A branch-and-bound algorithm for the computation of optimal point mappings of parametric optimization problems

Stefan Schwarze* Oliver Stein#

September 22, 2026

Abstract

We propose a novel branch-and-bound algorithm that constructs rigorous outer approximations of the optimal point mapping for parametric optimization problems with guaranteed feasibility and optimality tolerances. The method uses the improvement-function reformulation to define discarding and inclusion tests on subboxes, constructing a rigorous outer approximation. Under the same regularity conditions that ensure exactness of this reformulation, we prove finite termination of the resulting algorithm (P-ICGO). We provide an implementation in Julia and evaluate the algorithm in two studies: first, we illustrate the approximation behavior on some visual examples; second, we compute the follower's optimal point mapping on 99 bilevel optimization problems with up to four variables.

1 Introduction

This paper presents an algorithmic solution approach for parametric optimization problems of the form

$$P(t): \qquad \min_x \ f(x,t) \qquad \text{s.t.} \qquad x \in X, \quad \omega(x,t) \le 0$$

with $t \in T$, where f and ω are real-valued continuous functions on the nonempty and compact set $X \times T$. In fact, let $S(t)$ denote the minimal point set of $P(t)$ for $t \in T$. We introduce a branch-and-bound method which generates a rigorous outer approximation of the graph

$$\mathcal{S} \ := \ \operatorname{gph} S \ = \ \{(x,t) \in X \times T \mid x \in S(t)\}$$

*Institute for Operations Research (IOR), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, stefan.schwarze@kit.edu, Corresponding author

#Institute for Operations Research (IOR), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, stein@kit.edu

of the minimal point mapping $S : T \rightrightarrows X$ in the sense that it is not only valid, but that it also meets prescribed feasibility and optimality tolerances.

Graphs of minimal point mappings serve as fundamental components in complex optimization frameworks. For instance, they define the feasible sets in bilevel optimization problems through the lower-level minimal point mapping [2], and they characterize the set of equilibria in generalized Nash equilibrium problems via the intersection of graphs of the players' response mappings [6]. Similarly, in generalized semi-infinite optimization problems, which are reducible to bilevel structures [8], such graphs characterize the feasible sets. In the numerical experiments in Section 4.2, we focus on the application to feasible sets of bilevel optimization problems.

The paper is structured as follows. After providing necessary preliminaries in Section 2, in Section 3 we define our method, the parametric improvement function based complete global optimization (P-ICGO) algorithm. Under mild conditions, P-ICGO computes a list of boxes $\mathcal{O}$, such that $\mathcal{S} \subseteq \bigcup_{B\in\mathcal{O}} B \subseteq \mathcal{S}_{\varepsilon,\delta}$, where $\mathcal{S}_{\varepsilon,\delta}$ is a relaxation of $\mathcal{S}$ with specified optimality and feasibility tolerances. We explain the functionality of P-ICGO in detail and prove its finite termination. Finally, Section 4 is devoted to numerical experiments with P-ICGO, in order to discuss its properties and assess its capabilities.

2 Preliminaries

In Section 2.1 the parametric optimization problem $P(t)$ and the graph of its optimal point mapping $\mathcal{S}$ are defined in more detail, and the relaxation $\mathcal{S}_{\varepsilon,\delta}$ of $\mathcal{S}$ with specified optimality and feasibility tolerances is provided. Then, in Section 2.2, the improvement function reformulation of $\mathcal{S}$ is introduced and, in Section 2.3, extended to $\mathcal{S}_{\varepsilon,\delta}$.

2.1 Optimal point mappings and their relaxation

As introduced in Section 1, for nonempty compact sets $X \subseteq \mathbb{R}^n$ and $T \subseteq \mathbb{R}^r$, and for real-valued and continuous functions f and ω on $X \times T$, we consider the parametric optimization problem

$$P(t): \qquad \min_x \; f(x,t) \qquad \text{s.t.} \qquad x \in \Omega(t)$$

with the (possibly empty) compact sets

$$\Omega(t) \;=\; \{x \in X \mid \omega(x,t) \le 0\}$$

and $t \in T$. With the graph

$$\begin{aligned} G \;:=\; \operatorname{gph}\Omega \;&=\; \{(x,t) \in X \times T \mid x \in \Omega(t)\} \\ &=\; \{(x,t) \in X \times T \mid \omega(x,t) \le 0\} \end{aligned}$$

of the set-valued mapping $\Omega : T \rightrightarrows X$ we assume the parametric problem to be nontrivial in the sense that G is nonempty. Under these assumptions, $P(t)$ is solvable for each t from the (effective) domain of the set-valued mapping $\Omega : T \rightrightarrows X$,

$$\operatorname{dom}\Omega \;:=\; \{t \in T \mid \Omega(t) \neq \emptyset\}.$$

For each $t \in T$, let $S(t)$ and $v(t)$ be the set of minimal points and the minimal value of $P(t)$, respectively (with $S(t) = \emptyset$ and $v(t) = +\infty$ for $\Omega(t) = \emptyset$). Due to $S(t) \subseteq \Omega(t)$, with the graph

$$\mathcal{S} \;=\; \operatorname{gph} S \;:=\; \{(x,t) \in X \times T \mid x \in S(t)\}$$

of the minimal point mapping $S : T \rightrightarrows X$ one obtains $\mathcal{S} \subseteq G$, so that the graph may as well be written as $\mathcal{S} = \{(x,t) \in G \mid x \in S(t)\}$.

