Operations Research Perspectives 13 (2024) 100315

Contents lists available at ScienceDirect

Operations Research Perspectives

journal homepage: www.elsevier.com/locate/orp

δ-perturbation of bilevel optimization problems: An error bound analysis

Margarita Antoniou [a,b,*], Ankur Sinha [c], Gregor Papa [a,b]

[a] Computer Systems Department, Jožef Stefan Institute, Ljubljana, Slovenia
[b] Jožef Stefan International Postgraduate School, Ljubljana, Slovenia
[c] Operations & Decision Sciences, Indian Institute of Management, Ahmedabad, Gujarat, India

ARTICLE INFO

Keywords:
Bilevel optimization
Optimistic bilevel problem
Pessimistic bilevel problem
Perturbation method
Error bound
Iterative heuristics
Population-based methods
Evolutionary Algorithms

ABSTRACT

In this paper, we analyze a perturbed formulation of bilevel optimization problems, which we refer to as δ-perturbed formulation. The δ-perturbed formulation allows to handle the lower level optimization problem efficiently when there are multiple lower level optimal solutions. By using an appropriate perturbation strategy for the optimistic or pessimistic formulation, one can ensure that the optimization problem at the lower level contains only a single (approximate) optimal solution for any given decision at the upper level. The optimistic or the pessimistic bilevel optimal solution can then be efficiently searched for by algorithms that rely on solving the lower level optimization problem multiple times during the solution search procedure. The δ-perturbed formulation is arrived at by adding the upper level objective function to the lower level objective function after multiplying the upper level objective by a small positive/negative δ. We provide a proof that the δ-perturbed formulation is approximately equivalent to the original optimistic or pessimistic formulation and give an error bound for the approximation. We apply this scheme to a class of algorithms that attempts to solve optimistic and pessimistic variants of bilevel optimization problems by repeatedly solving the lower level optimization problem.

1. Introduction

In the realm of optimization, the search for efficient solutions to complex decision-making problems is difficult in numerous fields such as engineering, economics, transport, and management. Among the diverse set of optimization frameworks, bilevel optimization [1,2] stands out as a fundamental approach to modeling hierarchical decision-making processes. At its core, bilevel optimization involves two tiers of decision-makers: an upper level decision-maker, often regarded as the leader, and a lower level decision-maker, referred to as the follower. To optimize the objective function, the leader initially determines certain values for its decision vector, x. This process establishes the feasible region for the follower's problem. Subsequently, the follower proceeds to determine its own optimal decision vector, y, within the lower level feasible region, which is parameterized with respect to x. In order to optimize its decision vector, the leader must take into account the follower's response to any given x, as this generally impacts the feasibility and also the objective of the leader. Similarly to a Stackelberg game [3], this hierarchical structure exemplifies real-world scenarios in which upper level entities, acting as leaders, attempt to solve upper level optimization problems, while lower level entities, acting as followers, make decisions to optimize their own objectives within defined constraints.

Bilevel optimization is of significant importance because of its wide-ranging applicability in various practical domains. Applications include scenarios involving Stackelberg games, such as the toll-setting problem within the literature on transportation policy [4–6]. Another example is presented in the study by [7], where a bilevel approach is proposed for retailers to invest in renewable-based systems, integrating capacity sizing and pricing strategies while ensuring profitability and safety. Furthermore, bilevel optimization has been widely employed in various other fields, including chemical engineering [8,9], taxation [10,11], security protocols [12–14], production planning [15], and firm behavior analysis [16]. It also finds applications in feature selection algorithms [17] and mobile cloud computing architectures [18]. In machine learning, bilevel optimization has been applied to address hyperparameter tuning challenges [19,20]. For a comprehensive overview of bilevel optimization, interested readers can refer to the following sources: [2,21–23].

Bilevel problems are known to be NP-hard,even in the simplest formulation [24]. Various classical approaches are used to solve several classes of bilevel problems, such as the Karush–Kuhn–Tucker (KKT) method [25,26], trust region [27–29], and penalty methods [30–32]. Gradient descent methods are commonly used in bilevel optimization when certain regularity conditions are met [33,34]. These methods

* Corresponding author.
E-mail addresses: margarita.antoniou@ijs.si (M. Antoniou), asinha@iima.ac.in (A. Sinha), gregor.papa@ijs.si (G. Papa).

https://doi.org/10.1016/j.orp.2024.100315
Received 14 April 2024; Received in revised form 30 July 2024; Accepted 26 August 2024
Available online 31 August 2024

can struggle in the presence of non-convexities and discontinuities due to their reliance on differentiable functions and attractiveness to local minima [35]. In bilevel problems where gradients may be difficult to compute or may not exist or the problem is black-box, population-based methods offer advantages. Unlike gradient-based methods, population-based approaches do not depend on gradient information, rather converge based on function and constraint evaluations allowing greater flexibility for a variety of bilevel optimization problems [36]. This includes evolutionary algorithms (EAs) such as genetic algorithm [37–40], differential evolution [41,42], Bayesian optimization [43], Random Search [44], other metaheuristics [45–47], hybrid methods [48–50], surrogate-assisted methods [51,52] etc. Such algorithms use an efficient sampling strategy to sample upper level vectors and solve the corresponding lower level optimization problems to converge towards the bilevel optimal solution. A survey on metaheuristics for bilevel optimization can be found here [53].

The optimization techniques mentioned above often struggle to effectively address the complexities inherent in bilevel optimization. Specifically, when the set of solutions to the follower's reaction is not a singleton, the problem becomes ill-posed, and the leader cannot simply optimize her choice but needs to make some assumptions. The two extreme assumptions commonly used in the bilevel programming community are optimistic and pessimistic approaches. The optimistic approach assumes that the lower level decision-maker, in case of indifference with respect to its own objective, will select the best possible solution for the upper level decision-maker, aiming to improve the upper level objective. In pessimistic approach, the follower aims to worsen the upper level objective function by considering the worst-case scenario for the leader, whenever it faces indifference between solutions while optimizing its own objective. The worst-case scenario, for a bilevel problem where the leader is a minimizer, is typically expressed as the minimization of the maximum of the upper level objective function over a feasible set dependent on the follower choices. This leads to the pessimistic bilevel problem, which is also known as the weak Stackelberg game [54].

A comprehensive review of pessimistic bilevel optimization problems can be found in the work by [55]. Additionally, [56] provide a detailed examination of various methods used to address these problems, and [57] discuss their optimality conditions. Primarily, classical mathematical techniques, including the kth best algorithm [58] and penalty-based approaches [59,60], have been employed to tackle pessimistic linear bilevel problems. Moreover, approximation methods have been explored to solve bilevel problems in the works of [61,62].

To address bilevel optimization problems with multiple lower level optimal solutions, we explore the perturbation-based approach, initially described by Molodtsov in [63]. The author uses a series of strong Stackelberg games to solve the weak problem or the pessimistic bilevel problem. In the paper [64], the concept of perturbation of the lower level is used on a bilevel linear resource control problem. The same formulation is then also mentioned in a review article later in [65]. In [66], the reformulation is used for linear bilevel problems to find optimistic, pessimistic, and intermediate solutions, assuming some cooperation between levels. In [67], the authors utilized the Molodtsov method to calculate a series of solutions for the strong Stackelberg problem, i.e. optimistic bilevel problem. Then, they examined the connections between these solutions and solutions for the pessimistic bilevel problem. To our best knowledge, none of the perturbation-based studies have analyzed the error bounds for the perturbed problem either mathematically or computationally. This paper contributes to the bilevel literature by performing a mathematical analysis of the perturbed problem and also demonstrating the results computationally for a number of test cases.

There are a number of studies for optimistic bilevel problems, however, only limited number of studies exist for pessimistic bilevel problems. For instance, many of the nested EA approaches *assume an optimistic position* while solving the problem mainly due to the difficulties resulting from the complex interplay between optimization tasks at the upper and lower levels.

It should be mentioned that the pessimistic bilevel problem, in its original form, requires an extra optimization step, hence it turns into a trilevel problem. In a bilevel problem, where the follower has a single optimal solution for any given leader decision, the leader optimizes its decision considering the follower's (unique) optimal response. In the presence of multiple optimal solutions for the follower's problem for any given leader's decision a pessimistic or optimistic position is considered. In the pessimistic case the leader additionally considers the worst-case scenario within the follower's optimal responses. This adds a third level of optimization task effectively making the optimization problem a three level problem. The use of nested approaches become prohibitively costly while trying to solve the pessimistic problem in its original form. In this paper we demonstrate the application of the perturbation-based formulation to pessimistic bilevel problems and solve them using a population-based algorithm. The perturbation-based formulation is beneficial for optimistic bilevel problems as well, as it transforms the lower level problem with non-unique solutions into one with approximately unique solutions.

In light of the contributions made in the literature and existing gaps, this article seeks to address the following research questions:

- What are the theoretical proofs and error bounds of the equivalent perturbed bilevel problem?
- How can the perturbation method be applied effectively to solve optimistic and pessimistic bilevel optimization problems via sampling-based methods?

The primary aim of this paper is to establish a framework for addressing ill-posed bilevel optimization problems. Specifically, we provide rigorous proofs for both optimistic and pessimistic bilevel formulations that relies on the lower level optimal value function mapping. Additionally, we use a nested differential evolution (DE) to solve a number of bilevel test problems and a principal–agent problem, wherein the lower level contains multiple optimal solutions. Through a computational study we offer insight into the accuracy and efficiency of the perturbed formulation for varying levels of perturbation. A comparison of the computational performance is drawn with the theoretical results.

The paper is structured as follows: In Section 2, we introduce the bilevel optimization problem and the related terminology. Thereafter, in Section 3, we provide an analysis of the δ-perturbation of ill-posed bilevel problems for both an optimistic and pessimistic position. In Section 4, we provide an illustrative example. We use a class of optimization algorithms that solve bilevel problems using a nested approach to show that both optimistic and pessimistic formulations can be effectively solved using perturbation-based approach, referred to as δ-perturbation, in Section 5. Section 6 concludes the paper and gives some ideas for future research.

2. Bilevel problem definition

In a bilevel problem, each level operates with its own distinct objectives, constraints, and variables. There are two categories of decision variables (vectors): upper level decision vectors and lower level decision vectors. The lower level optimization is essentially a parametric optimization problem, solved with respect to the lower level decision vectors while considering the upper level decision vectors as parameters. Importantly, the lower level optimization problem acts as a constraint for the upper level optimization problem. For the upper-level objective function $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and the lower level objective function $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, the bilevel problem is given by:

$$\text{“}\min_{x\in X, y\in Y}\text{”}\; F(x, y) \tag{1}$$

subject to (2)

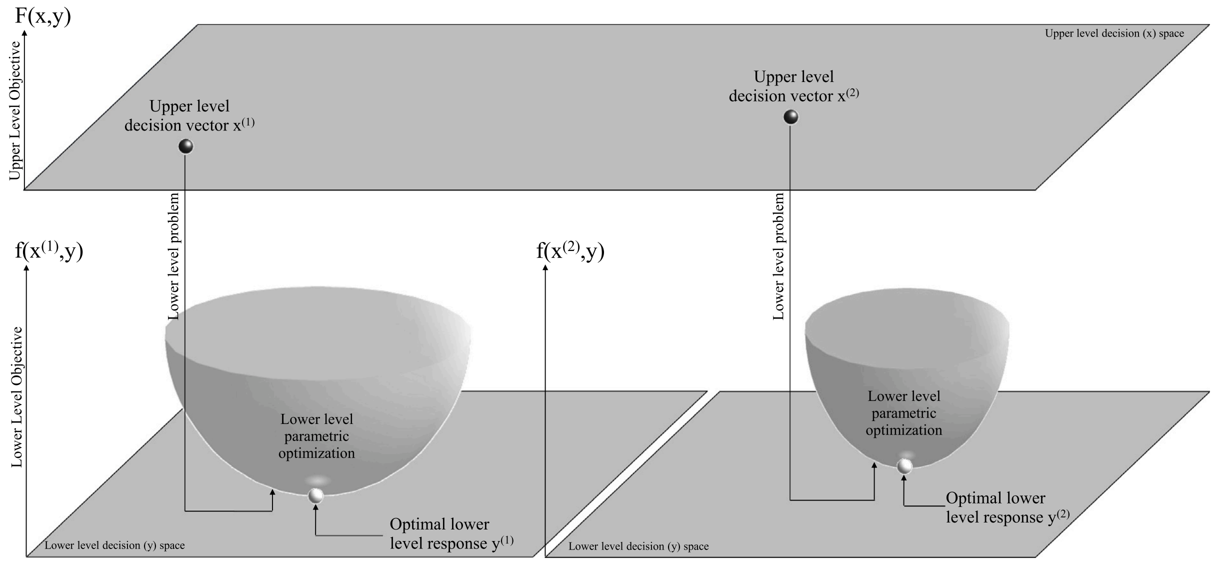


Fig. 1. A sketch of a bilevel problem showing the inter-linkage between the upper and lower levels.

$$y \in \underset{y}{\operatorname{argmin}} \{ f(x, y) : g_i(x, y) \leq 0, i = 1, \ldots, I \} \quad (3)$$

$$G_j(x, y) \leq 0, \quad j = 1, \ldots, J \quad (4)$$

where $g_i : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ denotes the lower level constraints, and $G_j : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ denotes the upper level constraints. Equality constraints may also exist but have been omitted for the sake of brevity. Note that $x \in X$ and $y \in Y$ in the above definition denote that variables can be real or integers, depending on X being $\mathbb{R}^n$ or $\mathbb{Z}^n$ and Y being $\mathbb{R}^m$ or $\mathbb{Z}^m$. However, in most of our discussions, we will avoid using the sets X and Y, and we will consider the variables to be real and continuous unless explicitly stated to be integers.

Fig. 1 illustrates a general bilevel problem involving nested optimization tasks. The figure illustrates that for any given upper level decision vector, there exists a corresponding lower level optimization problem to be addressed. This lower level problem, which is parametric in nature, yields the rational and optimal response of the lower level participant in response to any decision by the upper level participant. In this context, the upper level decision vector is denoted as x, while the lower level decision vector is denoted as y. Pairs $(x^{(1)}, y^{(1)})$ and $(x^{(2)}, y^{(2)})$, where $y^{(1)}$ and $y^{(2)}$ represent optimal responses to $x^{(1)}$ and $x^{(2)}$, respectively, are considered feasible solutions to the bilevel optimization problem, provided that they also adhere to the other constraints of the problem. In bilevel problems, the leader typically has knowledge of the follower's optimization problem , while the follower observes the leader's decisions and subsequently optimizes its own strategies accordingly.

