Valid Inequalities for Bilevel 0-1 Shortest-Path Problems

Bruno M. Pacheco, Margarida Carvalho, and Kim Yu

CIRRELT, Université de Montréal

Abstract. We characterize valid inequalities for bilevel programs in which the leader makes binary decisions, and the follower solves a shortest-path problem. We exploit a dualize-and-combine reformulation and show that even with ideal formulations for the leader and the follower primal solution sets, fractional solutions may still appear in the continuous relaxation. From that reformulation, we identify that the interaction between the leader variables and the follower dual variables, which we capture through a leader–potential set, is the key structure to be convexified to guarantee that a continuous relaxation recovers the optimal value of the original problem, closing the objective gap. Furthermore, we characterize the valid inequalities for the leader–potential set through network flows, and identify a family of inequalities we name *root cuts* that suffice to recover the optimal value. Complementary to the structural results, we prove that the cut generation problem is NP-hard even with partially fixed coefficients and a network with two nodes and two parallel arcs. Motivated by the complexity result, we propose a constructive method to fix coefficients in the valid inequalities while preserving separation guarantees, by exploiting the network structure.

Keywords: Binary bilevel optimization · Valid inequalities · Shortest-path follower · Cut generation · Separation.

1 Introduction

Bilevel programs model hierarchical decisions in which a *leader*, at the *upper level*, anticipates the optimal response of a *follower*, at the *lower level* [10,16]. In this work, we are interested in bilevel binary programs of the form

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{y}} \quad & \boldsymbol{c}^\top(\boldsymbol{x},\boldsymbol{y}) \\ \text{s.t.} \quad & \boldsymbol{x} \in \mathcal{X}, \\ & \boldsymbol{y} \in \arg\min_{\boldsymbol{y}' \in \mathcal{Y}} \boldsymbol{q}(\boldsymbol{x})^\top \boldsymbol{y}'. \end{aligned} \tag{1}$$

Here, $\boldsymbol{x} \in \mathcal{X} \subseteq \{0,1\}^n$ represents the leader's decisions, where $\mathcal{X}$ is a set of binary vectors described by linear inequalities. Vector $\boldsymbol{y}$ represents the follower's decisions, and $\boldsymbol{c}$ collects the coefficients of the leader's linear objective in $(\boldsymbol{x},\boldsymbol{y})$. The follower solves a shortest-path problem, and $\mathcal{Y}$ consists of the arc-incidence vectors of the feasible paths. The leader affects the follower's optimization solely through arc costs $\boldsymbol{q}(\boldsymbol{x})$ that are affine in $\boldsymbol{x}$. The formulation adopts the optimistic assumption, under which ties among optimal follower paths are resolved in the leader's favor.

Shortest-path followers as above arise in a wide range of optimization settings, and in many of these contexts, the leader naturally affects the follower's objective. Network pricing is one such example, where a leader anticipates how users (followers) will route through a network in response to her decisions [17,4,7]. The same structure also arises when a combinatorial follower is represented by a finite-stage dynamic program [6], such as a decision diagram [18,2]. In these representations, which may be exponentially large, paths through an acyclic state-transition graph encode follower solutions, and arc weights accumulate their objective values. When the follower's feasible set is fixed, and its objective coefficients depend affinely on binary leader decisions, this yields precisely the follower structure in (1). Similarly, our setting accommodates certain interdiction models after transferring linking constraints—lower-level constraints parametric on the leader's variables—to the follower's objective and constructing a network reformulation based on the follower's dynamic-programming representation. Knapsack interdiction provides such an example [8,14,6]. Related network formulations encode leader-dependent follower feasibility through arc capacities or represent the follower's value function over leader decisions [21,19].

Under common assumptions, the standard arc-flow formulation of the follower's shortest path is ideal. Thus, traditional *dualize-and-combine* can be directly applied to solve the bilevel problem, enforcing follower's optimality through primal feasibility, dual feasibility, and strong duality [7,18,6]. In our setting, this yields an extended single-level formulation in the leader decisions, network paths (follower's primal), and node potentials (follower's dual), linked by a bilinear optimality constraint. However, even with ideal descriptions of the leader and follower feasible sets, the resulting single-level relaxation can contain fractional points outside the convex hull of bilevel-feasible solutions. In other words, the relaxation may still contain fractional optimal solutions that satisfy follower's optimality and have strictly better objective values than any feasible solution to the original problem.

While cutting-planes have demonstrated success in separating fractional solutions in general mixed-integer programming, in bilevel settings we observe an absence of methods targeting fractional points that satisfy the follower-optimality condition. Existing cut generation for mixed-integer bilevel optimization targets, in particular, points that satisfy all leader and follower constraints, including integrality, but violate follower optimality [12,13]. When a point with a fractional leader decision already satisfies the follower-optimality condition, Tahernejad and Ralphs [20] discuss the use of standard mixed-integer linear programming cuts. Recent work by Vásquez et al. [22] derives cuts in the original variable space from networks that jointly encode the binary leader assignments and the follower's value function, and strengthens them via network-based tilting. In contrast, our bilevel problem (1) contains a network structure only in the follower, and we explore how this structure can be leveraged to characterize the convex hull of feasible solutions. Davarnia et al. [9] convexify bilinear terms over a dual network polytope and a simplex leader to strengthen dual formulations of network interdiction. In our setting, the leader decisions interact with the follower's dual through the arc costs, and not through bilinear terms. Furthermore, we do not assume the leader to be restricted to simplex variables.

Contributions and organization. We describe our problem setting in Section 2, and we make the following contributions in Sections 3–4.

(i) We establish the role of leader–potential convexification in the full single-level formulation (Section 3.1). With the follower's path polytope and the bilinear optimality constraint retained, we show that convexifying the leader–potential set makes every remaining feasible point a convex combination of feasible single-level solutions.
(ii) We characterize all valid linear inequalities to the leader–potential set (Section 3.2). Specifically, we provide necessary and sufficient conditions for the coefficients of node potential variables through network flow.
(iii) For a subclass of valid inequalities with special structures, which we call *root cuts*, we prove that adding them to the continuous relaxation already suffices to recover the optimal value of the original problem (Section 3.3).
(iv) We propose a constructive coefficient fixing for the node potentials enhanced by the network structure (Section 4.1), and we prove that the cut generation problem is NP-hard even with fixed node potential coefficients and a network consisting of two parallel arcs (Section 4.2).

2 Problem Setting

We first detail the follower's problem in Section 2.1 and then derive the single-level reformulation of (1) via dualize-and-combine in Section 2.2.

2.1 Network assumptions and path formulation

Let $\mathcal{D} = (\mathcal{V}, \mathcal{A})$ be the directed graph for the follower's problem, with distinct root $r \in \mathcal{V}$ and destination $t \in \mathcal{V}$. We also assume $\mathcal{D}$ to be acyclic, we allow parallel arcs and retain only nodes and arcs on $r \rightsquigarrow t$ paths. This preprocessing preserves all follower choices and makes r and t the unique source and sink, respectively. Two network examples are illustrated in Figure 1.

The above graph assumptions are not stringent; they naturally accommodate the representations from finite-stage dynamic programs and decision diagrams [6,2], but also broader routing settings through equivalent representations. When we have multiple independent shortest-path followers, connecting copies of their networks in series yields a single path problem of size linear in their original size, with each path selecting one route per follower. For cyclic networks, acyclic representations preserve all optimal follower paths if every directed cycle has strictly positive cost for every $\boldsymbol{x} \in \mathcal{X}$. Both constructions preserve the leader's linear objective and optimistic tie-breaking under projection. Common assumptions such as positive base costs and nonnegative tolls [7] or absence of nonpositive cost cycles [17,5] satisfy the condition for an acyclic representation.

The follower's path set is independent of $\boldsymbol{x}$, and its arc weights are given by $\boldsymbol{q}(\boldsymbol{x}) = Q\boldsymbol{x} + \boldsymbol{q}_0$, where $Q \in \mathbb{R}^{\mathcal{A} \times n}$ and $\boldsymbol{q}_0 \in \mathbb{R}^{\mathcal{A}}$. For each arc $a \in \mathcal{A}$, let $u(a)$ and $v(a)$ denote its tail and head. The node–arc incidence matrix $D \in \mathbb{R}^{\mathcal{V} \times \mathcal{A}}$ has column $D_a = \boldsymbol{e}_{u(a)} - \boldsymbol{e}_{v(a)}$, where $\boldsymbol{e}_u$ is the unit vector indexed by node u.

