A Penalty Branch-and-Bound Method for Mixed-Integer Quadratic Bilevel Problems

Andreas Horländer, Martin Schmidt

Abstract. We propose an algorithm for solving bilevel problems with mixed-integer convex-quadratic upper level as well as convex-quadratic and continuous lower level. The method is based on a classic branch-and-bound procedure, where branching is performed on the integer constraints and on the complementarity constraints resulting from the KKT reformulation of the lower-level problem. However, instead of branching on constraints as usual, suitably chosen penalty terms are added to the objective function in order to create new subproblems in the tree. We prove the correctness of the method and present its applicability by some first numerical results.

1. Introduction

Bilevel optimization gained increasing attention over the last years and decades mainly because bilevel models form a powerful tool for hierarchical decision making as it appears in the energy sector [9] or in security applications [8]. At the same time, bilevel problems are very hard to solve both in theory and practice [3]. Most of the solution techniques rely on branch-and-bound (B&B) or branch-and-cut techniques; see [2] for a state-of-the-art method and [4] for a recent survey.

In this paper, we follow the idea of [7] for solving mixed-integer linear complementarity problems and present a B&B method that uses penalizations of violated constraints instead of constraint branching. To be more specific, we discuss bilevel problems with convex mixed-integer quadratic problems in the upper and convex-quadratic problems in the lower level, i.e., we consider mixed-integer bilevel problems of the form

$$\min_{x,y} \quad Q(x,y) := \frac{1}{2}x^\top H_x x + \frac{1}{2}y^\top H_y y + c_x^\top x + c_y^\top y \tag{1a}$$

$$\text{s.t.} \quad Ax + By \geq a,\ x_i \in \{0,1\},\ i \in I \subseteq [n_x] := \{1,\dots,n_x\},\ y \in S(x), \tag{1b}$$

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

$$\min_y \quad \frac{1}{2}y^\top G_y y + x^\top G_{xy} y + d_y^\top y \quad \text{s.t.} \quad Cx + Dy \geq b \tag{2}$$

with $A \in \mathbb{R}^{m \times n_x}$, $B \in \mathbb{R}^{m \times n_y}$, $C \in \mathbb{R}^{\ell \times n_x}$, $D \in \mathbb{R}^{\ell \times n_y}$, $G_{xy} \in \mathbb{R}^{n_x \times n_y}$, $c_x \in \mathbb{R}^{n_x}$, $c_y, d_y \in \mathbb{R}^{n_y}$, $a \in \mathbb{R}^m$, and $b \in \mathbb{R}^\ell$. The matrices $H_x \in \mathbb{R}^{n_x \times n_x}$ and $H_y, G_y \in \mathbb{R}^{n_y \times n_y}$ are symmetric and positive semidefinite. The upper-level problem (1) thus has a convex-quadratic objective function, linear constraints, and mixed-integer variables. The objective function of the lower level is also convex and quadratic because the term $x^\top G_{xy} y$ is linear in y. The constraints of the lower level are linear and the variables are continuous. Since the upper level is a mixed-integer quadratic problem and the lower level is a quadratic problem, we call Problem (1) an MIQP-QP bilevel problem.

2020 *Mathematics Subject Classification.*

Key words and phrases. Bilevel optimization, Branch-and-bound, Penalty methods, Mixed-integer optimization.

We reformulate these problems using the classic KKT reformulation and tackle the integrality constraints as well as the KKT complementarity conditions using a novel penalty B&B method. By doing so, we introduce a new class of algorithms for solving bilevel optimization problems. We prove the correctness of our method (Sect. 2) and discuss some preliminary numerical results (Sect. 3). The latter show that, although the method is not yet outperforming state-of-the-art MILP solvers, it has the potential to become competitive if further algorithmic techniques such as cutting planes or tailored heuristics can be developed in the future.

