INDUSTRIAL AND SYSTEMS ENGINEERING

On the Complexity of Finding Locally Optimal Solutions in Bilevel Linear Optimization

Oleg A. Prokopyev

Plattenstrasse 14, Department of Business Administration, University of Zurich, Zurich 8032, Switzerland

Ted K. Ralphs

Department of Industrial and System Engineering, Lehigh University, Bethlehem, PA USA

COR@L Technical Report 24T-015-R1

LEHIGH UNIVERSITY

COR@L
COMPUTATIONAL OPTIMIZATION RESEARCH AT LEHIGH

On the Complexity of Finding Locally Optimal Solutions in Bilevel Linear Optimization

Oleg A. Prokopyev *[1] and Ted K. Ralphs †[2]

[1]Plattenstrasse 14, Department of Business Administration, University of Zurich, Zurich 8032, Switzerland

[2]Department of Industrial and System Engineering, Lehigh University, Bethlehem, PA USA

Original Publication: October 24, 2024

Last Revised: September 4, 2025

Abstract

We consider the computational complexity of finding locally optimal solutions to bilevel linear optimization problems (BLPs), from the leader's perspective. We show that, for any constant $c > 0$, the problem of finding a leader's solution that is within Euclidean distance c^n of any locally optimal leader's solution, where n is the total number of variables, is NP-hard. Our derivations exploit techniques similar to those used for the analogous result for quadratic optimization problems (QPs). As a side observation, we also provide a BLP reformulation of the celebrated Motzkin-Straus QP model for the maximum clique problem and thereby illuminate the close connection of combinatorial optimization problems to both BLPs and QPs.

1 Introduction

In this paper, we consider bilevel linear optimization problems (BLPs) of the form:

$$\max_{x} \quad a^\top x + d_1^\top y \tag{\textbf{BLP}-a}$$

$$\text{s.t.} \quad A_1 x \le b_1, \; x \in \mathbb{R}_+^{n_1}, \tag{\textbf{BLP}-b}$$

*oleg.prokopyev@business.uzh.ch

†ted@lehigh.edu

$$y \in \underset{\hat{y}}{\operatorname{argmax}} \left\{ d_2^\top \hat{y} \mid A_2 x + B\hat{y} \leq b_2,\ \hat{y} \in \mathbb{R}_+^{n_2} \right\}, \tag{BLP-c}$$

where $A_1 \in \mathbb{Q}^{m_1 \times n_1}$, $A_2 \in \mathbb{Q}^{m_2 \times n_1}$, $B \in \mathbb{Q}^{m_2 \times n_2}$, $a \in \mathbb{Q}^{n_1}$, $d_1 \in \mathbb{Q}^{n_2}$, $d_2 \in \mathbb{Q}^{n_2}$, $b_1 \in \mathbb{Q}^{m_1}$, and $b_2 \in \mathbb{Q}^{m_2}$. BLPs provide a general framework for hierarchical or game-theoretic decision-making in which a *leader* chooses the values of x and a *follower* sets the values of variables $y := y(x)$ in reaction to the leader's choice of x. The problem (**BLP**) is considered from the leader's point of view, and the goal is to select the best solution for the leader under certain assumptions about the follower's behavior.

Bilevel programming is an important class of mathematical optimization problems with multiple application domains, including network design [Baggio et al., 2021, Marcotte et al., 2009], pricing [Marcotte et al., 2009], and defense [Borrero et al., 2019]; see the recent surveys by Kleinert et al. [2021] and Beck et al. [2023]. Within bilevel optimization, BLPs are the simplest canonical class, analogous to classical linear programs (LPs) in single-level mathematical optimization.

Define $\mathcal{X} := \{x \in \mathbb{R}_+^{n_1} \mid A_1 x \leq b_1\}$, to be the set of the leader's decisions satisfying constraints (**BLP**-b), i.e., the *upper-level constraints.* Then, the set of possible follower's reactions is:

$$\mathcal{R}(x) := \operatorname{argmax} \left\{ d_2^\top y \mid By \leq b_2 - A_2 x,\ y \in \mathbb{R}_+^{n_2} \right\}, \tag{1}$$

i.e., (**BLP**-c) can be re-written as $y \in \mathcal{R}(x)$. We refer to $\mathcal{R}(x)$ it as the *follower's rational reaction set* with respect to x. Naturally, $\mathcal{R}(x)$ is not necessarily a singleton and different versions of (**BLP**), commonly known as the *optimistic* and *pessimistic* versions, are obtained under different assumptions on the part of the leader about how the follower selects an element of $\mathcal{R}(x)$; see, e.g., Lagos and Prokopyev [2023]. Our results apply to both these variants, since in our reductions below, the rational reaction set is always a singleton.

In the considered class of BLPs, the follower's variables do not appear in the upper-level constraints (**BLP**-b). As we demonstrate in this paper, it is difficult to construct locally optimal solutions even for this restricted class of problems. In the remainder of the paper, we also assume:

Assumption 1 (**A1**)**.** *The leader's feasible set $\mathcal{X}$ is non-empty and bounded.*

Assumption 2 (**A2**)**.** *For any $x \in \mathcal{X}$, the follower's feasible region is non-empty and bounded.*

These assumptions are standard in the related literature and are made for clarity of exposition. They guarantee that an optimal solution exists [Dempe, 2002] and that the follower's problem always has a finite optimal solution. In practice, some (or all) of these assumptions can be relaxed.

Computational complexity. Classical single-level LPs form a special case of BLPs and are known to be polynomial-time solvable [Khachiyan, 1979]. In contrast, general BLPs are computationally difficult. Ben-Ayed and Blair [1990] proved that BLPs are NP-hard, while Buchheim [2023] recently showed that the decision version of BLP is in NP and hence NP-complete. Hansen et al. [1992] showed that BLPs are strongly NP-hard even for the min-max case. A general complexity hierarchy for multilevel optimization problems is explored by Jeroslow [1985].

BLPs and more general classes of bilevel problems with LPs at the lower level can be reformulated as single-level mixed integer linear optimization problems (MILPs) via the LP optimality conditions. In fact, MILPs can also be reformulated as BLPs [Audet et al., 1997]. These reformulations enable solving BLPs using off-the-shelf MILP solvers. However, this approach typically requires a big-M parameter that is sufficiently large; see, e.g., Audet et al. [1997], Yang et al. [2023], Zare et al. [2019] for multiple examples of such reformulations. Buchheim [2023] showed that an appropriate big-M parameter can be computed efficiently for BLPs, though its magnitude is impractically large, despite being polynomially representable. In contrast, whether a given big-M yields an equivalent reformulation of a BLP cannot be verified in polynomial time unless $P = NP$; see Kleinert et al. [2020] and the related discussion by Buchheim [2023].

To date, we are not aware of any results that concern the computational complexity of finding locally optimal solutions to BLPs. To the best of our knowledge, the only relevant results are provided by Marcotte and Savard [2005] and Vicente et al. [1994], where it is shown that the problem of checking local optimality and the problem of checking strict local optimality of a given point are both NP-hard. However, these results do not rule out that for BLP there may exist a locally optimal leader's decision, which can be found efficiently. Here, we close that existing gap in the literature.

Local optimality (for the leader). The notion of local optimality must be carefully defined in the bilevel setting, since there are optimality conditions at both the upper and lower levels that can each be modified, in principle, to arrive at different notions of local optimality. For example, Shi et al. [2023] explore local optimality of the follower's response in order to provide upper and lower bounds for globally optimal solutions of the optimistic bilevel problem. The definition used in the present study relaxes global optimality of only the leader's decisions, which is meaningful from the modeling perspective, given that BLPs focus on optimizing the leader's objective function.

Let $\mathcal{B}(\hat{x}, \epsilon) = \{x \in \mathbb{R}^{n_1} \mid ||x - \hat{x}||_2 \leq \epsilon\}$ be the Euclidean ball of radius ϵ centered at $\hat{x}$. Then, we say that $\hat{x} \in \mathcal{X}$ is a *locally optimal leader's optimistic solution* to (**BLP**) if there exists $\epsilon > 0$ such that for any $x \in \mathcal{X} \cap \mathcal{B}(\hat{x}, \epsilon)$, we have that

$$a^\top \hat{x} + \max_{y\in\mathcal{R}(\hat{x})} d_1^\top y \geq a^\top x + \max_{y\in\mathcal{R}(x)} d_1^\top y. \tag{2}$$

Similarly, we say that $\hat{x} \in \mathcal{X}$ is a *locally optimal leader's pessimistic solution* to (**BLP**) if there exists $\epsilon > 0$ such that for any $x \in \mathcal{X} \cap \mathcal{B}(\hat{x}, \epsilon)$, we have that

$$a^\top \hat{x} + \min_{y\in\mathcal{R}(\hat{x})} d_1^\top y \geq a^\top x + \min_{y\in\mathcal{R}(x)} d_1^\top y. \tag{3}$$

As briefly mentioned above, the optimistic and pessimistic cases coincide in our reductions and our main result thus holds for both versions of BLP. In the remainder of the paper, we therefore drop the terms "optimistic" and "pessimistic." Also, to streamline the discussion, whenever we refer to a locally optimal solution of a BLP, we mean a locally optimal *leader's* solution.

For a motivating numerical example, consider the following BLP instance:

$$z^* = \max_{x,y} \quad z(x) := x - \frac{3}{2}y \tag{4a}$$

$$\text{s.t.} \quad 0 \leq x \leq 1,\ y \in \underset{\hat{y}}{\operatorname{argmax}}\{\hat{y} \mid \hat{y} \leq x,\ \hat{y} \leq 1 - x,\ \hat{y} \in \mathbb{R}_+\}. \tag{4b}$$

One can observe that if the leader's decision is $x \in [0, 0.5]$, then the leader's objective function $z(x) = -\frac{1}{2}x$; also, if $x \in [0.5, 1]$, then $z(x) = \frac{5}{2}x - \frac{3}{2}$. Clearly, $x^* = 1$ is the unique (globally) optimal solution for the leader, with $z^* := z(x^*) = 1$. However, observe that $\hat{x} = 0$, with $z(\hat{x}) = 0$, is locally optimal for the leader according to both (2) and (3). We refer to Appendix A for an illustration.

Contribution and outline. As noted in the discussion above, based on current knowledge, we cannot rule out the existence of a polynomial time algorithm that finds a locally optimal solution for BLP, or the existence of *some* easily verifiable locally optimal solution for BLP. Our main contribution is to show that neither of these is possible unless $P = NP$. Formally, our main result is stated as follows.