Recall that $\mathcal{Y}$ is the set of arc-incidence vectors of $r \rightsquigarrow t$ paths. By flow decomposition and acyclicity, every nonnegative unit flow is a convex combination of

path-incidence vectors. Thus, the follower has a bounded ideal linear formulation and a finite optimal value for every leader assignment. In particular,

$$\operatorname{conv}(\mathcal{Y}) = \{\boldsymbol{y} \in \mathbb{R}_+^{\mathcal{A}} : D\boldsymbol{y} = \boldsymbol{e}_r - \boldsymbol{e}_t\}. \tag{2}$$

2.2 Strong-duality reformulation

In what follows, let

$$\Omega := \left\{(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y}) \in \mathcal{X} \times \mathbb{R}^{\mathcal{V}} \times \mathcal{Y} : \ \boldsymbol{q}(\boldsymbol{x})^\top \boldsymbol{y} \le \theta_r, \quad D^\top \boldsymbol{\theta} \le \boldsymbol{q}(\boldsymbol{x}), \quad \theta_t = 0\right\}.$$

The constraints $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{y} \in \mathcal{Y}$ describe leader feasibility and follower path-incidence vectors, respectively. The *dual arc inequalities* $D^\top \boldsymbol{\theta} \le \boldsymbol{q}(\boldsymbol{x})$ couple leader decisions and node potentials. The bilinear *optimality constraint* $\boldsymbol{q}(\boldsymbol{x})^\top \boldsymbol{y} \le \theta_r$ enforces strong duality and couples all three variable blocks, $\boldsymbol{x}$, $\boldsymbol{\theta}$ and $\boldsymbol{y}$. Here, we normalize the dual node potentials $\boldsymbol{\theta}$ by $\theta_t = 0$. Following from (2) and strong duality, (1) admits the following equivalent single-level formulation:

$$\min_{(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y}) \in \Omega} \quad \boldsymbol{c}^\top (\boldsymbol{x}, \boldsymbol{y}). \tag{3}$$

This formulation is exact for binary $(\boldsymbol{x}, \boldsymbol{y})$ [18], but its continuous relaxation may have fractional extreme points. We seek valid inequalities to strengthen it.

3 Valid Inequalities

This section identifies which valid inequalities must be added to the relaxation of (3) to recover the bilevel optimum. Throughout this section, we assume $\mathcal{X} \neq \emptyset$, since otherwise (1) is infeasible and every inequality is valid. Section 3.1 shows that it suffices to convexify the set coupling leader decisions and node potentials. Section 3.2 characterizes its valid inequalities through network supplies and affine upper bounds of minimum-cost flow values over $\mathcal{X}$. Section 3.3 shows that the root supply $\boldsymbol{e}_r - \boldsymbol{e}_t$ suffices for the bilevel objective and establishes the limits of more general unit supplies for full convexification.

We first justify seeking a linear description of the feasible hull despite the bilinear optimality constraint. Although the extended formulation (3) introduces continuous variables, the convex hull of its feasible solutions $\operatorname{conv}(\Omega)$ is a polytope. This can be seen from a disjunctive programming perspective, which we formalize in the following proposition.

Proposition 1. $\operatorname{conv}(\Omega)$ *is a polytope.*

Consequently, finitely many valid linear inequalities suffice to recover the bilevel optimum from the continuous relaxation of (3). To identify a structured source of such inequalities, we first account for leader and follower integrality separately. The follower's flow formulation already describes $\operatorname{conv}(\mathcal{Y})$, and we may also replace $\mathcal{X}$ by $\operatorname{conv}(\mathcal{X})$. As the next example shows, however, the interaction between the variable blocks can still leave fractional points outside $\operatorname{conv}(\Omega)$ and an objective gap.

Example 1. Let $\mathcal{X} = \{0, 1\}$ and consider two parallel arcs a_1, a_2 from r to t, with weights $q_{a_1}(x) = 1 - 2x$ and $q_{a_2}(x) = 0$. The network is illustrated in Figure 1. The follower's path set is $\mathcal{Y} = \{(1, 0), (0, 1)\}$, with y_i indicating arc a_i. Since $\theta_t = 0$ and $y_2 = 1 - y_1$, we retain only (x, θ_r, y_1). Replacing $\mathcal{X}$ and $\mathcal{Y}$ by their convex hulls in (3) gives $(1-2x)y_1 \le \theta_r \le \min\{1-2x, 0\}$ for $0 \le x, y_1 \le 1$. The fractional point $(1/2, 0, 1)$ is feasible, but lies outside the segment joining the only binary-feasible solutions $(0, 0, 0)$ and $(1, -1, 1)$. For the objective $x - y_1$, it has value $-1/2$, whereas both binary solutions have value zero.

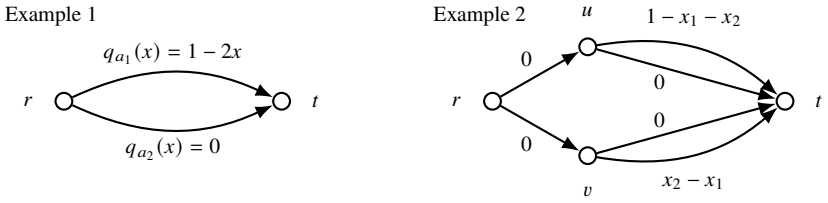


Fig. 1. Follower networks. Arc labels give costs.

Let $\phi(\boldsymbol{x})$ denote the follower's optimal value for $\boldsymbol{x} \in \mathrm{conv}(\mathcal{X})$. In the example, the dual arc inequalities impose $\theta_r \le \phi(x) = \min\{1 - 2x, 0\}$, while convexifying the leader–potential set gives the stronger inequality $\theta_r \le -x$. This motivates studying that set and the effect of its convexification on the full formulation.

3.1 The dual network polyhedron

For fixed leader decisions, the dual arc inequalities define a dual network flow polyhedron [3]. Allowing the arc weights to depend on the leader gives the mixed-integer set

$$\mathcal{S} = \left\{(\boldsymbol{x}, \boldsymbol{\theta}) \in \mathcal{X} \times \mathbb{R}^{\mathcal{V}} : D^\top \boldsymbol{\theta} \le \boldsymbol{q}(\boldsymbol{x}),\ \theta_t = 0\right\}$$

and its relaxation

$$\mathcal{P} = \left\{(\boldsymbol{x}, \boldsymbol{\theta}) \in \mathrm{conv}(\mathcal{X}) \times \mathbb{R}^{\mathcal{V}} : D^\top \boldsymbol{\theta} \le \boldsymbol{q}(\boldsymbol{x}),\ \theta_t = 0\right\}.$$

We refer to $\mathcal{S}$ as the *leader–potential* set, and to $\mathcal{P}$ as its continuous relaxation. We note that structures similar to $\mathcal{P}$ have been explored under the names *dual network polytope* [9] and *dual network flow polyhedra* [3] in different settings and with objectives distinct from ours.

Using $\mathrm{conv}(\mathcal{X})$ in $\mathcal{P}$ isolates the interaction between leader decisions and potentials. For a set C in $(\boldsymbol{x}, \boldsymbol{\theta})$-space, write $C(\bar{\boldsymbol{x}}) = \{\boldsymbol{\theta} : (\bar{\boldsymbol{x}}, \boldsymbol{\theta}) \in C\}$ for its *potential fiber*. Every fiber $\mathcal{P}(\bar{\boldsymbol{x}})$ with $\bar{\boldsymbol{x}} \in \mathrm{conv}(\mathcal{X})$ is nonempty, as witnessed by shortest-path costs from each node to t. Moreover, $\mathcal{S}(\hat{\boldsymbol{x}}) = \mathcal{P}(\hat{\boldsymbol{x}})$ for $\hat{\boldsymbol{x}} \in \mathcal{X}$, so $\mathrm{Proj}_{\boldsymbol{x}}\, \mathcal{S} = \mathcal{X}$. The finitely many nonempty fibers of $\mathcal{S}$ are pointed polyhedra with a common recession cone, hence $\mathrm{conv}(\mathcal{S})$ is also a polyhedron.

With the bilinear optimality constraint retained, convexifying $\mathcal{S}$ excludes every point outside $\text{conv}(\Omega)$. Define

$$\mathcal{E} = \left\{(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y}) : (\boldsymbol{x}, \boldsymbol{\theta}) \in \text{conv}(\mathcal{S}),\ \boldsymbol{y} \in \text{conv}(\mathcal{Y}),\ \boldsymbol{q}(\boldsymbol{x})^\top \boldsymbol{y} \leq \theta_r\right\}.$$

Proposition 2. $\text{conv}(\mathcal{E}) = \text{conv}(\Omega)$.

Thus every point of $\mathcal{E}$ is a convex combination of feasible single-level solutions, and every linear objective has the same optimal value over $\mathcal{E}$ and Ω. This justifies focusing on valid inequalities for $\mathcal{S}$. Although $\mathcal{E}$ may be nonconvex, these inequalities also strengthen the McCormick relaxation of the bilinear terms. In Example 1, adding $\text{conv}(\mathcal{S}) = \{(x, \theta_r) : 0 \leq x \leq 1,\ \theta_r \leq -x\}$ to that relaxation even yields an ideal formulation.

3.2 Valid inequalities and network flows

We now characterize all valid inequalities for $\mathcal{S}$. Consider valid inequalities for $\text{conv}(\mathcal{S})$ of the form

$$\boldsymbol{\beta}^\top \boldsymbol{\theta} \leq \boldsymbol{\pi}^\top \boldsymbol{x} + \pi_0. \tag{4}$$

With $\boldsymbol{\beta} = \mathbf{0}$, those are precisely the valid inequalities for $\text{Proj}_{\boldsymbol{x}}\, \mathcal{S} = \mathcal{X}$. Note that we assume only that the leader has a linear formulation, thus there is no structure to be exploited in generating the valid inequalities with $\boldsymbol{\beta} = \mathbf{0}$, and one can resort to general integer-programming techniques. Our focus is on inequalities with $\boldsymbol{\beta} \neq \mathbf{0}$, which account for the interaction between leader decisions and node potentials. Since $\theta_t = 0$, we may choose $\beta_t = -\sum_{u \in \mathcal{V} \setminus \{t\}} \beta_u$ without changing the inequality. Throughout, we use this representative, so that $\sum_{u \in \mathcal{V}} \beta_u = 0$.