An equivalent formulation of the above problem can be stated in terms of set-valued mappings, where $\Psi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is a set-valued mapping.

$$\Psi(x) = \underset{y}{\operatorname{argmin}} \{ f(x, y) : g(x, y) \leq 0 \}$$

The bilevel optimization problem, in terms of $\Psi(x)$, can be expressed as a constrained optimization problem as follows:

$$\text{“}\min_{x,y}\text{”} \quad F(x, y)$$

subject to

$$y \in \Psi(x)$$

$$G(x, y) \leq 0$$

A summary of the terminologies commonly used in the context of bilevel optimization is given in Table 1. In the preceding two definitions, quotation marks are employed when describing the upper level minimization problem due to an inherent ambiguity that emerges when multiple lower level optimal solutions exist for a given upper level decision vector. When confronted with several lower level optimal solutions, it becomes unclear at the upper level which specific optimal solution from the lower level should be employed. To address this ambiguity, different positions that the leader can adopt can be defined. Two frequently explored positions are the optimistic and pessimistic positions, which we will discuss next.

2.1. Optimistic position

In an optimistic position, when faced with multiple lower level optimal solutions, the leader anticipates that the follower will select the solution within the optimal set $\Psi(x)$ that yields the best objective function value at the upper level. The follower's choice function in this context (best-case for leader) can be defined as follows:

$$\psi_o(x) = \underset{y}{\operatorname{argmin}} \{ F(x, y) : y \in \Psi(x) \}$$

This formulation assumes a degree of cooperation between both participants. The bilevel optimization problem, operating under an optimistic position, is outlined below:

$$\min_{x,y} \quad F(x, y) \quad (5)$$

subject to

$$y = \psi_o(x) \quad (6)$$

$$G(x, y) \leq 0 \quad (7)$$

The optimistic position is more tractable as compared to the pessimistic position; therefore, most of the studies handle the optimistic version of the bilevel optimization problem. An alternative way to write the optimistic position is through the use of the lower level optimal value function φ mapping, defined as:

$$\varphi(x) = \min_{y} \{ f(x, y) : g(x, y) \leq 0 \}$$

Lower level optimal value function mapping gives the optimal lower level function value corresponding to any given upper level decision vector. In terms of φ-mapping, the bilevel problem can be written as follows:

$$\min_{x,y} \quad F(x, y) \quad (8)$$

subject to

Table 1
Important terminology and notations in bilevel optimization.

Name		Description
Lower level feasible region	$\Omega : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$	$\Omega(x) = \{y : g(x,y) \le 0\}$, represents the lower level feasible region for any given upper level decision vector
Relaxed feasible set	$\Phi = \text{gph}\Omega$	$\Phi = \{(x,y) : G(x,y) \le 0, g(x,y) \le 0\}$, represents the region satisfying both upper and lower level constraints
Lower level reaction set	$\Psi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$	$\Psi(x) = \{y : y \in \underset{y}{\text{argmin}}\{f(x,y) : g(x,y) \le 0\}\}$, represents the lower level optimal solution(s) for an upper level decision vector
Lower level feasible region	S(x)	$S(x) = \{y : g(x,y) \le 0\}$ represent the lower level feasible region corresponding to a given upper level decision vector.
Choice function	$\psi : \mathbb{R}^n \to \mathbb{R}^m$	$\psi(x)$ represents the solution chosen by the follower for any upper level decision vector. It becomes important in case of multiple lower level optimal solutions.
Lower level optimal value function	$\varphi : \mathbb{R}^n \to \mathbb{R}$	$\varphi(x) = \min_y\{f(x,y) : g(x,y) \le 0\}$ represents the minimum lower level function value corresponding to a given upper level decision vector.
Upper level pessimistic function	k: $\mathbb{R}^n \to \mathbb{R}$	$k(x) = \max_y\{F(x,y) : g(x,y) \le 0, f(x,y) \ge \varphi(x)\}$ represents the maximum Upper level function value corresponding to lower level decisions for a given upper level decision vector.

$$f(x,y) \le \varphi(x) \tag{9}$$

$$g(x,y) \le 0 \tag{10}$$

$$G(x,y) \le 0 \tag{11}$$

2.2. Pessimistic position

In a pessimistic position, when faced with multiple optimal solutions at the lower level, the leader seeks to optimize for the worst-case scenario. In other words, she assumes that the follower might choose the solution from the optimal set that results in the least favorable objective function value at the upper level. We can express the follower's choice function in this context (worst-case for leader) as follows:

$$\psi_p(x) = \underset{y}{\text{argmax}}\{F(x,y) : y \in \Psi(x)\}$$

This formulation does not assume any form of cooperation. The bilevel optimization problem under a pessimistic position has been defined below:

$$\min_{x,y} \quad F(x,y) \tag{12}$$

subject to

$$y = \psi_p(x) \tag{13}$$

$$G(x,y) \le 0 \tag{14}$$

The pessimistic position [57,68,69] is relatively less tractable when compared to the optimistic position. The pessimistic formulation is guaranteed to have optimal solutions under stronger assumptions compared to the optimistic formulation. In the case of the optimistic position, we provide a formulation based on φ mapping, which is a popular formulation and is commonly used in algorithms to solve the optimistic formulation. We argue that it is possible to write an alternative formulation of the pessimistic position in terms of the φ mapping. However, it requires an additional constraint to ensure that the upper level chooses the worst solution from the lower level's optimal set for any given x.

$$\min_{x,y} \quad F(x,y) \tag{15}$$

subject to

$$f(x,y) \le \varphi(x) \tag{16}$$

$$F(x,y) \ge k(x) \tag{17}$$

$$g(x,y) \le 0 \tag{18}$$

$$G(x,y) \le 0 \tag{19}$$

where $k(x) = \max_y\{F(x,y) : g(x,y) \le 0, f(x,y) \le \varphi(x)\}$.

3. δ-perturbation of bilevel problems

Consider the bilevel problem (1)–(4) for which one may take an optimistic or a pessimistic position while solving it. It is proposed that the δ-perturbed bilevel formulation for the optimistic position is given as follows:

$$\min_{x,y} \quad F(x,y)$$

subject to

$$y \in \underset{y}{\text{argmin}}\{f(x,y) + \delta F(x,y) : g_i(x,y) \le 0, i = 1,\dots,I\}$$

$$G_j(x,y) \le 0, \quad j = 1,\dots,J$$

Similarly, the δ-perturbed bilevel formulation for the pessimistic position is given as follows:

$$\min_{x,y} \quad F(x,y)$$

subject to

$$y \in \underset{y}{\text{argmin}}\{f(x,y) - \delta F(x,y) : g_i(x,y) \le 0, i = 1,\dots,I\}$$

$$G_j(x,y) \le 0, \quad j = 1,\dots,J$$

In both the above formulations, δ is a small constant that leads to a minor perturbation in the optimistic and the pessimistic bilevel problems. We assume that a minor perturbation with some δ modifies the lower level problem such that it has a unique global optimal solution for any given upper level decision.

We explain the perturbations visually through Fig. 2. This figure shows a bilevel optimization problem with multiple lower level optimal solutions on the left hand side. The feasible region of the lower level is shown in gray for four decision vectors of the upper level, x_0, x_1, x_2 and x_3. Note that there are multiple lower level optimal solutions for all of these four upper level decision vectors. Among the multiple lower level optimal solutions, the optimistic and pessimistic selections for x_0, x_1, x_2 and x_3 are marked in white and black circles, respectively. A dotted line marks the paths traced by the optimistic and pessimistic lower level selections for all upper level decision vectors, making it clear that x_0 represents the optimistic bilevel optimum (as x_0 corresponds to minimum of $F(x,y)$ for optimistic selection), while x_3 represents the pessimistic bilevel optimum (as x_3 corresponds to minimum of $F(x,y)$ for pessimistic selection).

The perturbation of the lower level problem with δ slightly changes the optimistic and pessimistic selections that have been shown on the right-hand side of Fig. 2. Solving this perturbed problem, therefore, leads to an approximate optimistic and pessimistic bilevel optimum. In this paper, we show the relationship between the choice of δ and

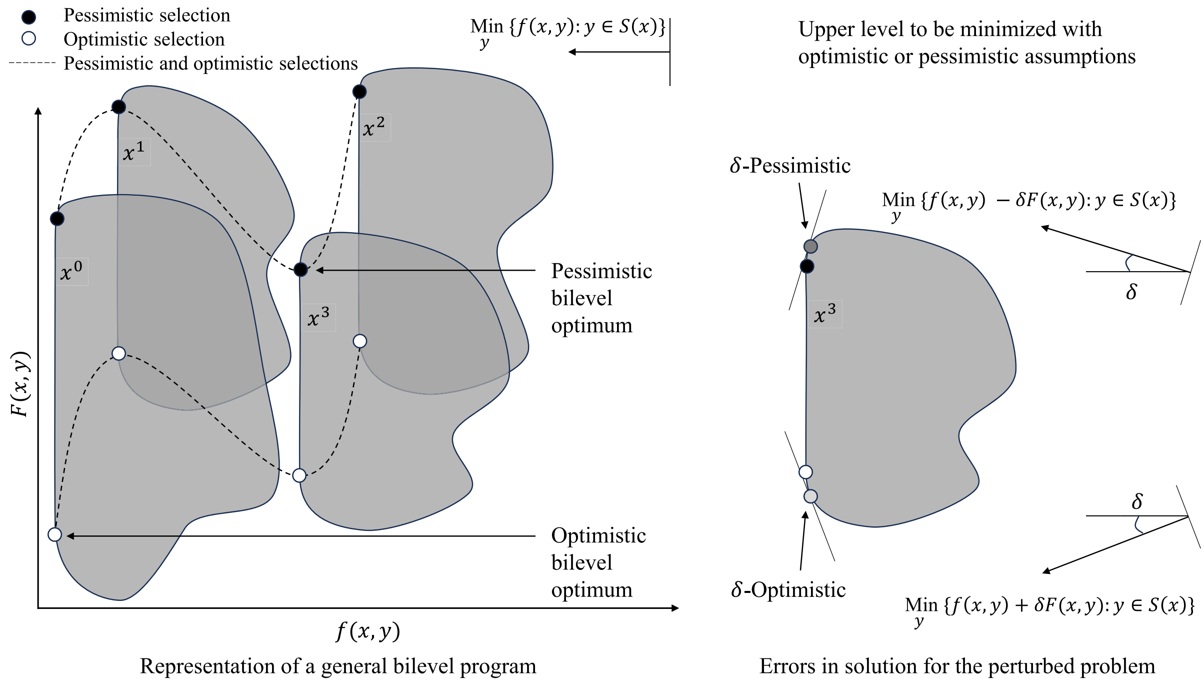


Fig. 2. Each shaded region in the left figure shows the feasible region of the lower level optimization problem in the F-f space for a given x_0, x_1, x_2 and x_3. The left vertical part of each of the shaded regions represent multiple lower level optimal solutions. The right figure gives a graphical representation for the errors induced by δ-perturbed lower level objective function corresponding to x_3.

the maximum error that it leads to when the δ-perturbed formulation is solved instead of the true bilevel (optimistic or pessimistic) formulation.

In the following subsections, we explore the δ-perturbed formulation for the pessimistic position first followed by the δ-perturbed formulation for the optimistic position. The reason for this is that the results obtained for the pessimistic case, can be easily extended to arrive at the results for the optimistic case.

3.1. δ-perturbation of the pessimistic bilevel problem

We consider the δ-perturbation of the pessimistic bilevel problem in (15)–(19). We make the assumption that all functions in the bilevel problem are continuous and relaxed feasible set $\Phi = \{(x, y) : G(x, y) \leq 0, g(x, y) \leq 0\}$ is closed and bounded. We also assume that there exists a feasible and bounded lower level optimal solution for any given upper level decision vector. Let M be the maximum value of $F(x, y)$ and m be the minimum value of $F(x, y)$ in the set Φ, i.e. $M = \max\{F(x, y) : (x, y) \in \Phi\}$ and $m = \min\{F(x, y) : (x, y) \in \Phi\}$. Note that (15)–(19) contains

$$k(x) = \max_y\{F(x, y) : G(x, y) \leq 0, f(x, y) \leq \varphi(x)\} \text{ and}$$
$$\varphi(x) = \min_y\{f(x, y) : g(x, y) \leq 0\}$$

that are unknown *a priori*. From the definition of $k(x)$ and $\varphi(x)$, for any feasible pessimistic bilevel solution (x, y) the following holds:

$$f(x, y) \leq \varphi(x) \tag{20}$$
$$F(x, y) \geq k(x) \tag{21}$$

Introducing $\delta > 0$, from (20) and (21) we can write the following:

$$f(x, y) - \delta F(x, y) \leq \varphi(x) - \delta k(x) \tag{22}$$

Now let us consider δ-perturbation of the lower level problem in (1)–(4):

$$\min_y \quad f(x, y) - \delta F(x, y)$$

subject to

$$g(x, y) \leq 0$$

and we denote the optimal solution of the above perturbed problem as:

$$y^p \in \underset{y}{\operatorname{argmin}}\{f(x, y) - \delta F(x, y) : g(x, y) \leq 0\}$$

Since $f(x, y^p) - \delta F(x, y^p)$ is the optimal function value of the above problem for any given x, we can write the following for any y that is lower level feasible for the given x:

$$f(x, y^p) - \delta F(x, y^p) \leq f(x, y) - \delta F(x, y) \tag{23}$$

From (22) and (23), the following holds:

$$f(x, y^p) - \delta F(x, y^p) \leq \varphi(x) - \delta k(x) \tag{24}$$

From the definition of $\varphi(x)$, it is obvious that:

$$\varphi(x) \leq f(x, y^p) \tag{25}$$

Therefore, from (24) and (25), we can write:

$$k(x) \leq F(x, y^p) \tag{26}$$

From (24) and (25), we can also write the following:

$$\varphi(x) \leq f(x, y^p) \leq \varphi(x) + \delta[F(x, y^p) - k(x)] \tag{27}$$
$$\varphi(x) \leq f(x, y^p) \leq \varphi(x) + \delta[M - m] \tag{28}$$

Note that as $\delta \to 0$, $f(x, y^p) \to \phi(x)$. Therefore, the error in $\varphi(x)$ is bounded by $\delta[M - m]$ for all x.

