A Dantzig–Wolfe Single-Level Reformulation for Mixed-Integer Linear Bilevel Optimization: Exact and Heuristic Approaches

Henri Lefebvre, Martin Schmidt

Abstract. Bilevel optimization problems arise in numerous real-world applications. While single-level reformulations are a common strategy for solving convex bilevel problems, such approaches usually fail when the follower's problem includes integer variables. In this paper, we present the first single-level reformulation for mixed-integer linear bilevel optimization, which does not rely on the follower's value function. Our approach is based on convexifying the follower's problem via a Dantzig–Wolfe reformulation and exploits strong duality of the reformulated problem. By doing so, we derive a nonlinear single-level problem, which is equivalent to the original bilevel model. Moreover, we show that this problem can be transformed into a mixed-integer linear problem using standard linearization techniques and bounds on the dual variables of the convexified follower's problem. Notably, we show that these bounds can be computed in practice via a polynomial-time-solvable problem, which is purely based on the primal problem's data. This results in a new branch-and-cut approach for mixed-integer linear bilevel optimization. In addition to this exact solution approach, we also present a penalty alternating direction method, which computes high-quality feasible points. Numerical experiments on instances from the BOBILib show that we are able to improve the best known solution of 427 instances out of a test set of 2216 instances.

1. Introduction

In the last decades, bilevel optimization has emerged as a prominent tool to tackle decision-making problems in which a hierarchical decision-making process is at stake. Originally introduced by von Stackelberg (1934) and von Stackelberg (1952) in the context of game theory, bilevel optimization has found many applications in fields such as the energy sector, logistics, or strategic defense. In short, bilevel problems model a situation in which a decision-maker, the leader, makes a decision while anticipating the optimal reaction of another decision-maker, the follower.

Despite its high practical relevance, bilevel optimization presents significant challenges both from a theoretical and a practical viewpoint. Even in the most simple case in which all decision-makers solve a linear optimization problem, Jeroslow (1985) showed that the resulting bilevel problem is NP-hard. This result was strengthened by Hansen et al. (1992) who showed strong NP-hardness and by Vicente et al. (1994) who showed that even checking local optimality of a given leader decision is NP-hard. More recently, Buchheim (2023) established that linear bilevel problems belong to NP. However, it is shown in Lodi et al. (2013) that bilevel problems in which the follower solves a mixed-integer linear problem are Σ_2^P-hard in general.

A classic approach to solve bilevel problems is to derive a single-level reformulation. If the follower's problem is convex and satisfies standard regularity conditions,

Date: June 12, 2025.

2020 *Mathematics Subject Classification.* 90-XX, 90-08, 90C11, 90C46.

Key words and phrases. Mixed-integer linear bilevel optimization, Penalty methods, Alternating direction methods, Primal heuristics, Single-level reformulations.

this can often be achieved using the Karush–Kuhn–Tucker (KKT) optimality conditions or strong duality. The resulting single-level problem then includes general bilinearities, which can be linearized under certain conditions (Zare et al. 2017), or complementarity constraints, which can be linearized using auxiliary binary variables (Fortuny-Amat and McCarl 1981). To do so, valid bounds on the follower's dual variables are necessary. Buchheim (2023) showed that valid bounds on the dual variables can be computed in polynomial time for general LP-LP bilevel problems. Unfortunately, they are too large to be useful in practice. Moreover, Kleinert et al. (2020) showed that computing tight valid bounds is as hard as solving the original bilevel problem. For this reason, a common approach in linear bilevel optimization is to compute valid bounds, which are tied to a specific application and use them for linearization. Alternatively, the complementarity constraints can also be encoded using special ordered sets of type 1 (SOS1); see Kleinert and Schmidt (2023) and Aussel et al. (2024) for a detailed discussion.

In the more general case in which the follower solves a mixed-integer linear problem (MILP), these classic optimality-based reformulations are no longer applicable. While it is still theoretically possible to reformulate the problem using the value function of the follower, this typically leads to an intractable problem. Indeed, the value function of a given mixed-integer linear problem is known to be nonconvex, discontinuous and has no closed-form expression in general (Hassanzadeh and Ralphs 2014). As a result, most existing computational methods rely on branch-and-cut techniques that avoid explicit modeling of the value function. Instead, these methods reconstruct the value function only locally once a leader's decision is fixed or exploit local information to derive cutting planes to speed up the solution process; see Fischetti et al. (2017a,b), Tahernejad et al. (2020), and Xu and Wang (2014).

In this paper, we consider general mixed-integer linear bilevel optimization problems and contribute to the literature in the following ways.

(i) We introduce the first single-level reformulation that does not use the value function of the follower's problem for general mixed-integer linear bilevel optimization problems. Our approach is based on a Dantzig–Wolfe reformulation that allows us to convexify the follower's problem. We then exploit strong duality of this reformulated problem to derive a single-level nonlinear problem (NLP) that is equivalent to the original bilevel model. To the best of our knowledge, this is the first application of a Dantzig–Wolfe reformulation in a bilevel context.

(ii) We show that this new NLP reformulation can be converted into a MILP (with exponentially many constraints) by exploiting bounds on the dual variables of the convexified follower's problem. Crucially, we show that such bounds can be computed by solving a polynomial-time solvable problem solely based on the data of the follower's primal problem. This is an important difference to methods for computing dual bounds that rely on the follower's dual problem, which usually suffer from unboundedness issues.

(iii) We derive a new exact approach for general mixed-integer linear bilevel problems based on the MILP single-level reformulation. The resulting algorithm is a branch-and-cut method that can be easily implemented in most state-of-the-art mixed-integer linear solvers using standard cut-generation callback routines.

(iv) We further exploit the single-level reformulation to develop a heuristic approach based on the penalty alternating direction method (Geißler et al. 2017). We first describe the algorithm and then further present a novel interpretation of this method in the context of bilevel optimization.

(v) Finally, we report very encouraging numerical results and show that the developed methods improve the best known solution for 427 instances over the 2216 instances from the BOBILib (Thürauf et al. 2024) that fulfill our assumptions. In particular, we found feasible points for 66 instances for which no general-purpose bilevel solver was able so far to find one.

The remainder of this paper is organized as follows. In Section 2, we define the problem setting and introduce the required notation. Section 3 recalls some standard results on linear bilevel optimization as well as exact and heuristic approaches. In Section 4, we present the main contribution of this paper, which is the first single-level reformulation of general mixed-integer linear bilevel optimization problems that is not relying on the value function of the follower's problem. We show in Section 5 how this reformulation can be solved using standard MILP techniques. A heuristic approach to solve this reformulation is presented in Section 6. We conclude the theoretical part in Section 7, which deals with interdiction problems and shows how our results can be tailored to this class of problems. Finally, Section 8 presents the numerical results on instances taken from the BOBILib (Thürauf et al. 2024).

2. Problem Statement

We consider optimistic bilevel problems of the form

$$\min_{x \in X, y} \quad c^\top x + d^\top y \tag{1a}$$

$$\text{s.t.} \quad Ax + By \geq a, \tag{1b}$$

$$y \in \operatorname*{arg\,min}_{y' \in Y} \left\{ f^\top y' \colon Cx + Dy' \geq b \right\}. \tag{1c}$$

We refer to x as the leader's decision and to y as the follower's decision. Constraints in (1b) that explicitly depend on the follower's decision are referred to as coupling constraints. Their use in bilevel optimization has been recently discussed in Henke et al. (2024, 2025). Here, vectors $a \in \mathbb{Q}^{m_x}$, $b \in \mathbb{Q}^{m_y}$, $c \in \mathbb{Q}^{n_x}$, $d \in \mathbb{Q}^{n_y}$, $f \in \mathbb{Q}^{n_y}$ as well as matrices $A \in \mathbb{Q}^{m_x \times n_x}$, $B \in \mathbb{Q}^{m_x \times n_y}$, $C \in \mathbb{Q}^{m_y \times n_x}$, and $D \in \mathbb{Q}^{m_y \times n_y}$ are given as input. The sets X and Y are also given and used to impose further restrictions on the leader's and follower's decision such as integrality constraints. If $X = \mathbb{R}^{n_x}$ and $Y = \mathbb{R}^{n_y}$, we say that Problem (1) is a linear bilevel problem.

Throughout the paper, we denote by $\operatorname{conv}(S)$ the convex hull of a set S, i.e., the set of points that can be expressed as a convex combination of points in S. When clear from the context, we use $\operatorname{proj}_x(S)$ to denote the projection of the set S onto the x variables. For a given function $f : \mathcal{X} \to \mathcal{Y}$, we denote by f^* its convex conjugate, i.e.,

$$f^*(y) = \max \left\{ y^\top x - f(x) \colon x \in \mathcal{X} \right\}.$$

We also recall the Fenchel inequality $y^\top x \leq f(x) + f^*(y)$ for all x and y. Equality holds in the previous inequality if and only if $y \in \partial f(x)$, where $\partial f(x)$ denotes the set of all subgradients of f at point x.

