



# Robust bilevel optimization for near-optimal lower-level solutions

Mathieu Besançon<sup>1</sup> · Miguel F. Anjos<sup>2,3</sup>  · Luce Brotcorne<sup>4</sup>

Received: 28 January 2023 / Accepted: 11 July 2024  
© The Author(s) 2024

## Abstract

Bilevel optimization problems embed the optimality of a subproblem as a constraint of another optimization problem. We introduce the concept of near-optimality robustness for bilevel optimization, protecting the upper-level solution feasibility from limited deviations from the optimal solution at the lower level. General properties and necessary conditions for the existence of solutions are derived for near-optimal robust versions of general bilevel optimization 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 exact and heuristic methods and the impact of valid inequalities on the solution time.

**Keywords** Bilevel optimization · Robust optimization · Decision-dependent uncertainty · Bounded rationality · Duality · Bilinear constraints · Extended formulation

**Mathematics Subject Classification** 90C33 · 90C46 · 91A65 · 90C26 · 90C34

## 1 Introduction

Bilevel optimization problems embed the optimality conditions of a subproblem into the constraints of another one. They can faithfully model various decision-making problems such as Stackelberg or leader-follower games, market equilibria, or pricing and revenue management. A recent review of methods and applications of bilevel problems is presented

---

✉ Miguel F. Anjos  
anjos@stanfordalumni.org

Mathieu Besançon  
mathieu.besancon@inria.fr

Luce Brotcorne  
luce.brotcorne@inria.fr

<sup>1</sup> Université Grenoble Alpes, Inria, LIG, Grenoble, France

<sup>2</sup> School of Mathematics, University of Edinburgh, James Clerk Maxwell Building, Peter Guthrie Tait Road, Edinburgh EH9 3FD, UK

<sup>3</sup> GERAD, 3000, chemin de la Côte-Sainte-Catherine, Montréal, QC H3T 2A7, Canada

<sup>4</sup> Centre Inria de l'Université de Lille, Inria, France and UMR 9189 - CRISTAL, Univ. Lille, CNRS, Centrale Lille, France

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, by which we mean decisions that respect bilevel feasibility but with the corresponding objective value different from the optimal value by up to an additive constant. An important issue in this setting is the definition of the robustness of the upper-level decisions with respect to such near-optimal lower-level solutions.

In some engineering applications [2–4], 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.

In economics and decision theory, the concept of *bounded rationality* [5] or  $\varepsilon$ -rationality [6] 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, and it is through these lenses that we will view it in this paper. Therefore, we use the terms “near-optimality robustness” and “near-optimal robust bilevel problem” or NRB in the rest of the paper to refer to this form of robustness.

The introduction of uncertainty and robustness in games has been approached from different points of view in the literature, with for instance [7] for the existence of robust counterparts of Nash equilibria without the knowledge of probability distributions associated with the uncertainty. For uncertainty in hierarchical games and bilevel problems which is our focus in this work, we refer to the recent survey [8] for an overview of approaches and formulations. 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 a heuristic or approximation algorithm or may deviate from its optimal value in a way that penalizes the objective of the upper level. In [15], a bilevel model is developed with the lower-level agent facing limit observability of the upper-level decisions, resulting in another form of uncertainty for the upper level; the uncertainty is formulated in a fashion inspired by the concept of near-optimality robustness presented in this work. In [16], a robust version of the bilevel continuous knapsack is considered, where the upper level does not have a complete knowledge of the lower-level objective function. They also establish complexity results for the problem which depend on the discrete or continuous nature of the uncertainty set of the lower-level objective coefficients.

Solving bilevel problems under limited deviations of the lower-level response was introduced in [17] under the term “ $\varepsilon$ -approximation” of the pessimistic bilevel problem. The authors focus on the independent case, i.e. problem settings 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 [17, Theorem 2.2]. A custom algorithm is designed for the independent case, solving a sequence of non-convex non-linear problems relying on global nonlinear solvers. We must highlight that other bilevel formulations such as the one presented in [18] have used concepts of approximate lower level solutions, but that these formulations are relaxations of the original problem, allowing solutions that are  $\varepsilon$ -optimal for the second level and therefore not necessarily bilevel-feasible unlike our approach or that of [17] which *restrict* the feasible space of the standard (both optimistic and pessimistic) bilevel optimization formulation and robustify the solution.

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 “ $\varepsilon$ -approximation” from [17]. 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 NRB where the lower-level problem is convex conic or linear in Sect. 3, resulting in a finite-dimensional single-level optimization problem.
3. An extended formulation for the linear-linear NRB in Sect. 4, linearizing the bilinear constraints of the single-level model using disjunctive constraints.
4. Exact and heuristic solution methods for the linear-linear NRB in Sect. 5 using the extended formulation and its properties.

The paper is organized as follows. In Sect. 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 Sects. 3 and 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. Exact and heuristic solution algorithms are provided in Sect. 5 and computational experiments are conducted in Sect. 6, comparing the extended formulation to the compact one and studying the impact of valid inequalities. Finally, in Sect. 7 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 general bilevel problem is classically defined as:

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

$$\text{s.t. } G_k(x, v) \leq 0 \quad \forall k \in \llbracket m_u \rrbracket \tag{1b}$$

$$x \in \mathcal{X} \quad (1c)$$

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

The upper- and lower-level objective functions are noted  $F, f : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$  respectively. We denote with  $\llbracket n \rrbracket$  the set of running indices  $1 \dots n$ ,  $n_U, n_L$  denote the number of variables of the upper and lower level respectively,  $m_U, m_L$  similarly denote the number of constraints. Constraint (1b) and  $g_i(x, y) \leq 0 \forall i \in \llbracket m_l \rrbracket$  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:

$$\begin{aligned} \phi : \mathbb{R}^{n_u} &\rightarrow \{-\infty\} \cup \mathbb{R} \cup \{+\infty\} \\ \phi(x) &= \min_y \{f(x, y) \text{ s.t. } g(x, y) \leq 0\}. \end{aligned} \quad (2)$$

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 feasible upper-level decision.

When, for a feasible upper-level decision, the solution  $v$  to the lower-level 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 [19, Chapter 1] for further details on these two approaches. The optimistic case is the most straightforward to formulate using the value function:

$$\begin{aligned} \min_{x, v} F(x, v) & \quad (B) \\ \text{s.t. } G_k(x, v) &\leq 0 & k \in \llbracket m_u \rrbracket \\ x &\in \mathcal{X} \\ f(x, v) &\leq \phi(x) \\ g_i(x, v) &\leq 0 & \forall i \in \llbracket m_l \rrbracket. \end{aligned}$$

Note that a near-optimal robust problem can be constructed from the original “unspecified” problem (1) or from the optimistic formulation (B).

Instead of making an explicit optimistic/pessimistic assumption about the reaction  $v$  of the lower level, we propose a means for the upper level to protect itself against possible follower deviations from its optimality. In other words, we seek a decision  $x$  at the upper level that is *robust* in the sense that it remains feasible even if the lower level deviates from its own optimality. For this purpose, for a given upper-level decision  $x$  and tolerance  $\delta$ , we define the near-optimal set of the lower level  $\mathcal{Z}(x; \delta)$  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 NRB, of parameter  $\delta$  is defined as a bilevel problem with the additional constraints (3e) below ensuring that the upper-level constraints must be satisfied for any lower-level solution  $z$  in the near-optimal set  $\mathcal{Z}(x; \delta)$ :

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

$$\text{s.t. } G_k(x, v) \leq 0 \quad \forall k \in \llbracket m_u \rrbracket \quad (3b)$$

$$f(x, v) \leq \phi(x) \quad (3c)$$

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

$$G_k(x, z) \leq 0 \quad \forall z \in \mathcal{Z}(x; \delta) \quad \forall k \in \llbracket m_u \rrbracket \tag{3e}$$

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

Each  $k$  constraint in (3b) is satisfied if the corresponding constraint set in (3e) is non-empty and 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.

NRB captures, unifies, and extends several common formulations in bilevel optimization. The special case  $\mathcal{Z}(x; 0)$  is the set of optimal solutions to the original lower-level problem, NRB with  $\delta = 0$  is therefore equivalent to the constraint-based pessimistic bilevel problem as formulated in [17]:

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

For  $\delta < 0$ ,  $\mathcal{Z}(x; \delta)$  is the empty set, in which case NRB is equivalent to the original optimistic bilevel problem. 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  $x$ . It therefore results in an optimistic bilevel formulation with a classical robustness constraints at the upper level. We also note the connection or near-optimality robustness to the uncertainty model proposed in [14], in which the lower-level decision is assumed to be derived from an exact or heuristic method known from a fixed set of known algorithms. In contrast, we do not make assumptions on a solution process but consider that the lower-level problem may be solved to near-optimality with a fixed additive constant  $\delta$ . This corresponds naturally to several classes of exact algorithms which provide a guarantee on the optimality gap that depends on the computational effort (e.g. first-order, interior point methods, branch-and-bound algorithms), and to all approximation algorithms which would provide solutions with a bound on the optimality gap.

