

# Bilevel linear optimization belongs to NP and admits polynomial-size KKT-based reformulations

Christoph Buchheim, Fakultät für Mathematik,  
Technische Universität Dortmund, Germany  
`christoph.buchheim@math.tu-dortmund.de`

October 10, 2023

## Abstract

It is a well-known result that bilevel linear optimization is NP-hard. In many publications, reformulations as mixed-integer linear optimization problems are proposed, which suggests that the decision version of the problem belongs to NP. However, to the best of our knowledge, a rigorous proof of membership in NP has never been published, so we close this gap by reporting a simple but not entirely trivial proof. A related question is whether a large enough “big M” for the classical KKT-based reformulation can be computed efficiently, which we answer in the affirmative. In particular, our big M has polynomial encoding length in the original problem data.

## 1 Introduction

We consider a general bilevel optimization problem with linear constraints at both the upper and lower level. Given as input the data  $c \in \mathbb{Q}^n$ ,  $d, q \in \mathbb{Q}^m$ ,  $A \in \mathbb{Q}^{k \times n}$ ,  $B \in \mathbb{Q}^{k \times m}$ ,  $a \in \mathbb{Q}^k$ ,  $T \in \mathbb{Q}^{r \times n}$ ,  $W \in \mathbb{Q}^{r \times m}$ , and  $h \in \mathbb{Q}^r$ , such a problem can be written as

$$\begin{aligned} \min \quad & c^\top x + d^\top y \\ \text{s.t.} \quad & Ax + By = a, \ x \geq 0 \\ & y \in \operatorname{argmin} \{q^\top \bar{y} \mid T\bar{x} + W\bar{y} = h, \ \bar{y} \geq 0\}. \end{aligned} \tag{BLP}$$

In this formulation, we implicitly assume the *optimistic* scenario: whenever the lower level problem

$$\begin{aligned} \min \quad & q^\top y \\ \text{s.t.} \quad & T\bar{x} + Wy = h, \ y \geq 0, \end{aligned} \tag{LL}$$

for some given upper level choice  $\bar{x} \geq 0$ , has more than one optimal solution satisfying the *coupling constraints*  $A\bar{x} + By = a$ , we assume that one minimizing  $d^\top y$  is chosen, i.e., a best possible one for the upper level problem. In the *pessimistic scenario*, it is instead assumed that the chosen  $y$  *maximizes*  $d^\top y$  over all optimal solutions of (LL). For a further discussion of the optimistic and the pessimistic scenario as well as structural properties of (BLP) and (LL), as a parametric optimization problem, we refer to [2, Chapters 2 and 3] and the references therein.

In the optimistic formulation (BLP), any  $\bar{x} \geq 0$  that renders the lower level problem (LL) infeasible or unbounded is an infeasible choice for the upper level by definition, and the same is true when no optimal solution of (LL) satisfies the coupling constraints. In the pessimistic case, it is natural to assume that  $\bar{x}$  is feasible only if *all* optimal choices of (LL) satisfy the coupling constraints. This is in line with the interpretation that the follower, i.e., the optimizer of (LL), is an adversary to the leader, i.e., the optimizer of (BLP), in the pessimistic case.

Problem (BLP) is known to be strongly NP-hard [7]. Essentially, one can model implicit binarity constraints for the upper level variables  $x$  by an appropriate lower level construction [1], so that every single-level binary linear optimization problem can be represented in the form (BLP): for each binary variable  $x_i$ , a new variable  $y_i$  together with constraints  $y_i \leq x_i$  and  $y_i \leq 1 - x_i$  and the objective  $\max y_i$  are introduced on the lower level. This implies  $y_i = \min\{x_i, 1 - x_i\}$  for every bilevel-feasible solution. On the upper level, a constraint  $y_i = 0$  is added, then  $\min\{x_i, 1 - x_i\} = 0$  and hence  $x_i \in \{0, 1\}$ . To avoid the coupling constraint, one may also minimize  $y_i$  in the upper level and add linear constraints  $x_i \in [0, 1]$ . Vicente et al. [12] show that even checking local optimality of a given feasible solution to (BLP) is NP-hard.

The complexity of (BLP) is further investigated by Deng [3]. Besides discussing hardness results and efficiently solvable special cases, he claims that bilevel linear optimization is NP-easy, i.e., can be Turing-reduced to some NP-complete problem [3, Theorem 6.2]. However, instead of a proof, only a reference to an unpublished (and untitled) manuscript is given. This result is related to, but weaker than the assertion that the decision version of (BLP) belongs to NP. To the best of our knowledge, no rigorous proof of NP-membership has appeared in the literature yet, a gap that we will fill in Section 2 below.

