Near-optimal robust bilevel optimization

Mathieu Besançon · Miguel F. Anjos ·
Luce Brotcorne

Abstract Bilevel optimization problems embed the optimality conditions of
a sub-problem into the constraints of another optimization problem. We intro-
duce the concept of near-optimality robustness for bilevel problems, protecting
the upper-level solution feasibility from limited deviations at the lower level.
General properties and necessary conditions for the existence of solutions are
derived for near-optimal robust versions of generic bilevel problems. A duality-
based solution method is defined when the lower level is convex, leveraging the
methodology from the robust and bilevel literature. Numerical results assess
the efficiency of the proposed algorithm and the impact of valid inequalities
on the solution time.

Keywords bilevel optimization, robust optimization, game theory, bounded
rationality, duality, bilinear constraints, extended formulation

Mathematics Subject Classification (2010) 90C33 · 90C46 · 91A65 ·
90C26 · 90C34 ·

1 Introduction

Bilevel optimization problems embed the optimality conditions of a sub-problem
into the constraints of another one. They can model various decision-making
problems such as Stackelberg or leader-follower games, market equilibria, or

Mathieu Besançon
Department of Mathematics & Industrial Engineering,
Polytechnique Montréal, Montréal, QC, Canada H3C 3A7
& INRIA Lille Nord-Europe & Centrale Lille
E-mail: mathieu.besancon@polymtl.ca

Miguel F. Anjos
School of Mathematics, University of Edinburgh, James Clerk Maxwell Building, Peter
Guthrie Tait Road, Edinburgh EH9 3FD, UK

Luce Brotcorne
INOCS, INRIA Lille Nord-Europe, 40 avenue Halley, 59650 Villeneuve-d'Ascq, France

pricing and revenue management. A review of methods and applications of bilevel problems is presented in [1]. In the classical bilevel setting, when optimizing its objective function, the upper level anticipates an optimal reaction of the lower level to its decisions. However, in many practical cases, the lower level can make near-optimal decisions [2]. An important issue in this setting is the definition of the robustness of the upper-level decisions with respect to near-optimal lower-level solutions.

In some engineering applications [3,4,5], the decision-maker optimizes an outcome over a dynamical system (modelled as the lower level). For stable systems, the rate of change of the state variables decreases as the system converges towards the minimum of a potential function. If the system is stopped before reaching the minimum, the designer of the system would require that the upper-level constraints be feasible for near-optimal lower-level solutions.

The concept of bounded rationality initially proposed in [6], sometimes referred as ε-rationality [7], defines an economic and behavioural interpretation of a decision-making process where an agent aims to take *any* solution associated with a "satisfactory" objective value instead of the optimal one.

Protecting the upper level from a violation of its constraints by deviations of the lower level is a form of robust optimization. Indeed, it corresponds to a protection of some constraints against uncertain parameters of the problem. Therefore, we use the terms "near-optimality robustness" and "near-optimal robust bilevel problem" or *NORBiP* in the rest of the paper.

The introduction of uncertainty and robustness in games has been approached from different points of view in the literature. In [8], the authors prove the existence of robust counterparts of Nash equilibria under standard assumptions for simultaneous games without the knowledge of probability distributions associated with the uncertainty. In [9], the robust version of a network congestion problem is developed. Users are assumed to make decisions under bounded rationality, leading to a robust Wardrop equilibrium. Robust versions of bilevel problems modelling specific Stackelberg games have been studied in [10,11], using robust formulations to protect the leader against non-rationality or partial rationality of the follower. A stochastic version of the pessimistic bilevel problem is studied in [12], where the realization of the random variable occurs after the upper level and before the lower level. The authors then derive lower and upper bounds on the pessimistic and optimistic versions of the stochastic bilevel problem as MILPs, leveraging an exact linearization by assuming the upper-level variables are all binary. The models developed in [13] and [14] explore different forms of bounded or partial rationality of the lower level, where the lower level either makes a decision using an imperfect algorithm or may deviate from its optimal value in a way that penalizes the objective of the upper level.

Solving bilevel problems under limited deviations of the lower-level response was introduced in [2] under the term "ε-approximation" of the pessimistic bilevel problem. The authors focus on the independent case, *i.e.* cases where the lower-level feasible set is independent of the upper-level decisions.

Problems in such settings are shown to be simpler to handle than the dependent case and can be solved in polynomial time when the lower-level problem is linear under the optimistic and pessimistic assumptions. A custom algorithm is designed for the independent case, solving a sequence of non-convex non-linear problems relying on global optimization solvers.

We consider bilevel problems involving upper- and lower-level variables in the constraints and objective functions at both levels, thus more general than the independent "ε-approximation" from [2]. Unlike the independent case, the dependent bilevel problem is $\mathcal{NP}$-hard even when the constraints and objectives are linear. By defining the uncertainty in terms of a deviation from optimality of the lower level, our formulation offers a novel interpretation of robustness for bilevel problems and Stackelberg games. In the case of a linear lower level, we derive an exact MILP reformulation while not requiring the assumption of pure binary upper-level variables.

The main contributions of the paper are:

1. The definition and formulation of the dependent near-optimal robust bilevel problem, resulting in a generalized semi-infinite problem and its interpretation as a special case of robust optimization applied to bilevel problems.
2. The study of duality-based reformulations of $NORBiP$ where the lower-level problem is convex conic or linear in Section 3, resulting in a finite-dimensional single-level optimization problem.
3. An extended formulation for the linear-linear $NORBiP$ in Section 4, linearizing the bilinear constraints of the single-level model using disjunctive constraints.
4. A solution algorithm for the linear-linear $NORBiP$ in Section 5 using the extended formulation and its implementation with several variants.

The paper is organized as follows. In Section 2, we define the concepts of near-optimal set and near-optimal robust bilevel problem. We study the near-optimal bilevel problems with convex and linear lower-level problems in Section 3 and Section 4 respectively. In both cases, the near-optimal robust bilevel problem can be reformulated as a single level. For a linear lower level, an extended formulation can be derived from the single-level problem. A solution algorithm is provided and computational experiments are conducted for the linear case in Section 5, comparing the extended formulation to the compact one and studying the impact of valid inequalities. Finally, in Section 6 we draw some conclusions and highlight research perspectives on near-optimality robustness.

2 Near-optimal set and near-optimal robust bilevel problem

In this section, we first define the near-optimal set of the lower level and near-optimality robustness for bilevel problems. Next, we illustrate the concepts

on an example and highlight several properties of general near-optimal robust bilevel problems before focusing on the convex and linear cases in the following sections.

The generic bilevel problem is classically defined as:

$$\min_{x} F(x, v) \tag{1a}$$

$$\text{s.t. } G_k(x, v) \leq 0 \qquad\qquad\qquad\qquad \forall k \in [\![m_u]\!] \tag{1b}$$

$$x \in \mathcal{X} \tag{1c}$$

$$v \in \arg\min_{y \in \mathcal{Y}} \{f(x, y) \text{ s.t. } g_i(x, y) \leq 0 \,\forall i \in [\![m_l]\!]\}. \tag{1d}$$

The upper- and lower-level objective functions are noted $F, f : \mathcal{X} \times \mathcal{Y} \mapsto \mathbb{R}$ respectively. Constraint (1b) and $g_i(x, y) \leq 0 \,\forall i \in [\![m_l]\!]$ are the upper- and lower-level constraints respectively. In this section, we assume that $\mathcal{Y} = \mathbb{R}^{n_l}$ in order that the lower-level feasible set can be only determined by the g_i functions. The optimal value function $\phi(x)$ is defined as follows:

$$\phi : \mathbb{R}^{n_u} \to \{-\infty\} \cup \mathbb{R} \cup \{+\infty\}$$

$$\phi(x) = \min_{y} \{f(x, y) \text{ s.t. } g(x, y) \leq 0\}. \tag{2}$$

To keep the notation succinct, the indices of the lower-level constraints g_i are omitted when not needed as in Constraint (2). Throughout the paper, it is assumed that the lower-level problem is feasible and bounded for any given upper-level decision.

When, for a feasible upper-level decision, the solution to the lower-level problem is not unique, the bilevel problem is not well defined and further assumptions are required [1]. In the optimistic case, we assume that the lower level selects the optimal solution favouring the upper level and the optimal solution disfavouring them the most in the pessimistic case. We refer the reader to [15, Chapter 1] for further details on these two approaches. The near-optimal set of the lower level $\mathcal{Z}(x; \delta)$ is defined for a given upper-level decision x and tolerance δ as:

$$\mathcal{Z}(x; \delta) = \{y \mid g(x, y) \leq 0, \ f(x, y) \leq \phi(x) + \delta\}.$$

A Near-Optimal Robust Bilevel Problem, *NORBiP*, of parameter δ is defined as a bilevel problem where the upper-level constraints are satisfied for any lower-level solution z in the near-optimal set $\mathcal{Z}(x; \delta)$.

$$\min_{x,v} \ F(x,v) \tag{3a}$$

$$\text{s.t. } G_k(x,v) \leq 0 \qquad\qquad \forall k \in [\![m_u]\!] \tag{3b}$$

$$f(x,v) \leq \phi(x) \tag{3c}$$

$$g(x,v) \leq 0 \tag{3d}$$

$$G_k(x,z) \leq 0 \ \forall z \in \mathcal{Z}(x;\delta) \qquad\qquad \forall k \in [\![m_u]\!] \tag{3e}$$

$$x \in \mathcal{X}. \tag{3f}$$

Each k constraint in (3b) is satisfied if the corresponding constraint set in (3e) holds and is therefore redundant, since $v \in \mathcal{Z}(x;\delta)$. However, we mention Constraint (3b) in the formulation to highlight the structure of the initial bilevel problem in the near-optimal robust formulation.