2. A Penalty Branch-and-Bound Method

2.1. **Main Ideas and Derivation of the Algorithm.** We first apply the KKT reformulation to the bilevel problem (1), i.e., we replace the lower level with its KKT conditions, which are necessary and sufficient for global optimality in our setup. They are given by

$$G_y y + G_{xy}^\top x + d_y - D^\top \lambda = 0, \quad 0 \leq \lambda \perp Cx + Dy - b \geq 0 \tag{3}$$

and the KKT reformulation of the bilevel problem (1) thus reads

$$\begin{aligned}
\min_{x,y,\lambda} \quad & Q(x,y) && \text{(4a)}\\
\text{s.t.} \quad & Ax + By \geq a, && \text{(4b)}\\
& G_y y + G_{xy}^\top x + d_y - D^\top \lambda = 0, && \text{(4c)}\\
& 0 \leq \lambda \perp Cx + Dy - b \geq 0, && \text{(4d)}\\
& x_i \in \{0,1\}, \quad i \in I. && \text{(4e)}
\end{aligned}$$

From Thm. 2.3 in [1], it follows that (x^*, y^*) is a global optimal solution of the problem (1) if (x^*, y^*, λ^*) is a global optimal solution of (4). Hence, we will work with the latter problem in what follows. Note that Problem (4) contains two complicating aspects: the integrality and the KKT complementarity constraints. To address these two aspects, we again reformulate the problem. We remove the problematic constraints and penalize their violation by extending the objective function with additional piecewise-linear penalty terms and obtain the continuous problem

$$\begin{aligned}
\min_{x,y,\lambda} \quad & Q(x,y) + \alpha \sum_{i\in I} \min\{x_i, 1-x_i\} + \beta \sum_{j=1}^{\ell} \min\{\lambda_j, (Cx+Dy-b)_j\} && \text{(5a)}\\
\text{s.t.} \quad & Ax + By \geq a, \quad Cx + Dy \geq b, && \text{(5b)}\\
& G_y y + G_{xy}^\top x + d_y - D^\top \lambda = 0, && \text{(5c)}\\
& \lambda \geq 0, \quad 0 \leq x_i \leq 1, \quad i \in I, && \text{(5d)}
\end{aligned}$$

where $\alpha, \beta > 0$ are suitably chosen penalty parameters. For what follows, let $\pi(x,y,\lambda)$ denote the objective function of the penalty reformulation (5) and let $\mathcal{P}$ denote its feasible set. Problem (5) can thus be written as $\min\{\pi(x,y,\lambda) : (x,y,\lambda) \in \mathcal{P}\}$. In order to measure the violation of the integrality and complementarity constraints, we make use of the following definition.

Definition 1. *For a given point $(x,y,\lambda) \in \mathcal{P}$, we call*

$$F_w(x,y,\lambda;\alpha,\beta) := \alpha \sum_{i\in I} \min\{x_i, 1-x_i\} + \beta \sum_{j=1}^{\ell} \min\{\lambda_j, (Cx+Dy-b)_j\}$$

the weighted infeasibility measure*. Moreover, $F_u(x,y,\lambda) := F_w(x,y,\lambda;1,1)$ is called the* unweighted infeasibility measure*.*

The following theorem establishes the connection between the penalty reformulation (5) and the bilevel problem (1).

Theorem 2. *Let (x^*, y^*, λ^*) be a global optimal solution of the penalty reformulation (5) for arbitrarily chosen $\alpha, \beta > 0$. If the solution is feasible for the KKT reformulation (4), i.e., $F_w(x^*, y^*, \lambda^*; \alpha, \beta) = 0$, then it is also globally optimal for Problem (4). Furthermore, (x^*, y^*) is globally optimal for the bilevel problem (1).*

Proof. Suppose that (x^*, y^*, λ^*) is not a global optimum of the KKT reformulation (4), i.e., there exists a point $(\bar{x}, \bar{y}, \bar{\lambda})$ with $Q(\bar{x}, \bar{y}) < Q(x^*, y^*)$. Then,

$$\pi(\bar{x}, \bar{y}, \bar{\lambda}) = Q(\bar{x}, \bar{y}) < Q(x^*, y^*) = \pi(x^*, y^*, \lambda^*)$$

holds, which is in contrast to the global optimality of (x^*, y^*, λ^*) w.r.t. the penalty reformulation (5). According to Thm. 2.3 in [1], (x^*, y^*) is also global optimal for Problem (1). □

According to Thm. 2, we can obtain globally optimal solutions of Problem (1) by solving the penalty reformulation (5). However, this problem is still hard to solve. Compared to the KKT reformulation (4), Problem (5) has only linear constraints but has a nonsmooth and nonconvex objective function due to the penalty terms. This seems to be obstructive w.r.t. Thm. 2, which requires a global solution of Problem (5). Fortunately, we will show that the method described in the following can tackle these aspects.