3. Continuous and Linear Lower-Level Problems

In this section, we discuss some standard techniques for tackling bilevel optimization problems in which the follower's problem is linear and continuous. To this end, we temporarily assume that $Y = \mathbb{R}^{n_y}$. Additionally, we let φ denote the value function of the follower's problem, i.e., we let

$$\varphi(x) := \min_{y \in Y} \left\{ f^\top y \colon Cx + Dy \geq b \right\}.$$

A well-known reformulation of the bilevel Problem (1) is the so-called value-function reformulation given by

$$\min_{x\in X, y\in Y} \quad c^\top x + d^\top y \tag{2a}$$
$$\text{s.t.} \quad Ax + By \geq a, \tag{2b}$$
$$Cx + Dy \geq b, \tag{2c}$$
$$f^\top y \leq \varphi(x). \tag{2d}$$

Here, the two additional Constraints (2c) and (2d) serve two purposes. First, it requires that $y \in Y$ is a feasible decision of the follower. Second, it ensures that it is chosen so that its associated follower's objective function value is not worse than the follower's optimal objective function value $\varphi(x)$, i.e., y is indeed an optimal decision of the follower for the given x. In practice, the value-function reformulation (2) remains challenging to solve, because φ is a convex function that appears on the right-hand side of a less-than-or-equal-to constraint. Hence, the optimality constraint "$f^\top y \leq \varphi(x)$" is a nonconvex constraint. Moreover, no closed-form expression of φ is known in general. To circumvent this, linear optimization duality can be used. Let x be a leader's decision such that the follower's problem is feasible, i.e., let x be such that there exists a vector y satisfying $Cx + Dy \geq b$. Strong duality implies that

$$\varphi(x) = \max_{\lambda\in\mathbb{R}^{m_y}} \left\{ (b - Cx)^\top \lambda \colon D^\top \lambda = f,\ \lambda \geq 0 \right\}. \tag{3}$$

Here, the maximization problem is the dual of the follower's (primal) problem. Using (3) leads to the so-called strong-duality reformulation of (1):

$$\begin{aligned} \min_{x\in X, y\in Y, \lambda} \quad & c^\top x + d^\top y \\ \text{s.t.} \quad & Ax + By \geq a, \\ & Cx + Dy \geq b, \quad f^\top y \leq (b - Cx)^\top \lambda, \\ & D^\top \lambda = f, \quad \lambda \geq 0. \end{aligned} \tag{4}$$

This problem is a nonconvex but single-level optimization problem. Moreover, the only nonlinearities are in the optimality condition and only involve products of variables. Hence, standard optimization techniques and software can be used to tackle Problem (4) with the property that if (x^*, y^*, λ^*) is a global solution to (4), then (x^*, y^*) is a global solution to (1). A re-writing of Problem (4) leads to the so-called KKT reformulation, which can also be obtained by replacing the follower's problem by its KKT optimality conditions. Indeed, exploiting the equality "$D^\top \lambda = f$", one can show that

$$0 \leq (b - Cx - Dy)^\top \lambda \iff \lambda \perp (b - Cx - Dy)$$

holds. For more details, we refer to Fortuny-Amat and McCarl (1981) and to the book by Dempe (2002) for an overview over existing techniques in linear bilevel optimization.

Remark 1. *In the above discussion, we have shown how the nonconvex strong-duality reformulation (4) can be reformulated as a bilinear optimization problem. To achieve this goal, we explicitly re-wrote the value-function φ as the maximum of affine functions. In this case, this representation is given by the dual of the follower's problem, i.e., by $\varphi(x) = \max_\lambda \left\{ (b - Cx)^\top \lambda \colon D^\top \lambda = f, \lambda \geq 0 \right\}$. More generally, any convex function f can be represented as $\max_\lambda \left\{ \lambda^\top x - f^*(\lambda) \right\}$ with f^* being the convex conjugate of f; see Theorem 12.2 from Rockafellar (1970). This representation allows us to transform a general nonconvex constraint like "$a^\top x \leq f(x)$" with a given convex function f as a constraint in which the only nonconvex terms are products of* *variables, i.e., constraints of the form: there exists* λ *such that* $a^\top x + f^*(\lambda) \le \lambda^\top x$. *This key observation is later used in Section 4 in the context of mixed-integer linear bilevel problems.*

While Problem (4) can be solved by standard approaches from the nonconvex optimization literature, it remains an NP-hard problem as Problem (1) is. Hence, solely considering global solution approaches may not be reasonable for large-scale problems. A heuristic approach that is based on the strong-duality reformulation (4) is proposed by Kleinert and Schmidt (2021) and has been recently extended by Lefebvre and Schmidt (2024). The idea is to decompose Problem (4) into two blocks: a primal block (x and y) and a dual block (λ). Then, the method alternatively solves a sub-problem over a given block, leaving the variables of the other block fixed to a current estimate. To avoid any infeasibility during the process and to speed-up the convergence of the method, the optimality condition "$f^\top y \le (b - Cx)^\top \lambda$" is moved to the objective function by means of a penalty function. This is motivated by the fact that this is the only constraint coupling the two blocks and the only constraint involving nonlinearities. Then, the penalty parameter is sequentially updated to ensure feasibility in the limit.

More formally, given a penalty parameter $\rho > 0$ and a current dual estimate $\bar\lambda$, the first sub-problem to be solved reads

$$\begin{aligned} \min_{x \in X, y \in Y} \quad & c^\top x + d^\top y + \rho(b^\top \bar\lambda - \bar\lambda^\top Cx - f^\top y) \\ \text{s.t.} \quad & Ax + By \ge a, \quad Cx + Dy \ge b. \end{aligned} \tag{5}$$

Note that if $\bar\lambda$ is dual feasible, i.e., if $\bar\lambda \ge 0$ and $D^\top \bar\lambda = f$, and if the penalty term "$b^\top \bar\lambda - \bar\lambda^\top Cx - f^\top y$" is zero for a given (x, y) solving (5), then (x, y) is a feasible point of (4). Otherwise, the second sub-problem is solved, i.e., given a solution $(\bar x, \bar y)$ to the first sub-problem, we solve

$$\begin{aligned} \min_{\lambda} \quad & (b - C\bar x)^\top \lambda \\ \text{s.t.} \quad & D^\top \lambda = f, \quad \lambda \ge 0. \end{aligned}$$

The process is then iterated until either a feasible solution is obtained or the algorithm stagnates. In the latter case, the penalty parameter ρ is updated to favor feasibility. It is known that feasible points generated by this procedure are stationary points of the single-level Problem (4). However, stationary points of the single-level problem do not need to be stationary points of the original bilevel Problem (1). Nevertheless, computational experiments show that the procedure typically produces high-quality feasible points in practice. For more details, we refer to Kleinert and Schmidt (2021) and Lefebvre and Schmidt (2024).

4. A Single-Level Reformulation for Mixed-integer Problems

We now consider the setting in which Problem (1) is a mixed-integer linear bilevel problem. To this end, we make the following assumption that holds for the rest of the paper.

Assumption 1. *It holds* $X = \{0,1\}^{n_x}$ *and* $Y = \mathbb{Z}^{n_y - p_y} \times \mathbb{R}^{p_y}$. *Moreover, the follower's feasible region* $\{y \in Y : Cx + Dy \ge b\}$ *is bounded for all* $x \in X$.

Note that Assumption 1 is a standard assumption in mixed-integer linear bilevel optimization. In particular, it ensures that a solution to (1) exists. Moreover, this assumption can be weakened to allow X to be mixed-integer but assuming that all linking variables are binary, i.e., all leader variables that effectively parameterize the follower's problem need to be binary. Finally, Assumption 1 also allows for bounded integer linking variables by resorting to a binary expansion. However, this typically leads to large models which are harder to solve in practice; see Owen and Mehrotra (2002) for more details.

We again consider the value-function reformulation of Problem (1):

$$\min_{x\in X, y\in Y} \quad c^\top x + d^\top y \tag{6a}$$

$$\text{s.t.} \quad Ax + By \geq a, \tag{6b}$$

$$Cx + Dy \geq b, \tag{6c}$$

$$f^\top y \leq \varphi(x). \tag{6d}$$

As before, the main difficulty in solving this reformulation is Constraint (6d) because it involves the value function φ of the follower's problem. Since the follower's problem is a mixed-integer linear optimization problem, φ is generally a nonconvex and discontinuous function; see, e.g., Hassanzadeh and Ralphs (2014). This is in contrast to the continuous case in which φ is a continuous and convex function. In the next lemma, we therefore introduce a convex function that can be used instead of φ while keeping the same value for all feasible leader's decision in Problem (1).

Before stating the lemma, however, we need some more notation. We denote by Z the set of (x, y) decisions such that y is a feasible decision of the follower's problem if the leader decision is $x \in X$, i.e., we let

$$Z := \{(x, y) \in X \times Y : Cx + Dy \geq b\}.$$

This set can also be seen as the graph of the set-valued map associated to the parameterized follower's feasible region. Its projection onto the x-space is denoted by Z_x, i.e., $Z_x := \{x \colon \exists y \text{ with } (x, y) \in Z\}$. Additionally, we let $\hat{Z} := \{(\hat{x}^k, \hat{y}^k) \colon k = 1, \dots, |\hat{Z}|\}$ be the (finite) set of extreme points of $\operatorname{conv}(Z)$, i.e., the set of points in Z that cannot be expressed as proper convex combinations of other points in Z.

We now state and prove the lemma.