In the following, we aim to compute an outer approximation of $\mathcal{S}$ that meets specified tolerances on feasibility and optimality. Consequently, we define the relaxation of the feasible set

$$\Omega_\delta(t) := \{x \in X \mid \omega(x,t) \le \delta\}$$

and the set of ε-minimal and δ-feasible points

$$S_{\varepsilon,\delta}(t) := \{x \in \Omega_\delta(t) \mid f(x,t) \le v(t) + \varepsilon\}$$

for each $t \in T$ and nonnegative $\varepsilon, \delta \ge 0$. Together with the graph

$$G_\delta := \operatorname{gph} \Omega_\delta = \;\{(x,t) \in X \times T \mid \omega(x,t) \le \delta\}$$

of the set-valued mapping $\Omega_\delta \colon T \rightrightarrows X$ we finally define the graph of the set-valued mapping $S_{\varepsilon,\delta} \colon T \rightrightarrows X$ as

$$\begin{aligned} \mathcal{S}_{\varepsilon,\delta} := \operatorname{gph} S_{\varepsilon,\delta} &= \{(x,t) \in G_\delta \mid x \in S_{\varepsilon,\delta}(t)\} \\ &= \{(x,t) \in G_\delta \mid f(x,t) \le v(t) + \varepsilon\}. \end{aligned} \tag{1}$$

Note that it is defined by the optimal value $v(t)$ of $P(t)$ without relaxing the feasible set $\Omega(t)$. On the one hand, this formulation allows us to relax feasibility, without perturbing the original optimal points. On the other hand, for $t \notin \operatorname{dom} \Omega$ we have $S_{\varepsilon,\delta}(t) = \Omega_\delta(t)$, and thus no information on the objective function. However, with respect to the later introduced regularity condition, we will see that $\mathcal{S}_{\varepsilon,\delta}$ is still an appropriate approximation of $\mathcal{S}$.

2.2 The improvement function reformulation

The improvement function reformulation for $P(t)$ was introduced and extensively discussed in [6]. By using a maximum-term, it moves functional constraints from the description of the feasible set to the objective function without the need to specify appropriate penalty parameters, as in penalty (or barrier) approaches, or multipliers, as in Lagrange methods. The price to pay is the introduction of an additional parameter ‘s’, a nonsmooth maximum-term, and the fact that the exactness of the reformulation requires some mild assumptions.

Indeed, for the treatment of the parametric optimization problem $P(t)$, $t \in T$, the improvement function is defined as

$$\psi(x,s,t) \;:=\; \max\{\omega(x,t), f(x,t) - f(s,t)\}$$

for $x, s \in X$ and $t \in T$. This function has a negative value, if the point x is an improvement of s in the sense that it is strictly feasible and has a lower objective value. In short, we obtain a reformulation of $\mathcal{S}$ by collecting all feasible points (s,t) from the graph G that can not be improved in their s-component:

$$\mathcal{R} \ := \ \{(s,t) \in G \mid \min_{x\in X} \psi(x,s,t) \geq 0\}. \tag{2}$$

Note that, by a standard result from parametric optimization, the function $\min_{x\in X} \psi(x,\cdot,\cdot)$ is continuous on G, so that (2) yields the closedness of the set $\mathcal{R}$. In [6], the above reformulation is investigated and constraint qualifications are derived, under which $\mathcal{S}$ and $\mathcal{R}$ coincide. Without these mild conditions, at least $\mathcal{S} \subseteq \mathcal{R}$ is true. Therefore, under these assumptions methods for the outer approximation of $\mathcal{R}$ may be used for the approximation of $\mathcal{S}$. We will see in Section 2.3 that the constraint qualifications from [6] that ensure $\mathcal{R} = \mathcal{S}$ additionally imply $\mathcal{R}_{\varepsilon,0} = \mathcal{S}_{\varepsilon,0}$ for an appropriately defined set $\mathcal{R}_{\varepsilon,0}$.

This reformulation is particularly advantageous when X and T are boxes, enabling $\mathcal{R}$ to be computed effectively by a branch-and-bound method. We remark that the nonsmoothness of the maximum is not an issue for branch-and-bound methods, since they do not rely on smoothness of the involved functions but on the existence of convergent lower bounding procedures. The latter are available for the function ψ, if they are for f and ω. The description (2) of $\mathcal{R}$ also forms the basis for its algorithmic treatment. A branch-and-bound method can approximate $\mathcal{R}$ by removing all points (s,t) from $X \times T$ that satisfy $\omega(s,t) > 0$ or $\min_{x\in X} \psi(x,s,t) < 0$ on the one hand, and, on the other hand prove the inclusion inside a suitable relaxation $\mathcal{R}_{\varepsilon,\delta}$ with some tolerances $\varepsilon, \delta > 0$ after finitely many iterations. The relaxation $\mathcal{R}_{\varepsilon,\delta}$ will be defined in the next section.

The two following branch-and-bound methods worked on related structures. In [4], the right hand side of $\psi(x,s,t)$ was (implicitly) used to compute inner and outer approximations of the solution set of box-constrained Nash equilibrium problems. Hence, it implicitly worked on sets of the form

$$\{(s,t) \in X \times T \mid \min_{x\in X} f(x,t) - f(s,t) \geq 0\}$$

for each player. The framework for complete global optimization in [7] captures general constraints (without parameters). Therein, without parameters, the set $\mathcal{R}$ was defined as

$$\{s \in \Omega \mid \min_{x} \psi(x,s) \geq 0\}.$$

We will bring together those frameworks to compute an outer approximation of (2).

2.3 Incorporating tolerances into the reformulation

As an appropriate extension of the constructions from the previous section to general $\varepsilon, \delta \geq 0$, we define the ε-improvement function

$$\psi_\varepsilon(x,s,t) \ := \ \max\{\omega(x,t), f(x,t) - f(s,t) + \varepsilon\}$$

and the set

$$\mathcal{R}_{\varepsilon,\delta} \ := \ \{(s,t) \in G_\delta \mid \min_{x\in X} \psi_\varepsilon(x,s,t) \geq 0\}. \tag{3}$$

For every $\varepsilon, \delta \geq 0$, in view of the continuity of $\min_{x\in X} \psi_\varepsilon(x, \cdot, \cdot)$ and the closedness of G_δ, (3) yields the closedness of $\mathcal{R}_{\varepsilon,\delta}$.

To characterize the relations between the sets $\mathcal{R}_{\varepsilon,\delta}$ and $\mathcal{S}_{\varepsilon,\delta}$, we consider the optimization problem

$$P_<(t): \quad \min_x \ f(x,t) \qquad \text{s.t.} \qquad x \in \operatorname{cl}\Omega_<(t) := \operatorname{cl}\{x \in X \mid \omega(x,t) < 0\}$$

and denote its optimal value by $v_<(t)$. The following lemma uses similar arguments as [6, Lemma 2.2], and its assertion covers [6, Lemma 2.9], if we set $\varepsilon = \delta = 0$.