2.2. **A Multi-Tree Penalty Branch-and-Bound Method.** We now derive a B&B method for solving (5) that branches by adding suitably chosen penalty terms instead of adding constraints or refining variable bounds. As in classic B&B, we first solve the root node relaxation of Problem (5), i.e., $\min\{Q(x, y) \colon (x, y, \lambda) \in \mathcal{P}\}$. If integrality or complementarity constraints are violated in the obtained solution, we choose one among them for branching and construct the subproblems

$$S_1(x, y, \lambda) := \min_{(x,y,\lambda)\in\mathcal{P}} Q(x, y) + \alpha x_i,$$
$$S_2(x, y, \lambda) := \min_{(x,y,\lambda)\in\mathcal{P}} Q(x, y) + \alpha(1 - x_i),$$

if we choose an integrality constraint or

$$\tilde{S}_1(x, y, \lambda) := \min_{(x,y,\lambda)\in\mathcal{P}} Q(x, y) + \beta\lambda_j,$$
$$\tilde{S}_2(x, y, \lambda) := \min_{(x,y,\lambda)\in\mathcal{P}} Q(x, y) + \beta(Cx + Dy - b)_j$$

otherwise. By doing so, the penalty terms in the objective of (5) are split in linear parts, which leads to convex-quadratic subproblems S_1, S_2, $\tilde{S}_1$, and $\tilde{S}_2$. An arbitrary node problem in the B&B tree is then defined as

$$\min_{(x,y,\lambda)\in\mathcal{P}} \pi_N(x, y, \lambda) \tag{6}$$

with

$$\pi_N(x, y, \lambda) := Q(x, y) + \alpha \left(\sum_{i\in Z} x_i + \sum_{i\in O} (1 - x_i) \right) + \beta \left(\sum_{j\in D} \lambda_j + \sum_{j\in P} (Cx + Dy - b)_j \right)$$

and the tuple $N := (Z, O, D, P)$ of index sets containing the indices for which binary variables are driven to 0 (Z) or to 1 (O) as well as for which either the dual variable λ_j is driven to 0 (D) or for which the primal constraint is driven to 0 (P). The overall method is given in Alg. 1. In Line 1 of Alg. 1 the set $\mathcal{N}$ of

Algorithm 1: A penalty B&B method to solve Problem (5)

Data: A bilevel problem of the form (1) and $\alpha, \beta > 0$.
Result: A globally optimal solution (x^*, y^*, λ^*) of Problem (5).
1 Set $\mathcal{N} \leftarrow \{(\emptyset, \emptyset, \emptyset, \emptyset)\}$ and $u \leftarrow \infty$.
2 **while** $\mathcal{N} \neq \emptyset$ **do**
3 Choose an $N \in \mathcal{N}$ and set $\mathcal{N} \leftarrow \mathcal{N} \backslash \{N\}$.
4 Solve Problem (6) for N and obtain the solution (x_N, y_N, λ_N).
5 **if** $\pi(x_N, y_N, \lambda_N) < u$ **then**
6 Set $(x^*, y^*, \lambda^*) \leftarrow (x_N, y_N, \lambda_N)$ and $u \leftarrow \pi(x_N, y_N, \lambda_N)$.
7 **if** $\pi_N(x_N, y_N, \lambda_N) < u$ ***and*** *(*$\exists i \in I \backslash (Z \cup O)$ ***or*** $\exists j \in [\ell] \backslash (D \cup P)$*)* **then**
8 Choose either an $i \in I \backslash (Z \cup O)$ and set $\mathcal{N} \leftarrow \mathcal{N} \cup \{(Z \cup \{i\}, O, D, P), (Z, O \cup \{i\}, D, P)\}$ or a $j \in [\ell] \backslash (D \cup P)$ and set $\mathcal{N} \leftarrow \mathcal{N} \cup \{(Z, O, D \cup \{j\}, P), (Z, O, D, P \cup \{j\})\}$.
9 **end**
10 **return** (x^*, y^*, λ^*).

open nodes is initialized with the root node and the incumbent u is set to infinity. Every time a node in the B&B tree is solved, the objective function of the penalty reformulation (5) is evaluated at the solution to check if a new incumbent is found (Line 5–6). The branching step is performed in the Lines 7–8. Note that we, of course, only choose integer variables or complementarity constraints as branching candidates that are fractional or violated.

To verify the correctness of the method, we show that a complete evaluation of the B&B tree yields an optimal solution of Problem (5).