Unlike the constraint-based pessimistic bilevel problem presented in [17], 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 modelling choice is enabled by the NRB formulation we chose which uses the optimistic lower-level response  $v$  explicitly. It also implies that the upper level chooses the best optimistic decision which also 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, without an epigraph formulation of the objective function. 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:

$$(C\text{-NRB}) \min_{x, v, \tau} \tau \tag{4a}$$

$$\text{s.t. } G_k(x, v) \leq 0 \quad \forall k \in \llbracket m_u \rrbracket \tag{4b}$$

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

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

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

$$G_k(x, z) \leq 0 \quad \forall z \in \mathcal{Z}(x; \delta) \quad \forall k \in \llbracket m_u \rrbracket \tag{4f}$$

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

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

$$\text{opt(B)} \leq \text{opt(NRB)} \leq \text{opt(C-NRB)},$$

where  $\text{opt(P)}$  denotes the optimal value of problem P. Both near-optimal robust formulations NRB and C-NRB can be of interest to model decision-making applications. It can also be noted that NRB includes the special case of interdiction problems, 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. We will focus on Problem (3), NRB, in the rest of this article, but all results and formulations extend to Problem (4).

Constraint (3e) is a generalized semi-infinite constraint [20]. The dependence of the set of constraints  $\mathcal{Z}(x; \delta)$  on the decision variables leads to the characterization of NRB as a robust problem with decision-dependent uncertainty [21]. 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  $(x, v, \delta)$ :

$$z_k \in \arg \max_y G_k(x, y) \quad (5a)$$

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

$$g(x, y) \leq 0. \quad (5c)$$

The addition of the semi-infinite near-optimal robustness constraint increases the computational difficulty of the bilevel optimization problem. Nonetheless, NRB and multilevel optimization problems with similar forms of uncertainty on lower-level decisions do not increase the complexity of the multilevel problem in the polynomial hierarchy under mild conditions [22]. For bilevel knapsack problems with an uncertain lower-level objective, [23] establishes complexity results in the discrete and continuous lower level cases.

It is common in the robust optimization literature to present models with either uncertainty on the constraints and/or on the objective function [24]. As for these, we show that the Objective-Robust Near-Optimal Bilevel Problem (O-NRB), is a special case of NRB:

$$\text{(O-NRB)} \min_{x \in X} \sup_{z \in \mathcal{Z}(x; \delta)} F(x, z)$$

$$\text{s.t. } X = \{x \in \mathcal{X}, G_k(x) \leq 0 \forall k \in \llbracket m_u \rrbracket\}$$

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

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

O-NRB is a special case of NRB, following a reformulation from objective uncertainty to constraint uncertainty with an epigraph variable, O-NRB is equivalent to:

$$\min_{x, \tau} \tau$$

$$\text{s.t. } x \in X$$

$$F(x, z) \leq \tau$$

$$\forall z \in \mathcal{Z}(x; \delta),$$

this formulation is a special case of NRB with an upper-level objective independent of lower-level variables. The pessimistic bilevel optimization problem defined in [25] is both a special case and a relaxation of O-NRB. For  $\delta = 0$ , the adversarial problem of O-NRB is equivalent to finding the worst lower-level decision with respect to the upper-level objective amongst

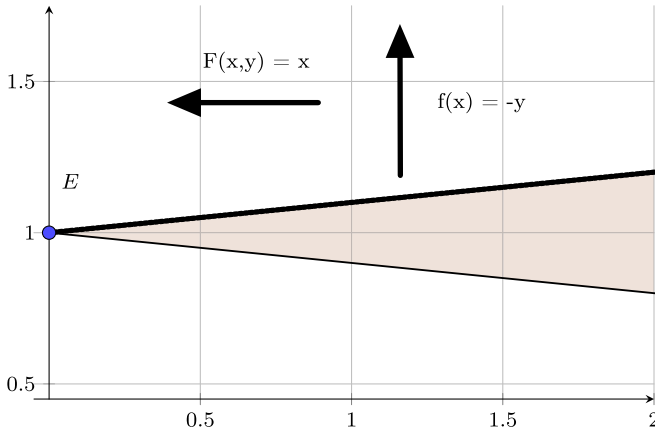


Fig. 1 Linear bilevel problem

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 O-NRB.

We illustrate the concept of near-optimal set and near-optimal robust solution, first with a simple linear bilevel problem represented in Fig. 1 for a geometric intuition, and then in two applications from the literature to show the interest of near-optimality robustness in practical applications.

$$\begin{aligned}
 & \min_{x,v} x \\
 & \text{s.t. } x \geq 0 \\
 & \quad v \geq 1 - \frac{x}{10} \\
 & \quad v \in \arg \max_y \{y \text{ s.t. } y \leq 1 + \frac{x}{10}\}.
 \end{aligned} \tag{6}$$

The high-point relaxation of Problem (6), obtained by relaxing the optimality constraint of the lower level, while maintaining feasibility, is:

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

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  $(\hat{x}, \hat{y}) = (0, 1)$ .

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

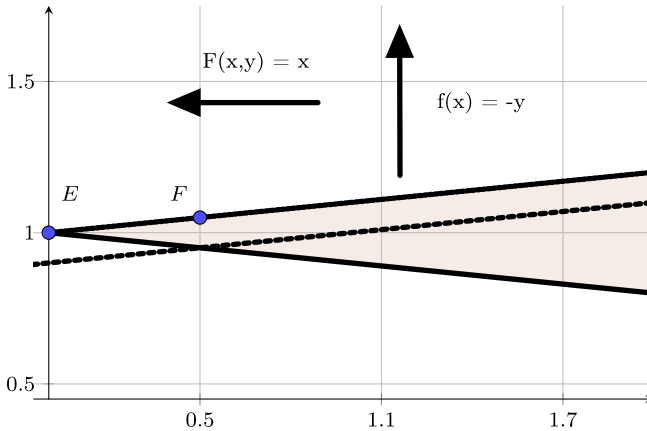


Fig. 2 Linear bilevel problem with a near-optimality robustness constraint

be reformulated as:

$$\begin{aligned}
 & \min_{x,v} x \\
 & \text{s.t. } x \geq 0 \\
 & v \geq 1 - \frac{x}{10} \\
 & v \in \arg \max_y \{y \text{ s.t. } y \leq 1 + \frac{x}{10}\} \\
 & z \geq 1 - \frac{x}{10} \quad \forall z \in \{z \mid z \leq 1 + \frac{x}{10}, z \geq v - \delta\}.
 \end{aligned}$$

Figure 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  $y = 1 - 0.1 \cdot 0.5 = 0.95$ .

We next present the interpretation of near-optimality robustness in the context of two applications. In [26], a bilevel model is introduced to determine the parameters of a pricing scheme for electricity in the context of demand response. The pricing scheme can be summarized as offering users two alternatives for a given time frame, either:

- the user remains on the baseline price for the given time frame, in which case the price is “flat”, i.e. independent of the consumption,
- the user books a certain capacity for a time frame, benefitting from a price lower than the baseline if their consumption remains below the booked capacity, and having to pay a price higher than the baseline if they overconsume.

The energy supplier, acting as the leader, seeks to offer a price that incentivizes the user to book a capacity in a certain range, for instance to make the consumption more predictable. Formulating the problem with the optimistic hypothesis implies that we assume the user will commit to booking a certain capacity and risking to potentially pay a higher price, even if the expected total cost is identical to the simpler flat price. The pessimistic hypothesis also



appears very limited in its mitigation of the problem, since it implies that the user choice preferred by the supplier need to be better than all others by any arbitrarily small gain. Bounded rationality appears like a more natural representation of the user decision-making process, who would probably not switch to a more complex pricing scheme unless the net gain reaches a certain amount, therefore making near-optimality robustness the natural modelling framework for this problem.

Another application in last-mile delivery can be proposed based on the problem introduced in [27]. In this model, a delivery platform, acting as the leader, must deliver parcels to customers by contracting intermediate delivery carriers acting as followers. The platform must ensure that the parcels are delivered, and chooses a compensation to offer to each carrier for delivering a particular parcel. The carriers then select a subset of parcels to deliver maximizing their profit subject to time or space constraints. It is however natural to assume that some carriers may have constraints or costs that are unknown from the leader, e.g., preferred routes, types of customers, or time windows, but could amount to a difference in objective bounded by some quantity  $\delta$ . Near-optimality robustness offers a framework to require solutions ensuring that all the parcels are delivered, even if some carriers choose a near-optimal route.

Generalizing from these application, near-optimal robust bilevel optimization is a natural modelling framework for leader-follower games where:

- the leader has constraints that depend on the follower decisions,
- these constraints cannot be imposed directly to the follower, whose choice can mainly be changed through incentives on their objective,
- the user could make a decision with bounded rationality.

Typically, when the follower objective amounts to a financial loss or profit, the optimistic and pessimistic hypotheses assume that the follower will adopt a particular behaviour, even if the relative benefit of that behaviour is infinitesimal for them. In contrast, NRB incorporates a minimum incentive of  $\delta$  directly in the model, ensuring the feasibility of the solution for the upper level despite bounded rationality. Other applications following this structure include pricing problems in networks and shared systems, e.g. passenger trains and flights, where the price is the only incentive mechanism, and in which the upper level needs to ensure the system is not used more than its capacity, corresponding to overbooking or overutilization of some resources. We also note that one group of problems in bilevel optimization that is not suited to NRB is that of interdiction games, since the lower level is already adversarially trying to impact the upper level negatively.

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 1** *For a given pair  $(x, \delta)$ , 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  $y$ , i.e.  $\|y\| \rightarrow \infty \Rightarrow f(x, y) \rightarrow \infty$ .
3.  $f(x, \cdot)$  is radially bounded such that:

$$\lim_{r \rightarrow +\infty} f(x, rs) > f(x, v) + \delta \quad \forall s \in S,$$

with  $S$  the unit sphere in the space of lower-level variables.