Lemma 2.1. *For all $\varepsilon, \delta \geq 0$ the identity*

$$\mathcal{R}_{\varepsilon,\delta} = \{(s,t) \in G_\delta \mid f(s,t) \leq v_<(t) + \varepsilon\} \tag{4}$$

holds.

Proof. Choose some $(s,t) \in \mathcal{R}_{\varepsilon,\delta}$, then by (3),

$$\forall x \in X: \qquad \max\{\omega(x,t), f(x,t) - f(s,t) + \varepsilon\} \geq 0.$$

If there exists some $x \in \Omega_<(t)$, then

$$f(s,t) \leq f(x,t) + \varepsilon$$

is true, and for all $x \in \operatorname{cl}\Omega_<(t)$ the above inequality holds by continuity of f in x. Altogether, we can state

$$f(s,t) \leq v_<(t) + \varepsilon.$$

Otherwise, we have $\Omega_<(t) = \emptyset$. Hence, we have $v_<(t) = +\infty$ and clearly $f(s,t) \leq v_<(t)+\varepsilon$.

Now choose some $(s,t) \notin \mathcal{R}_{\varepsilon,\delta}$. For $(s,t) \notin G_\delta$ we are done. Otherwise, there exists some $x \in X$ with $0 > \psi_\varepsilon(x,s,t)$. This implies $x \in \Omega_<(t)$ and

$$f(s,t) > f(x,t) + \varepsilon \geq v_<(t) + \varepsilon,$$

so that (4) is shown. $\square$

While the inequality $v(t) \leq v_<(t)$ holds for all $t \in \operatorname{dom}\Omega$, it may be strict for some values of t. Therefore, from (4) we obtain

$$\mathcal{S}_{\varepsilon,\delta} \subseteq \mathcal{R}_{\varepsilon,\delta}, \tag{5}$$

but identity may not hold without further assumptions. At least, any outer approximation of $\mathcal{R}_{\varepsilon,\delta}$ is also a valid outer approximation of $\mathcal{S}_{\varepsilon,\delta}$. However, the identity $\mathcal{S}_{\varepsilon,0} = \mathcal{R}_{\varepsilon,0}$ is true under the following additional assumption.

Assumption 2.2 (see also [6], Assumption 2.1)**.** *For all $t \in \operatorname{dom}\Omega$, at least one of the following two identities holds:*

- $v_<(t) = v(t)$,

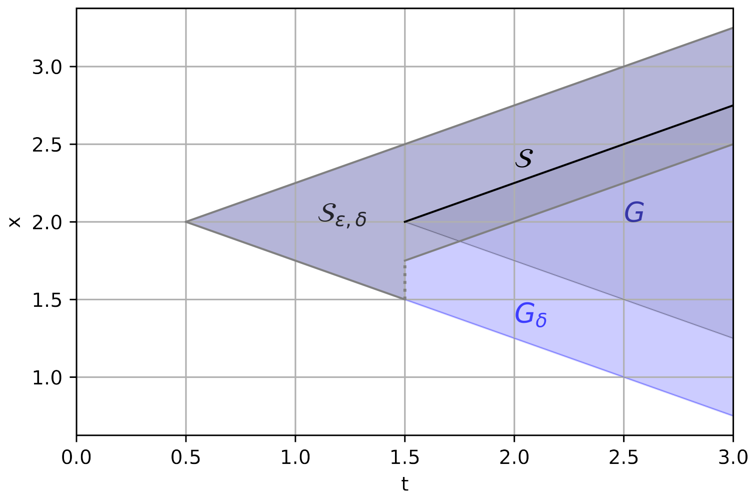

Figure 1: Relaxation of $\mathcal{S}$ in Example 2.4

- $S(t) = \Omega(t)$.

Under the above assumption, we see that the aforementioned property

$$S_{\varepsilon,\delta}(t) = \Omega_\delta(t) \text{ for all } t \notin \operatorname{dom}\Omega$$

is reasonable. At the boundary points of $\operatorname{dom}\Omega$ we already have $S(t) = \Omega(t)$ because $v(t) < v_<(t) = +\infty$. The following theorem is directly implied by the characterizations (1) and (4).

Theorem 2.3. *Assumption 2.2 implies*

$$\mathcal{R}_{\varepsilon,0} = \mathcal{S}_{\varepsilon,0}$$

for all $\varepsilon \geq 0$

Mild constraint qualifications that imply Assumption 2.2 as well as examples for

$$\mathcal{R}_{0,0} = \mathcal{R} \neq \mathcal{S} = \mathcal{S}_{0,0}$$

are provided in [6]. We emphasize that one cannot expect $\mathcal{R}_{\varepsilon,\delta} = \mathcal{S}_{\varepsilon,\delta}$ for $\delta > 0$ even under mild constraint qualifications. This can be seen in the following polyhedral example, where $\mathcal{S}_{\varepsilon,\delta}$ is not closed, and therefore $\mathcal{R}_{\varepsilon,\delta} \neq \mathcal{S}_{\varepsilon,\delta}$ holds.

Example 2.4. *Let* $n = r = 1$, $X = T = [0, 4]$, $f(x, t) := -2x$, *and* $\omega := \max\{\omega_1, \omega_2\}$, $\omega_1(x, t) := x - \frac{1}{2}t - \frac{5}{4}$, $\omega_2(x, t) := -x - \frac{1}{2}t + \frac{11}{4}$. *The sets* $\mathcal{S} = \mathcal{R}$ *and* $\mathcal{S}_{\varepsilon,\delta}$ *with* $\varepsilon = \delta = \frac{1}{2}$ *are displayed in Figure 1. The nonempty set difference* $\mathcal{R}_{\varepsilon,\delta} \setminus \mathcal{S}_{\varepsilon,\delta} = [\frac{3}{2}, \frac{7}{4}] \times \{\frac{3}{2}\}$ *is a boundary segment of* $\mathcal{R}_{\varepsilon,\delta}$ *that is not included in* $\mathcal{S}_{\varepsilon,\delta}$*, since*

$$\mathcal{S}_{\varepsilon,\delta}(\tfrac{3}{2}) = [\tfrac{7}{4}, \tfrac{5}{2}] \neq [\tfrac{3}{2}, \tfrac{5}{2}] = \Omega_\delta(\tfrac{3}{2}) = \mathcal{R}_{\varepsilon,\delta}(\tfrac{3}{2}).$$