Theorem 1. *For any constant $c > 0$, the problem of finding $x \in \mathcal{X}$ such that $x \in \mathcal{B}(\hat{x}, c^{\Theta(n_1+n_2)})$ for some $\hat{x} \in \mathcal{X}$ that is locally optimal to* (**BLP**) *is* NP*-hard.*

In fact, the constant c in Theorem 1 indicates that we obtain a slightly more general result: it is hard not only to find a locally optimal leader's decision, but even to compute any leader's decision that is sufficiently close to such a local optimum. The problem is difficult for essentially the same reasons that make finding globally optimal solutions challenging (see further discussion in Appendix B). The aforementioned study by Buchheim [2023] shows that BLP can be interpreted as a combinatorial search over bases of the follower's reaction problem. Similarly, Deng [1998] prove that

BLPs are polynomial-time solvable for the fixed number of the follower's variables. We show that the problem of finding locally optimal solutions for BLPs is analogously combinatorial in its essence.

Specifically, we prove the main result in Theorem 1 by reducing the problem of determining whether a given graph has an independent set of size k to that of finding a locally optimal solution to a particular BLP. Our approach is inspired by that of Ahmadi and Zhang [2022], who show that the problem of determining whether a given graph has an independent set of size k, can be reduced to that of finding a locally optimal solution to a particular quadratic optimization problem (QP). Ahmadi and Zhang exploit the Motzkin-Straus reformulation of the maximum clique problem as a QP [Motzkin and Straus, 1965]. In this paper, we also derive a BLP reformulation of the maximum clique problem. While the existence of some reformulation is fairly trivial from a complexity-theoretic perspective, non-convex continuous reformulations of combinatorial problems are often exploited for both theoretical insights and algorithmic design. Some notable examples include the approaches for finding maximum cliques [Abello et al., 2001, Busygin, 2006] and related clique relaxation problems [Stozhkov et al., 2022], which rely on the Motzkin-Straus QP and its generalizations. Thus, from our perspective, we view a new reformulation as noteworthy in its own right.

The remainder of this paper is organized as follows. In Section 2, we provide the necessary preliminary results along with some related observations. In particular, in Section 2.1, we describe the celebrated Motzkin-Straus QP reformulation for solving the maximum clique problem. Next, in Section 2.2, we briefly summarize how this reformulation is employed to arrive at the aforementioned complexity result for QPs by Ahmadi and Zhang [2022]. Following that, in Section 2.3 we provide our BLP reformulation of Motzkin-Straus QP, which also highlights some essential technical results. Section 3 then contains the proof of the main result given by Theorem 1. Finally, in Section 4 we conclude the paper with some remarks and observations.

Additional notation. For any $n \in \mathbb{Z}_{>0}$, let $[n] := \{1, \dots, n\}$. Denote by $e = (1, \dots, 1)^\top$ a vector of all ones, and by $e_i = (0, \dots, 0, 1, 0, \dots, 0)^\top$ the i^{th} unit vector. Let I and J denote the identity matrix and the matrix of all ones, respectively. For a symmetric matrix $M \in \mathbb{R}^{n \times n}$, we say that M is *copositive* if $x^\top M x \geq 0$ for any $x \geq 0$. In the remainder of the paper, all vectors and matrices are assumed to be of conformable dimension if not specified. Finally, throughout the paper an *optimal solution* is assumed to be globally optimal, while locally optimal solutions are always designated as such.

2 Preliminaries and related observations

2.1 Motzkin-Straus QP

Given a simple undirected unweighted graph $G = (V, E)$, a *clique* is a subset of vertices $S \subseteq V$ such that the subgraph induced by S in G is complete. Finding a clique of maximum cardinality is one of the classical combinatorial optimization problems [Pardalos and Xue, 1994]. This problem is known to be NP-hard, while its decision version is NP-complete [Garey and Johnson, 1979]. The cardinality of the maximum clique in G, also known as the *clique number* of G, is denoted by $\omega(G)$.

Let $A \in \mathbb{B}^{n\times n}$ be the *adjacency matrix* of a graph $G = (V, E)$ with $n := |V|$. That is, A is a symmetric matrix, where $a_{ij} = 1$ if $\{i, j\} \in E$, and $a_{ij} = 0$, otherwise, for all $i, j \in [n]$. Then, Motzkin and Straus showed that the following QP yields a reformulation of the maximum clique problem:

$$z^*_{\mathrm{QP}} := \max\ \{x^\top Ax \mid e^\top x = 1,\ x \in \mathbb{R}^n_+\} \qquad (\mathbf{MS^{QP}})$$

and the formal statement of the corresponding result is as follows.

Theorem 2 (Motzkin and Straus [1965])**.** *Let G be a graph on n vertices with adjacency matrix A and a clique number $\omega(G)$. The optimal objective function value of* $(\mathbf{MS^{QP}})$ *is* $z^*_{\mathrm{QP}} = 1 - \frac{1}{\omega(G)}$.

The QP model given by $(\mathbf{MS^{QP}})$ is a well-known non-convex continuous reformulation of a difficult combinatorial optimization problem [Abello et al., 2001]. We also refer to De Santis and Rinaldi [2012] and the references therein for the related discussion and other examples.

2.2 Ahmadi-Zhang result for QP

Given graph $G = (V, E)$, denote its *complement* as $\bar{G} = (V, \bar{E})$, where $\{i, j\} \in \bar{E}$ if and only if $\{i, j\} \notin E$. An *independent set* S of graph G is a subset $S \subseteq V$ of the vertices such that for any pair $i, j \in S$, $i \neq j$, we have that $\{i, j\} \notin E$. That is, the subgraph induced by S in G is edgeless. The cardinality of the largest independent set in G, also known as the *independence number* of G, is denoted by $\alpha(G)$. Clearly, $\alpha(G) = \omega(\bar{G})$, which means that the *maximum independent set problem* is also NP-hard, while its decision version is NP-complete [Garey and Johnson, 1979].

Given a scalar k and a symmetric matrix $A \in \mathbb{B}^{n\times n}$, let

$$M_{A,k} := kA + kI - J.$$

Then, the following result reduces the problem of determining whether G does not contain an independent set of size at least $k+1$ to that of determining whether $M_{A,k}$ is copositive.

Proposition 1 (Ahmadi and Zhang [2022]). *For a scalar $k > 0$ and a graph G with adjacency matrix A, matrix $M_{A,k} := kA + kI - J$ is copositive if and only if $\alpha(G) \leq k$.*

The proof of Proposition 1 exploits Theorem 2 and this proposition is, in turn, one of the key components in establishing the main result by Ahmadi and Zhang [2022]. It has long been known that finding a global minimizer of $x^\top Qx$ over a polyhedral set is an NP-hard problem, even when Q has only a single negative eigenvalue; see Pardalos and Vavasis [1991]. Also, Pardalos and Schnitger [1988] show that even the problem of checking whether a given point is a local minimizer of a quadratic function over a polyhedral set is NP-hard. As a side note, the aforementioned studies by Marcotte and Savard [2005] and Vicente et al. [1994], which establish an analogous result for BLPs, rely on closely related ideas. However, the computational complexity of finding a locally optimal solution to a nonconvex QP, posed by Pardalos and Vavasis [1992] as one of seven open questions regarding complexity in optimization in 1992, remained open until it was resolved by the aforementioned result by Ahmadi and Zhang [2022], stated formally as follows.

Theorem 3 (Ahmadi and Zhang [2022]). *If there is a polynomial-time algorithm that finds a point within Euclidean distance c^n (for any constant $c \geq 0$) of a local minimizer of an n-variate quadratic function over a polytope, then $P = NP$.*

This result is proved by exploiting Proposition 1 and the QP of the form:

$$\min\{x^\top M_{A,k} x \mid e^\top x \leq \lceil 3c^n \sqrt{n} \rceil,\ x \in \mathbb{R}^n_+\}. \tag{5}$$

There are clear similarities between the Motzkin-Straus QP formulation ($\mathbf{MS^{QP}}$) and the QP model (5). The latter provides a reduction from the decision version of the maximum clique problem by showing that $\alpha(G) \leq k$ if and only if the optimal objective function value of (5) is zero. Indeed, the feasible region of (5) is a 1-norm ball with radius $3c^n\sqrt{n}$. The proof is to first observe that when the optimal value of (5) is zero, zero is also the only local optimizer. Otherwise, the ball constraint in (5) must be binding for all locally optimal solutions. Indeed, without this constraint, there are no locally optimal solutions and the problem is unbounded. By enclosing the feasible region in a ball, the unbounded case instead becomes the case of a locally optimal solution on the bounding ball. Thus, the problem is reduced to simply determining which of the two aforementioned scenarios holds. The constant $3c^n\sqrt{n}$ ensures that no point can be within Euclidean distance c^n of both the origin and a point on the boundary of the feasible region. Hence, knowledge of any point that is within a Euclidean ball of radius c^n of any locally optimal solution is sufficient to decide whether or not $\alpha(G) \le k$.

Our main result depends on a reformulation of the maximum clique problem as a BLP that is analogous to the Motzkin-Straus reformulation of maximum clique as a QP. A slightly modified version of this reformulation is then exploited to prove Theorem 1, our main result, in Section 3. Specifically, the proposed approach uses a BLP analogous to (5) and bears a similar relation to the BLP reformulation of Motzkin-Straus QP as (5) does to ($\mathbf{MS^{QP}}$).

2.3 BLP reformulation of Motzkin-Straus QP

The Karush-Kuhn-Tucker (KKT) optimality conditions for ($\mathbf{MS^{QP}}$) are given by:

$$-2Ax - \pi + e\lambda = 0, \tag{6a}$$

$$e^\top x = 1, \tag{6b}$$

$$x^\top \pi = 0, \tag{6c}$$

$$x \ge 0,\ \pi \ge 0, \tag{6d}$$

and we refer to any $(x, \pi, \lambda) \in \mathbb{R}^{2n+1}$, which satisfies (6), as a KKT point of ($\mathbf{MS^{QP}}$).