The special case $\mathcal{Z}(x;0)$ is the set of optimal solutions to the original lower-level problem, NORBiP with $\delta = 0$ is therefore equivalent to the pessimistic bilevel problem as formulated in [2]:

$$f(x,y) \leq \phi(x) \ \forall y \in \mathcal{Z}(x;0).$$

For $\delta < 0$, $\mathcal{Z}(x;\delta)$ is the empty set, in which case Problem (3) is equivalent to the original optimistic bilevel problem while the set $\mathcal{Z}(x;\infty)$ corresponds to the complete lower-level feasible set, assuming the lower-level optimal solution is not unbounded for the given upper-level decision x.

Unlike the constraint-based pessimistic bilevel problem presented in [2], the upper-level objective $F(x,v)$ depends on both the upper- and lower-level variables, but is only evaluated with the optimistic lower-level variable v and not with a worst-case near-optimal solution. This implies the upper level chooses the best optimistic decision which protects its feasibility from near-optimal deviations. One implication for the modeller is that a near-optimal robust problem can be constructed directly from a bilevel instance where the objective function often depends on the variables of the two levels. Alternatively, the near-optimal robust formulation can protect both the upper-level objective value and constraints from near-optimal deviations of the lower level using an epigraph formulation introducing an additional variable:

$$\min_{x,v,\tau} \ \tau \tag{4a}$$

$$\text{s.t. } G_k(x,v) \leq 0 \qquad\qquad \forall k \in [\![m_u]\!] \tag{4b}$$

$$f(x,v) \leq \phi(x) \tag{4c}$$

$$g(x,v) \leq 0 \tag{4d}$$

$$F(x,z) \leq \tau \qquad\qquad \forall z \in \mathcal{Z}(x;\delta) \tag{4e}$$

$$G_k(x,z) \leq 0 \qquad \forall z \in \mathcal{Z}(x;\delta) \ \forall k \in [\![m_u]\!] \tag{4f}$$

$$x \in \mathcal{X}. \tag{4g}$$

The two models define different levels of conservativeness and risk. Indeed:

$$\text{opt}(1a\text{-}1d) \leq \text{opt}(3a\text{-}3f) \leq \text{opt}(4a\text{-}4g),$$

where $\text{opt}(P)$ denotes the optimal value of problem P. Both near-optimal robust formulations can be of interest to model decision-making applications. It can also be noted that Problem 3 includes the special case of opposite objectives between the two levels, i.e. problems for which $F(x,v) = -f(x,v)$. The two models offer different levels of conservativeness and risk and can both be of interest when modelling decision-making applications.

Constraint (3e) is a generalized semi-infinite constraint, based on the terminology from [16]. The dependence of the set of constraints $\mathcal{Z}(x;\delta)$ on the decision variables leads to the characterization of Problem (3) as a robust problem with decision-dependent uncertainty [17]. Each constraint in the set (3e) can be replaced by the corresponding worst-case second-level decision z_k obtained as the solution of the *adversarial problem*, parameterized by (x,v,δ):

$$z_k \in \arg\max_y \; G_k(x,y) \tag{5a}$$

$$\text{s.t. } f(x,y) \leq \phi(x) + \delta \tag{5b}$$

$$g(x,y) \leq 0. \tag{5c}$$

Finally, the near-optimal robust bilevel optimization problem can be expressed as:

$$\min_{x,v} \; F(x,v) \tag{6a}$$

$$\text{s.t. } f(x,v) \leq \phi(x) \tag{6b}$$

$$g(x,v) \leq 0 \tag{6c}$$

$$0 \geq \max_y \{G_k(x,y) \text{ s.t. } y \in \mathcal{Z}(x;\delta)\} \qquad \forall k \in [\![m_u]\!] \tag{6d}$$

$$x \in \mathcal{X}. \tag{6e}$$

In the robust optimization literature, models can present uncertainty on the constraints and/or on the objective function [18]. In bilevel optimization, the first case corresponds to *NORBiP*, where the impact of near-optimal lower-level solutions on the upper-level constraints is studied. The second case corresponds to the impact of near-optimal lower-level decisions on the upper-level objective value.

We next prove that the model including uncertainty on the objective, named Objective-Robust Near-Optimal Bilevel Problem (*ORNOBiP*), is a special case of *NORBiP*.

ORNOBiP is defined as:

$$\min_{x \in \mathcal{X}} \sup_{z \in \mathcal{Z}(x;\delta)} F(x,z) \tag{7a}$$

$$\text{s.t. } \mathcal{Z}(x;\delta) = \{y \text{ s.t. } g(x,y) \leq 0, f(x,y) \leq \phi(x) + \delta\}. \tag{7b}$$

In contrast to most objective-robust problem formulations, the uncertainty set $\mathcal{Z}$ depends on the upper-level solution x, qualifying Problem (7) as a problem with decision-dependent uncertainty.

Proposition 1 *ORNOBiP is a special case of NORBiP.*

Proof The reduction of the objective-uncertain robust problem to a constraint-uncertain robust formulation is detailed in [19]. In particular, Problem (7) is equivalent to:

$$\min_{x,\tau} \tau$$
$$\text{s.t. } x \in \mathcal{X}$$
$$\tau \geq F(x,z) \qquad\qquad \forall z \in \mathcal{Z}(x,\delta),$$

this formulation is a special case of *NORBiP*.   □

The pessimistic bilevel optimization problem defined in [20] is both a special case and a relaxation of *ORNOBiP*. For $\delta = 0$, the inner problem of *ORNOBiP* is equivalent to finding the worst lower-level decision with respect to the upper-level objective amongst the lower-level-optimal solutions. For any $\delta > 0$, the inner problem can select the worst solutions with respect to the upper-level objective that are not optimal for the lower level. The pessimistic bilevel problem is therefore a relaxation of *ORNOBiP*.

We illustrate the concept of near-optimal set and near-optimal robust solution with the following linear bilevel problem, represented in Fig. 1.

$$\min_{x,v} x \tag{8}$$
$$\text{s.t. } x \geq 0$$
$$v \geq 1 - \frac{x}{10}$$
$$v \in \arg\max_y \{y \text{ s.t. } y \leq 1 + \frac{x}{10}\}.$$

The high-point relaxation of Problem (8), obtained by relaxing the optimality constraint of the lower-level, while maintaining feasibility, is:

$$\min_{x,v} x$$
$$\text{s.t. } x \geq 0$$
$$v \geq 1 - \frac{x}{10}$$
$$v \leq 1 + \frac{x}{10}.$$

The shaded area in Fig. 1 represents the interior of the polytope, which is feasible for the high-point relaxation. The induced set, resulting from the optimal lower-level reaction, is given by:

$$\{(x, y) \in (\mathbb{R}_+, \mathbb{R}) \,\text{s.t.}\, y = 1 + \frac{x}{10}\}.$$

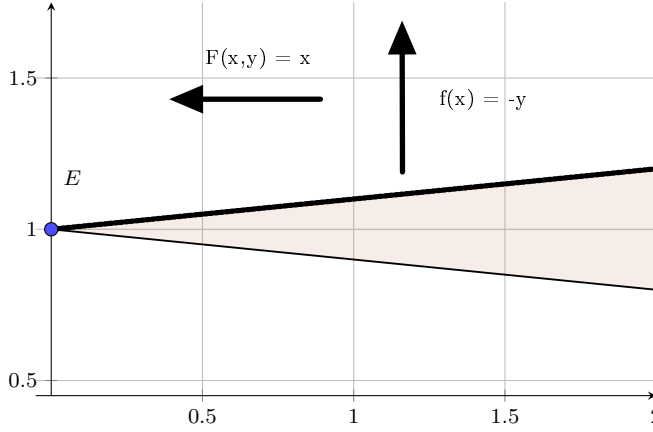The unique optimal point is $(\hat{x}, \hat{y}) = (0, 1)$.

Fig. 1 Linear bilevel problem

Let us now consider a near-optimal tolerance of the follower with $\delta = 0.1$. If the upper-level decision is $\hat{x}$, then the lower level can take any value between $1 - \delta = 0.9$ and 1. All these values except 1 lead to an unsatisfied upper-level constraint problem. The problem can be reformulated as:

$$\min_{x,v} \; x$$
$$\text{s.t.} \; x \geq 0$$
$$v \geq 1 - \frac{x}{10}$$
$$v \in \arg\max_{y}\{y \;\text{s.t.}\; y \leq 1 + \frac{x}{10}\}$$
$$z \geq 1 - \frac{x}{10} \; \forall z \;\text{s.t.}\; \{z \leq 1 + \frac{x}{10}, z \geq v - \delta\}.$$

Fig. 2 illustrates the near-optimal equivalent of the problem with an additional constraint ensuring the satisfaction of the upper-level constraint for all near-optimal responses of the lower level.

This additional constraint is represented by the dashed line. The optimal upper-level decision is $x = 0.5$, for which the optimal lower-level reaction is $y = 1 + 0.1 \cdot 0.5 = 1.05$. The boundary of the near-optimal set is $y =$
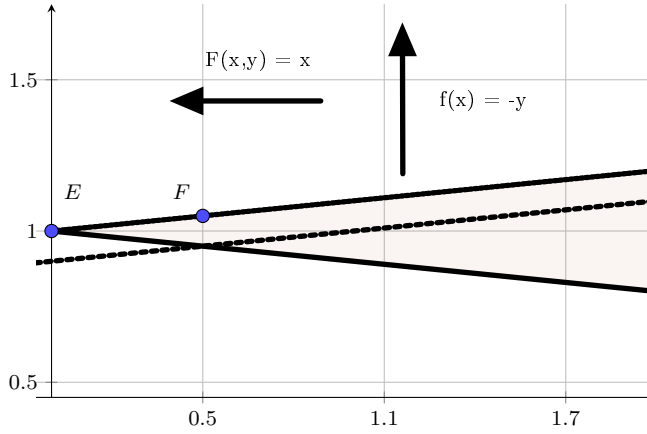
Fig. 2 Linear bilevel problem with a near-optimality robustness constraint

$1 - 0.1 \cdot 0.5 = 0.95$.