One of the classical approaches for solving Problem (BLP) in practice is to reformulate it as a mixed integer linear optimization problem, using the Karush-Kuhn-Tucker (KKT) conditions to model optimality for the lower level problem [4]. More precisely, in a first step, one can rewrite (BLP) equivalently as

$$\begin{aligned} \min \quad & c^\top x + d^\top y \\ \text{s.t.} \quad & Ax + By = a, \quad x \geq 0 \\ & Tx + Wy = h, \quad y \geq 0 \\ & W^\top \lambda \leq q \\ & y^\top (q - W^\top \lambda) = 0. \end{aligned} \tag{BLP-KKT}$$

The main difficulty here is the non-linear complementarity constraint  $y^\top (q - W^\top \lambda) = 0$ . It is typically linearized by introducing binary variables  $z_i \in \{0, 1\}$  and constraints

$$y_i \leq M_p(1 - z_i), \quad (q - W^\top \lambda)_i \leq M_d z_i \tag{1}$$

for sufficiently large  $M_p, M_d \in \mathbb{R}$ . This ensures that either  $y_i = 0$  or  $(q - W^\top \lambda)_i = 0$  for all  $i$ .

In general, choosing  $M_p$  and  $M_d$  correctly is not a trivial task: too small values of  $M_p$  and  $M_d$  may cut off lower level optimal solutions of (BLP), as (1) implies that  $y_i \leq M_p$  and  $(q - W^\top \lambda)_i \leq M_d$  for all  $i$ . This problem is highlighted in [10]. In fact, Kleinert et al. [9] recently showed that it is a co-NP-complete problem to decide whether given values for  $M_p$  and  $M_d$  are *bilevel-correct*, i.e., yield an equivalent reformulation of (BLP). They claim that their results “*imply that there is no hope for an efficient, i.e., polynomial-time, general-purpose method for verifying or computing a correct big-M in bilevel optimization unless P=NP*” [9]. However, it is not true that the second part of this claim follows from the first one: it may still be possible to compute *some* bilevel-correct  $M_p$  and  $M_d$  in polynomial time, even if verifying bilevel-correctness for *given*  $M_p$  and  $M_d$  is not possible. Stated differently, the result of Kleinert et al. rules out that the *smallest possible* bilevel-correct  $M_p$  and  $M_d$  can be found efficiently, if  $P \neq NP$ , but it does *not* rule out that there exist bilevel-correct  $M_p$  and  $M_d$  that can be computed efficiently, and thus in particular have polynomial encoding length – even though they may be much larger than necessary.

In fact, devising an approach for constructing bilevel-correct  $M_p$  and  $M_d$  in polynomial time is closely related to showing that (the decision version of) Problem (BLP) is contained in NP, i.e., that there exist certificates of polynomial size. It is quite natural to define the certificate as a feasible solution to either (BLP) or the KKT reformulation (BLP-KKT). However, it is not obvious that there exists such a solution of polynomial encoding length. In the following, we first give a short proof that the decision version of Problem (BLP) indeed belongs to NP (Section 2) and then show how to compute bilevel-correct  $M_p$  and  $M_d$  in polynomial time (Section 3).

## 2 Membership in NP

We first show that the decision version of Problem (BLP) belongs to NP. For a precise definition of the class NP and the corresponding certificates, as well as encoding lengths and other complexity-theoretic concepts, we refer to [6]. As is common for optimization problems, we formally define the decision version of Problem (BLP) as follows:

### Decision version of (BLP), optimistic scenario

Given: Problem data  $c \in \mathbb{Q}^n$ ,  $d, q \in \mathbb{Q}^m$ ,  $A \in \mathbb{Q}^{k \times n}$ ,  $B \in \mathbb{Q}^{k \times m}$ ,  $a \in \mathbb{Q}^k$ ,  $T \in \mathbb{Q}^{r \times n}$ ,  $W \in \mathbb{Q}^{r \times m}$ ,  $h \in \mathbb{Q}^r$ , and a number  $\alpha \in \mathbb{Q}$ .

Task: Decide whether there exists  $\bar{x} \geq 0$  such that (LL) has an optimal solution and such that at least one optimal solution  $\bar{y}$  of (LL) satisfies  $A\bar{x} + B\bar{y} = a$  and  $c^\top \bar{x} + d^\top \bar{y} \leq \alpha$ .