For a given graph $G = (V, E)$, with its adjacency matrix $A \in \mathbb{B}^{n\times n}$, consider the following BLP:

$$
\begin{aligned}
z^*_{\mathrm{BP}} := \max_{(x,q,s),t} \quad & s - \mu e^\top t && (\mathbf{MS^{BP}}\text{-a})\\
\text{s.t.} \quad & es = Ax + q, && (\mathbf{MS^{BP}}\text{-b})\\
& e^\top x = 1, && (\mathbf{MS^{BP}}\text{-c})\\
& q \le e, && (\mathbf{MS^{BP}}\text{-d})\\
& x, q \in \mathbb{R}^n_+,\ s \in \mathbb{R}_+, && (\mathbf{MS^{BP}}\text{-e})\\
& t \in \operatorname*{argmax}_{\hat t} \quad e^\top \hat t && (\mathbf{MS^{BP}}\text{-f})\\
& \qquad \text{s.t.} \qquad \hat t \le x, \quad \hat t \le q, \quad \hat t \in \mathbb{R}^n_+, && (\mathbf{MS^{BP}}\text{-g})
\end{aligned}
$$

where μ is a sufficiently large "penalty" parameter that ensures that $t = 0$ in any optimal solution; see Proposition 3 below. Note that assumptions **A1** and **A2** from Section 1 hold for ($\mathbf{MS^{BP}}$).

Before formally showing that ($\mathbf{MS^{BP}}$) is a reformulation of ($\mathbf{MS^{QP}}$), we first briefly summarize the steps in the proof to highlight the underlying intuition. The main idea is that the constraints and the objective function of ($\mathbf{MS^{BP}}$) ensure that the optimal values of the leader's variables

(x, q, s) correspond to a KKT point of ($\mathbf{MS^{QP}}$). Indeed, by setting $q = \pi/2$ and $s = \lambda/2$, we see that (6a) and (6b) correspond to ($\mathbf{MS^{BP}}$-b) and ($\mathbf{MS^{BP}}$-c), respectively. Then, the follower's problem enforces $t_i = \min\{x_i, q_i\}$ for all $i \in [n]$, which implies, due to ($\mathbf{MS^{BP}}$-a), that if the penalty parameter μ is sufficiently large, then t must take value zero, i.e., x and q are complementary, which corresponds to (6c). Finally, ($\mathbf{MS^{BP}}$-b) and ($\mathbf{MS^{BP}}$-c) ensure that $s = x^\top Ax + x^\top q$. As such, for an optimal solution (x^*, q^*, s^*, t^*) to ($\mathbf{MS^{BP}}$), we have that:

$$s^* - \mu e^\top t^* = {x^*}^\top Ax^* + {x^*}^\top q^* - \mu e^\top t^* = {x^*}^\top Ax^*,$$

and hence, the optimal solutions to ($\mathbf{MS^{BP}}$) correspond to the KKT points of ($\mathbf{MS^{QP}}$).

Next, we present the full details of the proof, following the above outline. The proof is broken into three simpler results (Lemma 1, Propositions 2 and 3) to make it easier to digest. We first ensure that there always exists a solution to ($\mathbf{MS^{BP}}$) in which the corresponding leader's objective function is non-negative, while the positive components of x and q, are bounded away from zero.

Lemma 1. *Problem* ($\mathbf{MS^{BP}}$) *has a feasible solution* $(\hat{x}, \hat{q}, \hat{s}, \hat{t})$ *such that* $\hat{s} - \mu e^\top \hat{t} \geq 0$.

Proof. We construct a feasible solution for ($\mathbf{MS^{BP}}$) as follows. Consider any clique $S \subseteq V$ in G. That is, $1 \leq |S| \leq \omega(G)$. Let $(\hat{x}, \hat{q}, \hat{s}, \hat{t})$ be as follows:

$$\hat{x}_i = \begin{cases} |S|^{-1} & \text{if } i \in S, \\ 0 & \text{otherwise,} \end{cases} \qquad \hat{q}_i = \begin{cases} 0 & \text{if } i \in S, \\ \hat{s} - (A\hat{x})_i & \text{otherwise,} \end{cases}$$

$$\hat{s} = 1 - |S|^{-1} \geq 0, \qquad \hat{t}_i = 0 \quad \forall i \in [n].$$

Then, ($\mathbf{MS^{BP}}$-c) is clearly satisfied for $\hat{x}$. Also, for $i \in S$, we have $(A\hat{x})_i + \hat{q}_i = (A\hat{x})_i = (|S| - 1)|S|^{-1} = \hat{s}$. On the other hand, for $i \notin S$, we have $(A\hat{x})_i = \sum_{j \in S \setminus \{i\}} a_{ij}\hat{x}_j \leq (|S| - 1)|S|^{-1} = \hat{s}$, which means that $0 \leq \hat{q}_i = \hat{s} - (A\hat{x})_i \leq 1$. Consequently, ($\mathbf{MS^{BP}}$-b)-($\mathbf{MS^{BP}}$-e) are all satisfied for $(\hat{x}, \hat{q}, \hat{s})$. Finally, observe from ($\mathbf{MS^{BP}}$-g) that the corresponding follower's decision $\hat{t} = 0$ is optimal for the follower's problem since either $\hat{x}_i = 0$ if $i \notin S$, or $\hat{q}_i = 0$ if $i \in S$. □

Proposition 2. *Let* $\Delta := (4n)^{-2n}$*. Problem* ($\mathbf{MS^{BP}}$) *has an optimal solution* (x^*, q^*, s^*, t^*) *such that for all* $i \in [n]$ *either* $x_i^* = 0$ *or* $x_i^* \geq \Delta$ *and, similarly, for all* $i \in [n]$ *either* $q_i^* = 0$ *or* $q_i^* \geq \Delta$.

Proof. Given $(x, q, t) \in \mathbb{R}_+^{3n}$, t is optimal for the lower-level LP ($\mathbf{MS^{BP}}$-f)–($\mathbf{MS^{BP}}$-g) if and only if for some $Q \subseteq [n]$, $t_i = x_i$ for all $i \in Q$ and $t_i = q_i$ for all $i \in [n] \setminus Q$. As such, given $Q \subseteq [n]$, denote by $P(Q) \subseteq \mathbb{R}_+^{3n+1}$ a polytope that is described by constraints ($\mathbf{MS^{BP}}$-b)-($\mathbf{MS^{BP}}$-d) along with $t_i = x_i$ for all $i \in Q$, $t_i = q_i$ for all $i \in [n] \setminus Q$, and the non-negativity constraints for all variables, including s. Then, every feasible solution to ($\mathbf{MS^{BP}}$) must be contained in $P(Q)$ for some $Q \subseteq [n]$.

For $Q \subseteq [n]$, $P(Q)$ is non-empty and bounded, since $x \in [0,1]^n$ from ($\mathbf{MS^{BP}}$-c) and ($\mathbf{MS^{BP}}$-e), and $q \in [0,1]^n$ from ($\mathbf{MS^{BP}}$-d) and ($\mathbf{MS^{BP}}$-e). Furthermore, recall that each element of A is either 0 or 1; hence, $s \in [0,2]$. Also, t_i is equal to either x_i or q_i for all $i \in [n]$; thus, $t \in [0,1]^n$. Consequently, all such LPs have a finite optimal solution, which implies that ($\mathbf{MS^{BP}}$) has a finite optimal solution.

Given any matrix $B \in \mathbb{R}^{m \times m}$ for which $|b_{ij}| \leq 1$ for $i,j \in [m]$, $|det(B)| \leq m^{m/2}$ by Hadamard's Inequality [Brenner and Cummings, 1972]. For each of the aforementioned LPs, a basic feasible solution (including an optimal one) can be found by solving a system of linear equations with $m := 3n+1$ equality constraints and $m := 3n+1$ variables. Moreover, from the constraint set in ($\mathbf{MS^{BP}}$-b)-($\mathbf{MS^{BP}}$-g) and the above discussion, it is clear that each element of the corresponding square matrix is in $\{-1, 0, 1\}$. Hence, from Cramer's rule, we conclude that for $i \in [n]$, x_i is either 0 or at least

$$1/det(B) \geq 1/(3n+1)^{(3n+1)/2} \geq (4n)^{-2n},$$

where the second inequality holds because $n \geq 1$. Finally, we note that the same observation holds for vector q. That is, for $i \in [n]$, q_i is either 0 or at least $1/det(B) \geq (4n)^{-2n}$. □

Next, we show that if the penalty parameter μ is sufficiently large, then x^* and q^* considered in the discussion above, must be complementary.

Proposition 3. *Let $\mu \geq (4n)^{2n}+3$. Then, there exists an optimal solution (x^*, q^*, s^*, t^*) of* ($\mathbf{MS^{BP}}$) *such that ${x^*}^\top q^* = 0$, $t^* = 0$ and $s^* - \mu e^\top t^* = {x^*}^\top A x^*$.*

Proof. Assume that ${x^*}^\top q^* > 0$. First, note that by left-multiplying ($\mathbf{MS^{BP}}$-b) by ${x^*}^\top$, we obtain:

$${x^*}^\top e s^* = {x^*}^\top A x^* + {x^*}^\top q^*,$$

which, by taking into account ($\mathbf{MS^{BP}}$-c), reduces to:

$$s^* = {x^*}^\top A x^* + {x^*}^\top q^*, \tag{7}$$

and the corresponding leader's objective function is given by:

$$s^* - \mu e^\top t^* = {x^*}^\top A x^* + {x^*}^\top q^* - \mu e^\top t^*. \tag{8}$$

From ($\mathbf{MS^{BP}}$-c)–($\mathbf{MS^{BP}}$-e), we have that both $x^*, q^* \in [0,1]^n$ and thus, $\min\{x_i^*, q_i^*\} \geq x_i^* \cdot q_i^*$ for every $i \in [n]$. By the initial assumption, there exists $j \in [n]$ such that $x_j^* \cdot q_j^* > 0$. Then:

$$\mu e^\top t^* - {x^*}^\top q^* = \sum_{i=1}^n \left(\mu \min\{x_i^*, q_i^*\} - x_i^* \cdot q_i^*\right) \tag{9a}$$

$$\geq \mu \min\{x_j^*, q_j^*\} - x_j^* \cdot q_j^* \geq (\mu - 1) \cdot \min\{x_j^*, q_j^*\} \geq (\mu - 1) \cdot (4n)^{-2n}, \tag{9b}$$

where the equality in (9a) follows from ($\mathbf{MS^{BP}}$-f)–($\mathbf{MS^{BP}}$-g), and the last inequality in (9b) follows from Proposition 2. Next, combining (8) with (9), we obtain that:

$$\begin{aligned} s^* - \mu e^\top t^* &= {x^*}^\top A x^* + {x^*}^\top q^* - \mu e^\top t^* \\ &\leq {x^*}^\top A x^* - (\mu - 1) \cdot (4n)^{-2n} \leq {x^*}^\top J x^* - (\mu - 1) \cdot (4n)^{-2n} \leq 1 - \frac{(4n)^{2n} + 2}{(4n)^{2n}} < 0, \end{aligned} \tag{10}$$

where the second inequality follows because $A \leq J = ee^\top$ and ($\mathbf{MS^{BP}}$-c) holds for x^*; the third inequality follows from $\mu \geq (4n)^{2n} + 3$. From Lemma 1, the optimal objective function value of ($\mathbf{MS^{BP}}$) is non-negative. Hence, (10) results in a contradiction, which implies that

$${x^*}^\top q^* = 0 \tag{11}$$

for any optimal solution of ($\mathbf{MS^{BP}}$).