However, in Example 2.4 we have at least $\mathcal{R}_{\varepsilon,\delta} = \operatorname{cl} S_{\varepsilon,\delta}$, and the effect vanishes for $\delta \leq \frac{1}{4}$. Further investigations on whether $\mathcal{R}_{\varepsilon,\delta} = S_{\varepsilon,\delta}$ holds for sufficiently small $\delta > 0$ or if generally $\mathcal{R}_{\varepsilon,\delta} = \operatorname{cl} S_{\varepsilon,\delta}$ holds will not be conducted here, but may be of interest.

In the remainder of the article, for the computational outer approximation of $\mathcal{S}$, we focus on the outer approximation of $\mathcal{R}$.

3 The branch-and-bound algorithm

This section presents the branch-and-bound algorithm for approximating optimal point mappings $\mathcal{S}$ and is structured as follows. First, Section 3.1 introduces the necessary foundation of convergent bounding procedures. Then, Section 3.2 defines the method and describes its high-level structure. Afterwards, the central building blocks, so-called discarding tests and inclusion tests are explained in detail in Sections 3.3 and 3.4. Finally, the finite termination is proven in Section 3.5.

3.1 Convergent bounding procedures

Let $B \subset \mathbb{R}^{n+r}$ be a compact box. For any sub-box $B' \subseteq B$, we use the following standard notations

$$B' \;=\; [\underline{b}', \overline{b}'], \quad \operatorname{diag}(B') \;=\; \|\overline{b}' - \underline{b}'\|_2.$$

To define the convergence of bounding procedures, we need to introduce the following notion of sequences of boxes.

Definition 3.1 (Exhaustive and weakly exhaustive sequences). *A sequence of sub-boxes (B^k) of B is called*

1. exhaustive *if $B^{k+1} \subseteq B^k$ for all $k \in \mathbb{N}$ and $\lim_k \operatorname{diag}(B^k) = 0$.*
2. weakly exhaustive *if there exists an exhaustive sequence (Z^k) such that $B^k \subseteq Z^k$ for all $k \in \mathbb{N}$.*

Next, we introduce lower and upper bounding procedures.

Definition 3.2 (Bounding procedures and convergence). *Let $\mathcal{F}$ be a set of continuous functions $f\colon B \to \mathbb{R}$.*

- *A* lower bounding procedure *for f is a map $\ell_f : \{$ sub-boxes of $B\} \to \mathbb{R}$ such that*

$$\ell_f(B') \;\leq\; \min_{x \in B'} f(x) \quad \textit{for all } B' \subseteq B \textit{ and all } f \in \mathcal{F}.$$

- *An* upper bounding procedure *for f is a map $u_f : \{$ sub-boxes of $B\} \to \mathbb{R}$ such that*

$$u_f(B') \;\geq\; \max_{x \in B'} f(x) \quad \textit{for all } B' \subseteq B \textit{ and all } f \in \mathcal{F}.$$

- *A lower (resp. upper) bounding procedure is called* convergent *if, for every weakly exhaustive sequence (B^k) and all $f \in \mathcal{F}$,*

$$\lim_k \ell_f(B^k) \;=\; \lim_k \min_{x \in B^k} f(x) \quad \Big(\textit{resp. } \lim_k u_f(B^k) = \lim_k \max_{x \in B^k} f(x)\Big).$$

We deliberately require convergence only over weakly exhaustive sequences. The following branch-and-bound construction does not produce perfectly nested boxes, but they can always be embedded in a truly nested (exhaustive) sequence. As long as $f \in \mathcal{F}$ is continuous, this weaker notion of exhaustiveness still guarantees that each bounding procedure converges to the common limit point of the boxes. The following result is taken from [7, Lem. 3.2], where its proof also can be found.

Lemma 3.3 (Convergence over weakly exhaustive sequences)**.** *Assume $f\colon B \to \mathbb{R}$ is continuous and let (B^k) be weakly exhaustive in the sense of Definition 3.1. Then there exists a unique $\widetilde{x} \in B$ with $x^k \to \widetilde{x}$ for any sequence of points $x^k \in B^k$, and every convergent lower bounding procedure ℓ_f and every convergent upper bounding procedure u_f satisfies*

$$\lim_{k\to\infty} \ell_f(B^k) = \lim_{k\to\infty} \min_{x\in B^k} f(x) = f(\widetilde{x}), \tag{6}$$

$$\lim_{k\to\infty} u_f(B^k) = \lim_{k\to\infty} \max_{x\in B^k} f(x) = f(\widetilde{x}). \tag{7}$$

All of the standard bounding-procedure constructions (interval arithmetic, αBB relaxations, McCormick relaxations, dual- and linearization-based bounds, etc.) can be implemented so as to converge on every weakly exhaustive sequence; see [7] for details.

3.2 Definition of P-ICGO

The algorithm for approximating the optimal point mapping is called parametric improvement function based complete global optimization (P-ICGO) and is stated below as Algorithm 1. Firstly, we describe its high-level structure. Then we explain its core functionalities, namely, discarding tests and inclusion tests, in detail, before we prove its finite convergence in Section 3.5.

From now on, we make the following assumption.

Assumption 3.4. *The sets X and T are bounded boxes.*

Note that the assumptions on the functions f and ω are not sharpened. In the algorithm we keep track of two lists, $\mathcal{W}$ and $\mathcal{O}$. The lists contain $(n+r)$-dimensional boxes that are subsets of the box $X \times T$. The working list $\mathcal{W}$ can be initialized with any list of disjoint boxes $B_1^0, \ldots, B_q^0$, such that

$$\mathcal{R} \subseteq \cup_{i=1}^q B_i^0 \subseteq X \times T, \tag{8}$$

or simply with one box $B^0 := X \times T$.