Note that this decision version of (BLP) is equivalent to the problem of determining whether (BLP) admits any feasible solution, since the constraint  $c^\top x + d^\top y \leq \alpha$  can be seen as an additional coupling constraint.

The idea of the following proof is to reformulate the feasibility of a given solution to Problem (BLP) that also satisfies  $c^\top x + d^\top y \leq \alpha$  by a system of linear constraints. However, in order to enforce the optimality of the lower level solution, we need to choose an optimal basis for (LL), which will then form the desired certificate. In other words, the NP-hardness of (BLP) is only due to the required choice of the optimal lower level basis. An important technical ingredient in the following proof is that the reduced costs of a given basis for (LL) do not depend on  $\bar{x}$ .

**Theorem 1.** *The decision version of Problem (BLP) in the optimistic scenario belongs to NP.*

*Proof.* Assume that a yes-instance is given, i.e., that there exist feasible  $x' \in \mathbb{R}^n$  and  $y' \in \mathbb{R}^m$  for (BLP) with  $c^\top x' + d^\top y' \leq \alpha$ . Let  $\mathcal{B}$  be any optimal basis for the lower level problem for  $x'$ . Then  $\mathcal{B}$  has non-negative reduced costs w.r.t.  $q$  and  $(x', y')$  together satisfy the linear constraints

$$c^\top x + d^\top y \leq \alpha \tag{2a}$$

$$Ax + By = a, \quad x \geq 0 \tag{2b}$$

$$Tx + Wy = h, \quad y \geq 0 \tag{2c}$$

$$q^\top y = q_{\mathcal{B}}^\top W_{\mathcal{B}}^{-1}(h - Tx) \tag{2d}$$

$$W_{\mathcal{B}}^{-1}(h - Tx) \geq 0. \tag{2e}$$

The constraint (2d) follows from the optimality of  $y'$ , since  $W_{\mathcal{B}}^{-1}(h - Tx')$  is the basic part of the optimal solution corresponding to  $\mathcal{B}$ . The constraint (2e) expresses that  $\mathcal{B}$  is a feasible basis for the lower level problem for  $x'$ .

Conversely, assume that there exists a basis  $\mathcal{B}$  of  $W$  with non-negative reduced costs w.r.t.  $q$  such that (2) is satisfied by some  $\bar{x}$  and  $\bar{y}$ . By (2e), the basis  $\mathcal{B}$  is feasible for the lower level problem for  $\bar{x}$ , hence also optimal because of the non-negative reduced costs w.r.t.  $q$ . It thus follows from (2d) that  $\bar{y}$  is an optimal lower level solution for  $\bar{x}$ . Together with the remaining constraints in (2), we derive that  $\bar{x}$  and  $\bar{y}$  form a feasible solution to (BLP) with  $c^\top \bar{x} + d^\top \bar{y} \leq \alpha$ .

A polynomial-size certificate for a yes-instance thus consists of a basis  $\mathcal{B} \subseteq \{1, \dots, r\}$  with non-negative reduced costs w.r.t.  $q$  such that the linear system (2) is solvable. The latter can be tested in time polynomial in the coefficients of (2), which are in turn polynomial in the coefficients of the original problem (BLP).  $\square$

Note that the basis  $\mathcal{B}$  in the above proof does not necessarily correspond to the solution  $y'$  or  $\bar{y}$ , if the lower level solution is not unique. In fact, in the optimistic setting considered here, it may happen that no basic solution of the follower's problem is feasible. The constraints (2c) and (2d)

in the proof are thus needed to model the optimal face of the lower level problem. As a (somewhat pathological) example, which does not contain any leader variables, consider

$$\begin{aligned} \min \quad & y \\ \text{s.t.} \quad & y = 1 \\ & y \in \operatorname{argmin} \{0 \mid \bar{y} \leq 2, \bar{y} \geq 0\} = [0, 2]. \end{aligned}$$

The optimal value of this problem is 1, since  $y = 1$  is the unique bilevel-feasible solution. However, the only two basic solutions of the follower's subproblem are  $y = 0$  and  $y = 2$ , which do not satisfy the upper level constraint  $y = 1$ . On the other hand, if there are no upper level coupling constraints, i.e., if  $B = 0$ , it is easy to see that there always exists an optimal lower level *basic* solution; see also [2, Theorem 2.1].

