

COMPUTING FEASIBLE POINTS OF BILEVEL PROBLEMS WITH A PENALTY ALTERNATING DIRECTION METHOD

THOMAS KLEINERT^{1,2}, MARTIN SCHMIDT^{3,2}

ABSTRACT. Bilevel problems are highly challenging optimization problems that appear in many applications of energy market design, critical infrastructure defense, transportation, pricing, etc. Often, these bilevel models are equipped with integer decisions, which makes the problems even harder to solve. Typically, in such a setting in mathematical optimization one develops primal heuristics in order to obtain feasible points of good quality quickly or to enhance the search process of exact global methods. However, there are comparably few heuristics for bilevel problems. In this paper, we develop such a primal heuristic for bilevel problems with mixed-integer linear or quadratic upper level and linear or quadratic lower level. The heuristic is based on a penalty alternating direction method, which allows for a theoretical analysis. We derive a convergence theory stating that the method converges to a stationary point of an equivalent single-level reformulation of the bilevel problem and extensively test the method on a test set of more than 2800 instances—which is one of the largest computational test sets ever used in bilevel programming. The study illustrates the very good performance of the proposed method, both in terms of running times and solution quality. This renders the method a suitable sub-routine in global bilevel solvers as well as a reasonable standalone approach.

1. INTRODUCTION

Bilevel optimization problems have become an increasingly important class of optimization problems during the last years and decades. They often appear in practice, where they are used to model hierarchical decision processes in which a leader makes a decision anticipating the rational reaction of a follower. These leader-follower or Stackelberg games date back to the seminal papers [66, 67]. Such situations arise, e.g., in energy markets [1, 19, 39–42, 45, 49], in critical infrastructure defense [15, 25], or in pricing models [51]. Since bilevel problems can model highly complicated decision processes it is not surprising that they are also hard to solve. In [24, 43] it is shown that even the easiest instantiation, namely bilevel problems with linear upper and lower level, is strongly NP hard. Other hardness results can be found in [6]. Going even further, the hardness of p -level linear problems is studied in [46]. The more general case of mixed-integer linear bilevel problems is Σ_p^2 -hard; cf. [54]. Moreover, checking local optimality for a given point is NP hard as well [72]. For general surveys of bilevel optimization see [5, 22–24] and [73] for a survey focusing on linear-linear bilevel problems.

For other optimization problems with similar hardness characteristics it is common to consider the computation of local optima or to develop approximation methods or heuristics that quickly deliver feasible points of good quality. See, e.g., the PhD thesis [8] for an overview about primal heuristics for mixed-integer (non)linear optimization problems. Such heuristics can then be applied as standalone procedures

Date: November 6, 2019.

2010 Mathematics Subject Classification. 90-08, 90C11, 90C46.

Key words and phrases. Bilevel optimization, Mixed-integer bilevel optimization, Stationary points, Penalty methods, Alternating direction methods.

or as sub-procedures that are used to enhance the search process of exact, i.e., global, optimization methods. However, the development of primal heuristics for bilevel problems is much harder than for non-hierarchical problems. The reason is that a bilevel feasible point already needs to be a global optimum of an optimization problem—namely the follower’s problem. Thus, only a few heuristics are known for bilevel problems and most of them are tailored for a specific class of applications [12, 16, 27, 32, 47] or rely on meta-heuristic ideas; see, e.g., [52, 53, 57, 60], the book [69], or the references in Section 6.7 of [22]. In [3], the authors consider an approximation method in which only near-optimal solutions are considered for the costly lower-level problem. Recently, neural networks also have been applied to solve linear bilevel problems [44]. In summary, at least to the best of our knowledge, there seem to be much less primal heuristics for bilevel optimization compared to other NP hard optimization tasks like mixed-integer programming. Moreover, only a few algorithms are known for computing a local optimum of a linear bilevel problem; see, e.g., [21, 68] for problems without upper-level constraints or [25] for two heuristics for general mixed-integer linear bilevel problems. One of the latter heuristics is the so-called “Stationary Point Heuristic” that alternately solves the high-point relaxation for fixed lower-level variables and the primal lower-level problem for a fixed upper-level solution of the high-point relaxation. Thus, the method applies an upper level/lower level splitting and solves two different problems alternately. To the best of our knowledge, there is no convergence theory for this approach. In contrast, our proposed approach uses a single-level reformulation and exploits its block separability, which results in a primal/dual splitting for which we provide a detailed convergence analysis and an extensive numerical study. For penalty methods for bilevel optimization see, e.g., [2, 14].

The discussed “gap” in the bilevel literature is the main motivation for this paper. We develop a general-purpose primal heuristic for bilevel problems with (mixed-integer) linear or quadratic upper and linear or quadratic lower-level problems. The algorithm is based on the penalty alternating direction method (PADM) developed in [35] that has also been applied to other fields like gas transport optimization or supply chain management in [36, 37, 65]. This also allows for a theoretical analysis of the method. We prove that the PADM converges to stationary points of an equivalent single-level reformulation of the bilevel problem using the strong duality theorem for the lower level. Finally, we evaluate the algorithm in a very extensive numerical study with over 2800 instances for each different class of bilevel problems—namely with (mixed-integer) linear or quadratic upper levels and linear lower levels as well as for the same problems but with quadratic lower levels. We compare the outcomes of the PADM with the results of the classical KKT (Karush–Kuhn–Tucker) approach; cf., e.g., Chapter 3 of the book [24]. This comparison is carried out both in terms of running times and solution quality. It turns out that the PADM-based method clearly outperforms the KKT approach in both measures. Due to the good solution quality, the method can be used as a reasonable standalone algorithm and due to the very fast running times, it can also serve as a sub-procedure in global methods for solving mixed-integer bilevel problems like, e.g., [7, 28, 56].

The remainder of the paper is structured as follows. In Section 2, we start by introducing linear-linear (LP-LP) bilevel problems and illustrate the strong-duality based single-level reformulation for this class of problems. The following Section 3 introduces the basics of alternating direction methods (ADMs) and reviews the extension of penalty ADMs (PADMs). The main convergence results of ADMs and PADMs are also given in this section. Section 4 then shows how to apply the PADM to LP-LP bilevel problems, states the convergence results for PADMs applied to bilevel problems, and also explains the extension to bilevel problems with

a mixed-integer quadratic upper level (MIQP-LP bilevel problems). In Section 5, we show how to extend the previously discussed ideas for bilevel problems with quadratic follower problems. The numerical results are presented in Section 6 and the paper closes with a conclusion and some comments on possible future research directions in Section 7.

2. PROBLEM STATEMENT AND SINGLE-LEVEL REFORMULATION

We start by considering linear bilevel problems of the form

$$\min_{x,y} c^\top x + d^\top y \quad (1a)$$

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

$$y \in \arg \max_{\bar{y}} \{e^\top \bar{y} : Cx + D\bar{y} \leq b\}, \quad (1c)$$

with objective function coefficient vectors $c \in \mathbb{R}^n$, $d, e \in \mathbb{R}^m$, right-hand side vectors $a \in \mathbb{R}^k$, $b \in \mathbb{R}^\ell$, and matrices $A \in \mathbb{R}^{k \times n}$, $B \in \mathbb{R}^{k \times m}$, $C \in \mathbb{R}^{\ell \times n}$, $D \in \mathbb{R}^{\ell \times m}$. We refer to x as the upper-level variables (or leader decisions) and to y as the lower-level variables (or follower decisions). The Problem (1a), (1b) is called the upper-level problem and (1c) is the lower-level problem.

Considered as a parametric linear optimization problem (LP), the lower-level problem reads

$$\max_y e^\top y \quad \text{s.t. } Dy \leq b - C\bar{x}, \quad (2)$$

where the leader's decision \bar{x} is the parameter of the problem. For more information on parametric optimization in the context of bilevel optimization we refer to Chapter 4 of the book [23]. The dual problem of Problem (2) is given by

$$\min_{\lambda} (b - C\bar{x})^\top \lambda \quad (3a)$$

$$\text{s.t. } D^\top \lambda = e, \quad (3b)$$

$$\lambda \geq 0. \quad (3c)$$

Note that the dual polyhedron $D := \{\lambda \in \mathbb{R}^\ell : D^\top \lambda = e, \lambda \geq 0\}$ does not depend on primal upper-level variables. By applying the strong duality theorem to the lower-level problem, one can equivalently reformulate the bilevel problem (1) to the single-level problem

$$\min_{x,y,\lambda} c^\top x + d^\top y \quad (4a)$$

$$\text{s.t. } Ax + By \geq a, \quad (4b)$$

$$Cx + Dy \leq b, \quad (4c)$$

$$D^\top \lambda = e, \quad (4d)$$

$$\lambda \geq 0, \quad (4e)$$

$$e^\top y - b^\top \lambda + x^\top C^\top \lambda \geq 0. \quad (4f)$$

Note that every point (x, y, λ) that satisfies Constraints (4b)–(4e) always satisfies

$$e^\top y - b^\top \lambda + x^\top C^\top \lambda \leq 0 \quad (5)$$

by weak duality of linear optimization; see, e.g., [18]. This together with Constraint (4f) thus implies the strong duality condition and hence lower-level optimality. As a consequence, Problem (1) and (4) are equivalent in the following way: (x^*, y^*) is a solution of Problem (1) if and only if there exists a λ^* such that (x^*, y^*, λ^*) is a solution of Problem (4). We also remark that, in the case of multiplicities in the solution set of the lower level, one typically has to take care about optimistic

Algorithm 1 A Standard Alternating Direction Method

1: Choose initial values $(x^0, y^0) \in X \times Y$.

2: **for** $i = 0, 1, \dots$ **do**

3: Compute

$$x^{i+1} \in \arg \min_x \{f(x, y^i) : g(x, y^i) = 0, h(x, y^i) \geq 0, x \in X\}.$$

4: Compute

$$y^{i+1} \in \arg \min_y \{f(x^{i+1}, y) : g(x^{i+1}, y) = 0, h(x^{i+1}, y) \geq 0, y \in Y\}.$$

5: **end for**

vs. pessimistic solutions [23]. In the case of the single-level reformulation (4), we however always obtain an optimistic solution.

Note finally that the single-level reformulation consists of a linear objective function, linear constraints, and the single nonconvex constraint (4f). In the latter, the nonconvexity is due to the bilinear term $x^\top C^\top \lambda$ of primal upper-level variables x and dual lower-level variables λ . Hence, just like the bilevel problem (1), the single-level problem (4) is nonconvex.

3. A PENALTY ALTERNATING DIRECTION METHOD

In this section, we review the alternating direction method (ADM) and an extension of this method; the penalty ADM (PADM). Afterward, we discuss how the latter method can be used to compute a stationary point of the classical strong-duality based single-level reformulation of the linear bilevel problem (1) in Section 4.

We start with discussing general ADMs. To this end, we consider an optimization problem in the specific form

$$\min_{x,y} f(x, y) \tag{6a}$$

$$\text{s.t. } g(x, y) = 0, \quad h(x, y) \geq 0, \tag{6b}$$

$$x \in X, \quad y \in Y. \tag{6c}$$

Here, $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ are variable vectors. The feasible set of this problem is abbreviated by

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

For discussing the theoretical properties of ADMs, we need the following assumption.

Assumption 1. *The objective function $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ and the constraint functions $g : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^k$ and $h : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^\ell$ are continuous and the sets X and Y are non-empty and compact.*

A standard ADM proceeds as follows. Given an iterate (x^i, y^i) we first solve Problem (6) for y fixed to y^i . Thus, we obtain a new x -iterate x^{i+1} . We now fix x to this new iterate x^{i+1} , solve Problem (6) again, and obtain y^{i+1} . Repeating these two steps yields the method that is listed in Algorithm 1.

Under certain mild assumptions one can show that the ADM of Algorithm 1 converges to a partial minimum, i.e., to a point (x^*, y^*) for which

$$f(x^*, y^*) \leq f(x, y^*) \quad \text{for all } (x, y^*) \in \Omega,$$

$$f(x^*, y^*) \leq f(x^*, y) \quad \text{for all } (x^*, y) \in \Omega$$

holds. The following general convergence result is taken from [38].

