathbf{BiMIP_d}$. Basically, consider a given (upper-level) solution $\mathbf{x}^*$. By solving the subproblem, which is the lower-level problem, we derive an optimal $(\mathbf{y}^*, \mathbf{z}^*)$. As $(\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}^*)$ is a feasible solution, its value, $\mathbf{fx}^* + \mathbf{gy}^* + \mathbf{hz}^*$, provides an upper bound to $\mathbf{BiMIP_d}$. Then, we update the set $\underline{\mathbf{Z}}$ by including $\mathbf{z}^*$ and expand our master problem $\Sigma_{\underline{\mathbf{Z}}}$. Solving the augmented master problem leads to a new $\mathbf{x}^*$, as well as a stronger lower bound, for a new iteration. We anticipate that, by iteratively solving those master and subproblems, lower and upper bounds converge to the optimal value.

Let UB and LB be the upper and lower bounds respectively, l be the iteration index and ϵ be the optimality tolerance. Next, we provide the implementation details.

The Column-and-Constraint Generation Algorithm for Bilevel MIP

---

1. Set $LB = -\infty$, $UB = +\infty$, and $l = 0$.

2. Solve the following master problem

$$
\begin{aligned}
\mathbf{MP}: \quad \underline{\Theta}^* = \min \quad & \mathbf{fx} + \mathbf{gy}^0 + \mathbf{hz}^0 && (40)\\
\text{s.t.} \quad & (13-14) \\
& \mathbf{wy}^0 + \mathbf{vz}^0 \geq \mathbf{vz}^j + \mathbf{wy}^j, \;\; 1 \leq j \leq l && (41)\\
& \mathbf{Py}^j \leq \mathbf{R} - \mathbf{Kx} - \mathbf{Nz}^j, \; \mathbf{P}^t \pi^j \geq \mathbf{w}^t, \; 1 \leq j \leq l && (42)\\
& \mathbf{y}^j \perp (\mathbf{P}^t \pi^j - \mathbf{w}^t), \; \pi^j \perp (\mathbf{R} - \mathbf{Kx} - \mathbf{Nz}^j - \mathbf{Py}^j), \; 1 \leq j \leq l && (43)\\
& \mathbf{y}^j \in \mathbb{R}_+^{n_c}, \pi^j \in \mathbb{R}_+^{n_1}, \; 1 \leq j \leq l. && (44)
\end{aligned}
$$

Derive an optimal solution $(\mathbf{x}^*, \mathbf{y}^{0*}, \mathbf{z}^{0*}, \mathbf{y}^{1*}, \ldots, \mathbf{y}^{l*}, \pi^{1*}, \ldots, \pi^{l*})$, and update $LB = \underline{\Theta}^*$.

3. If $UB - LB \leq \epsilon$, return UB and the corresponding (incumbent) solution. Terminate. Otherwise, go to Step 4.

4. Solve the following lower-level problem for given $\mathbf{x}^*$, which serves as the first subproblem.

$$\theta(\mathbf{x}^*) = \max\{\mathbf{wy} + \mathbf{vz} : \mathbf{Py} + \mathbf{Nz} \leq \mathbf{R} - \mathbf{Kx}^*, \mathbf{y} \in \mathbb{R}_+^{n_c}, \mathbf{z} \in \mathbb{Z}_+^{n_d}\}.$$

Then, compute the next one, which is the second subproblem.

$$\Theta_o(\mathbf{x}^*) = \min\{\mathbf{gy} + \mathbf{hz} : \mathbf{wy} + \mathbf{vz} \geq \theta(\mathbf{x}^*), \mathbf{Py} + \mathbf{Nz} \leq \mathbf{R} - \mathbf{Kx}^*, \mathbf{y} \in \mathbb{R}_+^{n_c}, \mathbf{z} \in \mathbb{Z}_+^{n_d}\}.$$

Derive an optimal solution $(\mathbf{y}^*, \mathbf{z}^*)$, and update $UB = \min\{UB, \mathbf{fx}^* + \Theta_o(\mathbf{x}^*)\}$.