We emphasize that the proof of Theorem 1 does not make any assumptions on the feasibility or boundedness of the lower level problem (LL): since we start with feasible  $x'$  and  $y'$ , we know by definition that  $y'$  is an optimal lower level response to the upper level decision  $x'$ . Hence the lower level problem cannot be unbounded or infeasible for the given  $x'$ . Moreover, the second part of the proof shows that a certificate cannot exist if there is no upper level decision  $\bar{x}$  such that the lower level problem (LL) has an optimal solution.

**Remark 2.** *The proof of Theorem 1 can be easily extended to show that the decision version of an  $\ell$ -level linear optimization problem (in the optimistic scenario) belongs to  $\Sigma_{\ell-1}^P$ , for all  $\ell \geq 1$ . For  $\ell = 1$ , this just means that linear optimization is tractable. Assume recursively that the claim holds for some  $\ell \geq 1$  and consider an  $(\ell + 1)$ -level linear optimization problem. Again using as certificate a basis  $\mathcal{B}$  with non-negative reduced costs for the lowest level problem, and applying the same reformulation as in the proof of Theorem 1, we obtain a feasibility problem for an  $\ell$ -level linear optimization problem, depending on  $\mathcal{B}$ , in place of (2). By our recursive assumption, the latter feasibility problem belongs to  $\Sigma_{\ell-1}^P$ . In summary, the  $(\ell + 1)$ -level problem can be decided by an NP-algorithm using an oracle for a problem in  $\Sigma_{\ell-1}^P$ , so that it belongs to  $\Sigma_{\ell}^P$  by definition.*

We now turn our attention to the pessimistic case, where the follower tries to violate the coupling constraints and, if not possible, returns a worst-possible solution for the upper level problem. We first define the decision version again:

Decision version of (BLP), pessimistic scenario

Given: Problem data  $c \in \mathbb{Q}^n$ ,  $d, q \in \mathbb{Q}^m$ ,  $A \in \mathbb{Q}^{k \times n}$ ,  $B \in \mathbb{Q}^{k \times m}$ ,  $a \in \mathbb{Q}^k$ ,  $T \in \mathbb{Q}^{r \times n}$ ,  $W \in \mathbb{Q}^{r \times m}$ ,  $h \in \mathbb{Q}^r$ , and a number  $\alpha \in \mathbb{Q}$ .

Task: Decide whether there exists  $\bar{x} \geq 0$  such that (LL) has an optimal solution and such that all optimal solutions  $\bar{y}$  of (LL) satisfy  $A\bar{x} + B\bar{y} = a$  and  $c^\top \bar{x} + d^\top \bar{y} \leq \alpha$ .

The universal quantifier in this formulation creates some additional difficulty: for a yes-instance, we have to make sure that  $\bar{x}$  is chosen such that all optimal solutions of the lower level problem satisfy the coupling constraints and  $c^\top \bar{x} + d^\top \bar{y} \leq \alpha$ . More abstractly, we thus have to enforce an inclusion relation between two polyhedra that are both parametrized by  $\bar{x}$ . In the following proof, we need a second basis in our certificate to achieve this.

**Theorem 3.** *The decision version of Problem (BLP) in the pessimistic scenario belongs to NP.*

*Proof.* Assume that a yes-instance is given. Then, in particular, some  $\bar{x} \geq 0$  exists such that (LL) has an optimal solution. This means that there exists a basis  $\mathcal{B}$  of  $W$  with non-negative reduced costs w.r.t.  $q$  such that the corresponding basic solution is feasible, i.e., such that  $\bar{x}$  satisfies

$$W_{\mathcal{B}}^{-1}(h - Tx) \geq 0, \quad x \geq 0. \quad (3)$$

Moreover, every optimal solution  $y$  of (LL) satisfies both  $A\bar{x} + By = a$  and  $c^\top \bar{x} + d^\top y \leq \alpha$ . In other words, we have that all solutions  $y$  of

$$T\bar{x} + Wy = h, \quad q^\top y = q_B^\top W_B^{-1}(h - T\bar{x}), \quad y \geq 0 \quad (4)$$