In the rest of this section, we establish properties of the near-optimal set and near-optimal robust bilevel problems. If the lower-level optimization problem is convex, then the near-optimal set $\mathcal{Z}(x;\theta)$ is convex as the intersection of two convex sets:

- $\{y \mid g(x,y) \leq 0\}$
- $\{y \mid f(x,y) \leq \phi(x) + \delta\}$.

In robust optimization, the characteristics of the uncertainty set sharply impact the difficulty of solving the problem. The near-optimal set of the lower-level is not always bounded; this can lead to infeasible or ill-defined near-optimal robust counterparts of bilevel problems. In the next proposition, we define conditions under which the uncertainty set $\mathcal{Z}(x;\delta)$ is bounded.

Proposition 2 *For a given pair (x,δ), any of the following properties is sufficient for $\mathcal{Z}(x;\delta)$ to be a bounded set:*

1. *The lower-level feasible domain is bounded.*
2. *$f(x,\cdot)$ is radially unbounded with respect to y.*
3. *$f(x,\cdot)$ is radially bounded such that:*

$$\lim_{r\in\mathbb{R}, r\to+\infty} f(x,rs) > f(x,v) + \delta \; \forall s \in \mathcal{S},$$

   with $\mathcal{S}$ the unit sphere in the space of lower-level decisions.

Proof The first case is trivially satisfied since $\mathcal{Z}(x;\delta)$ is the intersection of sets including the lower-level feasible set. If $f(x,\cdot)$ is radially unbounded, for any finite $\delta > 0$, there is a maximum radius around v beyond which any value of the objective function is greater than $f(x,v) + \delta$. The third case follows the same line of reasoning as the second, with a lower bound in any direction $\|y\| \to \infty$, such that this lower bound is above $f(x,v) + \delta$. □

The radius of robust feasibility is defined as the maximum "size" of the uncertain set [21,22], such that the robust problem remains feasible. In the case of near-optimality robustness, the radius can be interpreted as the maximum deviation of the lower-level objective from its optimal value, such that the near-optimal robust bilevel problem remains feasible.

Definition 1 For a given optimization problem BiP, let $\mathcal{NO}(BiP; \delta)$ be the optimum value of the near-optimal robust problem constructed from BiP with a tolerance δ. The radius of near-optimal feasibility $\hat{\delta}$ is defined by:

$$\hat{\delta} = \arg\max_{\delta} \{\delta \text{ s.t. } \mathcal{NO}(BiP; \delta) < \infty\}. \tag{9}$$

It is interesting to note that the radius as defined in Definition 1 can be interpreted as a maximum robustness budget in terms of the objective value of the lower level. It represents the maximum level of tolerance of the lower level on its objective, such that the upper level remains feasible.

Proposition 3 *The standard optimistic bilevel problem BiP is a relaxation of the equivalent near-optimal robust bilevel problem for any $\delta > 0$.*

Proof By introducing additional variables $z_{jk}, j \in [\![n_l]\!], k \in [\![m_u]\!]$ in the optimistic bilevel problem, we obtain:

$$\min_{x,v,z} F(x,v) \tag{10}$$
$$\text{s.t. } G_k(x,v) \leq 0 \qquad\qquad \forall k \in [\![m_u]\!]$$
$$f(x,v) \leq \phi(x)$$
$$g(x,v) \leq 0$$
$$x \in \mathcal{X}, v \in \mathbb{R}^{n_l}, z \in \mathbb{R}^{n_l \times m_u}.$$

Problem (10) is strictly equivalent to the optimistic bilevel problem with additional variables z that are not used in the objective nor constraints. Furthermore, Problem (10) is a relaxation of Problem (6), which has similar variables but additional constraints (6d). At each point where the bilevel problem is feasible, either the objective value of the two problems are the same or $NORBiP$ is infeasible. □

Proposition 4 *If the bilevel problem is feasible, then the adversarial problem (5) is feasible.*

Proof If the bilevel problem is feasible, then the solution $z = v$ is feasible for the primal adversarial problem. □

Proposition 5 *If $(\hat{x}, \hat{y})$ is a bilevel-feasible point, and $G_k(\hat{x}, \cdot)$ is K_k-Lipschitz continuous for a given $k \in [\![m_u]\!]$ such that:*

$$G_k(\hat{x}, \hat{y}) < 0,$$

then the constraint $G_k(\hat{x}, y) \leq 0$ is satisfied for all $y \in \mathcal{F}_L^{(k)}$ such that:

$$\mathcal{F}_L^{(k)}(\hat{x}, \hat{y}) = \{y \in \mathbb{R}^{n_l} \mid \|y - \hat{y}\| \leq \frac{|G_k(\hat{x}, \hat{y})|}{K_k}\}.$$

Proof As $G_k(\hat{x}, \hat{y}) < 0$, and $G_k(\hat{x}, \cdot)$ is continuous, there exists a ball $\mathcal{B}_r(\hat{y})$ in $\mathbb{R}^{n_l}$ centered on $(\hat{y})$ of radius $r > 0$, such that

$$G(\hat{x}, y) \leq 0 \ \forall y \in \mathcal{B}_r(\hat{y}).$$

Let us define:

$$r_0 = \arg\max_r \ r \tag{11}$$
$$\text{s.t. } G(\hat{x}, y) \leq 0 \qquad\qquad \forall y \in \mathcal{B}_r(\hat{y}).$$

By continuity, Problem (11) always admits a feasible solution. If the feasible set is bounded, there exists a point y_0 on the boundary of the ball, such that $G_k(\hat{x}, y_0) = 0$. It follows from Lipschitz continuity that:

$$|G_k(\hat{x}, \hat{y}) - G_k(\hat{x}, y_0)| \leq K_k \|y_0 - \hat{y}\|$$
$$\frac{|G_k(\hat{x}, \hat{y})|}{K_k} \leq \|y_0 - \hat{y}\|.$$

$G_k(\hat{x}, y) \leq G_k(\hat{x}, y_0) \ \forall y \in \mathcal{B}_{r_0}(\hat{y})$, therefore all lower-level solutions in the set

$$\mathcal{F}_L^{(k)}(\hat{x}, \hat{y}) = \{y \in \mathbb{R}^{n_l} \text{ s.t. } \|y - \hat{y}\| \leq \frac{|G_k(\hat{x}, \hat{y})|}{K_k}\}$$

satisfy the k-th constraint. $\square$

Corollary 1 *Let $(\hat{x}, \hat{y})$ be a bilevel-feasible solution of a near-optimal robust bilevel problem of tolerance δ, and*

$$\mathcal{F}_L(\hat{x}, \hat{y}) = \bigcap_{k=1}^{m_u} \mathcal{F}_L^{(k)}(\hat{x}, \hat{y}),$$

then $\mathcal{Z}(x; \delta) \subseteq \mathcal{F}_L(\hat{x}, \hat{y})$ is a sufficient condition for near-optimality robustness of $(\hat{x}, \hat{y})$.

Proof Any lower-level solution $y \in \mathcal{F}_L(\hat{x}, \hat{y})$ satisfies all m_u upper-level constraints, thus $\mathcal{Z}(x; \delta) \subseteq \mathcal{F}_L(\hat{x}, \hat{y})$ is a sufficient condition for the near-optimality robustness of $(\hat{x}, \hat{y})$. $\square$

Corollary 2 *Let $(\hat{x}, \hat{y})$ be a bilevel-feasible solution of a near-optimal robust bilevel problem of tolerance δ, R be the radius of the lower-level feasible set and $G_k(\hat{x}, \cdot)$ be K_k-Lipschitz for a given k, then the k-th constraint is robust against near-optimal deviations if:*

$$|G_k(\hat{x}, \hat{y})| \leq K_k R.$$

Proof The inequality can be deduced from the fact that $\|y - \hat{y}\| \leq R$. $\square$

Corollary 2 can be used when the lower level feasible set is bounded to verify near-optimality robustness of incumbent solutions.

3 Near-optimal robust bilevel problems with a convex lower level

In this section, we study near-optimal robust bilevel problems where the lower-level problem (1d) is a parametric convex optimization problem with both a differentiable objective function and differentiable constraints. If Slater's constraint qualifications hold, the KKT conditions are necessary and sufficient for the optimality of the lower-level problem and strong duality holds for the adversarial subproblems. These two properties are leveraged to reformulate *NORBiP* as a single-level closed-form problem.

Given a bilevel solution (x, v), the adversarial problem associated with constraint k can be formulated as:

$$\max_{y} \ G_k(x, y) \tag{12a}$$

$$\text{s.t.} \ g(x, y) \leq 0 \tag{12b}$$

$$f(x, y) \leq f(x, v) + \delta. \tag{12c}$$

Even if the upper-level constraints are convex with respect to y, Problem (12) is in general non-convex since the function to maximize is convex over a convex set. First-order optimality conditions may induce several non-optimal critical points and the definition of a solution method needs to rely on global optimization techniques [23,24].