Since each box B in those lists consists of an 'x-component' $B_x \subseteq X$ and a 't-component' $B_t \subseteq T$, we denote $B := (B_x, B_t)$. Given the 't-component' B'_t of some box B', we also define the sub-lists

$$\mathcal{W}(B'_t) := \{B \in \mathcal{W} \mid B_t \cap B'_t \neq \emptyset\}$$

and

$$\mathcal{O}(B'_t) := \{B \in \mathcal{O} \mid B_t \cap B'_t \neq \emptyset\}$$

containing all boxes that overlap with B'_t in their t-component. From these sets, we compute special boxes (based on bounds on ψ or ψ_ε, respectively) to either

Algorithm 1: (P-ICGO) Parametric improvement function based complete global optimization

Input: $P(t)$, tolerances $\varepsilon \geq \delta > 0$, boxes $\mathcal{W} = \{B_1^0, \ldots, B_q^0\}$ with (8)
Output: List $\mathcal{O}$, such that (9)

Initialize list $\mathcal{O} = \{\}$;
while $\mathcal{W} \neq \emptyset$ **do**
  select an element $B' \in \mathcal{W}$;
  if $\ell_\omega(B') > 0$ **then**
    discard B';
  else
    find a box $\widehat{B} \in \operatorname{argmin}_{B \in \mathcal{W}(B'_t) \cup \mathcal{O}(B'_t)} \ell_\psi(B_x, B')$;
    select a point $\widehat{y} \in \widehat{B}_x$;
    if $u_\psi(\{\widehat{y}\}, B') < 0$ **then**
      discard B';
    else
      if $u_\omega(B') \leq \delta$ **then**
        find a box $\breve{B} \in \operatorname{argmin}_{B \in \mathcal{W}(B'_t) \cup \mathcal{O}(B'_t)} \ell_{\psi_\varepsilon}(B_x, B')$;
        if $\ell_{\psi_\varepsilon}(\breve{B}_x, B') \geq 0$ **then**
          remove B' from $\mathcal{W}$;
          append B' to $\mathcal{O}$;
        end
      end
      if $B' \in \mathcal{W}$ **then**
        remove B' from $\mathcal{W}$;
        append a partition of B' to $\mathcal{W}$;
        if $\widehat{B} \neq B'$ **then**
          remove $\widehat{B}$ from $\mathcal{W}$ or $\mathcal{O}$, respectively;
          append a partition of $\widehat{B}$ to $\mathcal{W}$ or $\mathcal{O}$, respectively;
        end
      end
    end
  end
end

- prove that, in the given box, no $s' \in B'_x$ is optimal, given any parameter $t' \in B'_t$, or
- prove that any $s' \in B'_x$ is ε-optimal and δ-feasible given any $t' \in B'_t$.

Together, we terminate if the working list $\mathcal{W}$ is empty and the output list $\mathcal{O}$ satisfies the following approximation property

$$\mathcal{R} \subseteq \cup_{B \in \mathcal{O}} B \subseteq \mathcal{R}_{\varepsilon,\delta}. \tag{9}$$

To achieve this, we firstly discard all boxes $B' \in \mathcal{W}$ that can be proven to not be a part of $\mathcal{R}$, i.e. $B' \cap \mathcal{R} = \emptyset$. And, secondly, we move all boxes $B' \in \mathcal{W}$, for which $B' \subseteq \mathcal{R}_{\varepsilon,\delta}$ can be proven, from $\mathcal{W}$ to $\mathcal{O}$. In the following sections, we will see how both properties are verified by P-ICGO and prove the finite termination.

3.3 Discarding tests

For a set A, we denote its complement by A^c. The algorithm may remove all sub-boxes B' of B^0 that are included in the set complement $\mathcal{R}^c$ of $\mathcal{R}$. Based on the description (2), we may decompose it as follows

$$\mathcal{R}^c = \{(s,t) \in \mathbb{R}^{n+r} \mid \omega(s,t) > 0\} \ \cup \ \{(s,t) \in \mathbb{R}^{n+r} \mid \min_{x \in X} \psi(x,s) < 0\}$$

Then $B' \subseteq \mathcal{R}^c$ is implied by either of the two conditions

$$\min_{(s,t) \in B'} \omega(s,t) > 0, \tag{10}$$

$$\max_{(s,t) \in B'} \min_{x \in X} \psi(x,s,t) < 0. \tag{11}$$

For (10), we employ a convergent lower bounding procedure ℓ_ω and test $\ell_\omega(B') > 0$ in line 4.

On the other hand, for (11), we proceed analogously to [7] and firstly observe that, for any point $\widehat{y} \in X$, the term

$$u_\psi(\{\widehat{y}\}, B') := \max\{u_\omega(\{\widehat{y}\}, B'_t), u_f(\{\widehat{y}\}, B'_t) - \ell_f(B')\}$$

satisfies

$$\begin{aligned}
u_\psi(\{\widehat{y}\}, B') &\geq \max\{\max_{t \in B'_t} \omega(\widehat{y},t), \max_{t \in B'_t} f(\widehat{y},t) - \min_{(s,t) \in B'} f(s,t)\} \\
&\geq \max\{\max_{t \in B'_t} \omega(\widehat{y},t), \max_{(s,t) \in B'} f(\widehat{y},t) - f(s,t)\} \\
&\geq \max_{(s,t) \in B'} \max\{\omega(\widehat{y},t), f(\widehat{y},t) - f(s,t)\} \\
&= \max_{(s,t) \in B'} \psi(\widehat{y},s,t) \\
&\geq \max_{(s,t) \in B'} \min_{x \in X} \psi(x,s,t).
\end{aligned}$$

Therefore $u_\psi(\{\widehat{y}\}, B') < 0$ implies (11) and is a valid discarding test for B'. It ensures that the improvement function is negative for the singleton $\{\widehat{y}\}$, given any $(s,t) \in B'$.