**Proof** The first case is trivially satisfied since  $Z(x; \delta)$  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  $\|y\| \rightarrow \infty$ , such that this lower bound is above  $f(x, v) + \delta$ .  $\square$

The radius of robust feasibility is defined as the maximum “size” of the uncertain set [28, 29], 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** (Radius of near-optimal feasibility) For a given optimistic bilevel optimization problem (B), let  $\text{opt}_\delta(\text{NRB})$  be the optimal value of the near-optimal robust problem (3) constructed from (B) with a tolerance  $\delta$ . The radius of near-optimal feasibility  $\hat{\delta}$  is defined by:

$$\hat{\delta} = \arg \max_{\delta} \{ \delta \text{ s.t. } \text{opt}_\delta(\text{NRB}) < \infty \}. \tag{7}$$

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 2** The optimistic bilevel problem (B) is a relaxation of the corresponding near-optimal robust bilevel problem for any  $\delta$ .

**Proof** The optimistic bilevel problem (B) is equivalent to Problem (3) without Constraints (3e) and has the same variables as Problem (3).  $\square$

**Proposition 3** If the optimistic bilevel problem (B) 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.  $\square$

**Proposition 4** If  $(\hat{x}, \hat{y})$  is a bilevel-feasible point, and  $G_k(\hat{x}, \cdot)$  is  $K_k$ -Lipschitz continuous for a given  $k \in \llbracket m_u \rrbracket$  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  $\mathbb{R}^m$  centered on  $(\hat{y})$  of radius  $r > 0$ , such that

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

Let us define:

$$r_0 = \arg \max_r \{r \text{ s.t. } G(\hat{x}, y) \leq 0 \quad \forall y \in \mathcal{B}_r(\hat{y})\}. \tag{8}$$

By continuity, Problem (8) 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:

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

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

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

satisfy the  $k$ -th constraint. □

**Corollary 1** Let  $(\hat{x}, \hat{y})$  be a bilevel-feasible solution to the optimistic bilevel problem (B),  $\delta$  a tolerance value, 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 solution  $(\hat{x}, \hat{y})$  to be near-optimal robust. □

**Corollary 2** Let  $(\hat{x}, \hat{y})$  be a bilevel-feasible solution to the optimistic bilevel problem (B),  $\delta$  a tolerance value, let  $R$  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  $\|y - \hat{y}\| \leq R$ . □

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 qualification holds, 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 NRB as a single-level closed-form problem.

Given a pair  $(x, v)$ , the adversarial problem associated with the  $k$ -th constraint of Problem (3) can be formulated as:

$$\max_y G_k(x, y) \tag{9a}$$

$$\text{s.t. } g(x, y) \leq 0 \tag{9b}$$

$$f(x, y) \leq f(x, v) + \delta. \tag{9c}$$

Even if the upper-level constraints are convex with respect to  $y$ , Problem (9) 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 [30, 31].

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, the upper-level constraint can be re-written as:

$$G_k(x, y) \leq 0 \Leftrightarrow G_k(x) + H_k^T y \leq q_k. \tag{10}$$

The  $k$ -th adversarial problem is then expressed as:

$$\max_y \langle H_k, y \rangle \tag{11a}$$

$$\text{s.t. } g_i(x, y) \leq 0 \quad \forall i \in \llbracket m_l \rrbracket \tag{11b} \quad (\alpha_i)$$

$$f(x, y) \leq f(x, v) + \delta \tag{11c} \quad (\beta)$$

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 (11) 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 (11) and its dual respectively.  $\mathcal{D}_k$  takes values in the extended real set to account for infeasible and unbounded cases. Proposition 3 holds for Problem (11). The feasibility of the upper-level constraint with the dual adversarial objective value as formulated in Constraint (12) 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 [32] by strong duality:

$$\mathcal{A}_k \leq \mathcal{D}_k \leq q_k - G_k(x). \tag{12}$$

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

$$\min_{x,v} F(x, v) \tag{13a}$$

$$\text{s.t. } G(x) + H v \leq q \tag{13b}$$

$$f(x, v) \leq \phi(x) \tag{13c}$$

$$g(x, v) \leq 0 \tag{13d}$$

$$\mathcal{D}_k \leq q_k - G_k(x) \quad \forall k \in \llbracket m_u \rrbracket \tag{13e}$$

$$x \in \mathcal{X}. \tag{13f}$$

In order to write Problem (13) in a closed form, the lower-level problem (13c–13d) is reduced to its KKT conditions:

$$\nabla_v f(x, v) - \sum_{i=1}^{m_l} \lambda_i \nabla_v g_i(x, v) = 0 \tag{14a}$$

$$g_i(x, v) \leq 0 \quad \forall i \in \llbracket m_l \rrbracket \quad (14b)$$

$$\lambda_i \geq 0 \quad \forall i \in \llbracket m_l \rrbracket \quad (14c)$$

$$\lambda_i g_i(x, v) = 0 \quad \forall i \in \llbracket m_l \rrbracket. \quad (14d)$$

Constraint (14d) derived from the KKT conditions cannot be tackled directly by non-linear solvers [33]. Specific reformulations, such as relaxations of the equality Constraints (14d) into inequalities or branching on combinations of variables (as developed in [34, 35]) 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.

$$\begin{aligned} \min_y \quad & \langle d, y \rangle \\ \text{s.t.} \quad & Ax + By = b \\ & y \in \mathcal{K} \end{aligned} \quad (15)$$

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 [32, Chapter 2]. This class of problems encompasses a broad class of convex optimization problems of practical interest [36, 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:

$$\max_{y,r} \langle H_k, y \rangle \quad (16a)$$

$$\text{s.t. } By = b - Ax \quad (16b)$$

$$\langle d, y \rangle + r = \langle d, v \rangle + \delta \quad (16c)$$

$$y \in \mathcal{K} \quad (16d)$$

$$r \geq 0 \quad (16e)$$

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

$$\begin{aligned} \hat{y} &= \begin{bmatrix} y \\ r \end{bmatrix} \quad \hat{B} = [B \ 0] \quad \hat{d} = [d \ 1] \quad \hat{H}_k = \begin{bmatrix} H_k \\ 0 \end{bmatrix}, \\ \hat{\mathcal{K}} &= \{(y, r), y \in \mathcal{K}, r \geq 0\}, \end{aligned}$$

$\hat{\mathcal{K}}$  is a cone as the Cartesian product of  $\mathcal{K}$  and the nonnegative orthant. Problem (16) is reformulated as:

$$\begin{aligned} \max_{\hat{y}} \quad & \langle \hat{H}_k, \hat{y} \rangle \\ \text{s.t.} \quad & (\hat{B}\hat{y})_i = b_i - (Ax)_i \quad \forall i \in \llbracket m_l \rrbracket \quad (\alpha_i) \\ & \langle \hat{d}, \hat{y} \rangle = \langle d, v \rangle + \delta \quad (\beta) \\ & \hat{y} \in \hat{\mathcal{K}} \end{aligned}$$

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

$$\min_{\alpha, \beta, s_k} \langle (b - Ax), \alpha \rangle + (\langle d, v \rangle + \delta)\beta \quad (17a)$$

$$\text{s.t. } \hat{B}^\top \alpha + \beta \hat{d} + s = \hat{H}_k \quad (17b)$$

$$s \in -\hat{\mathcal{K}}^*, \quad (17c)$$

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:

$$\min_{\alpha, \beta, s, p, o} \langle p, \alpha \rangle + (o + \delta)\beta \tag{18a}$$

$$\text{s.t. } p = b - Ax \tag{18b}$$

$$o = \langle d, v \rangle \tag{18c}$$

$$\hat{B}^\top \alpha + \beta \hat{d} + s = \hat{H}_k \tag{18d}$$

$$s \in -\hat{\mathcal{K}}^*. \tag{18e}$$

The number of bilinear terms in the set of constraints is thus reduced from  $n_u m_l + n_l$  to  $m_l + 1$  terms in (18a). Problem (17) or equivalently Problem (18) have a convex feasible set but a bilinear non-convex objective function. The KKT conditions of the follower Problem (15) are given for the primal-dual pair  $(y, \lambda)$ :

$$By = b - Ax \tag{19a}$$

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

$$d - B^\top \lambda \in \mathcal{K}^* \tag{19c}$$

$$\langle d - B^\top \lambda, y \rangle = 0. \tag{19d}$$

The single-level problem is:

$$\min_{x, v, \lambda, \alpha, \beta, s} F(x, v) \tag{20a}$$

$$\text{s.t. } G(x) + Hv \leq q \tag{20b}$$

$$Ax + Bv = b \tag{20c}$$

$$d - B^\top \lambda \in \mathcal{K}^* \tag{20d}$$

$$\langle d - B^\top \lambda, v \rangle = 0 \tag{20e}$$

$$\langle Ax - b, \alpha_k \rangle + \beta_k (\langle v, d \rangle + \delta) \leq q_k - (Gx)_k \quad \forall k \in \llbracket m_u \rrbracket \tag{20f}$$

$$\hat{B}^\top \alpha_k + \hat{d} \beta_k + s_k = \hat{H}_k \quad \forall k \in \llbracket m_u \rrbracket \tag{20g}$$

$$x \in \mathcal{X}, v \in \mathcal{K} \tag{20h}$$

$$s_k \in -\hat{\mathcal{K}}^* \quad \forall k \in \llbracket m_u \rrbracket. \tag{20i}$$

The Mangasarian–Fromovitz constraint qualification is violated at every feasible point of Problem (20a) because of the complementarity Constraints (20e) [37]. In non-linear approaches to complementarity constraints [33, 34], parameterized successive relaxations which respect constraint qualifications are used:

$$\langle d - B^\top \lambda, v \rangle \leq \varepsilon \tag{21a}$$

$$-\langle d - B^\top \lambda, v \rangle \leq \varepsilon. \tag{21b}$$

Constraints (20f) and (21) are both bilinear non-convex inequalities, the other ones added by the near-optimal robust model are conic and linear constraints. In general and unlike the complementarity constraints, the feasible region defined by near-optimal robust constraints admits strictly feasible solutions and therefore respect constraint qualifications.