5. Set $\mathbf{z}^{l+1} = \mathbf{z}^*$, create variables $(\mathbf{y}^{l+1}, \pi^{l+1})$, and add the following constraints

$$
\begin{aligned}
& \mathbf{wy}^0 + \mathbf{vz}^0 \geq \mathbf{vz}^{l+1} + \mathbf{wy}^{l+1}, \\
& \mathbf{Py}^{l+1} \leq \mathbf{R} - \mathbf{Kx} - \mathbf{Nz}^{l+1}, \; \mathbf{P}^t \pi^{l+1} \geq \mathbf{w}^t, \\
& \mathbf{y}^{l+1} \perp (\mathbf{P}^t \pi^{l+1} - \mathbf{w}^t), \; \pi^{l+1} \perp (\mathbf{R} - \mathbf{Kx} - \mathbf{Nz}^{l+1} - \mathbf{Py}^{l+1}), \\
& \mathbf{y}^{l+1} \in \mathbb{R}_+^{n_c}, \pi^{l+1} \in \mathbb{R}_+^{n_1}
\end{aligned}
$$

to **MP**. Set $l = l + 1$, and go to Step 2. $\square$

We point out that the second subproblem in Step 4 is generally necessary. Note that the lower-level problem may have multiple optimal solutions for $\mathbf{x}^*$. By computing the second subproblem, which is constructed by the *lexicographic method*, we will be able to select an optimal solution that is in favor of the upper-level DM, a reflection of the optimistic consideration. Nevertheless, for structured interdiction problems or formulation in the form of (11), because upper-level and lower-level DMs are of opposite interest, computing the second subproblem is not needed. Next, we show that this algorithm converges in finite iterations.

Proposition 8. *Let $\epsilon = 0$ and assume that $\mathbf{Z}$ is finite. The presented column-and-constraint generation algorithm converges to the optimal value of* $\mathbf{BiMIP_d}$ *within $O(|\mathbf{Z}|)$ iterations.*

Proof. Clearly, it is sufficient to show that a repeated $\mathbf{z}^*$ leads to $LB = UB$. Assume that the current iteration index is l_1, $(\mathbf{x}^*, \mathbf{y}^{0*}, \mathbf{z}^{0*})$ is obtained in Step 2 with $LB < UB$, and $\mathbf{z}^*$ is obtained in Step 4. We further assume that $\mathbf{z}^*$ was also derived in some previous iteration $l_0 (< l_1)$.

Because $UB - LB > 0$, as in Step 5, **MP** will be augmented with a set of new variables and constraints associated with $\mathbf{z}^* (= \mathbf{z}^{l_1+1})$. Nevertheless, as those variables and constraints are same as those created and included in iteration l_0, the augmentation essentially does not change MP. So, it yields the same optimal value in iteration $l_1 + 1$ as that of iteration l_1. Hence, LB does not change when the algorithm proceeds from iteration l_1 to $l_1 + 1$.

In the following, we show that $LB \geq UB$ in iteration $l_1 + 1$. Note that $\mathbf{z}^{l_1+1} = \mathbf{z}^*$.

$$
\begin{aligned}
LB &= \mathbf{f}\mathbf{x}^* + \mathbf{g}\mathbf{y}^{0*} + \mathbf{h}\mathbf{z}^{0*} \\
&= \mathbf{f}\mathbf{x}^* + \min\{\mathbf{g}\mathbf{y}^0 + \mathbf{h}\mathbf{z}^0 : (13-14), (41-44), \mathbf{x} = \mathbf{x}^*\} \\
&\geq \mathbf{f}\mathbf{x}^* + \min\{\mathbf{g}\mathbf{y}^0 + \mathbf{h}\mathbf{z}^0 : \mathbf{P}\mathbf{y}^0 + \mathbf{N}\mathbf{z}^0 \leq \mathbf{R} - \mathbf{K}\mathbf{x}^*,\ \mathbf{w}\mathbf{y}^0 + \mathbf{v}\mathbf{z}^0 \geq \mathbf{v}\mathbf{z}^{l_1+1} + \mathbf{w}\mathbf{y}^{l_1+1}, \\
&\qquad\qquad \mathbf{P}\mathbf{y}^{l_1+1} \leq \mathbf{R} - \mathbf{K}\mathbf{x}^* - \mathbf{N}\mathbf{z}^{l_1+1},\ \mathbf{P}^t\pi^{l_1+1} \geq \mathbf{w}^t, \\
&\qquad\qquad \mathbf{y}^{l_1+1} \perp (\mathbf{P}^t\pi^{l_1+1} - \mathbf{w}^t),\ \pi^{l_1+1} \perp (\mathbf{R} - \mathbf{K}\mathbf{x}^* - \mathbf{N}\mathbf{z}^{l_1+1} - \mathbf{P}\mathbf{y}^{l_1+1}), \\
&\qquad\qquad \mathbf{z}^0 \in \mathbb{Z}_+^{n_d},\ \mathbf{y}^0, \mathbf{y}^{l_1+1} \in \mathbb{R}_+^{n_c},\ \pi^{l_1+1} \in \mathbb{R}_+^{n_1}\} \\
&\geq \mathbf{f}\mathbf{x}^* + \min\{\mathbf{g}\mathbf{y}^0 + \mathbf{h}\mathbf{z}^0 : \mathbf{P}\mathbf{y}^0 + \mathbf{N}\mathbf{z}^0 \leq \mathbf{R} - \mathbf{K}\mathbf{x}^*, \mathbf{w}\mathbf{y}^0 + \mathbf{v}\mathbf{z}^0 \geq \theta(\mathbf{x}^*),\ \mathbf{y}^0 \in \mathbb{R}_+^{n_c}, \mathbf{z}^0 \in \mathbb{Z}_+^{n_d}\} \\
&= \mathbf{f}\mathbf{x}^* + \Theta_o(\mathbf{x}^*)
\end{aligned}
$$