All valid inequalities of interest have node potential coefficients $\boldsymbol{\beta}$ that define valid supplies over the network. Intuitively, $\boldsymbol{\beta}$ derives from nonnegative combinations of the follower's dual inequalities, thus columns of D. Combining these columns with nonnegative weights can therefore be viewed as assigning flows to the arcs. We define the supply cone and its associated flow sets by

$$\mathcal{B} = \{D\boldsymbol{\lambda} : \boldsymbol{\lambda} \in \mathbb{R}_+^{\mathcal{A}}\}, \qquad \Lambda(\boldsymbol{\beta}) = \{\boldsymbol{\lambda} \in \mathbb{R}_+^{\mathcal{A}} : D\boldsymbol{\lambda} = \boldsymbol{\beta}\}.$$

For $\boldsymbol{\beta} \in \mathcal{B}$, acyclicity makes $\Lambda(\boldsymbol{\beta})$ a nonempty polytope, and linear programming duality gives the finite, attained value

$$f_{\boldsymbol{\beta}}(\boldsymbol{x}) := \max_{\boldsymbol{\theta} \in \mathcal{P}(\boldsymbol{x})} \boldsymbol{\beta}^\top \boldsymbol{\theta} = \min_{\boldsymbol{\lambda} \in \Lambda(\boldsymbol{\beta})} \boldsymbol{q}(\boldsymbol{x})^\top \boldsymbol{\lambda}, \qquad \forall\, \boldsymbol{x} \in \text{conv}(\mathcal{X}). \tag{5}$$

Proposition 3. *Let $\boldsymbol{\beta}$ satisfy $\sum_{u \in \mathcal{V}} \beta_u = 0$. Inequality* (4) *is valid for $\mathcal{S}$ if and only if $\boldsymbol{\beta} \in \mathcal{B}$ and*

$$\boldsymbol{\pi}^\top \boldsymbol{x} + \pi_0 \geq f_{\boldsymbol{\beta}}(\boldsymbol{x}) \qquad \forall\, \boldsymbol{x} \in \mathcal{X}.$$

Equivalently, for every $\boldsymbol{x} \in \mathcal{X}$ there exists a flow certificate $\boldsymbol{\lambda}^{\boldsymbol{x}} \in \Lambda(\boldsymbol{\beta})$ such that

$$\boldsymbol{\pi}^\top \boldsymbol{x} + \pi_0 \geq \boldsymbol{q}(\boldsymbol{x})^\top \boldsymbol{\lambda}^{\boldsymbol{x}}. \tag{6}$$

By Proposition 3, the right-hand side of any valid inequality defines an affine upper bound for $f_{\boldsymbol{\beta}}$ on $\mathcal{X}$. Since the graph is acyclic, $\Lambda(\boldsymbol{\beta})$ is a polytope, and $f_{\boldsymbol{\beta}}$ can be expressed as a minimum of finitely many affine functions of $\boldsymbol{x}$, one for each extreme point of $\Lambda(\boldsymbol{\beta})$. Although $f_{\boldsymbol{\beta}}$ is therefore concave, we are interested in upper bounds for the function over $\mathcal{X}$ only, which means they might imply stronger bounds at fractional leader coordinates than the function itself. For $\boldsymbol{\beta} \in \mathcal{B}$, let $\hat{f}_{\boldsymbol{\beta}}$ be the smallest concave function on $\operatorname{conv}(\mathcal{X})$ bounded below by $f_{\boldsymbol{\beta}}$ on $\mathcal{X}$, that is, let $\hat{f}_{\boldsymbol{\beta}}$ be the *concave envelope* of $f_{\boldsymbol{\beta}}$ over $\mathcal{X}$.

Corollary 1. *For $\boldsymbol{\beta} \in \mathcal{B}$,* (4) *is valid for $\mathcal{S}$ if and only if $\boldsymbol{\pi}^\top \boldsymbol{x} + \pi_0$ is an affine upper bound of $\hat{f}_{\boldsymbol{\beta}}$ on* $\operatorname{conv}(\mathcal{X})$.

Consequently, at every $\bar{\boldsymbol{x}} \in \operatorname{conv}(\mathcal{X})$, there exists a supporting hyperplane of the hypograph of $\hat{f}_{\boldsymbol{\beta}}$ whose associated valid inequality has minimal right-hand side at $\bar{\boldsymbol{x}}$ among all valid inequalities with potential coefficients $\boldsymbol{\beta}$.

3.3 Root supply and optimality gap

The root potential θ_r is the only potential in the follower's optimality constraint of (3), and no potential appears in the leader's objective. This motivates studying inequalities involving only $\boldsymbol{x}$ and θ_r. We show that this family suffices to recover the optimal value of the bilevel problem.

The *root supply* $\boldsymbol{\beta}^r = \boldsymbol{e}_r - \boldsymbol{e}_t$ supplies one unit of flow at the root r and consumes it at the target t. Since $\Lambda(\boldsymbol{\beta}^r) = \operatorname{conv}(\mathcal{Y})$, (5) gives

$$f_{\boldsymbol{\beta}^r}(\boldsymbol{x}) = \min_{\boldsymbol{y} \in \mathcal{Y}} \boldsymbol{q}(\boldsymbol{x})^\top \boldsymbol{y} = \phi(\boldsymbol{x}).$$

Thus, the associated *root cuts*[1] have the form $\theta_r \le \boldsymbol{\pi}^\top \boldsymbol{x} + \pi_0$, where the right-hand side is an affine upper bound of the follower's value function on $\mathcal{X}$. Let $\hat{\phi} = \hat{f}_{\boldsymbol{\beta}^r}$ denote its concave envelope over $\mathcal{X}$. By Corollary 1, adding all root cuts to $\mathcal{P}$ gives the closure

$$\mathcal{R}_r = \{(\boldsymbol{x}, \boldsymbol{\theta}) \in \mathcal{P} : \theta_r \le \hat{\phi}(\boldsymbol{x})\}. \tag{7}$$

Lemma 1. *The root closure satisfies*

$$\operatorname{Proj}_{(\boldsymbol{x}, \theta_r)} \mathcal{R}_r = \operatorname{Proj}_{(\boldsymbol{x}, \theta_r)} \operatorname{conv}(\mathcal{S}) = \{(\boldsymbol{x}, \eta) : \boldsymbol{x} \in \operatorname{conv}(\mathcal{X}),\ \eta \le \hat{\phi}(\boldsymbol{x})\}.$$

The projection equality implies that, for every $\boldsymbol{a} \in \mathbb{R}^n$ and $\gamma \le 0$,

$$\min_{(\boldsymbol{x}, \boldsymbol{\theta}) \in \mathcal{R}_r} \boldsymbol{a}^\top \boldsymbol{x} + \gamma \theta_r = \min_{(\boldsymbol{x}, \boldsymbol{\theta}) \in \mathcal{S}} \boldsymbol{a}^\top \boldsymbol{x} + \gamma \theta_r. \tag{8}$$

Thus, root cuts close the gap for objectives on $\mathcal{S}$ involving the potentials only through θ_r. This has an important consequence for problems with interdiction objectives [11], which can be formulated [8,14,6] as in (8) with $\boldsymbol{a} = 0$ and $\gamma = -1$. Precisely, for bilevel shortest-path problems with interdiction objectives, root cuts yield an LP relaxation that recovers the optimal objective value.

[1] Here "root" refers to the root node r of the graph, not to the root of a branch-and-bound tree.

More interestingly, this same family actually suffices for the general case, which we prove below. Define the relaxation

$$\mathcal{E}_r = \{(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y}) : (\boldsymbol{x}, \boldsymbol{\theta}) \in \mathcal{R}_r,\ \boldsymbol{y} \in \operatorname{conv}(\mathcal{Y}),\ \boldsymbol{q}(\boldsymbol{x})^\top \boldsymbol{y} \le \theta_r\}.$$

For a point of $\mathcal{E}_r$, take a convex decomposition of its leader component attaining the root envelope and a convex decomposition of its follower path component. The $\hat{\phi}(\boldsymbol{x})$ bound and the bilinear optimality constraint force the follower optimality gaps to zero (under the product weights), so every (positively weighted) leader–path pair is bilevel feasible.

Theorem 1. *Every point of* $\operatorname{Proj}_{(\boldsymbol{x},\boldsymbol{y})} \mathcal{E}_r$ *is a convex combination of bilevel-feasible leader–path pairs. Moreover, for every linear objective* $\boldsymbol{c}^\top(\boldsymbol{x}, \boldsymbol{y})$*,*

$$\min_{(\boldsymbol{x},\boldsymbol{\theta},\boldsymbol{y}) \in \mathcal{E}_r} \boldsymbol{c}^\top(\boldsymbol{x}, \boldsymbol{y}) = \min_{(\boldsymbol{x},\boldsymbol{\theta},\boldsymbol{y}) \in \Omega} \boldsymbol{c}^\top(\boldsymbol{x}, \boldsymbol{y}).$$

The theorem strengthens the conclusion of Proposition 2 for the objective of interest: convexifying the entire leader–potential set $\mathcal{S}$ is unnecessary to close the bilevel gap. The result concerns the complete family of root cuts with the bilinear optimality constraint retained; it does not assert gap closure after McCormick relaxation.