By assuming that the constraints of the upper-level problem $G_k(x, y)$ can be decomposed and that the projection onto the lower variable space is affine:

$$G_k(x, y) \leq 0 \Leftrightarrow G_k(x) + H_k^T y \leq q_k. \tag{13}$$

The k-th adversarial problem is then expressed as:

$$\max_{y} \ \langle H_k, y \rangle \tag{14a}$$

$$\text{s.t.} \ g_i(x, y) \leq 0 \ \ \forall i \in [\![m_l]\!] \qquad (\alpha_i) \tag{14b}$$

$$f(x, y) \leq f(x, v) + \delta \qquad (\beta) \tag{14c}$$

and is convex for a fixed pair (x, v). Satisfying the upper-level constraint in the worst-case requires that the objective value of Problem (14) is lower than $q_k - G_k(x)$. We denote by $\mathcal{A}_k$ and $\mathcal{D}_k$ the objective values of the adversarial problem (14) and its dual respectively. $\mathcal{D}_k$ takes values in the extended real set to account for infeasible and unbounded cases. Proposition 4 holds for Problem (14). The feasibility of the upper-level constraint with the dual adversarial objective value as formulated in Constraint (15) is, by weak duality of convex problems, a sufficient condition for the feasibility of a near-optimal solution. If Slater's constraint qualifications hold, it is also a necessary condition [25] by strong duality:

$$\mathcal{A}_k \leq \mathcal{D}_k \leq q_k - G_k(x). \tag{15}$$

The generic form for the single-level reformulation of the near-optimal robust problem can then be expressed as:

$$\min_{x,v,\alpha,\beta} \quad F(x,v) \tag{16a}$$

$$\text{s.t.} \quad G(x) + Hv \leq q \tag{16b}$$

$$f(x,v) \leq \phi(x) \tag{16c}$$

$$g(x,v) \leq 0 \tag{16d}$$

$$\mathcal{D}_k \leq q_k - G_k(x) \qquad \forall k \in [\![m_u]\!] \tag{16e}$$

$$x \in \mathcal{X}, \tag{16f}$$

where (α, β) are certificates of the near-optimality robustness of the solution. In order to write Problem (16) in a closed form, the lower-level problem (16c-16d) is reduced to its KKT conditions:

$$\nabla_v f(x,v) - \sum_{i=1}^{m_l} \lambda_i \nabla_v g_i(x,v) = 0 \tag{17a}$$

$$g_i(x,v) \leq 0 \qquad \forall i \in [\![m_l]\!] \tag{17b}$$

$$\lambda_i \geq 0 \qquad \forall i \in [\![m_l]\!] \tag{17c}$$

$$\lambda_i g_i(x,v) = 0 \qquad \forall i \in [\![m_l]\!]. \tag{17d}$$

Constraint (17d) derived from the KKT conditions cannot be tackled directly by non-linear solvers [26]. Specific reformations, such as relaxations of the equality constraints (17d) into inequalities or branching on combinations of variables (as developed in [27,28]) are often used in practice.

In the rest of this section, we focus on bilevel problems such that the lower level is a conic convex optimization problem. Unlike the convex version developed above, the dual of a conic optimization problem can be written in closed form.

$$\min_y \quad \langle d, y \rangle \tag{18}$$

$$\text{s.t.} \quad Ax + By = b$$

$$y \in \mathcal{K}$$

where $\langle \cdot, \cdot \rangle$ is the inner product associated with the space of the lower-level variables and $\mathcal{K}$ is a proper cone [25, Chapter 2]. This class of problems encompasses a broad class of convex optimization problems of practical interest [29, Chapter 4], while the dual problem can be written in a closed-form if the dual cone is known, leading to a closed-form single-level reformulation. The

$k-$th adversarial problem is given by:

$$\max_{y,r} \ \langle H_k, y \rangle \tag{19a}$$

$$\text{s.t.} \ By = b - Ax \tag{19b}$$

$$\langle d, y \rangle + r = \langle d, v \rangle + \delta \tag{19c}$$

$$y \in \mathcal{K} \tag{19d}$$

$$r \geq 0 \tag{19e}$$

with the introduction of a slack variable r. With the following change of variables:

$$\hat{y} = \begin{bmatrix} y \\ r \end{bmatrix} \ \ \hat{B} = \begin{bmatrix} B \ 0 \end{bmatrix} \ \ \hat{d} = \begin{bmatrix} d \ 1 \end{bmatrix} \ \ \hat{H}_k = \begin{bmatrix} H_k \\ 0 \end{bmatrix},$$

$$\hat{\mathcal{K}} = \{(y, r), \ y \in \mathcal{K}, \ r \geq 0\},$$

$\hat{\mathcal{K}}$ is a cone as the Cartesian product of $\mathcal{K}$ and the nonnegative orthant. Problem (19) is reformulated as:

$$\max_{\hat{y}} \ \langle \hat{H}_k, \hat{y} \rangle$$

$$\text{s.t.} \ (\hat{B}\hat{y})_i = b_i - (Ax)_i \ \ \forall i \in [\![m_l]\!] \tag{α_i}$$

$$\langle \hat{d}, \hat{y} \rangle = \langle d, v \rangle + \delta \tag{β}$$

$$\hat{y} \in \hat{\mathcal{K}}$$

which is a conic optimization problem, for which the dual problem is:

$$\min_{\alpha, \beta, s_k} \ \langle (b - Ax), \alpha \rangle + (\langle d, v \rangle + \delta)\beta \tag{20a}$$

$$\text{s.t.} \ \hat{B}^T \alpha + \beta \hat{d} + s = \hat{H}_k \tag{20b}$$

$$s \in -\hat{\mathcal{K}}^*, \tag{20c}$$

with $\hat{\mathcal{K}}^*$ the dual cone of $\hat{\mathcal{K}}$. In the worst case (maximum number of non-zero coefficients), there are $(m_l \cdot n_u + n_l)$ bilinear terms in m_u non-linear non-convex constraints. This number of bilinear terms can be reduced by introducing the following variables (p, o), along with the corresponding constraints:

$$\min_{\alpha, \beta, s, p, o} \ \langle p, \alpha \rangle + (o + \delta)\beta \tag{21a}$$

$$\text{s.t.} \ p = b - Ax \tag{21b}$$

$$o = \langle d, v \rangle \tag{21c}$$

$$\hat{B}^T \alpha + \beta \hat{d} + s = \hat{H}_k \tag{21d}$$

$$s \in -\hat{\mathcal{K}}^*. \tag{21e}$$

The number of bilinear terms in the set of constraints is thus reduced from $n_u \cdot m_l + n_l$ to $m_l + 1$ terms in (21a). Problem (20) or equivalently Problem (21) have a convex feasible set but a bilinear non-convex objective function.

The KKT conditions of the follower problem (18) are given for the primal-dual pair (y, λ):

$$By = b - Ax \tag{22a}$$

$$y \in \mathcal{K} \tag{22b}$$

$$d - B^T \lambda \in \mathcal{K}^* \tag{22c}$$

$$\langle d - B^T \lambda, y \rangle = 0. \tag{22d}$$

The single-level problem is:

$$\min_{x,v,\lambda,\alpha,\beta,s} \quad F(x, v) \tag{23a}$$

$$\text{s.t.} \quad G(x) + Hv \leq q \tag{23b}$$

$$Ax + Bv = b \tag{23c}$$

$$d - B^T \lambda \in \mathcal{K}^* \tag{23d}$$

$$\langle d - B^T \lambda, v \rangle = 0 \tag{23e}$$

$$\langle Ax - b, \alpha_k \rangle + \beta_k \left(\langle v, d \rangle + \delta \right) \leq q_k - (Gx)_k \quad \forall k \in [\![m_u]\!] \tag{23f}$$

$$\hat{B}^T \alpha_k + \hat{d} \beta_k + s_k = \hat{H}_k \quad \forall k \in [\![m_u]\!] \tag{23g}$$

$$x \in \mathcal{X}, v \in \mathcal{K} \tag{23h}$$

$$s_k \in -\hat{\mathcal{K}}^* \quad \forall k \in [\![m_u]\!]. \tag{23i}$$

The Mangasarian-Fromovitz constraint qualification is violated at every feasible point of Constraint (23e) [30]. In non-linear approaches to complementarity constraints [27,26], parameterized successive relaxations of the complementarity constraints are used:

$$\langle d - B^T \lambda, v \rangle \leq \varepsilon \tag{24a}$$

$$-\langle d - B^T \lambda, v \rangle \leq \varepsilon. \tag{24b}$$

Constraints (23f) and (24) are both bilinear non-convex inequalities, the other ones added by the near-optimal robust model are conic and linear constraints.

In conclusion, near-optimal robustness has only added a finite number of constraints of the same nature (bilinear inequalities) to the reformulation proposed in [26]. Solution methods used for bilevel problems with convex lower-level thus apply to their near-optimal robust counterpart.

4 Linear near-optimal robust bilevel problem

In this section, we focus on near-optimal robust linear-linear bilevel problems. More precisely, the structure of the lower-level problem is exploited to derive an extended formulation leading to an efficient solution algorithm. We consider that all vector spaces are subspaces of $\mathbb{R}^n$, with appropriate dimensions. The

inner product of two vectors $\langle a, b \rangle$ is equivalently written $a^T b$.

The linear near-optimal robust bilevel problem is formulated as:

$$\min_{x,v} \ c_x^T x + c_y^T v \tag{25a}$$

$$\text{s.t.} \ Gx + Hv \le q \tag{25b}$$

$$d^T v \le \phi(x) \tag{25c}$$

$$Ax + Bv \le b \tag{25d}$$

$$Gx + Hz \le q \qquad \forall z \in \mathcal{Z}(x; \delta) \tag{25e}$$

$$v \in \mathbb{R}_+^{n_l} \tag{25f}$$

$$x \in \mathcal{X}. \tag{25g}$$

For a given pair (x, v), each semi-infinite robust constraint (25e) can be reformulated as the objective value of the following adversarial problem:

$$\max_{y} \ H_k^T y \tag{26a}$$

$$\text{s.t.} \ (By)_i \le b_i - (Ax)_i \ \ \forall i \in [\![m_l]\!] \qquad (\alpha_i) \tag{26b}$$

$$d^T y \le d^T v + \delta \qquad (\beta) \tag{26c}$$

$$y \in \mathbb{R}_+^{n_l}. \tag{26d}$$

Let (α, β) be the dual variables associated with each group of constraints (26b-26c). The near-optimal robust version of Problem (25) is feasible only if the objective value of each k-th adversarial subproblem (26) is lower than $q_k - (Gx)_k$. The dual of Problem (26) is defined as:

$$\min_{\alpha,\beta} \alpha^T (b - Ax) + \beta \left(d^T v + \delta \right) \tag{27a}$$