Lemma 3 (See Lemma 1 in [7])**.** *Let (x^*, y^*, λ^*) be a globally optimal solution of Problem* (5)*. Then, $\pi(x^*, y^*, \lambda^*) = \min\{\pi_N(x_N, y_N, \lambda_N) \colon N \in \mathcal{N}'\}$ holds, where the minimum is taken over all tuples in $\mathcal{N}'$, which contains those $N = (Z, O, D, P)$ with $Z \cup O = I$, $Z \cap O = \emptyset$, $D \cup P = [\ell]$, and $D \cap P = \emptyset$.*

Furthermore, we state that the implicit pruning in Line 7 of Alg. 1 is correct. To this end, we show that the optimal objective value of a node is not larger then the optimal objective value of any node below in the tree.

Lemma 4 (See Lemma 2 in [7])**.** *Let $N' = (Z', O', D', P')$ be a successor node of $N = (Z, O, D, P)$ in the B&B tree, i.e., $Z \subseteq Z'$, $O \subseteq O'$, $D \subseteq D'$, and $P \subseteq P'$ holds. For an optimal solution (x_N, y_N, λ_N) of the problem in node N, it holds $\pi_N(x_N, y_N, \lambda_N) \leq \pi_{N'}(x_{N'}, y_{N'}, \lambda_{N'})$.*

We can now establish a correctness theorem for Alg. 1.

Theorem 5 (See Thm. 1 in [7])**.** *Suppose that the root node relaxation of Problem* (5) *is feasible and bounded from below. Then, Alg. 1 terminates after finitely many steps with a globally optimal solution of Problem* (5)*.*

We have just seen that Alg. 1 computes globally optimal points for the penalty reformulation (5). However, the point is only optimal for the bilevel problem (1) if $F_u(x^*, y^*, \lambda^*) = 0$ holds, i.e., if the solution satisfies all integrality and complementarity constraints. To address this issue, we first show that the unweighted infeasibility measure F_u converges to zero if α and β tend to infinity.

Theorem 6. *Let the root node relaxation of* (5) *be feasible and bounded from below and let the KKT reformulation* (4) *be solvable. Furthermore, let $(\bar{x}(\alpha, \beta)$,*

$\bar{y}(\alpha,\beta), \bar{\lambda}(\alpha,\beta))$ be a global solution of Problem (5) in dependence of the penalty parameters α and β. Then, it holds

$$\lim_{\alpha\to\infty,\beta\to\infty} F_u(\bar{x}(\alpha,\beta), \bar{y}(\alpha,\beta), \bar{\lambda}(\alpha,\beta)) = 0.$$

Proof. Let (x^*, y^*) be a globally optimal solution of Problem (4). Due to optimality, it holds

$$\begin{aligned} &\pi(\alpha,\beta,\bar{x}(\alpha,\beta), \bar{y}(\alpha,\beta), \bar{\lambda}(\alpha,\beta)) \\ \leq\ &\pi(x^*, y^*, \lambda^*) = Q(x^*, y^*) + F_w(\alpha,\beta,x^*,y^*,\lambda^*) = Q(x^*, y^*). \end{aligned} \tag{7}$$

Let (x_r, y_r, λ_r) be the optimal solution and $Q_r := Q(x_r, y_r) > -\infty$ be the optimal objective value of the root node relaxation. Note that the latter is independent of the penalty parameters α and β. Using Inequality (7), we get

$$\begin{aligned} Q(x^*, y^*) &\geq \pi(\alpha,\beta,\bar{x}(\alpha,\beta), \bar{y}(\alpha,\beta), \bar{\lambda}(\alpha,\beta)) \\ &= Q(\bar{x}(\alpha,\beta), \bar{y}(\alpha,\beta)) + F_w(\alpha,\beta,\bar{x}(\alpha,\beta), \bar{y}(\alpha,\beta), \bar{\lambda}(\alpha,\beta)) \\ &\geq Q_r + F_w(\alpha,\beta,\bar{x}(\alpha,\beta), \bar{y}(\alpha,\beta), \bar{\lambda}(\alpha,\beta)). \end{aligned}$$