**Remark 1** If  $H_k$  belongs to the interior of the polar set of  $\mathcal{K}$ , and if the sufficient conditions for applying KKT to the lower level hold, then the k-th dual adversarial problem admits strictly feasible solutions.

**Proof** From the assumptions on the lower-level KKT conditions, the dual lower level admits strictly feasible solutions, i.e.

$$\exists \lambda_0 \in \mathbb{R}^{m_u}, d - B^\top \lambda_0 \in \text{int}(\mathcal{K}^*). \tag{22}$$

The dual adversarial feasible set is described by the constraints:

$$\begin{aligned} B^\top \alpha + \beta d - H_k &\in \mathcal{K}^* \\ \beta &\geq 0. \end{aligned}$$

From (22), by setting  $\alpha = -\lambda_0, \beta = 1$ , we have  $B^\top \alpha + \beta d \in \text{int}(\mathcal{K}^*)$ .  $\mathcal{K}^*$  is a closed convex cone and is therefore closed under addition:

$$\begin{aligned} \frac{-H_k}{2} + \frac{1}{2}(B^\top \alpha + \beta d) &\in \mathcal{K}^* \Leftrightarrow \\ 2\left(\frac{-H_k}{2} + \frac{1}{2}(B^\top \alpha + \beta d)\right) &\in \mathcal{K}^*. \end{aligned}$$

Since  $B^\top \alpha + \beta d \in \text{int}(\mathcal{K}^*)$  is in the interior of  $\mathcal{K}^*$ , so is  $\frac{1}{2}(B^\top \alpha + \beta d)$ .

$$2\left(\frac{-H_k}{2} + \frac{1}{2}(B^\top \alpha + \beta d)\right)$$

then lies on the open segment between a point in  $\text{int}(\mathcal{K}^*)$  and a point in the cone (that may or be not be in the interior), it is therefore an interior point.  $\square$

In conclusion, near-optimal robustness has only added a finite number of constraints of the same nature (bilinear inequalities) to the reformulation proposed in [33]. 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  $a^\top b$ .

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

$$\min_{x,v} c_x^\top x + c_y^\top v \tag{23a}$$

$$\text{s.t. } Gx + Hv \leq q \tag{23b}$$

$$d^\top v \leq \phi(x) \tag{23c}$$

$$Ax + Bv \leq b \tag{23d}$$

$$Gx + Hz \leq q \quad \forall z \in \mathcal{Z}(x; \delta) \tag{23e}$$

$$v \in \mathbb{R}_+^{m_l} \tag{23f}$$

$$x \in \mathcal{X}. \tag{23g}$$

For a given pair  $(x, v)$ , each semi-infinite robust constraint (23e) can be reformulated as the objective value of the following adversarial problem:

$$\max_y H_k^T y \tag{24a}$$

$$\text{s.t. } (By)_i \leq b_i - (Ax)_i \quad \forall i \in \llbracket m_l \rrbracket \tag{24b} \quad (\alpha_i)$$

$$d^T y \leq d^T v + \delta \tag{24c} \quad (\beta)$$

$$y \in \mathbb{R}_+^{n_l}. \tag{24d}$$

Let  $(\alpha, \beta)$  be the dual variables associated with each group of constraints (24b–24c). The near-optimal robust version of Problem (23) is feasible only if the objective value of each  $k$ -th adversarial subproblem (24) is lower than  $q_k - (Gx)_k$ . The dual of Problem (24) is defined as:

$$\min_{\alpha, \beta} \alpha^T (b - Ax) + \beta (d^T v + \delta) \tag{25a}$$

$$\text{s.t. } B^T \alpha + \beta d \geq H_k \tag{25b}$$

$$\alpha \in \mathbb{R}_+^{m_l}, \beta \in \mathbb{R}_+. \tag{25c}$$

Based on Problem (3) 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 (23e). This is equivalent to the existence of a feasible dual solution  $(\alpha, \beta)$  certifying the feasibility of  $(x, v)$  within the near-optimal set  $\mathcal{Z}(x; \delta)$ . We obtain one pair of certificates  $(\alpha, \beta)$  for each upper-level constraint in  $\llbracket m_u \rrbracket$ , resulting in the following problem:

$$\min_{x, v, \alpha, \beta} c_x^T x + c_y^T v \tag{26a}$$

$$\text{s.t. } Gx + Hv \leq q \tag{26b}$$

$$d^T v \leq \phi(x) \tag{26c}$$

$$Ax + Bv \leq b \tag{26d}$$

$$\alpha_k^T (b - Ax) + \beta_k (d^T v + \delta) \leq q_k - (Gx)_k \quad \forall k \in \llbracket m_u \rrbracket \tag{26e}$$

$$B^T \alpha_k + \beta_k d \geq H_k \quad \forall k \in \llbracket m_u \rrbracket \tag{26f}$$

$$\alpha_k \in \mathbb{R}_+^{m_l}, \beta_k \in \mathbb{R}_+ \quad \forall k \in \llbracket m_u \rrbracket \tag{26g}$$

$$v \in \mathbb{R}_+^{n_l} \tag{26h}$$

$$x \in \mathcal{X}. \tag{26i}$$

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

$$d_j + \sum_i B_{ij} \lambda_i - \sigma_j = 0 \quad \forall j \in \llbracket n_l \rrbracket \tag{27a}$$

$$0 \leq b_i - (Ax)_i - (Bv)_i \perp \lambda_i \geq 0 \quad \forall i \in \llbracket m_l \rrbracket \tag{27b}$$

$$0 \leq v_j \perp \sigma_j \geq 0 \quad \forall j \in \llbracket n_l \rrbracket \tag{27c}$$

$$\sigma \geq 0, \lambda \geq 0 \tag{27d}$$

where  $\perp$  defines a complementarity constraint. A common technique to linearize Constraints (27b–27c) is the “big-M” reformulation, introducing auxiliary binary variables with primal



and dual upper bounds. The resulting formulation has a weak continuous relaxation. Furthermore, a correct choice of bounds is itself an  $\mathcal{NP}$ -hard Problem [38], and the incorrect choice of these bounds can lead to cutting valid and potentially optimal solutions [39]. 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:

$$\min_{x,v,\lambda,\sigma,\alpha,\beta} c_x^\top x + c_v^\top v \tag{28a}$$

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

$$Ax + Bv \leq b \tag{28c}$$

$$d_j + \sum_i \lambda_i B_{ij} - \sigma_j = 0 \quad \forall j \in \llbracket n_l \rrbracket \tag{28d}$$

$$0 \leq \lambda_i \perp A_i x + B_i v - b_i \leq 0 \quad \forall i \in \llbracket m_l \rrbracket \tag{28e}$$

$$0 \leq \sigma_j \perp v_j \geq 0 \quad \forall j \in \llbracket n_l \rrbracket \tag{28f}$$

$$x \in \mathcal{X} \tag{28g}$$

$$\alpha_k^\top (b - Ax) + \beta_k (d^\top v + \delta) \leq q_k - (Gx)_k \quad \forall k \in \llbracket m_u \rrbracket \tag{28h}$$

$$\sum_{i=1}^{m_l} B_{ij} \alpha_{ki} + \beta_k d_j \geq H_{kj} \quad \forall k \in \llbracket m_u \rrbracket, \forall j \in \llbracket n_l \rrbracket \tag{28i}$$

$$\alpha_k \in \mathbb{R}_+^{m_l}, \beta_k \in \mathbb{R}_+ \quad \forall k \in \llbracket m_u \rrbracket. \tag{28j}$$

Problem (28) is a single-level problem and has a closed form. However, constraints (28h) 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 (28h) 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 the rest of this paper, when the value of a variable  $a$  is fixed in a given problem, we will denote it with  $\bar{a}$ . For fixed values  $(\bar{x}, \bar{v})$  of  $(x, v)$ ,  $m_u$  dual adversarial subproblems (25) are defined. The optimal value of each  $k$ -th subproblem must be lower than  $q_k - (G\bar{x})_k$  for near-optimality robustness to hold. The feasible region of each subproblem is defined by (28h–28j) and is independent of  $(\bar{x}, \bar{v})$ ; their objective functions are linear in  $(\alpha, \beta)$ . Following Proposition 3 and by weak duality, Problem (25) is bounded. If, moreover, Problem (25) is feasible, at least a vertex of the polytope (28i–28j) is an optimal solution. Following these observations, Constraints (28h–28j) 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. We must highlight that disjunctive formulations are well established in the bilevel literature to express the complementarity constraints from the lower-level KKT conditions [40–42]. However, the bilinear reformulation of near-optimality robustness constraints does not possess the same structure and thus cannot leverage similar techniques.