Let us write the following problems for the ease of discussion:

$$F(x, y^p) = \max_y \quad F(x, y)$$
subject to
$$g(x, y) \leq 0$$
$$f(x, y) \leq f(x, y^p)$$
$$k(x) = \max_y \quad F(x, y)$$
subject to
$$g(x, y) \leq 0$$
$$f(x, y) \leq \varphi(x, y)$$
$$l(x) = \max_y \quad F(x, y)$$
subject to
$$g(x, y) \leq 0$$
$$f(x, y) \leq \varphi(x) + \delta(M - m)$$

From (28), we can write $k(x) \leq F(x, y^p) \leq l(x)$.

From the assumptions made at the beginning of the section, $k(x)$ and $l(x)$ will always be feasible and bounded. Note that $l(x)$ can be obtained by perturbing the constraint $f(x, y) \leq \varphi(x, y)$ in $k(x)$ by $\delta(M - m)$. If λ is the dual for constraint $f(x, y) \leq \varphi(x, y)$ and the right-hand side of the constraint is perturbed by $\delta(M - m)$, then for $\delta \to 0$, we can write $l(x) - k(x) = \delta(M - m)\lambda$ with first-order approximation.

Since $F(x, y)$ is assumed to be continuous and bounded in Φ, $k(x)$ and $l(x)$ are feasible and bounded, it implies that λ will always be finite. Therefore, we can state the following:

$$l(x) = k(x) + \delta(M - m)\lambda$$
$$\Rightarrow k(x) \leq F(x, y^p) \leq k(x) + \delta(M - m)\lambda$$
$$\Rightarrow \text{as } \delta \to 0 \text{ , } F(x, y^p) \to k(x)$$

We have shown that as $\delta \to 0$,

$$\begin{aligned} f(x, y^p) &\to \phi(x) \quad \text{with error } \delta[M - m] \\ F(x, y^p) &\to k(x) \quad \text{with error } \delta\lambda[M - m] \end{aligned} \tag{29}$$

This implies that the errors in the lower and upper objective level function values on solving a δ-perturbed pessimistic bilevel problem are bounded by $\delta[M - m]$ and $\delta\lambda[M - m]$, respectively.

3.2. δ-perturbation of the optimistic bilevel problem

The δ-perturbation of the optimistic bilevel problem may also be necessary at times to ensure that whenever the lower level optimization problem is solved, it returns an (approximate) optimistic selection for the algorithm to work efficiently. The error bounds for the δ-perturbed optimistic bilevel problem can be obtained in a similar manner as in the pessimistic case. Note that the formulation based on the lower level optimal value function (8)–(11) for the optimistic bilevel problem does not involve $k(x)$. However, for the sake of finding the error bounds in a similar manner as that for the δ-perturbed pessimistic formulation, we will introduce a slightly modified $k(x)$ and write the lower level optimal value function-based formulation as follows:

$$\min_{x,y} \quad F(x, y) \tag{30}$$
subject to
$$f(x, y) \leq \varphi(x) \tag{31}$$
$$F(x, y) \leq k(x) \tag{32}$$
$$g(x, y) \leq 0 \tag{33}$$
$$G(x, y) \leq 0 \tag{34}$$

where $\varphi(x) = \min_y\{f(x, y) \; : \; g(x, y) \leq 0\}$ and $k(x) = \min_y\{F(x, y) \; : \; g(x, y) \leq 0, f(x, y) \leq \varphi(x)\}$. Obviously, (32) is a redundant constraint, and once again, from the definition of $k(x)$ and $\varphi(x)$, the following holds:

$$f(x, y) \leq \varphi(x) \tag{35}$$
$$F(x, y) \leq k(x) \tag{36}$$

Introducing $\delta > 0$, from (35) and (36) we can write the following:

$$f(x, y) + \delta F(x, y) \leq \varphi(x) + \delta k(x) \tag{37}$$

The δ-perturbation of the lower level problem in (1)–(4) in the optimistic case is given as follows:

$$\min_y \quad f(x, y) + \delta F(x, y)$$
subject to
$$g(x, y) \leq 0$$

with the optimal solution of the above perturbed problem denoted as:

$$y^o \in \underset{y}{\operatorname{argmin}}\{f(x, y) + \delta F(x, y) \; : \; g(x, y) \leq 0\}$$

Since $f(x, y^o) + \delta F(x, y^o)$ is the optimal function value of the above problem for any given x, we can write the following for any y that is lower level feasible for the given x:

$$f(x, y^o) + \delta F(x, y^o) \leq f(x, y) + \delta F(x, y) \tag{38}$$

Building up along the same lines as in the pessimistic case, one can write the following. The definitions, $M = \max\{F(x, y) : (x, y) \in \Phi\}$ and $m = \min\{F(x, y) : (x, y) \in \Phi\}$ are consistent with those defined earlier.

$$f(x, y^o) + \delta F(x, y^o) \leq \varphi(x) + \delta k(x) \tag{39}$$
$$\varphi(x) \leq f(x, y^o) \tag{40}$$
$$F(x, y^o) \leq k(x) \tag{41}$$

From (39) and (40),:

$$\varphi(x) \leq f(x, y^o) \leq \varphi(x) + \delta[k(x) - F(x, y^o)] \tag{42}$$
$$\varphi(x) \leq f(x, y^o) \leq \varphi(x) + \delta[M - m] \tag{43}$$

Note that as $\delta \to 0$, $f(x, y^o) \to \phi(x)$. Therefore, the error in $\varphi(x)$ is bounded by $\delta[M - m]$ for all x.

Now let us write the following minimization problems, noting that in the pessimistic case, similar problems were defined with maximization of the objective function:

$$F(x, y^o) = \min_y \quad F(x, y)$$
subject to
$$g(x, y) \leq 0$$
$$f(x, y) \leq f(x, y^o)$$
$$k(x) = \min_y \quad F(x, y)$$
subject to
$$g(x, y) \leq 0$$
$$f(x, y) \leq \varphi(x, y)$$
$$l(x) = \min_y \quad F(x, y)$$
subject to
$$g(x, y) \leq 0$$
$$f(x, y) \leq \varphi(x) + \delta(M - m)$$

From (43), we can write $l(x) \leq F(x, y^o) \leq k(x)$. Next, following similar arguments as in the pessimistic case, the error bounds for the lower and upper objective functions again turn out to be $\delta[M - m]$ and $\delta\lambda[M - m]$, respectively, with λ being the dual variable with a finite value for the constraint $f(x, y) \leq \varphi(x, y)$.

Table 2
Optimistic and pessimistic solutions for the example problem [70].

Solutions	x	y	F
Optimistic	0.2106	1.799	−1.755
Pessimistic	0	0.2929	−0.2929

Table 3
Parameters used in DE.

	Upper level	Lower level
Crossover rate	0.9	0.9
Mutation rate	0.8	0.8
Population size	50	50
Number of generations	30	50

4. Illustrative example

In this section, we provide an illustrative example to explain the ideas discussed in the previous sections. The example problem, which is taken from [70], features multiple global minima at the lower level. The problem formulation is as follows:

$$\begin{aligned} &\min_{x,y} \quad x^2 - y \\ &\text{s.t.} \\ &\quad y \in \underset{y}{\operatorname{argmin}}\{((y-1-x/10)^2 - x/2 - 1/2)^2 \,:\, 0 \le y \le 3\} \\ &\quad 0 \le x \le 1 \end{aligned} \tag{44}$$

For a fixed upper level decision x, the set of optimal lower level decisions is given by:

$$\Psi(x) = \{1 + x/10 \pm \sqrt{2+2x}/2\} \tag{45}$$

In Table 2 the known global optima for the optimistic and pessimistic variant of the problem are reported. It is clear that the optimistic and pessimistic decisions for the leader and the follower are different.

In Fig. 3(a), we show the lower level objective function for $x = 0.2106$, which has two lower level optimal points shown with filled circles. A leader solving this problem from an optimistic perspective will prefer the right-filled circle based on the upper level objective function. On using the δ-perturbation, the lower level objective function gets slightly modified such that an approximate optimistic solution is selected automatically by solving the lower level problem. In Fig. 3(b), the same is shown from a pessimistic perspective, where the δ-perturbation allows the selection of an approximate pessimistic solution automatically. In Fig. 4, we show the actual choice functions $\psi_o(x)$ and $\psi_p(x)$ for the optimistic and pessimistic cases, respectively. When the optimistic and the pessimistic problems are δ-perturbed, the actual functions change to $\psi_o^\delta(x)$ and $\psi_p^\delta(x)$. The difference between the actual choice function and the approximate choice function represents the error induced because of the δ-perturbed formulation.

5. Numerical experiments

In this section, we provide an experimental study on various bilevel test problems from the literature to computationally demonstrate the workings of the δ-perturbation approach when a nested algorithm is used to solve the test problems. The chosen nested approach is differential evolution [71], which we describe in the next subsection, followed by the experimental setup and the results.

5.1. Experimental setup

We employ the DE algorithm with integrated constraint handling to solve the δ-perturbed formulations. DE is a popular population-based optimization algorithm that is effective at solving optimization problems. Operating through mutation, crossover, and selection, DE iteratively evolves a population of candidate solutions. Mutation introduces random changes to individuals, promoting exploration of the solution space, while crossover combines information from multiple individuals to generate new candidate solutions, facilitating the exploitation of promising regions. The selection mechanism decides which individuals survive and reproduce in the next generation. To ensure adherence to problem constraints, we implemented a constraint handling mechanism based on the following criteria [72]:

1. Any feasible solution is preferred to any infeasible solution.
2. Among the two feasible solutions, the one with a better objective function value is preferred.
3. Among two infeasible solutions, the one with a smaller overall constraint violation is preferred.

A nested approach requires a lower level optimization to be performed for any given upper level vector. We sample the upper level vectors using upper level DE, and utilize a lower level DE to handle the lower level optimization problem for any given upper level vector. We set the algorithm parameters to reasonable values for our numerical experiments following the algorithmic recommendations from the previous literature. The values of the parameters are stated in Table 3. For every test problem, we use the nested DE to solve the δ-perturbed formulations of optimistic and pessimistic variants separately. The only difference in the two variants is the choice of δ, which is expected to be a small positive number for the optimistic position and a small negative number for the pessimistic position.

We solve six test problems from an optimistic and pessimistic position. The definitions of the test problems are provided in Table 5. Five test problems have been taken from the *mb* suite [70], and an additional test problem that is a principal agent problem has been taken from [73].

For all test problems, we performed 20 independent runs on an Intel(R) Core(TM) i5-8365U CPU @ 1.60 GHz with 16 GB of RAM on the Windows 10 operating system. The algorithm is implemented in Matlab R2023b without any parallelization.

To evaluate the performance of our method, we use the absolute error (AE) in the upper level objective function and the lower level objective function as a metric. If the optimal upper and lower level objective function values for the optimistic bilevel problem are F_o^* and f_o^*, and the bilevel optimal solution obtained by solving the δ-perturbed formulation are F_o^δ and f_o^δ, then the absolute errors at the upper and lower levels are given as follows:

$$\mathrm{AE}_o^F = |F_o^\delta - F_o^*|$$
$$\mathrm{AE}_o^f = |f_o^\delta - f_o^*|$$

Similarly, for the pessimistic case, the absolute errors are as follows:

$$\mathrm{AE}_p^F = |F_p^\delta - F_p^*|$$
$$\mathrm{AE}_p^f = |f_p^\delta - f_p^*|$$

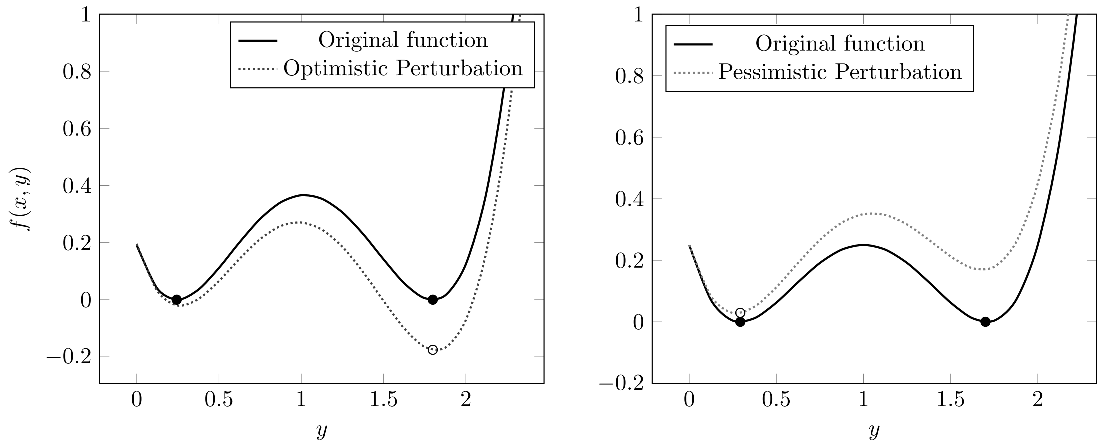
5.2. Results for bilevel test problems

First, we report the results of the bilevel problems taken from [70].

In Figs. 5, 7, 9, 11, and 13, we present the boxplots of the upper level and lower level AE obtained by the algorithm for different δ while solving the δ-perturbed formulations of the optimistic and pessimistic variants for each test problem. It is obvious that in almost all cases the higher the δ, the AE is expected to be higher. However, this may not always generalize computationally as can be seen in Fig. 7(b), where for $\delta = 0.5$, the AE is lower compared to $\delta = 0.05, 0.1$.

Furthermore, in Fig. 9(b), the cases of $\delta = 0.001$ and 0.1 once again seems to be an anomaly, but it should be noted that the AE for both these cases have a magnitude of 1e-6, which is a sufficiently low.

In Figs. 6, 8, 10, 12 and 14 we have plotted the lower level accuracy achieved for each δ for all runs, along with the theoretical bounds proposed in the previous section. The AE of the lower level function



(a) Lower level objective function for $x = 0.2106$

(b) Lower level objective function for $x = 0.0$

Fig. 3. δ-perturbation plots for the example problem showing true lower level optimal solutions and the δ-perturbed lower level solutions for optimistic and pessimistic cases for $\delta = 0.1$.

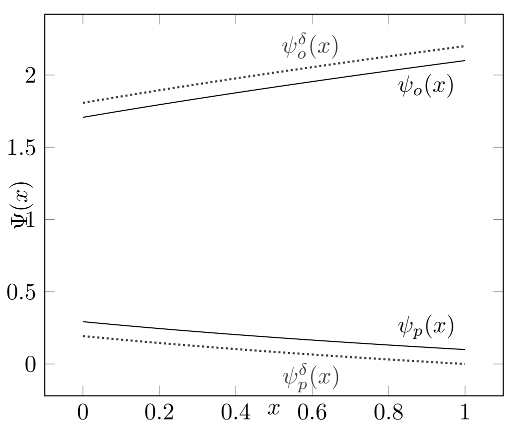


Fig. 4. $\Psi(x)$-x plot for the original and δ-perturbed example problem.

for different δ is shown to be always below the theoretical error bound (i.e. within the gray area). Further details of the results for these problems can be found in Appendix A.