Even the complete family of root cuts can fail to describe $\operatorname{conv}(\mathcal{S})$. This limitation persists for the larger family of all *unit supplies*, namely $\boldsymbol{e}_u - \boldsymbol{e}_v$ for distinct nodes u, v joined by a directed $u \rightsquigarrow v$ path, as the following example shows.

Example 2 (Unit supplies are insufficient). Let $\mathcal{X} = \{0, 1\}^2$ and $\mathcal{V} = \{r, u, v, t\}$. The arcs $r \to u$ and $r \to v$ have cost zero. There are two parallel $u \to t$ arcs of costs 0 and $1 - x_1 - x_2$, and two parallel $v \to t$ arcs of costs 0 and $x_2 - x_1$, as illustrated in Figure 1. The point

$$\bar{\boldsymbol{x}} = (1/2, 1/2), \qquad (\bar{\theta}_r, \bar{\theta}_u, \bar{\theta}_v, \bar{\theta}_t) = (-1/2, 0, 0, 0)$$

belongs to $\mathcal{P}$ and satisfies every cut with a unit supply, including every root cut. Nevertheless, the inequality

$$\theta_u + \theta_v \le -x_1$$

is valid for $\mathcal{S}$ and violated by this point. Its supply is $\boldsymbol{e}_u + \boldsymbol{e}_v - 2\boldsymbol{e}_t$.

Unit supplies contain a generator of every extreme ray of $\mathcal{B}$, since this cone is generated by the columns of D, each of which is a unit supply. Thus even inequalities from all extreme supply directions can leave points outside $\operatorname{conv}(\mathcal{S})$ unseparated.

4 Cut Generation

In this section, we explore the generation of linear valid inequalities (cuts) to separate a given point of a continuous relaxation of (3), possibly already strengthened by other valid inequalities. In particular, we study cut generation for the leader–potential set, construct node potential coefficients that separate fractional vertices of $\mathcal{P}$, and analyze the complexity of separation under partially-fixed coefficients.

Our structured approach to the single-level reformulation (3) can guide the choice of variables involved in separation. Let $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\theta}}, \bar{\boldsymbol{y}})$ be an optimal solution from any continuous relaxation of Ω. If $\bar{\boldsymbol{x}}$ is fractional and its active leader constraints have rank n, then an inequality for $\mathcal{X}$ alone suffices to separate it. As previously discussed, we will not concentrate on this case.

For candidates with $\bar{\boldsymbol{x}} \in \operatorname{conv}(\mathcal{X})$, Proposition 2 justifies focusing on the leader–potential set. Define Ξ to be a continuous relaxation of Ω as follows:

$$\Xi := \left\{(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y}) \in \operatorname{conv}(\mathcal{X}) \times \mathbb{R}^{\mathcal{V}} \times \operatorname{conv}(\mathcal{Y}) : \ \boldsymbol{q}(\boldsymbol{x})^\top \boldsymbol{y} \le \theta_r, \ D^\top \boldsymbol{\theta} \le \boldsymbol{q}(\boldsymbol{x}), \ \theta_t = 0\right\}.$$

The proposition also shows that inequalities of the form (4) with $\boldsymbol{\beta} \neq \mathbf{0}$ are sufficient to separate any point in $\Xi \setminus \operatorname{conv}(\Omega)$. We explore this result in the following lemma.

Lemma 2. *Every extreme point $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\theta}}, \bar{\boldsymbol{y}}) \in \Xi$ satisfies $\bar{\boldsymbol{y}} \in \mathcal{Y}$, and $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\theta}})$ is a vertex of $\mathcal{P}$.*

We therefore focus on vertices of $\mathcal{P}$ with fractional leader coordinates, which necessarily lie outside $\operatorname{conv}(\mathcal{S})$ and admit a separating inequality.

4.1 Fixing node potential coefficients $\boldsymbol{\beta}$

Let $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\theta}})$ be a vertex of $\mathcal{P}$ with fractional leader coordinates. By Proposition 3, a cut-generating linear program (CGLP) can jointly select $(\boldsymbol{\beta}, \boldsymbol{\pi}, \pi_0)$ and the flow certificates, subject to a linear normalization fixing the coefficient scale. Its explicit formulation requires one flow certificate (6) per leader assignment, hence $|\mathcal{X}||\mathcal{A}|$ flow variables, and may be exponentially large. We instead construct the node potential coefficients $\boldsymbol{\beta}$ from the arc inequalities active at $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\theta}})$, preserving the existence of a separating cut while leaving the leader coefficients to be optimized.

Proposition 3 restricts admissible $\boldsymbol{\beta}$ to the supply cone $\mathcal{B}$, but does not determine which supplies admit separation at the candidate. For fixed $\boldsymbol{\beta} \in \mathcal{B} \setminus \{\mathbf{0}\}$, Corollary 1 gives the maximum attainable violation as

$$v_{\boldsymbol{\beta}}(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\theta}}) := \boldsymbol{\beta}^\top \bar{\boldsymbol{\theta}} - \hat{f}_{\boldsymbol{\beta}}(\bar{\boldsymbol{x}}). \tag{9}$$

A separating inequality with this $\boldsymbol{\beta}$ exists if and only if this value is positive.

For any $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\theta}}) \in \mathcal{P}$, define its active arcs by

$$\mathcal{A}_0 = \{a \in \mathcal{A} : D_a^\top \bar{\boldsymbol{\theta}} = q_a(\bar{\boldsymbol{x}})\}.$$

Connectivity and spanning trees below refer to the underlying undirected graph; each selected arc retains its original orientation.

Lemma 3. *For $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\theta}}) \in \mathcal{P}$, the active-arc graph $(\mathcal{V}, \mathcal{A}_0)$ is connected if and only if $\bar{\boldsymbol{\theta}}$ is a vertex of $\mathcal{P}(\bar{\boldsymbol{x}})$. In particular, it is connected at every vertex of $\mathcal{P}$.*

Figure 2 illustrates the connected active-arc graph at a fractional vertex of $\mathcal{P}$.

Our construction is based on a standard geometric observation: a positive combination of full-rank active inequality rows yields an optimization direction uniquely

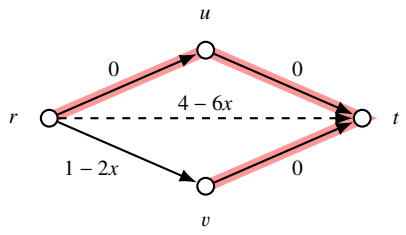


Fig. 2. Active arcs at a vertex of $\mathcal{P}$ with $\mathcal{X} = \{0, 1\}$ and $(\bar{x}, \bar{\boldsymbol{\theta}}) = (1/2, \mathbf{0})$. Arc labels give costs. Solid arcs are active, and the dashed arc has slack 1. The red highlight marks a spanning tree of active arcs. The four active equalities uniquely determine $(x, \theta_r, \theta_u, \theta_v)$.

maximized by the associated basic solution. We apply this observation to the fiber $\mathcal{P}(\boldsymbol{x})$. The network structure makes the required row selection inexpensive: after eliminating θ_t, a spanning tree of active arcs provides a minimum-cardinality full-rank collection of $|\mathcal{V}| - 1$ rows. Once the active-arc graph is available, such a tree can be found by graph traversal in $O(|\mathcal{V}| + |\mathcal{A}_0|)$ time, without matrix rank computations. Positive weights on its arcs then determine a supply vector $\boldsymbol{\beta}$; the following lemma establishes the resulting separation guarantee.

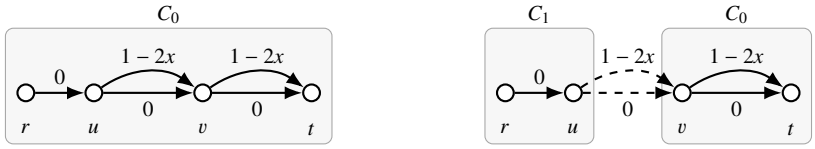
Lemma 4. *Suppose $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\theta}}) \in \mathcal{P}$ has a connected active-arc graph. Choose a spanning tree $T \subseteq \mathcal{A}_0$ and weights $w_a > 0$ for $a \in T$. Then*

$$\boldsymbol{\beta} = \sum_{a \in T} w_a D_a$$

belongs to $\mathcal{B} \setminus \{\mathbf{0}\}$. Moreover, $v_{\boldsymbol{\beta}}(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\theta}}) > 0$ if and only if $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\theta}}) \notin \operatorname{conv}(\mathcal{S})$. In particular, this construction guarantees separation at every vertex of $\mathcal{P}$ with fractional leader coordinates.

The weighted sum of the tree inequalities is valid for $\mathcal{P}$ and active at $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\theta}})$, so it does not itself separate. It selects the potential coefficients; separation comes from strengthening the affine right-hand side over the discrete leader set $\mathcal{X}$. The guarantee concerns the existence of a cut, not its strength relative to other supply vectors.