$$\text{s.t.} \ B^T \alpha + \beta d \ge H_k \tag{27b}$$

$$\alpha \in \mathbb{R}_+^{m_l} \beta \in \mathbb{R}_+. \tag{27c}$$

Based on Problem (4) and weak duality results, the dual problem is either infeasible or feasible and bounded. By strong duality, the objective value of the dual and primal problems are equal. This value must be smaller than $q_k - (Gx)_k$ to satisfy Constraint (25e). This is equivalent to the existence of a feasible dual solution (α, β) certifying the feasibility of (x, v) within the near-optimal set $\mathcal{Z}(x; \delta)$. We obtain one pair of certificates (α, β) for each upper-level constraint

in $[\![m_u]\!]$, resulting in the following problem:

$$
\min_{x,v,\alpha,\beta} \quad c_x^T x + c_y^T v \tag{28a}
$$

$$
\text{s.t.} \quad Gx + Hv \le q \tag{28b}
$$

$$
d^T v \le \phi(x) \tag{28c}
$$

$$
Ax + Bv \le b \tag{28d}
$$

$$
\alpha_k^T (b - Ax) + \beta_k \left(d^T v + \delta\right) \le q_k - (Gx)_k \ \forall k \in [\![m_u]\!] \tag{28e}
$$

$$
B^T \alpha_k + \beta_k d \ge H_k \ \forall k \in [\![m_u]\!] \tag{28f}
$$

$$
\alpha_k \in \mathbb{R}_+^{m_l} \beta_k \in \mathbb{R}_+ \ \forall k \in [\![m_u]\!] \tag{28g}
$$

$$
v \in \mathbb{R}_+^{n_l} \tag{28h}
$$

$$
x \in \mathcal{X}. \tag{28i}
$$

Lower-level optimality is guaranteed by the corresponding KKT conditions:

$$
d_j + \sum_i B_{ij} \lambda_i - \sigma_j = 0 \qquad\qquad \forall j \in [\![n_l]\!] \tag{29a}
$$

$$
0 \le b_i - (Ax)_i - (Bv)_i \perp \lambda_i \ge 0 \qquad\qquad \forall i \in [\![m_l]\!] \tag{29b}
$$

$$
0 \le v_j \perp \sigma_j \ge 0 \qquad\qquad \forall j \in [\![n_l]\!] \tag{29c}
$$

$$
\sigma \ge 0, \lambda \ge 0 \tag{29d}
$$

where $\perp$ defines a complementarity constraint. A common technique to linearize Constraints (29b-29c) is the "big-M" reformulation, introducing auxiliary binary variables with primal and dual upper bounds. The resulting formulation has a weak continuous relaxation. Furthermore, the correct choice of bounds is itself an NP-hard problem [31], and the incorrect choice of these bounds can lead to cutting valid and potentially optimal solutions [32]. Other modelling and solution approaches, such as special ordered sets of type 1 (SOS1) or indicator constraints avoid the need to specify such bounds in a branch-and-bound procedure.

The aggregated formulation of the linear near-optimal robust bilevel problem is:

$$\min_{x,v,\lambda,\sigma,\alpha,\beta} \quad c_x^T x + c_y^T v \tag{30a}$$

$$\text{s.t.} \quad Gx + Hv \leq q \tag{30b}$$

$$Ax + Bv \leq b \tag{30c}$$

$$d_j + \sum_i \lambda_i B_{ij} - \sigma_j = 0 \qquad \forall j \in [\![n_l]\!] \tag{30d}$$

$$0 \leq \lambda_i \perp A_i x + B_i v - b_i \leq 0 \qquad \forall i \in [\![m_l]\!] \tag{30e}$$

$$0 \leq \sigma_j \perp v_j \geq 0 \qquad \forall j \in [\![n_l]\!] \tag{30f}$$

$$x \in \mathcal{X} \tag{30g}$$

$$\alpha_k \cdot (b - Ax) + \beta_k(d^T v + \delta) \leq q_k - (Gx)_k \quad \forall k \in [\![m_u]\!] \tag{30h}$$

$$\sum_{i=1}^{m_l} B_{ij}\alpha_{ki} + \beta_k d_j \geq H_{kj} \qquad \forall k \in [\![m_u]\!], \ \forall j \in [\![n_l]\!] \tag{30i}$$

$$\alpha_k \in \mathbb{R}_+^{m_l}, \beta_k \in \mathbb{R}_+ \qquad \forall k \in [\![m_u]\!]. \tag{30j}$$

Problem (30) is a single-level problem and has a closed form. However, constraints (30h) contain bilinear terms, which cannot be tackled as efficiently as convex constraints by branch-and-cut based solvers. Therefore, we exploit the structure of the dual adversarial problem and its relation to the primal lower level to design a new efficient reformulation and solution algorithm.

4.1 Extended formulation

The bilinear constraints (30h) involve products of variables from the upper and lower level (x, v) as well as dual variables of each of the m_u dual-adversarial problems. For fixed values of (x, v), m_u dual adversarial sub-problems (27) are defined. The optimal value of each k-th subproblem must be lower than $q_k - (Gx)_k$. The feasible region of each sub-problem is defined by (30h-30j) and is independent of (x, v). The objective functions are linear in (α, β). Following Proposition 4, Problem (27) is bounded. If, moreover, Problem (27) is feasible, a vertex of the polytope (30h-30j) is an optimal solution. Following these observations, Constraints (30h-30j) can be replaced by disjunctive constraints, such that for each k, at least one extreme vertex of the k-th dual polyhedron is feasible. This reformulation of the bilinear constraints has, to the best of our knowledge, never been developed in the literature. Let $\mathcal{V}_k$ be the number of vertices of the k-th sub-problem and α_k^l, β_k^l be the l-th vertex of the k-th sub-problem. Constraints (30h-30j) can be written as:

$$\bigvee_{l=1}^{\mathcal{V}_k} \sum_{i=1}^{m_l} \alpha_{ki}^l (b - Ax)_i + \beta_k^l \cdot (d^T v + \delta) \leq q_k - (Gx)_k \qquad \forall k \in [\![m_u]\!], \tag{31}$$

where $\bigvee_{i=1}^{N} \mathcal{C}_i$ is the disjunction (logical "OR") operator, expressing the constraint that at least one of the constraints $\mathcal{C}_i$ must be satisfied. These disjunctions are equivalent to indicator constraints [33].

This reformulation of bilinear constraints based on the polyhedral description of the (α, β) feasible space is similar to the Benders decomposition. Indeed in the near-optimal robust extended formulation, at least one of the vertices must satisfy a constraint (a disjunction) while Benders decomposition consists in satisfying a set of constraints for all extreme vertices and rays of the dual polyhedron (a constraint described with a universal quantifier). Disjunctive constraints (31) are equivalent to the following formulation, using set cover and SOS1 constraints:

$$\theta_k^l \in \mathbb{B} \qquad\qquad\qquad \forall k, \forall l \qquad (32a)$$

$$\omega_k^l \geq 0 \qquad\qquad\qquad \forall k, \forall l \qquad (32b)$$

$$(b - Ax)^T \alpha_k^l + \beta_k^l (d^T v + \delta) - \omega_k^l \leq q_k - (Gx)_k \qquad \forall k, \forall l \qquad (32c)$$

$$\sum_{l=1}^{\mathcal{V}_k} \theta_k^l \geq 1 \qquad\qquad\qquad \forall k \qquad (32d)$$

$$SOS1(\theta_k^l, \omega_k^l) \qquad\qquad\qquad \forall k, \ \forall l. \qquad (32e)$$

In conclusion, using disjunctive constraints over the extreme vertices of each dual polyhedron and SOS1 constraints to linearize the complementarity constraints leads to an equivalent reformulation of Problem (30). The finite solution property holds even though the boundedness of the dual feasible set is not required. This single-level extended reformulation can be solved by any off-the-shelf MILP solver. Nevertheless, to decrease the computation time, we design a specific algorithm based on necessary conditions for the existence of a solution.

First, we illustrate the extended formulation on the following example.

4.2 Bounded example

Consider the bilevel linear problem defined by the following data:

$$x \in \mathbb{R}_+, y \in \mathbb{R}_+$$

$$G = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \ H = \begin{bmatrix} 4 \\ 2 \end{bmatrix} \ q = \begin{bmatrix} 11 \\ 13 \end{bmatrix} \ c_x = \begin{bmatrix} 1 \end{bmatrix} \ c_y = \begin{bmatrix} -10 \end{bmatrix}$$

$$A = \begin{bmatrix} -2 \\ 5 \end{bmatrix} \ B = \begin{bmatrix} -1 \\ -4 \end{bmatrix} \ b = \begin{bmatrix} -5 \\ 30 \end{bmatrix} \ d = \begin{bmatrix} 1 \end{bmatrix}.$$

The optimal solution of the high-point relaxation $(x,v) = (5,4)$ is not bilevel-feasible. The optimal value of the optimistic bilevel problem is reached at $(x,v) = (1,3)$. These two points are respectively represented by the blue diamond and red cross in Fig. 3. The dotted segments represent the upper-level constraints and the solid lines represent the lower-level constraints.
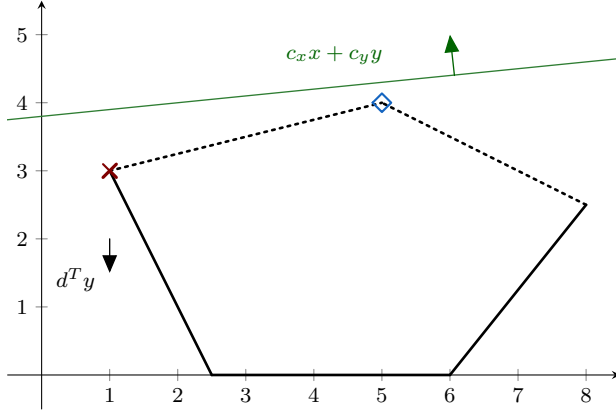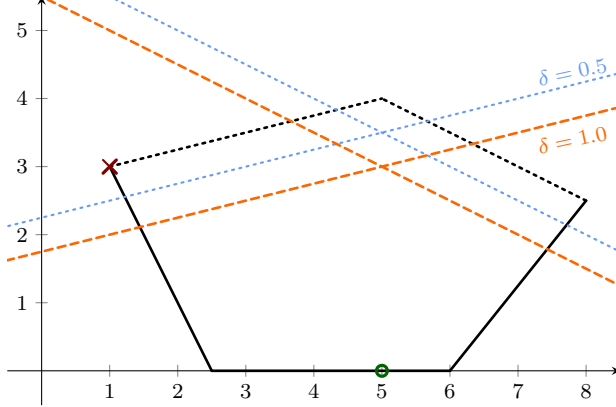

Fig. 3 Representation of the bilevel problem.