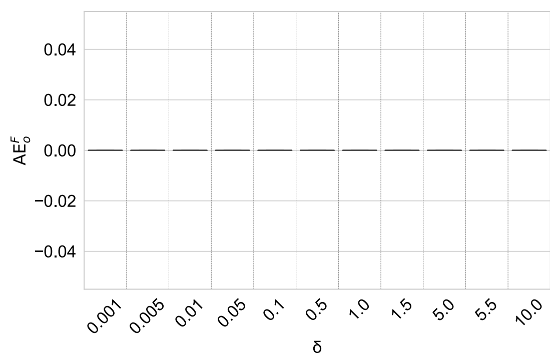
In our experiments, we have selected both small and large values of δ to demonstrate the impact of δ on the solution. However, for practical problems, one may simply choose a small value of δ which is a few orders of magnitude smaller than either the desired accuracy or the magnitude of the values that the upper and lower-level objectives take in their domain.

5.3. Results for a principal–agent problem

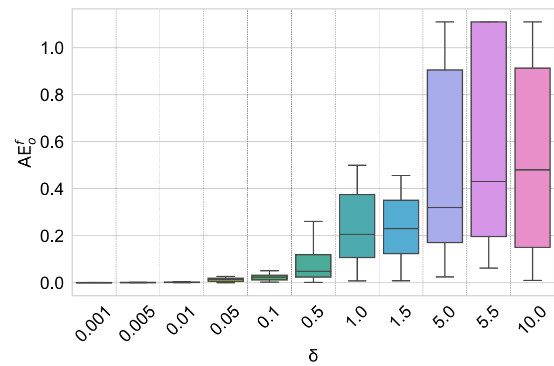
In the domain of game theory, the principal–agent (PA) problem surfaces when a principal delegates a task to an agent in return for compensation. In this scenario, the principal devises a contract, where the decision of the principal is x, while the agent selects an effort or action, denoted as y, to execute the task. The agent's objective is to maximize their utility function, denoted as $f(x, y)$, whereas the principal aims to maximize its utility function, denoted as $F(x, y)$. The definition of the problem considered in the paper can be found in Table 5. The problem is taken from [73]. In Table 4, the statistical results from 20 runs of the algorithm for various values of δ are reported for the optimistic and pessimistic variants. More specifically, we report the minimum, maximum, median, mean, and standard deviation of the accuracy of the objective functions. The results are also shown in Fig. 15. It is noted that the results are robust, as the algorithm converges to similar solutions in all 20 runs for various values of δ. For small values of δ we observe the solution obtained from the δ-perturbed formulation is close to the true solution for both optimistic and pessimistic case. However, the algorithm converges to an inaccurate upper level solution when $\delta \geq 0.5$ for optimistic case, and $\delta \geq 1.5$ for pessimistic case. In Fig. 16 we have once again plotted the (median) lower level accuracy achieved for each δ, along with the theoretical limits that we have reported in the previous section. The behavior is found to be the same, such that the median error is well below the theoretical line for all the cases.

6. Conclusion and future work

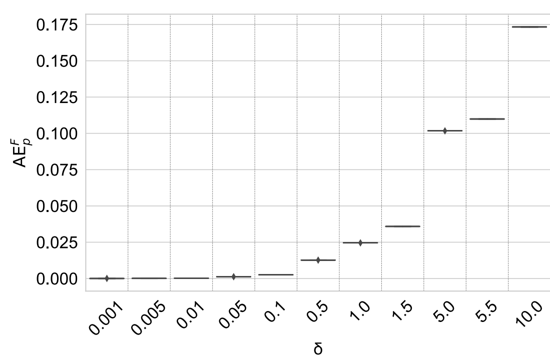
In this paper, we conducted an analysis of the δ-perturbed formulation for ill-posed bilevel optimization problems. This formulation proves to be useful in efficiently managing bilevel scenarios where multiple optimal solutions exist at the lower level. By using a tailored perturbation strategy for the optimistic or pessimistic formulation, it can be ensured that there is only a single optimal solution in the lower level optimization problem for any given upper level decision. This facilitates the efficient search for optimistic or pessimistic bilevel optimal solutions through algorithms that iteratively solve the bilevel problem by exploiting the nested structure. Importantly, we have provided a proof demonstrating the approximate equivalence of the δ-perturbed formulation with the original optimistic and pessimistic formulations, accompanied by an error bound analysis. Furthermore, we applied this scheme to a number of test problems and solved them with DE. Overall, our results highlight the effectiveness of the δ-perturbed formulation in addressing bilevel optimization problems with multiple lower level optimal solutions, and the computational study performed verifies the theoretical bounds proposed in the study. As a future work, one can create δ-formulations for the multiobjective bilevel optimization problems as well and perform a similar error bound analysis for the optimistic and pessimistic frontiers.



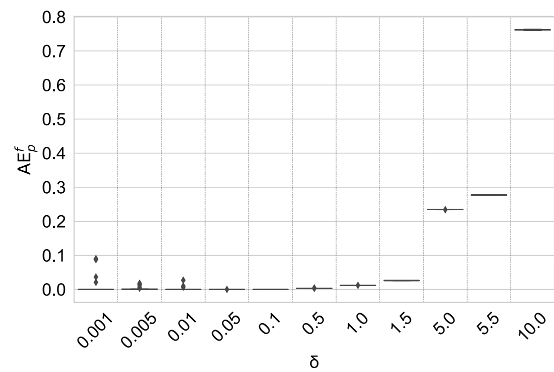
(a) Absolute error (optimistic) for upper level objective.



(b) Absolute error (optimistic) for lower level objective.

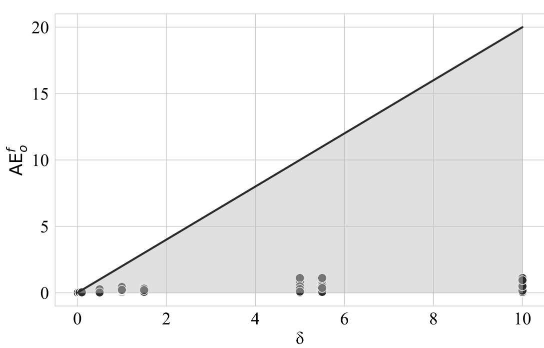


(c) Absolute error (pessimistic) for upper level objective.

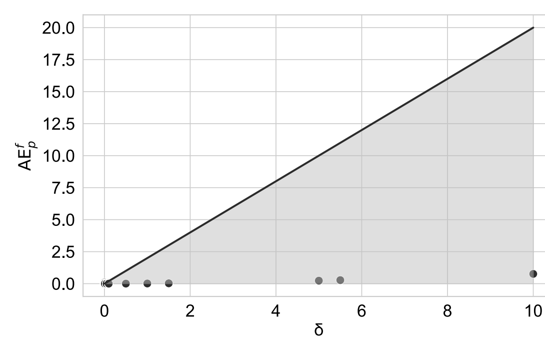


(d) Absolute error (pessimistic) for lower level objective.

Fig. 5. Boxplots of absolute error (AE) for various values of δ for Problem 1 from 20 runs.



(a) Optimistic case



(b) Pessimistic case

Fig. 6. Theoretical error bound (shaded) and absolute error for lower level objective from 20 runs of Problem 1 for various values of δ.

CRediT authorship contribution statement

Margarita Antoniou: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Ankur Sinha:** Writing – review & editing, Writing – original draft, Validation, Methodology, Formal analysis, Conceptualization. **Gregor Papa:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

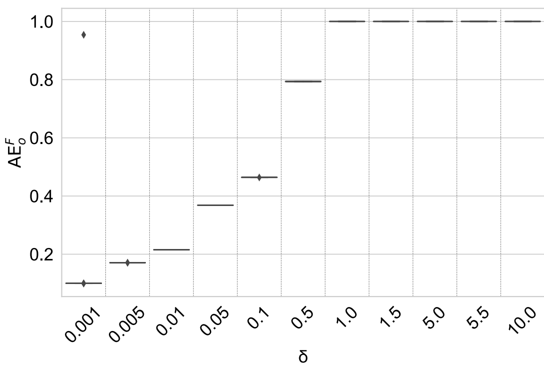
Data will be made available on request.

Acknowledgment

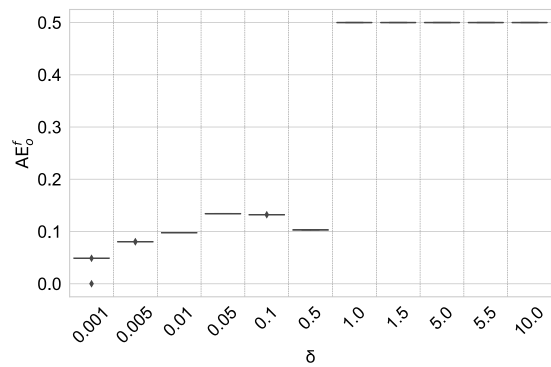
This work was supported by the Slovenian Research Agency (research core funding No. P2-0098).

Appendix A

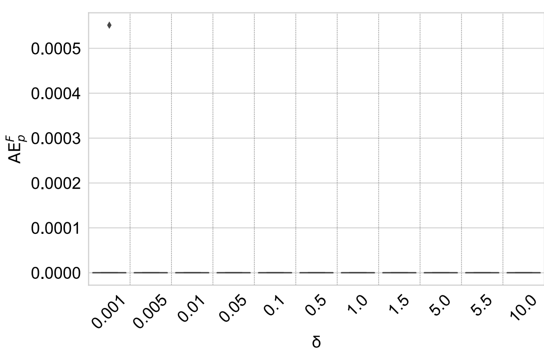
In this section, we report the statistical results for the bilevel test problems (see Tables 6–10).



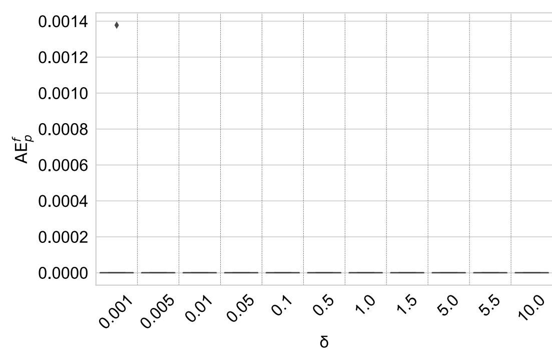
(a) Absolute error (optimistic) for upper level objective.



(b) Absolute error (optimistic) for lower level objective.

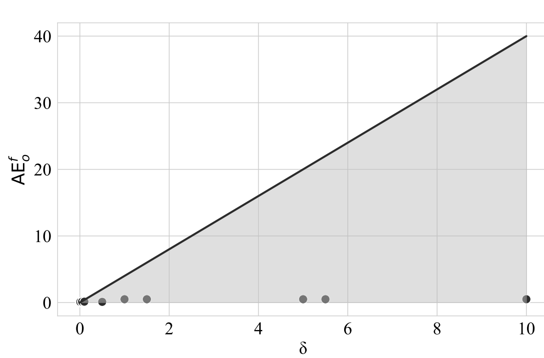


(c) Absolute error (pessimistic) for upper level objective.

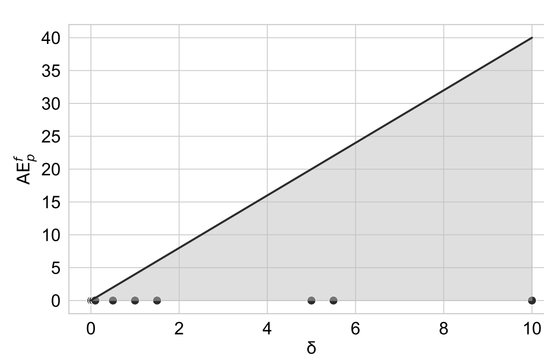


(d) Absolute error (pessimistic) for lower level objective.

Fig. 7. Boxplots of absolute error (AE) for various values of δ for Problem 2 from 20 runs.

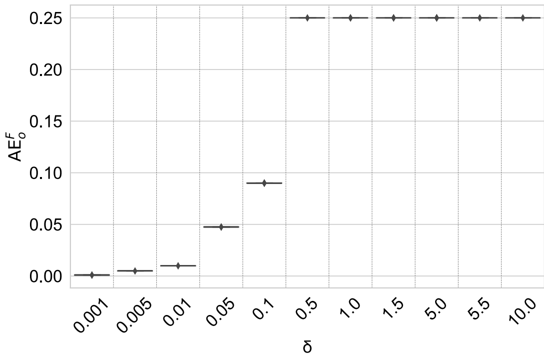


(a) Optimistic case

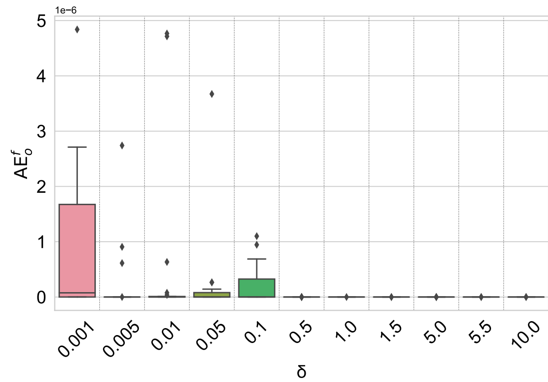


(b) Pessimistic case

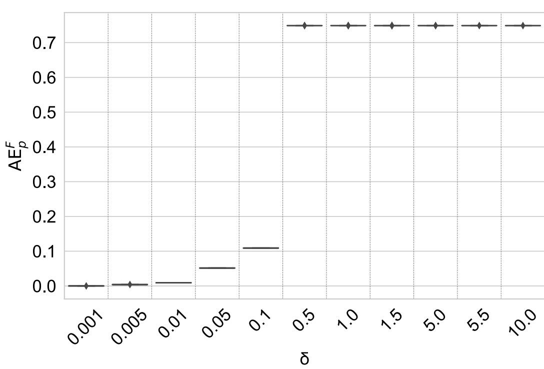
Fig. 8. Theoretical error bound (shaded) and absolute error for lower level objective from 20 runs of Problem 2 for various values of δ.