Furthermore, combining (11) with ($\mathbf{MS^{BP}}$-g), we conclude that $t^* = 0$. Finally, from the latter and (8) we also observe that $s^* - \mu e^\top t^* = {x^*}^\top A x^*$, which completes the proof. □

We are ready to establish the main result of this section, which is analogous to Theorem 2 and shows that ($\mathbf{MS^{BP}}$) is a reformulation of the Motzkin-Straus QP model ($\mathbf{MS^{QP}}$).

Theorem 4. *Let G be a graph on n vertices with adjacency matrix A and a clique number $\omega(G)$. If $\mu \geq (4n)^{2n} + 3$, then the optimal objective function value of* ($\mathbf{MS^{BP}}$) *is $z^*_{\mathrm{BP}} = 1 - \frac{1}{\omega(G)}$.*

Proof. Let $\mu \geq (4n)^{2n} + 3$. Using Proposition 3, we can rewrite ($\mathbf{MS^{BP}}$) as the following QP:

$$\max_{x,q,s} \quad x^\top A x \tag{12a}$$

$$\text{s.t.} \quad Ax + q - es = 0, \tag{12b}$$

$$e^\top x = 1, \tag{12c}$$

$$q \leq e, \tag{12d}$$

$$x^\top q = 0, \tag{12e}$$

$$x \geq 0, \; q \geq 0, \; s \geq 0. \tag{12f}$$

By setting $\pi = 2q$ and $\lambda = 2s$ in (6), and then comparing (6) with (12b)-(12f), we observe that (12) is nothing more than ($\mathbf{MS^{QP}}$) augmented with the associated KKT conditions (12b) and (12e), as well as an upper bound on q in (12d) and non-negativity of s in (12f).

The addition of the necessary optimality (KKT) conditions to a QP cannot eliminate any optimal solution. Furthermore, given (x^*, π^*, λ^*), an optimal KKT solution to ($\mathbf{MS^{QP}}$), we have that:

$$\frac{1}{2}\pi^* = q^* \leq es^* = ex^{*\top} A x^* \leq ex^{*\top} ee^\top x^* \leq e,$$

where the first inequality follows from (12b), the second equality follows from (7) and (12e), the second inequality follows from $A \leq J = ee^\top$, and the last inequality follows from (12c). Hence, adding (12d) cannot eliminate any optimal solution of ($\mathbf{MS^{QP}}$). Finally, non-negativity of s follows from (12b) and non-negativity of A, x and q. In view of the above discussion, the required result follows and we conclude that the optimal objective function values of ($\mathbf{MS^{QP}}$) and ($\mathbf{MS^{BP}}$) coincide. □

In conclusion, we point out the following two observations. First, the value of μ that is sufficiently large in our derivations, is $(4n)^{2n} + 3$, which has a bit representation of polynomial size, since $\log[(4n)^{2n} + 3] = O(n \log n)$. Hence, the outlined derivations provide yet another proof that solving BLP is a strongly NP-hard problem. The other observation is concerned with the fact that (11) holds whenever μ is sufficiently large. That is, "favorable" decisions for the leader in the outlined BLP correspond to the KKT points of ($\mathbf{MS^{QP}}$) and hence, the underlying bilevel optimization problem can be interpreted as the combinatorial problem of selecting the best KKT points of ($\mathbf{MS^{QP}}$). A similar in spirit idea is used in the literature for reducing non-convex QPs to MILPs; see, e.g., Xia et al. [2020].

3 Proof of the main result

The proof of Theorem 1 combines the ideas used in proving Theorem 3 by Ahmadi and Zhang [2022] and Theorem 4 in Section 2.3. Note that the proof by Ahmadi and Zhang [2022] relies on the complexity of finding a local minimizer for a degree-4 polynomial, which, in turn, exploits a Hessian of the corresponding quartic form. In contrast, our proof requires a particular constraint structure of the underlying BLP, along with Proposition 1 and some additional technical observations.

Formally, given some positive $k \in \mathbb{Q}_{>0}$ and $c \in \mathbb{Q}_{>0}$, we consider a BLP of the form:

$$
\begin{aligned}
z^*_{k,c} := \max_{(x,q,s,h),t} \quad & s - \frac{1}{n\mu_{k,c}} \cdot h - \mu_{k,c} \cdot e^\top t && \text{(13a)}\\
\text{s.t.} \quad & es = -(kA + kI - J)x + q, && \text{(13b)}\\
& e^\top x = \lceil 3c^n \sqrt{n} \rceil h, && \text{(13c)}\\
& h \leq 1, && \text{(13d)}\\
& q \leq \big(2k \lceil 3c^n \sqrt{n} \rceil e\big) h, && \text{(13e)}\\
& x, q \in \mathbb{R}^n_+, \ \ h, s \in \mathbb{R}_+, && \text{(13f)}\\
& t \in \operatorname*{argmax}_{\hat{t}} \quad e^\top \hat{t} && \text{(13g)}\\
& \qquad \text{s.t.} \qquad \hat{t} \leq x, \quad \hat{t} \leq q, \quad \hat{t} \in \mathbb{R}^n_+, && \text{(13h)}
\end{aligned}
$$

where $\mu_{k,c}$ is a positive parameter that depends on k and c; below, we specify one particular value for $\mu_{k,c}$, which is of polynomial size with respect to the problem's parameters. The leader's variables are $(x, q, s, h) \in \mathbb{R}^{2n+2}$, and the follower's variables are given by $t \in \mathbb{R}^n$.

Two comments regarding this bilevel formulation are in order. First, there are clear similarities with (5), but also some differences. We introduce a variable h in the right-hand side of (13c) to make the proof cleaner. Substituting out h using (13c), one can observe that (13d) then becomes identical to the ball constraint in (5). The role of this constraint is the same here as it is in (5), to constrain x to be in a 1-norm ball, so that locally optimal solutions lie either at the origin or on the boundary of the ball. Without this constraint, the problem could simply have no locally optimal solutions. Although the same conclusions would ultimately be reached, dealing with the possibility of unboundedness would make the proof more difficult and we choose to impose the bounding ball, as Ahmadi and Zhang [2022] similarly did in their work. Our second comment is about the constant parameter $\lceil 3c^n \sqrt{n} \rceil$, which determines the radius of the bounding ball. This constant could have a simpler form by using different norms, which would simplify some of the equations to follow. However, we keep it as is to stay as close to the approach taken by Ahmadi and Zhang [2022], as possible.

The proof of our main result, Theorem 1, is performed in the following three main steps:

(i) First, we show that if parameter $\mu_{k,c}$ in (13) is sufficiently large, then for all locally optimal solutions $(\hat{x}, \hat{q}, \hat{s}, \hat{h})$ to (13), $\hat{x}$ and $\hat{q}$ must be complementary; see Proposition 4 below.

(ii) Next, in Proposition 5, we show that for all locally optimal solutions $(\hat{x}, \hat{q}, \hat{s}, \hat{h})$ to (13), we must have that $\hat{h} \in \{0, 1\}$. Given (13c), it implies that either $\hat{x} = 0$ or $e^\top \hat{x} = \lceil 3c^n \sqrt{n} \rceil$.

(iii) Finally, in Propositions 6 and 7, we show that deciding whether $\alpha(G) < k$, where $k \notin \mathbb{Z}$, is equivalent to determining whether leader's decision $(\hat{x}, \hat{q}, \hat{s}, \hat{h}) = (0, 0, 0, 0)$ is locally optimal. Because the only other possible locally optimal solutions are outside a Euclidean ball around the origin of radius c^n, constructing any point within Euclidean distance c^n of any locally optimal solution is sufficient to decide whether $\alpha(G) < k$.

Each of the above steps requires additional technical lemmas along with some assumptions outlined next. Specifically, in the remainder of this section, we assume that for some integer $r \geq 1$

$$k := r + 0.5, \tag{14}$$

which implies that k is not integer, $k > 1$ and $2k$ is a positive integer. Furthermore, throughout this section we also define the following parameters, which have bit representation of polynomial size:

$$\begin{aligned}
&\sigma_{k,c} := (2k+1) \cdot \lceil 3c^n \sqrt{n} \rceil, \\
&\Delta_{k,c} := (2k \lceil 3c^n \sqrt{n} \rceil)^{-(3n+2)} (3n+2)^{-(3n+2)/2}, \text{ and} \\
&\mu_{k,c} := 2\sigma_{k,c} (\Delta_{k,c})^{-1}.
\end{aligned} \tag{15}$$

We first establish the existence of a finite optimal solution for (13). Then, we derive lower and upper bounds on the leader's objective function values for globally and locally optimal solutions.

Lemma 2. *Problem* (13) *has an optimal solution* $(x^*, q^*, s^*, h^*, t^*)$ *such that*

$$\sigma_{k,c} \geq z^*_{k,c} = s^* - \frac{1}{n\mu_{k,c}} \cdot h^* - \mu_{k,c} \cdot e^\top t^* \geq 0. \tag{16}$$

Proof. One can verify that setting all variables to zero is feasible; hence, $z^*_{k,c} \geq 0$. From (13b), we observe that for any (x, q, s, h, t) feasible for (13), we have

$$s \leq \max_{i \in [n]} (Jx + q)_i \leq (2k+1) \cdot \lceil 3c^n \sqrt{n} \rceil = \sigma_{k,c}, \tag{17}$$

where the second inequality follows from (13c)–(13e). Recall that h and t are non-negative; hence, the objective function value of any feasible solution of (13) is upper-bounded by $\sigma_{k,c}$.