also satisfy

$$A\bar{x} + By = a, \quad c^\top \bar{x} + d^\top y \leq \alpha.$$

In particular, we deduce

$$\max_{(4)} d^\top y \leq -c^\top \bar{x} + \alpha, \quad (5)$$

and this maximization problem has an optimal solution since it is bounded and (4) is feasible. Since (4) is in standard form, there exists a basis  $\hat{\mathcal{B}}$  for  $\hat{W} := \begin{pmatrix} W \\ q^\top \end{pmatrix}$  with non-positive reduced costs w.r.t.  $d$  such that  $\bar{x}$  satisfies the linear constraints

$$d_{\hat{\mathcal{B}}}^\top \hat{W}_{\hat{\mathcal{B}}}^{-1} \begin{pmatrix} h - T\bar{x} \\ q_B^\top W_B^{-1}(h - T\bar{x}) \end{pmatrix} \leq -c^\top \bar{x} + \alpha \quad (6a)$$

$$\hat{W}_{\hat{\mathcal{B}}}^{-1} \begin{pmatrix} h - T\bar{x} \\ q_B^\top W_B^{-1}(h - T\bar{x}) \end{pmatrix} \geq 0. \quad (6b)$$

Indeed, constraint (6b) states that the basis  $\hat{\mathcal{B}}$  is feasible for (4), while constraint (6a) states that the corresponding basic solution  $y$  satisfies the constraint  $c^\top \bar{x} + d^\top y \leq \alpha$ .

Proceeding to the coupling constraints  $A\bar{x} + By = a$ , let  $b^i$  and  $a^i$  denote the  $i$ -th row of  $A$  and  $B$ , respectively, and let  $a_i$  denote the  $i$ -th entry of  $a$ . Then, for all  $i = 1, \dots, k$ , the basis  $\hat{\mathcal{B}}$  is optimal for both minimizing and maximizing  $b^i y$  over (4), since  $b^i y = -a^i \bar{x} + a_i$  for all  $y$  satisfying (4). The reduced costs of  $\hat{\mathcal{B}}$  w.r.t.  $b^i$  are thus zero, and the linear constraints

$$b_{\hat{\mathcal{B}}}^i \hat{W}_{\hat{\mathcal{B}}}^{-1} \begin{pmatrix} h - T\bar{x} \\ q_B^\top W_B^{-1}(h - T\bar{x}) \end{pmatrix} = -a^i \bar{x} + a_i \quad (7)$$

hold for  $\bar{x}$ , for all  $i = 1, \dots, k$ .

To sum up, for any yes-instance, there exist a basis  $\mathcal{B}$  of  $W$  with non-negative reduced costs w.r.t.  $q$  and a basis  $\hat{\mathcal{B}}$  of  $\hat{W}$  with non-positive reduced costs w.r.t.  $d$  and zero reduced costs w.r.t.  $b^1, \dots, b^k$ , such that all linear constraints on  $x$  in (3), (6), and (7) can be satisfied simultaneously.

Conversely, if such bases exist, the instance is a yes-instance. Indeed, by (3) the corresponding feasible solution  $\bar{x}$  admits an optimal response. From (6) it follows that

$$\max_{(4)} d^\top y = d_{\hat{\mathcal{B}}}^\top \hat{W}_{\hat{\mathcal{B}}}^{-1} \begin{pmatrix} h - T\bar{x} \\ q_B^\top W_B^{-1}(h - T\bar{x}) \end{pmatrix} \leq -c^\top \bar{x} + \alpha,$$

as  $\hat{\mathcal{B}}$  is a feasible and hence optimal basis for the given maximization problem. Thus  $c^\top \bar{x} + d^\top y \leq \alpha$  for all  $y$  feasible for (4). From (7) we similarly derive  $A\bar{x} + By = a$  for all such  $y$ . The result thus follows with the certificate consisting of the two bases  $\mathcal{B}$  and  $\hat{\mathcal{B}}$ .  $\square$

We emphasize that, different from the proof of Theorem 1, the linear constraints constructed in the proof of Theorem 3 do not contain the lower level variables, they only restrict the upper level decision. This is in line with the fact that, in the pessimistic case, the lower level variables appear in a universal instead of an existential quantifier. In order to deal with this ‘‘adversarial’’ quantifier, the second basis  $\hat{\mathcal{B}}$  is needed in the certificate.

One may be tempted to simplify the above proof by replacing the maximization problem in (5) by its dual and then resolving the resulting minimization problem by introducing additional variables. However, this approach would lead to a non-linear model, because the new variables would be multiplied by the variables  $x$  appearing on the right hand side of (4). For the resulting quadratic system, testing feasibility would not be trivially polynomial any more.