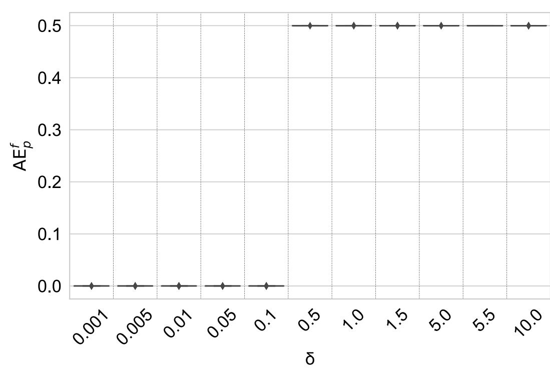
(a) Absolute error (optimistic) for upper level objective.



(b) Absolute error (optimistic) for lower level objective.

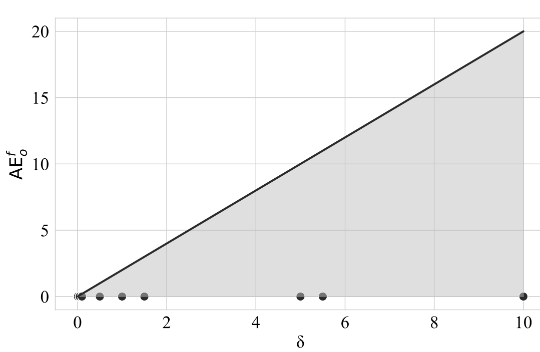


(c) Absolute error (pessimistic) for upper level objective.

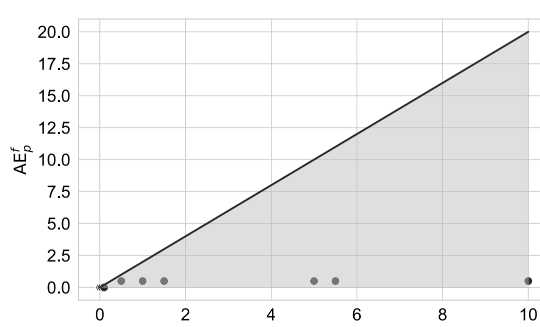


(d) Absolute error (pessimistic) for lower level objective.

Fig. 9. Boxplots of absolute error (AE) for various values of δ for Problem 3 from 20 runs.

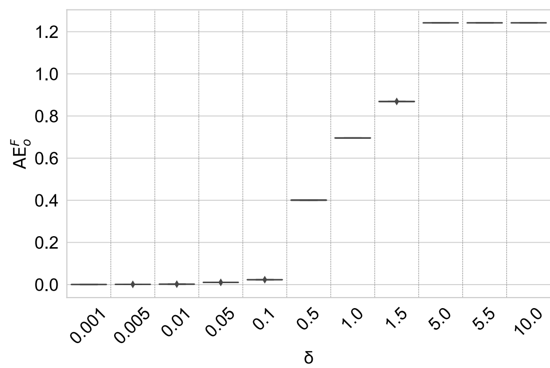


(a) Optimistic case

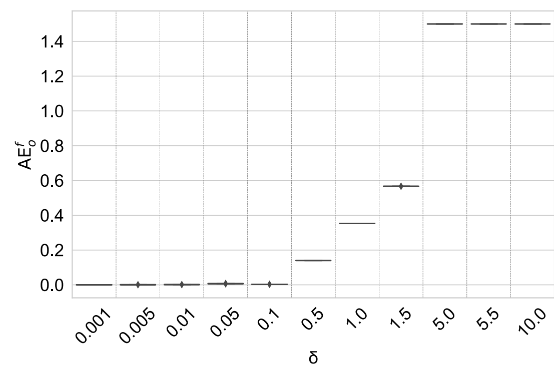


(b) Pessimistic case

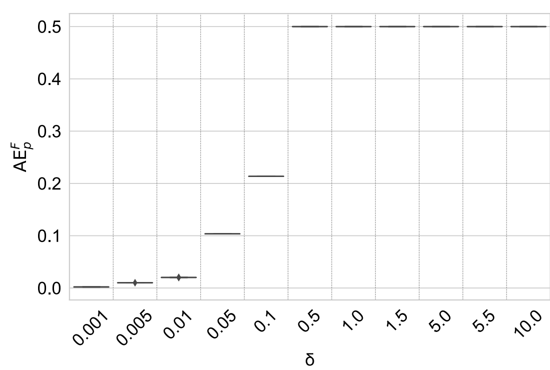
Fig. 10. Theoretical error bound (shaded) and absolute error for lower level objective from 20 runs of Problem 3 for various values of δ.



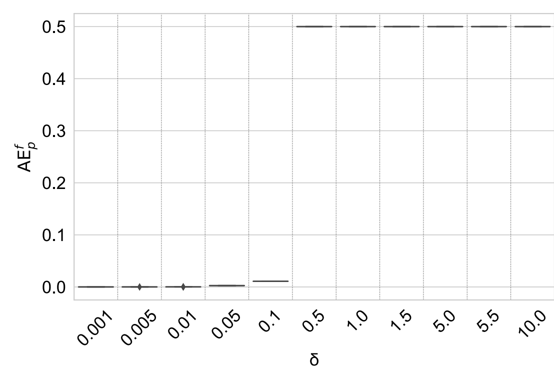
(a) Absolute error (optimistic) for upper level objective.



(b) Absolute error (optimistic) for lower level objective.

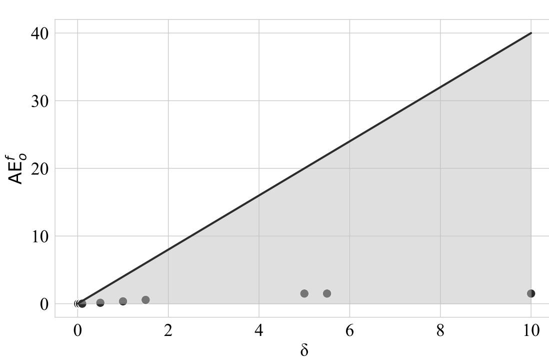


(c) Absolute error (pessimistic) for upper level objective.

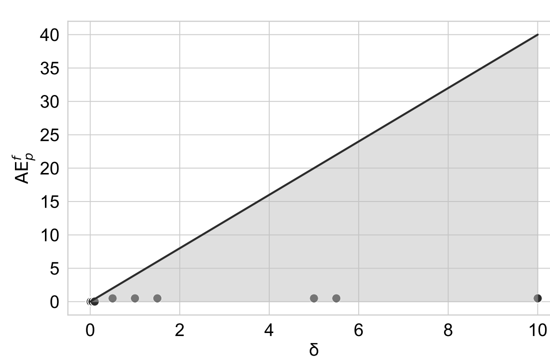


(d) Absolute error (pessimistic) for lower level objective.

Fig. 11. Boxplots of absolute error (AE) for various values of δ for Problem 4 from 20 runs.

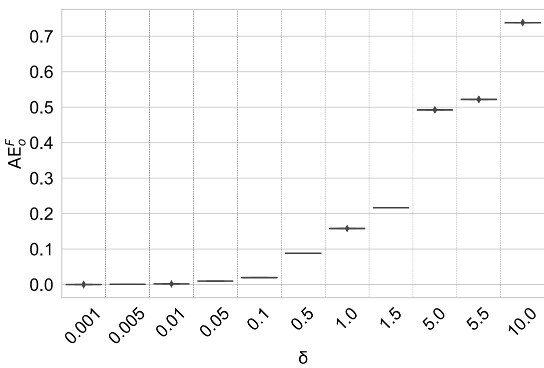


(a) Optimistic case

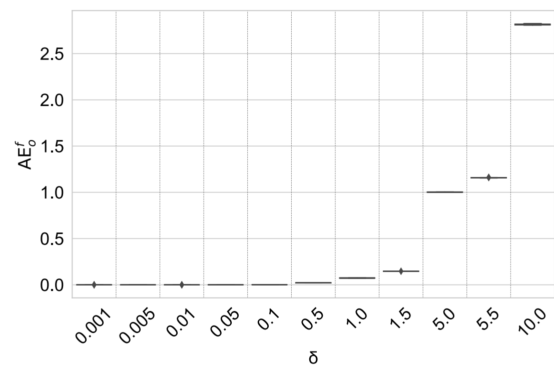


(b) Pessimistic case

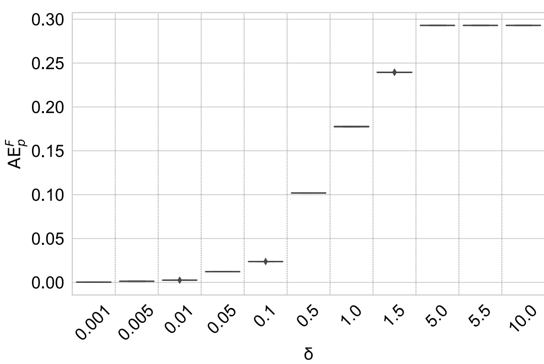
Fig. 12. Theoretical error bound (shaded) and absolute error for lower level objective from 20 runs of Problem 4 for various values of δ.



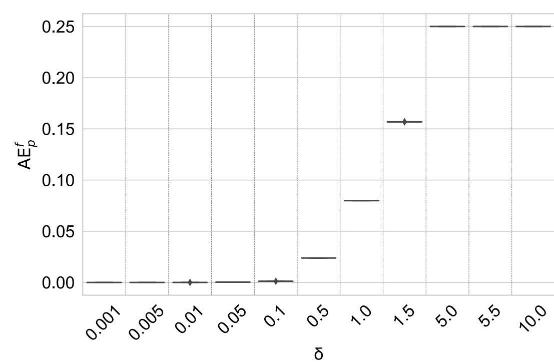
(a) Absolute error (optimistic) for upper level objective.



(b) Absolute error (optimistic) for lower level objective.

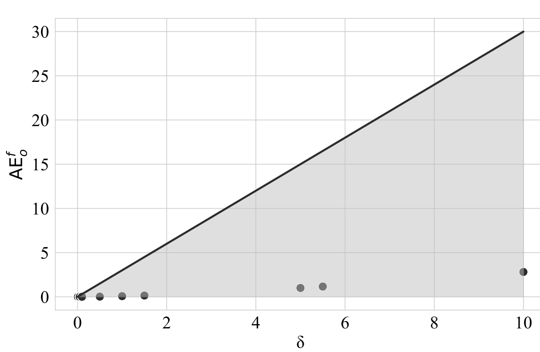


(c) Absolute error (pessimistic) for upper level objective.

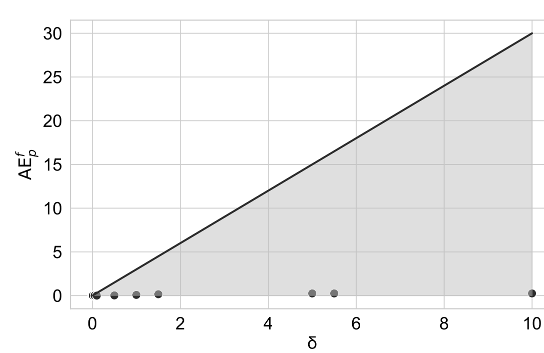


(d) Absolute error (pessimistic) for lower level objective.

Fig. 13. Boxplots of absolute error (AE) for various values of δ for Problem 5 from 20 runs.



(a) Optimistic case



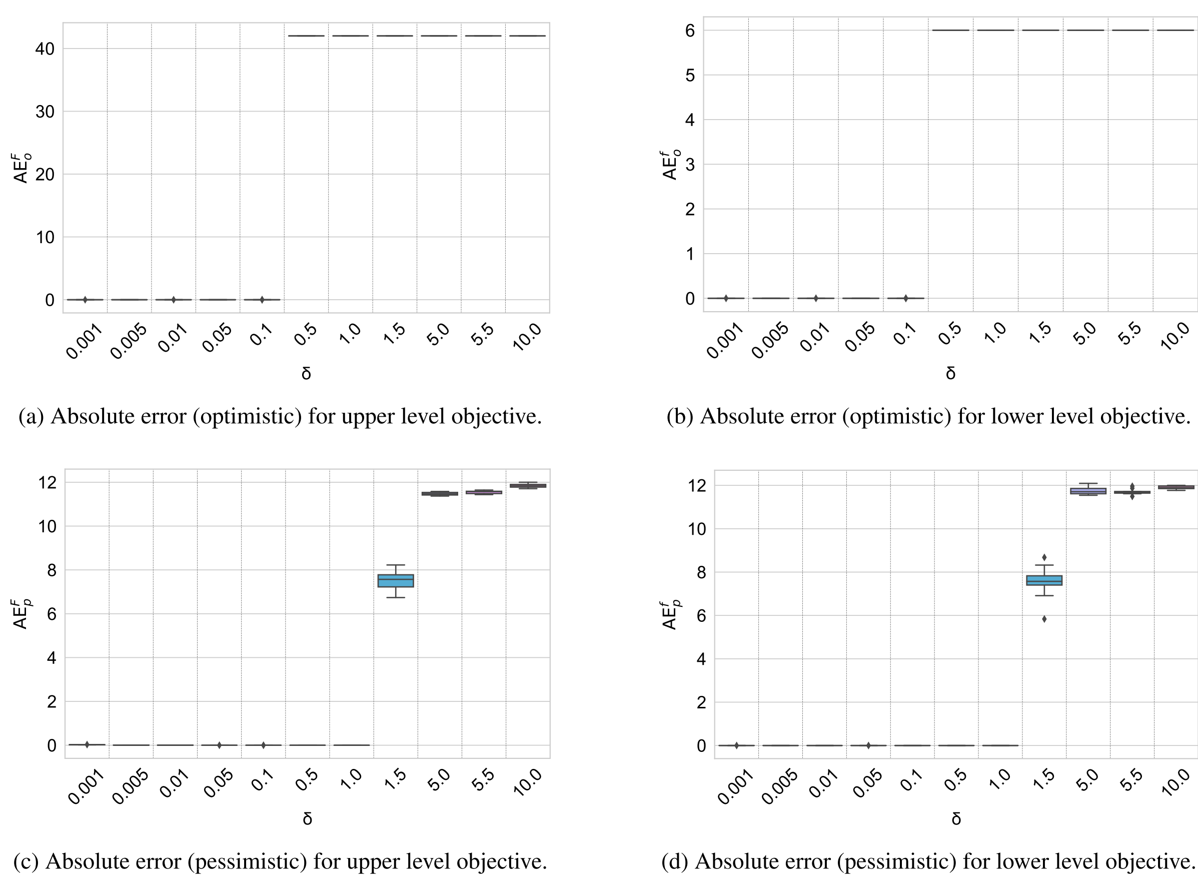
(b) Pessimistic case

Fig. 14. Theoretical error bound (shaded) and absolute error for lower level objective from 20 runs of Problem 5 for various values of δ.