Let  $\mathcal{V}_k$  be the number of vertices of the  $k$ -th subproblem and  $\bar{\alpha}_k^l, \bar{\beta}_k^l$  be the  $l$ -th vertex of the  $k$ -th subproblem. Constraints (28h–28j) can be written as:

$$\bigvee_{l=1}^{\mathcal{V}_k} \sum_{i=1}^{m_l} \bar{\alpha}_{ki}^l (b - Ax)_i + \bar{\beta}_k^l \cdot (d^\top v + \delta) \leq q_k - (Gx)_k \quad \forall k \in \llbracket m_u \rrbracket, \quad (29)$$

where  $\bigvee_{i=1}^N C_i$  is the disjunction (logical “OR”) operator, expressing the constraint that at least one of the constraints  $C_i$  must be satisfied. These disjunctions are equivalent to indicator constraints [43].

This reformulation of bilinear constraints based on the polyhedral description of the  $(\alpha, \beta)$  feasible space is similar to the Benders decomposition[44]. 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 (29) are equivalent to the following formulation, using set cover and SOS1 constraints:

$$\theta_k^l \in \{0, 1\} \quad \forall k \in \llbracket m_u \rrbracket, \forall l \in \mathcal{V}_k \quad (30a)$$

$$\omega_k^l \geq 0 \quad \forall k \in \llbracket m_u \rrbracket, \forall l \in \mathcal{V}_k \quad (30b)$$

$$(b - Ax)^\top \bar{\alpha}_k^l + \bar{\beta}_k^l (d^\top v + \delta) - \omega_k^l \leq q_k - (Gx)_k \quad \forall k \in \llbracket m_u \rrbracket, \forall l \in \mathcal{V}_k \quad (30c)$$

$$\sum_{l=1}^{\mathcal{V}_k} \theta_k^l \geq 1 \quad \forall k \in \llbracket m_u \rrbracket \quad (30d)$$

$$SOS1(\theta_k^l, \omega_k^l) \quad \forall k \in \llbracket m_u \rrbracket, \forall l \in \mathcal{V}_k, \quad (30e)$$

where  $SOS1(a, b)$  expresses a SOS1-type constraint between the variables  $a$  and  $b$ . 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 (28). 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. We next illustrate the extended formulation on the following example.

### 4.2 Bounded example

Consider the bilevel linear problem defined by the following data:

$$\begin{aligned} x \in \mathbb{R}_+, y \in \mathbb{R}_+ \\ G = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \quad H = \begin{bmatrix} 4 \\ 2 \end{bmatrix} \quad q = \begin{bmatrix} 11 \\ 13 \end{bmatrix} \quad c_x = [1] \quad c_y = [-10] \\ A = \begin{bmatrix} -2 \\ 5 \end{bmatrix} \quad B = \begin{bmatrix} -1 \\ -4 \end{bmatrix} \quad b = \begin{bmatrix} -5 \\ 30 \end{bmatrix} \quad d = [1]. \end{aligned}$$

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.

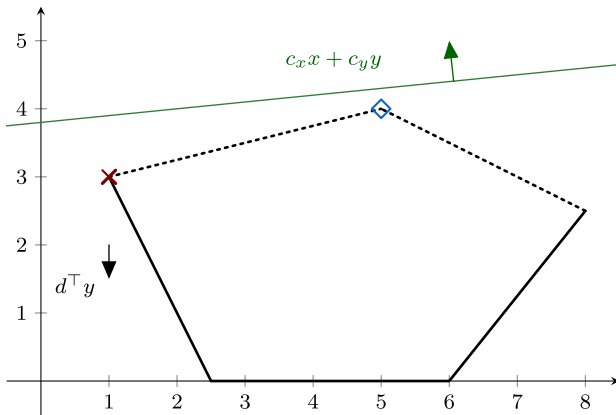


Fig. 3 Representation of the bilevel problem

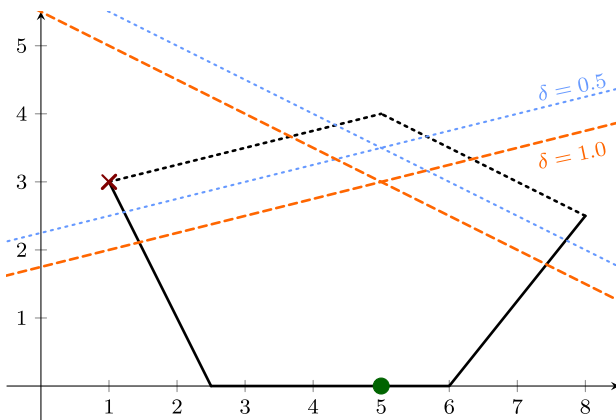


Fig. 4 Near-optimal robustness constraints

The feasible space for  $(\alpha, \beta)$  is given by:

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

This feasible space can be described as a set of extreme points and rays. It consists in this case of one extreme point  $(\bar{\alpha}_{ki} = 0, \bar{\beta}_1 = 4, \bar{\beta}_2 = 2)$  and 4 extreme rays. The  $(x, v)$  solution needs to be valid for the corresponding near-optimality conditions:

$$\begin{aligned}
 & \bar{\beta}_1 (v + \delta) \leq 11 + x \\
 & \bar{\beta}_2 (v + \delta) \leq 13 - x.
 \end{aligned}$$

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 disk at  $(5, 0)$  in Fig. 4.

### 4.3 Valid inequalities

The extended formulation can be tackled directly in a branch-and-cut framework. Nevertheless, we propose two groups of valid inequalities to tighten the formulation.

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

$$(Gx)_k + (Hv)_k \leq q_k \quad \forall k \in \llbracket m_u \rrbracket.$$

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 [45] and based on strong duality of the lower level. We only implement the valid inequalities for the root node, which are the primary focus of [45]:

$$\langle \lambda, b \rangle + \langle v, d \rangle \leq \langle A^+, \lambda \rangle, \quad (31)$$

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:

$$A_i^+ = \max_{x, v, \lambda} \langle A_i, x \rangle \quad (32a)$$

$$\text{s.t. } Gx + Hv \leq q \quad (32b)$$

$$Ax + Bv \leq b \quad (32c)$$

$$d + B^\top \lambda \geq 0 \quad (32d)$$

$$x \in \mathcal{X}, v \geq 0, \lambda \geq 0 \quad (32e)$$

$$(x, v, \lambda) \in \mathcal{Y}, \quad (32f)$$

where  $\mathcal{Y}$  is the set containing all valid inequalities (31).

The method proposed in [45] 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, similar to domain propagation techniques:

1. Solve Problem (32a)  $\forall i \in \llbracket m_l \rrbracket$  and obtain  $A^+$ ;
2. If  $\exists i$ ,  $A_i^+$  is unbounded, terminate;
3. Otherwise, add Constraint (31) to (32f) 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 (31) is  $+\infty$ , the constraint is trivial and cannot be improved upon. Otherwise, each iteration improves the bound until the convergence of  $A^+$ .

## 5 Solution algorithm

In this section, we describe the direct solution method for the extended formulation, derive an exact variant lazifying the subproblem exploration and provide a heuristic providing a near-optimal robust solution.

Solving the extended formulation can be done by first optimizing the dual adversarial subproblems, enumerating their vertices; if any subproblem is infeasible, the procedure can be terminated as the instance cannot have a robust solution.

### 5.1 Lazy subproblem expansion

Directly solving the extended formulation of NRB quickly becomes computationally demanding when the problem size increases, more lower-level constraints imply more vertices in each dual adversarial subproblem and thus larger disjunctive constraints, more upper-level constraints imply more dual adversarial subproblems and thus more disjunctive constraints. We next provide an exact method that builds upon this formulation while avoiding adding the whole set of disjunctive constraints upfront.

Each dual adversarial subproblem must be feasible with an objective value of less than  $q_k - G_k^\top x$  for the NRB instance to be near-optimal robust. For  $\mathcal{S} \subseteq \llbracket m_u \rrbracket$ , we denote by  $\overline{\text{NRB}}(\mathcal{S})$  the NRB formulation with the near-optimality robustness constraints replaced with:

$$G_k^\top x + H_k^\top v \leq q_k \quad \forall z \in \mathcal{Z}(x; \delta) \quad \forall k \in \mathcal{S}. \tag{33}$$

$\overline{\text{NRB}}(\emptyset)$  does not have any near-optimality robustness constraint and corresponds to the optimistic bilevel problem, while  $\overline{\text{NRB}}(\llbracket m_u \rrbracket)$  is equivalent to the extended formulation of NRB and integrates all near-optimality robustness constraints.  $\overline{\text{NRB}}(\mathcal{S})$  is trivially a relaxation of NRB since only a subset of the robust constraints is applied.

Furthermore, for a given pair  $(\bar{x}, \bar{v})$ , verifying its near-optimality robustness with respect to the  $k$ -th upper-level constraint can be done with an auxiliary linear optimization problem:

$$\begin{aligned} \min_{\alpha, \beta} & (b - A\bar{x})^\top \alpha + (d^\top \bar{v} + \delta)\beta \\ \text{s.t.} & B^\top \alpha + \beta d \geq H_k \\ & \alpha \in \mathbb{R}_+^{m_l}, \beta \in \mathbb{R}_+. \end{aligned} \tag{34}$$

We consider a solution to the auxiliary problem as ‘‘robust’’ if the optimal value is below  $q_k - G_k^\top \bar{x}$ , in which case the constraint is robust to near-optimal deviations of the lower level. Algorithm 1 starts from  $\mathcal{S} = \emptyset$  and iteratively adds violated constraints from (33) that ensure the near-optimality robustness of some upper-level constraints. We thus qualify it as lazy in contrast to the extended formulation using all disjunctive constraints upfront.