Theorem 1. Let $\{(x^i, y^i)\}_{i=0}^{\infty}$ be a sequence with $(x^{i+1}, y^{i+1}) \in \Theta(x^i, y^i)$, where $\Theta(x^i, y^i) := \{(x^*, y^*): f(x^*, y^i) \leq f(x, y^i) \forall x \in X; f(x^*, y^*) \leq f(x^*, y) \forall y \in Y\}$. Suppose that Assumption 1 holds and that the solution of the first optimization problem is always unique. Then, every convergent subsequence of $\{(x^i, y^i)\}_{i=0}^{\infty}$ converges to a partial minimum. For two limit points z, z' of such subsequences it holds that $f(z) = f(z')$.

For what follows, we also note that stronger convergence results can be obtained if stronger assumptions on f and Ω are made. For later reference, we state these results as a corollary; see [35, 36, 38, 74] for the proofs and more detailed discussions.

Corollary 1. Suppose that the assumptions of Theorem 1 are satisfied. Then, the following holds:

- (i) If f is continuously differentiable, then every convergent subsequence of $\{(x^i, y^i)\}_{i=0}^{\infty}$ converges to a stationary point of Problem (6).
- (ii) If f is continuously differentiable and if f and Ω are convex, then every convergent subsequence of $\{(x^i, y^i)\}_{i=0}^{\infty}$ converges to a global minimum of Problem (6).

Let us comment on the main rationale of the alternating direction method discussed so far. The considered Problem (6) can be seen as a quasi block-separable problem, where the blocks are given by the variables x and y as well as their respective feasible sets X and Y . We add the notion “quasi” here since there still are the constraints g and h that couple the feasible sets of the two blocks. The main idea of an ADM is to alternately solve in the directions of the blocks separately until the method stagnates.

In practice it can often be observed that an even stronger decoupling of Problem (6) is favorable [10, 35–37]. Thus, we now go one step further and relax the coupling constraints g and h . To this end, we introduce the weighted ℓ_1 penalty function

$$\phi_1(x, y; \mu, \rho) := f(x, y) + \sum_{t=1}^k \mu_t |g_t(x, y)| + \sum_{t=1}^{\ell} \rho_t [h_t(x, y)]^-.$$

Here, $[\alpha]^- := \max\{0, -\alpha\}$ holds and μ and ρ are vectors of penalty parameters of size k and ℓ , respectively. The penalty ADM consists of an inner and an outer loop. In the inner loop we apply a standard ADM like in Algorithm 1 to the penalty problem

$$\min_{x, y} \phi_1(x, y; \mu, \rho) \quad \text{s.t.} \quad x \in X, y \in Y. \quad (7)$$

If this inner loop iteration terminates with a partial minimum of Problem (7), we check whether the coupling constraints are satisfied. If they are, we terminate. If not, we increase the penalty parameters and proceed with computing a partial minimum of the new penalty problem in the next inner loop. This method is formally stated in Algorithm 2. For later reference, we also state the convergence results for the PADM algorithm 2, which have been derived in [35]. There, all details and proofs can be found.

Theorem 2. Suppose that Assumption 1 holds and that $\mu_t^j \nearrow \infty$ for all $t = 1, \dots, k$ and $\rho_t^j \nearrow \infty$ for all $t = 1, \dots, \ell$. Moreover, let $\{(x^j, y^j)\}_{j=0}^{\infty}$ be a sequence of partial minima of (7) (for $\mu = \mu^j$ and $\rho = \rho^j$) generated by Algorithm 2 with $(x^j, y^j) \rightarrow (x^*, y^*)$. Then, there exist weights $\bar{\mu}, \bar{\rho} \geq 0$ such that (x^*, y^*) is a partial minimizer of the weighted ℓ_1 feasibility measure

$$\chi_{\bar{\mu}, \bar{\rho}}(x, y) := \sum_{t=1}^k \bar{\mu}_t |g_t(x, y)| + \sum_{t=1}^{\ell} \bar{\rho}_t [h_t(x, y)]^-.$$

Algorithm 2 The ℓ_1 Penalty Alternating Direction Method

```

1: Choose initial values  $(x^{0,0}, y^{0,0}) \in X \times Y$  and penalty parameters  $\mu^0, \rho^0 \geq 0$ .
2: for  $j = 0, 1, \dots$  do
3:   Set  $i = 0$ .
4:   while  $(x^{j,i}, y^{j,i})$  is not a partial minimum of (7) with  $\mu = \mu^j$  and  $\rho = \rho^j$  do
5:     Compute  $x^{j,i+1} \in \arg \min_x \{\phi_1(x, y^{j,i}; \mu^j, \rho^j) : x \in X\}$ .
6:     Compute  $y^{j,i+1} \in \arg \min_y \{\phi_1(x^{j,i+1}, y; \mu^j, \rho^j) : y \in Y\}$ .
7:     Set  $i \leftarrow i + 1$ .
8:   end while
9:   Choose new penalty parameters  $\mu^{j+1} \geq \mu^j$  and  $\rho^{j+1} \geq \rho^j$ .
10: end for

```

If, in addition, (x^*, y^*) is feasible for the original problem (6), the following holds:

- (i) If f is continuous, then (x^*, y^*) is a partial minimum of (6).
- (ii) If f is continuously differentiable, then (x^*, y^*) is a stationary point of (6).
- (iii) If f is continuously differentiable and if f and Ω are convex, then (x^*, y^*) is a global optimum of (6).

4. APPLYING THE PADM TO LINEAR BILEVEL PROBLEMS

In this section, we apply the PADM to the single-level reformulation (4) of the original linear bilevel problem (1). As already discussed at the end of Section 2, the problematic constraint is the strong duality inequality (4f). This is due to two reasons. On the one hand, it is the only nonlinear constraint and thus the reason for the nonconvexity of the problem. On the other hand, it is the only constraint that couples the variable blocks (x, y) and λ . Thus, we relax this constraint and obtain the penalty problem reformulation

$$\min_{x,y,\lambda} \quad c^\top x + d^\top y + \rho [e^\top y - b^\top \lambda + x^\top C^\top \lambda]^- \quad (8a)$$

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

$$Cx + Dy \leq b, \quad (8c)$$

$$D^\top \lambda = e, \quad (8d)$$

$$\lambda \geq 0. \quad (8e)$$

Moreover, we smoothen the penalty term by exploiting weak duality (5) of the lower level that is equivalent to that

$$b^\top \lambda - x^\top C^\top \lambda - e^\top y \geq 0$$

holds for every feasible point of Problem (8). Thus,

$$[e^\top y - b^\top \lambda + x^\top C^\top \lambda]^- = \max\{0, b^\top \lambda - x^\top C^\top \lambda - e^\top y\} = b^\top \lambda - x^\top C^\top \lambda - e^\top y$$

holds and we obtain the equivalent penalty problem

$$\min_{x,y,\lambda} \quad c^\top x + d^\top y + \rho (b^\top \lambda - x^\top C^\top \lambda - e^\top y) \quad \text{s.t.} \quad (8b)-(8e), \quad (9)$$

which is a smooth but still nonconvex optimization problem. To be more specific, Problem (9) is an indefinite quadratic optimization problem. A closer look also reveals that Problem (9) is exactly of the form in (7) if the first block of variables is (x, y) and if the second block of variables is λ . Thus, the splitting of the feasible

set is obtained by identifying¹

$$x \in X \quad \longleftrightarrow \quad \text{Constraints (8b), (8c)}, \quad (10a)$$

$$y \in Y \quad \longleftrightarrow \quad \text{Constraints (8d), (8e)}. \quad (10b)$$

Note further that this splitting corresponds to a primal-dual splitting of the single-level reformulation (4). Given this splitting, the first sub-problem that needs to be solved if we apply Algorithm 2 to Problem (9) reads

$$\min_{x,y} \quad c^\top x + d^\top y - \rho \left((C^\top \bar{\lambda})^\top x + e^\top y \right) \quad (11a)$$

$$\text{s.t.} \quad Ax + By \geq a, \quad Cx + Dy \leq b, \quad (11b)$$

where $\bar{\lambda}$ is a given constant vector and where we already omitted the constant objective function term $b^\top \bar{\lambda}$. This problem has the same feasible set as the classical high-point relaxation of the original bilevel problem (1); see, e.g., [56]. However, the objective function coefficients are modified in dependence of the penalty parameter ρ , the current dual estimate $\bar{\lambda}$, and the lower-level objective function coefficients e .

The second sub-problem is equivalent to

$$\min_{\lambda} \quad (b - C\bar{x})^\top \lambda \quad (12a)$$

$$\text{s.t.} \quad D^\top \lambda = e, \quad (12b)$$

$$\lambda \geq 0, \quad (12c)$$

which is exactly the dual lower-level problem (3). Here, three interesting aspects can be observed. First, the second sub-problem only depends on the upper level's primal variables \bar{x} —not on the lower level's primal variables \bar{y} . The reason is that we again omit constant terms in the objective function, i.e., $c^\top \bar{x} + d^\top \bar{y} - \rho e^\top \bar{y}$. Second, sub-problem (12) does not depend on the penalty parameter ρ anymore since it only scales the remaining objective function. Third, the second sub-problem may be unbounded for a given estimate for \bar{x} . This means that the primal lower-level problem is infeasible for the upper-level decision \bar{x} . One can think of several ideas to resolve this situation, e.g., starting the PADM again from a different initial solution or excluding the upper-level solution \bar{x} from sub-problem (11). However, in our numerical studies (cf. Section 6), it turns out that this problem happens only rarely. We thus refrain from a detailed analysis.

By using Corollary 1 and Theorem 2 we now state two theoretical results for the PADM algorithm 2 applied to Problem (4).

Theorem 3. *Consider the inner loop of Algorithm 2 applied to Problem (9) for a fixed penalty parameter $\rho > 0$ and let $\{(x^{j,i}, y^{j,i})\}_{i=0}^\infty$ be the generated sequence of iterates. Moreover, let one of the two sub-problems (11) or (12) always have a unique solution. Then, every convergent subsequence of $\{(x^{j,i}, y^{j,i})\}_{i=0}^\infty$ converges to a stationary point of the penalty problem (9).*

Proof. The feasible set of Problem (9) is a polyhedron and thus convex and the objective function is continuously differentiable but nonconvex. Thus, Case (i) of Corollary 1 applies. \square

Regarding the obtained stationary points, two possible situations may appear. First, the stationary point may have a strong duality error

$$\chi^{\text{sd}}(x, y, \lambda) := |e^\top y - b^\top \lambda + x^\top C^\top \lambda|$$

¹Note that in (10), x and y denote the variable blocks of the general problem formulation (6). All other occurrences of x and y in this and the following sections stand for the respective upper- and lower-level variables of the considered bilevel problem.

of zero, which means that the stationary point is also bilevel feasible. Second, the stationary point has a non-zero strong duality error $\chi^{\text{sd}}(x, y, \lambda) > 0$ and is thus not bilevel feasible. The latter case then motivates to proceed with a larger penalization of the strong duality term.

By applying the same arguments and that the single-level reformulation Problem (4) is equivalent to the original bilevel problem (1), Theorem 2 implies the following main convergence result.

Theorem 4. *Suppose that $\rho^j \nearrow \infty$ holds and let $\{(x^j, y^j)\}_{j=0}^{\infty}$ be a sequence of stationary points of (9) (for $\rho = \rho^j$) generated by Algorithm 2 applied to Problem (4) with $(x^j, y^j) \rightarrow (x^*, y^*)$ for $j \rightarrow \infty$. Then, (x^*, y^*) is a stationary point of the strong duality error χ^{sd} . If, in addition, (x^*, y^*) is bilevel feasible, i.e., the strong duality error χ^{sd} is zero, then (x^*, y^*) is a stationary point of Problem (4).*

We close the discussion of applying an PADM to linear bilevel problems with three remarks.