Let $\mathcal{P}^+$ be obtained from $\mathcal{P}$ by adding finitely many valid inequalities $(\boldsymbol{\beta}^j)^\top \boldsymbol{\theta} \leq \ell_j(\boldsymbol{x})$, with affine ℓ_j and $\boldsymbol{\beta}^j \in \mathcal{B}$. Unlike for $\mathcal{P}$, the active-arc graph at a vertex of $\mathcal{P}^+$ need not be connected. Consider first the case in which $(\mathcal{V}, \mathcal{A}_0)$ consists of two components, C_0 containing t and C_1. A spanning tree in each component fixes all differences of node potential variables, allowing only for a common shift of the potentials in C_1. An active added cut fixes this shift whenever its total supply $\sum_{u \in C_1} \beta^j_u$ is nonzero. At a vertex, at least one such cut must exist; otherwise, a small shift in either direction would preserve feasibility. Thus two trees and one active added cut suffice to determine the node potentials. Figure 3 illustrates how adding a cut can produce a new fractional vertex with two active-arc components. The following proposition extends this construction to any number of components.



$\bar{x} = 1/2, \quad \bar{\boldsymbol{\theta}} = (0, 0, 0, 0)$

(a) Vertex of $\mathcal{P}$

$\bar{x} = 1/2, \quad \bar{\boldsymbol{\theta}} = (-1, -1, 0, 0)$

(b) Vertex after adding $\theta_r \le -2x$

Fig. 3. Connected components before and after adding one cut, with $\mathcal{X} = \{0, 1\}$ and potentials ordered as (r, u, v, t). Solid arcs are active, dashed arcs are inactive, and shading identifies components of the original active-arc graph. The root cut is valid since $\phi(0) = 0$ and $\phi(1) = -2$. In (b), it is active and fixes the relative shift between C_1 and C_0.

Proposition 4. *Let $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\theta}})$ be a vertex of $\mathcal{P}^+$, let $C_0, \ldots, C_{c-1}$ be its active-arc components, with $t \in C_0$, and let $\mathcal{J}_0$ index its active added cuts. The matrix*

$$\Gamma_{j\ell} = \sum_{u \in C_\ell} \beta_u^j, \qquad j \in \mathcal{J}_0, \quad \ell = 1, \ldots, c-1,$$

has rank $c - 1$. Choose a spanning tree in each component, with union F, and a set $J \subseteq \mathcal{J}_0$ indexing $c - 1$ independent rows of Γ. The corresponding $|\mathcal{V}| - 1$ original arc inequalities and added cuts form a minimum-cardinality full-rank collection in the node potential coordinates, excluding θ_t. For any positive weights w_a and μ_j, the supply

$$\boldsymbol{\beta} = \sum_{a \in F} w_a D_a + \sum_{j \in J} \mu_j \boldsymbol{\beta}^j$$

belongs to $\mathcal{B} \setminus \{\mathbf{0}\}$. Moreover, $v_{\boldsymbol{\beta}}(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\theta}}) > 0$ if and only if $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\theta}}) \notin \operatorname{conv}(\mathcal{S})$. In particular, it admits a separating inequality whenever $\bar{\boldsymbol{x}}$ is fractional.

The network structure therefore keeps row selection inexpensive within each component: graph traversal selects the forest, and $c - 1$ linearly independent active cuts fix the remaining degrees of freedom. For $c = 1$, no added cut is needed and the spanning-tree construction applies directly.

On a practical setting, there is strong reason to start with root supplies before attempting the constructions suggested above. By Theorem 1, root cuts suffice to close the bilevel objective gap when the bilinear optimality constraint is retained. Moreover, each root cut has the node potential coefficients of an additional $r \to t$ arc, so adding root cuts preserves the spanning-tree construction on the augmented active graph (Lemma 4). Root cuts can therefore be pursued first while retaining an inexpensive way to select other supply vectors when needed.

4.2 Complexity of separation

Selecting a supply vector $\boldsymbol{\beta}$ that admits separation leaves the leader variable coefficients to be determined. For fixed $\boldsymbol{\beta} \in \mathcal{B} \setminus \{\mathbf{0}\}$ and a point $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\theta}}) \in \mathcal{P}$, Corollary 1 gives the

attained maximum violation as

$$\begin{aligned} v_{\boldsymbol{\beta}}(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\theta}}) = \max_{\boldsymbol{\pi}, \pi_0} \quad & \boldsymbol{\beta}^\top \bar{\boldsymbol{\theta}} - \boldsymbol{\pi}^\top \bar{\boldsymbol{x}} - \pi_0 \\ \text{s.t.} \quad & \boldsymbol{\pi}^\top \boldsymbol{x}^k + \pi_0 \geq f_{\boldsymbol{\beta}}(\boldsymbol{x}^k), \qquad k = 1, \ldots, K, \end{aligned} \tag{10}$$

where $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^K$ is an enumeration of the feasible points in $\mathcal{X}$. Each value $f_{\boldsymbol{\beta}}(\boldsymbol{x}^k)$ can be computed by a minimum-cost flow problem. Given these values, LP (10) has only $n+1$ variables, but still one constraint per leader assignment. Replacing each constraint by its flow certificate (6) gives an equivalent CGLP without precomputing the values. The resulting formulation can however still be exponential in size.

We propose two ways to address the exponential size of the CGLP. With respect to the number of leader assignments, we can, for example, consider a split disjunction on one of the fractional coordinates [1]. With respect to the size of the underlying graph, we can restrict the minimum-cost flow problem to the active arcs $\mathcal{A}_0$. For a vertex of $\mathcal{P}$ with fractional leader coordinates and a supply constructed by Lemma 4, these reductions also preserve the existence of a separating inequality. The two reductions can decrease the maximum violation compared with the original CGLP with fixed $\boldsymbol{\beta}$. When a linear description of $\operatorname{conv}(\mathcal{X})$ is supplied, the reduced CGLP has polynomial size in that description and the active graph.

To motivate the split relaxation, we also examine the computational complexity of the separation problem for a particular fixed supply vector on a small graph. By *exact separation*, we mean to determine occurrence of a violation, and if so, to find a violated cut with the fixed supply vector $\boldsymbol{\beta}$.

Proposition 5. *Membership in the root-cut closure $\mathcal{R}_r$ is NP-hard, even for candidates in $\mathcal{P}$ when $\mathcal{X} = \{0,1\}^n$ and the graph consists of two parallel $r \to t$ arcs, one with constant cost. Consequently, exact separation of root cuts and exact evaluation of $\hat{\phi}$ are NP-hard under the same restrictions.*

The proof, given in Appendix A, reduces Partition to membership at $\bar{\boldsymbol{x}} = \frac{1}{2}\mathbf{1}$. Both arc inequalities are active at the constructed candidate, so restricting to active arcs leaves the hardness construction intact. The result concerns arbitrary candidates in $\mathcal{P}$ and also establishes hardness of general fixed-supply separation.

Exact separation therefore remains NP-hard even with an unconstrained leader and a graph consisting of two parallel arcs. The split relaxation addresses a computational difficulty that persists after graph reduction. Combined with the supply construction of Lemma 4, it retains a separating inequality at fractional vertices while admitting a CGLP of polynomial size in the supplied leader formulation and the active graph. To see this, each branch of the split disjunction is described by the given linear description of $\operatorname{conv}(\mathcal{X})$, the retained arc inequalities, and one split equation. Dualizing cut validity on each disjunction branch introduces one multiplier per row, one coefficient-matching equation per original variable, and one inequality for the right-hand side. Since the disjunction has two branches, the resulting CGLP has size polynomial in the supplied leader description and the active graph.

5 Conclusions

In this work, motivated by network pricing, follower's dynamic programming formulations and certain interdiction models, we developed a polyhedral characterization of the bilevel 0–1 shortest-path problem by studying the interaction between leader decisions and follower node potentials. Network flows characterize the valid inequalities for this joint set. With the leader set convexified and the bilinear optimality constraint retained, root cuts strengthen the continuous relaxation of the problem and recover the bilevel optimum for every linear leader objective. Full convexification of the leader–potential set can require coupled supplies even after all unit-supply cuts are added.

We proposed cut generation by constructively fixing supplies. Spanning trees of active arcs yield separating supplies at vertices of $\mathcal{P}$ with fractional leader coordinates, and active-arc components extend this construction after cuts are added. Nonetheless, we showed that exact fixed-supply separation is NP-hard even for networks with two parallel arcs. For the initial relaxation, split disjunctions and restriction to active arcs reduce the CGLP while preserving separation with the constructed supplies.

Acknowledgments

This work was funded by the FRQNT Doctoral Scholarship No. B2X-367536, and the NSERC Grants No. 2024-04051 and 2024-05059.

OpenAI Codex (GPT-6) was used to assist with literature searches, LaTeX edits, and figure generation. The authors reviewed the content as needed and take full responsibility for this manuscript.