Table 4
Statistical results of Absolute Errors (AE) for PA.

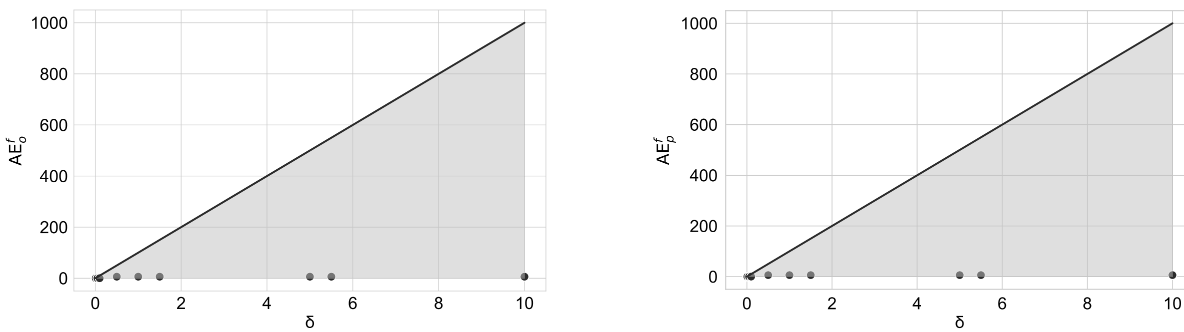
δ	AE_o^F					AE_o^f				
	Min	Max	Median	Mean	Std	Min	Max	Median	Mean	Std
0.001	1.73e−09	2.84e−07	6.70e−08	8.74e−08	7.57e−08	1.00e−03	1.00e−03	1.00e−03	1.00e−03	0.00e+00
0.005	3.74e−09	1.43e−07	2.62e−08	4.71e−08	4.75e−08	1.00e−03	1.00e−03	1.00e−03	1.00e−03	0.00e+00
0.01	2.50e−09	1.42e−07	2.58e−08	3.17e−08	3.06e−08	1.00e−03	1.00e−03	1.00e−03	1.00e−03	0.00e+00
0.05	9.95e−14	2.36e−11	3.50e−12	7.09e−12	7.12e−12	1.00e−03	1.00e−03	1.00e−03	1.00e−03	0.00e+00
0.10	0.00e+00	5.40e−12	6.96e−13	1.00e−12	1.41e−12	1.00e−03	1.00e−03	1.00e−03	1.00e−03	0.00e+00
0.50	4.20e+01	4.20e+01	4.20e+01	4.20e+01	0.00e+00	6.00e+00	6.00e+00	6.00e+00	6.00e+00	0.00e+00
1.00	4.20e+01	4.20e+01	4.20e+01	4.20e+01	0.00e+00	6.00e+00	6.00e+00	6.00e+00	6.00e+00	0.00e+00
1.50	4.20e+01	4.20e+01	4.20e+01	4.20e+01	0.00e+00	6.00e+00	6.00e+00	6.00e+00	6.00e+00	0.00e+00
5.00	4.20e+01	4.20e+01	4.20e+01	4.20e+01	0.00e+00	6.00e+00	6.00e+00	6.00e+00	6.00e+00	0.00e+00
5.50	4.20e+01	4.20e+01	4.20e+01	4.20e+01	0.00e+00	6.00e+00	6.00e+00	6.00e+00	6.00e+00	0.00e+00
10.00	4.20e+01	4.20e+01	4.20e+01	4.20e+01	0.00e+00	6.00e+00	6.00e+00	6.00e+00	6.00e+00	0.00e+00

δ	AE_p^F					AE_p^f				
	Min	Max	Median	Mean	Std	Min	Max	Median	Mean	Std
0.001	1.77e−02	2.97e−02	2.09e−02	2.19e−02	3.38e−03	2.71e−07	9.83e−06	1.34e−06	2.17e−06	2.85e−06
0.005	4.96e−03	7.11e−03	5.65e−03	5.72e−03	6.78e−04	5.58e−07	1.50e−05	4.72e−06	6.72e−06	5.52e−06
0.01	1.36e−03	2.95e−03	1.97e−03	2.17e−03	5.40e−04	5.76e−07	5.55e−06	2.56e−06	2.74e−06	1.83e−06
0.05	2.99e−07	6.58e−06	9.53e−07	1.62e−06	1.91e−06	1.08e−08	2.99e−07	2.92e−08	6.31e−08	8.73e−08
0.10	9.53e−09	5.81e−08	1.96e−08	2.29e−08	1.39e−08	9.20e−11	2.29e−08	1.11e−08	1.07e−08	7.30e−09
0.50	5.84e−10	1.40e−09	8.78e−10	9.02e−10	2.42e−10	5.84e−10	1.40e−09	8.78e−10	9.02e−10	2.42e−10
1.00	2.26e−09	7.58e−09	4.05e−09	4.30e−09	1.80e−09	2.26e−09	7.58e−09	4.05e−09	4.30e−09	1.80e−09
1.50	6.74e+00	8.23e+00	7.57e+00	7.52e+00	4.52e−01	5.84e+00	8.68e+00	7.57e+00	7.53e+00	7.73e−01
5.00	1.14e+01	1.16e+01	1.15e+01	1.15e+01	7.32e−02	1.15e+01	1.21e+01	1.17e+01	1.17e+01	1.72e−01
5.50	1.14e+01	1.16e+01	1.16e+01	1.15e+01	7.06e−02	1.15e+01	1.20e+01	1.17e+01	1.17e+01	1.31e−01
10.00	1.17e+01	1.20e+01	1.18e+01	1.18e+01	1.04e−01	1.18e+01	1.20e+01	1.19e+01	1.19e+01	7.89e−02



(a) Absolute error (optimistic) for upper level objective.
(b) Absolute error (optimistic) for lower level objective.
(c) Absolute error (pessimistic) for upper level objective.
(d) Absolute error (pessimistic) for lower level objective.

Fig. 15. Boxplots of Absolute Error (AE) for each different δ for PA, for Optimistic and Pessimistic Upper and Lower Level Functions over 20 runs.



(a) Optimistic case. (b) Pessimistic case.

Fig. 16. Theoretical error bound (shaded) and absolute error for lower level objective from 20 runs of PA for various values of δ.

Table 5
Test problems and their optimistic and pessimistic global solutions. [70,73].

Problem	Formulation	Best known solution	
		Optimistic	Pessimistic
Problem 1 ($mb\_1\_1\_4$)	$\min_{x,y} \ y$ s.t. $y \in \underset{y}{\operatorname{argmin}} \{x(16y^4 + 2y^3 - 8y^2 - 3y/2 + 1/2)$: $-4/5 \leq y \leq 1\}$ $-1 \leq x \leq 1$	$F = -4/5$ $x = 0$ $y = -0.8$	$F = 1/2$ $x = (0, 1]$ $y = 0.5$
Problem 2 ($mb\_1\_1\_6$)	$\min_{x,y} \ x - y$ s.t. $y \in \underset{y}{\operatorname{argmin}} \{xy^2/2 - yx^3$: $-1 \leq y \leq 1\}$ $-1 \leq x \leq 1$	$F = -1$ $x = 0$ $y = 1$	$F = 0$ $(x, y) \in$ $\{(-1, -1), (1, 1)\}$
Problem 3 ($mb\_1\_1\_11$)	$\min_{x,y} \ x^2 + y^2$ s.t. $y \in \underset{y}{\operatorname{argmin}} \{xy^2 - y^4/2$: $-1 \leq y \leq 1\}$ $-1 \leq x \leq 1$	$F = 1/4$ $x = 1/2$ $y = 0$	$F = 1/4 + a^2 + a$ $x = 1/2 + a$ $y = 0$
Problem 4 ($mb\_1\_1\_12$)	$\min_{x,y} \ xy - y + y^2/2$ s.t. $y \in \underset{y}{\operatorname{argmin}} \{-xy^2 + y^4/2$: $-1 \leq y \leq 1\}$ $-1 \leq x \leq 1$	$F = -0.258$ $x = 0.189$ $y = \sqrt{0.189}$	$F = 0$ $x = [-1, 0]$ $y = 0$
Problem 5 ($mb\_1\_1\_17$)	$\min_{x,y} \ x^2 - y$ s.t. $y \in \underset{y}{\operatorname{argmin}} \{[(y - 1 - x/10)^2 - x/2 - 1/2]^2$: $0 \leq y \leq 3\}$ $0 \leq x \leq 1$	$F = -1.755$ $x = 0.2106$ $y = 1.799$	$F = -0.2929$ $x = 0$ $y = 0.2929$
Principal Agent	$\max_{x,y} \ 5x_1 + 8x_2 + 4y_1 - y_2$ s.t. $y \in \underset{y}{\operatorname{argmax}} \{y_1 + y_2$: $y_1 + y_2 \leq 15 - x_1 - 2x_2, y_1, y_2 \geq 0\}$ $x_1 + x_2 \leq 6, x_1, x_2 \geq 0$	$F = 66$ $x = 6, 0$ $y = 9, 0$	$F = 45$ $x = 0, 6$ $y = 0, 3$

* where a is a small positive value.

Table 6
Statistical results of Absolute Errors (AE) for Problem 1.

δ	AE_o^F					AE_o^f				
	Min	Max	Median	Mean	Std	Min	Max	Median	Mean	Std
0.001	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	6.56e−06	4.19e−04	1.57e−04	1.76e−04	1.27e−04
0.005	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	8.03e−05	2.70e−03	5.55e−04	1.03e−03	8.23e−04
0.01	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	2.44e−04	4.32e−03	1.51e−03	1.91e−03	1.24e−03
0.05	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	8.79e−05	2.66e−02	1.41e−02	1.31e−02	8.21e−03
0.10	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	2.92e−03	5.10e−02	2.42e−02	2.28e−02	1.32e−02
0.50	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	1.43e−03	2.61e−01	4.88e−02	8.48e−02	7.94e−02
1.00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	7.98e−03	5.00e−01	2.06e−01	2.40e−01	1.70e−01
1.50	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	8.29e−03	4.56e−01	2.30e−01	2.29e−01	1.40e−01
5.00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	2.48e−02	1.11e+00	3.20e−01	4.99e−01	4.14e−01
5.50	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	6.25e−02	1.11e+00	4.30e−01	5.80e−01	4.34e−01
10.00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	9.71e−03	1.11e+00	4.80e−01	5.30e−01	4.04e−01
δ	AE_p^F					AE_p^f				
	Min	Max	Median	Mean	Std	Min	Max	Median	Mean	Std
0.001	1.44e−05	6.40e−05	2.44e−05	2.83e−05	1.20e−05	1.00e−06	9.01e−02	1.00e−06	1.17e−02	2.79e−02
0.005	6.39e−05	1.11e−04	8.28e−05	8.43e−05	1.16e−05	1.00e−06	1.79e−02	1.00e−06	2.76e−03	5.59e−03
0.01	1.82e−04	2.37e−04	2.16e−04	2.12e−04	1.72e−05	1.00e−06	2.70e−02	1.00e−06	2.17e−03	6.39e−03
0.05	1.24e−03	1.29e−03	1.27e−03	1.27e−03	1.22e−05	2.93e−05	3.18e−05	3.09e−05	3.09e−05	5.91e−07
0.10	2.56e−03	2.60e−03	2.58e−03	2.58e−03	9.25e−06	1.26e−04	1.29e−04	1.27e−04	1.27e−04	9.13e−07
0.50	1.27e−02	1.27e−02	1.27e−02	1.27e−02	1.02e−05	3.11e−03	4.28e−03	3.13e−03	3.18e−03	2.59e−04
1.00	2.46e−02	2.46e−02	2.46e−02	2.46e−02	1.68e−05	1.20e−02	1.20e−02	1.20e−02	1.20e−02	1.68e−05
1.50	3.58e−02	3.59e−02	3.59e−02	3.59e−02	1.54e−05	2.60e−02	2.61e−02	2.60e−02	2.60e−02	2.31e−05
5.00	1.02e−01	1.02e−01	1.02e−01	1.02e−01	2.93e−05	2.34e−01	2.35e−01	2.35e−01	2.34e−01	1.46e−04
5.50	1.10e−01	1.10e−01	1.10e−01	1.10e−01	3.54e−05	2.76e−01	2.77e−01	2.77e−01	2.77e−01	1.94e−04
10.00	1.73e−01	1.73e−01	1.73e−01	1.73e−01	7.22e−05	7.61e−01	7.63e−01	7.62e−01	7.62e−01	7.21e−04

Table 7
Statistical results of Absolute Errors (AE) for Problem 2.