The last inequality follows from the optimality of (x_r, y_r, λ_r) for the root node. By using the definition of F_w, we get

$$\alpha \sum_{i\in I} \min\{\bar{x}_i, 1-\bar{x}_i\} + \beta \sum_{j=1}^{\ell} \min\{\bar{\lambda}_j, (C\bar{x} + D\bar{y} - b)_j\} \leq Q(x^*, y^*) - Q_r$$

with constant right-hand site $Q(x^*, y^*) - Q_r$. Now, we have

$$\begin{aligned} \sum_{i\in I} \min\{\bar{x}_i, 1-\bar{x}_i\} &\leq \frac{Q(x^*, y^*) - Q_r}{\alpha}, \\ \sum_{j=1}^{\ell} \min\{\bar{\lambda}_j, (C\bar{x} + D\bar{y} - b)_j\} &\leq \frac{Q(x^*, y^*) - Q_r}{\beta} \end{aligned} \tag{8}$$

and taking the limits yields

$$\lim_{\alpha\to\infty,\beta\to\infty} \left(\sum_{i\in I} \min\{\bar{x}_i, 1-\bar{x}_i\} + \sum_{j=1}^{\ell} \min\{\bar{\lambda}_j, (C\bar{x} + D\bar{y} - b)_j\} \right) = 0. \qquad \square$$

To make use of Thm. 6, we modify Alg. 1 so that the penalty parameters are increased after every evaluation of the B&B tree. Furthermore, we relax the condition $F_u(x^*, y^*, \lambda^*) = 0$ to $F_u(x^*, y^*, \lambda^*) \leq t$ for a given tolerance $t > 0$.

Theorem 7. *Let the root node relaxation of (5) be feasible and bounded from below, let the KKT reformulation (4) be solvable, and let $t > 0$ be given. Furthermore, let $(\bar{x}(\alpha,\beta), \bar{y}(\alpha,\beta), \bar{\lambda}(\alpha,\beta))$ be the solution of Problem (5) for penalty parameters α and β. Then, there exists finite penalty parameters α and β with*

$$F_u(\bar{x}(\alpha,\beta), \bar{y}(\alpha,\beta), \bar{\lambda}(\alpha,\beta)) \leq t.$$

Proof. Let (x^*, y^*) be a globally optimal solution of Problem (4). Using (8), for $\alpha, \beta \geq 2(Q(x^*, y^*) - Q_w)/t$, we get

$$\begin{aligned} \sum_{i\in I} \min\{\bar{x}_i, 1-\bar{x}_i\} &\leq \frac{(Q(x^*, y^*) - Q_w)t}{2(Q(x^*, y^*) - Q_w)} = \frac{t}{2}, \\ \sum_{j=1}^{\ell} \min\{\bar{\lambda}_j, (C\bar{x} + D\bar{y} - b)_j\} &\leq \frac{(Q(x^*, y^*) - Q_w)t}{2(Q(x^*, y^*) - Q_w)} = \frac{t}{2} \end{aligned}$$

and, thus, $F_u(\bar{x}(\alpha,\beta), \bar{y}(\alpha,\beta), \bar{\lambda}(\alpha,\beta)) \leq t$. $\square$

Since the value of $Q(x^*, y^*)$ is unknown in practice, one can also use the value $Q(\tilde{x}, \tilde{y})$ of an already obtained point $(\tilde{x}, \tilde{y})$, which is feasible for the KKT reformulation (4). This yields a valid and finite, but also worse, upper bound for the unweighted infeasibility F_u since $Q(x^*, y^*) \leq Q(\tilde{x}, \tilde{y})$ holds.

Definition 8. *A point $(x^*, y^*, \lambda^*) \in \mathcal{P}$ is called* approximate solution *of Problem* (1)*, if it minimizes $Q(x, y)$ and satisfies the condition $F_u(x^*, y^*, \lambda^*) \leq t$.*

We can now present a multi-tree penalty B&B method that computes an approximate solution to the bilevel problem (1) by repeatedly applying Alg. 1 for increasing penalty parameters α and β; see Alg. 2.

Algorithm 2: A penalty B&B method for MIQP-QP bilevel problems

Data: A bilevel problem of the form (1), $\alpha^1 > 0$, $\beta^1 > 0$, $t > 0$, and $\tau > 0$.
Result: An approximate solution (x^*, y^*, λ^*) of the bilevel problem (1).