Lemma 1. *Let $\check{\varphi}$ be defined as*

$$\check{\varphi}(x) := \max_{\pi\in\mathbb{R}, \lambda\in\mathbb{R}^{n_x}} \left\{ \pi + \lambda^\top x \colon \pi + \lambda^\top \hat{x}^k \leq f^\top \hat{y}^k,\ k = 1, \dots, |\hat{Z}| \right\}. \tag{7}$$

Then, for all $x \in Z_x$, it holds $\varphi(x) = \check{\varphi}(x)$.

Proof. Let $x \in Z_x$ be given. Then, it holds

$$\varphi(x) = \min_y \left\{ f^\top y \colon (x, y) \in Z \right\} = \min_y \left\{ f^\top y \colon (x, y) \in \operatorname{conv}(Z) \right\}.$$

Note that the last inequality is due to Proposition 2 in Sherali and Fraticelli (2002) and requires that x is binary. This is indeed the case since $Z_x \subseteq X = \{0, 1\}^{n_x}$; see Assumption 1. Using the Dantzig–Wolfe reformulation, it holds that any point $(x, y) \in \operatorname{conv}(Z)$ can be expressed as

$$(x, y) = \sum_{k=1}^{|\hat{Z}|} \alpha_k (\hat{x}^k, \hat{y}^k), \quad \sum_{k=1}^{|\hat{Z}|} \alpha_k = 1, \quad \alpha_k \geq 0, \quad k = 1, \dots, |\hat{Z}|.$$

Moreover, note that $\operatorname{conv}(Z)$ does not have extreme rays since it is compact by Assumption 1. Hence, we can express $\varphi(x)$ as

$$\varphi(x) = \min_\alpha \left\{ f^\top \sum_{k=1}^{|\hat{Z}|} \alpha_k \hat{y}^k \colon \sum_{k=1}^{|\hat{Z}|} \alpha_k \hat{x}^k = x,\ \sum_{k=1}^{|\hat{Z}|} \alpha_k = 1,\ \alpha \in \mathbb{R}^{|\hat{Z}|}_{\geq 0} \right\}.$$

By construction and since $x \in Z_x$, this problem is feasible and attains the same value as its dual. Its dual is exactly the optimization problem in (7) with λ and π being the dual vector associated to the first set of primal constraints and π the dual scalar value associated to the last primal constraint. □

From the definition of $\breve{\varphi}$, one can see that it is a convex function. Next, we show that it is actually the best convex function underestimating the value function φ over Z_x.

Proposition 1. *The convex envelope of φ over Z_x is $\breve{\varphi}$.*

Proof. It is well-known from standard results in convex analysis that the double convex conjugate of a given function is its convex envelope; see Corollary 12.1.1 in Rockafellar (1970). By definition of convex conjugates,

$$\varphi^{**}(x) = \max_{\lambda} \left\{ \lambda^\top x - \varphi^*(\lambda) \colon \lambda \in \mathbb{R}^{n_x} \right\}$$

holds. Again by definition, we also have $-\varphi^*(\lambda) = \min_{\hat{x} \in Z_x} \varphi(\hat{x}) - \lambda^\top \hat{x}$. With the observation that $\min_{\hat{x} \in Z_x} \varphi(\hat{x}) = \min_{(\hat{x}, y) \in Z} f^\top y$, one obtains

$$\begin{aligned} \varphi^{**}(x) &= \max_{\lambda} \ \lambda^\top x + \min_{\hat{x}, \hat{y}} \left\{ f^\top \hat{y} - \lambda^\top \hat{x} \colon (\hat{x}, \hat{y}) \in Z \right\} \\ &= \max_{\pi \in \mathbb{R}, \lambda} \left\{ \pi + \lambda^\top x \colon \pi + \lambda^\top \hat{x}^k \leq f^\top \hat{y}^k, \ k = 1, \dots, |\hat{Z}| \right\} = \breve{\varphi}(x). \end{aligned}$$ □

We now use the convex envelope $\breve{\varphi}$ of φ to derive our main theoretical result, which is the first single-level reformulation of the mixed-integer linear bilevel Problem (1), which does not make use of the value function of the follower's problem. This new formulation is a nonlinear problem with exponentially many constraints.

Theorem 2. *Let $\mathcal{S}$ denote the set of optimal points of Problem (1) and let $\hat{\mathcal{S}}$ denote the set of optimal points of the single-level nonlinear problem*

$$\min_{x \in X, y \in Y, \pi, \lambda} \quad c^\top x + d^\top y \tag{8a}$$

$$\text{s.t.} \quad Ax + By \geq a, \tag{8b}$$

$$Cx + Dy \geq b, \tag{8c}$$

$$f^\top y \leq \pi + \lambda^\top x, \tag{8d}$$

$$\pi + \lambda^\top \hat{x}^k \leq f^\top \hat{y}^k, \quad k = 1, \dots, |\hat{Z}|. \tag{8e}$$

Then, it holds $\mathcal{S} = \operatorname{proj}_{x,y}(\hat{\mathcal{S}})$.

Proof. Consider the value-function reformulation (6) of Problem (1). By Lemma 1, for any $x \in Z_x$, $\varphi(x) = \breve{\varphi}(x)$ holds. Hence, since the projection on the x-space of the feasible region of (6) is a subset of Z_x, i.e., since

$$\operatorname{proj}_x \left(\left\{ (x, y) \in X \times Y : Ax + By \geq a, \ Cx + Dy \geq b, \ f^\top y \leq \varphi(x) \right\} \right) \subseteq Z_x,$$

one can replace Constraint (6d) by "$f^\top y \leq \breve{\varphi}(x)$". The latter holds if and only if there exists $(\pi, \lambda) \in \mathbb{R}^{n_x + 1}$ satisfying (8d)–(8e); see Lemma 1. □

Theorem 2 introduces a single-level nonlinear reformulation of Problem (1) that paves the way to new solution approaches. The remainder of this paper is dedicated to such methods. First, note that, in itself, Model (8) can be tackled directly by existing state-of-the-art nonlinear optimization solvers. This could be achieved, e.g., by a priori enumerating Constraints (8e) and solving the resulting problem. While this is theoretically possible, it may not be a viable approach in practice since the number of such constraints is typically too large. Another option is to generate such constraints on-the-fly via a cutting-plane approach.

It is folklore knowledge that computational methods for tackling global optimization problems perform best if all variables involved in nonlinear constraints have tight bounds. In Problem (8), the only nonlinear constraint is (8d), which involves products between x and λ. Because x is binary, it is trivially bounded. In the next proposition, we show that λ and π can also be bounded.

Proposition 3. *Let* $(x^*, y^*) \in \mathcal{S}$. *Then, there exists* (π^*, λ^*) *such that* $(x^*, y^*, \pi^*, \lambda^*) \in \hat{\mathcal{S}}$, $\|\lambda^*\|_1 \le u - \ell$, *and* $2\ell - u \le \pi^* \le 2u - \ell$ *with*

$$\ell \le \min_{x\in X} \max_{y\in Y(x)} f^\top y \le \max_{x\in X} \min_{y \in Y(x)} f^\top y \le u,$$

where $Y(x) := \{y \in Y : Cx + Dy \ge b\}$. *Conversely, for any* $(x^*, y^*, \pi^*, \lambda^*) \in \hat{\mathcal{S}}$ *satisfying the above condition,* $(x^*, y^*) \in \mathcal{S}$ *holds.*

Proof. We first derive bounds on $\|\lambda^*\|_1$. To this end, we consider the optimization problem

$$\check{\varphi}_M(x) := \max_{\|\lambda\|_1 \le M} \lambda^\top x + \min_{(\hat{x},\hat{y})\in Z} f^\top \hat{y} - \lambda^\top \hat{x}, \tag{9}$$

where $M > 0$ is a given parameter. Note that $\check{\varphi}_M = \check{\varphi}$ holds if $M = \infty$. Our goal is to show that this equality also holds for $M = u - \ell$. By linearity of the objective function of the inner minimization problem in (9), we have

$$\check{\varphi}_M(x) = \max_{\|\lambda\|_1 \le M} \lambda^\top x + \min_{(\hat{x},\hat{y})\in \operatorname{conv}(Z)} f^\top \hat{y} - \lambda^\top \hat{x}.$$

Using the von Neumann min-max theorem, see, e.g., Section 4.1 in Bertsekas (2009), leads to

$$\begin{aligned}\check{\varphi}_M(x) &= \min_{(\hat{x},\hat{y})\in \operatorname{conv}(Z)} f^\top \hat{y} + \max_{\|\lambda\|_1 \le M} \lambda^\top (x - \hat{x}) \\ &= \min_{(\hat{x},\hat{y})\in \operatorname{conv}(Z)} f^\top \hat{y} + M \|x - \hat{x}\|_\infty .\end{aligned}$$

Note that the last equality holds by definition of dual norms. Moreover, for a fixed $x \in Z_x$, $\|x - \hat{x}\|_\infty$ is a linear function of $\hat{x}$ since it can be written as

$$\|x - \hat{x}\|_\infty = \sum_{j : x_j = 0} \hat{x}_j + \sum_{j : x_j = 1} (1 - \hat{x}_j).$$

Here, we exploit the fact that the leader variables x are binary. Hence, it holds

$$\check{\varphi}_M(x) = \min_{(\hat{x},\hat{y})\in Z} f^\top \hat{y} + M \|x - \hat{x}\|_\infty .$$

The proof is achieved by noting that $\|x - \hat{x}\|_\infty$ is greater or equal than 1 if and only if $x \ne \hat{x}$. Hence, choosing $M = u - \ell$ enforces that $x = \hat{x}$ holds at any optimal point since any other choice for x would lead to a worse objective function value. This readily shows that, for any $x \in Z_x$, $\varphi(x) = \check{\varphi}(x) = \check{\varphi}_{u-\ell}(x)$ holds.