### 3 Computation of bilevel-correct bounds

For sake of simplicity, we focus on the optimistic scenario in the remainder of this paper. From an abstract point of view, it already follows from Theorem 1 that there exists a polynomial-time algorithm that reformulates any bilevel linear optimization problem into an equivalent mixed-integer linear optimization problem. Indeed, as mixed-integer linear optimization is NP-hard, any decision problem in NP can be polynomially reduced to it. We now show that this task can actually be achieved by the KKT-reformulation (BLP-KKT), after deriving bilevel-correct bounds. In fact, the proof of Theorem 1 already implies that we can compute a bound on the entries of  $x$  and  $y$  in polynomial time such that this bound does not cut off all optimal solutions. We now make this bound more explicit and extend this statement to the dual variables. For the convenience of the reader, we first report the following elementary result. We will denote by  $[X]$  the maximal absolute value of any entry in an integer matrix or vector  $X$ .

**Lemma 4.** *Let  $P = \{x \in \mathbb{R}^n \mid Ax = b, x \geq 0\}$  be a polyhedron with  $A \in \mathbb{Z}^{m \times n}$  and  $b \in \mathbb{Z}^m$ . Let  $\bar{x}$  be a vertex of  $P$ . Then  $\bar{x}_i \leq m! [b] [A]^{m-1}$  for all  $i = 1, \dots, n$ .*

*Proof.* By removing redundant constraints, we may assume that  $A$  has full row rank. Let  $\mathcal{B}$  be a basis yielding  $\bar{x}$ . Then  $\bar{x}_i = 0$  for  $i \notin \mathcal{B}$  and  $x_{\mathcal{B}} = A_{\mathcal{B}}^{-1}b$ . Hence, for  $i \in \mathcal{B}$ , we have  $|\bar{x}_i| \leq |\det(A_{\mathcal{B}}^i)|$  by Cramer's rule, where  $A_{\mathcal{B}}^i$  arises from  $A_{\mathcal{B}}$  by replacing column  $i$  by  $b$ ; note that  $|\det(A_{\mathcal{B}})| \geq 1$  by integrality and regularity of  $A_{\mathcal{B}}$ . Hence

$$|\bar{x}_i| \leq \left| \sum_{\sigma \in S_m} \text{sgn}(\sigma) \prod_{j=1}^m (A_{\mathcal{B}}^i)_{j, \sigma(j)} \right| \leq \sum_{\sigma \in S_m} \prod_{j=1}^m |(A_{\mathcal{B}}^i)_{j, \sigma(j)}|,$$

which directly implies the result.  $\square$

Note that the number  $m!$  appearing in the bound above has polynomial encoding length in  $m \in \mathbb{N}$ , since  $m! \leq m^m$  and the latter has encoding length at most  $m(\lceil \log(m+1) \rceil + 1)$ . In particular, the naive algorithm to compute  $m!$  runs in polynomial time, but there exist faster algorithms performing this task in  $O(m \log^2 m)$  time [8, 11].

**Theorem 5.** *Bilevel-correct values for  $M_p$  and  $M_d$  can be computed in polynomial time from the data of Problem (BLP).*

*Proof.* We may assume that (BLP) admits an optimal solution  $x', y'$ . As in the proof of Theorem 1, we choose  $\mathcal{B}$  as an optimal basis of the lower level problem for  $x'$ . Let  $(\bar{x}, \bar{y}, \bar{z})$  be an optimal vertex of the polyhedron defined by

$$\begin{aligned} Ax + By &= a \\ Tx + Wy &= h \\ q_{\mathcal{B}}^{\top} W_{\mathcal{B}}^{-1} Tx + q^{\top} y &= q_{\mathcal{B}}^{\top} W_{\mathcal{B}}^{-1} h \\ W_{\mathcal{B}}^{-1} Tx + z &= W_{\mathcal{B}}^{-1} h \\ x, y, z &\geq 0, \end{aligned} \tag{8}$$

with respect to the objective function  $c^{\top}x + d^{\top}y$ . Then  $\bar{\lambda} := W_{\mathcal{B}}^{-\top} q_{\mathcal{B}}$  is a dual optimal solution for the lower level problem for  $\bar{x}$ . In particular, we have  $(h - T\bar{x})^{\top} \bar{\lambda} = q^{\top} \bar{y}$ , which implies the complementarity constraint  $\bar{y}^{\top} (q - W^{\top} \bar{\lambda}) = 0$ . In summary, the constructed solution consisting of  $\bar{x}$ ,  $\bar{y}$ , and  $\bar{\lambda}$  is feasible and hence optimal for (BLP-KKT). It thus suffices to bound the entries of  $\bar{y}$  and  $q - W^{\top} \bar{\lambda}$  in terms of the problem data of (BLP), independently of the choice of  $\mathcal{B}$ . By scaling all constraints and objective functions, we may assume that all coefficients in (BLP) are integer. This can be done in polynomial time, as it suffices to scale by the product of all numerators of these coefficients. In particular, the encoding length of the scaled instance is polynomial in the original encoding length.