Fig. 4 Near-optimal robustness constraints.

The feasible space for (α, β) is given by:

$$-1\alpha_{11} - 4\alpha_{12} + \beta_1 \geq 4$$
$$-1\alpha_{21} - 4\alpha_{22} + \beta_2 \geq 2$$
$$\alpha_{ki} \geq 0, \beta_k \geq 0.$$

This feasible space can be described as a set of extreme points and rays. It consists in this case of one extreme point $(\alpha_{ki} = 0, \beta_1 = 4, \beta_2 = 2)$ and 4 extreme rays. The (x,v) solution needs to be valid for the corresponding

near-optimality conditions:

$$\beta_1 \ (v + \delta) \leq 11 + x$$
$$\beta_2 \ (v + \delta) \leq 13 - x.$$

This results in two constraints in the (x, v) space, represented in Fig. 4 for $\delta = 0.5$ and $\delta = 1.0$ in dotted blue and dashed orange respectively. The radius of near-optimal feasibility $\hat{\delta} = 5$ can be computed using the formulation provided in Definition 1, for which the feasible domain at the upper-level is reduced to the point $x = 5$, for which $v = 0$, represented as a green circle at $(5, 0)$ in Fig. 4.

4.3 Solution algorithm

The solution procedure is defined as follows based on the structure of the extended formulation. The main goal of the algorithm is to prove infeasibility early in the resolution process and to solve the extended formulation only in the last step. Let $\mathcal{P}_0(BiP)$, $\mathcal{P}_1(BiP)$, $\text{FEAS}_k((BiP)$, $\mathcal{P}_{no}(BiP; \delta)$ be the high-point relaxation, optimistic bilevel problem, dual feasibility and near-optimal robust problem respectively. Let C_k be the list of extreme vertices of the k-th dual adversarial polyhedron.

Algorithm 1 Near-Optimal Robust Vertex Enumeration Procedure (NORVEP)

1: **function** *near_optimal_bilevel*(BiP, δ)
2: {**Step 1**: dual subproblems expansion & pre-solving}
3: **for** $k \in [\![m_u]\!]$ **do**
4:     Solve dual adversarial problem
5:     **if** $\text{feas}_k = $ Infeasible **then**
6:         **Terminate**: k-th dual adversarial infeasible
7:     **else**
8:         $C_k \leftarrow (\alpha_k^l, \beta_k^l)_{l \in \mathcal{V}_k}$
9:     **end if**
10: **end for**
11: {**Step 2**: high-point relaxation $\mathcal{P}_0(BiP)$}
12: **if** $\mathcal{P}_0(BiP)$ infeasible **then**
13:     **Terminate**: high-point relaxation infeasible
14: **end if**
15: {**Step 3**: optimistic relaxation $\mathcal{P}_1(BiP)$}
16: **if** $\mathcal{P}_1(BiP)$ infeasible **then**
17:     **Terminate**: optimistic bilevel infeasible
18: **end if**
19: {**Step 4**: extended formulation $\mathcal{P}_{no}(BiP, (C_k)_{k \in [\![m_u]\!]}; \delta)$}
20: **return** Terminate and return solution information
21: **end function**

Each step consists in solving a problem that must be feasible for the feasibility of NORBiP to hold. The algorithm terminates without proceeding to the subsequent steps in an infeasibility is detected.

4.4 Valid inequalities

The extended formulation and Algorithm 1 can be directly applied. Nevertheless, we propose two groups of valid inequalities to tighten the formulation.

The first group of inequalities consists of the primal upper-level constraints:

$$(Gx)_k + (Hv)_k \leq q_k \qquad\qquad \forall z \in [\![m_u]\!].$$

These constraints are necessary for the optimistic formulation but not for the near-optimal robust one since they are always redundant with and included in the near-optimal robust constraints. However, their addition can strengthen the linear relaxation of the extended formulation and lead to faster convergence.

The second group of inequalities is defined in [34] and based on strong duality of the lower level. We only implement the valid for the root node, which are the primary focus of [34]:

$$\langle \lambda, b \rangle + \langle v, d \rangle \leq \langle A^+, \lambda \rangle, \qquad\qquad (33)$$

where A_i^+ is an upper bound on $\langle A_i, x \rangle$. The computation of each upper bound A_i^+ relies on solving an auxiliary problem:

$$A_i^+ = \max_{x,v,\lambda} \langle A_i, x \rangle \qquad\qquad (34\text{a})$$

$$\text{s.t.} \ \ Gx + Hv \leq q \qquad\qquad (34\text{b})$$

$$Ax + Bv \leq b \qquad\qquad (34\text{c})$$

$$d + B^T \lambda \geq 0 \qquad\qquad (34\text{d})$$

$$x \in \mathcal{X}, v \geq 0, \lambda \qquad\qquad (34\text{e})$$

$$(x, v, \lambda) \in \Upsilon, \qquad\qquad (34\text{f})$$

where Υ is the set containing all valid inequalities (33).

The method proposed in [34] relies on solving each i-th auxiliary problem once and using the resulting bound A^+. We define a new iterative procedure to improve the bounds computed at the root node:

1. Solve Problem (34a) $\forall i \in [\![m_l]\!]$ and obtain A^+;
2. If $\exists i$, A_i^+ is unbounded, terminate;
3. Otherwise, add Constraint (33) to (34f) and go to step 1;
4. Stopping criterion: when an iteration does not improve any of the bounds, terminate and return the last inequality with the sharpest bound.

This procedure allows tightening the bound as long as improvement can be made in one of the A_i^+. If the procedure terminates with one A_i^+ unbounded, the right-hand side of (33) is $+\infty$, the constraint is trivial and cannot be improved upon. Otherwise, each iteration improves the bound until the convergence of A^+.

5 Computational experiments

In this section, we demonstrate the applicability of our approach through numerical experiments on instances of the linear-linear near-optimal robust bilevel problem. We first describe the sets of test instances and the computational setup and then the experiments and their results.

5.1 Instance sets

Two sets of data are considered. For the first one, a total number of 1000 small, 200 medium and 100 large random instances are generated and characterized as follows:

$$(m_u, m_l, n_l, n_u) = (5, 5, 5, 5) \qquad \text{(small)}$$
$$(m_u, m_l, n_l, n_u) = (10, 10, 10, 10) \qquad \text{(medium)}$$
$$(m_u, m_l, n_l, n_u) = (20, 10, 20, 20) \qquad \text{(large)}.$$

All matrices are randomly generated with each coefficient having a 0.6 probability of being 0 and uniformly distributed on $[0, 1]$ otherwise. High-point feasibility and the vertex enumeration procedures are run after generating each tuple of random parameters to discard infeasible instances. Collecting 1000 small instances required generating 10532 trials, the 200 medium-sized instances were obtained with 18040 trials and the 100 large instances after 90855 trials. A second dataset is created from the 50 MIPS/Random instances of the Bilevel Problem library [35], where integrality constraints are dropped. All of these instances contain 20 lower-level constraints and no upper-level constraints. For each of them, two new instances are built by moving either the first 6 or the last 6 constraints from the lower to the upper level, resulting in 100 instances. We will refer to the first set of instances as the small/medium/large instances and the second as the MIPS instances. All instances are available in [36] in JLD format, along with a reader to import them in Julia programs.

5.2 Computational setup

Algorithm 1 is implemented in Julia [37] using the JuMP v0.21 modelling framework [38,39]; the MILP solver is SCIP 6.0 [40] with SoPlex 4.0 as the inner LP solver, both with default solving parameters. SCIP handles indicator constraints in the form of linear inequality constraints activated only if a binary variable is equal to one. *Polyhedra.jl* [41] is used to model the dual subproblem polyhedra with CDDLib [42] as a solver running the double-description algorithm, computing the list of extreme vertices and rays from the constraint-based representation. The exact rational representation of numbers is used in CDDLib instead of floating-point types to avoid rounding errors. Moreover, CDDLib fails to produce the list of vertices for some instances when set in floating-point mode. All experiments are performed on a consumer-end laptop with 15.5GB of RAM and an Intel i7 1.9GHz CPU running Ubuntu 18.04LTS.

5.3 Bilinear and extended formulation

To assess the efficiency of the extended formulation, we compare its solution time to that of the non-extended formulation including bilinear constraints (25). The bilinear formulation is implemented with SCIP using SoPlex as the linear optimization solver and Ipopt as the non-linear solver. SCIP handles the bilinear terms through bound computations and spatial branching. We test the two methods on 100 small instances. The bilinear version only manages to solve the small random instances and runs out of time or memory for all other instance sets. A time limit of 3600 seconds and a memory limit of 5000MB were fixed. The distribution of runtimes is presented in Fig. 5.
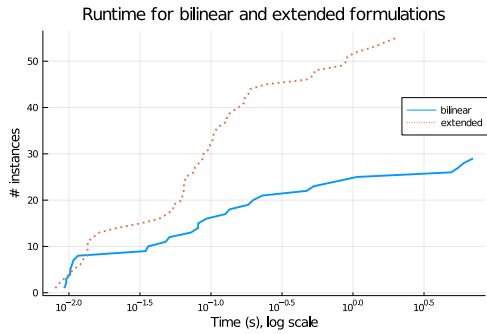

Fig. 5 Runtime of the two methods on 100 of the small instances

The extended formulation dominates at almost any time the bilinear formulation that uses spatial branching. The latter runs out of time or memory for most instances.