We conclude with trivial bounds on π^*:

$$\begin{aligned}&(8\text{d}) \implies f^\top y - \lambda^\top x \le \pi \implies f^\top y - \|\lambda\|_1 \le \pi \implies \ell - (u - \ell) \le \pi, \\ &(8\text{e}) \implies \pi \le f^\top \hat{y}^k - \lambda^\top \hat{x}^k \implies \pi \le f^\top \hat{y}^k + \|\lambda\|_1 \implies \pi \le u + (u - \ell). \quad \square\end{aligned}$$

Remark 2. *In Proposition 3, we give bounds on the dual variables* λ *and* π. *Note that the resulting bounds may still be challenging to compute. However, they can be under- or over-estimated easily as follows:*

$$\min_{x\in X, y\in Y(x)} f^\top y \le \min_{x\in X} \max_{y\in Y(x)} f^\top y \quad \text{and} \quad \max_{x\in X} \min_{y\in Y(x)} f^\top y \le \max_{x\in X, y\in Y(x)} f^\top y.$$

Here, the optimization problems on the right- and on the left-hand side are, in general, at a lower level in the polynomial hierarchy compared to Problem (1). *Also note that relaxing the integrality requirements on both the leader's and the follower's variables leads to linear optimization problems, which can be solved in polynomial time in the input data of Problem* (1).

Another key aspect of Proposition 3 is to show that bounds on the dual variables can be computed by means of a problem stated in the primal space. Note that this problem always admits a solution if the standard compactness assumption holds. This is in contrast to what is typically done in the linear bilevel optimization literature, *where computing such bounds is often expressed in the dual space, which is known to be unbounded if the primal feasible set is compact; see Williams (1970) and Kleinert et al. (2020) for a detailed discussion.*

In practice, solving a mixed-integer nonlinear optimization problem is harder than solving a mixed-integer linear one. For that reason, we derive a mixed-integer linear formulation of Problem (8) by linearizing products of variables. Indeed, by Proposition 3, the only nonlinear terms in (8) involve products of binary variables x and bounded continuous variables λ. Hence, the well-known McCormick inequalities lead to the following corollary; see McCormick (1976).

Corollary 1. *Let $\tilde{\mathcal{S}}$ denote the set of optimal points of the mixed-integer linear problem*

$$\begin{aligned} \min_{x\in X, y\in Y, \pi, \lambda, w} \quad & c^\top x + d^\top y \\ \text{s.t.} \quad & Ax + By \geq a, \quad Cx + Dy \geq b, \\ & f^\top y \leq \pi + 1^\top w, \\ & -Mx \leq w \leq Mx, \quad \lambda - M(1-x) \leq w \leq \lambda + M(1-x), \\ & \pi + \lambda^\top \hat{x}^k \leq f^\top \hat{y}^k, \quad k = 1, \dots, |\hat{Z}|, \end{aligned} \tag{10}$$

with $M = u - \ell$ as defined in Proposition 3. Then, it holds $\mathcal{S} = \operatorname{proj}_{x,y}(\tilde{\mathcal{S}})$.

We conclude this section with a remark on the implications of Corollary 1.

Remark 3. *Corollary 1 states that the bilevel Problem (1) can be stated as a mixed-integer linear optimization problem with an exponential number of constraints. Note that this was to be expected since mixed-integer linear bilevel problems are known to be Σ_2^P-hard (Lodi et al. 2013). Moreover, Remark 2 shows that the M values can be computed in polynomial time and, thus, have an encoding length that is bounded from above by a polynomial of the encoding length of the input data $(c, d, A, B, a, C, D, b, f)$. Therefore, this also implies that Problem (1) is Σ_2^P-complete since checking a solution can be done by a polynomial-time routine with a single call to an oracle solving an NP-complete problem, namely, checking if all constraints of the form (8e) are satisfied.*

5. An Exact Branch-and-Cut Algorithm

To the best of our knowledge, Corollary 1 introduces the first mixed-integer linear single-level reformulation of general mixed-integer bilevel problems. As discussed in Remark 3, this formulation contains an exponential yet finite number of constraints. Unless P = NP, no polynomially-sized single-level reformulation of Problem (1) exists and Corollary 1 is the best result one can hope for in general. Nevertheless, standard approaches from mixed-integer linear optimization can be employed to tackle Problem (10). In this section, we discuss a new branch-and-cut approach to solve Problem (1).

This branch-and-cut algorithm is based on a linear and polynomially-sized relaxation of Problem (10). Let $\bar{X}$ and $\bar{Y}$ denote the continuous relaxation of X and Y, i.e., the sets obtained by dropping the integrality requirements imposed in the definition of X and Y. Then, each node j of the branch-and-cut algorithm solves the optimization problem

$$
\begin{aligned}
\min_{x\in\bar{X},y\in\bar{Y},\pi,\lambda,w} \quad & c^\top x + d^\top y && \text{(11a)}\\
\text{s.t.} \quad & Ax + By \geq a, \quad Cx + Dy \geq b, && \text{(11b)}\\
& f^\top y \leq \pi + 1^\top w, && \text{(11c)}\\
& -Mx \leq w \leq Mx, \quad \lambda - M(1-x) \leq w \leq \lambda + M(1-x), && \text{(11d)}\\
& \pi + \lambda^\top \hat{x} \leq f^\top \hat{y} \quad \text{for all } (\hat{x},\hat{y}) \in \mathcal{Z}^j, && \text{(11e)}\\
& \alpha^\top x + \beta^\top y \leq \gamma \quad \text{for all } (\alpha,\beta,\gamma) \in \mathcal{C}^j, && \text{(11f)}
\end{aligned}
$$

where $\mathcal{Z}^j \subseteq \hat{Z}$ denotes a subset of constraints of type (8e) and $\mathcal{C}^j$ denotes a set of constraints that are valid inequalities or branching decisions inherited from parent nodes. Initially, we set $\mathcal{Z}^0 = \emptyset$ and $\mathcal{C}^0 = \emptyset$. Algorithm 1 presents how node j shall be processed inside a classic branch-and-bound algorithm.

Algorithm 1 Processing of node j in the proposed branch-and-cut algorithm

1: Solve Problem (11).
2: **if** it is infeasible **then**
3: Fathom the node by infeasibility and **return**.
4: **end if**
5: Let $(x^j, y^j, \pi^j, \lambda^j, w^j)$ denote an optimal point.
6: **if** $x^j \notin X$ or $y^j \notin Y$ **then**
7: Try to enrich $\mathcal{C}^j$ with valid inequalities to cut off the fractional point.
8: **if** no valid inequality is found **then**
9: Add node j to the list of active nodes and **return**.
10: **end if**
11: **go to** Line 1.
12: **end if**
13: Solve the separation problem $(\hat{x},\hat{y}) \in \arg\min \left\{ f^\top y - (\lambda^j)^\top x \colon (x,y) \in Z \right\}$.
14: **if** $\pi^j + (\lambda^j)^\top \hat{x} > f^\top \hat{y}$ **then**
15: Set $\mathcal{Z}^j \leftarrow \mathcal{Z}^j \cup \{(\hat{x},\hat{y})\}$ and **go to** Line 1.
16: **end if**
17: Prune the node and possibly update the incumbent to $(x^j, y^j, \pi^j, \lambda^j, w^j)$.

In Line 1, the node's problem is solved. If it is infeasible, we fathom the node in Line 3. Otherwise, in Line 6, we check if the node's solution satisfies the integrality requirements, i.e., we check if $x^j \in X$ and $y^j \in Y$ holds. If some integer variables have fractional values, we use standard cut-generation routines from mixed-integer linear optimization to strengthen the relaxation and resolve the relaxation; see, e.g., Conforti et al. (2014). At this point, if no cut can be generated, we resort to branching and the node is added to the list of active nodes in Line 9. Otherwise, the node's solution is feasible w.r.t. integrality requirements. We now need to check if this point is bilevel feasible, i.e., if $f^\top y^j \leq \varphi(x^j)$ holds. To do so, we solve the extended follower's problem in Line 13. Recall that $\varphi(x^j) = \check{\varphi}(x^j)$ since $x^j \in X$. If the condition in Line 14 is not satisfied, then it shows $f^\top y^j > \varphi(x^j)$ and a new cut of type (8e) is added and the relaxation is resolved. Otherwise, the current solution is bilevel feasible and the incumbent solution is potentially updated.

It is worth highlighting that the proposed scheme is a classic branch-and-cut algorithm, which can be easily implemented in state-of-the-art MILP solvers. Moreover, it is based on the generation of cuts, which are globally valid in contrast to the state-of-the-art methods based on intersection cuts which are only locally valid; see, e.g., Fischetti et al. (2017b).

We conclude this section by showing that the obtained method is correct and that it terminates after a finite number of iterations.