While the objective function value of any feasible solution satisfies the bounds in (16), however, it is not sufficient to show that the optimal value is both finite *and* attained. Given (17) along with (13c)–(13f), we conclude that the upper-level feasible set is bounded. Also, for any leader's decision (x, q, s, h), the follower's feasible set is bounded. These observations are sufficient to ensure that the BLP given by (13) has a finite optimal value that is attained; recall assumptions **A1** and **A2**. □

Lemma 3. *If $(\hat{x}, \hat{q}, \hat{s}, \hat{h})$ is a locally optimal solution to* (13)*, then the leader's objective function value is non-negative, i.e.,*

$$\hat{s} - \frac{1}{n\mu_{k,c}}\hat{h} - \mu_{k,c}e^\top \hat{t} \geq 0, \tag{18}$$

where $\hat{t}$ is the corresponding follower's solution.

Proof. For any $\delta \in [0, 1)$, one can verify that the vector $(1-\delta) \cdot (\hat{x}, \hat{q}, \hat{s}, \hat{h}, \hat{t})$ is also feasible for (13) and has the corresponding leader's objective function value given by

$$(1-\delta) \cdot (\hat{s} - \frac{1}{n\mu_{k,c}}\hat{h} - \mu_{k,c}e^\top \hat{t}).$$

Recall our assumption that $(\hat{x}, \hat{q}, \hat{s}, \hat{h})$ is locally optimal. Thus, for sufficiently small $\delta \in (0, 1)$:

$$(1-\delta) \cdot (\hat{s} - \mu_{k,c}e^\top \hat{t} + \frac{1}{n\mu_{k,c}}\hat{h}) \leq \hat{s} - \mu_{k,c}e^\top \hat{t} + \frac{1}{n\mu_{k,c}}\hat{h},$$

which can hold only if (18) is satisfied. $\square$

The next result is analogous to Proposition 2. However, instead of globally optimal solutions, it is concerned with locally optimal ones. Also, recall that $\mu_{k,c}$ is set to be sufficiently large; see (15).

Proposition 4. *If $(\hat{x}, \hat{q}, \hat{s}, \hat{h})$ is a locally optimal solution of* (13)*, then*

$$\hat{x}^\top \hat{q} = 0, \tag{19}$$

and the corresponding follower's solution is $\hat{t} = 0$. Furthermore, if $\hat{s} > 0$, then

$$\hat{s} > \frac{1}{n\mu_{k,c}}\hat{h}. \tag{20}$$

Proof. As in the proof of Proposition 2, we first note that given $(x, q, t) \in \mathbb{R}^{3n}_+$, t is optimal for the lower-level LP (13g)– (13h) if and only if for some $Q \subseteq [n]$, $t_i = x_i$ for all $i \in Q$ and $t_i = q_i$ for all $i \in [n] \setminus Q$. As such, given $Q \subseteq [n]$, denote by $P(Q) \subseteq \mathbb{R}^{3n+2}_+$ a polytope that is described by constraints (13b)-(13e) along with $t_i = x_i$ for all $i \in Q$, $t_i = q_i$ for all $i \in [n] \setminus Q$, and the non-negativity constraints for all variables, including h and s. Then, every feasible solution to (13) must be contained in $P(Q)$ for some $Q \subseteq [n]$.

Now, let $(\hat{x}, \hat{q}, \hat{s}, \hat{h})$ be locally optimal for (13) and let $\hat{Q} \subseteq [n]$ be such that $(\hat{x}, \hat{q}, \hat{s}, \hat{h}, \hat{t}) \in P(\hat{Q})$. There are two cases to be considered.

Case 1: $(\hat{x}, \hat{q}, \hat{s}, \hat{h}, \hat{t})$ is a vertex of $P(\hat{Q})$. This case is similar to Proposition 2. Specifically, recall that for any matrix $B \in \mathbb{R}^{m \times m}$ for which $|b_{ij}| \leq \beta$ for $i, j \in [n]$, we have that $|det(B)| \leq \beta^m m^{m/2}$ by

Hadamard's inequality [Brenner and Cummings, 1972]. A vertex of $P(\hat{Q})$ can be found by solving a system of linear equations with $m := 3n+2$ equality constraints and $m := 3n+2$ variables.

If we multiply both sides of constraint (13b) by 2, then by our choice of k as in (14), all coefficients in the constraints that define $P(\hat{Q})$ are integer. Moreover, from the constraints (13b)-(13e) and the above discussion, it is clear that each element of the corresponding square matrix takes values in $\{\pm 2k\lceil 3c^n\sqrt{n}\rceil, \pm\lceil 3c^n\sqrt{n}\rceil, \pm 1, 0, \pm 2(k-1)\}$, which are all integers.

Hence, from Cramer's rule, $\hat{x}_i$ is either 0, or at least

$$\frac{1}{det(B)} \geq \Delta_{k,c},$$

for all $i \in [n]$, where $\Delta_{k,c}$ is defined in (15). Similarly, $\hat{q}_i$ is either 0 or at least $\Delta_{k,c}$ for all $i \in [n]$. Furthermore, $\hat{t}_i$ is also either 0 or at least $\Delta_{k,c}$ for all $i \in [n]$. Next, consider the corresponding leader's objective function value, for which we have that

$$0 \leq \hat{s} - \frac{1}{n\mu_{k,c}}\hat{h} - \mu_{k,c}e^\top\hat{t} \leq \sigma_{k,c} - \mu_{k,c}e^\top\hat{t}, \tag{21}$$

where the first inequality follows from Lemma 3 and the second inequality follows from (17) in Lemma 2, and the non-negativity of $\hat{h}$. Then, we have that

$$e^\top\hat{t} \leq \frac{\sigma_{k,c}}{\mu_{k.c}} = \frac{\Delta_{k,c}}{2}, \tag{22}$$

which implies that $e^\top\hat{t} = 0$, since, otherwise, we must have $e^\top\hat{t} \geq \Delta_{k,c}$. Thus, (19) holds.

To establish (20), using the arguments as in the discussion above, we observe that if $\hat{s} > 0$, then $\hat{s} \geq \Delta_{k,c} > 0$. Recall that by its definition, $\sigma_{k,c} \geq 1$. Then, we have that

$$\hat{s} \geq \Delta_{k,c} > \frac{1}{2\sigma_{k,c}}\Delta_{k,c} \geq \frac{1}{\mu_{k,c}} \geq \frac{\hat{h}}{n\mu_{k,c}},$$

as $0 \leq \hat{h} \leq 1$ and $n \geq 1$. Thus, (20) holds.

Case 2: $(\hat{x}, \hat{q}, \hat{s}, \hat{h}, \hat{t})$ is not a vertex of $P(\hat{Q})$. In this case, $(\hat{x}, \hat{q}, \hat{s}, \hat{h}, \hat{t})$ should be a convex combination of at least two vertices of $P(\hat{Q})$. Without loss of generality, assume that $(\hat{x}, \hat{q}, \hat{s}, \hat{h}, \hat{t}) = \alpha v^1 + (1-\alpha)v^2$, where $v^1 := (x^1, q^1, s^1, h^1, t^1)$ and $v^2 := (x^2, q^2, s^2, h^2, t^2)$ for some $\alpha \in (0,1)$. Since $(\hat{x}, \hat{q}, \hat{s}, \hat{h})$ is locally optimal and the leader's objective function is linear, the leader's objective function values must be the same for all three solutions and all three solutions must be locally optimal. By the same logic as in the proof of Case 1, we must have $e^\top t^1 = e^\top t^2 = 0$ and so also, $e^\top\hat{t} = 0$. Thus, (19) holds. Finally, (20) can be established similarly to Case 1. $\square$

Next, we show that locally optimal solutions may occur only for $h \in \{0,1\}$. Formally:

Proposition 5. *If $(\hat{x}, \hat{q}, \hat{s}, \hat{h})$ is a locally optimal solution of* (13)*, then $\hat{h} \in \{0,1\}$.*

Proof. Assume for the sake of contradiction that $(\hat{x}, \hat{q}, \hat{s}, \hat{h})$ is locally optimal, but $0 < \hat{h} < 1$. Then, one of the following two cases holds.

Case 1: $\hat{x}^\top (kA + kI - J)\hat{x} < 0$. After left-multiplying both sides of (13b) by $\hat{x}$ we obtain that:

$$\hat{x}^\top e\hat{s} = -\hat{x}^\top (kA + kI - J)\hat{x} + \hat{x}^\top \hat{q}, \tag{23}$$

which, in view of (13c) and (19), reduces to

$$\lceil 3c^n \sqrt{n} \rceil \hat{h}\hat{s} = -\hat{x}^\top (kA + kI - J)\hat{x}, \tag{24}$$

which, in turn, implies that $\hat{s} > 0$.

Let $(\tilde{x}, \tilde{q}, \tilde{s}, \tilde{h}) := (1 + \epsilon)(\hat{x}, \hat{q}, \hat{s}, \hat{h})$, where $\epsilon > 0$ is chosen so that $\epsilon < (1 - \hat{h})/\hat{h}$ and thus, $0 < \tilde{h} < 1$. Then, it is easy to verify that (13b)–(13e) hold. Furthermore, $\tilde{x}^\top \tilde{q} = (1 + \epsilon)^2 \hat{x}^\top \hat{q} = 0$, which implies that $\tilde{t} = 0$ forms the corresponding follower's decision.

By our assumption that the leader's decision $(\hat{x}, \hat{q}, \hat{s}, \hat{h})$ is locally optimal and the fact that $\hat{s} > 0$ (as shown above), we have from Proposition 4, see (20), that:

$$\epsilon\big(\hat{s} - \frac{1}{n\mu_{k,c}}\hat{h}\big) > 0,$$

which, in turn, implies that:

$$(1 + \epsilon)\hat{s} - (1 + \epsilon)\frac{1}{n\mu_{k,c}}\hat{h} > \hat{s} - \frac{1}{n\mu_{k,c}}\hat{h},$$

and hence, $(\tilde{x}, \tilde{q}, \tilde{s}, \tilde{h})$ provides a strictly better objective function value than $(\hat{x}, \hat{q}, \hat{s}, \hat{h})$. Finally, $(\tilde{x}, \tilde{q}, \tilde{s}, \tilde{h})$ can be made arbitrarily close to $(\hat{x}, \hat{q}, \hat{s}, \hat{h})$ by choosing ϵ sufficiently small. This observation contradicts our assumption that $(\hat{x}, \hat{q}, \hat{s}, \hat{h})$ is locally optimal.