1 Set $\mathcal{N} \leftarrow \{(\emptyset, \emptyset, \emptyset, \emptyset)\}$, $u \leftarrow \infty$, $u_{\max} \leftarrow \infty$, $\alpha_{\max} \leftarrow \infty$, $\beta_{\max} \leftarrow \infty$, and $k \leftarrow 0$.
2 **while** $\alpha^k \leq \alpha_{max}, \beta^k \leq \beta_{max}$ **do**
3   Apply Alg. 1.
4   **if** $\pi^k(x_N^k, y_N^k, \lambda_N^k) < u$ *for node N in iteration k* **then**
5   Set $(x_{\text{opt}}^k, y_{\text{opt}}^k, \lambda_{\text{opt}}^k) \leftarrow (x_N^k, y_N^k, \lambda_N^k)$ and $u \leftarrow \pi^k(x_N^k, y_N^k, \lambda_N^k)$.
6   **if** $F_u(x_N^k, y_N^k, \lambda_N^k) \leq t$ **then**
7   Set $\alpha_{\max}, \beta_{\max} \leftarrow \frac{2(Q(x_N^k, y_N^k) - Q_r)}{t}$.
8   **if** $F_u(x_N^k, y_N^k, \lambda_N^k) = 0$ **then** set $u_{\max} \leftarrow u$.
9   **end**
10   **if** $F_u(x_{opt}^k, y_{opt}^k, \lambda_{opt}^k) > t$ **then**
11   Set $u \leftarrow u_{\max} + \tau$ and $\mathcal{N} \leftarrow \{(\emptyset, \emptyset, \emptyset, \emptyset)\}$.
12   Choose new penalty parameters $\alpha^{k+1} > \alpha^k$ and $\beta^{k+1} > \beta^k$.
13   **else**
14   Set $\alpha^{k+1} > \alpha_{\max}$ and $\beta^{k+1} > \beta_{\max}$ to terminate the algorithm.
15   **end**
16   Set $k \leftarrow k + 1$.
17 **end**
18 **return** $(x^*, y^*, \lambda^*) := (x_{\text{opt}}^k, y_{\text{opt}}^k, \lambda_{\text{opt}}^k)$

In Line 3, Alg. 1 is applied to obtain a globally optimal solution of Problem (5) for α^k and β^k. If the solution is not an approximate solution of Problem (1), the penalty parameters are increased and the procedure is repeated (Lines 10–12). Further steps are added in Lines 6–8, where every node solution that yields a better objective value than the incumbent is checked for feasibility w.r.t. the KKT reformulation (4). Note that Q_r in Line 7 stands for Q evaluated at the root node's solution. If the unweighted infeasibility measure F_u for this point is within the tolerance t, we update the upper bounds of the penalty parameters according to the proof of Thm. 7. Moreover, if $F_u = 0$ holds, we update the upper bound of the incumbent, which can be used in further evaluations of the B&B tree to initialize the incumbent instead of setting it to infinity; see Line 11. Therefore, it has to be modified by the value τ, which is also shown in the following theorem.

Theorem 9. *Let the KKT reformulation* (4) *be solvable, the root node relaxation of* (5) *be feasible and bounded from below, and let $t, \tau > 0$ be given. Furthermore, suppose that $\alpha^{k+1} - \alpha^k \geq \varepsilon$ and $\beta^{k+1} - \beta^k \geq \varepsilon$ holds for a fixed $\varepsilon > 0$ and every k. Then, Alg.* 2 *terminates after finitely many steps in an iteration k^* with a globally* *optimal solution* $(x_{opt}^{k^*}, y_{opt}^{k^*}, \lambda_{opt}^{k^*})$ *of the penalty reformulation with* α^{k^*} *and* β^{k^*} *that satisfy the condition* $F_u(x_{opt}^{k^*}, y_{opt}^{k^*}, \lambda_{opt}^{k^*}) \leq t$.

Proof. According to Thm. 7, the algorithm terminates with finite penalty parameters $\alpha^k \leq \alpha_{\max}$ and $\beta^k \leq \beta_{\max}$ with a solution that satisfies the condition of the theorem. Using the claim that the increase of the penalty parameters is not arbitrarily small, we obtain

$$k_{\max} \leq \max\left\{\frac{\alpha_{\max} - \alpha^1}{\epsilon}, \frac{\beta_{\max} - \beta^1}{\epsilon}\right\} < \infty,$$

where $k_{\max}$ denotes the number of maximally required penalty parameter updates. According to Thm. 5, a complete evaluation of the B&B tree takes only finitely many steps so that Alg. 2 also terminates in finite time. It remains to show that the update $u \leftarrow u_{\max} + \tau$ at the end of iteration k cannot lead to overlooking a globally optimal solution of Problem (5) with new penalty parameters α^{k+1} and β^{k+1} in Line 4 of the next iteration $k+1$. We obtain $u_{\max}$ in iteration k by the objective value π^k of a node's solution that satisfies all integrality and complementarity conditions. Let $(\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ be this solution, i.e., $(\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ is the optimal solution of a node N for iteration k for which $u_{\max} = \pi^k(\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ holds. Additionally, it is feasible for the KKT reformulation (4), i.e., it holds

$$F_u(\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) = \sum_{i \in I} \min\left\{\tilde{x}_i^k, 1 - \tilde{x}_i^k\right\} + \sum_{j=1}^{\ell} \min\left\{\tilde{\lambda}_i^k, (C\tilde{x} + D\tilde{y} - b)_i\right\} = 0.$$