At any given iteration, the set  $\hat{\mathcal{S}}$  is the complement of  $\mathcal{S}$ ,

$$\hat{\mathcal{S}} \equiv \llbracket m_u \rrbracket \setminus \mathcal{S}$$

and  $\mathcal{S}_{opt}$  contains the set of upper-level constraint indices that are robust for a current iterate. If  $\hat{\mathcal{S}}$  is empty, there is no upper-level constraint that is not either added (already in  $\mathcal{S}$ ) or already optimal (in  $\mathcal{S}_{opt}$ ). Given that  $\overline{\text{NRB}}$  is a relaxation of NRB,  $\hat{\mathcal{S}} = \emptyset$  implies that the optimum is reached.

At any given iterate of Algorithm 1,  $\overline{\text{NRB}}(\mathcal{S})$ , a MILP with disjunctive constraints, is solved from scratch. The advantage over the extended formulation is that each of these MILPs is smaller and contains fewer indicator constraints than the extended formulation.

**Algorithm 1** Lazy Subproblem Expansion

```

1: function LAZYSUBPROBLEMEXPANSION( $G, H, q, c_x, c_y, A, B, b, d, \delta$ )
2:    $\overline{\text{NRB}} \leftarrow$  optimistic model with parameters  $G, H, q, c_x, c_y, A, B, b, d$ 
3:   if Optimistic infeasible then
4:     Terminate with optimistic infeasible
5:   end if
6:    $\hat{S} \leftarrow \llbracket m_u \rrbracket$ 
7:    $w_k \leftarrow \{\} \forall k \in \llbracket m_u \rrbracket$ 
8:   while  $\hat{S}$  not empty do ▷ subproblem exploration phase
9:      $\mathcal{S}_{opt} \leftarrow \emptyset$ 
10:     $(\hat{x}, \hat{v}) \leftarrow$  current solution
11:    Choose  $k \in \hat{S}$ 
12:    Solve  $k$ -th dual subproblem (34) parameterized by  $(\hat{x}, \hat{v})$ 
13:    if Solution is robust then
14:       $\mathcal{S}_{opt} \leftarrow \mathcal{S}_{opt} \cup \{k\}$ 
15:       $\hat{S} \leftarrow \hat{S} \setminus \{k\}$ 
16:    else ▷ subproblem expansion phase
17:      Add new variables  $w_{kl} \in \{0, 1\} \forall l \in 1 \dots |\mathcal{V}_k|$  to  $\overline{\text{NRB}}$ 
18:      Add constraint  $\sum_{l=1}^{|\mathcal{V}_k|} w_{kl} \geq 1$  to  $\overline{\text{NRB}}$ 
19:      for  $l \in 1 \dots |\mathcal{V}_k|$  do
20:         $(\bar{\alpha}_k^l, \bar{\beta}_k^l) \leftarrow \mathcal{V}_k^l$ 
21:        Add indicator constraint to  $\overline{\text{NRB}}$ :
22:         $w_{kl} = 1 \Rightarrow (b - Ax)^\top \bar{\alpha}_k^l + (d^\top v + \delta) \bar{\beta}_k^l \leq q_k - G_k^\top x$ 
23:      end for
24:    end if
25:    Solve current iterate  $\overline{\text{NRB}}$ 
26:    if Infeasible then
27:      Terminate with infeasible  $k$ -th subproblem
28:    end if
29:     $\hat{S} \leftarrow \hat{S} \cap \mathcal{S}_{opt}$ 
30:  end while
31:  return  $(\hat{x}, \hat{v}, w)$ 
32: end function

```

**5.2 Single vertex heuristic**

We now present a heuristic method with a computational cost close to that of the canonical bilevel problem and computing a high-quality bilevel-feasible near-optimal robust solution. It only requires optimizing a sequence of at most  $m_u$  MILPs with the same variables as the canonical bilevel problem and at most  $m_u$  additional linear constraints, instead of the disjunctive constraints with a number of terms equal to the number of vertices of the dual adversarial polyhedron.

At any given iterate,  $\mathcal{C}$  is the set of upper-level constraint indices that were already added to the model. New constraints are added in a batched fashion, with the batch size controlled by the  $\eta$  parameter. Each  $k$ -th subproblem is added only once, by selecting a single vertex  $(\alpha, \beta)$  and using it to enforce the constraint

$$(b - Ax)^\top \alpha + (d^\top v + \delta) \beta \leq q_k - G_k^\top x.$$

Unlike the exact algorithms, Algorithm 2 initializes a MILP model and iteratively adds linear constraints to it. This procedure therefore lends itself to warm starts and single-tree formulations.

**Algorithm 2** Single-Vertex Heuristic

```

1: function SINGLEVERTEXHEURISTIC( $G, H, q, c_x, c_y, A, B, b, d, \delta, \eta$ )
2:    $P \leftarrow$  optimistic model with parameters  $(G, H, q, c_x, c_y, A, B, b, d)$ 
3:   if Optimistic infeasible then
4:     Terminate with optimistic infeasible
5:   end if
6:   if  $\exists k$ ,  $k$ -th dual adversarial subproblem is infeasible then
7:     Terminate, no robust solution exists
8:   end if
9:   count  $\leftarrow$  1
10:   $C \leftarrow \emptyset$ 
11:  while count  $>$  0 do ▷ Outer loop
12:     $k \leftarrow$  1
13:    count  $\leftarrow$  0
14:     $(\hat{x}, \hat{v}) \leftarrow$  current solution of  $P$ 
15:    while count  $\leq \eta$  and  $k \leq m_u$  do
16:       $(\alpha, \beta) \leftarrow$  solution to  $k$ -th subproblem (34) parameterized by  $(\hat{x}, \hat{v})$ 
17:      if  $(b - A\hat{x})^\top \alpha + (d^\top \hat{v} + \delta)\beta > q_k - G_k^\top \hat{x}$  then
18:         $C \leftarrow C \cup \{k\}$ 
19:        Add constraint to  $P$ :  $(b - Ax)^\top \alpha + (d^\top v + \delta)\beta \leq q_k - G_k^\top x$ 
20:        count  $\leftarrow$  count + 1
21:        while  $k \in C$  do
22:           $k \leftarrow k + 1$ 
23:        end while
24:      end if
25:    end while
26:    Re-optimize  $P$ 
27:    if Current iterate infeasible then
28:      Terminate with no found solution
29:    end if
30:  end while
31:  return  $(\hat{x}, \hat{v})$ 
32: end function

```

**Proposition 5** Algorithm 2 terminates in at most  $m_u$  iterations of the outer loop beginning Line 11 and solves optimization problems with the same variables as the canonical optimization problem and at most  $m_u$  linear constraints.

**Proof** Algorithm 2 maintains a cache  $C$  of the subproblems that have been explored. For each subproblem, exactly one vertex is chosen, which minimizes

$$(b - A\hat{x})^\top \alpha + (d^\top \hat{v} + \delta)\beta$$

with  $(\hat{x}, \hat{v})$  the current iterate. For a chosen  $\hat{x}$ , the lower-level problem is feasible since  $\hat{v}$  is computed, so the dual problem cannot be unbounded. It cannot be infeasible since its feasibility domain depends only on the problem data and is verified Line 6. Therefore, a vertex  $(\alpha, \beta)_k$  is computed. If the objective is greater than  $q_k - G_k^\top \hat{x}$ , then the current iterate is not near-optimal robust with respect to the  $k$ -th constraint, and the linear constraint:

$$(b - Ax)^\top \alpha + (d^\top v + \delta)\beta \leq q_k - G_k^\top x$$

is added to the model. If no addition is made, the current iterate is near-optimal robust, the count variable remains at 0, and the outer loop exits, with the function returning the current iterate. Otherwise, the iterate was not near-optimal robust with respect to at least one upper-level constraint, which is turned into a constraint and added to the cache. The outer loop adds at least one constraint in a non-terminating iteration, therefore,  $m_u$  iterations suffice to add

all constraints. Moreover, each subproblem adds exactly one linear constraint, so at most  $m_u$  linear constraints are added to the formulation of the canonical problem.  $\square$

The  $\eta$  parameter controls the maximum number of linear constraints added for each outer iteration.  $\eta = 1$  implies that the algorithm reoptimizes the problem after adding a single constraint, while  $\eta \geq m_u$  will add all the linear constraints that correspond to upper-level constraints that are not near-robust at each iterate.

Finally, it can be noted that the single-vertex algorithm can be applied even when the number of vertices in the dual adversarial is infinite, i.e. when the lower-level problem is a convex optimization problem. The only modification is the optimization of the dual adversarial problem for fixed values of  $(x, v)$  at Line 16, where a convex optimization problem is solved instead of a linear one.

## 6 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.

### 6.1 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:

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

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 10,532 trials, the 200 medium-sized instances were obtained with 18,040 trials and the 100 large instances after 90,855 trials. A second dataset is created from the 50 MIPS/Random instances of the Bilevel Problem library [46], 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 [47] in JLD format, along with a reader to import them in Julia programs.

### 6.2 Computational setup

All experiments are carried out in Julia 1.6 [48] using the JuMP v0.21 modelling framework [49, 50]; the MILP solver is SCIP 7.0 [51] with SoPlex 5.0 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 [52] is used



to model the dual subproblem polyhedra with `CDDLib` [53] 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 workstation with 32GB of RAM and Intel Xeon 3.5GHz CPUs running Ubuntu 18.04LTS.