Case 2: $\hat{x}^\top (kA + kI - J)\hat{x} \geq 0$. In this case, using (23) and (24), we conclude that $\hat{h}\hat{s} = \hat{x}^\top (kA + kI - J)\hat{x} = 0$. Recall that $\hat{h} > 0$; hence, $\hat{s} = 0$. Also, $\hat{x}^\top \hat{q} = 0$ by Proposition 4. Thus, the corresponding leader's objective function is equal to

$$-\frac{1}{n\mu_{k,c}}\hat{h} < 0,$$

which, in view of, Lemma 3, implies that $(\hat{x}, \hat{q}, \hat{s}, \hat{h})$ is not locally optimal. Therefore, we have a contradiction in this case as well. $\square$

From Proposition 5, one can see that for a given $Q \subset [n]$, $P(Q)$ is a polyhedron with one extreme point in the origin and the other on the bounding ball. Fixing the value of h determines which of these two extreme points is a candidate for being locally optimal. Thus, the problem of determining whether $\alpha(G) < k$ amounts to enumerating the subsets of $[n]$. It is also interesting to note that choosing Q is equivalent to choosing a basis of the follower's LP. Thus, the problem of finding a locally optimal solution reduces to enumerating bases of the follower's problem; see also Appendix B. It is shown by Buchheim [2023] that the same is true for the problem of finding globally optimal solutions to general BLPs, which validates our finding that these two problems are essentially of equal difficulty from a complexity theoretic standpoint.

The next lemma is technical and assumes that $M_{A,k} := kA+kI-J$ is not copositive. The lemma is needed for exploring the local optimality conditions of the leader's decision given by $(\hat{x}, \hat{q}, \hat{s}, \hat{h}) = (0, 0, 0, 0)$. The lemma's proof relies on a QP of a particular form along with the corresponding KKT conditions for deriving the required matrix properties; this approach is similar to the one used by Hiriart-Urruty and Seeger [2010], who exploit it in the context of Pareto eigenvalues.

Lemma 4. *Let $M \in \mathbb{R}^{n\times n}$ be a symmetric matrix that is not copositive. Then, there exist $\delta \in \mathbb{R}_{>0}$ and $y \in \mathbb{R}^n_+$ such that $e^\top y = 1$ and for any $i \in [n]$, one of the two following statements holds:*

(i) $(My)_i \geq 0$ and $y_i = 0$; or

(ii) $(My)_i = -\delta < 0$,

Furthermore, if $M := M_{A,k}$ for some $k > 1$, then δ can be chosen such that $\delta \geq \Delta_{k,c}$.

Proof. Assume $M \in \mathbb{R}^{n\times n}$ is not copositive. Then, there exists $\hat{x} \in \mathbb{R}^n$ such that $\hat{x} \geq 0$ and $\hat{x}^\top M\hat{x} < 0$. Clearly, we have that $\hat{x} \neq 0$. Next, consider a QP of the form:

$$\bar{z} = \frac{1}{2} \min_y \{y^\top My \mid e^\top y = 1,\ y \in \mathbb{R}^n_+\}, \tag{25}$$

which is feasible and has a finite optimal solution as the feasible region is bounded.

Define $\hat{y} = (e^\top \hat{x})^{-1} \cdot \hat{x}$. Then, we observe that $\hat{y}$ is feasible for (25) and $\hat{y}^\top M\hat{y} < 0$. Hence, $\bar{z} < 0$. Let $\bar{y} \in \mathbb{R}^n_+$ be the corresponding optimal solution for (25). As such, it must satisfy the KKT optimality conditions. That is, there exist $\delta \in \mathbb{R}$ and $\mu \in \mathbb{R}^n_+$ such that:

$$e\delta = -M\bar{y} + \mu, \tag{26a}$$

$$e^\top \bar{y} = 1, \tag{26b}$$

$$\mu^\top \bar{y} = 0, \tag{26c}$$

and similar to the derivations above, by left-multiplying (26a) by $\bar{y}^\top$, we conclude that $\delta = -2\bar{z} > 0$. Next, we make the following two observations.

- Let $(My)_i \geq 0$. From (26a) and the fact that $\delta > 0$, we conclude that $\mu_i > 0$. Hence, from (26c) we have that $\bar{y}_i = 0$, which implies that the statement (i) holds.
- Let $0 < \bar{y}_i \leq 1$. Then, from (26c) and the fact that $\mu \geq 0$, we have that $\mu_i = 0$. Moreover, $(My)_i = -\delta < 0$ from (26a), which implies that the statement (ii) holds.

To show that we can assume w.l.o.g that $\delta \geq \Delta_{k,c}$ under the given assumptions, observe that $(\delta, \mu, \bar{y})$ is defined by a polyhedral set given by (26) and an additional constraint $\delta \geq 0$. This polyhedral set does not contain $\{0\}$ and also does not contain a line; hence, it should have at least one vertex. Next, using the arguments similar to those used in the proof of Proposition 4, by Cramer's rule we conclude that there exists a vertex of this polyhedral set, where a non-zero δ is at least $\Delta_{k,c}$. (In fact, this bound can be tightened further; however, it is not needed for our derivations below.) $\square$

In the next two propositions we provide necessary and sufficient conditions (assuming k is not integer) for local optimality of the leader's decision given by $(\hat{x}, \hat{q}, \hat{s}, \hat{h}) = (0, 0, 0, 0)$.

Proposition 6. *If $\alpha(G) < k$, then $(\hat{x}, \hat{q}, \hat{s}, \hat{h}) = (0, 0, 0, 0)$ is the only locally optimal solution of* (13)*.*

Proof. Let $\alpha(G) < k$ and let $(\hat{x}, \hat{q}, \hat{s}, \hat{h})$ be any locally optimal solution. Due to Proposition 5, we must have $\hat{h} \in \{0, 1\}$. We consider each of these cases separately. First, we show that if $\hat{h} = 0$, then we must have $(\hat{x}, \hat{q}, \hat{s}, \hat{h}) = (0, 0, 0, 0)$ and $(\hat{x}, \hat{q}, \hat{s}, \hat{h})$ is locally optimal. Afterwards we demonstrate that $\hat{h} = 1$ cannot correspond to a locally optimal solution.

Case 1: $\hat{h} = 0$. From (13c), (13e) and (13b) we conclude that $\hat{x} = \hat{q} = \hat{s} = 0$. Hence, we need only to show that $(0, 0, 0, 0)$ is locally optimal. For the sake of contradiction, suppose that it is not the case. That is, we can construct another leader's solution $(\tilde{x}, \tilde{q}, \tilde{s}, \tilde{h})$, which is arbitrarily close to $(0, 0, 0, 0)$ and provides a strictly better objective function.

Note that the leader's solution $(0, 0, 0, 0)$ has an objective function value of zero. Hence, we must show that $(\tilde{x}, \tilde{q}, \tilde{s}, \tilde{h})$ has a leader's objective function value that is strictly positive. Because $e^\top t \geq 0$ and $h \geq 0$ in (13a) for all bilevel feasible solutions, we have that

$$\tilde{s} > 0. \tag{27}$$

Similar to the discussion above, by left-multiplying (13b) by x and then using (13c) we have:

$$\lceil 3c^n\sqrt{n}\rceil hs = -x^\top(kA + kI - J)x + x^\top q, \tag{28}$$

which we exploit below. Next, as in the proof of Proposition 4, let $Q \subseteq \{1, \dots, n\}$ be such that $(\tilde{x}, \tilde{q}, \tilde{s}, \tilde{h}, \tilde{t}) \in P(Q)$. By construction $P(Q)$ is a polytope; hence, $(\tilde{x}, \tilde{q}, \tilde{s}, \tilde{h}, \tilde{t})$ can be represented as a convex combination of the vertices of $P(Q)$.

Recall from the proof of Proposition 4 that any vertex $\bar{v} := (\bar{x}, \bar{q}, \bar{s}, \bar{h}, \bar{t})$ of $P(Q)$, for which $e^\top\bar{t} > 0$ must have a strictly negative leader's objective function value (see (21) and the corresponding discussion). Hence, $(\tilde{x}, \tilde{q}, \tilde{s}, \tilde{h}, \tilde{t})$ can provide a strictly positive leader's objective function only if there exists a vertex $\bar{v} := (\bar{x}, \bar{q}, \bar{s}, \bar{h}, \bar{t})$ of $P(Q)$ such that $e^\top\bar{t} = 0$ and $\bar{s} > 0$; recall also (27). In view of (28) and Proposition 1 (recall $\alpha(G) < k$), we conclude that $\bar{h} = 0$. Then, from (13c), (13e) and (13b) we conclude that $\bar{x} = \bar{q} = \bar{s} = 0$, which results in a contradiction.

Case 2: $\hat{h} = 1$. In view of (28), Proposition 1 (recall $\alpha(G) < k$) and Proposition 4, as well as (13c), we conclude that $\hat{s} = 0$. Thus, the corresponding leader's solution is strictly negative, which is a contradiction due to Lemma 3; recall that $(\hat{x}, \hat{q}, \hat{s}, \hat{h})$ is assumed to be locally optimal. □

Proposition 7. *If $\alpha(G) > k$, then $(\hat{x}, \hat{q}, \hat{s}, \hat{h}) = (0, 0, 0, 0)$ cannot be a locally optimal solution of* (13).

Proof. The proof is to construct a leader's solution $(\tilde{x}, \tilde{q}, \tilde{s}, \tilde{h})$, which is feasible to (13), has a strictly positive leader's objective function value, and can be made arbitrarily close to zero, thereby showing that the leader's solution $(0, 0, 0, 0)$ cannot be locally optimal. From Proposition 1, we observe that matrix $M_{A,k} := kA + kI - J$ is not copositive. Hence, there exist $\delta > 0$ and $y \in \mathbb{R}^n_+ \setminus \{0\}$ such that $e^\top y = 1$ and the properties (i) and (ii) in Lemma 4 hold. Given such y, let

$$\tilde{x} := \epsilon y, \qquad \tilde{h} := \frac{1}{\lceil 3c^n\sqrt{n}\rceil}\epsilon, \qquad \tilde{s} := \epsilon\delta, \qquad \tilde{t} := 0, \tag{29}$$

where $\epsilon > 0$ is chosen so that $\epsilon < \lceil 3c^n\sqrt{n}\rceil$. Hence, $0 < \tilde{h} < 1$ and (13d) holds.

Next, given that $e^\top y = 1$, we have the following sequence of inequalities:

$$0 < e^\top\tilde{x} = \epsilon e^\top y = \lceil 3c^n\sqrt{n}\rceil \cdot \frac{1}{\lceil 3c^n\sqrt{n}\rceil}\epsilon = \lceil 3c^n\sqrt{n}\rceil \cdot \tilde{h}, \tag{30}$$

which implies that (13c) also holds for $\tilde{x}$ and $\tilde{h}$. It remains to construct $\tilde{q}$ such that $(\tilde{x}, \tilde{q}, \tilde{s}, \tilde{h})$

satisfies (13b) and (13e), and to show that $\tilde{t}$ satisfies (13g)–(13h).