Theorem 4. *If we embed Algorithm 2 into a standard branch-and-bound framework, we obtain a correct method that terminates after a finite number of iterations and that either*

(i) *correctly identifies Problem (1) as infeasible or*
(ii) *returns a point $(x^*, y^*, \lambda^*, \pi^*)$ such that (x^*, y^*) is an optimal solution to Problem (1).*

Proof. Consider an arbitrary node j in the branch-and-bound tree. We show that Algorithm 1 terminates in a finite number of iterations. First, note that Line 15 cannot be reached an infinite number of times since the solutions to the separation problem in Line 13 are (w.l.o.g.) extreme points of $\operatorname{conv}(Z)$, which are only finitely many. Also note that the separation problem is a MILP that can be solved in finite time. Moreover, any cut added in Line 15 cuts off the current point $(x^j, y^j, \pi^j, \lambda^j, w^j)$, which implies that this point cannot be visited again. Second, Line 7 can also not be reached an infinite number of times if a fixed tolerance is used for checking the violation of valid inequalities in Line 7. Indeed, by compactness of the node's feasible region, see Assumption 1, its volume is finite. Thus, any valid inequality added in Line 7 effectively reduces the volume of the node's feasible region by a constant, which is bounded away from zero. Thus, after a sufficiently large but finite number of iterations, the node's feasible region must be empty. Finally, the finite termination of the overall branch-and-bound method is due to the fact that Problem (11) has finitely many integer variables, which induces a finite number of nodes—each solved by Algorithm 1, which also finitely terminates. Correctness is due to Theorem 1. □

6. A Heuristic Based on the Penalty Alternating Direction Method

In this section, we derive a heuristic approach for mixed-integer linear bilevel problems. This heuristic is based on the single-level reformulation (8) introduced in Theorem 2 and generalizes the penalty direction method approach (PADM) presented in Section 3. Looking at Problem (8), the only complicating constraint is the strong-duality inequality (8d) since it is the only nonlinear constraint. Moreover, it is the only constraint that couples the primal variables x and y with the dual variables π and λ. Thus, we suggest to relax these constraints to obtain a penalty problem. In the next theorem, we show that for a sufficiently large but finite penalty parameter, the penalty problem and the single-level reformulation (8) are equivalent on the level of global solutions.

Theorem 5. *Let $\mathcal{S}_\rho$ denote the set of optimal points to*

$$\begin{aligned} \min_{x \in X, y \in Y, \pi, \lambda} \quad & c^\top x + d^\top y + \rho(f^\top y - \pi - \lambda^\top x) \\ \text{s.t.} \quad & Ax + By \geq a, \quad Cx + Dy \geq b, \\ & \pi + \lambda^\top \hat{x}^k \leq f^\top \hat{y}^k, \quad k = 1, \dots, |\hat{Z}|. \end{aligned}$$

Then, there exists a finite $\rho > 0$ such that $\mathcal{S} = \operatorname{proj}_{x,y}(\mathcal{S}_\rho)$ holds.

Proof. By Corollary 1, Problem (1) can be formulated as a MILP. Then, using the augmented Lagrangian duality framework from Feizollahi et al. (2016), there exists a finite $\rho > 0$ so that Problem (1) is equivalent to

$$\begin{aligned} \min_{x \in X, y \in Y, \pi, \lambda, w} \quad & c^\top x + d^\top y + \rho[f^\top y - \pi - 1^\top w]^+ \\ \text{s.t.} \quad & Ax + By \geq a, \quad Cx + Dy \geq b, \\ & -Mx \leq w \leq Mx, \quad \lambda - M(1-x) \leq w \leq \lambda + M(1-x), \\ & \pi + \lambda^\top \hat{x}^k \leq f^\top \hat{y}^k, \quad k = 1, \dots, |\hat{Z}|, \end{aligned}$$

Reverting back the McCormick inequalities, it holds that $w = \lambda^\top x$. The proof is achieved by noting that $f^\top y \geq \pi + \lambda^\top x$ always holds by weak duality. □

Following the discussion in Section 3, given a penalty parameter $\rho > 0$, the PADM first solves the sub-problem

$$\begin{aligned} \min_{x \in X, y \in Y} \quad & c^\top x + d^\top y + \rho(f^\top y - \bar{\pi} - x^\top \bar{\lambda}) \\ \text{s.t.} \quad & Ax + By \geq a, \quad Cx + Dy \geq b, \end{aligned}$$

where $\bar{\pi}$ and $\bar{\lambda}$ are given. Note that this problem has the same feasible region as the classic high-point relaxation of the original bilevel Problem (1). However, the objective function is modified depending on the penalty parameter value, the current dual estimate $\bar{\lambda}$, and the follower's objective function. Also note that the term $\rho\bar{\pi}$ is constant and can be omitted without changing the solutions to this sub-problem.

The second sub-problem is given by

$$\max_{\lambda, \pi} \left\{ \pi + \lambda^\top \bar{x} \colon \pi + \lambda^\top \hat{x}^j \leq f^\top \hat{y}^j, \, j = 1, \dots, |\hat{Z}| \right\}. \tag{13}$$

Here, we have directly omitted the constant term $c^\top \bar{x} + f^\top \bar{y} + \rho f^\top \bar{y}$ for better readability. Also note that this problem does not depend on the penalty parameter ρ because it only scales the objective function value without altering the set of solutions. Hence, Problem (13) simply corresponds to computing $\check{\varphi}(\bar{x})$. In the next proposition, we give another interpretation in terms of sub-differentials of $\check{\varphi}$.

Proposition 6. *Let $\bar{x} \in Z_x$ be given and let $\bar{\mathcal{S}}(\bar{x})$ denote the set of global solutions of (13). Then, it holds $\partial\check{\varphi}(\bar{x}) = \operatorname{proj}_\lambda(\bar{\mathcal{S}}(\bar{x}))$.*

Proof. Recall the proof of Proposition 1 and observe that the optimal function value of Problem (13) is $\check{\varphi}(\bar{x}) = \varphi^{**}(\hat{x}) = \max_\lambda\{\lambda^\top \bar{x} - \varphi^*(\lambda)\}$. Hence, for any $(\lambda, \pi) \in \bar{\mathcal{S}}(\bar{x})$, it holds $\varphi^{**}(x) = \lambda^\top \bar{x} - \varphi^*(\lambda)$. It then follows from Theorem 23.5 in Rockafellar (1970) that

$$\bar{x}^\top \lambda = \varphi^{**}(x) + \varphi^*(\lambda) \iff \lambda \in \partial\varphi^{**}(x).$$

With $\varphi^{**} = \check{\varphi}$, we conclude the proof. □

Proposition 6 leads to a new intuition for the PADM applied to Problem (8). To present it, let $\bar{x}$ be a feasible leader decision, i.e., let $\bar{x}$ be such that there exists y satisfying $A\bar{x} + By \geq a$, $C\bar{x} + Dy \geq b$, and $f^\top y \leq \varphi(\bar{x})$. We first consider the optimization problem

$$\begin{aligned} \min_{x \in X, y \in Y} \quad & c^\top x + d^\top y \\ \text{s.t.} \quad & Ax + By \geq a, \quad Cx + Dy \geq b, \\ & f^\top y \leq \varphi(\bar{x}) + \lambda^\top (x - \bar{x}) \end{aligned} \tag{$*$}$$

with $\lambda \in \partial\check{\varphi}(\bar{x})$. Clearly, $\bar{x}$ is feasible for this problem. Recall that $\check{\varphi}(\bar{x}) = \varphi(\bar{x})$ holds according to Lemma 1 because $\bar{x} \in Z_x$. Hence, this problem looks for a better decision for the leader given a local estimate on how the follower's optimal objective function value φ changes. Penalizing Constraint ($*$) yields the first sub-problem in the PADM (up to some constant terms). The second sub-problem corresponds to the computation of a subgradient of $\check{\varphi}$ at a current point $\bar{x}$.

7. Exploiting the Structure of Interdiction Problems

A prominent special case of mixed-integer linear bilevel problems are interdiction games; see, e.g., Israeli and Wood (2002), Fischetti et al. (2019), Schmidt and Thürauf (2024), or the survey Kleinert et al. (2021) and the references therein. In this setting, the leader and the follower typically share a common set of resources, which the leader can interdict so that the follower cannot use them anymore. For such problems, another single-level reformulation akin to the one of Theorem 2 can be derived. Before we do so, we make the following assumption.

Assumption 2 (Interdiction linking constraints)**.** *It holds $C = D = -I_{n_x}$ and $b = -1$, i.e., the follower's problem is $\min_y\{f^\top y \colon y \in Y,\ x + y \le 1\}$.*

Note that we neither require that the bilevel problem is a min-max problem nor that any structural assumption on Y holds. One specificity of interdiction games lies in that the linking constraints $x + y \le 1$ are always faces of the convex hull of the follower's feasible region, i.e., it holds

$$\operatorname{conv}(\{y \in Y \colon x + y \le 1\}) = \operatorname{conv}(Y) \cap \{y \in \mathbb{R}^{n_y} \colon x + y \le 1\}. \tag{14}$$

Note that this requires x to be binary and $y \in [0,1]^{n_y}$. This property has been recently exploited in the context of two-stage robust optimization; see, e.g., Arslan and Detienne (2022), Detienne et al. (2024), and Lefebvre et al. (2023). Hence, a single-level reformulation can be derived without the need of lifting the follower's problem as it is done in Theorem 2. This is what we do in the next theorem.