Then, the point $(\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ is also feasible for the node $\tilde{N} = (\tilde{Z}, \tilde{O}, \tilde{D}, \tilde{P})$ with $\tilde{Z} := \{i \in I \colon \tilde{x}_i^k \leq 1 - \tilde{x}_i^k\}$, $\tilde{O} := \{i \in I \colon \tilde{x}_i^k > 1 - \tilde{x}_i^k\}$, $\tilde{D} := \{j \in \{1, \dots, \ell\} \colon \tilde{\lambda}_j^k \leq (C\tilde{x}^k + D\tilde{y}^k - b)_j\}$, and $\tilde{P} := \{j \in \{1, \dots, \ell\} \colon \tilde{\lambda}_j^k > (C\tilde{x}^k + D\tilde{y}^k - b)_j\}$. Furthermore, it holds

$$\alpha^k \left(\sum_{i \in \tilde{N}} \tilde{x}_i^k + \sum_{i \in \tilde{O}} (1 - \tilde{x}_i^k)\right) + \beta^k \left(\sum_{j \in \tilde{D}} \tilde{\lambda}_j^k + \sum_{j \in \tilde{P}} (C\tilde{x}^k + D\tilde{y}^k - b)_j\right) = 0,$$

which is why the point $(\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ is also an optimal solution of the node $\tilde{N}$ for every other $\alpha^{k+1} > \alpha^k$ and $\beta^{k+1} > \beta^k$. If we use the incumbent

$$u = u_{\max} + \tau > \pi^k(\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) = \pi_{\tilde{N}}^k(\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k), \tag{9}$$

in Alg. 2 in iteration $k+1$, then for an optimal solution $(x_{\text{opt}}^{k+1}, y_{\text{opt}}^{k+1}, \lambda_{\text{opt}}^{k+1})$ of Problem (5) with penalty parameters α^{k+1} and β^{k+1}, it holds

$$\pi^{k+1}(x_{\text{opt}}^{k+1}, y_{\text{opt}}^{k+1}, \lambda_{\text{opt}}^{k+1}) \leq \pi_{\tilde{N}}^{k+1}(\tilde{x}^{k+1}, \tilde{y}^{k+1}, \tilde{\lambda}^{k+1}) = \pi_{\tilde{N}}^k(\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) < u. \tag{10}$$

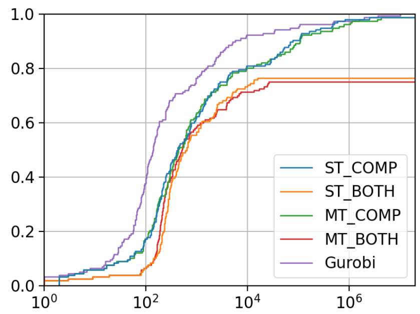
The first inequality follows from Lemma 3. Due to the strict inequality in (10), the optimal solution $(x_{\text{opt}}^{k+1}, y_{\text{opt}}^{k+1}, \lambda_{\text{opt}}^{k+1})$ cannot be overlooked in Line 4 of iteration $k+1$ of Alg. 2. To guarantee this, τ has to be chosen strictly positive; see (9). □

Remark 10. *Let us close this section with two remarks.*

(1) *We also derived a single-tree version of Alg. 2 in which we check for every node if the obtained solution fulfills all constraints that have been added by branching so far, i.e., if*

$$\sum_{i \in Z} x_{N,i} + \sum_{i \in O} (1 - x_{N,i}) + \sum_{j \in D} \lambda_{N,j} + \sum_{j \in P} (Cx_N + Dy_N - b)_j = 0 \tag{11}$$

holds. If this is not the case, we set $Q(x, y) = 0$ *and solve the node again to check if the node is infeasible w.r.t. the integrality and complementarity constraints. If it is, we prune the node and, otherwise, we increase the*



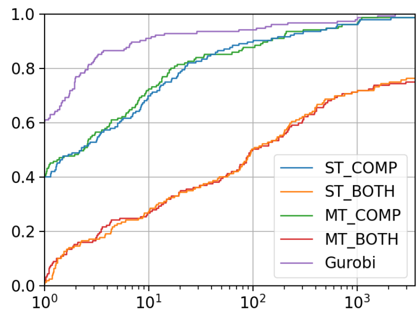


FIGURE 1. Node counts (left) and running times (in s; right) vs. percentage of instances

parameters until condition (11) is satisfied in the associated solution. By doing so, we can prune nodes more efficiently and only a single B&B tree is required. However, we have to solve nodes multiple times so that it is not clear a priori if the single-tree is performing better. We will discuss this in more detail in the next section.