First, let $\tilde{q}_i = 0$ for all $i \in [n]$ such that $0 < y_i \leq 1$, $(My)_i < 0$ and $(My)_i = -\delta < 0$ (property (ii) in Lemma 4 holds). Then:

$$\tilde{s} = -\epsilon \cdot (M_{A,k}y)_i + \tilde{q}_i = -(M_{A,k}\tilde{x})_i + \tilde{q}_i. \tag{31}$$

Hence, constraints (13b) and (13e), as well as (13g)–(13h) are satisfied for all such i.

From the property (ii) in Lemma 4 and the definition of $M_{A,k}$ we also have that:

$$\delta = -(M_{A,k}y)_i = (-kAy - kIy + Jy)_i \leq (Jy)_i \leq e^\top y = 1,$$

which, in turn, implies by (29) that:

$$\tilde{s} = \epsilon \cdot \delta \leq \lceil 3c^n \sqrt{n} \rceil \cdot \tilde{h}, \tag{32}$$

and this upper bound is exploited further in the proof.

Next, consider all $i \in [n]$ such that $(M_{A,k}y)_i \geq 0$ and $y_i = 0$ (property (i) in Lemma 4 holds). The latter implies that $\tilde{x}_i = \epsilon y_i = 0$, so we can always select a sufficiently large $\tilde{q}_i > 0$ such that

$$\tilde{s} = -\epsilon \cdot (M_{A,k}y)_i + \tilde{q}_i = -(M_{A,k}\tilde{x})_i + \tilde{q}_i, \tag{33}$$

is satisfied. Also, from the above equation, it is clear that we can always pick $\tilde{q}_i$ such that

$$\tilde{q}_i \leq \lceil 3c^n \sqrt{n} \rceil \cdot \tilde{h} + ((kA + kI - J)\tilde{x})_i \leq \lceil 3c^n \sqrt{n} \rceil \cdot \tilde{h} + k\epsilon e^\top y \leq 2k\lceil 3c^n \sqrt{n} \rceil \cdot \tilde{h},$$

where we use (32) and (30) for the first and the second terms, respectively, in the right-hand side of the first inequality, as well as the fact that $k > 1$. Hence, (13b) and (13e) are satisfied for all such i.

As the above discussion considers all $i \in [n]$, the constructed leader's solution $(\tilde{x}, \tilde{q}, \tilde{s}, \tilde{h})$ is feasible and has the corresponding objective function value given by:

$$\begin{aligned}
\tilde{s} - \frac{1}{n\mu_{k,c}} \cdot \tilde{h} &= \epsilon \cdot \delta - \frac{1}{n\mu_{k,c}\lceil 3c^n \sqrt{n} \rceil}\epsilon = \epsilon\big(\delta - \frac{1}{n\mu_{k,c}\lceil 3c^n \sqrt{n} \rceil}\big) \\
&\geq \epsilon\big(\Delta_{k,c} - \frac{\Delta_{k,c}}{2\sigma_{k,c}\lceil 3c^n \sqrt{n} \rceil}\big) = \epsilon \cdot \Delta_{k,c} \cdot \big(1 - \frac{1}{2\sigma_{k,c}\lceil 3c^n \sqrt{n} \rceil}\big) > 0,
\end{aligned}$$

where we use (15) and (29) as well as the property (iii) in Lemma 4 and the fact that $\sigma_{k,c} \geq 1$.

Finally, as $\tilde{t}_i = 0$ by our construction for all $i \in [n]$, the outlined solution is bilevel feasible. By changing ϵ in (29) this leader's solution can be made arbitrarily close to $(\hat{x}, \hat{q}, \hat{s}, \hat{h}) = (0, 0, 0, 0)$ with its objective function better than $\hat{s} = 0$ as indicated in the derivations above. Hence,

$(\hat{x}, \hat{q}, \hat{s}, \hat{h}) = (0, 0, 0, 0)$ is not locally optimal. $\square$

Finally, we are ready to prove the main result.

Proof of Theorem 1: Given graph G and integer r, we need to decide whether $\alpha(G) \geq r$, which is clearly an NP-hard problem. Define $k := r + 0.5$; recall (14). Next, we construct a BLP instance as in (13). Note that all parameters of the instance, see (15), are of polynomial size including $\mu_{k,c}$ as $O(\log \mu_{k,c}) := O(\log k + n \log k + n^2 \log c + n \log n)$. Furthermore, we observe that $\alpha(G) = k$ can never occur. Thus, we need to consider only two mutually exclusive cases that is, either $\alpha(G) < k$, or $\alpha(G) > k$. Suppose $(\hat{x}, \hat{q}, \hat{s}, \hat{h})$ is locally optimal for (13). Thus, one of the two cases holds:

(i) $\alpha(G) < k$. Then, from Proposition 6 we conclude that the leader's solution $(\hat{x}, \hat{q}, \hat{s}, \hat{h}) = (0, 0, 0, 0)$ is the only locally optimal solution.

(ii) $\alpha(G) > k$. Then, from Propositions 5 and 7 we conclude that $\hat{h} = 1$ and $e^\top \hat{x} = \lceil 3c^n \sqrt{n} \rceil$.

There exists *no* leader's solution that has x within Euclidean distance c^n from both the origin $x = 0$ and the hyperplane $e^\top x = \lceil 3c^n \sqrt{n} \rceil$. Hence, if there exists a polynomial-time algorithm that finds a leader's solution that is within Euclidean distance of c^n of any locally optimal leader's solution, then we can answer the question whether $\alpha(G) \geq r$ in polynomial time. $\square$

4 Concluding remarks

Our main result shows that producing a locally optimal leader's decision for a BLP is no easier (in the worst case) than producing a globally optimal one. Note that the existence of a locally (and globally) optimal leader's decision is guaranteed because the feasible regions for both the leader and follower are non-empty and bounded. However, certifying local optimality for the leader seems to require producing a (partial) basis of the lower-level LP, which, in turn, involves a combinatorial search. Thus, it is not surprising in hindsight that the considered problem is also difficult. Our proof illustrates this point concretely by showing that, for some specific BLP class, locally optimal leader's solutions are all associated with a particular basis of the lower-level LP. Finally, our construction also shows that it is difficult to decide whether there exists a locally optimal solution in the interior of the leader's feasible region (or on its boundary).

On the positive side, however, the close connection that has been demonstrated between combinatorial, quadratic and bilevel optimization problems, is compelling. It is clear from the existing reductions that these three problem classes are fundamentally equivalent. This observation could lead to some interesting cross-fertilization.

References

J. Abello, S. Butenko, P. M. Pardalos, and M. G. Resende. Finding independent sets in a graph using continuous multivariable polynomial formulations. *Journal of Global Optimization*, 21: 111–137, 2001.

A. A. Ahmadi and J. Zhang. On the complexity of finding a local minimizer of a quadratic function over a polytope. *Mathematical Programming*, 195(1-2):783–792, 2022.

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(2):273–300, 1997.

A. Baggio, M. Carvalho, A. Lodi, and A. Tramontani. Multilevel approaches for the critical node problem. *Operations Research*, 69(2):486–508, 2021.

Y. Beck, I. Ljubić, and M. Schmidt. A survey on bilevel optimization under uncertainty. *European Journal of Operational Research*, 311(2):401–426, 2023.

O. Ben-Ayed and C. E. Blair. Computational difficulties of bilevel linear programming. *Operations Research*, 38(3):556–560, 1990.

J. S. Borrero, O. A. Prokopyev, and D. Sauré. Sequential interdiction with incomplete information and learning. *Operations Research*, 67(1):72–89, 2019.

J. Brenner and L. Cummings. The Hadamard maximum determinant problem. *The American Mathematical Monthly*, 79(6):626–630, 1972.

C. Buchheim. Bilevel linear optimization belongs to NP and admits polynomial-size KKT-based reformulations. *Operations Research Letters*, 51(6):618–622, 2023.

S. Busygin. A new trust region technique for the maximum weight clique problem. *Discrete Applied Mathematics*, 154(15):2080–2096, 2006.

M. De Santis and F. Rinaldi. Continuous reformulations for zero–one programming problems. *Journal of Optimization Theory and Applications*, 153:75–84, 2012.

S. Dempe. *Foundations of bilevel programming*. Springer Science & Business Media, 2002.

X. Deng. Complexity issues in bilevel linear programming. In P. Pardalos, A. Migdalas, and P. Värbrand, editors, *Multilevel Optimization: Algorithms and Applications*, pages 149–164. Springer, 1998.

M. Garey and D. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W.H. Freeman, San Francisco, 1979.

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.

J.-B. Hiriart-Urruty and A. Seeger. A variational approach to copositive matrices. *SIAM Review*, 52(4):593–629, 2010.

R. Jeroslow. The polynomial hierarchy and a simple model for competitive analysis. *Mathematical Programming*, 32:146–164, 1985.

L. G. Khachiyan. A polynomial algorithm in linear programming. In *Doklady Akademii Nauk*, volume 244, pages 1093–1096. Russian Academy of Sciences, 1979.

T. Kleinert, M. Labbé, F. a. 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.

T. Kleinert, M. Labbe, I. Ljubic, and M. Schmidt. A survey on mixed-integer programming techniques in bilevel optimization. *EURO Journal on Computational Optimization*, 9:100007, 2021. ISSN 2192-4406. doi: https://doi.org/10.1016/j.ejco.2021.100007. URL `https://www.sciencedirect.com/science/article/pii/S2192440621001349`.

T. Lagos and O. A. Prokopyev. On complexity of finding strong-weak solutions in bilevel linear programming. *Operations Research Letters*, 51(6):612–617, 2023.

P. Marcotte and G. Savard. Bilevel programming: A combinatorial perspective. In D. Avis, A. Hertz, and O. Marcotte, editors, *Graph theory and combinatorial optimization*, pages 191–217. Springer, 2005.

P. Marcotte, A. Mercier, G. Savard, and V. Verter. Toll policies for mitigating hazardous materials transport risk. *Transportation Science*, 43(2):228–243, 2009.

T. S. Motzkin and E. G. Straus. Maxima for graphs and a new proof of a theorem of Turán. *Canadian Journal of Mathematics*, 17:533–540, 1965.