The second inequality follows from the fact that $\mathbf{z}^{l_1+1}$ is optimal to $\theta(\mathbf{x}^*)$ and constraints from

KKT conditions ensure that $\mathbf{v}\mathbf{z}^{l_1+1} + \mathbf{w}\mathbf{y}^{l_1+1} = \theta(\mathbf{x}^*)$. Then, in Step 3, it is easy to see that $LB \geq UB$, which terminates the whole algorithm. $\square$

We mention that the actual implementation of the algorithm does not depend on the cardinality of $\mathbf{Z}$, which could be infinite. Provided that a finite optimal solution exists, the algorithm will converge to an optimal solution through finite iterations. Indeed, as shown in Section 6, the algorithm often leads to an optimal solution within a small number of iterations, which could be drastically less than the cardinality of $\mathbf{Z}$. In addition to the convergence and computational complexity, we observe that the whole solution scheme, including single-level equivalent reformulations, has several features that distinguish itself from existing methods.

Remarks:

(*i*) First, the underlying mathematical basis of the solution scheme is our single-level equivalent reformulations, which are very simple. They just involve KKT conditions (and strong duality) and an enumeration of possible discrete values. There is no any sophisticated mathematical theory or concepts, or complicated algorithm operations. The decomposition structure simply reflects the bilevel logic, which does not involve any subjective design or customization. Due to such simple structure and its connection to KKT conditions (and strong duality), we believe that the solution scheme provides a fundamental platform to solve bilevel MIP problems.

(*ii*) Second, the complete decomposition algorithm is easy to implement. Again, master problem **MP** is an MIP with complementarity constraints, which can be converted into a regular MIP by the technique in (8). Hence, both master problem and subproblem can basically be computed by any popular MIP solver, which is conveniently accessible to many researchers and practitioners in practice. Certainly, algorithms or packages [28, 29, 23] specializing in mathematical program with complementarity constraints could bring more computational advantages.

(*iii*) The decomposition algorithm is an open and flexible framework that supports further improvements. As an example, we can make use of domain knowledge to develop fast heuristic procedures for better upper bounds, and take advantage of non-trivial $\mathbf{z}$ identified by those heuristics to derive better lower bounds and therefore to reduce the needed iterations. Also, one bottleneck of this method is to solve master problem **MP**, which grows with iterations and demonstrates a dual block angular structure. Hence, instead of using off-the-shelf solvers, it would be beneficial to develop customized algorithms to make use of such structure for fast computing. Next, we present a computationally efficient enhancement approach to strengthen master problem **MP**, whose validity is straightforward.

Proposition 9. *Let* $\hat{\mathbf{y}}$ *and* $\hat{\pi}$ *represent the primal and dual variables of the lower-level LP corresponding to* $(\mathbf{x}, \mathbf{z}^0)$*. The master problem* **MP** *in Step 2 can be augmented as the following* $\mathbf{MP}_{aug}$ *problem.*

$$\mathbf{MP}_{aug}: \quad \underline{\Theta}^* = \min \quad \mathbf{fx} + \mathbf{gy}^0 + \mathbf{hz}^0 \tag{45}$$

$$s.t. \quad (13-14),\ (41-44)$$

$$\mathbf{wy}^0 + \mathbf{vz}^0 \geq \mathbf{vz}^0 + \mathbf{w}\hat{\mathbf{y}} \tag{46}$$