References

1. Balas, E.: Lift-and-project cuts for mixed 0–1 programs, pp. 79–89. Springer International Publishing, Cham (2018). https://doi.org/10.1007/978-3-030-00148-3_6
2. Bergman, D., Cire, A.A., Van Hoeve, W.J., Hooker, J.: Decision diagrams for optimization. Artificial Intelligence: Foundations, Theory, and Algorithms, Springer International Publishing, Cham (2016). https://doi.org/10.1007/978-3-319-42849-9
3. Borgwardt, S., Finhold, E., Hemmecke, R.: Quadratic diameter bounds for dual network flow polyhedra. Mathematical Programming **159**(1), 237–251 (2016). https://doi.org/10.1007/s10107-015-0956-4
4. Brotcorne, L., Labbé, M., Marcotte, P., Savard, G.: A bilevel model for toll optimization on a multicommodity transportation network. Transportation science **35**(4), 345–358 (2001)
5. Brotcorne, L., Labbé, M., Marcotte, P., Savard, G.: Joint design and pricing on a network. Operations Research **56**(5), 1104–1115 (2008). https://doi.org/10.1287/opre.1080.0617
6. Bui, Q.M., Carvalho, M., Neto, J.: Solving combinatorial pricing problems using embedded dynamic programming models. INFORMS Journal on Computing **38**(2), 548–567 (2026). https://doi.org/10.1287/ijoc.2024.0686
7. Bui, Q.M., Gendron, B., Carvalho, M.: A catalog of formulations for the network pricing problem. INFORMS Journal on Computing **34**(5), 2658–2674 (2022). https://doi.org/10.1287/ijoc.2022.1198

8. Caprara, A., Carvalho, M., Lodi, A., Woeginger, G.J.: Bilevel knapsack with interdiction constraints. INFORMS Journal on Computing **28**(2), 319–333 (May 2016). https://doi.org/10.1287/ijoc.2015.0676
9. Davarnia, D., Richard, J.P.P., Tawarmalani, M.: Simultaneous convexification of bilinear functions over polytopes with application to network interdiction. SIAM Journal on Optimization **27**(3), 1801–1833 (2017). https://doi.org/10.1137/16M1066166
10. Dempe, S.: Bilevel optimization: theory, algorithms, applications and a bibliography. In: Dempe, S., Zemkoho, A. (eds.) Bilevel Optimization, Springer Optimization and Its Applications, vol. 161, pp. 581–672. Springer International Publishing, Cham (2020). https://doi.org/10.1007/978-3-030-52119-6_20
11. DeNegre, S.: Interdiction and discrete bilevel linear programming. Ph.D. thesis, Lehigh University, USA (2011)
12. Fischetti, M., Ljubić, I., Monaci, M., Sinnl, M.: A new general-purpose algorithm for mixed-integer bilevel linear programs. Operations Research **65**(6), 1615–1637 (2017)
13. Fischetti, M., Ljubić, I., Monaci, M., Sinnl, M.: On the use of intersection cuts for bilevel optimization. Mathematical Programming **172**(1), 77–103 (2018)
14. Fischetti, M., Ljubić, I., Monaci, M., Sinnl, M.: Interdiction games and monotonicity, with application to knapsack problems. INFORMS Journal on Computing **31**(2), 390–410 (2019). https://doi.org/10.1287/ijoc.2018.0831
15. Karp, R.M.: Reducibility among combinatorial problems. In: Miller, R.E., Thatcher, J.W., Bohlinger, J.D. (eds.) Complexity of Computer Computations: Proceedings of a Symposium on the Complexity of Computer Computations, pp. 85–103. Springer US, Boston, MA (1972). https://doi.org/10.1007/978-1-4684-2001-2_9
16. Kleinert, T., Labbé, M., Ljubić, I., Schmidt, M.: A survey on mixed-integer programming techniques in bilevel optimization. EURO Journal on Computational Optimization **9**, 100007 (2021). https://doi.org/10.1016/j.ejco.2021.100007
17. Labbé, M., Marcotte, P., Savard, G.: A bilevel model of taxation and its application to optimal highway pricing. Management Science **44**(12-part-1), 1608–1622 (Dec 1998). https://doi.org/10.1287/mnsc.44.12.1608
18. Lozano, L., Bergman, D., Cire, A.A.: Constrained shortest-path reformulations via decision diagrams for structured two-stage optimization problems (Jul 2024). https://doi.org/10.48550/arXiv.2206.12962
19. Lozano, L., Bergman, D., Cire, A.A.: Network relaxations for combinatorial bilevel optimization under linear interactions. Operations Research (Aug 2026). https://doi.org/10.1287/opre.2025.2457
20. Tahernejad, S., Ralphs, T.K.: Valid inequalities for mixed integer bilevel linear optimization problems. arXiv preprint arXiv:2510.02998 (2025). https://doi.org/10.48550/arXiv.2510.02998
21. Vásquez, S., Lozano, L., van Hoeve, W.J.: A single-level reformulation of binary bilevel programs using decision diagrams. Mathematical Programming (Dec 2025). https://doi.org/10.1007/s10107-025-02294-1
22. Vásquez, S., Lozano, L., van Hoeve, W.J.: Network flow-based cuts for binary bilevel programs. arXiv preprint arXiv:2610.39261 (2026). https://doi.org/10.48550/arXiv.2610.39261

A Deferred Proofs

A.1 Polyhedrality of the single-level feasible hull

Proof of Proposition 1. Fix $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) \in \mathcal{X} \times \mathcal{Y}$ for which the potential constraints in (3) are feasible. Weak duality and the optimality constraint give $\theta_r = \boldsymbol{q}(\hat{\boldsymbol{x}})^\top \hat{\boldsymbol{y}}$. For each node u,

choose an $r \rightsquigarrow u$ path ρ_u and a $u \rightsquigarrow t$ path τ_u, allowing the empty path at an endpoint. Summing the dual arc inequalities along these paths yields

$$\boldsymbol{q}(\hat{\boldsymbol{x}})^\top \hat{\boldsymbol{y}} - \sum_{a \in \rho_u} q_a(\hat{\boldsymbol{x}}) \le \theta_u \le \sum_{a \in \tau_u} q_a(\hat{\boldsymbol{x}}).$$

Thus the feasible potentials for each fixed pair form a bounded polyhedron. Since $\mathcal{X}$ and $\mathcal{Y}$ are finite, Ω is a finite union of polytopes, and its convex hull is a polytope.

A.2 Effect of convexifying the dual network polyhedron

Proof of Proposition 2. It suffices to show $\Omega \subseteq \mathcal{E} \subseteq \operatorname{conv}(\Omega)$. Taking the convex hull gives $\operatorname{conv}(\mathcal{E}) = \operatorname{conv}(\Omega)$.

The inclusion $\Omega \subseteq \mathcal{E}$ follows from $\mathcal{S} \subseteq \operatorname{conv}(\mathcal{S})$ and $\mathcal{Y} \subseteq \operatorname{conv}(\mathcal{Y})$. For the reverse containment in $\operatorname{conv}(\Omega)$, take $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\theta}}, \bar{\boldsymbol{y}}) \in \mathcal{E}$ and choose finite convex representations

$$(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\theta}}) = \sum_i \lambda_i (\boldsymbol{x}^i, \boldsymbol{\theta}^i), \qquad \bar{\boldsymbol{y}} = \sum_j \mu_j \boldsymbol{y}^j,$$

where $(\boldsymbol{x}^i, \boldsymbol{\theta}^i) \in \mathcal{S}$, $\boldsymbol{y}^j \in \mathcal{Y}$, $\lambda_i, \mu_j \ge 0$, and $\sum_i \lambda_i = \sum_j \mu_j = 1$. Since the path set $\mathcal{Y}$ is independent of the leader assignment, every pair (i, j) combines a feasible follower path with dual-feasible potentials at $\boldsymbol{x}^i$. Weak duality then gives us $\boldsymbol{q}(\boldsymbol{x}^i)^\top \boldsymbol{y}^j - \theta^i_r \ge 0$ for each pair (i, j). Because the candidate solution satisfies the follower's optimality constraint and $\boldsymbol{q}(\boldsymbol{x})$ is affine, we have

$$0 \ge \boldsymbol{q}(\bar{\boldsymbol{x}})^\top \bar{\boldsymbol{y}} - \bar{\theta}_r = \sum_{i,j} \lambda_i \mu_j (\boldsymbol{q}(\boldsymbol{x}^i)^\top \boldsymbol{y}^j - \theta^i_r) \ge 0.$$

For every (i, j) pair with $\lambda_i \mu_j > 0$ the optimality gap must vanish, that is, $\boldsymbol{q}(\boldsymbol{x}^i)^\top \boldsymbol{y}^j - \theta^i_r = 0$. Hence each positively weighted triple $(\boldsymbol{x}^i, \boldsymbol{\theta}^i, \boldsymbol{y}^j)$ belongs to Ω. Finally, the weights $\lambda_i \mu_j$ sum to one and satisfy

$$(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\theta}}, \bar{\boldsymbol{y}}) = \sum_{i,j} \lambda_i \mu_j (\boldsymbol{x}^i, \boldsymbol{\theta}^i, \boldsymbol{y}^j),$$

proving $\mathcal{E} \subseteq \operatorname{conv}(\Omega)$.