Theorem 7. *Let Assumption 2 hold and let $\mathcal{S}$ denote the set of optimal points of Problem (1). Let $\hat{\mathcal{S}}$ denote the set of optimal points of the single-level nonlinear problem*

$$\begin{aligned}
\min_{x\in X, y\in Y, \pi, \lambda} \quad & c^\top x + d^\top y && \text{(15a)}\\
\text{s.t.} \quad & Ax + By \ge a, \quad x + y \le 1, && \text{(15b)}\\
& f^\top y \le \pi + \lambda^\top (1 - x), && \text{(15c)}\\
& \pi - \lambda^\top \hat{y}^j \le f^\top \hat{y}^j, \quad j = 1, \dots, |\hat{Y}|, && \text{(15d)}\\
& \lambda \ge 0, && \text{(15e)}
\end{aligned}$$

where $\hat{Y}$ denotes the set of extreme points of $\operatorname{conv}(Y)$*. Then, it holds $\mathcal{S} = \operatorname{proj}_{x,y}(\hat{\mathcal{S}})$.*

Proof. Following the proof of Theorem 2, we first write $\varphi(x)$ as

$$\begin{aligned}
\varphi(x) &= \min_y \left\{ f^\top y \colon y \in \operatorname{conv}(\{y \in Y \colon x + y \le 1\}) \right\}\\
&= \min_y \left\{ f^\top y \colon y \in \operatorname{conv}(Y),\ x + y \le 1 \right\}.
\end{aligned}$$

The first equality is due to the linearity of the objective function while the second is due to (14). Using the Dantzig–Wolfe reformulation, any point $y \in \operatorname{conv}(Y)$ can be expressed as

$$y = \sum_{k=1}^{|\hat{Y}|} \alpha_k \hat{y}^k, \quad \sum_{k=1}^{|\hat{Y}|} \alpha_k = 1, \quad \alpha_k \ge 0, \quad k = 1, \dots, |\hat{Y}|.$$

Hence, one can readily express $\varphi(x)$ as

$$\varphi(x) = \min_\alpha \left\{ f^\top \sum_{k=1}^{|\hat{Y}|} \alpha_k \hat{y}^k \colon \sum_{k=1}^{|\hat{Y}|} \alpha_k \hat{y}^k + x \le 1,\ \sum_{k=1}^{|\hat{Y}|} \alpha_k = 1,\ \alpha \in \mathbb{R}^{|\hat{Y}|}_{\ge 0} \right\}.$$

TABLE 1. Summary of our test set

BOBILib class	BOBILib type	# Instances
general-bilevel	mixed-integer	166
	pure-integer	92
interdiction	assignment	24
	clique	219
	generalized	90
	knapsack	599
	multidimensional-knapsack	954
	network	72

By construction and since $x \in Z_x$, this problem is feasible and attains the same value as its dual. Its dual is exactly the optimization problem in (15) with λ and π being the dual vector associated to the linking constraints and the scalar dual value associated to the last primal constraint. □

Similar to the mixed-integer linear reformulation presented in Corollary 1, a mixed-integer linear reformulation of Problem (1) under Assumption 2 can be derived by linearizing the products between λ and x. One important class of interdiction problems are so-called downward monotone interdiction problems; see Fischetti et al. (2019). We recall that Y is said to be downward monotone if for any $y \in Y$, any y' with $0 \leq y' \leq y$ satisfies $y' \in Y$. For this class of problems, Fischetti et al. (2019) shows that the dual variables λ are bounded by f.

8. Numerical Results

In this section, we present numerical results to computationally evaluate the performance of the proposed methods.

8.1. Instances and Experimental Setup.

We consider instances from the BOBILib (Thürauf et al. 2024); see https://bobilib.org as well. More precisely, we include all the instances from the "Collection" that fulfill our assumptions, namely, all problems in which the linking variables are binary and the shared constraint set is bounded. Overall, our test set consists of 2216 instances. A summary of the characteristics of the instances can be found in Table 1. We can see that most of our test set is composed by interdiction problems. However, this is in line with the overall collection of instances from BOBILib.

All models presented in this paper are implemented in C++20 using the idol library (Lefebvre 2025) and Gurobi 11.0 as the underlying solver. All parameters are left at their default settings apart from the LazyConstraints parameter for those problems for which lazy constraints need to be generated. Indeed, cuts are separated inside the Gurobi callback triggered by the MIPSOL event, i.e., when an integer point is found. Thus, turning the LazyConstraints parameter on is necessary to deactivate certain reductions and transformations that are incompatible with lazy constraints.[1] We compare the PADM and the MILP formulation w.r.t. computation time to the bilevel solver MibS 1.2.1 (Tahernejad et al. 2020) equipped with CPLEX 22.1 as the underlying solver for solving the follower's problem and for checking bilevel feasibility. Moreover, we also consider solving the MILP formulation introduced in

[1] See the online documentation at https://docs.gurobi.com/projects/optimizer/en/current/reference/parameters.html#lazyconstraints.

this paper with Gurobi and stopping the solver after the first feasible point is found. Overall, we therefore consider the following approaches for the entire test set.

- MILP solves the MILP formulation from Section 4 using Gurobi.
- MILP-First stops MILP after the first feasible point is found.
- PADM runs the PADM described as in Section 6.
- MibS solves the bilevel problem with MibS (Tahernejad et al. 2020).

For those approaches based on the new single-level reformulation (MILP, MILP-First, and PADM), we compute bounds on the dual variables according to Remark 2, i.e., we use Gurobi with a time limit of 5 minutes to compute bounds on the optimal objective function value of

$$\min_{x \in X, y \in Y(x)} f^\top y \quad \text{and} \quad \max_{x \in X, y \in Y(x)} f^\top y.$$

All computations were carried out on a single node of a compute server with Intel XEON SP 6126 CPUs.[2] We use a time limit of 1 h, a memory limit of 32 GB, and limit the number of threads to 4 for all approaches.

8.2. **Results for the Entire Test Set.** Figure 1 depicts the empirical cumulative distribution function (ECDF) of computation time over our test set, for all four methods. First, it can be seen that the bilevel solver MibS outperforms MILP. Indeed, MibS solves approximately twice as many instances to global optimality as the MILP formulation. Nevertheless, MILP still is able to find a global solution to 12 % of the test set, mainly in the first 15 minutes of computation. This result is encouraging since this approach solely relies on an existing MILP solver used as a black-box and does not require any extra implementation effort. Moreover, we believe that this new formulation can be useful from a computational viewpoint in at least two directions. First, it can be combined with existing cutting-plane approaches such as intersection cuts; see Fischetti et al. (2017b). Second, the performance could, in principle, be further improved by computing tighter bounds on the dual variables λ and π. Indeed, these bounds are used to linearize products between primal and dual variables and it is well-known that tighter bounds lead to faster solution times. In our experiments, 15 % of the computed bounds are less than $3 \cdot 10^4$ while 75 % are less than $2 \cdot 10^5$.

Regarding heuristic approaches, Figure 1 also shows that MILP-First and PADM are able to produce feasible points to 48 % and 70 % of the instances, respectively. In most of the cases, this occurs within 15 minutes of computation. To assess the quality of the computed feasible points, we compare the obtained objective function value to the best known objective function value reported in the BOBILib. To this end, Figure 2 depicts the ECDF of optimality gap[3] for those instances that have the status "optimal" in the BOBILib and for which a feasible point can be found by both MILP-First and PADM. It can be seen that both methods tend to produce high-quality solutions. For instance, 85 % of the solutions returned by MILP-First are below 1 % of optimality gap. Similarly, 80 % of those points found by the PADM are below 4 % of optimality gap. While it is tempting to argue that MILP-First performs better than PADM, it should be kept in mind that PADM finds substantially more feasible points than MILP-First.

The ECDF of MILP-First presented in Figure 1 is rather untypical with a "jump" around 1 % of optimality gap. It seems like Gurobi is able to compute very good heuristic solutions early in the search process. This can be explained by the fact that

[2] See https://hpc.rz.rptu.de/elwetritsch/hardware.shtml

[3] Optimality gaps are computed according to https://www.ibm.com/docs/en/icos/22.1.1?topic=parameters-relative-mip-gap-tolerance

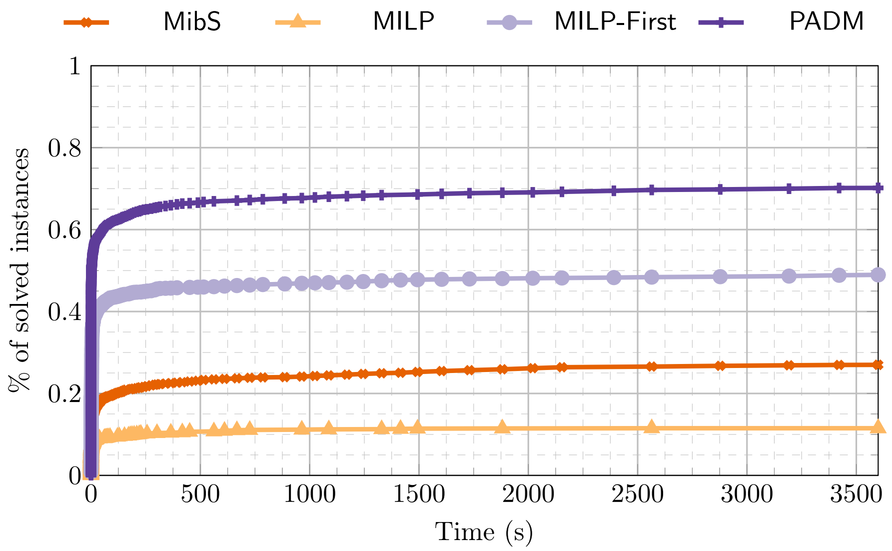

FIGURE 1. ECDF of computation time over all instances and for all methods.