We propose to find such points deterministically by considering the boxes from $\mathcal{W}$ and $\mathcal{O}$ that yield the smallest lower bound for the improvement function, and overlap with B' in their 't-component'. This amounts to computing

$$\ell_\psi(\widehat{B}_x, B') := \max\{\ell_\omega(\widehat{B}_x, B'_t), \ell_f(\widehat{B}_x, B'_t) - u_f(B')\}$$

for all $\widehat{B}_x$ such that a corresponding box $\widehat{B} = (\widehat{B}_x, \widehat{B}_t)$ is contained in one of the lists $\mathcal{W}(B'_t)$ or $\mathcal{O}(B'_t)$. Hence, we first compute

$$\widehat{B} \in \operatorname{argmin}_{B \in \mathcal{W}(B'_t) \cup \mathcal{O}(B'_t)} \ell_\psi(B_x, B')$$

in line 7 and then try to compute some $\widehat{y} \in \widehat{B}_x$ with (11). If this does not work and a partition of B' persists in $\mathcal{W}$, we have to refine $\widehat{B}$ to compute an appropriate $\widehat{y}$ after finitely many iterations (whenever it exists). In practice, this works by simply selecting any $\widehat{y} \in \widehat{B}_x$ in line 8, as long as the following assumption on the partitioning of a box in lines 21 and 24 holds.

Assumption 3.5. *Partitioning a box B into $B^1, \ldots, B^k$ yields for at least one longest edge l of B and all $i \in \{1, \ldots, k\}$:*

a) $|\overline{b}^i_l - \underline{b}^i_l| < |\overline{b}_l - \underline{b}_l|$,

b) $\operatorname{int} B^i \neq \emptyset$,

c) $\operatorname{int} B^i \cap \operatorname{int} B^j = \emptyset,\ \forall j \in \{1, \ldots, k\} \setminus \{i\}$.

Given that assumption, we are able to prove a central property of the algorithm, to which we come back in the next section.

Lemma 3.6. *Let $\widetilde{t} \in \operatorname{dom} \Omega_<$ and let $\widetilde{y}$ be an optimal point of $P_<(\widetilde{t})$. Then, in every iteration of P-ICGO, there exists a box $B \in \mathcal{W} \cup \mathcal{O}$, such that $(\widetilde{y}, \widetilde{t}) \in B$ and $B \cap \Omega_< \neq \emptyset$.*

Proof. We clearly have $f(\widetilde{y}, \widetilde{t}) = v_<(t)$ and thus $(\widetilde{y}, \widetilde{t}) \in \mathcal{R}$ by (4). As shown in this section, no box containing an element of $\mathcal{R}$ is discarded by P-ICGO. Thus, all boxes containing $(\widetilde{y}, \widetilde{t})$ are in one of the lists $\mathcal{W}$ or $\mathcal{O}$. Moreover, according to Assumption 3.5, every box has a nonempty interior. Together with $\widetilde{y} \in \operatorname{cl} \Omega_<(\widetilde{t})$, we conclude that there must exist at least one box such that $(\widetilde{y}, \widetilde{t}) \in B$ and $B_x \cap \Omega_<(\widetilde{t}) \neq \emptyset$ is satisfied. $\square$

3.4 Inclusion tests

To prove the inclusion of some box $B' \subseteq X \times T$ inside the relaxed set

$$\mathcal{R}_{\varepsilon,\delta} := \{(s,t) \in G_\delta \mid f(s,t) \leq v_<(t) + \varepsilon\}$$

the properties

$$\max_{(s,t) \in B'} \omega(s,t) \leq \delta \tag{12}$$

and

$$f(s,t) \leq v_<(t) + \varepsilon\,,\ \forall (s,t) \in B' \tag{13}$$

must hold. The first property is verified by evaluating the upper bound $u_\omega(B')$ in line 12 of P-ICGO. For the second property we use the lower bound on ψ_ε over the box B' and the 'x-part' of other non-discarded boxes B_x. The bound is defined as

$$\ell_{\psi_\varepsilon}(B_x, B') := \max\{\ell_\omega(B_x, B'_t), \ell_f(B_x, B'_t) - u_f(B') + \varepsilon\}.$$

The following lemma shows how these bounds can be employed to verify the property (13) in line 14 of P-ICGO.

Lemma 3.7. *In P-ICGO, for some box B', the bound*

$$0 \le \min_{B \in \mathcal{W}(B'_t) \cup \mathcal{O}(B'_t)} \ell_{\psi_\varepsilon}(B_x, B')$$

implies (13).

Proof. Firstly, we obtain for all $B \in \mathcal{W}(B'_t) \cup \mathcal{O}(B'_t)$:

$$\begin{aligned} 0 &\le \ell_{\psi_\varepsilon}(B_x, B') = \max\{\ell_\omega(B_x, B'_t), \ell_f(B_x, B'_t) - u_f(B') + \varepsilon\} \\ &\le \max\{\min_{y \in B_x} \min_{t \in B'_t} \omega(y,t), \min_{y \in B_x} \min_{t \in B'_t} f(y,t) - \max_{(s,t) \in B'} f(s,t) + \varepsilon\}. \end{aligned}$$

We now prove (13) considering all $t \in B'_t$ in the following cases.

Case 1: For any $t^0 \in B'_t$ such that $\Omega_<(t^0) = \emptyset$, we have (by convention) $v_<(t^0) = +\infty$ and (13) fulfilled for all $s \in B'_x$, given t^0.

Case 2: For any $t^\star \in B'_t$, such that $\Omega_<(t^\star) \neq \emptyset$, we may choose an optimal point $y^\star$ of $P_<(t^\star)$. Then, by Lemma 3.6, there exists a box $B^\star \in \mathcal{W} \cup \mathcal{O}$ containing $(y^\star, t^\star)$ with $B^\star_x \cap \Omega_<(t^\star) \neq \emptyset$. Since $t^\star \in B'_t$, we also have $B^\star \in \mathcal{W}(B'_t) \cup \mathcal{O}(B'_t)$. For this box,

$$\min_{y \in B^\star_x} \min_{t \in B'_t} \omega(y,t) < 0$$

and thus, the second term in the above maximum has to be nonnegative, i.e.

$$\begin{aligned} 0 &\le \min_{y \in B^\star_x} \min_{t \in B'_t} f(y,t) - \max_{(s,t) \in B'} f(s,t) + \varepsilon \\ &\le \min_{y \in B^\star_x} f(y, t^\star) - \max_{s \in B'_x} f(s, t^\star) + \varepsilon \\ &\le v_<(t^\star) - \max_{s \in B'_x} f(s, t^\star) + \varepsilon. \end{aligned}$$

We may rewrite this as

$$f(s, t^\star) \le v_<(t^\star) + \varepsilon, \ \forall s \in B'_x,$$

which corresponds to (13) for $t = t^\star$. Since $t^\star$ was arbitrarily chosen from B'_t, (13) is proven. □

If the statements on line 12 and 14 hold true, P-ICGO moves the box B' to the output list $\mathcal{O}$. For the sake of efficiency, the second test is only computed, if the first held true.