δ	AE_o^F					AE_o^f				
	Min	Max	Median	Mean	Std	Min	Max	Median	Mean	Std
0.0001	2.90e−03	4.64e−02	4.64e−02	1.51e−02	3.94e−02	1.45e−03	2.31e−02	2.31e−02	1.96e−02	7.50e−03
0.001	9.95e−02	9.54e−01	1.00e−01	1.43e−01	1.91e−01	3.09e−05	4.90e−02	4.90e−02	4.65e−02	1.09e−02
0.005	1.71e−01	1.71e−01	1.71e−01	1.71e−01	7.17e−05	8.04e−02	8.05e−02	8.05e−02	8.05e−02	2.96e−05
0.01	2.15e−01	2.15e−01	2.15e−01	2.15e−01	8.61e−05	9.76e−02	9.77e−02	9.77e−02	9.77e−02	3.11e−05
0.05	3.68e−01	3.68e−01	3.68e−01	3.68e−01	5.85e−05	1.34e−01	1.34e−01	1.34e−01	1.34e−01	5.44e−06
0.10	4.64e−01	4.64e−01	4.64e−01	4.64e−01	7.51e−05	1.32e−01	1.32e−01	1.32e−01	1.32e−01	1.10e−05
0.50	7.93e−01	7.94e−01	7.94e−01	7.94e−01	8.39e−05	1.03e−01	1.03e−01	1.03e−01	1.03e−01	1.17e−04
1.00	1.00e+00	1.00e+00	1.00e+00	1.00e+00	0.00e+00	5.00e−01	5.00e−01	5.00e−01	5.00e−01	0.00e+00
1.50	1.00e+00	1.00e+00	1.00e+00	1.00e+00	0.00e+00	5.00e−01	5.00e−01	5.00e−01	5.00e−01	0.00e+00
5.00	1.00e+00	1.00e+00	1.00e+00	1.00e+00	0.00e+00	5.00e−01	5.00e−01	5.00e−01	5.00e−01	0.00e+00
5.50	1.00e+00	1.00e+00	1.00e+00	1.00e+00	0.00e+00	5.00e−01	5.00e−01	5.00e−01	5.00e−01	0.00e+00
10.00	1.00e+00	1.00e+00	1.00e+00	1.00e+00	0.00e+00	5.00e−01	5.00e−01	5.00e−01	5.00e−01	0.00e+00
δ	AE_p^F					AE_p^f				
	Min	Max	Median	Mean	Std	Min	Max	Median	Mean	Std
0.001	1.00e−06	5.52e−04	1.00e−06	2.85e−05	1.23e−04	1.00e−06	1.38e−03	1.00e−06	6.98e−05	3.08e−04
0.005	1.00e−06	1.00e−06	1.00e−06	1.00e−06	0.00e+00	1.00e−06	1.00e−06	1.00e−06	1.00e−06	0.00e+00
0.01	1.00e−06	1.00e−06	1.00e−06	1.00e−06	0.00e+00	1.00e−06	1.00e−06	1.00e−06	1.00e−06	0.00e+00
0.05	1.00e−06	1.00e−06	1.00e−06	1.00e−06	0.00e+00	1.00e−06	1.00e−06	1.00e−06	1.00e−06	0.00e+00
0.10	1.00e−06	1.00e−06	1.00e−06	1.00e−06	0.00e+00	1.00e−06	1.00e−06	1.00e−06	1.00e−06	0.00e+00
0.50	1.00e−06	1.00e−06	1.00e−06	1.00e−06	0.00e+00	1.00e−06	1.00e−06	1.00e−06	1.00e−06	0.00e+00
1.00e	1.00e−06	1.00e−06	1.00e−06	1.00e−06	0.00e+00	1.00e−06	1.00e−06	1.00e−06	1.00e−06	0.00e+00
1.50	1.00e−06	1.00e−06	1.00e−06	1.00e−06	0.00e+00	1.00e−06	1.00e−06	1.00e−06	1.00e−06	0.00e+00
5.00	1.00e−06	1.00e−06	1.00e−06	1.00e−06	0.00e+00	1.00e−06	1.00e−06	1.00e−06	1.00e−06	0.00e+00
5.50	1.00e−06	1.00e−06	1.00e−06	1.00e−06	0.00e+00	1.00e−06	1.00e−06	1.00e−06	1.00e−06	0.00e+00
10.00	1.00e−06	1.00e−06	1.00e−06	1.00e−06	0.00e+00	1.00e−06	1.00e−06	1.00e−06	1.00e−06	0.00e+00

Table 8
Statistical results of Absolute Errors (AE) for Problem 3.

δ	AE_o^F					AE_o^f				
	Min	Max	Median	Mean	Std	Min	Max	Median	Mean	Std
0.001	9.26e−04	9.96e−04	9.83e−04	9.80e−04	1.93e−05	1.00e−06	4.84e−06	1.00e−06	1.50e−06	9.23e−07
0.005	4.88e−03	4.97e−03	4.96e−03	4.95e−03	2.14e−05	1.00e−06	2.74e−06	1.00e−06	1.09e−06	3.89e−07
0.01	9.86e−03	9.90e−03	9.89e−03	9.89e−03	1.08e−05	1.00e−06	4.77e−06	1.00e−06	1.37e−06	1.15e−06
0.05	4.75e−02	4.75e−02	4.75e−02	4.75e−02	1.29e−05	1.00e−06	3.67e−06	1.00e−06	1.13e−06	5.98e−07
0.10	9.00e−02	9.00e−02	9.00e−02	9.00e−02	7.70e−06	1.00e−06	1.10e−06	1.00e−06	1.01e−06	2.25e−08
0.50	2.50e−01	2.50e−01	2.50e−01	2.50e−01	2.10e−12	1.00e−06	1.00e−06	1.00e−06	1.00e−06	0.00e+00
1.00	2.50e−01	2.50e−01	2.50e−01	2.50e−01	7.50e−15	1.00e−06	1.00e−06	1.00e−06	1.00e−06	0.00e+00
1.50	2.50e−01	2.50e−01	2.50e−01	2.50e−01	1.54e−15	1.00e−06	1.00e−06	1.00e−06	1.00e−06	0.00e+00
5.00	2.50e−01	2.50e−01	2.50e−01	2.50e−01	2.14e−15	1.00e−06	1.00e−06	1.00e−06	1.00e−06	0.00e+00
5.50	2.50e−01	2.50e−01	2.50e−01	2.50e−01	1.41e−15	1.00e−06	1.00e−06	1.00e−06	1.00e−06	0.00e+00
10.00	2.50e−01	2.50e−01	2.50e−01	2.50e−01	4.05e−15	1.00e−06	1.00e−06	1.00e−06	1.00e−06	0.00e+00
δ	AE_p^F					AE_p^f				
	Min	Max	Median	Mean	Std	Min	Max	Median	Mean	Std
0.001	1.66e−06	7.65e−05	1.87e−05	2.25e−05	1.75e−05	1.00e−06	1.52e−06	1.00e−06	1.03e−06	1.16e−07
0.005	4.03e−03	4.11e−03	4.04e−03	4.04e−03	1.92e−05	1.00e−06	1.29e−06	1.00e−06	1.04e−06	9.22e−08
0.01	9.10e−03	9.14e−03	9.12e−03	9.12e−03	1.09e−05	1.00e−06	1.98e−06	1.00e−06	1.10e−06	2.67e−07
0.05	5.15e−02	5.16e−02	5.15e−02	5.15e−02	1.68e−05	1.00e−06	1.40e−05	1.00e−06	1.70e−06	2.90e−06
0.10	1.09e−01	1.09e−01	1.09e−01	1.09e−01	2.18e−05	1.00e−06	3.78e−06	1.00e−06	1.31e−06	8.05e−07
0.50	7.49e−01	7.49e−01	7.49e−01	7.49e−01	3.07e−09	5.00e−01	5.00e−01	5.00e−01	5.00e−01	4.71e−05
1.00	7.49e−01	7.49e−01	7.49e−01	7.49e−01	3.05e−09	5.00e−01	5.00e−01	5.00e−01	5.00e−01	3.80e−05
1.50	7.49e−01	7.49e−01	7.49e−01	7.49e−01	4.81e−09	5.00e−01	5.00e−01	5.00e−01	5.00e−01	4.38e−05
5.00	7.49e−01	7.49e−01	7.49e−01	7.49e−01	2.21e−09	5.00e−01	5.00e−01	5.00e−01	5.00e−01	3.12e−05
5.50	7.49e−01	7.49e−01	7.49e−01	7.49e−01	6.12e−10	5.00e−01	5.00e−01	5.00e−01	5.00e−01	2.13e−05
10.00	7.49e−01	7.49e−01	7.49e−01	7.49e−01	1.47e−09	5.00e−01	5.00e−01	5.00e−01	5.00e−01	3.15e−05

Table 9
Statistical results of Absolute Errors (AE) for Problem 4.

δ	AE_o^F					AE_o^f				
	Min	Max	Median	Mean	Std	Min	Max	Median	Mean	Std
0.001	2.65e−04	2.65e−04	2.65e−04	2.65e−04	5.64e−08	2.13e−04	2.66e−04	2.43e−04	2.42e−04	1.54e−05
0.005	1.03e−03	1.03e−03	1.03e−03	1.03e−03	8.00e−08	7.00e−04	7.84e−04	7.41e−04	7.41e−04	1.80e−05
0.01	2.00e−03	2.00e−03	2.00e−03	2.00e−03	5.69e−08	1.35e−03	1.44e−03	1.38e−03	1.38e−03	2.20e−05
0.05	1.04e−02	1.04e−02	1.04e−02	1.04e−02	9.56e−08	6.78e−03	6.88e−03	6.83e−03	6.83e−03	2.37e−05
0.10	2.32e−02	2.32e−02	2.32e−02	2.32e−02	1.22e−07	3.03e−03	3.13e−03	3.09e−03	3.09e−03	2.29e−05
0.50	4.01e−01	4.01e−01	4.01e−01	4.01e−01	2.44e−06	1.40e−01	1.40e−01	1.40e−01	1.40e−01	1.22e−06
1.00	6.96e−01	6.96e−01	6.96e−01	6.96e−01	7.52e−06	3.53e−01	3.53e−01	3.53e−01	3.53e−01	7.52e−06
1.50	8.69e−01	8.69e−01	8.69e−01	8.69e−01	1.57e−05	5.67e−01	5.67e−01	5.67e−01	5.67e−01	2.36e−05
5.00	1.24e+00	1.24e+00	1.24e+00	1.24e+00	0.00e+00	1.50e+00	1.50e+00	1.50e+00	1.50e+00	0.00e+00
5.50	1.24e+00	1.24e+00	1.24e+00	1.24e+00	0.00e+00	1.50e+00	1.50e+00	1.50e+00	1.50e+00	0.00e+00
10.00	1.24e+00	1.24e+00	1.24e+00	1.24e+00	0.00e+00	1.50e+00	1.50e+00	1.50e+00	1.50e+00	0.00e+00
δ	AE_p^F					AE_p^f				
	Min	Max	Median	Mean	Std	Min	Max	Median	Mean	Std
0.001	2.00e−03	2.00e−03	2.00e−03	2.00e−03	2.28e−08	1.00e−06	1.00e−06	1.00e−06	1.00e−06	2.28e−11
0.005	1.00e−02	1.00e−02	1.00e−02	1.00e−02	2.03e−08	2.51e−05	2.51e−05	2.51e−05	2.51e−05	1.01e−10
0.01	2.01e−02	2.01e−02	2.01e−02	2.01e−02	1.66e−08	1.01e−04	1.01e−04	1.01e−04	1.01e−04	1.66e−10
0.05	1.04e−01	1.04e−01	1.04e−01	1.04e−01	7.16e−08	2.62e−03	2.62e−03	2.62e−03	2.62e−03	3.58e−09
0.10	2.14e−01	2.14e−01	2.14e−01	2.14e−01	1.88e−07	1.09e−02	1.09e−02	1.09e−02	1.09e−02	1.88e−08
0.50	5.00e−01	5.00e−01	5.00e−01	5.00e−01	0.00e+00	5.00e−01	5.00e−01	5.00e−01	5.00e−01	0.00e+00
1.00	5.00e−01	5.00e−01	5.00e−01	5.00e−01	0.00e+00	5.00e−01	5.00e−01	5.00e−01	5.00e−01	0.00e+00
1.50	5.00e−01	5.00e−01	5.00e−01	5.00e−01	0.00e+00	5.00e−01	5.00e−01	5.00e−01	5.00e−01	0.00e+00
5.00	5.00e−01	5.00e−01	5.00e−01	5.00e−01	0.00e+00	5.00e−01	5.00e−01	5.00e−01	5.00e−01	0.00e+00
5.50	5.00e−01	5.00e−01	5.00e−01	5.00e−01	0.00e+00	5.00e−01	5.00e−01	5.00e−01	5.00e−01	0.00e+00
10.00	5.00e−01	5.00e−01	5.00e−01	5.00e−01	0.00e+00	5.00e−01	5.00e−01	5.00e−01	5.00e−01	0.00e+00

Table 10
Statistical results of Absolute Errors (AE) for Problem 5.

δ	AE_o^F					AE_o^f				
	Min	Max	Median	Mean	Std	Min	Max	Median	Mean	Std
0.001	7.56e−05	7.57e−05	7.56e−05	7.56e−05	1.76e−08	1.00e−06	1.00e−06	1.00e−06	1.00e−06	0.00e+00
0.005	7.49e−04	7.49e−04	7.49e−04	7.49e−04	7.56e−08	2.57e−06	2.58e−06	2.58e−06	2.58e−06	6.14e−10
0.01	1.78e−03	1.78e−03	1.78e−03	1.78e−03	2.02e−07	1.03e−05	1.03e−05	1.03e−05	1.03e−05	3.10e−09
0.05	9.86e−03	9.86e−03	9.86e−03	9.86e−03	8.35e−07	2.52e−04	2.53e−04	2.52e−04	2.52e−04	9.28e−08
0.10	1.97e−02	1.97e−02	1.97e−02	1.97e−02	1.73e−06	9.87e−04	9.89e−04	9.88e−04	9.88e−04	5.73e−07
0.50	8.84e−02	8.84e−02	8.84e−02	8.84e−02	8.15e−06	2.12e−02	2.13e−02	2.12e−02	2.12e−02	2.58e−05
1.00	1.58e−01	1.58e−01	1.58e−01	1.58e−01	2.23e−05	7.31e−02	7.34e−02	7.32e−02	7.33e−02	7.09e−05
1.50	2.16e−01	2.17e−01	2.16e−01	2.16e−01	2.94e−05	1.46e−01	1.47e−01	1.47e−01	1.47e−01	2.02e−04
5.00	4.92e−01	4.92e−01	4.92e−01	4.92e−01	7.52e−05	1.00e+00	1.00e+00	1.00e+00	1.00e+00	8.95e−04
5.50	5.22e−01	5.22e−01	5.22e−01	5.22e−01	1.05e−04	1.16e+00	1.16e+00	1.16e+00	1.16e+00	1.57e−03
10.00	7.38e−01	7.39e−01	7.38e−01	7.38e−01	2.65e−04	2.81e+00	2.82e+00	2.81e+00	2.81e+00	4.83e−03
δ	AE_p^F					AE_p^f				
	Min	Max	Median	Mean	Std	Min	Max	Median	Mean	Std
0.001	2.57e−04	2.57e−04	2.57e−04	2.57e−04	1.16e−08	1.00e−06	1.00e−06	1.00e−06	1.00e−06	0.00e+00
0.005	1.25e−03	1.25e−03	1.25e−03	1.25e−03	2.08e−08	3.11e−06	3.11e−06	3.11e−06	3.11e−06	1.04e−10
0.01	2.49e−03	2.49e−03	2.49e−03	2.49e−03	4.68e−08	1.24e−05	1.24e−05	1.24e−05	1.24e−05	4.67e−10
0.05	1.22e−02	1.22e−02	1.22e−02	1.22e−02	1.09e−07	3.02e−04	3.02e−04	3.02e−04	3.02e−04	5.44e−09
0.10	2.38e−02	2.38e−02	2.38e−02	2.38e−02	3.39e−07	1.17e−03	1.17e−03	1.17e−03	1.17e−03	3.39e−08
0.50	1.02e−01	1.02e−01	1.02e−01	1.02e−01	8.44e−07	2.39e−02	2.39e−02	2.39e−02	2.39e−02	4.22e−07
1.00	1.78e−01	1.78e−01	1.78e−01	1.78e−01	1.72e−06	7.99e−02	7.99e−02	7.99e−02	7.99e−02	1.72e−06
1.50	2.40e−01	2.40e−01	2.40e−01	2.40e−01	2.95e−06	1.57e−01	1.57e−01	1.57e−01	1.57e−01	4.42e−06
5.00	2.93e−01	2.93e−01	2.93e−01	2.93e−01	0.00e+00	2.50e−01	2.50e−01	2.50e−01	2.50e−01	0.00e+00
5.50	2.93e−01	2.93e−01	2.93e−01	2.93e−01	0.00e+00	2.50e−01	2.50e−01	2.50e−01	2.50e−01	0.00e+00
10.00	2.93e−01	2.93e−01	2.93e−01	2.93e−01	0.00e+00	2.50e−01	2.50e−01	2.50e−01	2.50e−01	0.00e+00