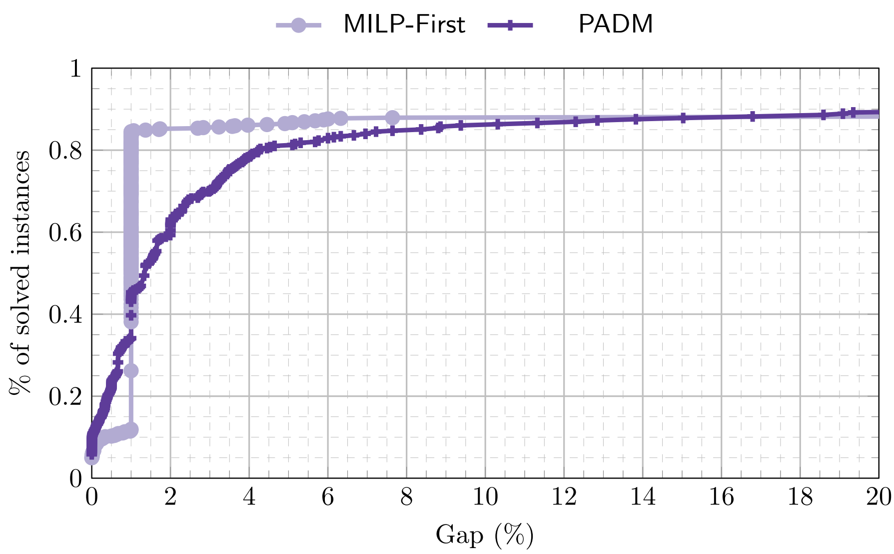

FIGURE 2. ECDF of optimality gap over those instances for which both heuristic approaches MILP-First and PADM found a feasible point.

the search for feasible points is driven by an approximation of the value function φ, which is tight if x has a binary value; see Lemma 1.

Table 2 and Table 3 give a more detailed description of those instances for which MILP-First and PADM are able to find feasible points, respectively. It can be seen that PADM computes a feasible point for 48 instances, which were open in BOBILib (out of 285) and improves the best known solution of 12 instances (out of 1132). Moreover, while MILP-First "only" finds a feasible point to 43 open instances, it is able to improve the best known solution for 355 instances.

Table 2. Outcome of PADM

BOBILib status	# Instances	PADM outcome	# Instances
open	285	found feasible	48
open with feasible point	1132	found improving feasible	12
		found non-improving feasible	747
optimal	770	found optimal (without proof)	40
		found feasible	707
infeasible	29	—	—
Total	2216		1554 (70.1 %)

Table 3. Outcome of MILP-First

BOBILib status	# Instances	MILP-First outcome	# Instances
open	285	found feasible	43
open with feasible point	1132	found improving feasible	355
		found non-improving feasible	5
optimal	770	found optimal (without proof)	34
		found feasible	645
infeasible	29	—	—
Total	2216		1082 (48.8 %)

Finally, we take interest in the size of the instances that can be solved by each method. Here, an instance is considered as "solved" if

- for MibS and MILP, an optimal point can be provably reported;
- for MILP-First and PADM, a feasible point can be reported.

We define the size of an instance as the sum of its number of variables and its number of constraints (both in the upper and lower level). Figure 3 depicts the number of instances that can be solved by each approach depending on its size gathered into 5 groups of approximately equal size. Understandably, the easiest instances for all four methods are the smallest ones. Moreover, we see that PADM is less sensitive to the size than the three other approaches.

8.3. **Results for Interdiction Instances.** In this section, we consider the PADM from Section 7, which is tailored to interdiction problems. This approach is referred to as PADM-Interdiction. Here, we again take interest in computation time and in the quality of the returned feasible point. Figure 4 depicts the ECDF of computation time for a time limit of 1 h over those instances fulfilling Assumption 2. It can be seen that PADM-Interdiction approach outperforms PADM since it is able to compute feasible points to 80 % of the instances in contrast to 70 % for PADM.

Figure 5 shows the ECDF of optimality gap over those instances with a status "optimal" in the BOBILib and for which a feasible point could be computed by both PADM and PADM-Interdiction approach. It can be seen that the quality of the returned solution by both approaches are similar. Hence, this indicates that exploiting the structure of interdiction problems is beneficial in terms of computation time without altering the quality of the returned feasible points.

9. Conclusion

A common strategy to tackle bilevel optimization problems in which the follower's problem is convex is to derive a single-level reformulation based on optimality conditions. However, if the follower's problem includes integer variables, such techniques are no longer applicable and enumerative methods based on cutting

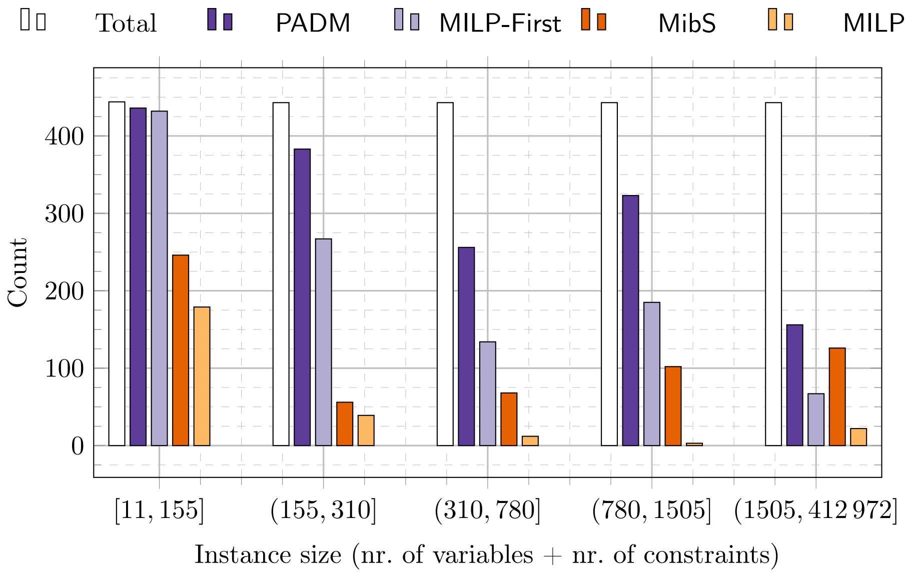

FIGURE 3. Number of solved instances by each method depending on the instance size. (Note that PADM and MILP-first are heuristic approaches.)

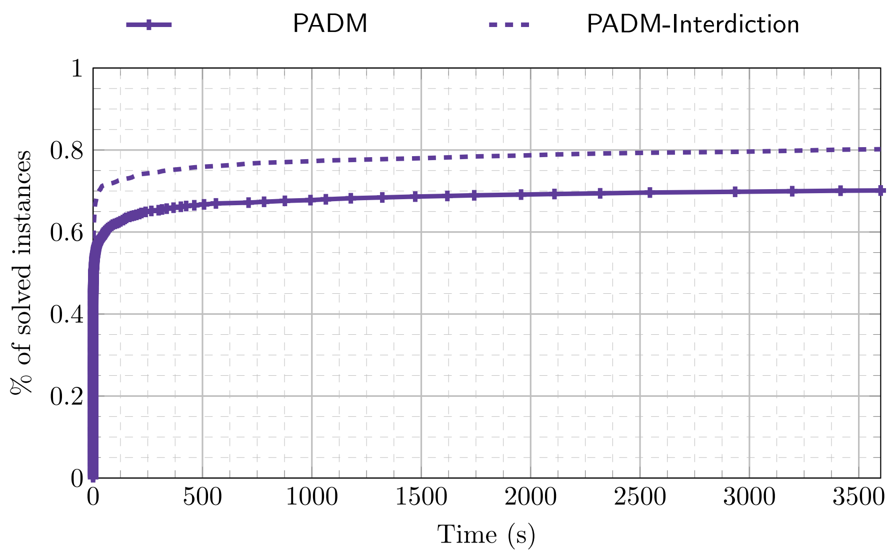

FIGURE 4. ECDF of computational time over the interdiction-type instances for two variants of the PADM.

planes have been the predominant approach. In this paper, we introduced the first single-level reformulation of general mixed-integer linear bilevel optimization problems, which does not explicitly rely on the value function of the follower's problem. Moreover, we discussed both exact and heuristic approaches.