5.4 Robustness of optimistic solutions and influence of δ

We solve the MIPS instances to bilevel optimality and verify the near-optimal robustness of the obtained solutions. We use various tolerance values:

$$\delta = \max(0.05, \delta_r \times \mathrm{opt}(L))$$

with $\mathrm{opt}(L)$ the lower-level objective value at the found solution and

$$\delta_r \in (0.01, 0.05, 0.1, 0.5, 3.0).$$

Out of the 100 instances, 57 have canonical solutions that are not robust to even the smallest near-optimal deviation $0.01\mathrm{opt}(L)$. Twelve more instances that have a near-optimal robust solution with the lowest tolerance are not near-optimal robust when the tolerance is increased to $3\mathrm{opt}(L)$ Out of the 57 instances that are not near-optimal robust with the lowest tolerance, 40 have

exactly one upper-level constraint that is violated by near-optimal deviations of the lower level and 17 that have more than one. Finally, we observe that the number of violated constraints changes across the range of tolerance values for 31 out of 100 instances. For the other 69 instances, the number of violated upper-level constraints remains identical for all tolerance values.

Table 1 summarizes the number of infeasible instances for different values of δ. As δ increases, so does the proportion of infeasible problems. This is due to the increase in the left-hand side in constraints (31).

δ	0.01	0.1	0.2	1	3	5	7	10	12
Small (/1000)	366	423	466	595	658	670	672	674	676
Medium (/200)	78	88	95	118	122	123	123	123	123

Table 1 Number of infeasible problems for various tolerance levels δ

In Fig. 6, we present the runtime difference between the canonical bilevel problem and its near-optimal robust counterpart.
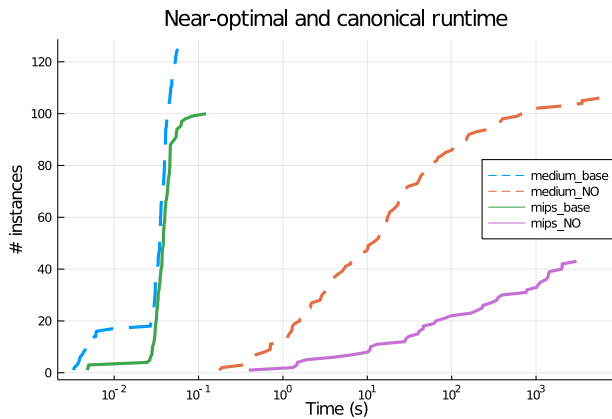

Fig. 6 Runtime cost of adding near-optimality robustness constraints.

Scaling up the dimension of the tackled problems is limited not only due to high computation time but also due to memory requirements since the formulation of the problem requires allocating binary variables and a disjunctive constraint over all vertices of the dual polyhedron of each of the $k \in [\![m_u]\!]$ subproblems.

These runtime profiles highlight the fact that near-optimality robustness implemented using the extended formulation adds a significant runtime cost to the resolution of linear-linear bilevel problems. Nonetheless, the study of near-optimal lower-level decisions on optimistic solutions shows that these optimistic solutions are not robust, even for small tolerance values. This time

difference also motivates the design of Algorithm 1. Indeed, since the optimistic bilevel problem is solved in a much shorter time, it is interesting to verify its feasibility before solving the near-optimal robust version.

5.5 Computational time of Algorithm 1

Statistics on the computation times of the two phases of Algorithm 1 for each instance size are provided in Table 2 and Table 3.

Size	mean	10% quant.	50% quant.	90% quant.
Small	0.023	0.014	0.019	0.046
Medium	1.098	0.424	0.956	2.148
MIPS	21.061	0.231	3.545	65.004

Table 2 Runtime statistics for the vertex enumeration (s).

Instance type	# optimized	mean	10% quant.	50% quant.	90% quant.
Small	577	0.205	0.004	0.064	0.596
Medium	106	207.399	0.797	14.451	317.624
MIPS	70	909.302	57.592	344.202	2613.404

Table 3 Runtime statistics for the optimization phase (s).

The solution time, corresponding to Steps 2-4 of Algorithm 1, is greater than the vertex enumeration phase, corresponding to Step 1, but does not dominate it completely for any of the problem sizes.

Figure 7 shows the distribution of the upper-level objective values across small and medium-sized instances. The number of problems solved to optimality monotonically decreases when δ increases (Table 1); greater δ values indeed make reduce the set of feasible solutions to *NORBiP*. The optimal values only slightly increase with δ and the lower-level objective value does not vary significantly with δ.

Even though more instances become infeasible as δ increases, the degradation of the objective value is in general insignificant for the optimal near-optimal robust solution compared to the optimistic solution.

5.6 Implementation of valid inequalities

In the last group of experiments, we implement and investigate the impact of the valid inequalities defined in Section 4.4.
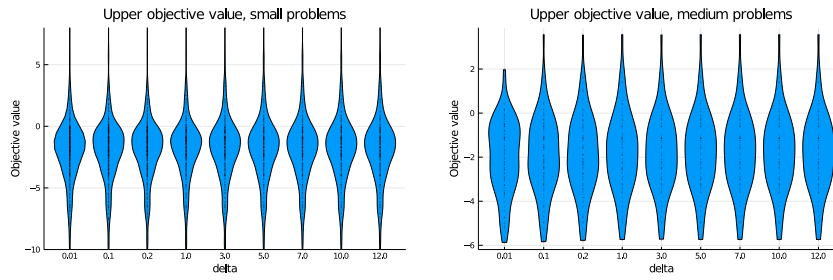
Fig. 7 Violin plots of the upper objective value distributions versus δ.

On the 200 medium-sized instances, adding the valid inequality (33) is enough to prove the infeasibility of 61 instances out of 68 that are infeasible but possess a feasible high-point relaxation. On 100 large instances, adding the valid inequality proves the infeasibility of 29 out of 45 infeasible instances for which the high-point relaxation is feasible. For all medium and large instances, a non-trivial valid inequality i.e. where all A_i^+ are finite was computed.

These results highlight the improvement of the model tightness with the addition of the valid inequalities, compared to the high-point relaxation where primal and dual variables are subject to distinct groups of constraints. These inequalities thus discard infeasible instances without the need to solve the complete MILP reformulation. In Fig. 8, the distribution of the number of iterations of the inequality-finding procedure is presented for the medium and large instances. For the majority of instances of both sizes (about 80% and 60% of instances for the medium and large instances), a single iteration is sufficient to find the best valid inequality (33). The number of iteration, however, goes up to 40 and 50 for the medium and large instances respectively (truncated on the graph for clarity).
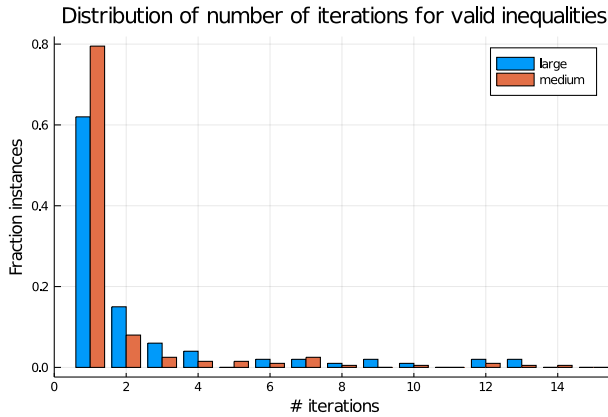

Fig. 8 Distribution of the number of iterations for the computation of valid inequalities.

In Fig. 9, we compare the total runtime for MIPS and medium instances under near-optimality robustness constraints using $\delta = 0.1$ with and without valid inequalities for all instances solved to optimality. The runtime for instances with valid inequalities includes the runtime of the inequality computation.
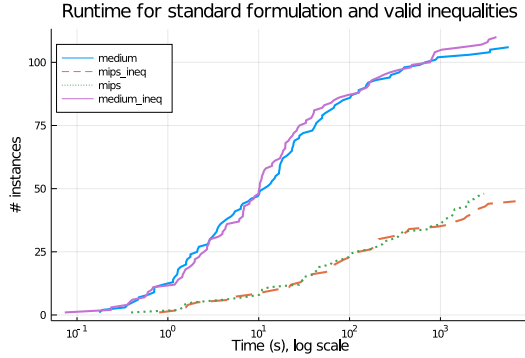

Fig. 9 Runtime for MIPS and medium instances with and without valid inequalities.

Valid inequalities do not improve the runtime for NORBiP in either group of instances, This result is similar to the observations in [34] for instances of the canonical bilevel linear problem without near-optimality robustness.

We next study the inequalities based on the upper-level constraints on the small, medium and MIPS instances.
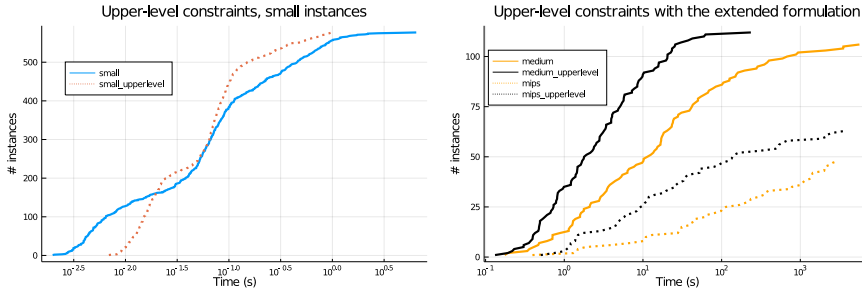

Fig. 10 Runtime for small, medium and MIPS instances with and without upper-level constraints.