Remark 1. *Note that Theorem 4 “only” makes a statement regarding stationary points of the single-level reformulation (4) and not about the original bilevel problem. In general, a stationary point of the Problem (4) does not need to be a stationary point of the bilevel problem, which is shown in [20] for the equivalent setting of a single-level reformulation based on KKT conditions of the lower level. Although we thus do not have any theoretical quality guarantee for the bilevel feasible points obtained by our method, we later show in our numerical results in Section 6 that, in practice, the quality of the obtained solutions is very good.*

Remark 2. *A crucial assumption of Theorem 3 (and that is also implicitly present in Theorem 4) is that one of the two PADM sub-problems always needs to have a unique solution. In the context of bilevel optimization this is strongly connected to the topic of unique lower-level solutions. It is well known that a bilevel problem can be very ill-behaved if its lower-level problem does not possess a unique solution for all possible decisions of the leader. Moreover, it is also folklore knowledge that the development of stable solution methods is much harder in the case of ambiguities in the lower level. Almost the same situation appears in the previous theorems despite the fact that dual uniqueness is required instead of primal uniqueness of the lower level if one considers the uniqueness of the second PADM sub-problem. Uniqueness of the first PADM sub-problem translates to unique solutions of the extended high-point relaxation (11). For a more thorough discussion of this topic we refer to Chapter 7 of the book [23].*

Remark 3. *The approach described in this section can also be applied to bilevel problems for which the upper level contains integrality constraints and a convex-quadratic objective function, i.e., problems of the form*

$$\min_{x,y} \frac{1}{2}x^\top H_u x + c^\top x + \frac{1}{2}y^\top G_u y + d^\top y \quad (13a)$$

$$\text{s.t. } Ax + By \geq a, \quad (13b)$$

$$x_i \in \mathcal{Z} \subset \mathbb{Z} \quad \text{for all } i \in I \subseteq \{1, \dots, n\}, \quad (13c)$$

$$y \in \arg \max_{\bar{y}} \{e^\top \bar{y} : Cx + D\bar{y} \leq b\}, \quad (13d)$$

with symmetric and positive semidefinite matrices $H_u \in \mathbb{R}^{n \times n}$ and $G_u \in \mathbb{R}^{m \times m}$. This does not affect the second PADM sub-problem at all. However, the first PADM sub-problem (11) is a convex-quadratic problem (QP) for $I = \emptyset$ and a mixed-integer convex-quadratic problem (MIQP) for $I \neq \emptyset$. Solving (MI)QPs to global optimality in every iteration may have significant impact on the performance of the PADM. This is analyzed in Section 6.

Up to now, we only considered bilevel problems with a linear follower problem. It is quite natural to extend the developed approach to bilevel problems with quadratic follower problems. This is addressed in the following section.

5. EXTENSION TO STRICTLY CONCAVE QUADRATIC FOLLOWER PROBLEMS

Consider a bilevel problem with a quadratic upper level and a strictly concave quadratic lower level:

$$\min_{x,y} \quad q_u(x,y) = \frac{1}{2}x^\top H_u x + c^\top x + \frac{1}{2}y^\top G_u y + d^\top y \quad (14a)$$

$$\text{s.t.} \quad Ax + By \geq a, \quad (14b)$$

$$y \in \arg \max_{\bar{y}} \left\{ q_l(\bar{y}) = \frac{1}{2}\bar{y}^\top G_l \bar{y} + e^\top \bar{y}; Cx + D\bar{y} \leq b \right\}. \quad (14c)$$

According to Remark 3, $H_u \in \mathbb{R}^{n \times n}$ and $G_u \in \mathbb{R}^{m \times m}$ are symmetric and positive semidefinite matrices. Further, $G_l \in \mathbb{R}^{m \times m}$ is a symmetric and negative definite matrix, which is required in the derivation of the dual lower-level problem; see the appendix. In addition, it ensures that the follower's problem has a unique solution.

In this setup, $q_u(\cdot, \cdot)$ is convex and $q_l(\cdot)$ is strictly concave. As a result, the upper-level problem is a convex QP and, for fixed upper-level variables, the lower level is a strictly concave QP, i.e., it is a parametric concave quadratic problem. Using the same ideas as before leads to the single-level reformulation

$$\min_{x,y,\lambda,\mu} \quad q_u(x,y) = \frac{1}{2}x^\top H_u x + c^\top x + \frac{1}{2}y^\top G_u y + d^\top y \quad (15a)$$

$$\text{s.t.} \quad Ax + By \geq a, \quad (15b)$$

$$Cx + Dy \leq b, \quad (15c)$$

$$G_l \mu + D^\top \lambda = e, \quad (15d)$$

$$\lambda \geq 0, \quad (15e)$$

$$\frac{1}{2}y^\top G_l y + \frac{1}{2}\mu^\top G_l \mu + e^\top y - b^\top \lambda + x^\top C^\top \lambda \geq 0. \quad (15f)$$

See the appendix for a detailed derivation of this reformulation. Relaxing the strong duality inequality (15f) and applying weak duality exactly in the same way as in the linear case in Section 4 yields the following smoothed convex-quadratic penalty reformulation:

$$\min_{x,y,\lambda,\mu} \quad \frac{1}{2}x^\top H_u x + c^\top x + \frac{1}{2}y^\top G_u y + d^\top y \quad (16a)$$

$$- \rho \left(\frac{1}{2}y^\top G_l y + \frac{1}{2}\mu^\top G_l \mu + e^\top y - b^\top \lambda + x^\top C^\top \lambda \right) \quad (16b)$$

$$\text{s.t.} \quad Ax + By \geq a, \quad (16c)$$

$$Cx + Dy \leq b, \quad (16d)$$

$$G_l \mu + D^\top \lambda = e, \quad (16e)$$

$$\lambda \geq 0. \quad (16f)$$

This results in the two PADM sub-problems

$$\min_{x,y} \frac{1}{2}x^\top H_u x + c^\top x + \frac{1}{2}y^\top G_u y + d^\top y \quad (17a)$$

$$- \rho \left(\frac{1}{2}y^\top G_l y + (C^\top \bar{\lambda})^\top x + e^\top y \right) \quad (17b)$$

$$\text{s.t. } Ax + By \geq a, \quad (17c)$$

$$Cx + Dy \leq b, \quad (17d)$$

and

$$\min_{\lambda,\mu} -\frac{1}{2}\mu^\top G_l \mu - (C\bar{x} - b)^\top \lambda \quad (18a)$$

$$\text{s.t. } G_l \mu + D^\top \lambda = e, \quad (18b)$$

$$\lambda \geq 0. \quad (18c)$$

Still, the first PADM sub-problem (17) has the same feasible region as the high point relaxation and the second PADM sub-problem (18) is exactly the dual lower-level problem. Compared to the case of linear follower problems, two differences can be observed:

- (i) The penalty term in the objective function (17b) is now convex-quadratic.
- (ii) The second PADM sub-problem is a convex QP (compared to an LP in the linear case) with additional variables μ . The latter results in a weaker coupling of the two PADM sub-problems.

Since the objective function of the first PADM sub-problem is already quadratic, (i) is not expected to have a strong impact on the performance of the entire method. However, it is not clear a-priorily, how the structural change described in (ii) affects the practical performance of the PADM. This is discussed in more detail in Section 6. We close this section with some final remarks.

Remark 4. *Looking back on the last sections, we see that the described approach makes use of two main structural properties of the tackled bilevel problems:*

- (i) *The possibility of replacing the follower's problem by its primal and dual constraint set and an additional inequality, which is based on strong duality and that ensures lower-level optimality. Moreover, the reformulation needs to be decomposable w.r.t. the respective variable blocks.*
- (ii) *The ability of solving the PADM sub-problems to global optimality.*

The first condition holds for convex lower-level problems. The second prerequisite is related to what is considered to be solvable to global optimality in practice. From the point of view of complexity theory, convex problems can be solved efficiently, i.e., in polynomial time. However, in practice, many mixed-integer linear (or even quadratic) problems can also be solved rather effectively. Going further, even the class of MINLP-QP bilevel problems can be tackled with the approach presented in this paper if the arising MINLPs can be solved effectively in practice.

6. NUMERICAL RESULTS

In this section, we present the numerical results obtained with the algorithm developed in the preceding sections. In Section 6.1, we first describe our computational setup, discuss some implementation details, and present the instance sets. Section 6.2 contains the numerical results for bilevel problems with a linear lower-level problem. First, we present results for LP-LP bilevel problems and we afterward also analyze the proposed PADM applied to QP-LP, MILP-LP, and MIQP-LP bilevel problems. In Section 6.3, we present numerical results for the extension to quadratic follower problems as discussed in Section 5.

The PADM of Algorithm 2 can be seen as a heuristic for bilevel optimization problems. Thus, its goal is to compute bilevel feasible points of good quality as fast as possible. In order to compare the performance of the PADM with a standard approach for solving bilevel problems with a convex lower level, we also implemented the classical KKT reformulation of the bilevel problems. Here, the resulting KKT complementarity constraints are interpreted as disjunctive constraints and are reformulated using binary variables; see [33] or [34]. Given the bilevel problem (14), the KKT reformulation leads to the MIQP

$$\min_{x,y,\lambda,z} \frac{1}{2}x^\top H_u x + c^\top x + \frac{1}{2}y^\top G_u y + d^\top y \quad (19a)$$

$$\text{s.t. } Ax + By \geq a, \quad Cx + Dy \leq b, \quad (19b)$$

$$-G_u y + D^\top \lambda = e, \quad \lambda \geq 0, \quad (19c)$$

$$z_i \in \{0,1\}, \quad \lambda_i \leq Mz_i, \quad (b - Cx - Dy)_i \leq (1 - z_i)M, \quad i = 1, \dots, \ell, \quad (19d)$$

which is equivalent to (14) for a sufficiently large constant M . Note that the latter mixed-integer constraint models KKT complementarity conditions. Further, additional integrality constraints in the upper level lead to a straightforward extension of the KKT reformulation (19). Solving this reformulation with a state-of-the-art MILP solver and terminating with the first bilevel feasible solution mimics a heuristic that we use as a benchmark for the running times. Of course, MILP solvers are not specifically designed to compute feasible points of good quality of a special bilevel reformulation quickly. However, state-of-the-art solvers like **Gurobi** are equipped with many powerful primal heuristics, which renders this approach also a feasible benchmark for both running times and solution quality. In the following, we label this benchmark approach with KKT-FF (where “FF” stands for “first feasible”). To further evaluate the overall quality of the feasible points computed with the PADM, we also consider a second benchmark by running the MILP solver applied to the KKT reformulation significantly longer. In the best case, this allows to compare the objective function values obtained by PADM with the optimal objective function values. If the optimal solutions cannot be computed within a reasonable time limit, this approach still delivers a best (lower) bound that can be used for comparison. We label this second benchmark approach with KKT-OPT.

6.1. Computational Setup, Implementation Details, and Test Sets. The PADM and all models have been implemented using C++11 and have been compiled using GCC 7.3.0. The (MI)QP and (MI)LP models are solved with **Gurobi** 8.0.1 using its C interface. All computational experiments have been executed on a compute cluster; see [62] for the details about installed hardware.

For our computations we parameterized the penalty ADM of Algorithm 2 as follows. For solving the first PADM sub-problem, we initialize the lower-level’s dual variable by the dual feasible point that we obtain by solving (12) with a zero objective function. The initial penalty parameter is 1 and the update is realized by doubling the parameter. We impose a total iteration limit of 100 and a time limit of 300 s. For numerical reasons, we also terminate the method if the penalty parameter’s upper bound of 10^{10} is achieved. Strong duality is checked with a relative tolerance of 10^{-4} , i.e., we terminate with a bilevel feasible point if $\chi^{\text{sd}}(x, y, \lambda)/|e^\top y| \leq 10^{-4}$ holds. The check for partial minima uses a tolerance of 10^{-6} . This means, we terminate the j th penalty iteration if $\|(x^{j,i}, y^{j,i}) - (x^{j,i-1}, y^{j,i-1})\|_\infty \leq 10^{-6}$ holds for some inner iteration i . All **Gurobi** tolerances and parameters are set to their default values.