To obtain a valid choice of  $M_d$ , we need  $M_d \geq (q - W^{\top} \bar{\lambda})_i = (q - W^{\top} W_{\mathcal{B}}^{-\top} q_{\mathcal{B}})_i$  for all bases  $\mathcal{B}$  of  $W$ . For  $i \in \mathcal{B}$  we have  $(q - W^{\top} \bar{\lambda})_i = 0$  by construction, while the remaining entries contain the

reduced costs of  $\mathcal{B}$ . It thus suffices to compute a bound on the reduced costs of any basis of  $W$  in terms of  $W$  and  $q$ . For  $i \in \mathcal{N}$ , we have

$$|q_i - W_{\cdot i}^\top W_{\mathcal{B}}^{-\top} q_{\mathcal{B}}| \leq |q_i| + |W_{\cdot i}^\top W_{\mathcal{B}}^{-\top} q_{\mathcal{B}}| \leq |q_i| + \sum_{j \in \mathcal{B}} |W_{ij}| |(W_{\mathcal{B}}^{-\top} q_{\mathcal{B}})_j|.$$

Now, since  $W$  is assumed to be integer, we have  $|\det(W_{\mathcal{B}}^\top)| \geq 1$ , so by Cramer's rule, each entry of  $W_{\mathcal{B}}^{-\top} q_{\mathcal{B}}$  is bounded by  $\det(\bar{W})$ , where  $\bar{W}$  arises from  $W_{\mathcal{B}}$  by replacing some column by  $q_{\mathcal{B}}$ . Hence  $|(W_{\mathcal{B}}^{-\top} q_{\mathcal{B}})_j| \leq r! [q][W]^{r-1}$  and we can choose  $M_d$  as

$$[q] + r[W]r! [q][W]^{r-1} = [q](1 + r! r[W]^r).$$

To determine a bilevel-correct value for  $M_p$ , it suffices to bound the entries of  $y$  in all vertices of (8). Using Lemma 4, we can choose  $M_p$  as  $\ell! [f][L]^{\ell-1}$  where  $\ell := k + r + 1 + m$  is the number of equations in (8) and

$$\begin{aligned} [L] &:= \max\{[A], [B], [T], [W], [q_{\mathcal{B}}^\top W_{\mathcal{B}}^{-1} T], [q], [W_{\mathcal{B}}^{-1} T], 1\} \\ [f] &:= \max\{[a], [h], [q_{\mathcal{B}}^\top W_{\mathcal{B}}^{-1} h], [W_{\mathcal{B}}^{-1} h]\}. \end{aligned}$$

We further have

$$\begin{aligned} [W_{\mathcal{B}}^{-1} h] &\leq r! [W]^{r-1} [h] \\ [q_{\mathcal{B}}^\top W_{\mathcal{B}}^{-1} h] &\leq r! r[q][W]^{r-1} [h] \\ [W_{\mathcal{B}}^{-1} T] &\leq r! [W]^{r-1} [T] \\ [q_{\mathcal{B}}^\top W_{\mathcal{B}}^{-1} T] &\leq r! r[q][W]^{r-1} [T]. \end{aligned}$$

Altogether, we can compute bilevel-correct bounds  $M_d$  and  $M_p$  in polynomial time from the coefficients of (BLP).  $\square$

Note that the bound  $M_d$  derived in the proof of Theorem 5 only depends on  $W$  and  $q$ , while  $M_p$  depends on all problem data, including the upper level coefficients. If  $r \geq 1$  and none of  $q, W, T, h$  is all zero, the bounds can be simplified to

$$\begin{aligned} [L] &\leq \max\{[A], [B], r! r[q][W]^{r-1} [T]\} \\ [f] &\leq \max\{[a], r! r[q][W]^{r-1} [h]\}. \end{aligned}$$

Both bounds constructed in Theorem 5 have polynomial encoding length. However, their values are exponential in general. The latter cannot be avoided, even without the bilevel structure. As a simple example, consider constraints  $y_1 = 1$  and  $2y_i - y_{i+1} = 0$  for  $i = 1, \dots, m-1$ . Then all coefficients have size at most 2, but in the unique feasible solution, the value of  $y_n$  is  $2^{n-1}$ .