As shown in Fig. 10, the addition of primal upper-level constraints accelerates the resolution of the MIPS and medium instances and dominates the standard extended formulation. For the small instances, we observe smaller runtimes for the first instances solved. This can be due to the upper-level constraints making the linear relaxation larger by adding constraints, thus creating overhead for

smaller problems. This overhead is compensated for instances that are harder to optimize, i.e. that require more than 0.02 seconds to solve.

6 Conclusion

In this work, we introduce near-optimal robust bilevel optimization, a specific formulation of the bilevel optimization problem where the upper-level constraints are protected from deviations of the lower level from optimality. Near-optimality robustness challenges the assumption that the lower-level problem is solved to optimality, resulting in a generalized, more conservative formulation including the optimistic and pessimistic bilevel problems as special cases. We formulate NORBiP in the dependent case, i.e. where the upper- and lower-level constraints depend on both upper- and lower-level variables, thus offering a framework applicable to many bilevel problems of practical interest.

We derive a closed-form, single-level expression of NORBiP for convex lower-level problems, based on dual adversarial certificates to guarantee near-optimality robustness. In the linear case, we derive an extended formulation that can be represented as a MILP with indicator constraints. Numerical experiments highlight the efficiency of the extended method compared to the compact bilinear formulation and the impact of some valid inequalities on both solution time and tightness of the linear relaxation.

References

1. S. Dempe, "Bilevel optimization: theory, algorithms and applications," 2018.
2. W. Wiesemann, A. Tsoukalas, P.-M. Kleniati, and B. Rustem, "Pessimistic Bilevel Optimization," *SIAM Journal on Optimization*, vol. 23, pp. 353–380, jan 2013.
3. S. Dempe, *Foundations of bilevel programming*. Springer Science & Business Media, 2002.
4. G. E. Stavroulakis and H. Günzel, "Optimal structural design in nonsmooth mechanics," in *Multilevel Optimization: Algorithms and Applications*, pp. 91–115, Springer, 1998.
5. M. G. Nicholls, "Optimizing the operations of an aluminium smelter using non-linear bilevel programming," in *Multilevel Optimization: Algorithms and Applications*, pp. 117–148, Springer, 1998.
6. H. A. Simon, "Theories of bounded rationality," *Decision and organization*, vol. 1, no. 1, pp. 161–176, 1972.
7. R. J. Aumann, "Rationality and bounded rationality," in *Cooperation: Game-Theoretic Approaches*, pp. 219–231, Springer, 1997.
8. M. Aghassi and D. Bertsimas, "Robust game theory," *Mathematical Programming*, vol. 107, pp. 231–273, Jun 2006.
9. F. Ordóñez and N. E. Stier-Moses, "Robust wardrop equilibrium," in *Network Control and Optimization* (T. Chahed and B. Tuffin, eds.), (Berlin, Heidelberg), pp. 247–256, Springer Berlin Heidelberg, 2007.
10. M. Jain, F. Ordónez, J. Pita, C. Portway, M. Tambe, C. Western, P. Paruchuri, and S. Kraus, "Robust solutions in stackelberg games: Addressing boundedly rational human preference models," 2008.
11. J. Pita, M. Jain, M. Tambe, F. Ordóñez, and S. Kraus, "Robust solutions to stackelberg games: Addressing bounded rationality and limited observations in human cognition," *Artificial Intelligence*, vol. 174, no. 15, pp. 1142–1171, 2010.

12. I. Yanıkoğlu and D. Kuhn, "Decision rule bounds for two-stage stochastic bilevel programs," *SIAM Journal on Optimization*, vol. 28, no. 1, pp. 198–222, 2018.
13. M. H. Zare, O. Y. Özaltın, and O. A. Prokopyev, "On a class of bilevel linear mixed-integer programs in adversarial settings," *Journal of Global Optimization*, vol. 71, no. 1, pp. 91–113, 2018.
14. M. H. Zare, O. A. Prokopyev, and D. Sauré, "On bilevel optimization with inexact follower," *Decision Analysis*, vol. 17, no. 1, pp. 74–95, 2020.
15. S. Dempe, V. Kalashnikov, G. A. Prez-Valds, and N. Kalashnykova, *Bilevel Programming Problems: Theory, Algorithms and Applications to Energy Networks*. Springer Publishing Company, Incorporated, 2015.
16. G. Still, "Generalized semi-infinite programming: Theory and methods," *European Journal of Operational Research*, vol. 119, no. 2, pp. 301–313, 1999.
17. V. Goel and I. E. Grossmann, "A class of stochastic programs with decision dependent uncertainty," *Mathematical Programming*, vol. 108, pp. 355–394, Sep 2006. DOI: 10.1007/s10107-006-0715-7.
18. V. Gabrel, C. Murat, and A. Thiele, "Recent advances in robust optimization: An overview," *European Journal of Operational Research*, vol. 235, pp. 471–483, jun 2014.
19. D. Bertsimas, D. B. Brown, and C. Caramanis, "Theory and Applications of Robust Optimization," *SIAM Review*, vol. 53, pp. 464–501, jan 2011.
20. P. Loridan and J. Morgan, "Weak via strong Stackelberg problem: New results," *Journal of Global Optimization*, vol. 8, pp. 263–287, Apr 1996.
21. F. Liers, L. Schewe, and J. Thürauf, "Radius of robust feasibility for mixed-integer problems," 2019. `http://www.optimization-online.org/DB_FILE/2019/05/7219.pdf`.
22. M.A. Goberna and V. Jeyakumar and G. Li and N. Linh, "Radius of robust feasibility formulas for classes of convex programs with uncertain polynomial constraints," *Operations Research Letters*, vol. 44, no. 1, pp. 67 – 73, 2016. DOI: 10.1016/j.orl.2015.11.011.
23. P. Pardalos and J. Rosen, "Methods for global concave minimization: A bibliographic survey," *SIAM Review*, vol. 28, no. 3, pp. 367–379, 1986.
24. H. Benson and R. Horst, "A branch and bound-outer approximation algorithm for concave minimization over a convex set," *Computers & Mathematics with Applications*, vol. 21, no. 6, pp. 67 – 76, 1991.
25. S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
26. S. Dempe, "Computing locally optimal solutions of the bilevel optimization problem using the kkt approach," in *International Conference on Mathematical Optimization Theory and Operations Research*, pp. 147–157, Springer, 2019.
27. S. Scholtes, "Convergence properties of a regularization scheme for mathematical programs with complementarity constraints," *SIAM Journal on Optimization*, vol. 11, no. 4, pp. 918–936, 2001.
28. L. Schewe and M. Schmidt, "Computing feasible points for binary MINLPs with MPECs," *Mathematical Programming Computation*, vol. 11, pp. 95–118, Mar 2019. DOI: 10.1007/s12532-018-0141-x.
29. Y. Nesterov and A. Nemirovskii, *Interior-point polynomial algorithms in convex programming*, vol. 13. SIAM, 1994.
30. H. Scheel and S. Scholtes, "Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity," *Mathematics of Operations Research*, vol. 25, no. 1, pp. 1–22, 2000.
31. T. Kleinert, M. Labbé, F. Plein, and M. Schmidt, "There's no free lunch: On the hardness of choosing a correct big-m in bilevel optimization," 2019. `http://www.optimization-online.org/DB_FILE/2019/04/7172.pdf`.
32. S. Pineda and J. M. Morales, "Solving linear bilevel problems using big-ms: Not all that glitters is gold," *IEEE Transactions on Power Systems*, vol. 34, pp. 2469–2471, May 2019.
33. P. Bonami, A. Lodi, A. Tramontani, and S. Wiese, "On mathematical programming with indicator constraints," *Mathematical Programming*, vol. 151, pp. 191–223, Jun 2015. `10.1007/s10107-015-0891-4`.
34. T. Kleinert, M. Labbé, M. Schmidt, and F. Plein, "Closing the gap in linear bilevel optimization: A new valid primal-dual inequality," 2020.
35. T. Ralphs and S. Tahernejad, "Bilevel Optimization Problem Library v0.1," 2020. `https://coral.ise.lehigh.edu/data-sets/bilevel-instances/`.

36. M. Besançon, M. F. Anjos, and L. Brotcorne, "Near-optimal robust bilevel linear instances," Aug. 2020. https://doi.org/10.5281/zenodo.4009108.

37. J. Bezanson, A. Edelman, S. Karpinski, and V. Shah, "Julia: A Fresh Approach to Numerical Computing," *SIAM Review*, vol. 59, pp. 65–98, 01 2017.

38. I. Dunning, J. Huchette, and M. Lubin, "JuMP: A Modeling Language for Mathematical Optimization," *SIAM Review*, vol. 59, no. 2, pp. 295–320, 2017.

39. B. Legat, O. Dowson, J. D. Garcia, and M. Lubin, "MathOptInterface: a data structure for mathematical optimization problems," *arXiv preprint arXiv:2002.03447*, 2020.

40. A. Gleixner, M. Bastubbe, L. Eifler, T. Gally, G. Gamrath, R. L. Gottwald, G. Hendel, C. Hojny, T. Koch, M. E. Lübbecke, S. J. Maher, M. Miltenberger, B. Müller, M. E. Pfetsch, C. Puchert, D. Rehfeldt, F. Schlösser, C. Schubert, F. Serrano, Y. Shinano, M. Viernickel, M. Walter, F. Wegscheider, J. T. Witt, and J. Witzig, "The SCIP Optimization Suite 6.0," technical report, Optimization Online, July 2018. http://www.optimization-online.org/DB_HTML/2018/07/6692.html.

41. B. Legat, R. Deits, O. Evans, G. Goretkin, T. Koolen, J. Huchette, D. Oyama, M. Forets, guberger, R. Schwarz, E. Saba, and C. Coleman, "JuliaPolyhedra/Polyhedra.jl: v0.5.1," may 2019. DOI: 10.5281/zenodo.3066026.

42. B. Legat, R. Deits, M. Forets, D. Oyama, F. Pacaud, and E. Saba, "JuliaPolyhedra/CDDLib.jl: v0.5.2," may 2019. DOI: 10.5281/zenodo.2671760.