The implementation details of KKT-FF are the following. As for the PADM, we set the time limit to 300 s. Our preliminary numerical experiments revealed that

Gurobi’s default settings may lead to integer-feasible points (within the standard tolerances of Gurobi) of the KKT reformulation that numerically violate the KKT complementarity constraints. Thus, we tightened Gurobi’s integer feasibility tolerance to 10^{-9} . The default value of the parameter `IntFeasTol` is 10^{-5} . To further improve the numerical stability our preliminary numerical experiments also revealed that a `NumericFocus` parameter value of 3 is favorable, which results in increased numerical accuracy. In addition, we implemented a Gurobi callback that checks for every integer-feasible solution found if it indeed satisfies strong duality. As for the PADM, we only terminate, if $\chi^{\text{sd}}(x, y, \lambda)/|e^\top y| \leq 10^{-4}$. Note that we left the Gurobi parameter `MIPFocus` at its default value 0 (automatic). Our numerical experiments revealed that a setting that favors finding feasible points quickly (`MIPFocus` = 1), which is the main focus here, does not improve the running times of the KKT approach. One crucial issue for the KKT approach is the choice of the big- M constant used in the binary reformulation of the KKT complementarity constraints. Although there exist some strategies for choosing reasonable values, in general, this choice is heuristically. See the recent papers [48, 58, 59] for a thorough discussion of this issue. Here, we choose a value of 10^8 . Let us emphasize that, in contrast to the KKT reformulation, the PADM approach does not involve any big- M and is thus more reliable and stable from a numerical point of view.

For KKT-OPT, we set the time limit to 900s and return the best solution that satisfies the bilevel feasibility criterion $\chi^{\text{sd}}(x, y, \lambda)/|e^\top y| \leq 10^{-4}$ found within this limit. All other parameters are set to the same values as for KKT-FF.

Our test set is based on various MILP-MILP bilevel instance sets from the literature and on new MILP-MILP bilevel instances created from the MIPLIB2010 [50] and the MIPLIB2017 [55]. Every instance from the latter two libraries has been converted to three mixed-integer bilevel problems by considering the first 50% of the variables to be upper-level variables and the remaining variables to be lower-level variables, and by splitting the constraints in the following ways:

- (i) All constraints are considered to belong to the upper level.
- (ii) The first 50% of the constraints are considered to be upper-level constraints, the other 50% of the constraints are considered to be lower-level constraints.
- (iii) All constraints are considered to be lower-level constraints.

In all three cases, the lower-level objective function is the same as the upper-level objective function but restricted to the lower-level variables and with flipped optimization direction.

In Table 1 we list the details of the instance sets that we used. We specify—if available—the reference of the corresponding publication (“Ref.”), a reference to the underlying single-level test set (“Source”), and a reference to a website on which the instances can be downloaded (“Web”). In addition, we specify the number of instances per test set (“Size”) and the structure of the constraints (“Constr. Structure”). By 0/100, we indicate that at most one constraint belongs to the upper level. The “at most one” stems from interdiction instances for which all constraints belong to one level and an additional constraint models the interdiction budget of the other level; see, e.g., [13, 30, 64] for more details on interdiction problems. Similarly, 100/0 means that at most one constraint belongs to the lower level and 50/50 specifies that 50% of the constraints belong to each level. Consequently, “all” in Table 1 means that all of the three types of constraint structures appear in the respective instance set. We point out that we do not consider simple variable bounds as constraints in this context. The instance set OR does not fit to any of the mentioned patterns and is thus labeled with “—”. For each instance of the OR test set exist three variants. All three variants have a fixed number of constraints that are always assigned to the lower level. In addition there is a (smaller) number of

TABLE 1. The complete test set.

Instance Set	Ref.	Source	Web	Size	Constr. Structure
CLIQUE	[32]	—	[29]	60	0/100
GENERALIZED	[32]	—	[29]	90	50/50
GK	—	—	[29]	33	all
IMKP	[30]	[63]	[29]	144	all
INTOSUM	—	—	[61]	8	50/50
INTER-CLIQUE	[70]	—	[29]	80	0/100
INTER-FIRE	[4]	—	[29]	72	0/100
KP	[32]	—	[29]	450	0/100
MIPLIB	[31]	[9]	[29]	60	0/100
MIPLIB2010	—	[50]	—	261	all
MIPLIB2017	—	[55]	—	654	all
OR	—	[17]	[29]	810	—
XUWANG	[75]	—	[29, 76]	101	50/50
XUWANG-LARGE	[28]	[75]	[29]	60	50/50

constraints that is distributed in various ways to the upper and lower level, namely, (i) all but one constraint are assigned to the upper level, (ii) 50 % each are assigned to the upper and lower level, and (iii) all but one constraint are assigned to the lower level.

In total, our test set contains 2883 instances, which is—at least to the best of our knowledge—the largest test set of bilevel problems considered in the literature. To obtain suitable LP-LP or MILP-LP instances, we relax the integrality constraints in both levels or only in the lower level, respectively. Furthermore, we randomly generated symmetric positive semidefinite matrices H_u and G_u as well as symmetric and negative definite matrices G_l to extend (MI)LP-LP instances to (MI)QP-LP and (MI)QP-QP instances.² For each instance, we randomly generated quadratic matrices Q , R , and S of suitable sizes with integer entries in $[-\sqrt{\sigma}, \sqrt{\sigma}]$ with $\sigma = \max\{\|c\|_\infty, \|d\|_\infty\}$. We then computed $H_u = Q^\top Q$, $G_u = R^\top R$, and $G_l = -(S^\top S + D)$, where D is a diagonal matrix with entries in $[1, \sqrt{\sigma}]$ to ensure that G_l is negative definite. We display the sizes and densities of the generated matrices in Figure 1.

6.2. Bilevel Problems with a Linear Follower Problem. We first evaluate the PADM for LP-LP bilevel problems in terms of running times and solution quality by comparing the results with KKT-FF and KKT-OPT, respectively. For this analysis, we filter out some instances as summarized in Table 2. Out of the 2883 instances, 1705 instances are solved by both methods in less than 1 s (labeled as “Easy” in Table 2) and thus are left out. For 21 instances both methods could neither find a bilevel feasible point nor prove infeasibility. These instances are considered as too hard and are left out as well (“Hard”). In addition, 200 instances are infeasible (“Inf”). In fact, detecting infeasibility is a non-trivial task. For 80 instances, either the first or the second PADM sub-problem is infeasible. In the former case, the high point relaxation is infeasible. The latter case implies that the dual lower level is infeasible, which renders the primal lower-level problem infeasible or unbounded. In both cases we know for sure that the bilevel instance is infeasible. For the remaining 120 infeasible instances the KKT approaches terminated with infeasibility, while the

²The MATLAB function that we implemented to generate the matrices and a brief documentation thereof can be found in the GitHub repository under <https://github.com/m-schmidt-math-opt/qp-bilevel-matrix-generator>.

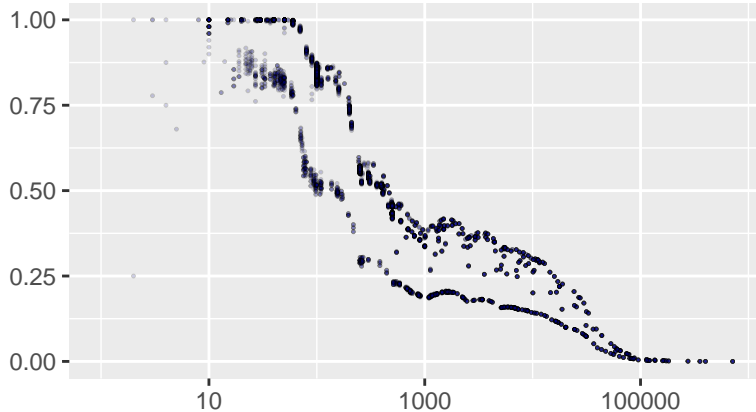


FIGURE 1. Sizes (n resp. m) and densities of the matrices $H_u \in \mathbb{R}^{n \times n}$, $G_u, G_l \in \mathbb{R}^{m \times m}$ for the (MI)QP-LP and (MI)QP-QP instances.

TABLE 2. Characteristic numbers of the different bilevel problem classes for the 2883 instances per class.

Class	General					KKT		PADM	
	Easy	Hard	Inf	FalseM	Rem	FInf	Feas	Unb2	Feas
LP-LP	1705	21	200	33	924	18	484	0	921
QP-LP	591	488	219	63	1522	119	1073	11	1478
MILP-LP	1178	127	205	17	1356	19	998	5	1112
MIQP-LP	248	768	220	12	1635	60	1364	5	672

PADM approach did not converge within the time and iteration limit. Although this is not a proof of infeasibility (potentially, the big- M of the KKT approach could be chosen too small), it is very likely that these instances are indeed infeasible and we thus exclude them from the test set. We also exclude 33 instances, for which the best bound provided by KKT-OPT is larger than the objective value of the bilevel feasible point provided by PADM. The reason for this behavior is a big- M that is too small (“FalseM”). The remaining 924 instances (“Rem”) are considered for the analysis of running times and solution qualities. For these instances, the case of an unbounded sub-problem (12) (“Unb2”) never occurred but we detected 18 instances for which the KKT approach wrongly reported infeasibility (“FInf”). We consider the latter instances as unsolved. Finally, column “Feas” denotes the number of instances for which the respective method found a bilevel feasible point.

Figure 2 (left) shows a log-scaled performance profile according to [26] comparing the running times of PADM and KKT-FF. Since KKT-OPT is totally dominated by KKT-FF by construction, we leave it out for the running times analysis. Obviously, the PADM significantly outperforms KKT-FF. Out of the 924 instances, 921 single-level reformulations can be solved to stationarity with the PADM and it is the faster method for almost every instance. KKT-FF is only able to find a bilevel feasible point for 484 instances, which is slightly more than 50% of the 924 instances, and for these instances it is almost never the faster approach. In order to further classify the running times, we provide additional information on mean and median running times; see Table 3. The means are 5.6s (PADM) vs. 142.9s (KKT-FF) and the

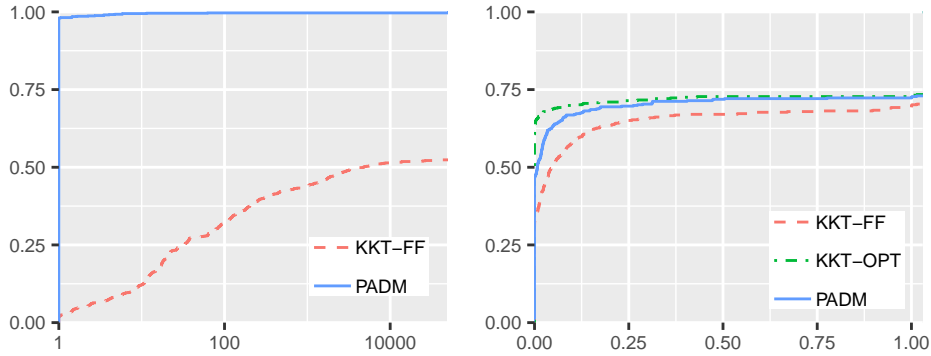


FIGURE 2. Log-scaled performance profile for the running times (left) and ECDFs for the relative gaps (right) of PADM (solid), KKT-FF (dashed) and KKT-OPT (dot-dashed) applied to the LP-LP bilevel instances.

TABLE 3. Running times of PADM and KKT-FF for all analyzed classes of bilevel problems with a linear lower level.

Class	LP-LP		QP-LP		MILP-LP		MIQP-LP	
	Mean	Med	Mean	Med	Mean	Med	Mean	Med
PADM	5.6	0.1	31.4	5.8	73.7	6.5	198.3	300.0
KKT-FF	142.9	50.3	82.9	11.5	80.7	3.7	56.1	5.3
KKT-OPT	570.9	900.0	577.7	900.0	626.6	900.0	716.0	900.0

medians are 0.1 s (PADM) vs. 50.3 s (KKT-FF). These numbers underline the clear dominance of KKT-FF by the PADM. We explicitly point out that the running times in Table 3 are not comparable across different instance classes. The reason is that each class uses a different instance subset; see Table 2.