**Remark 6.** From the proof of Theorem 1 it can be argued that the NP-hardness of (BLP) is due to the exponential number of bases of the lower level problem. More precisely, even though it is possible to efficiently find an optimal lower level basis for any fixed upper level decision, since this reduces to solving a linear optimization problem, it is not possible to efficiently decide whether there exists a basis  $\mathcal{B}$  that yields a feasible system (2), unless  $P = NP$ . Otherwise, it would follow from the proof of Theorem 1 that (BLP) is tractable. In terms of the bounds  $M_d$  and  $M_p$ , Theorem 1 implies that we can determine bilevel-correct values by considering all possible lower level bases, which is the core idea of the proof of Theorem 5. If, however, the task is to test bilevel-correctness of given values of  $M_d$  and  $M_p$  as in [9], we need to decide whether there exists a basis  $\mathcal{B}$  such that some feasible solution of (2), with  $\alpha$  being the optimal value of (BLP), has a small enough encoding length. As shown in [9], the latter problem is NP-complete. Both observations are related to the hardness of the problem OVP, which was shown in [5] and used in the main proof of [9]: while it is easy to find an optimal basis for a given linear optimization problem, if one exists, it is hard to test whether an (unbounded) linear optimization problem has a basic solution exceeding a certain threshold. From an abstract point of view, the common difficulty of (BLP) and OVP is thus to determine a suitable basis for the respective task, out of exponentially many candidates.

**Acknowledgments** This work was inspired by discussions held at Dagstuhl Seminar 22441 “Optimization at the Second Level”. We thank the anonymous reviewer for many valuable comments that helped to significantly improve the paper.

## References

- [1] C. Audet, P. Hansen, B. Jaumard, and G. Savard. Links between linear bilevel and mixed 0–1 programming problems. *Journal of Optimization Theory and Applications*, 93:273–300, 1997.
- [2] S. Dempe, V. Kalashnikov, G. Pérez-Valdés, and N. Kalashnykova. *Bilevel Programming Problems – Theory, Algorithms and Applications to Energy Networks*. Springer-Verlag Berlin Heidelberg, 2015.
- [3] X. Deng. *Multilevel Optimization: Algorithms and Applications*, chapter Complexity Issues in Bilevel Linear Programming. Kluwer Academic Publishers, 1998.
- [4] J. Fortuny-Amat and B. McCarl. A representation and economic interpretation of a two-level programming problem. *The Journal of the Operational Research Society*, 32(9):783–792, 1981.
- [5] K. Fukuda, T. M. Liebling, and F. Margot. Analysis of backtrack algorithms for listing all vertices and all faces of a convex polyhedron. *Computational Geometry*, 8(1):1–12, 1997.
- [6] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric Algorithms and Combinatorial Optimization*. Springer Berlin, Heidelberg, second edition, 1993.
- [7] P. Hansen, B. Jaumard, and G. Savard. New branch-and-bound rules for linear bilevel programming. *SIAM Journal on Scientific and Statistical Computing*, 13(5):1194–1217, 1992.
- [8] D. Harvey and J. Van Der Hoeven. Integer multiplication in time  $O(n \log n)$ . *Annals of Mathematics*, 193(2):563–617, 2021.
- [9] T. Kleinert, M. Labbé, F. Plein, and M. Schmidt. There’s no free lunch: On the hardness of choosing a correct big-M in bilevel optimization. *Operations Research*, 68(6):1716–1721, 2020.
- [10] S. Pineda and J. M. Morales. Solving linear bilevel problems using big-Ms: Not all that glitters is gold. *IEEE Transactions on Power Systems*, 34(3):2469–2471, 2019.
- [11] A. Schönhage, A.F.W. Grotefeld, and E. Vetter. *Fast Algorithms – A Multitape Turing Machine Implementation*. BI Wissenschafts-Verlag, 1994.
- [12] L. Vicente, G. Savard, and J. Júdice. Descent approaches for quadratic bilevel programming. *Journal of Optimization Theory and Applications*, 81(2):379–399, 1994.