$$\mathbf{P}\hat{\mathbf{y}} \leq \mathbf{R} - \mathbf{Kx} - \mathbf{Nz}^0,\ \ \mathbf{P}^t\hat{\pi} \geq \mathbf{w}^t \tag{47}$$

$$\hat{\mathbf{y}} \perp (\mathbf{P}^t\hat{\pi} - \mathbf{w}^t),\ \ \hat{\pi} \perp (\mathbf{R} - \mathbf{Kx} - \mathbf{Nz}^0 - \mathbf{P}\hat{\mathbf{y}}) \tag{48}$$

$$\hat{\mathbf{y}} \in \mathbb{R}^{n_c}_+, \hat{\pi} \in \mathbb{R}^{n_1}_+. \tag{49}$$

It has a larger optimal value and therefore produces a stronger lower bound than those of **MP***.* □

It is worth mentioning that, given that both $\mathbf{x}$ and $\mathbf{z}^0$ are variables, $\mathbf{MP}_{aug}$ includes some lower bound information that is parametric not only to $\mathbf{x}$ but also to $\mathbf{z}^0$. As shown in (46), although $\mathbf{z}^0$ might not reflect the optimal response from the lower-level DM towards $\mathbf{x}$, it provides, through $\hat{\mathbf{y}}$ and $\hat{\pi}$, an effective lower bound support to $\mathbf{wy}^0 + \mathbf{vz}^0$ which might not be available from any fixed $\mathbf{z}^1, \ldots, \mathbf{z}^l$. Indeed, we observe in numerical study that the benefit of this enhancement strategy could be very significant. Note that for instance without continuous variable in its lower-level problem, this enhancement strategy is basically ineffective, given that the artificial continuous variables in its extended formulation are penalized with big-M to remove its impact.

In the following subsection, we illustrate our solution scheme using Example 1.

5.2 An Illustration on Example 1

We continue to solve the instance in Example 1 to build a basic understanding on our solution scheme.

Example 1. *(continue:)*

To make the bilevel MIP formulation in (17) *with the relatively complete response property, we* *introduce continuous variables* $(\tilde{y}_1, \tilde{y}_2, \tilde{y}_3, \tilde{y}_4)$ *as in* (24) *for constraints in the lower-level problem. We have*

$$
\begin{aligned}
&\min_{(x,z^0)\in\mathbb{Z}_+^2} \ -x - 10z^0 \\
&-25x + 20z^0 \le 30,\ x + 2z^0 \le 10,\ 2x - z^0 \le 15,\ 2x + 10z^0 \ge 15, \\
&-z^0 \ge \max\{-z - M\sum_{i=1}^{4}\tilde{y}_i : -25x + 20z \le 30 + \tilde{y}_1,\ x + 2z \le 10 + \tilde{y}_2, \\
&\qquad\qquad 2x - z \le 15 + \tilde{y}_3,\ 2x + 10z \ge 15 + \tilde{y}_4,\ z \in \mathbb{Z}_+,\ \tilde{y}_i \in \mathbb{R}_+,\ i = 1, \ldots, 4\}.
\end{aligned}
$$

Next, we provide the detailed algorithm progress over iterations with bound information plotted in Figure 3. In our numerical study, big-M is set to 10,000 *and the computation platform is described in Section 6.*

Iteration $l = 0$*: Solving* ***MP****, which actually is high point problem, we have* $(x^*, z^{0*}) = (2, 4)$*, and* $LB = -42$*. Given* $x = 2$*, solving subproblems, we have* $z^* = 2$ *and* $UB = -22$*.*
Iteration $l = 1$*: Solving* ***MP*** *, we have* $(x^*, z^{0*}) = (6, 2)$*, and* $LB = -26$*. Given* $x = 6$*, solving subproblems, we have* $z^* = 1$ *and* $UB = \min\{-22, -16\} = -22$*.*
Iteration $l = 2$*: Solving* ***MP****, we have* $(x^*, z^{0*}) = (2, 2)$ *and* $LB = -22$*. Because* $LB = UB$*, it terminates with an optimal solution* $(x^*, z^*) = (2, 2)$*, which, according to [37], is optimal to the original bilevel MIP problem in* (16) *(and its equivalence* (17) *).*