Besides running times we are also interested in the quality of the bilevel feasible points computed by the PADM. We therefore compare relative objective function value gaps for the three methods. More precisely, we compare relative gaps $(q_u^* - q_u^{\text{LB}})/|q_u^{\text{LB}}|$ for $q_u^* \in \{q_u^{\text{PADM}}, q_u^{\text{KKT-FF}}, q_u^{\text{KKT-OPT}}\}$. The latter denote the objective function values of PADM, KKT-FF, and KKT-OPT, respectively, and q_u^{LB} denotes the best lower bound found by KKT-OPT. Of course, comparing to the optimal objective function value would be much more desirable, but due to the general hardness of the tested instances, in many cases proving bilevel optimality is out of reach. In case of $q_u^{\text{LB}} = 0$, the relative gap is set to ∞ . Note that we already measure the overall robustness of the PADM and KKT-FF in the running times analysis. Thus, for the solution quality analysis, we only consider the 481 instances, for which all methods provide a bilevel feasible point. Figure 2 (right) shows the empirical cumulative density functions (ECDFs) of the relative gaps. The plot shows the most relevant part between 0 and 1, i.e., gaps up to 100%. The figure reveals that the PADM heuristic also outperforms KKT-FF in terms of the solution quality. As mentioned before, MILP solvers are not specifically designed to find feasible solutions of good quality of special bilevel reformulations. Nevertheless, these solvers use many sophisticated primal heuristics, such that the clear dominance of KKT-FF by PADM was not to be expected. Figure 2 (right) also reveals that PADM is quite close to KKT-OPT. This means that PADM not only delivers bilevel feasible solutions very quickly. The solutions are also almost as good as the solutions of

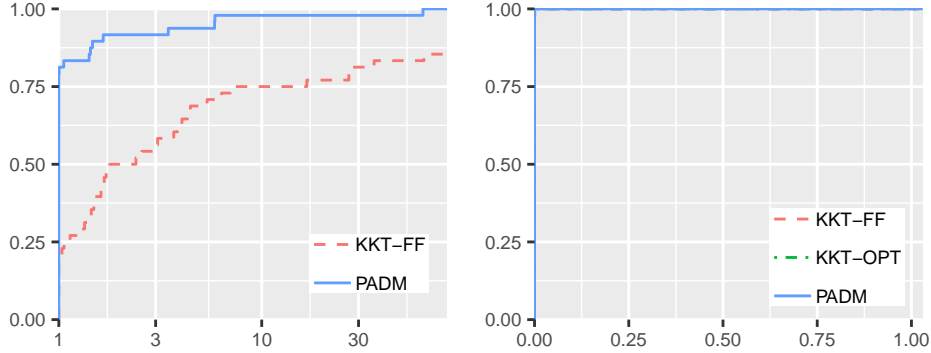


FIGURE 3. Log-scaled performance profiles for the running times (left) and ECDFs for the relative gaps (right) of PADM (solid), KKT-FF (dashed), and KKT-OPT (dot-dashed) for instances of MIPLIB2017 with structure 100/0.

KKT-OPT that runs significantly longer; see the running times (e.g., the means 5.6 vs. 570.9) in Table 3. For many instances, the solutions of the PADM are even globally optimal. In almost 50% of the instances considered for the solution quality analysis, the gap of the bilevel feasible point found by PADM is 0%, i.e., the PADM found the optimal solution. Note that, since we compare with best bounds and not with optimal solution values, the actual number of optimal solutions found by PADM might be even higher.

It is clear that the constraint structure has a strong impact on the performance of the KKT approach: Every lower-level constraint yields a KKT complementarity constraint that is linearized via a binary variable. Thus, the structure 0/100 is the hardest and the structure 100/0 is the easiest for the KKT approach. Figure 3 shows running times and the solution quality for the instance set MIPLIB2017 with structure 100/0. For this subset of 48 remaining instances, KKT-FF is the faster method for around 20% of the instances. Still, the PADM totally dominates KKT-FF in terms of running times. Surprisingly, both approaches always find the optimal solution (over those 41 instances for which all methods found a bilevel feasible point); cf. Figure 3 (right).

We now analyze the limits of the PADM by increasing the complexity of the bilevel problems step by step; see Remark 3. We start by taking a look at QP-LP bilevel problems. Some general numbers on the instances of the QP-LP test set are given in Table 2. In Figure 4 (left), we display the performance profile for the running times of the remaining 1522 QP-LP instances (after filtering according to Table 2). Again, the PADM totally dominates KKT-FF in terms of running times, although not as clear as for the LP-LP instances. This is supported by mean and median running times; see Table 3. Furthermore, Figure 4 (right) shows that the solution quality of the PADM solutions is comparable to the solutions provided by KKT-OPT for those 1029 instances for which all approaches computed a bilevel feasible point. Again, PADM is significantly faster than KKT-OPT; see Table 3. At this point, we explicitly mention that characteristics of the running time and solution quality plots (like percentage of solved instances, etc.) are not comparable across problem classes, since the aggregated instance sets differ; see Table 2. The behavior in Figure 4 can be expected since not much changes in the general structure of the two solution concepts. The KKT reformulation (19) becomes an MIQP instead of an MILP for quadratic leader problems. However, the feasible region of the KKT reformulation is exactly the same for LP-LP and QP-LP instances. Since the

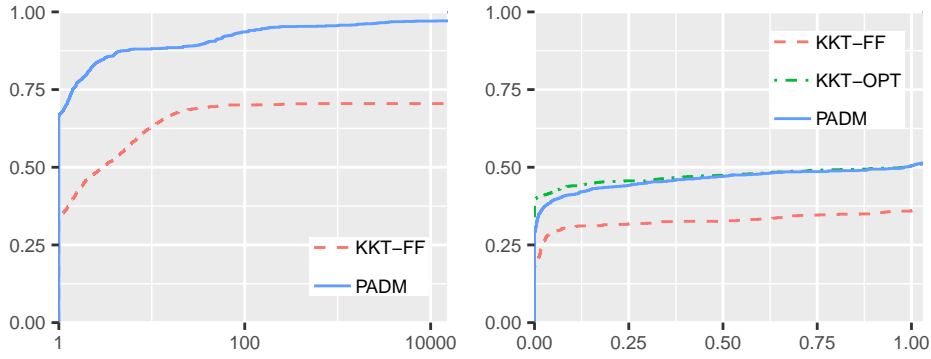


FIGURE 4. Log-scaled performance profile for the running times (left) and ECDFs for the relative gaps (right) of PADM (solid), KKT-FF (dashed), and KKT-OPT (dot-dashed) applied to the QP-LP bilevel instances.

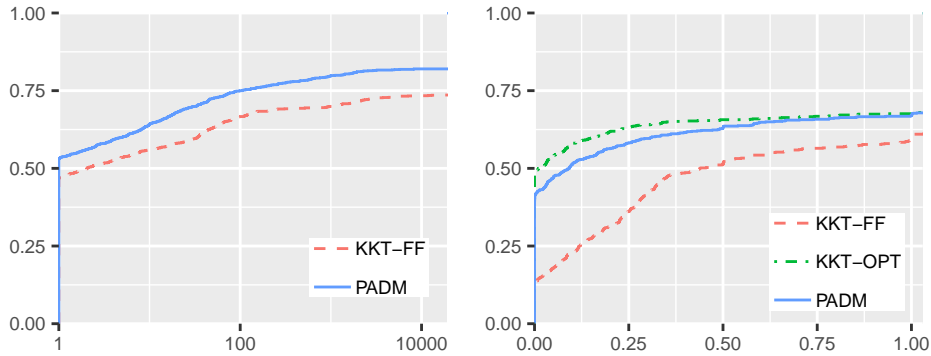


FIGURE 5. Log-scaled performance profile for the running times (left) and ECDFs for the relative gaps (right) of PADM (solid), KKT-FF (dashed), and KKT-OPT (dashed), applied to the MILP-LP bilevel instances.

KKT-based heuristic “only” finds a feasible point of the reformulated MI(Q)P, it does not make a strong difference whether the leader problem is an LP or QP. For the PADM, the second sub-problem is the same as for LP-LP bilevel problems but the first sub-problem becomes a QP instead of an LP that needs to be solved to optimality.

The situation changes when introducing integers to the leader problem. For MILP-LP problems, the KKT approach yields an MILP and thus remains in the same problem class compared to LP-LP bilevel problems, because binary variables are already present in the KKT reformulation to linearize the KKT complementarity constraints. In particular, the KKT approach can still profit from the many MILP heuristics implemented in Gurobi to find a feasible point quickly. In contrast, the first PADM sub-problem now is an MILP instead of an LP. Thus, solving this sub-problem to optimality in every iteration becomes more expensive and the question arises whether the clear dominance of the PADM as seen for LP-LP or QP-LP instances can still be observed for MILP-LP instances. Figure 5 compares running times and the solution quality of the two approaches for our MILP-LP test set. As before, details on the instances can be found in Table 2. The PADM still performs

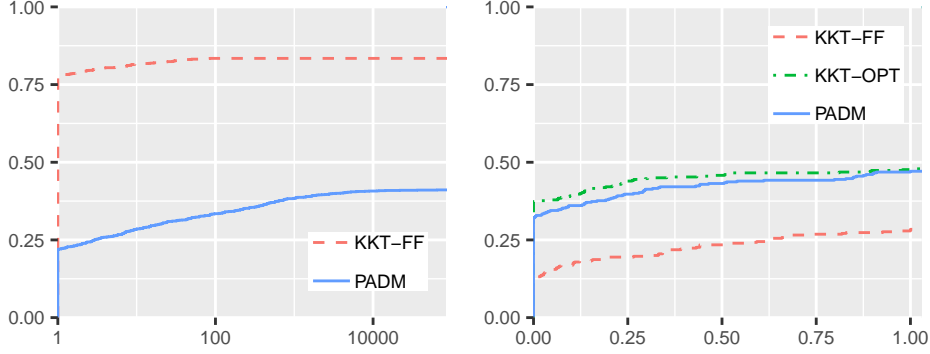


FIGURE 6. Log-scaled performance profile for the running times (left) and ECDFs for the relative gaps (right) of PADM (solid), KKT-FF (dashed), and KKT-OPT (dot-dashed) applied to the MIQP-LP bilevel instances.

better than KKT-FF, although both approaches are the faster method for around 50% of the 1356 remaining MILP-LP instances, but the mean and median runtimes are now more in favor of KKT-FF; see Table 3. However, PADM is still the more robust approach, i.e., it solves more instances to bilevel feasibility compared to KKT-FF (1112 vs. 998). As before, the solution quality provided by PADM is almost comparable to KKT-OPT (for the 754 instances for which all approaches found a bilevel feasible point), while PADM being clearly the faster method; see Table 3.

Finally, we analyze MIQP-LP instances, see again Table 2 for some numbers on the instances. Following the discussion for QP-LP and MILP-LP instances, it cannot be expected, that the PADM is still the faster method. Every sub-problem (12) is now an MIQP, while the feasible region of the KKT reformulation does not change compared to the MILP-LP case. The results for the MIQP-LP instances are shown in Figure 6. It turns out that KKT-FF clearly dominates the PADM in terms of running times. This is supported by the mean and median running times in Table 3. It is the more robust approach and the faster method for around 75% of the 1635 instances. Nevertheless, out of the 401 instances, for which all methods provide a bilevel feasible point, the quality of the solutions of PADM is significantly better than the solutions of KKT-FF and comparable to the solutions obtained by KKT-OPT (that runs noticeably longer; see the runtimes in Table 3).

In summary, the PADM by far outperforms the KKT reformulation for LP-LP and QP-LP bilevel problems. It is the faster method and finds bilevel feasible points for more instances. In addition, the quality of the bilevel feasible solutions provided by PADM is comparable to solutions obtained by running the KKT reformulation significantly longer. Surprisingly, also for MILP-LP problems, the PADM is the better method. As expected, for MIQP-LP bilevel problems, the KKT approach outperforms the PADM in terms of running times (but not in terms of the solution quality). However, we emphasize that we did not apply any enhancements to the PADM, whereas the KKT approach benefits from all the batteries included in Gurobi. The PADM allows for many possible tweaks that can greatly improve the running times. These are discussed in more detail in the following subsection as well as in the conclusion; see Section 7.

6.3. Bilevel Problems with a Quadratic Follower Problem. In this section, we evaluate the PADM for (MI)QP-QP bilevel problems as described in Section 5.

TABLE 4. Characteristic numbers of the different bilevel problem classes for the 2883 instances per class with quadratic lower level.