A.3 Flow characterization and fixed-supply convexification

Proof of Proposition 3. We first verify (5) and supply admissibility. For any $\boldsymbol{x} \in \operatorname{conv}(\mathcal{X})$, the potential maximization is feasible, and its linear programming dual is

$$\min\{\boldsymbol{q}(\boldsymbol{x})^\top \boldsymbol{\lambda} : D\boldsymbol{\lambda} + \mu \boldsymbol{e}_t = \boldsymbol{\beta},\ \boldsymbol{\lambda} \ge \boldsymbol{0},\ \mu \in \mathbb{R}\}.$$

Summing the balance equations gives $\mu = 0$ because $\sum_{u \in \mathcal{V}} \beta_u = 0$. Hence dual feasibility is equivalent to $\boldsymbol{\beta} \in \mathcal{B}$. If (4) is valid, its right-hand side bounds the potential objective on any fiber indexed by $\boldsymbol{x} \in \mathcal{X}$. Such a fiber exists because $\mathcal{X} \neq \emptyset$, so linear programming duality implies $\boldsymbol{\beta} \in \mathcal{B}$.

For $\boldsymbol{\beta} \in \mathcal{B}$, the recession cone of $\Lambda(\boldsymbol{\beta})$ consists of nonnegative circulations. Acyclicity implies that the only such circulation is zero, so $\Lambda(\boldsymbol{\beta})$ is a nonempty polytope. Strong duality now gives (5), with both optima finite and attained. Maximizing the left-hand side of (4) on each integer fiber proves the affine-upper-bound characterization. An optimal flow on that fiber supplies (6) whenever the upper bound condition holds. Conversely, any such certificate gives

$$\boldsymbol{\beta}^\top \boldsymbol{\theta} = (\boldsymbol{\lambda}^{\boldsymbol{x}})^\top D^\top \boldsymbol{\theta} \le \boldsymbol{q}(\boldsymbol{x})^\top \boldsymbol{\lambda}^{\boldsymbol{x}} \le \boldsymbol{\pi}^\top \boldsymbol{x} + \pi_0 \qquad \text{for all } \boldsymbol{\theta} \in \mathcal{S}(\boldsymbol{x}),$$

proving validity.

A.4 Root supply and the optimality gap

Proof of Lemma 1. We first determine the possible values of the root potential θ_r for a fixed leader decision $\boldsymbol{x} \in \operatorname{conv}(\mathcal{X})$. Assigning each node its shortest-path cost to t gives feasible potentials with $\theta_r = \phi(\boldsymbol{x})$. Since r has no incoming arcs, decreasing only θ_r preserves feasibility. Together with weak duality, this proves

$$\operatorname{Proj}_{\theta_r} \mathcal{P}(\boldsymbol{x}) = (-\infty, \phi(\boldsymbol{x})].$$

We next show that convexifying $\mathcal{S}$ replaces this upper bound by the concave envelope of the values at integer leader decisions. Let $\boldsymbol{x}^1, \dots, \boldsymbol{x}^K$ be an enumeration of $\mathcal{X}$ and fix $\bar{\boldsymbol{x}} \in \operatorname{conv}(\mathcal{X})$. The concave envelope has the representation

$$\hat{\phi}(\bar{\boldsymbol{x}}) = \max \left\{ \sum_{k=1}^{K} \alpha_k \phi(\boldsymbol{x}^k) : \sum_{k=1}^{K} \alpha_k \boldsymbol{x}^k = \bar{\boldsymbol{x}},\ \sum_{k=1}^{K} \alpha_k = 1,\ \boldsymbol{\alpha} \ge \mathbf{0} \right\}. \tag{11}$$

Indeed, the right-hand side is a concave upper bound of the values on $\mathcal{X}$, and every such upper bound dominates it by concavity. The maximum is attained because the feasible weight set is nonempty and compact and each $\phi(\boldsymbol{x}^k)$ is finite. Every convex combination of points of $\mathcal{S}$ with leader component $\bar{\boldsymbol{x}}$ has root potential at most the corresponding weighted sum of follower values, and hence at most $\hat{\phi}(\bar{\boldsymbol{x}})$. To attain this bound, choose weights α_k maximizing (11) and, for each k, let $\boldsymbol{\theta}^k$ be the shortest-path potentials for arc costs $\boldsymbol{q}(\boldsymbol{x}^k)$. Then $(\boldsymbol{x}^k, \boldsymbol{\theta}^k) \in \mathcal{S}$ and $\theta^k_r = \phi(\boldsymbol{x}^k)$, so

$$(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\theta}}) := \sum_k \alpha_k (\boldsymbol{x}^k, \boldsymbol{\theta}^k) \in \operatorname{conv}(\mathcal{S}), \qquad \bar{\theta}_r = \hat{\phi}(\bar{\boldsymbol{x}}).$$

Decreasing every θ^k_r by the same nonnegative amount preserves membership in $\mathcal{S}$ and yields every smaller value of $\bar{\theta}_r$. This proves the asserted projection of $\operatorname{conv}(\mathcal{S})$.

Finally, we compare this projection with that of the root closure. Since ϕ is concave, $\hat{\phi} \le \phi$ on $\operatorname{conv}(\mathcal{X})$. The first part of the proof therefore shows that imposing $\theta_r \le \hat{\phi}(\boldsymbol{x})$ on $\mathcal{P}$ leaves precisely the root potential values $(-\infty, \hat{\phi}(\boldsymbol{x})]$ for each $\boldsymbol{x} \in \operatorname{conv}(\mathcal{X})$. By (7), this proves the equality for $\mathcal{R}_r$.

Proof of Theorem 1. Take $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\theta}}, \bar{\boldsymbol{y}}) \in \mathcal{E}_r$. Choose weights α_k attaining (11) and write $\bar{\boldsymbol{y}} = \sum_j \mu_j \boldsymbol{y}^j$ with $\boldsymbol{y}^j \in \mathcal{Y}$, $\mu_j \ge 0$, and $\sum_j \mu_j = 1$. Feasibility and the affinity of $\boldsymbol{q}$ give

$$\begin{aligned} \hat{\phi}(\bar{\boldsymbol{x}}) &\ge \bar{\theta}_r \ge \boldsymbol{q}(\bar{\boldsymbol{x}})^\top \bar{\boldsymbol{y}} \\ &= \sum_{k,j} \alpha_k \mu_j \boldsymbol{q}(\boldsymbol{x}^k)^\top \boldsymbol{y}^j \ge \sum_k \alpha_k \phi(\boldsymbol{x}^k) = \hat{\phi}(\bar{\boldsymbol{x}}). \end{aligned}$$

All inequalities are equalities. Each difference $\boldsymbol{q}(\boldsymbol{x}^k)^\top \boldsymbol{y}^j - \phi(\boldsymbol{x}^k)$ is nonnegative, so it vanishes whenever $\alpha_k \mu_j > 0$. Every positively weighted pair $(\boldsymbol{x}^k, \boldsymbol{y}^j)$ is therefore bilevel feasible, and

$$(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}) = \sum_{k,j} \alpha_k \mu_j (\boldsymbol{x}^k, \boldsymbol{y}^j).$$

This proves the convex-combination claim and bounds every linear leader objective below by the bilevel optimum. Conversely, $\Omega \subseteq \mathcal{E}_r$ because root cuts are valid for $\mathcal{S}$. The optimal values consequently coincide.

A.5 Reduction to the dual network polyhedron

Proof of Lemma 2. Let $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\theta}}, \bar{\boldsymbol{y}})$ be an extreme point of $\mathcal{R}$. Decompose $\bar{\boldsymbol{y}}$ into path-incidence vectors. Weak duality gives $\boldsymbol{q}(\bar{\boldsymbol{x}})^\top \boldsymbol{y} \geq \bar{\theta}_r$ for each path, and the optimality constraint forces equality in their weighted average. Every positively weighted path therefore gives a feasible triple with the same $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\theta}})$. Extremality forces all those paths to equal $\bar{\boldsymbol{y}}$, proving $\bar{\boldsymbol{y}} \in \mathcal{Y}$.

Now suppose $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\theta}})$ is a nontrivial convex combination of two points of $\mathcal{P}$. With $\bar{\boldsymbol{y}}$ fixed, the gap $\boldsymbol{q}(\boldsymbol{x})^\top \bar{\boldsymbol{y}} - \theta_r$ is affine and nonnegative on $\mathcal{P}$. It vanishes at the candidate, so it vanishes at both endpoints. Appending $\bar{\boldsymbol{y}}$ to these endpoints gives a nontrivial convex representation in $\mathcal{R}$, a contradiction.

A.6 Constructive supply selection

Proof of Lemma 3. If the active-arc graph of a point $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\theta}}) \in \mathcal{P}$ is connected, a displacement of node potentials preserving its arc equalities must be constant on all nodes. The normalization at t makes that displacement zero. Hence the active equalities uniquely determine the node potentials from $\bar{\boldsymbol{x}}$, proving that $\bar{\boldsymbol{\theta}}$ is a vertex of the fiber $\mathcal{P}(\bar{\boldsymbol{x}})$. If the graph is disconnected, shift every node potential in a component not containing t by ε or by $-\varepsilon$. Active equalities are preserved, and the finitely many inactive arc inequalities remain feasible for sufficiently small $\varepsilon > 0$. These two distinct feasible node potential vectors have midpoint $\bar{\boldsymbol{\theta}}$, contradicting fiber extremality. Finally, a nontrivial convex representation within a fiber also gives one in $\mathcal{P}$, so every vertex of $\mathcal{P}$ is a fiber vertex.