The reformulation is based on a Dantzig–Wolfe reformulation to convexify the follower's problem, enabling the application of strong duality and the derivation of an equivalent nonlinear reformulation. Furthermore, we show that valid bounds on the dual variables that are required to linearize the resulting nonlinearities can be

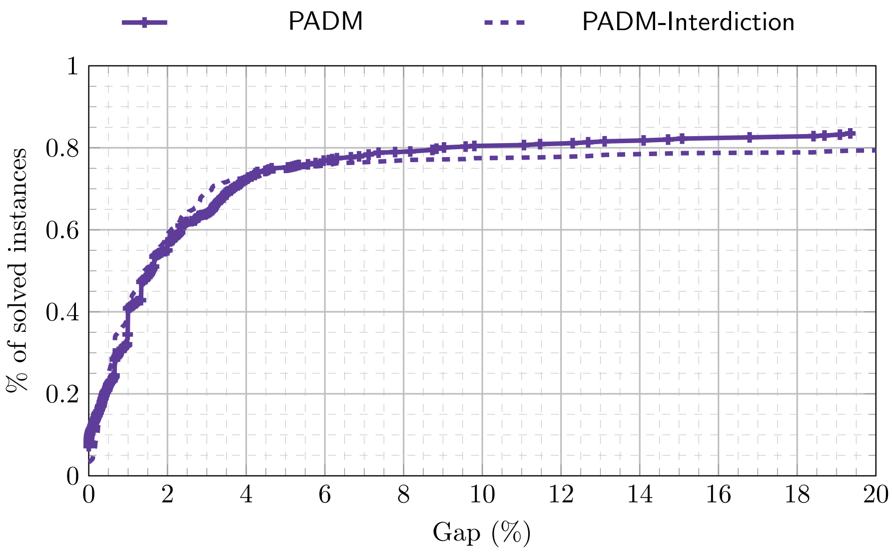

FIGURE 5. ECDF of optimality gap over our those instances for which both the PADM and the PADM-Interdiction approach found a feasible point.

computed by solving a polynomial-time-solvable optimization problem. Importantly, this problem is solely based on the follower's primal problem which, under standard assumptions, always admits a solution. As a result, we obtain an exact MILP reformulation of the original bilevel model, which contains an exponential number of constraints. From this model, we derived an exact approach as well as a heuristic based on the penalty alternating direction method.

Our numerical experiments show encouraging results if the single-level reformulation is used as-is and solved by a standard state-of-the-art mixed-integer linear solver. Additionally, we show that the proposed heuristics are able to produce high-quality feasible points. Future work includes the refinement of the bounds of the dual variables and the integration of existing locally valid cuts to improve the convergence of the solution process.

ACKNOWLEDGEMENTS

We acknowledge the support by the German Bundesministerium für Bildung und Forschung within the project "RODES" (Förderkennzeichen 05M22UTB). The computations were executed on the high performance cluster "Elwetritsch" at the TU Kaiserslautern, which is part of the "Alliance of High Performance Computing Rheinland-Pfalz" (AHRP). We kindly acknowledge the support of RHRK. Last but not least, we are grateful to Johannes Thürauf for communicating more detailed results regarding the BOBILib instances.

REFERENCES

Arslan, A. N. and B. Detienne (2022). "Decomposition-Based Approaches for a Class of Two-Stage Robust Binary Optimization Problems." In: *INFORMS Journal on Computing* 34.2, pp. 857–871. DOI: 10.1287/ijoc.2021.1061.

Aussel, D., C. Egea, and M. Schmidt (2024). "A tutorial on solving single-leader-multi-follower problems using SOS1 reformulations." In: *International Transactions in Operational Research* 32.3, pp. 1227–1250. DOI: 10.1111/itor.13466.

Bertsekas, D. (2009). *Convex Optimization Theory*. Athena Scientific.

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

Conforti, M., G. Cornuéjols, and G. Zambelli (2014). *Integer Programming*. Springer International Publishing. DOI: 10.1007/978-3-319-11008-0.

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

Detienne, B., H. Lefebvre, E. Malaguti, and M. Monaci (2024). "Adjustable robust optimization with objective uncertainty." In: *European Journal of Operational Research* 312.1, pp. 373–384. DOI: 10.1016/j.ejor.2023.06.042.

Feizollahi, M. J., S. Ahmed, and A. Sun (2016). "Exact augmented Lagrangian duality for mixed integer linear programming." In: *Mathematical Programming* 161.1–2, pp. 365–387. DOI: 10.1007/s10107-016-1012-8.

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

– (2017b). "On the use of intersection cuts for bilevel optimization." In: *Mathematical Programming* 172.1–2, pp. 77–103. DOI: 10.1007/s10107-017-1189-5.

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

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

Geißler, B., A. Morsi, L. Schewe, and M. Schmidt (2017). "Penalty Alternating Direction Methods for Mixed-Integer Optimization: A New View on Feasibility Pumps." In: *SIAM Journal on Optimization* 27.3, pp. 1611–1636. DOI: 10.1137/16M1069687.

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

Hassanzadeh, A. and T. Ralphs (2014). *On the Value Function of a Mixed Integer Linear Optimization Problem and an Algorithm for Its Construction*. Technical Report 14T-004. COR@L Laboratory, Lehigh University. URL: http://coral.ie.lehigh.edu/~ted/files/papers/MILPValueFunction14.pdf.

Henke, D., H. Lefebvre, M. Schmidt, and J. Thürauf (2024). "On coupling constraints in linear bilevel optimization." In: *Optimization Letters* 19.3, pp. 689–697. DOI: 10.1007/s11590-024-02156-3.

– (2025). *On Coupling Constraints in Pessimistic Linear Bilevel Optimization*. Tech. rep. URL: https://optimization-online.org/?p=29476.

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

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

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

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

Kleinert, T. and M. Schmidt (2021). "Computing Feasible Points of Bilevel Problems with a Penalty Alternating Direction Method." In: *INFORMS Journal on Computing* 33.1, pp. 198–215. DOI: 10.1287/ijoc.2019.0945.

– (2023). "Why there is no need to use a big-M in linear bilevel optimization: a computational study of two ready-to-use approaches." In: *Computational Management Science* 20.1. DOI: 10.1007/s10287-023-00435-5.

Lefebvre, H. (2025). *idol, A C++ Framework for Optimization.* publicly available online. URL: https://hlefebvr.github.io/idol/ (visited on 03/25/2025).

Lefebvre, H. and M. Schmidt (2024). *Computing Counterfactual Explanations for Linear Optimization: A New Class of Bilevel Models and a Tailored Penalty Alternating Direction Method.* URL: https://optimization-online.org/?p=28803.

Lefebvre, H., M. Schmidt, and J. Thürauf (2023). *Column Generation in Column-and-Constraint Generation for Adjustable Robust Optimization with Interdiction-Type Linking Constraints.* URL: https://optimization-online.org/?p=24462.

Lodi, A., T. K. Ralphs, and G. J. Woeginger (2013). "Bilevel programming and the separation problem." In: *Mathematical Programming* 146.1–2, pp. 437–458. DOI: 10.1007/s10107-013-0700-x.

McCormick, G. P. (1976). "Computability of global solutions to factorable nonconvex programs: Part I — Convex underestimating problems." In: *Mathematical Programming* 10.1, pp. 147–175. DOI: 10.1007/bf01580665.

Owen, J. H. and S. Mehrotra (2002). "On the Value of Binary Expansions for General Mixed-Integer Linear Programs." In: *Operations Research* 50.5, pp. 810–819. DOI: 10.1287/opre.50.5.810.370.

Rockafellar, R. T. (1970). *Convex Analysis.* Princeton University Press. DOI: doi:10.1515/9781400873173.

Schmidt, M. and J. Thürauf (2024). "An Exact Method for Nonlinear Network Flow Interdiction Problems." In: *SIAM Journal on Optimization* 34.4, pp. 3623–3652. DOI: 10.1137/22m152983x.

Sherali, H. D. and B. M. Fraticelli (2002). "A modification of Benders' decomposition algorithm for discrete subproblems: An approach for stochastic programs with integer recourse." In: *Journal of Global Optimization* 22.1/4, pp. 319–342. DOI: 10.1023/a:1013827731218.

Tahernejad, S., T. K. Ralphs, and S. T. DeNegre (2020). "A branch-and-cut algorithm for mixed integer bilevel linear optimization problems and its implementation." In: *Mathematical Programming Computation* 12.4, pp. 529–568. DOI: 10.1007/s12532-020-00183-6.

Thürauf, J., T. Kleinert, I. Ljubić, T. Ralphs, and M. Schmidt (2024). *BOBILib: Bilevel Optimization (Benchmark) Instance Library.* URL: https://optimization-online.org/?p=27063.

Vicente, L., G. Savard, and J. Júdice (1994). "Descent approaches for quadratic bilevel programming." In: *Journal of Optimization Theory and Applications* 81.2, pp. 379–399. DOI: 10.1007/BF02191670.

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

– (1952). *Theory of the market economy.* Oxford University Press.

Williams, A. (1970). "Boundedness relations for linear constraint sets." In: *Linear Algebra and its Applications* 3.2, pp. 129–141. DOI: 10.1016/0024-3795(70)90009-1.

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

Zare, M. H., J. S. Borrero, B. Zeng, and O. A. Prokopyev (2017). "A note on linearized reformulations for a class of bilevel linear integer problems." In: *Annals of Operations Research* 272.1–2, pp. 99–117. DOI: 10.1007/s10479-017-2694-x.

(Henri Lefebvre, Martin Schmidt) Trier University, Department of Mathematics, Universitätsring 15, 54296 Trier, Germany

Email address: henri.lefebvre@uni-trier.de

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