### 6.3 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 (23). The bilinear formulation is implemented with `SCIP` using `SoPlex` as the linear optimization solver and `Ipopt` as the nonlinear solver. `SCIP` handles the bilinear terms through bound computations and spatial branching. Out of all MIPS and ALTMIPS sets, only one instance is solved with the bilinear formulation within the time limit in half a second, a time similar to the extended formulation. The bilinear formulation not only runs out of time, but also of memory (a memory limit of 5000MB was fixed in this setting). The spatial branching required to handle the non-convex bilinear inequalities is thus more time- and memory-consuming than the branching over disjunctive constraints introduced by the extended formulation. Because of the vertices of the dual adversarial subproblems being optimal solutions as developed in Sect. 4.1, the disjunctive constraints explicitly constrain the optimality candidates to a discrete set instead of exploring a continuous set of  $(\alpha, \beta)$  solutions through spatial branching.

### 6.4 Robustness of optimistic solutions and influence of $\delta$

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

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

with  $\text{opt}(L)$  the lower-level objective value at the obtained optimistic 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  $0.01\text{opt}(L)$ . 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  $3\text{opt}(L)$ . 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.

### 6.5 Computation time of the different algorithms

Statistics on the computation times of the vertex enumeration and solution phases for the extended formulation on each set of instances are provided in Table 1 and Table 2.

**Table 1** Runtime statistics for the vertex enumeration (s)

Set	Mean	10% Quant	50% Quant	90% Quant
MIPS	15.79	0.15	2.38	53.37
ALTMIPS	0.17	0.05	0.13	0.42

**Table 2** Runtime statistics for the optimization phase (s)

Set	# Optimized	Mean	10% Quant	50% Quant	90% Quant
MIPS	47	474.657	2.583	71.219	1659.787
ALTMIPS	41	603.040	1.392	66.819	1751.774

The solution time is greater than the vertex enumeration phase which is thus not the bottleneck to solve NRB instances.

We also studied the sensitivity of NRB solutions to variations of  $\delta$  on small-size random instances. When  $\delta$  increases, the number of problems solved to optimality monotonically decreases; greater  $\delta$  values indeed reduce the set of feasible solutions to NRB. The optimal values of the upper-level only slightly increase with  $\delta$  and the lower-level objective value does not vary significantly with  $\delta$ .

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

The different solution methods are compared in Fig. 5. The single vertex heuristics with various batch sizes, labelled `random_sv_batchsize`, are the fastest to compute near-optimal robust solutions, slightly slower than solving the optimistic bilevel formulation. On exact methods, the lazy vertex enumeration Algorithm 1 outperforms the eager `extended` formulation and terminates within the time limit for all instances. One MIPS instance is bilevel feasible but does not possess a near-optimal robust solution. On 90 MIPS and 71 ALTMIPS instances, the single vertex heuristic successfully finds a solution while it terminates unsuccessfully in one and seven cases respectively. Furthermore, the heuristic runtime is fast enough compared to the exact methods that it could be improved by randomly removing some vertices that were added in the first iterations to then possibly reintroduce them to the formulation. Finally, we note that when the single-vertex heuristic finds a solution, it is one of high quality. On no MIPS instance and only two ALTMIPS instances is the objective value of the heuristic solution not equal to that of the exact method, verified with the lazy algorithm.

## 6.6 Implementation of valid inequalities

In the last group of experiments, we implement and investigate the impact of the valid inequalities defined in Sect. 4.3.

The valid inequality procedure found cuts for all 100 MIPS instances. For 98 of these, the procedure terminates after one iteration, the two other instances terminate with a cut after 4 and 8 iterations. On the 100 ALTMIPS instances, 18 are bilevel-infeasible, none of which had an infeasible high-point relaxation. For 12 of these instances, adding the valid inequalities was enough to prove their infeasibility. For all instances, the procedure computed non-trivial valid inequalities i.e. all coefficients of  $A^+$  are finite.

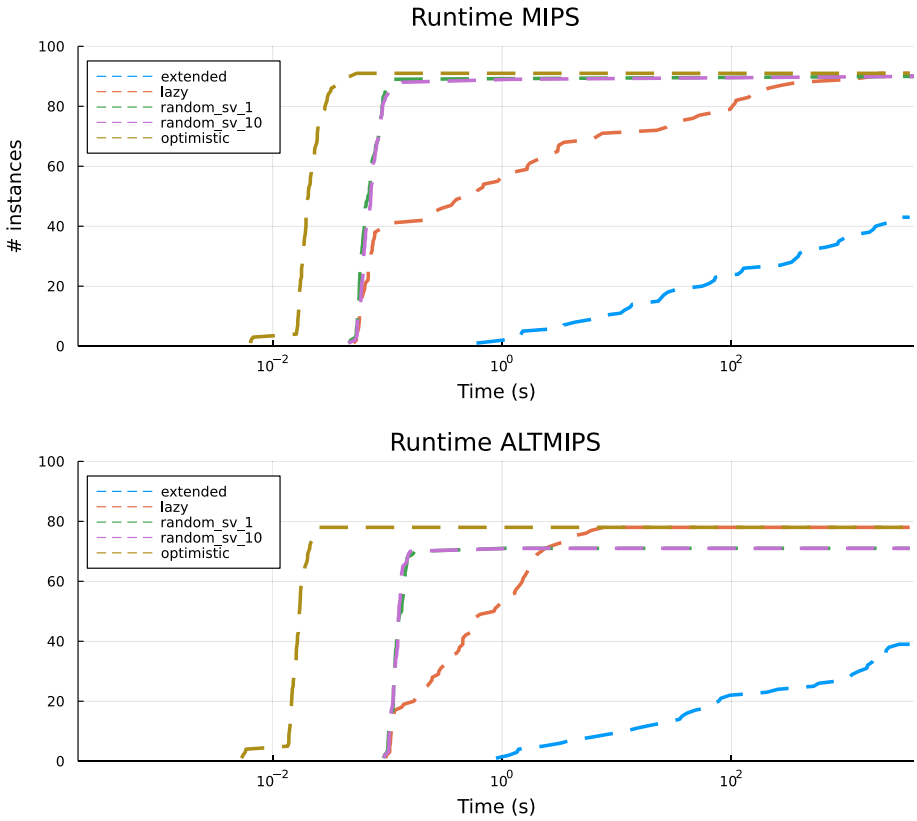


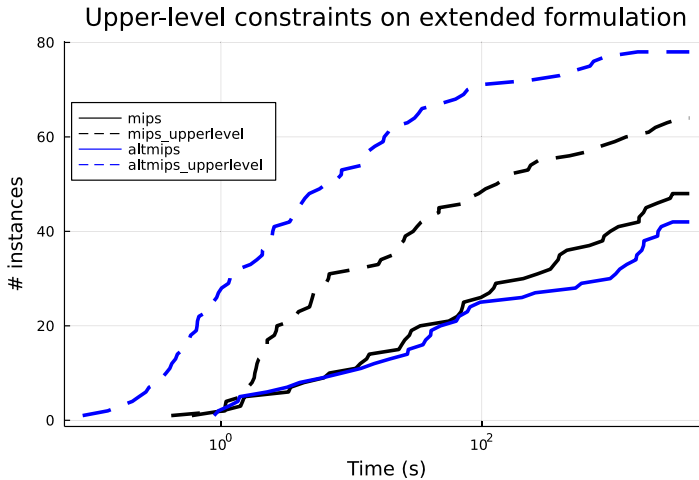
Fig. 5 Comparing the different solution methods on the two instance sets

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 a MILP. For all but two MIPS instances, a single iteration of the procedure computes the final valid inequality. 12 out of 100 ALTMIPS instances require more than one iteration with one instance requiring up to 32 iterations.

We also compared the total runtime for MIPS and ALTMIPS instances under near-optimality robustness constraints using  $\delta = 0.1$  with and without valid inequalities for all instances solved to optimality. Valid inequalities do not improve the runtime for NRB in either group of instances, This result is similar to the observations in [45] 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.

As shown in Fig. 6, the addition of primal upper-level constraints accelerates the resolution of the MIPS and even more of ALTMIPS instances and dominates the standard extended formulation.



**Fig. 6** Runtime for MIPS and ALTMIPS instances with and without upper-level constraints

## 7 Conclusion

In this work, we introduce near-optimal robust bilevel optimization, a specific formulation of the bilevel optimization problem does not make an explicit optimistic/pessimistic assumption about the reaction of the lower level, but instead seeks an optimal/decision at the upper level that is robust in the sense that it remains feasible even if the lower level deviates from its own optimality within a prescribed near-optimal set for the lower level. 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 NRB 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 NRB 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. This extended formulation lends itself to a faster exact lazy method and a single-vertex heuristic leveraging the problem structure to find high-quality feasible solutions. Numerical experiments highlight the efficiency of the extended method compared to the compact bilinear formulation, the impact of valid inequalities on both solution time and tightness of the linear relaxation. Finally, they highlight the interest of the lazy and heuristic formulation, showing a solution time only slightly slower than the corresponding