Class	General					KKT		PADM	
	Easy	Hard	Inf	FalseM	Rem	FInf	Feas	Unb2	Feas
QP-QP	73	745	130	20	1915	95	419	0	1885
MIQP-QP	49	1742	102	20	970	94	169	0	894

We already pointed out that, in this case, the second PADM sub-problem has a quadratic objective function and an additional set of variables μ . Since QPs can be solved with comparable efficiency as LPs, the quadratic objective function is not expected to lead to a significant decrease in the performance of the method. However, the additional variables may harm the good convergence properties of the PADM that we observe in the linear case. This is supported by our preliminary numerical experiments. It is straightforward to see that for a primal-dual solution (y^*, λ^*, μ^*) of Problem (18), $\mu^* = -y^*$ holds; see the derivation of the dual in the appendix. However, μ is not present in the first PADM sub-problem (17) and y is not present in the second sub-problem (18). Thus, the only coupling between the two sub-problems is—exactly like in the case of linear follower problems—between x and λ . Due to this weak coupling, the convergence to $\mu = -y$ can be expected to be rather slow. Unfortunately, in our opinion, this situation cannot be resolved. Simply substituting $\mu = -y$ in Problem (18) couples primal and dual variables and thus destroys the block-separability. Also other methods like penalizing $|\mu + y|$ in the objective function of the single-level reformulation (15) showed bad convergence properties in our preliminary numerical results.

One way to speed up convergence is to make a compromise between running times and the quality of the solutions. This can be done, e.g., by relaxing the tolerance for accepting partial minima or by limiting the number of inner ADM iterations. After some numerical experiments, we equipped the PADM for the (MI)QP-QP instances with the following tweaks:

- (i) We set the initial penalty parameter to 4.
- (ii) We limit the number of inner ADM iterations to 10.
- (iii) We relax the tolerance for the check for partial minima from 10^{-6} to 10^{-4} .
- (iv) We implemented the following additional termination criterion: Whenever the PADM finds a bilevel feasible point and the objective function value did not improve by more than $p\%$ compared to the last iteration, then the PADM terminates. For our computations we use $p = 0.01$.

All other settings are the same as described in Section 6.1. The aspects (i) and (ii) cause the penalty parameter to increase faster and thus speed up the convergence to a bilevel feasible point. Further, (iii) and (iv) may result in accepting solutions that are actually not stationary. It can thus be expected that the solution quality may not be as good as for linear lower levels. However, let us explicitly point out that we still always terminate with a bilevel feasible point.

We now present results for the QP-QP instances; see Table 4 for details on the instances. Figure 7 (left) shows the performance profile of the running times of the 1915 remaining QP-QP instances. It is obvious that the PADM approach outperforms the KKT approach in every aspect. It is the faster method for more than 90% of the instances and it finds a bilevel feasible point for almost all of them. KKT-FF is only able to compute a bilevel feasible point for a little more than 20% of the instances. The dominance is also underlined by the mean and median running times displayed in Table 5.

TABLE 5. Running times for the two analyzed classes of bilevel problems with quadratic lower level.

Class	QP-QP		MIQP-QP	
	Mean	Med	Mean	Med
PADM	41.2	18.2	49.8	10.4
KKT-FF	224.8	300.0	207.9	300.0
KKT-OPT	788.0	900.0	679.7	900.0

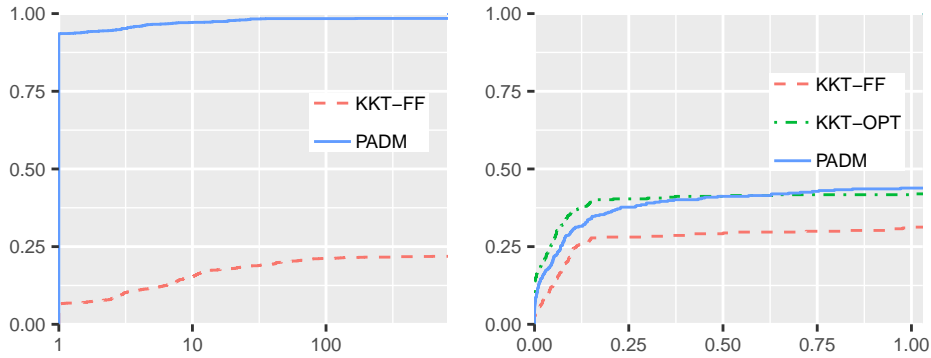


FIGURE 7. Log-scaled performance profile for the running times (left) and ECDFs for the relative gaps (right) of PADM (solid), KKT-FF (dashed), and KKT-OPT (dot-dashed) applied to the QP-QP bilevel instances.

Like before, Figure 7 (right) displays an ECDF plot of the relative gaps of the 389 instances, for which all three methods find a bilevel feasible point. Note that we cannot draw strong conclusions regarding the absolute quality of the solution due to the general hardness of the QP-QP instances. The bilevel optimal solution is very often not known and the best bounds obtained by KKT-OPT are rather weak. Still, Figure 7 (right) allows a comparison of the solution quality of PADM, KKT-FF, and KKT-OPT. Although we not necessarily terminate the PADM with a stationary point of the single-level reformulation but “only” with a bilevel feasible point, it still outperforms KKT-FF and is competitive w.r.t. KKT-OPT.

Finally, we turn to the MIQP-QP instances; see Table 4 for details. Figure 8 reveals that, also for this problem class, the PADM outperforms the KKT approach. The PADM finds a bilevel feasible point for much more instances (894 vs. 169) in significantly less time; see the running times in Table 5. Taking into account the results of the MIQP-LP instances, the dominance of the PADM is a bit surprising. The reasons might be two-fold. On the one hand, the KKT reformulation (19) has to deal with the extended constraint $-G_l y + D^\top \lambda = e$. On the other hand, we strongly trimmed the PADM approach for better running times. Still, the solution quality provided by the PADM is better compared to KKT-FF and again quite competitive w.r.t. KKT-OPT. Overall, for our bilevel instances with a quadratic lower level, the PADM can be considered the best approach among the three tested ones.

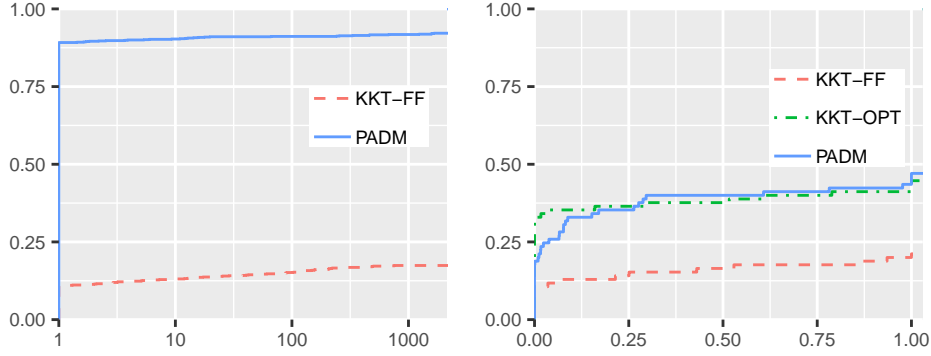


FIGURE 8. Log-scaled performance profile for the running times (left) and ECDFs for the relative gaps (right) of PADM (solid), KKT-FF (dashed), and KKT-OPT (dot-dashed) applied to the MIQP-QP bilevel instances.

7. CONCLUSION

In this paper, we applied a penalty alternating direction method to various bilevel problems that can be reformulated as a single-level problem via strong duality of convex programming. For these kinds of bilevel problems, we showed that the proposed method converges to stationary points of the classical strong-duality based single-level reformulation. We further evaluated the performance of the PADM in terms of running times and solution quality in a very extensive numerical study. The running times of the PADM rely on solving the two sub-problems effectively, which makes it especially suitable for LP-LP bilevel problems. For this problem class, we showed that the described method significantly outperforms the benchmark—a classical KKT reformulation of the lower level. We obtained similar results for QP-LP bilevel problems. Even for MILP-LP bilevel instances, our proposed method delivers the best results. Only for MIQP-LP instances, a problem class for which the PADM sub-problems become very expensive in terms of running times, the plain PADM (i.e., without any enhancements) performed worse than the KKT approach.

We extended the PADM to the class of (MI)QP-QP bilevel problems and introduced several enhancements of the method. Essentially, they compromise on the solution quality in favor of better running times. It turned out that the PADM equipped with these enhancements clearly outperforms the benchmark approach in terms of running times and solution quality for both QP-QP and MIQP-QP instances. Based on these very good numerical results, the PADM can serve as both a primal heuristic for global solution methods as well as a standalone method.

A careful analysis of techniques to speed up the convergence of the proposed method is subject to future research and may include to (i) try to check for bilevel feasibility in every iteration (disregarding partial minima) as well as (ii) analyze different initial penalty parameters and penalty parameter update rules like updating the penalty parameter after every iteration. Finally, another topic for future research is to apply the proposed PADM with suitable enhancements to more complex bilevel optimization problems like MINLP-QP problems and to integrate it into state-of-the-art exact solution methods for bilevel problems.

ACKNOWLEDGMENTS

This research has been performed as part of the Energie Campus Nürnberg and is supported by funding of the Bavarian State Government. We also thank the

Deutsche Forschungsgemeinschaft for their support within project A05 and B08 in the ‘‘Sonderforschungsbereich/Transregio 154 Mathematical Modelling, Simulation and Optimization using the Example of Gas Networks’’. Further, we thank Markus Sinnl for his help in organizing and sorting the bilevel instances of our numerical study, Lars Schewe for many valuable discussions, and Andrea Lodi for a fruitful discussion on the numerical results. Finally, we are very grateful to Stephan Dempe for giving us the important hint regarding the context of Remark 1.

APPENDIX. DERIVATION OF THE SINGLE-LEVEL REFORMULATION OF THE MIQP-QP BILEVEL PROBLEM

In this section, we derive the single-level reformulation (15) of the QP-QP bilevel problem (14). To this end, we need some more notation that is taken from [5, 56]. The bilevel constraint region is denoted by

$$P := \{(x, y) : Ax + By \geq a, Cx + Dy \leq b\}.$$

This is the feasible set of the high-point relaxation. Its projection on the decision space of the upper level is denoted by

$$P_u := \{x : \exists y \text{ such that } (x, y) \in P\}.$$

Moreover, for fixed $x \in P_u$, the lower-level feasible region is given by

$$P_l(x) := \{y : Dy \leq b - Cx\}.$$

Like in Section 2, we consider the follower’s problem as parametric optimization problem that now reads

$$\max_y q_l(y) \quad \text{s.t.} \quad Dy \leq b - C\bar{x}, \quad (20)$$

where the leader’s decision \bar{x} parameterizes the problem. The dual problem of Problem (20) is given by

$$\min_{\lambda \geq 0} g(\bar{x}, \lambda), \quad (21)$$

with $g(\bar{x}, \lambda) = \sup_y \mathcal{L}(\bar{x}, y, \lambda)$; see [11]. In our setup, the Lagrangian \mathcal{L} is given by

$$\mathcal{L}(\bar{x}, y, \lambda) = \frac{1}{2}y^\top G_l y + e^\top y - \lambda^\top (C\bar{x} + Dy - b).$$

Since $\mathcal{L}(\bar{x}, y, \lambda)$ is strictly concave and differentiable in y , its supremum is given by

$$\nabla_y \mathcal{L}(\bar{x}, y, \lambda) = G_l y + e - D^\top \lambda = 0 \iff y = -G_l^{-1}(e - D^\top \lambda).$$

The latter expression is valid, since G_l is negative definite and hence regular. We can now substitute this expression in the Lagrangian and obtain the dual objective function

$$g(\bar{x}, \lambda) = -\frac{1}{2}(e - D^\top \lambda)^\top G_l^{-1}(e - D^\top \lambda) - (C\bar{x} - b)^\top \lambda. \quad (22)$$

Since G_l is negative definite, the same holds for its inverse G_l^{-1} and $g(\bar{x}, \cdot)$ is a convex-quadratic function. Thus, the dual problem (21) is a parametric convex-quadratic problem in λ . By introducing additional dual variables $\mu = G_l^{-1}(e - D^\top \lambda)$, we obtain the following representation of (21):

$$\min_{\lambda, \mu} g(x, \lambda, \mu) = -\frac{1}{2}\mu^\top G_l \mu - (Cx - b)^\top \lambda \quad (23a)$$

$$\text{s.t.} \quad G_l \mu + D^\top \lambda = e, \quad \lambda \geq 0. \quad (23b)$$

This formulation may be favorable from a numerical point of view, since no inverse is involved (cf. (22))—and it is easier to read. Hence, we use this representation.