3.5 Convergence of P-ICGO

Theorem 3.8. *Let the bounding procedures $\ell_\omega, u_\omega, \ell_f, u_f$ be convergent, let Assumption 2.2 hold and let the partitioning of the boxes satisfy Assumption 3.5. Then P-ICGO terminates after finitely many iterations.*

Proof. We will assume that the assertion is wrong and derive a contradiction.

Part 1: Establish a weakly exhaustive sequence (B^k) with limit point $(\widetilde{x}, \widetilde{t})$.

Assume that the assertion is wrong, then $\mathcal{W}$ never becomes empty and, in infinitely many iterations, an element $B' \in \mathcal{W}$ can be selected. Inside the while-loop, a selected element is discarded (i.e. removed) from $\mathcal{W}$ or further partitioned into sub-boxes that are appended to $\mathcal{W}$. An infinite number of iterations is only possible, if the box B' is further partitioned in infinitely many iterations. Otherwise, $\mathcal{W}$ would become empty at some point. Thus, we may consider an infinite sequence of boxes (B^k), whose elements are neither discarded in lines 4 to 10, nor moved from $\mathcal{W}$ to $\mathcal{O}$ in line 15f. Consequently, for each element of the sequence (B^k) the following inequalities hold

$$\ell_\omega(B^k) \leq 0, \tag{14}$$

$$u_\psi(\{\widehat{y}^k\}, B^k) \geq 0, \tag{15}$$

where $\widehat{y}^k$ denotes the point '$\widehat{y}$' selected in line 8 in correspondence to B^k, and

$$u_\omega(B^k) > \delta \qquad \text{or} \qquad \ell_{\psi_\varepsilon}(\breve{B}_x^k, B^k) < 0, \tag{16}$$

where $\breve{B}_x^k$ denotes the sub-box of $\breve{B}$ computed in line 13 in correspondence to B^k. If one of the statements does not hold for any element of (B^k), this contradicts the generation of this infinite sequence of boxes. Since each element of (B^k) is removed from $\mathcal{W}$, before a partition of it is appended to $\mathcal{W}$, the elements of (B^k) are pairwise distinct. Therefore, Lemma A.1 implies that we may pass to a weakly exhaustive subsequence of (B^k) and, without loss of generality, assume that (B^k) itself is such a sequence. This means that there exists a point $(\widetilde{x}, \widetilde{t})$, such that for any sequence with $(x^k, t^k) \in B^k$ for all $k \in \mathbb{N}$, we have $\lim_k (x^k, t^k) = (\widetilde{x}, \widetilde{t})$.

Part 2: Show that $\ell_{\psi_\varepsilon}(\breve{B}_x^k, B^k) < 0$ holds in infinitely many iterations.

In the following, we focus on (16) and only come back to (14) and (15) if necessary. In particular, at least one of the expressions in (16) holds infinitely often. Assume that this is the case for $u_\omega(B^k) > \delta$. Then, by convergence of the bounding procedures, for k sufficiently large we have $\ell_w(B^k) > 0$ which contradicts (14). Thus $\ell_{\psi_\varepsilon}(\breve{B}_x^k, B^k) < 0$ must hold in infinitely many iterations.

Part 3: Construct a lower bound on $f(\widetilde{x}, \widetilde{t})$ and show $v(\widetilde{t}) = v_<(\widetilde{t}) < +\infty$

We first consider the infinite sequence $(\widehat{B}^k)$ produced in line 7 of the algorithm in correspondence to (B^k). Since the boxes $\widehat{B}^k$ also get refined in the corresponding iterations, by Lemma A.1 the sequence $(\widehat{B}^k)$ is weakly exhaustive, after possibly passing to a subsequence. Moreover, for the sequence of points $\widehat{y}^k \in \widehat{B}_x^k$, Lemma 3.3 implies in particular $\lim_k \widehat{y}^k = \widehat{y}$ with some $\widehat{y} \in X$. Since (15) must hold for all elements of (X^k), we obtain

$$0 \leq u_\psi(\{\widehat{y}^k\}, B^k) = \max\{u_\omega(\{\widehat{y}^k\}, B_t^k),\ u_f(\{\widehat{y}^k\}, B_t^k) - \ell_f(B^k)\} \tag{17}$$

and the convergence of the bounding procedures implies

$$0 \leq \lim_k u_\psi(\{\widehat{y}^k\}, B^k) = \max\{\omega(\widehat{y}, \widetilde{t}), f(\widehat{y}, \widetilde{t}) - f(\widetilde{x}, \widetilde{t}))\} = \psi(\widehat{y}, \widetilde{x}, \widetilde{t}). \tag{18}$$

We consider the sequence $\breve{B}^k$ generated in line 13. As shown in part 2, we may pass to an infinite subsequence such that

$$\ell_{\psi_\varepsilon}(\breve{B}_x^k, B^k) < 0 \tag{19}$$

for all $k \in \mathbb{N}$. By $\varepsilon > 0$, we obtain

$$\min_{B \in \mathcal{W}(B_t^k) \cup \mathcal{O}(B_t^k)} \ell_\psi(B_x, B^k) = \ell_\psi(\widehat{B}_x^k, B^k) \leq \ell_\psi(\breve{B}_x^k, B^k) \leq \ell_{\psi_\varepsilon}(\breve{B}_x^k, B^k) < 0.$$