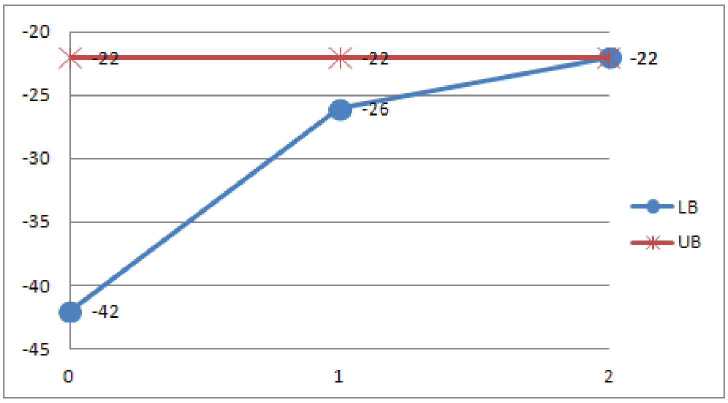


Figure 3: LB & UB vs. Iterations

6 Preliminary Computational Study

In this section, we present a preliminary numerical study on random bilevel MIP instances to evaluate our solution method. Our computational study is made through C++ on a PC desktop (with a single processor at 3GHz and 3.25 G memory), with IBM ILOG CPLEX 12.4 as the MIP solver. We set the optimality tolerance of master problem and the whole algorithm to 0.5%, and those of subproblems to 0.1%, and the computational time limit to 3,600 seconds.

Our random instances are generated according to following specifications. (1) All instances have 20 integer variables in total. Those integer variables are split for the upper-level DM and the lower-level DM. We consider two combinations, i.e., $15 + 5$ and $10 + 10$. (2) Three types of instances are included: a) Upper-level variables, i.e., $\mathbf{x}$, are binary. The lower-level problem has 5 continuous variables. b) Upper-level variables are nonnegative integer variables (bounded by 30). The lower-level problem has 5 continuous variables. c) Upper-level variables are nonnegative integer variables (bounded by 30). The lower-level problem has no continuous variables. (3) Two objective functions are introduced such that one for the upper-level DM, and one for the lower-level DM, where the latter one only involves lower-level variables. (4) As in [37, 22], the lower-level DM is subject to all constraints. Three different sizes are considered: 10, 20 or 30 constraints respectively. Coefficients are randomly chosen in the range of [-50, 50]. (5) To ensure the relatively complete response property, each constraint is associated with an artificial variable whose big-M coefficient is set to 10,000, which is also used for linearizing complementarity constraints.

Overall, there are 18 different combinations. For each of them, we randomly generate 10 instances, with $18\times 10=180$ instances in total. We then compute them using the presented solution scheme, including reformulations and the decomposition method. Except for instances with pure IP in the lower-level, the enhancement strategy presented in Proposition 9 is adopted. For instances with $\mathbf{x}$ being binary, the strong duality based reformulation, as shown in Section 4.3, is employed. Both modifications can lead to a clear reduction in computational time over the standard implementation. Indeed, for the difficult instances that demand for a larger number of iterations in the standard implementation, our enhancement strategy can reduce up to 90% iterations and computational time. The overall computational results, which are averages of 10 random instances of every combination, are reported in Table 1.

From Table 1, we note that:

Table 1: Numerical Study on General Instances

Type	# of Const.	# Int. Var.(up)	# Int. Var.(low)	Iter.	Time (s)	Gap (%)
Bilevel ($\mathbf{x}$ Binary) MIP	10	15	5	2.3	4.19	
		10	10	2.8	9.92	
	20	15	5	2.3	4.07	
		10	10	2.5	5.68	
	30	15	5	2.2	13.25	
		10	10	2.5	15.14	
Bilevel General MIP	10	15	5	2.2	2.68	
		10	10	3.5	427.7	0.3
	20	15	5	2.3	14.59	
		10	10	2.6	39.01	
	30	15	5	2.2	13.79	
		10	10	2.5	37.09	
Bilevel Pure IP	10	15	5	3.2	0.23	
		10	10	67.1	1065.41	3.4
	20	15	5	5.9	1.62	
		10	10	56.7	1094.55	3.0
	30	15	5	24.7	361.15	1.9
		10	10	55.6	1095.06	10.5

(*i*) Our solution method demonstrates a very strong capability, especially for those with an MIP problem in the lower-level. Comparing to existing computational study on similar instances [22, 37], a significant reduction in computational time or optimality gap can be observed. Certainly, due to different computation facilities, such comparison may not be fair. It is worth pointing out that an optimal solution can often be derived after a couple of iterations within the column-and-constraint generation algorithm. Those small numbers indicate that our method could compute bilevel MIP by solving just several (single-level) MIP problems (with complementarity constraints). We believe that such observation strongly confirms the effectiveness of our solution approach and its distinct advantages over existing methods.