References

[1] Bracken J, McGill J. Mathematical programs with optimization problems in the constraints. Oper Res 1973;21:37–44.
[2] Sinha A, Malo P, Deb K. A review on bilevel optimization: From classical to evolutionary approaches and applications. IEEE Trans Evol Comput 2018;22(2):276–95.
[3] Von Stackelberg Heinrich. Market structure and equilibrium. Springer Science & Business Media; 2010.
[4] Migdalas A. Bilevel programming in traffic planning: Models, methods and challenge. J Global Optim 1995;7(4):381–405.
[5] Brotcorne L, Labbe M, Marcotte P, Savard G. A bilevel model for toll optimization on a multicommodity transportation network. Transp Sci 2001;35(4):345–58.
[6] Sinha A, Malo P, Deb K. Transportation policy formulation as a multi-objective bilevel optimization problem. In: 2015 IEEE congress on evolutionary computation. IEEE Press; 2015.
[7] Beraldi Patrizia. Green retailer: A stochastic bi-level approach to support investment decisions in sustainable energy systems. Oper Res Perspect 2024;100300.
[8] Smith WR, Missen RW. Chemical reaction equilibrium analysis: Theory and algorithms. John Wiley & Sons, New York; 1982.
[9] Clark PA, Westerberg AW. Bilevel programming for steady-state chemical process design-I. Fundamentals and algorithms. Comput Chem Eng 1990;14(1):87–97.
[10] Labbé M, Marcotte P, Savard G. A bilevel model of taxation and its application to optimal highway pricing. Manage Sci 1998;44(12):1608–22.
[11] Sinha A, Malo P, Frantsev A, Deb K. Multi-objective Stackelberg game between a regulating authority and a mining company: A case study in environmental economics. In: 2013 IEEE congress on evolutionary computation. IEEE Press; 2013.
[12] Brown G, Carlyle M, Diehl D, Kline J, Wood K. A two-sided optimization for theater ballistic missile defense. Oper Res 2005;53(5):745–63.
[13] Wein L. Homeland Security: From Mathematical Models to Policy Implementation: The 2008 Philip McCord Morse Lecture. Oper Res 2009;57(4):801–11.
[14] Yang Hongjiu, Shi Min, Xia Yuanqing, Zhang Peng. Security research on wireless networked control systems subject to jamming attacks. IEEE Trans Cybern 2018;49(6):2022–31.
[15] Nicholls MG. Aluminium Production Modeling - A Nonlinear Bilevel Programming Approach. Oper Res 1995;43(2):208–18.
[16] Hu X, Ralph D. Using EPECs to Model Bilevel Games in Restructured Electricity Markets with Locational Prices. Oper Res 2007;55(5):809–27.
[17] Rodriguez-Martinez Eduardo, Mu Tingting, Goulermas John Yannis. Sequential projection pursuit with kernel matrix update and symbolic model selection. IEEE Trans Cybern 2014;44(12):2458–69.
[18] Huang Pei-Qiu, Wang Yong, Wang Kezhi, Liu Zhi-Zhong. A bilevel optimization approach for joint offloading decision and resource allocation in cooperative mobile edge computing. IEEE Trans Cybern 2019.
[19] Bennett Kristin P, Kunapuli Gautam, Hu Jing, Pang Jong-Shi. Bilevel optimization and machine learning. In: Computational intelligence: Research frontiers. Springer; 2008, p. 25–47.
[20] Sinha Ankur, Khandait Tanmay, Mohanty Raja. A gradient-based bilevel optimization approach for tuning regularization hyperparameters. Optim Lett 2023;1–22.
[21] Colson Benoît, Marcotte Patrice, Savard Gilles. An overview of bilevel optimization. Ann Oper Res 2007;153:235–56.
[22] Antoniou Margarita, Korošec Peter. Multilevel optimisation. In: Optimization under uncertainty with applications to aerospace engineering. Springer; 2021, p. 307–31.
[23] Dempe Stephan, Zemkoho Alain. Bilevel optimization. In: Springer optimization and its applications, vol. 161, Springer; 2020.
[24] Hansen Pierre, Jaumard Brigitte, Savard Gilles. New branch-and-bound rules for linear bilevel programming. SIAM J Sci Stat Comput 1992;13(5):1194–217.
[25] Allende Gemayqzel Bouza, Still Georg. Solving bilevel programs with the KKT-approach. Math Program 2013;138:309–32.
[26] Dempe Stephan, Zemkoho Alain B. KKT reformulation and necessary conditions for optimality in nonsmooth bilevel optimization. SIAM J Optim 2014;24(4):1639–69.
[27] Liu G, Han J, Wang S. A trust region algorithm for bilevel programing problems. Chin Sci Bull 1998;43(10):820–4.
[28] Marcotte P, Savard G, Zhu DL. A trust region algorithm for nonlinear bilevel programming. Oper Res Lett 2001;29(4):171–9.
[29] Colson B, Marcotte P, Savard G. A trust-region method for nonlinear bilevel programming: Algorithm and computational experience. Comput Optim Appl 2005;30(3):211–27.
[30] Ishizuka Y, Aiyoshi E. Double penalty method for bilevel optimization problems. Ann Oper Res 1992;34:73–88.
[31] White D, Anandalingam G. A penalty function approach for solving bi-level linear programs. J Global Optim 1993;3:397–419.
[32] Kleinert Thomas, Schmidt Martin. Computing feasible points of bilevel problems with a penalty alternating direction method. INFORMS J Comput 2021;33(1):198–215.
[33] Savard G, Gauvin J. The steepest descent direction for the nonlinear bilevel programming problem. Oper Res Lett 1994;15:275–82.
[34] Vicente L, Savard G, Júdice J. Descent approaches for quadratic bilevel programming. J Optim Theory Appl 1994;81:379–99.
[35] Colson Bruno, Marcotte Pierre, Savard Gilles. Bilevel programming: A survey. In: 10th anniversary volume of the annals of operations research, vol. 153, Springer; 2007, p. 235–56.
[36] Deb Kalyanmoy, Sinha Ankur, Malo Pekka, Lu Zhichao. Approximate bilevel optimization with population-based evolutionary algorithms. In: Bilevel optimization: Advances and next challenges. Springer; 2020, p. 361–402.
[37] Mathieu R, Pittard L, Anandalingam G. Genetic algorithm based approach to bi-level linear programming. Oper Res 1994;28(1):1–21.
[38] Yin Y. Genetic algorithm based approach for bilevel programming models. J Transp Eng 2000;126(2):115–20.

[39] Li H. A genetic algorithm using a finite search space for solving nonlinear/linear fractional bilevel programming problems. Ann Oper Res 2015;1–16.
[40] Sinha A, Malo P, Deb K. Evolutionary algorithm for bilevel optimization using approximations of the lower level optimal solution mapping. European J Oper Res 2017;257(2):395–411.
[41] Angelo J, Krempser E, Barbosa H. Differential evolution for bilevel programming. In: Proceedings of the 2013 congress on evolutionary computation. IEEE Press; 2013.
[42] Wu Lianghong, Liu Zhenzu, Wei Hua-Liang, Wang Rui. An efficient bilevel differential evolution algorithm with adaptation of lower level population size and search radius. Memetic Comput 2021;13(2):227–47.
[43] Kieffer Emmanuel, Danoy Grégoire, Bouvry Pascal, Nagih Anass. Bayesian optimization approach of general bi-level problems. In: Proceedings of the genetic and evolutionary computation conference companion. 2017, p. 1614–21.
[44] Karoonsoontawong Ampol, Waller S Travis. Dynamic continuous network design problem: Linear bilevel programming and metaheuristic approaches. Transp Res Rec 2006;1964(1):104–17.
[45] Li X, Tian P, Min X. A hierarchical particle swarm optimization for solving bilevel programming problems. In: Rutkowski Leszek, Tadeusiewicz Ryszard, Zadeh Lotfi A, Zurada Jacek M, editors. Artificial intelligence and soft computing - ICAISC 2006. Lecture notes in computer science, vol. 4029, Springer Berlin Heidelberg; 2006, p. 1169–78.
[46] Jiang Yan, Li Xuyong, Huang Chongchao, Wu Xianing. Application of particle swarm optimization based on CHKS smoothing function for solving nonlinear bilevel programming problem. Appl Math Comput 2013;219(9):4332–9.
[47] Calvete Herminia I, Galé Carmen, Iranzo José A, Toth Paolo. The school bus routing problem with student choice: a bilevel approach and a simple and effective metaheuristic. Int Trans Oper Res 2023;30(2):1092–119.
[48] Zhu X, Yu Q, Wang X. A hybrid differential evolution algorithm for solving nonlinear bilevel programming with linear constraints. In: Cognitive informatics, 2006. ICCI 2006. 5th IEEE international conference on, vol. 1, IEEE; 2006, p. 126–31.
[49] Salinas-Guerra Rocío, Mezura-Montes Efrén, Quiroz-Castellanos Marcela, Mejía-de Dios Jesús-Adolfo, Bechikh Slim. A hybrid evolutionary approach with group-based solution encoding for solving the constrained bilevel multi-depot vehicle routing problem. In: 2023 IEEE congress on evolutionary computation. IEEE; 2023, p. 1–8.
[50] Sinha A, Soun T, Deb K. Using karush-kuhn-tucker proximity measure for solving bilevel optimization problems. Swarm Evol Comput 2019;44:496–510.
[51] Islam Md Monjurul, Singh Hemant Kumar, Ray Tapabrata. A surrogate assisted approach for single-objective bilevel optimization. IEEE Trans Evol Comput 2017;21(5):681–96.
[52] Sinha Ankur, Shaikh Vaseem. Solving bilevel optimization problems using kriging approximations. IEEE Trans Cybern 2021;52(10):10639–54.
[53] Camacho-Vallejo José-Fernando, Corpus Carlos, Villegas Juan G. Metaheuristics for bilevel optimization: A comprehensive review. Comput Oper Res 2023;106410.
[54] Lignola Maria Beatrice, Morgan Jacqueline. Topological existence and stability for Stackelberg problems. J Optim Theory Appl 1995;84:145–69.
[55] Liu June, Fan Yuxin, Chen Zhong, Zheng Yue. Pessimistic bilevel optimization: A survey. Int J Comput Intell Syst 2018;11(1):725–36.
[56] Liu June, Fan Yuxin, Chen Zhong, Zheng Yue. Methods for pessimistic bilevel optimization. In: Bilevel optimization: Advances and next challenges. Springer; 2020, p. 403–20.
[57] Dempe Stephan, Mordukhovich Boris S, Zemkoho Alain Bertrand. Necessary optimality conditions in pessimistic bilevel programming. Optimization 2014;63(4):505–33.
[58] Zheng Yue, Fang Debin, Wan Zhongping. A solution approach to the weak linear bilevel programming problems. Optimization 2016;65(7):1437–49.
[59] Aboussoror Abdelmalek, Mansouri Abdelatif. Weak linear bilevel programming problems: Existence of solutions via a penalty method. J Math Anal Appl 2005;304(1):399–408.
[60] Zheng Yue, Wan Zhongping, Sun Kangtai, Zhang Tao. An exact penalty method for weak linear bilevel programming problem. J Appl Math Comput 2013;42(1–2):41–9.
[61] Lignola M Beatrice, Morgan Jacqueline. Inner regularizations and viscosity solutions for pessimistic bilevel optimization problems. J Optim Theory Appl 2017;173:183–202.
[62] Tsoukalas Angelos, Wiesemann Wolfram, Rustem Berc, et al. Global optimisation of pessimistic bi-level problems. In: Lectures on global optimization, vol. 55, American Mathematical Soc.; 2009, p. 215–43.
[63] Molodtsov DA. The solution of a class of non-antagonistic games. USSR Comput Math Math Phys 1976;16(6):67–72.
[64] Bialas Wayne F, Karwan Mark H. Two-level linear programming. Manage Sci 1984;30(8):1004–20.
[65] Wen Ue-Pyng, Hsu Shuh-Tzy. Linear bi-level programming problems—A review. J Oper Res Soc 1991;42:125–33.
[66] Cao Dong, Leung Lawrence C. A partial cooperation model for non-unique linear two-level decision problems. European J Oper Res 2002;140(1):134–41.
[67] Loridan Pierre, Morgan Jacqueline. Weak via strong Stackelberg problem: New results. J Global Optim 1996;8:263–87.
[68] Lucchetti R, Mignanego F, Pieri G. Existence theorem of equilibrium points in Stackelberg games with constraints. Optimization 1987;18:857–66.
[69] Loridan P, Morgan J. Weak via strong Stackelberg problems. J Global Optim 1996;8:263–87.
[70] Wiesemann Wolfram, Tsoukalas Angelos, Kleniati Polyxeni Margarita, Rustem Berç. Pessimistic Bilevel Optimization. SIAM J Optim 2013;23.
[71] Storn Rainer, Price Kenneth. Differential evolution–A simple and efficient heuristic for global optimization over continuous spaces. J Global Optim 1997;11:341–59.
[72] Deb Kalyanmoy. An efficient constraint handling method for genetic algorithms. Comput Methods Appl Mech Engrg 2000;186(2–4):311–38.
[73] Zheng Yue, Zhang Guangquan, Zhang Zhen, Lu Jie. A reducibility method for the weak linear bilevel programming problems and a case study in principal-agent. Inform Sci 2018;454:46–58.