In the following, we only consider $x \in P_u$, i.e., upper-level variables for which the parametric lower-level problem is feasible. The parametric lower-level problem is a concave maximization problem over affine-linear constraints. Consequently, duality conditions apply; see [11, Chap. 5.2.3] and [71, Theorem 24.1]. For every primal-dual feasible point (y, λ, μ) , i.e., $y \in P_l(x)$ and $\lambda \geq 0, \mu$ fulfill (23b), weak duality

$$q_l(y) \leq g(x, \lambda, \mu) \quad (24)$$

holds. Furthermore, for optimal primal-dual points (y^*, λ^*, μ^*) , strong duality holds, i.e., $q_l(y^*) = g(x, \lambda^*, \mu^*)$. Together with QP weak duality (24), strong duality can be ensured using the constraint $q_l(y) - g(x, \lambda, \mu) \geq 0$, i.e.,

$$\frac{1}{2}y^\top G_l y + e^\top y + \frac{1}{2}\mu^\top G_l \mu + (Cx - b)^\top \lambda \geq 0. \quad (25)$$

We remark that (25) is exactly the QP counterpart of (4f). Both inequalities only differ in the quadratic forms $\frac{1}{2}y^\top G_l y$ and $\frac{1}{2}\mu^\top G_l \mu$. The same holds true for the dual constraint (23b), which is the QP counterpart of (4d). Here, the only difference is the linear term $G_l \mu$.

Taking the above consideration into account, the bilevel problem (14) can be reformulated equivalently to the single-level problem

$$\begin{aligned} \min_{x, y, \lambda, \mu} \quad & q_u(x, y) = \frac{1}{2}x^\top H_u x + c^\top x + \frac{1}{2}y^\top G_u y + d^\top y \\ \text{s.t.} \quad & Ax + By \geq a, \\ & Cx + Dy \leq b, \\ & G_l \mu + D^\top \lambda = e, \\ & \lambda \geq 0, \\ & \frac{1}{2}y^\top G_l y + \frac{1}{2}\mu^\top G_l \mu + e^\top y - b^\top \lambda + x^\top C^\top \lambda \geq 0, \end{aligned}$$

which is (15).

REFERENCES

- [1] Mirjam Ambrosius, Veronika Grimm, Thomas Kleinert, Frauke Liers, Martin Schmidt, and Gregor Zöttl. *Endogenous Price Zones and Investment Incentives in Electricity Markets: An Application of Multilevel Optimization with Graph Partitioning*. Tech. rep. Oct. 2018. URL: http://www.optimization-online.org/DB_HTML/2018/10/6868.html.
- [2] G. Anandalingam and D.J. White. “A solution method for the linear static Stackelberg problem using penalty functions.” In: *IEEE Transactions on Automatic Control* 35.10 (1990), pp. 1170–1173. DOI: [10.1109/9.58565](https://doi.org/10.1109/9.58565).
- [3] Jaqueline S. Angelo and Helio J. C. Barbosa. “A study on the use of heuristics to solve a bilevel programming problem.” In: *International Transactions in Operational Research* 22.5 (2015), pp. 861–882. DOI: [10.1111/itor.12153](https://doi.org/10.1111/itor.12153).
- [4] Andrea Baggio, Margarida Carvalho, Andrea Lodi, and Andrea Tramontani. *Multilevel approaches for the critical node problem*. Tech. rep. Ecole Polytechnique de Montral, 2016.
- [5] Jonathan F. Bard. *Practical bilevel optimization: algorithms and applications*. Vol. 30. Springer Science & Business Media, 2013.
- [6] Jonathan F. Bard. “Some properties of the bilevel programming problem.” In: *Journal of optimization theory and applications* 68.2 (1991), pp. 371–378. DOI: [10.1007/BF00941574](https://doi.org/10.1007/BF00941574).

- [7] Jonathan F. Bard and James T. Moore. “A Branch and Bound Algorithm for the Bilevel Programming Problem.” In: *SIAM Journal on Scientific and Statistical Computing* 11.2 (1990), pp. 281–292. DOI: [10.1137/0911017](https://doi.org/10.1137/0911017).
- [8] Timo Berthold. “Heuristic algorithms in global MINLP solvers.” PhD thesis. Technische Universität Berlin, 2014.
- [9] Robert E. Bixby, Sebastian Ceria, Cassandra M. McZeal, and Martin W. P. Savelsbergh. *An updated mixed integer programming library: MIPLIB 3.0*. Tech. rep. 1998.
- [10] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. “Distributed optimization and statistical learning via the alternating direction method of multipliers.” In: *Foundations and Trends in Machine Learning* 3.1 (2011), pp. 1–122. DOI: [10.1561/22000000016](https://doi.org/10.1561/22000000016).
- [11] Stephen Boyd and Lieven Vandenbergh. *Convex Optimization*. Cambridge University Press, 2004. DOI: [10.1017/cbo9780511804441](https://doi.org/10.1017/cbo9780511804441).
- [12] Luce Brotcorne, Martine Labbé, Patrice Marcotte, and Gilles Savard. “A bilevel model for toll optimization on a multicommodity transportation network.” In: *Transportation Science* 35.4 (2001), pp. 345–358. DOI: [10.1287/trsc.35.4.345.10433](https://doi.org/10.1287/trsc.35.4.345.10433).
- [13] Gerald Brown, Matthew Carlyle, Javier Salmerón, and Kevin Wood. “Defending critical infrastructure.” In: *Interfaces* 36.6 (2006), pp. 530–544. DOI: [10.1287/inte.1060.0252](https://doi.org/10.1287/inte.1060.0252).
- [14] M. Campelo, S. Dantas, and S. Scheimberg. “A note on a penalty function approach for solving bilevel linear programs.” In: *Journal of Global Optimization* 16.3 (2000), pp. 245–255. DOI: [10.1023/A:1008308218364](https://doi.org/10.1023/A:1008308218364).
- [15] Alberto Caprara, Margarida Carvalho, Andrea Lodi, and Gerhard J. Woeginger. “Bilevel Knapsack with Interdiction Constraints.” In: *INFORMS Journal on Computing* 28.2 (2016), pp. 319–333. DOI: [10.1287/ijoc.2015.0676](https://doi.org/10.1287/ijoc.2015.0676).
- [16] Martha-Selene Casas-Ramírez and José-Fernando Camacho-Vallejo. “Solving the p -median bilevel problem with order through a hybrid heuristic.” In: *Applied Soft Computing* 60 (2017), pp. 73–86. DOI: [10.1016/j.asoc.2017.06.026](https://doi.org/10.1016/j.asoc.2017.06.026).
- [17] Paul C. Chu and John E. Beasley. “A genetic algorithm for the multidimensional knapsack problem.” In: *Journal of Heuristics* 4.1 (1998), pp. 63–86. DOI: [10.1023/A:1009642405419](https://doi.org/10.1023/A:1009642405419).
- [18] V. Chvátal. *Linear Programming*. Series of books in the mathematical sciences. W. H. Freeman, 1983.
- [19] O. Daxhelet and Yves Smeers. “The EU regulation on cross-border trade of electricity: A two-stage equilibrium model.” In: *European Journal of Operational Research* 181.3 (2007), pp. 1396–1412. DOI: [10.1016/j.ejor.2005.12.040](https://doi.org/10.1016/j.ejor.2005.12.040).
- [20] S. Dempe and J. Dutta. “Is Bilevel Programming a Special Case of a Mathematical Program with Complementarity Constraints?” In: *Mathematical Programming* 131.1 (2012), pp. 37–48. DOI: [10.1007/s10107-010-0342-1](https://doi.org/10.1007/s10107-010-0342-1).
- [21] Stephan Dempe. “A simple algorithm for the linear bilevel programming problem.” In: *Optimization* 18.3 (1987), pp. 373–385. DOI: [10.1080/02331938708843247](https://doi.org/10.1080/02331938708843247).
- [22] Stephan Dempe. *Bilevel Optimization: Theory, Algorithms, and Applications*. Tech. rep. 2018. URL: http://www.optimization-online.org/DB_HTML/2018/08/6773.html.
- [23] Stephan Dempe. *Foundations of bilevel programming*. Springer, 2002. DOI: [10.1007/b101970](https://doi.org/10.1007/b101970).

- [24] Stephan Dempe, Vyacheslav Kalashnikov, Gerardo A. Pérez-Valdés, and Nataliya Kalashnykova. *Bilevel Programming Problems*. Springer, 2015. DOI: [10.1007/978-3-662-45827-3](https://doi.org/10.1007/978-3-662-45827-3).
- [25] Scott DeNegre. “Interdiction and discrete bilevel linear programming.” PhD thesis. Lehigh University, 2011.
- [26] Elizabeth D. Dolan and Jorge J. Moré. “Benchmarking Optimization Software with Performance Profiles.” In: *Mathematical Programming* 91 (2 2002), pp. 201–213. DOI: [10.1007/s101070100263](https://doi.org/10.1007/s101070100263).
- [27] Jean-Pierre Dussault, Patrice Marcotte, Sébastien Roch, and Gilles Savard. “A smoothing heuristic for a bilevel pricing problem.” In: *European Journal of Operational Research* 174.3 (2006), pp. 1396–1413. DOI: [10.1016/j.ejor.2004.07.076](https://doi.org/10.1016/j.ejor.2004.07.076).
- [28] Matteo Fischetti, Ivana Ljubić, Michele Monaci, and Markus 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](https://doi.org/10.1287/opre.2017.1650).
- [29] Matteo Fischetti, Ivana Ljubić, Michele Monaci, and Markus Sinnl. *Instances and solver software for Mixed-Integer Bilevel Linear Problems*. <https://msinnl.github.io/pages/bilevel.html>. Accessed: 2018-12-13.
- [30] Matteo Fischetti, Ivana Ljubić, Michele Monaci, and Markus Sinnl. *Interdiction games and monotonicity, with application to knapsack problems*. 2. 2019, pp. 390–410. DOI: [10.1287/ijoc.2018.0831](https://doi.org/10.1287/ijoc.2018.0831).
- [31] Matteo Fischetti, Ivana Ljubić, Michele Monaci, and Markus Sinnl. “Intersection cuts for bilevel optimization.” In: *International Conference on Integer Programming and Combinatorial Optimization*. Springer. 2016, pp. 77–88. DOI: [10.1007/978-3-319-33461-5_7](https://doi.org/10.1007/978-3-319-33461-5_7).
- [32] Matteo Fischetti, Michele Monaci, and Markus Sinnl. “A dynamic reformulation heuristic for Generalized Interdiction Problems.” In: *European Journal of Operational Research* 267.1 (May 2018), pp. 40–51. DOI: [10.1016/j.ejor.2017.11.043](https://doi.org/10.1016/j.ejor.2017.11.043).
- [33] José Fortuny-Amat and Bruce McCarl. “A Representation and Economic Interpretation of a Two-Level Programming Problem.” In: *The Journal of the Operational Research Society* 32.9 (1981), pp. 783–792. JSTOR: [2581394](https://www.jstor.org/stable/2581394).
- [34] Steven A. Gabriel, Antonio J. Conejo, J. David Fuller, Benjamin F. Hobbs, and Carlos Ruiz. *Complementarity modeling in energy markets*. Vol. 180. Springer Science & Business Media, 2012.
- [35] Björn Geißler, Antonio Morsi, Lars Schewe, and Martin Schmidt. “Penalty Alternating Direction Methods for Mixed-Integer Optimization: A New View on Feasibility Pumps.” In: *SIAM Journal on Optimization* 27.3 (2017), pp. 1611–1636. DOI: [10.1137/16M1069687](https://doi.org/10.1137/16M1069687).
- [36] Björn Geißler, Antonio Morsi, Lars Schewe, and Martin Schmidt. “Solving Highly Detailed Gas Transport MINLPs: Block Separability and Penalty Alternating Direction Methods.” In: *INFORMS Journal on Computing* 30.2 (2018), pp. 309–323. DOI: [10.1287/ijoc.2017.0780](https://doi.org/10.1287/ijoc.2017.0780).
- [37] Björn Geißler, Antonio Morsi, Lars Schewe, and Martin Schmidt. “Solving power-constrained gas transportation problems using an MIP-based alternating direction method.” In: *Computers & Chemical Engineering* 82 (2015), pp. 303–317. DOI: [10.1016/j.compchemeng.2015.07.005](https://doi.org/10.1016/j.compchemeng.2015.07.005).
- [38] Jochen Gorski, Frank Pfeuffer, and Kathrin Klamroth. “Biconvex sets and optimization with biconvex functions: a survey and extensions.” In: *Mathematical Methods of Operations Research* 66.3 (2007), pp. 373–407. DOI: [10.1007/s00186-007-0161-1](https://doi.org/10.1007/s00186-007-0161-1).