(*ii*) As we expect, discrete variables in the lower-level clearly increase the computational burden, especially for those bilevel pure IP problems. Nevertheless, continuous variables and corresponding constraints in the lower-level MIP could be very helpful. When the number of constraints increases, the number of iterations or computational time actually do not have a clear increase. One explanation is that the rich primal and dual information from KKT conditions (or strong duality) helps the algorithm quickly identify critical values for discrete lower-level variables and hence leads to a fast convergence. This observation suggests that the presented algorithm probably is scalable to deal with practical instances.

(*iii*) Bilevel pure integer programming problem is very difficult to compute. Unlike the situation where computing those with mixed variables in the lower-level only needs a couple of iterations, it is often the case that a much larger number of iterations is incurred and a nontrivial gap exists before time limit. Such result is probably mainly due to the lack of support from KKT conditions. Hence, we believe that advanced Branch-and-Bound methods and other strategies are necessary for further improvement, which are our next research tasks.

We mention that a structured interdiction problem, which arises from power system research and includes binary variables in both levels [21], is solved by a multi-start Benders decomposition based heuristic method in [21]. By using the column-and-constraint generation method through the strong duality based reformulation [54], exact solutions generally can be obtained within a marginal computational time, compared to that Benders decomposition based heuristic. Also, a similar power grid protection problem [14, 51] is computed by a Branch-and-Bound method developed for bilevel MIP problem in [50], which produces an optimal solution with more than 100 seconds. Indeed, by directly applying the column-and-constraint generation method [52], we note that it can be computed within 2 seconds after a couple of iterations. Those comparisons, together with observations in Table 1, support that the presented solution approach is a very promising and general method to compute the challenging bilevel MIPs. Certainly, sophisticated enhancement strategies should be studied, evaluated and implemented to further strengthen its capacity as a practical tool.

7 Conclusions

In this paper, we study the challenging bilevel MIP problem and present a novel computing scheme based on reformulations and decomposition strategies. Note that it has several features that distinguish itself from existing methods: (1) The set of single-level reformulations provide a standard platform that is friendly to perform analytical study and build insights. (2) The decomposition algorithm based on column-and-constraint generation method can minimize the impact of enumeration and lends itself to a high computational efficiency. (3) The whole scheme can be easily implemented using popular MIP solvers, without involving any complicated algorithm operations or procedures. Indeed, it is probably the first algorithm that does not depend on customized Branch-and-Bound techniques to solve general bilevel MIP problems. Together with its superior computational performance over existing methods, we believe that this solution method makes an important progress in addressing the "unsolved" bilevel MIP problem.

We mention that this computing scheme provides a foundation to develop more efficient solution methods. As discussed in Section 5, future research directions include designing fast heuristics for better lower and upper bounds and developing customized algorithms to explore the structure of master problem **MP** (and $\mathbf{MP}_{aug}$) for fast computing. In addition, we observe that this computing scheme may not yield exact solutions in reasonable time, especially for instances with pure integer programs as their lower-level problems. Hence, enhancement strategies specializing in discrete structures, such as advanced Branch-and-Bound techniques and strong valid inequalities, should be investigated and integrated in the future study.

References

[1] José Manuel Arroyo. Bilevel programming applied to power system vulnerability analysis under multiple contingencies. *Generation, Transmission & Distribution, IET*, 4(2):178–190, 2010.

[2] José Manuel Arroyo and Francisco Galiana. On the solution of the bilevel programming formulation of the terrorist threat problem. *Power Systems, IEEE Transactions on*, 20(2):789–797, 2005.

[3] Charles Audet, Pierre Hansen, Brigitte Jaumard, and Gilles Savard. Links between linear bilevel and mixed 0–1 programming problems. *Journal of Optimization Theory and Applications*, 93(2):273–300, 1997.

[4] Jonathan F Bard. An efficient point algorithm for a linear two-stage optimization problem. *Operations Research*, 31(4):670–684, 1983.

[5] Jonathan F Bard. *Practical bilevel optimization: algorithms and applications*. Springer, 1998.

[6] Jonathan F Bard and James E Falk. An explicit solution to the multi-level programming problem. *Computers & Operations Research*, 9(1):77–100, 1982.