(2) *Furthermore, we derived a modified version of Alg. 2 in which we do not branch on the integrality conditions but only on the complementarity constraints and let the MILP solver handle the other ones. We compare these two different variants in the next section as well.*

3. Numerical Results

In this section, we present a brief and preliminary numerical comparison of the penalty B&B method with a benchmark approach. Since reformulating the bilevel problem using the KKT conditions of the lower level is still the most frequently used method in practice, we choose Gurobi 9.5.1 (with cuts, presolve, and heuristics deactivated to get a fair comparison) applied to this formulation as a benchmark, where we handle complementarity constraints via SOS1 conditions; see [6].

Our test set consists of a subset of the instance collections Denegre, Int0Sum, MIPLIB2010, MIPLIB2017, MIPLIB3, Small, and Xuwang as used in [5]. Moreover, we excluded all instances that all methods can solve in less than 1 s and all instances that no method can solve within the time limit of 1 h; leading to 157 remaining instances in total. Our algorithm has been implemented in Python 3.9.7 and all occurring subproblems have been solved using Gurobi 9.5.1. All 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).[1] We used a single Intel XEON SP 6126 core with 2.6 GHz and 32 GB RAM.

As discussed in Remark 10, we tested both a multi-tree (MT) and a single-tree (ST) variant of our method as well as a version in which we branch on complementarity and integrality constraints (BOTH; in this case, we first branch on integrality constraints and then on complementarity constraints) and on complementarity constraints only (COMP). While doing so, we always choose the most violated integrality or complementarity constraint for branching in a depth-first search. Moreover, we set $\alpha^1 = \beta^1 = 100$ with update factor 100 and $\tau = 10^{-1}$ and $t = 10^{-4}$. The results are given in Fig. 1.

[1]We kindly acknowledge the support of RHRK.

It can be seen that the penalty B&B method is more effective if we only apply the penalty branching to complementarity conditions and let Gurobi handle the integrality constraints. The proposed method is rather competitive with the benchmark when it comes to node counts, although it is still slightly outperformed by it. The running times are worse, which can be devoted to our prototypical implementation that is not comparable to a commercial software. Based on the results for the node counts, we see a clear potential of the new method. However, it needs to be improved–both w.r.t. further algorithmic techniques as well as w.r.t. its implementation—in order to obtain a method that is as powerful as the commercial state-of-the-art.

Acknowledgements

The first author thanks the DFG for their support within RTG 2126 "Algorithmic Optimization".

References

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

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

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

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

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

[6] T. Kleinert and M. Schmidt. *Why there is no need to use a big-M in linear bilevel optimization: A computational study of two ready-to-use approaches*. Tech. rep. 2020. URL: http://www.optimization-online.org/DB_HTML/2020/10/8065.html.

[7] M. de Santis, S. de Vries, M. Schmidt, and L. Winkel. "A Penalty Branch-and-Bound Method for Mixed-Binary Linear Complementarity Problems." In: *INFORMS Journal on Computing* (2022). Forthcoming.

[8] J. C. Smith and Y. Song. "A survey of network interdiction models and algorithms." In: *European Journal of Operational Research* 283.3 (2020), pp. 797–811. DOI: 10.1016/j.ejor.2019.06.024.

[9] S. Wogrin, S. Pineda, and D. A. Tejada-Arango. "Applications of Bilevel Optimization in Energy and Electricity Markets." In: *Bilevel Optimization: Advances and Next Challenges*. Ed. by S. Dempe and A. Zemkoho. Cham: Springer International Publishing, 2020, pp. 139–168. DOI: 10.1007/978-3-030-52119-6_5.

(A. Horländer, M. Schmidt) Trier University, Department of Mathematics, Universitätsring 15, 54296 Trier, Germany

Email address: horlaender@uni-trier.de

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