Proof of Lemma 4. The vector $\boldsymbol{\beta}$ is the supply of the nonnegative flow with value w_a on T and zero elsewhere. It is nonzero because a DAG has no nonzero nonnegative circulation. Summing the selected inequalities gives the valid inequality

$$\boldsymbol{\beta}^\top \boldsymbol{\theta} \leq \ell(\boldsymbol{x}), \qquad \ell(\boldsymbol{x}) = \sum_{a \in T} w_a q_a(\boldsymbol{x}),$$

with equality at the candidate.

The node potentials $\bar{\boldsymbol{\theta}}$ uniquely maximize $\boldsymbol{\beta}^\top \boldsymbol{\theta}$ on $\mathcal{P}(\bar{\boldsymbol{x}})$. To see this, fix $\boldsymbol{x} = \bar{\boldsymbol{x}}$ and take any $\boldsymbol{\theta} \in \mathcal{P}(\bar{\boldsymbol{x}})$. The slack of the aggregated inequality is

$$\ell(\bar{\boldsymbol{x}}) - \boldsymbol{\beta}^\top \boldsymbol{\theta} = \sum_{a \in T} w_a \big(q_a(\bar{\boldsymbol{x}}) - D_a^\top \boldsymbol{\theta}\big) \geq 0.$$

The candidate $\bar{\boldsymbol{\theta}}$ attains the upper bound $\ell(\bar{\boldsymbol{x}})$ because every arc of T is active there. Any potential vector $\boldsymbol{\theta}$ attaining the same bound must make every slack in the sum zero, since the slacks are nonnegative and all weights are strictly positive. Hence $D_a^\top(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}) = 0$ for every $a \in T$, so the displacement $\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}$ has equal values at the endpoints of every tree arc. Connectivity makes this displacement constant on all nodes, and $\theta_t = \bar{\theta}_t = 0$ makes the constant zero. In other words, $\boldsymbol{\theta}$ must be equal to $\overline{\boldsymbol{\theta}}$.

Applying the convex-combination representation in (11) to $f_{\boldsymbol{\beta}}$ gives the attained identity

$$\hat{f}_{\boldsymbol{\beta}}(\bar{\boldsymbol{x}}) = \max\{\boldsymbol{\beta}^\top \boldsymbol{\theta} : (\bar{\boldsymbol{x}}, \boldsymbol{\theta}) \in \operatorname{conv}(\mathcal{S})\}. \tag{12}$$

Indeed, any convex combination of points of $\mathcal{S}$ has objective at most the corresponding weighted integer-fiber values, and maximizing weights together with maximizing fiber potentials attain equality. Since $\operatorname{conv}(\mathcal{S}) \subseteq \mathcal{P}$, (12) gives $\hat{f}_{\boldsymbol{\beta}}(\bar{\boldsymbol{x}}) \le \boldsymbol{\beta}^\top \bar{\boldsymbol{\theta}}$. Equality holds if the candidate belongs to $\operatorname{conv}(\mathcal{S})$. Conversely, if equality holds, an attaining hull potential also maximizes the objective $\boldsymbol{\beta}^\top \boldsymbol{\theta}$ over $\mathcal{P}(\bar{\boldsymbol{x}})$, so uniqueness makes it $\bar{\boldsymbol{\theta}}$.

The difference $\boldsymbol{\beta}^\top \bar{\boldsymbol{\theta}} - \hat{f}_{\boldsymbol{\beta}}(\bar{\boldsymbol{x}})$ is the maximum attainable violation $v_{\boldsymbol{\beta}}(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\theta}})$ defined in (9). We have shown that this difference is nonnegative and vanishes precisely when the candidate belongs to $\operatorname{conv}(\mathcal{S})$. A vertex of $\mathcal{P}$ in $\operatorname{conv}(\mathcal{S})$ must itself belong to $\mathcal{S}$, since all positively weighted points in a convex representation from $\mathcal{S} \subseteq \mathcal{P}$ must equal that vertex. A vertex with fractional leader coordinates is therefore outside $\operatorname{conv}(\mathcal{S})$, proving the final assertion.

Proof of Proposition 4. A displacement preserving all active arc equalities is constant on each component C_ℓ and zero on C_0 after imposing $\theta_t = 0$. Write its remaining component values as $\delta_1, \dots, \delta_{c-1}$. It preserves the active added cuts precisely when $\Gamma \boldsymbol{\delta} = \mathbf{0}$. If Γ had rank less than $c - 1$, a nonzero such displacement of sufficiently small magnitude in either direction would preserve every constraint of $\mathcal{P}^+$, contradicting the vertex assumption. Thus $\operatorname{rank} \Gamma = c - 1$.

The selected forest has $|\mathcal{V}| - c$ arcs and forces componentwise constancy. The selected $c - 1$ rows of Γ force all remaining component shifts to vanish. Consequently, the $|\mathcal{V}| - 1$ selected rows have full rank in the free potential coordinates, which is the minimum possible cardinality. This also covers $c = 1$, when no added row is needed.

Each selected potential coefficient belongs to $\mathcal{B}$, so their positive combination $\boldsymbol{\beta}$ also belongs to $\mathcal{B}$. Choose a nonnegative flow representing each selected cut coefficient and aggregate these flows with the selected arc flows using the prescribed positive weights. Full rank ensures that at least one selected coefficient is nonzero, so the aggregate flow is nonzero. Acyclicity precludes it from being a circulation, proving $\boldsymbol{\beta} \neq \mathbf{0}$.

The weighted sum of the selected inequalities is

$$\boldsymbol{\beta}^\top \boldsymbol{\theta} \le \ell(\boldsymbol{x}), \qquad \ell(\boldsymbol{x}) = \sum_{a \in F} w_a q_a(\boldsymbol{x}) + \sum_{j \in J} \mu_j \ell_j(\boldsymbol{x}).$$

It is valid for $\mathcal{P}^+$ and active at the candidate. Equality in the fiber at $\bar{\boldsymbol{x}}$ forces every selected inequality to be active, and full rank makes $\bar{\boldsymbol{\theta}}$ its unique solution. Hence $\bar{\boldsymbol{\theta}}$ uniquely maximizes $\boldsymbol{\beta}^\top \boldsymbol{\theta}$ over $\mathcal{P}^+(\bar{\boldsymbol{x}})$. Because $\operatorname{conv}(\mathcal{S}) \subseteq \mathcal{P}^+$, the attained identity (12) gives exactly the same zero-violation equivalence as in the preceding proof. Finally, a vertex of $\mathcal{P}^+$ with fractional leader coordinates cannot belong to $\operatorname{conv}(\mathcal{S})$, since $\mathcal{S} \subseteq \mathcal{P}^+$ and its leader coordinates are binary.

A.7 Complexity of separation

Proof of Proposition 5. Assume rational data with binary encoding. Let $s_1, \dots, s_n$ be positive integer weights. The PARTITION problem [15] asks whether there exists a subset $I \subseteq \{1, \dots, n\}$ such that

$$\sum_{i \in I} s_i = \sum_{i \notin I} s_i.$$

Write $W = \sum_{i=1}^{n} s_i$. Set $\mathcal{X} = \{0,1\}^n$ and give the two parallel arcs costs

$$q_{a_1}(\boldsymbol{x}) = 2\boldsymbol{s}^\top \boldsymbol{x} - W, \qquad q_{a_2}(\boldsymbol{x}) = 0.$$

The point $\bar{\boldsymbol{x}} = \frac{1}{2}\mathbf{1}$, $\bar{\theta}_r = \bar{\theta}_t = 0$ belongs to $\mathcal{P}$, with both arc inequalities active. Its root-supply function is $\phi(\boldsymbol{x}) = \min\{2\boldsymbol{s}^\top \boldsymbol{x} - W, 0\}$, so $\hat{\phi} \le 0$. The candidate belongs to $\mathcal{R}_r$ if and only if $\hat{\phi}(\bar{\boldsymbol{x}}) = 0$.

If $\boldsymbol{x}^* \in \{0,1\}^n$ satisfies $2\boldsymbol{s}^\top \boldsymbol{x}^* = W$, then its complement also does, and both have follower value zero. Their average is $\bar{\boldsymbol{x}}$, so (11) gives $\hat{\phi}(\bar{\boldsymbol{x}}) = 0$. Conversely, suppose this envelope value is zero and take attaining weights in (11). Since all integer follower values are nonpositive, every positively weighted assignment satisfies $\phi(\boldsymbol{x}^k) = 0$, hence $2\boldsymbol{s}^\top \boldsymbol{x}^k \ge W$. Their weighted average satisfies

$$\sum_{k=1}^{K} \alpha_k \, 2\boldsymbol{s}^\top \boldsymbol{x}^k = 2\boldsymbol{s}^\top \bar{\boldsymbol{x}} = W.$$

Every positively weighted assignment therefore satisfies equality and solves the PARTITION instance with weights $(s_1, \dots, s_n)$. The construction has polynomial encoding length, and PARTITION is a well-known NP-complete problem, proving NP-hardness of membership under the stated restrictions. An exact root-cut separator decides this membership by reporting whether a violated cut exists, and exact evaluation of $\hat{\phi}$ decides it by comparison with zero. Both tasks are therefore NP-hard as well.