[7] Jonathan F Bard and James T Moore. A branch and bound algorithm for the bilevel programming problem. *SIAM Journal on Scientific and Statistical Computing*, 11(2):281–292, 1990.

[8] Jonathan F Bard, John Plummer, and Jean Claude Sourie. A bilevel programming approach to determining tax credits for biofuel production. *European Journal of Operational Research*, 120(1):30–46, 2000.

[9] Wayne Bialas and Mark Karwan. On two-level optimization. *Automatic Control, IEEE Transactions on*, 27(1):211–214, 1982.

[10] Daniel Bienstock and Abhinav Verma. The $N-k$ problem in power grids: New models, formulations, and numerical experiments. *SIAM Journal on Optimization*, 20(5):2352–2380, 2010.

[11] Jerome Bracken and James T McGill. Mathematical programs with optimization problems in the constraints. *Operations Research*, 21(1):37–44, 1973.

[12] Luce Brotcorne, Martine Labbé, Patrice Marcotte, and Gilles Savard. A bilevel model for toll optimization on a multicommodity transportation network. *Transportation Science*, 35(4):345–358, 2001.

[13] Luce Brotcorne, Martine Labbé, Patrice Marcotte, and Gilles Savard. Joint design and pricing on a network. *Operations Research*, 56(5):1104–1115, 2008.

[14] Gerald G Brown, Matthew Carlyle, Javier Salmeron, and Kevin Wood. Analyzing the vulnerability of critical infrastructure to attack and planning defenses. In *Tutorials in Operations Research. INFORMS*, pages 102–123. INFORMS, 2005.

[15] Anthony P Burgard, Priti Pharkya, and Costas D Maranas. Optknock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnology and bioengineering*, 84(6):647–657, 2003.

[16] Wilfred Candler and Roger Norton. Multilevel programming and development policy. Technical Report 258, World Bank Staff, Washington, D.C., 1977.

[17] Wilfred Candler and Robert Townsley. A linear two-level programming problem. *Computers & Operations Research*, 9(1):59–76, 1982.

[18] RG Cassidy, MJL Kirby, and WM Raike. Efficient distribution of resources through three levels of government. *Management Science*, 17(8):462–473, 1971.

[19] Benoît Colson, Patrice Marcotte, and Gilles Savard. Bilevel programming: A survey. *4OR*, 3(2):87–107, 2005.

[20] Jean-Philippe Côté, Patrice Marcotte, and Gilles Savard. A bilevel modelling approach to pricing and fare optimisation in the airline industry. *Journal of Revenue and Pricing Management*, 2(1):23–36, 2003.

[21] Andrés Delgadillo, José Manuel Arroyo, and Natalia Alguacil. Analysis of electric grid interdiction with line switching. *Power Systems, IEEE Transactions on*, 25(2):633–641, 2010.

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

[23] Steven P Dirkse and Michael C Ferris. The path solver: a nommonotone stabilization scheme for mixed complementarity problems. *Optimization Methods and Software*, 5(2):123–156, 1995.

[24] José Fortuny-Amat and Bruce McCarl. A representation and economic interpretation of a two-level programming problem. *Journal of the Operational Research Society*, 32(9):783–792, 1981.

[25] Antonio Frangioni. On a new class of bilevel programming problems and its use for reformulating mixed integer problems. *European Journal of Operational Research*, 82(3):615–646, 1995.

[26] Pierre Hansen, Brigitte Jaumard, and Gilles Savard. New branch-and-bound rules for linear bilevel programming. *SIAM Journal on Scientific and Statistical Computing*, 13(5):1194–1217, 1992.

[27] Benjamin F Hobbs, Carolyn B Metzler, and Jong-Shi Pang. Strategic gaming analysis for electric power systems: An MPEC approach. *Power Systems, IEEE Transactions on*, 15(2):638–645, 2000.

[28] Jing Hu, John E Mitchell, Jong-Shi Pang, Kristin P Bennett, and Gautam Kunapuli. On the global solution of linear programs with linear complementarity constraints. *SIAM Journal on Optimization*, 19(1):445–471, 2008.

[29] Jing Hu, John E Mitchell, Jong-Shi Pang, and Bin Yu. On linear programs with linear complementarity constraints. *Journal of Global Optimization*, 53(1):29–51, 2012.

[30] Xinmin Hu and Daniel Ralph. Convergence of a penalty method for mathematical programming with complementarity constraints. *Journal of Optimization Theory and Applications*, 123(2):365–390, 2004.