- [39] Veronika Grimm, Julia Grübel, Lars Schewe, Martin Schmidt, and Gregor Zöttl. “Nonconvex equilibrium models for gas market analysis: Failure of standard techniques and alternative modeling approaches.” In: *European Journal of Operational Research* 273.3 (2019), pp. 1097–1108. DOI: [10.1016/j.ejor.2018.09.016](https://doi.org/10.1016/j.ejor.2018.09.016).
- [40] Veronika Grimm, Thomas Kleinert, Frauke Liers, Martin Schmidt, and Gregor Zöttl. “Optimal price zones of electricity markets: a mixed-integer multilevel model and global solution approaches.” In: *Optimization Methods and Software* 34.2 (2019), pp. 406–436. DOI: [10.1080/10556788.2017.1401069](https://doi.org/10.1080/10556788.2017.1401069).
- [41] Veronika Grimm, Alexander Martin, Martin Schmidt, Martin Weibelzahl, and Gregor Zöttl. “Transmission and Generation Investment in Electricity Markets: The Effects of Market Splitting and Network Fee Regimes.” In: *European Journal of Operational Research* 254.2 (2016), pp. 493–509. DOI: [10.1016/j.ejor.2016.03.044](https://doi.org/10.1016/j.ejor.2016.03.044).
- [42] Veronika Grimm, Lars Schewe, Martin Schmidt, and Gregor Zöttl. “A Multi-level Model of the European Entry-Exit Gas Market.” In: *Mathematical Methods of Operations Research* 89.2 (2019), pp. 223–255. DOI: [10.1007/s00186-018-0647-z](https://doi.org/10.1007/s00186-018-0647-z).
- [43] Pierre Hansen, Brigitte Jaumard, and Gilles 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](https://doi.org/10.1137/0913069).
- [44] X. He, C. Li, T. Huang, C. Li, and J. Huang. “A Recurrent Neural Network for Solving Bilevel Linear Programming Problem.” In: *IEEE Transactions on Neural Networks and Learning Systems* 25.4 (2014), pp. 824–830. DOI: [10.1109/TNNLS.2013.2280905](https://doi.org/10.1109/TNNLS.2013.2280905).
- [45] Xinmin Hu and Daniel Ralph. “Using EPECs to Model Bilevel Games in Restructured Electricity Markets with Locational Prices.” In: *Operations Research* 55.5 (2007), pp. 809–827. DOI: [10.1287/opre.1070.0431](https://doi.org/10.1287/opre.1070.0431).
- [46] Robert G. Jeroslow. “The polynomial hierarchy and a simple model for competitive analysis.” In: *Mathematical Programming* 32.2 (1985), pp. 146–164. DOI: [10.1007/BF01586088](https://doi.org/10.1007/BF01586088).
- [47] Nataliya I. Kalashnykova, Vyacheslav V. Kalashnikov, and Roberto C. Herrera Maldonado. “Bilevel Toll Optimization Problems: A Heuristic Algorithm Based Upon Sensitivity Analysis.” In: *Intelligent Decision Technologies*. Springer, 2012, pp. 135–143. DOI: [10.1007/978-3-642-29977-3_14](https://doi.org/10.1007/978-3-642-29977-3_14).
- [48] Thomas Kleinert, Martine Labbé, Fränk Plein, and Martin Schmidt. “There’s No Free Lunch: On the Hardness of Choosing a Correct Big-M in Bilevel Optimization.” In: *Operations Research* (2019). URL: http://www.optimization-online.org/DB_HTML/2019/04/7172.html. Forthcoming.
- [49] Thomas Kleinert and Martin Schmidt. “Global Optimization of Multilevel Electricity Market Models Including Network Design and Graph Partitioning.” In: *Discrete Optimization* 33 (2019), pp. 43–69. DOI: [10.1016/j.disopt.2019.02.002](https://doi.org/10.1016/j.disopt.2019.02.002).
- [50] Thorsten Koch, Tobias Achterberg, Erling Andersen, Oliver Bastert, Timo Berthold, Robert E. Bixby, Emilie Danna, Gerald Gamrath, Ambros M. Gleixner, Stefan Heinz, Andrea Lodi, Hans Mittelmann, Ted Ralphs, Domenico Salvagnin, Daniel E. Steffy, and Kati Wolter. “MIPLIB 2010.” In: *Mathematical Programming Computation* 3.2 (2011), pp. 103–163. DOI: [10.1007/s12532-011-0025-9](https://doi.org/10.1007/s12532-011-0025-9).
- [51] Martine Labbé and Alessia Violin. “Bilevel programming and price setting problems.” In: *4OR* 11.1 (2013), pp. 1–30. DOI: [10.1007/s10288-012-0213-0](https://doi.org/10.1007/s10288-012-0213-0).

- [52] H. Li and L. Zhang. “A Differential Evolution with Two Mutation Strategies for Linear Bilevel Programming Problems.” In: *2013 Ninth International Conference on Computational Intelligence and Security*. 2013, pp. 55–60. DOI: [10.1109/CIS.2013.19](https://doi.org/10.1109/CIS.2013.19).
- [53] H. Li and L. Zhang. “Solving Linear Bilevel Programming Problems Using a Binary Differential Evolution.” In: *2015 11th International Conference on Computational Intelligence and Security (CIS)*. 2015, pp. 38–42. DOI: [10.1109/CIS.2015.18](https://doi.org/10.1109/CIS.2015.18).
- [54] Andrea Lodi, Ted K. Ralphs, and Gerhard J. Woeginger. “Bilevel programming and the separation problem.” In: *Mathematical Programming* 146.1-2 (2014), pp. 437–458. DOI: [10.1007/s10107-013-0700-x](https://doi.org/10.1007/s10107-013-0700-x).
- [55] *MIPLIB 2017*. <http://miplib.zib.de/>. Accessed: 2018-12-14. 2018.
- [56] James T. Moore and Jonathan F. Bard. “The mixed integer linear bilevel programming problem.” In: *Operations Research* 38.5 (1990), pp. 911–921. DOI: [10.1287/opre.38.5.911](https://doi.org/10.1287/opre.38.5.911).
- [57] K. Pan, Y. Yang, and J. Liu. “A New Approach for Solving Linear Bilevel Programming Using Differential Evolution.” In: *2012 Sixth International Conference on Genetic and Evolutionary Computing*. 2012, pp. 453–456. DOI: [10.1109/ICGEC.2012.24](https://doi.org/10.1109/ICGEC.2012.24).
- [58] S. Pineda, H. Bylling, and J. M. Morales. “Efficiently solving linear bilevel programming problems using off-the-shelf optimization software.” In: *Optimization and Engineering* 19.1 (2018), pp. 187–211. DOI: [10.1007/s11081-017-9369-y](https://doi.org/10.1007/s11081-017-9369-y).
- [59] Salvador Pineda and Juan Miguel Morales. *Solving Linear Bilevel Problems Using Big-Ms: Not All That Glitters Is Gold*. 2018. URL: <https://arxiv.org/abs/1809.10448>.
- [60] J. Rajesh, Kapil Gupta, Hari Shankar Kusumakar, V. K. Jayaraman, and B. D. Kulkarni. “A Tabu Search Based Approach for Solving a Class of Bilevel Programming Problems in Chemical Engineering.” In: *Journal of Heuristics* 9.4 (2003), pp. 307–319. DOI: [10.1023/A:1025699819419](https://doi.org/10.1023/A:1025699819419).
- [61] Ted Ralphs. *Cor@l: Bilevel Optimization Problem Library*. <http://coral.ise.lehigh.edu/data-sets/bilevel-instances/>. Accessed: 2018-12-13.
- [62] Regionales Rechenzentrum Erlangen. *Woodcrest Cluster*. URL: <https://www.rrze.fau.de/dienste/arbeiten-rechnen/hpc/systeme/woodcrest-cluster.shtml> (visited on 12/17/2018).
- [63] Jörg Heitkoetter S. Khuri T. Baeck. *SAC94 Suite: Collection of Multiple Knapsack Problems*. <http://www.cs.cmu.edu/Groups/AI/areas/genetic/ga/test/sac/0.html>. Accessed: 2018-12-13.
- [64] Maria P. Scaparra and Richard L. Church. “A bilevel mixed-integer program for critical infrastructure protection planning.” In: *Computers & Operations Research* 35.6 (2008), pp. 1905–1923. DOI: [10.1016/j.cor.2006.09.019](https://doi.org/10.1016/j.cor.2006.09.019).
- [65] Lars Schewe, Martin Schmidt, and Dieter Wening. *A Decomposition Heuristic for Mixed-Integer Supply Chain Problems*. Tech. rep. 2019. URL: http://www.optimization-online.org/DB_HTML/2019/03/7120.html.
- [66] Heinrich von Stackelberg. *Marktform und gleichgewicht*. Springer, 1934.
- [67] Heinrich von Stackelberg. *Theory of the market economy*. Oxford University Press, 1952.
- [68] Georg Still. “Linear bilevel problems: Genericity results and an efficient method for computing local minima.” In: *Mathematical Methods of Operations Research* 55.3 (2002), pp. 383–400. DOI: [10.1007/s001860200](https://doi.org/10.1007/s001860200).
- [69] El-Ghazali Talbi. *Metaheuristics for Bi-level Optimization*. Vol. 482. May 2013. DOI: [10.1007/978-3-642-37838-6](https://doi.org/10.1007/978-3-642-37838-6).

- [70] Yen Tang, Jean-Philippe P. Richard, and J. Cole Smith. “A class of algorithms for mixed-integer bilevel min–max optimization.” In: *Journal of Global Optimization* 66.2 (2015), pp. 225–262. DOI: [10.1007/s10898-015-0274-7](https://doi.org/10.1007/s10898-015-0274-7).
- [71] Robert J. Vanderbei. *Linear Programming*. Springer, 2014. DOI: [10.1007/978-1-4614-7630-6](https://doi.org/10.1007/978-1-4614-7630-6).
- [72] L. Vicente, G. Savard, and J. Júdice. “Descent approaches for quadratic bilevel programming.” In: *Journal of Optimization Theory and Applications* 81.2 (1994), pp. 379–399. DOI: [10.1007/BF02191670](https://doi.org/10.1007/BF02191670).
- [73] Ue-Pyng Wen and Shuh-Tzy Hsu. “Linear Bi-Level Programming Problems – A Review.” In: *The Journal of the Operational Research Society* 42.2 (1991), pp. 125–133. URL: <http://www.jstor.org/stable/2583177>.
- [74] Richard E. Wendell and Arthur P. Hurter Jr. “Minimization of a non-separable objective function subject to disjoint constraints.” In: *Operations Research* 24.4 (1976), pp. 643–657. DOI: [10.1287/opre.24.4.643](https://doi.org/10.1287/opre.24.4.643).
- [75] Pan Xu and Lizhi Wang. “An exact algorithm for the bilevel mixed integer linear programming problem under three simplifying assumptions.” In: *Computers & Operations Research* 41 (2014), pp. 309–318.
- [76] Pan Xu and Lizhi Wang. *BMILP Library*. <http://lzwang.public.iastate.edu/bmilplib.html>. Accessed: 2018-12-14.

(T. Kleinert, M. Schmidt) ¹FRIEDRICH-ALEXANDER-UNIVERSITÄT ERLANGEN-NÜRNBERG, DISCRETE OPTIMIZATION, CAUERSTR. 11, 91058 ERLANGEN, GERMANY; ²ENERGIE CAMPUS NÜRNBERG, FÜRTH STR. 250, 90429 NÜRNBERG, GERMANY; ³TRIER UNIVERSITY, DEPARTMENT OF MATHEMATICS, UNIVERSITÄTSRING 15, 54296 TRIER, GERMANY

Email address: thomas.kleinert@fau.de

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