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Dempe, S., Zemkoho, A.: (eds.) *Bilevel Optimization: Theory, Algorithms, Applications and a Bibliography*, pp. 581–672. Springer International Publishing, Cham (2020)
2. Dempe, S.: *Foundations of Bilevel Programming*. Springer Science & Business Media (2002)
3. Stavroulakis, G.E., Günzel, H.: Optimal structural design in nonsmooth mechanics. In: *Multilevel Optimization: Algorithms and Applications*, pp. 91–115. Springer (1998)
4. Nicholls, M.G.: Optimizing the operations of an aluminium smelter using non-linear bi-level programming. In: *Multilevel Optimization: Algorithms and Applications*, pp. 117–148. Springer (1998)
5. Simon, H.A.: Theories of bounded rationality. *Decis. Organ.* **1**(1), 161–176 (1972)
6. Aumann, R.J.: Rationality and bounded rationality. In: *Cooperation: Game-Theoretic Approaches*, pp. 219–231. Springer (1997)
7. Aghassi, M., Bertsimas, D.: Robust game theory. *Math. Program.* **107**, 231–273 (2006)
8. Beck, Y., Ljubić, I., Schmidt, M.: A survey on bilevel optimization under uncertainty. *Eur. J. Oper. Res.* (2023)
9. Ordóñez, F., Stier-Moses, N.E.: Robust Wardrop equilibrium. In: Chahed, T., Tuffin, B. (eds.) *Network Control and Optimization*, pp. 247–256. Springer Berlin Heidelberg, Berlin, Heidelberg (2007)
10. Jain, M., Ordóñez, F., Pita, J., Portway, C., Tambe, M., Western, C., Paruchuri, P., Kraus, S.: Robust solutions in Stackelberg games: addressing boundedly rational human preference models. In: *Proc. of the AAAI 4th Multidisciplinary Workshop on Advances in Preference Handling* (2008)
11. Pita, J., Jain, M., Tambe, M., Ordóñez, F., Kraus, S.: Robust solutions to Stackelberg games: addressing bounded rationality and limited observations in human cognition. *Artif. Intell.* **174**(15), 1142–1171 (2010)
12. Yamkoğlu, I., Kuhn, D.: Decision rule bounds for two-stage stochastic bilevel programs. *SIAM J. Optim.* **28**(1), 198–222 (2018)
13. Zare, M.H., Ózaltun, O.Y., Prokopyev, O.A.: On a class of bilevel linear mixed-integer programs in adversarial settings. *J. Glob. Optim.* **71**(1), 91–113 (2018)
14. Zare, M.H., Prokopyev, O.A., Sauré, D.: On bilevel optimization with inexact follower. *Decis. Anal.* **17**(1), 74–95 (2020)
15. Beck, Y., Schmidt, M.: A robust approach for modeling limited observability in bilevel optimization. *Oper. Res. Lett.* **49**(5), 752–758 (2021)
16. Buchheim, C., Henke, D.: The robust bilevel continuous knapsack problem with uncertain coefficients in the follower’s objective. *J. Glob. Optim.* 1–22 (2022)
17. Wiesemann, W., Tsoukalas, A., Kleniati, P.-M., Rustem, B.: Pessimistic bilevel optimization. *SIAM J. Optim.* **23**, 353–380 (2013)
18. Loridan, P., Morgan, J.:  $\epsilon$ -regularized two-level optimization problems: approximation and existence results. In: Dolecki, S. (ed.) *Optimization*, pp. 99–113. Springer Berlin Heidelberg, Berlin, Heidelberg (1989)
19. Dempe, S., Kalashnikov, V., Prez-Valds, G.A., Kalashnykova, N.: *Bilevel Programming Problems: Theory, Algorithms and Applications to Energy Networks*. Springer Publishing Company, Incorporated (2015)
20. Still, G.: Generalized semi-infinite programming: theory and methods. *Eur. J. Oper. Res.* **119**(2), 301–313 (1999)
21. Goel, V., Grossmann, I.E.: A class of stochastic programs with decision dependent uncertainty. *Math. Program.* **108**, 355–394 (2006). <https://doi.org/10.1007/s10107-006-0715-7>
22. Besançon, M., Anjos, M.F., Brotcorne, L.: Complexity of near-optimal robust versions of multilevel optimization problems. *Optim. Lett.* **15**(8), 2597–2610 (2021)
23. Buchheim, C., Henke, D., Hommelsheim, F.: On the complexity of robust bilevel optimization with uncertain follower’s objective. *Oper. Res. Lett.* **49**(5), 703–707 (2021)
24. Gabrel, V., Murat, C., Thiele, A.: Recent advances in robust optimization: an overview. *Eur. J. Oper. Res.* **235**, 471–483 (2014)
25. Loridan, P., Morgan, J.: Weak via strong Stackelberg problem: new results. *J. Glob. Optim.* **8**, 263–287 (1996)
26. Besançon, M., Anjos, M.F., Brotcorne, L., Gomez-Herrera, J.A.: A bilevel approach for optimal price-setting of time-and-level-of-use tariffs. *IEEE Trans. Smart Grid* **11**(6), 5462–5465 (2020)
27. Cerulli, M., Archetti, C., Fernandez, E., Ljubic, I.: A bilevel approach for compensation and routing decisions in last-mile delivery. *arXiv preprint arXiv:2304.09170* (2023)
28. Liers, F., Schewe, L., Thürauf, J.: Radius of robust feasibility for mixed-integer problems. *INFORMS J. Comput.* (2021)
29. Goberna, M.A., Jeyakumar, V., Li, G., Linh, N.: Radius of robust feasibility formulas for classes of convex programs with uncertain polynomial constraints. *Oper. Res. Lett.* **44**(1), 67–73 (2016). <https://doi.org/10.1016/j.orl.2015.11.011>

30. Pardalos, P., Rosen, J.: Methods for global concave minimization: a bibliographic survey. *SIAM Rev.* **28**(3), 367–379 (1986)
31. Benson, H., Horst, R.: A branch and bound-outer approximation algorithm for concave minimization over a convex set. *Comput. Math. Appl.* **21**(6), 67–76 (1991)
32. Boyd, S., Vandenberghe, L.: *Convex Optimization*. Cambridge University Press (2004)
33. Dempe, S.: 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)
34. Scholtes, S.: Convergence properties of a regularization scheme for mathematical programs with complementarity constraints. *SIAM J. Optim.* **11**(4), 918–936 (2001)
35. Schewe, L., Schmidt, M.: Computing feasible points for binary MINLPs with MPECs. *Math. Program. Comput.* **11**, 95–118 (2019). <https://doi.org/10.1007/s12532-018-0141-x>
36. Nesterov, Y., Nemirovskii, A.: *Interior-Point Polynomial Algorithms in Convex Programming*, vol. 13. SIAM (1994)
37. Scheel, H., Scholtes, S.: Mathematical programs with complementarity constraints: stationarity, optimality, and sensitivity. *Math. Oper. Res.* **25**(1), 1–22 (2000)
38. Kleinert, T., Labbé, M., Plein, F., Schmidt, M.: There’s no free lunch: on the hardness of choosing a correct big-M in bilevel optimization. *Oper. Res.* **68**(6), 1716–1721 (2020)
39. Pineda, S., Morales, J.M.: Solving linear bilevel problems using big-Ms: not all that glitters is gold. *IEEE Trans. Power Syst.* **34**, 2469–2471 (2019)
40. Audet, C., Haddad, J., Savard, G.: Disjunctive cuts for continuous linear bilevel programming. *Optim. Lett.* **1**(3), 259–267 (2007)
41. Siddiqui, S., Gabriel, S.A.: An SOS1-based approach for solving MPECs with a natural gas market application. *Netw. Spat. Econ.* **13**(2), 205–227 (2013)
42. Pineda, S., Bylling, H., Morales, J.: Efficiently solving linear bilevel programming problems using off-the-shelf optimization software. *Optim. Eng.* **19**(1), 187–211 (2018)
43. Bonami, P., Lodi, A., Tramontani, A., Wiese, S.: On mathematical programming with indicator constraints. *Math. Program.* **151**, 191–223 (2015). <https://doi.org/10.1007/s10107-015-0891-4>
44. Rahmaniani, R., Crainic, T.G., Gendreau, M., Rei, W.: The Benders decomposition algorithm: a literature review. *Eur. J. Oper. Res.* **259**(3), 801–817 (2017)
45. Kleinert, T., Labbé, M., Plein, F., Schmidt, M.: Closing the gap in linear bilevel optimization: a new valid primal-dual inequality. *Optim. Lett.* **15**, 1027–1040 (2021)
46. Ralphs, T., Tahernejad, S.: Bilevel optimization problem library v0.1 (2020). <https://coral.ise.lehigh.edu/data-sets/bilevel-instances/>
47. Besançon, M., Anjos, M.F., Brotcorne, L.: Near-optimal robust bilevel linear instances (2020). <https://doi.org/10.5281/zenodo.4009108>
48. Bezanson, J., Edelman, A., Karpinski, S., Shah, V.: Julia: a fresh approach to numerical computing. *SIAM Rev.* **59**, 65–98 (2017)
49. Dunning, I., Huchette, J., Lubin, M.: JuMP: a modeling language for mathematical optimization. *SIAM Rev.* **59**(2), 295–320 (2017)
50. Legat, B., Dowson, O., Garcia, J.D., Lubin, M.: MathOptInterface: a data structure for mathematical optimization problems. *INFORMS J. Comput.* **34**(2), 672–689 (2022)
51. Gamrath, G., Anderson, D., Bestuzheva, K., Chen, W.-K., Eifler, L., Gasse, M., Gemander, P., Gleixner, A., Gottwald, L., Halbig, K., et al.: The SCIP optimization suite 7.0 (2020). <https://optimization-online.org/2020/03/7705/>
52. Legat, B., Deits, R., Evans, O., Goretkin, G., Koolen, T., Huchette, J., Oyama, D., Forets, M., guberger, Schwarz, R., Saba, E., Coleman, C.: JuliaPolyhedra/Polyhedra.jl: v0.5.1 (2019). <https://doi.org/10.5281/zenodo.3066026>
53. Legat, B., Deits, R., Forets, M., Oyama, D., Pacaud, F., Saba, E.: JuliaPolyhedra/CDDL.jl: v0.5.2 (2019). <https://doi.org/10.5281/zenodo.2671760>

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.