[31] Shan Jin and Sara M Ryan. A tri-level model of centralized transmission and decentralized generation expansion planning for an electricity market - part I. *Power Systems, IEEE Transactions on*, 29(1):132–141, 2014.

[32] Matthias Köppe, Maurice Queyranne, and Christopher Thomas Ryan. Parametric integer programming algorithm for bilevel mixed integer programs. *Journal of Optimization Theory and Applications*, 146(1):137–150, 2010.

[33] Churlzu Lim and J Cole Smith. Algorithms for discrete and continuous multicommodity flow network interdiction problems. *IIE Transactions*, 39(1):15–26, 2007.

[34] Pierre Loridan and Jacqueline Morgan. Weak via strong Stackelberg problem: New results. *Journal of Global Optimization*, 8(3):263–287, 1996.

[35] Patrice Marcotte. Network design problem with congestion effects: A case of bilevel programming. *Mathematical Programming*, 34(2):142–162, 1986.

[36] Ayalew Getachew Mersha and Stephan Dempe. Linear bilevel programming with upper level constraints depending on the lower level solution. *Applied mathematics and computation*, 180(1):247–254, 2006.

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

[38] Alexis L Motto, José Manuel Arroyo, and Francisco D Galiana. A mixed-integer LP procedure for the analysis of electric grid security under disruptive threat. *Power Systems, IEEE Transactions on*, 20(3):1357–1365, 2005.

[39] George P Papavassilopoulos. Algorithms for static Stackelberg games with linear costs and polyhedra constraints. In *Decision and Control, 1982 21st IEEE Conference on*, volume 21, pages 647–652. IEEE, 1982.

[40] Arvind U Raghunathan and Lorenz T Biegler. An interior point method for mathematical programs with complementarity constraints (MPCCs). *SIAM Journal on Optimization*, 15(3):720–750, 2005.

[41] Shaogang Ren, Bo Zeng, and Xiaoning Qian. Adaptive bi-level programming for optimal gene knockouts for targeted overproduction under phenotypic constraints. *BMC Bioinformatics*, 14(Suppl 2):S17, 2013.

[42] Helena Sofia Rodrigues and M Teresa T Monteiro. Solving mathematical programs with complementarity constraints with nonlinear solvers. In *Recent Advances in Optimization*, pages 415–424. Springer, 2006.

[43] Carlos Ruiz and Antonio J Conejo. Pool strategy of a producer with endogenous formation of locational marginal prices. *Power Systems, IEEE Transactions on*, 24(4):1855–1866, 2009.

[44] Javier Salmeron, Kevin Wood, and Ross Baldick. Worst-case interdiction analysis of large-scale electric power grids. *Power Systems, IEEE Transactions on*, 24(1):96–104, 2009.

[45] Maria P Scaparra and Richard L Church. A bilevel mixed-integer program for critical infrastructure protection planning. *Computers & Operations Research*, 35(6):1905–1923, 2008.

[46] Holger Scheel and Stefan Scholtes. Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity. *Mathematics of Operations Research*, 25(1):1–22, 2000.

[47] Hoang Tuy, Athanasios Migdalas, and Peter Värbrand. A global optimization approach for the linear two-level program. *Journal of Global Optimization*, 3(1):1–23, 1993.

[48] Ue-Pyng Wen and YH Yang. Algorithms for solving the mixed integer two-level linear programming problem. *Computers & Operations Research*, 17(2):133–142, 1990.

[49] Pan Xu. *Three essays on bilevel optimization algorithms and applications*. PhD thesis, Iowa State University, 2012.

[50] Pan Xu and Lizhi Wang. An exact algorithm for the bilevel mixed integer linear programming problem under three simplifying assumptions. *Computers & Operations Research*, 41(1):309–318, 2014.

[51] Yiming Yao, Thomas Edmunds, Dimitri Papageorgiou, and Rogelio Alvarez. Trilevel optimization in power network defense. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 37(4):712–718, 2007.

[52] Wei Yuan, Long Zhao, and Bo Zeng. Optimal power grid protection through a defender–attacker–defender model. *Reliability Engineering & System Safety*, 121:83–89, 2014.

[53] Bo Zeng and Long Zhao. Solving two-stage robust optimization problems using a column-and-constraint generation method. *Operations Research Letters*, 41(5):457–461, 2013.

[54] Long Zhao and Bo Zeng. Vulnerability analysis of power grids with line switching. *Power Systems, IEEE Transactions on*, 28(3):2727–2736, 2013.