P. M. Pardalos and G. Schnitger. Checking local optimality in constrained quadratic programming is NP-hard. *Operations Research Letters*, 7(1):33–35, 1988.

P. M. Pardalos and S. A. Vavasis. Quadratic programming with one negative eigenvalue is NP-hard. *Journal of Global Optimization*, 1(1):15–22, 1991.

P. M. Pardalos and S. A. Vavasis. Open questions in complexity theory for numerical optimization. *Mathematical Programming*, 57(1-3):337–339, 1992.

P. M. Pardalos and J. Xue. The maximum clique problem. *Journal of Global Optimization*, 4: 301–328, 1994.

X. Shi, O. A. Prokopyev, and T. K. Ralphs. Mixed integer bilevel optimization with a k-optimal follower: a hierarchy of bounds. *Mathematical Programming Computation*, 15(1):1–51, 2023.

V. Stozhkov, A. Buchanan, S. Butenko, and V. Boginski. Continuous cubic formulations for cluster detection problems in networks. *Mathematical Programming*, 196(1):279–307, 2022.

L. Vicente, G. Savard, and J. Júdice. Descent approaches for quadratic bilevel programming. *Journal of Optimization theory and applications*, 81(2):379–399, 1994.

W. Xia, J. C. Vera, and L. F. Zuluaga. Globally solving nonconvex quadratic programs via linear integer programming techniques. *INFORMS Journal on Computing*, 32(1):40–56, 2020.

J. Yang, X. Shi, and O. A. Prokopyev. Exact solution approaches for a class of bilevel fractional programs. *Optimization Letters*, 17(1):191–210, 2023.

M. H. Zare, J. S. Borrero, B. Zeng, and O. A. Prokopyev. A note on linearized reformulations for a class of bilevel linear integer problems. *Annals of Operations Research*, 272:99–117, 2019.

Appendix A Illustration of a BLP instance given by the example in (4)

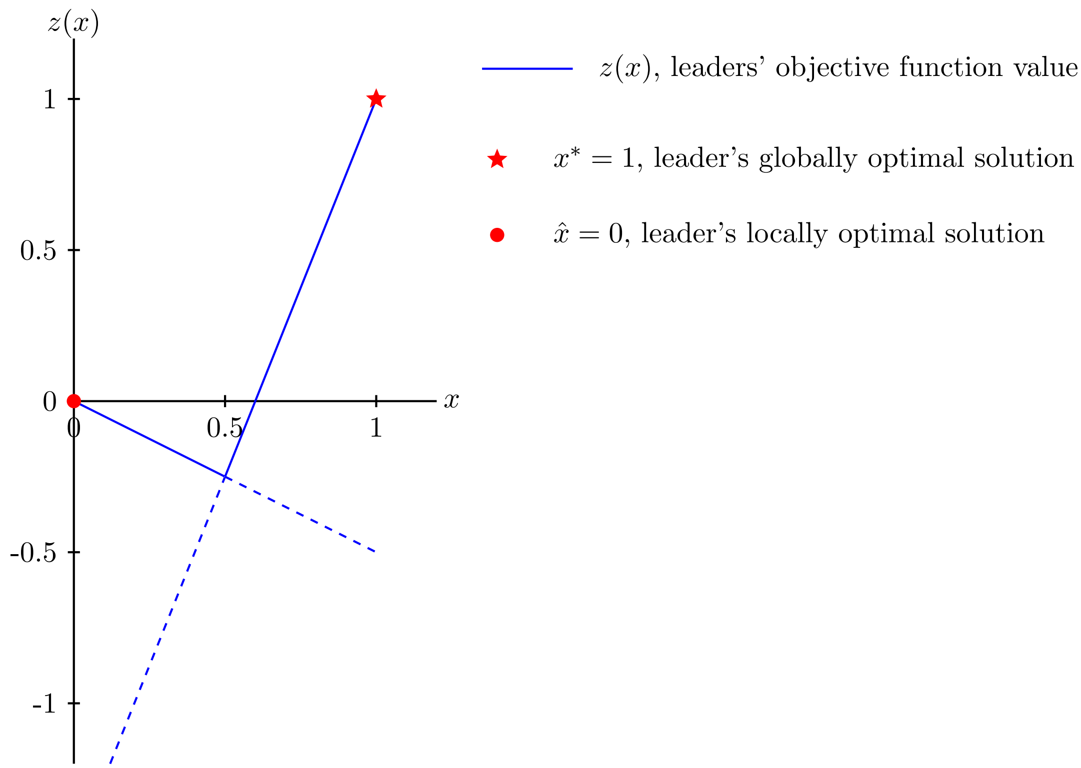

Figure 1: An illustration of the example given by (4) in Section 1. Given leader's decision $x \in [0, 1]$, the leader's objective function value is $z(x) = -\frac{1}{2}x$ for $x \in [0, 0.5]$ and $z(x) = \frac{5}{2}x - \frac{3}{2}$ for $x \in [0.5, 1]$.

Appendix B Motivating geometric illustration

Here, we briefly motivate the geometric intuition behind our main result with a simple illustration. In the following description, we denote by $\mathcal{P} = \{(x, y) \mid x \in \mathcal{X},\ By \leq b_2 - A_2x,\ y \in \mathbb{R}^{n_2}_+\}$ the feasible region of the single-level relaxation of (**BLP**), which is obtained by relaxing the optimality condition on the follower's decision. For a given $\hat{x} \in \mathcal{X}$, we further denote the set of follower's decisions that are feasible with respect to the linear constraints of (**BLP**) as $\mathcal{P}(\hat{x}) := \{y \in \mathbb{R}^{n_2}_+ \mid By \leq b_2 - A_2\hat{x}\}$.

Set $\mathcal{P}(\hat{x})$ is itself a polyhedron contained in $\mathcal{P}$ and as such, $\mathcal{R}(\hat{x})$, being the optimal face of $\mathcal{P}(\hat{x})$ with respect to the objective function vector d_2, see (1), is also a polyhedron. Next, assume for simplicity an optimistic model. Then, since the leader's objective function is linear and there are no coupling constraints, the bilevel feasible solutions of (**BLP**), which are contained in $\mathcal{R}(\hat{x})$, are those on the face $\mathcal{F}(\hat{x}) = \operatorname{argmax}\{d_1^\top y \mid y \in \mathcal{R}(\hat{x})\}$ of $\mathcal{R}(\hat{x})$ that are optimal with respect to the *leader's* objective function. Thus, $\mathcal{F}(\hat{x})$ is again a polyhedron, and also, a face of $\mathcal{P}(\hat{x})$, which consists of all bilevel feasible points contained in $\mathcal{R}(\hat{x})$. Consequently, optimal solutions to (**BLP**) must be contained in $\mathcal{F}(x)$ for *some* $x \in \mathcal{X}$, which is itself contained in a face of $\mathcal{P}$.

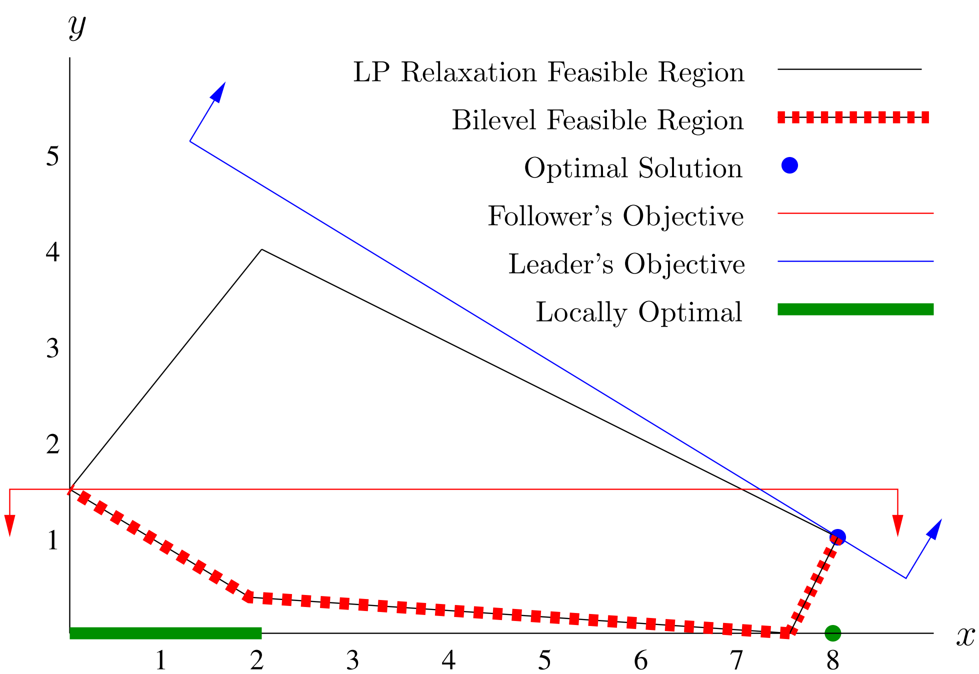

Figure 2: An illustrative example of a bilevel feasible region.

To make these concepts more concrete, consider the example in Figure 2, where the follower's objective function is to minimize the value of y. The bilevel feasible region is a union of faces of $\mathcal{P}$ that are shown as the dashed lines on the boundary of $\mathcal{P}$ in the figure. In this example, for a given $x \in \mathcal{X}$, set $\mathcal{F}(x)$ is a single point in the bilevel feasible region associated with x. Furthermore, for the considered leader's and the follower's objective functions, the globally optimal solution is $(8, 1)$.

For $\hat{x} \in \mathcal{X}$ to be locally optimal for the leader, we must have that for any $u \in \mathbb{R}^{n_1}$, the leader's objective function value associated with $\mathcal{F}(x + \delta u)$ is no larger than that of $\mathcal{F}(x)$ for $\delta \in \mathbb{R}_{>0}$ sufficiently small. Roughly speaking, $\mathcal{F}(x)$ must not be close to a face of $\mathcal{P}$ containing points having better leader's objective function values than those of $\mathcal{F}(x)$. In this example, the locally optimal leader's solutions are (i) the globally optimal solution $x = 8$ and (ii) $x \in [0, 2)$. Indeed, the face of $\mathcal{P}$ containing $\mathcal{F}(x)$ for $x \in [0, 2)$ is parallel to the leader's objective function as depicted in Figure 2. Hence, any $x \in \mathcal{X}$ for which $\mathcal{F}(x)$ is an inner point of this face must be locally optimal.