The convergence of the bounding procedures and (18) imply

$$\lim_k \min_{B \in \mathcal{W}(B_t^k) \cup \mathcal{O}(B_t^k)} \ell_\psi(B_x, B^k) = \lim_k \ell_\psi(\widehat{B}_x^k, B^k) = \psi(\widehat{y}, \widetilde{x}, \widetilde{t}) \geq 0.$$

Together, the sandwich theorem yields

$$\lim_k \ell_{\psi_\varepsilon}(\breve{B}_x^k, B^k) = 0, \tag{20}$$

which implies that $(\breve{B}_x^k)$ also contains a weakly exhaustive sequence of boxes. To see this, note that (19) holds for all $k \in \mathbb{N}$. This value can only converge to zero for infinitely many pairwise distinct boxes $(\breve{B}^k)$. Together with Lemma A.1 follows the weak exhaustiveness of $(\breve{B}^k)$, after possibly passing to a subsequence. Furthermore, with (6) and (7), (20) reads as

$$\begin{aligned} 0 &= \max\{\lim_k \ell_\omega(\breve{B}_x^k, B_t^k), \lim_k \ell_f(\breve{B}_x^k, B_t^k) - \lim_k u_f(B^k) + \varepsilon\} \\ &= \max\{\omega(\breve{y}, \widetilde{t}), f(\breve{y}, \widetilde{t}) - f(\widetilde{x}, \widetilde{t}) + \varepsilon\} \end{aligned}$$

with $\breve{y} \in X$ (see again Lemma 3.3). Due to $\omega(\breve{y}, \widetilde{t}) \leq 0$, we have $\Omega(\widetilde{t}) \neq \emptyset$ and immediately obtain $f(\breve{y}, \widetilde{t}) \geq v(\widetilde{t})$ as well as $v(\widetilde{t}) < +\infty$. Therefore, the following lower bound applies

$$f(\widetilde{x}, \widetilde{t}) \geq v(\widetilde{t}) + \varepsilon. \tag{21}$$

Moreover, from (14), we obtain

$$0 \geq \lim_k \ell_\omega(B^k) = \omega(\widetilde{x}, \widetilde{t})$$

and thus $\widetilde{x} \in \Omega(\widetilde{t})$. Then, (21) and $\varepsilon > 0$ imply $S(\widetilde{t}) \neq \Omega(\widetilde{t})$ and Assumption 2.2 yields

$$v_<(\widetilde{t}) = v(\widetilde{t}) < +\infty.$$

Part 4: Construct an upper bound on $f(\widetilde{x}, \widetilde{t})$ and derive the contradiction.

In the following, we again consider the weakly exhaustive sequence $\widehat{B}^k$ and the 'x-component' $\widehat{y} \in X$ of its limit point. Now let $y_<$ be a solution of $P_<(\widetilde{t})$, which means $f(y_<, \widetilde{t}) = v_<(\widetilde{t})$ and $y_< \in \operatorname{cl} \Omega_<(\widetilde{t})$. By Lemma 3.6, in every iteration of P-ICGO, there exists a box B^ρ in $\mathcal{W}(B_t^k)$ or $\mathcal{O}(B_t^k)$ such that $(y_<, \widetilde{t}) \in B^\rho$ and $B_x^\rho \cap \Omega_<(\widetilde{t}) \neq \emptyset$. For any given tolerance ρ, in every iteration, we may choose some $y^\rho \in \Omega_<(\widetilde{t}) \cap B_x^\rho$ near $y_<$ such that $f(y^\rho, \widetilde{t}) \leq v_<(\widetilde{t}) + \rho$. Subsequently, We will show that the relation

$$\psi(\widehat{y}, \widetilde{x}, \widetilde{t}) \leq \psi(y^\rho, \widetilde{x}, \widetilde{t}) \tag{22}$$

holds. Assume, for the sake of contradiction, $\psi(\widehat{y}, \widetilde{x}, \widetilde{t}) > \psi(y^\rho, \widetilde{x}, \widetilde{t})$. The convergence of the bounding procedures yields

$$\ell_\psi(\widehat{B}_x^k, B^k) = \max\{\ell_\omega(\widehat{B}_x^k, B_t^k), \ell_f(\widehat{B}_x^k, B_t^k) - u_f(B^k)\} > \psi(y^\rho, \widetilde{x}, \widetilde{t}) \tag{23}$$

for sufficiently large k. For B^ρ we have

$$\psi(y^\rho, \widetilde{x}, \widetilde{t}) \geq \max\{\ell_\omega(B_x^\rho, B_t^k), \ell_f(B_x^\rho, B_t^k) - u_f(B^k)\} = \ell_\psi(B_x^\rho, B^k).$$

Together with (23), we obtain a contradiction since, for sufficiently large k, the box $\widehat{B}^k$ would not have been chosen in line 7, which proves (22).

Consequently, together with (18), we can state

$$0 \leq \psi(\widehat{y}, \widetilde{x}, \widetilde{t}) \leq \psi(y^\rho, \widetilde{x}, \widetilde{t}) = \max\{\omega(y^\rho, \widetilde{t}), f(y^\rho, \widetilde{t}) - f(\widetilde{x}, \widetilde{t})\}. \tag{24}$$

From $\omega(y^\rho, \widetilde{t}) < 0$ we obtain

$$f(\widetilde{x}, \widetilde{t}) \leq f(y^\rho, \widetilde{t}) \leq v_<(\widetilde{t}) + \rho$$

and, since $\rho > 0$ can be chosen arbitrarily small, the upper bound is

$$f(\widetilde{x}, \widetilde{t}) \leq v_<(\widetilde{t}) = v(\widetilde{t}), \tag{25}$$

where the last identity has been shown in Part 3. Finally, the upper bound (25) contradicts the lower bound (21). $\square$

4 Numerical tests

This section is devoted to numerical experiments with P-ICGO. Firstly, in Section 4.1, we consider small examples with visualizations to illustrate the output and discuss the effect of specific degeneracies on its functionality. Secondly, in Section 4.2, we run the algorithm on a larger set of instances from a library to discuss its performance and verify its output.

We implemented P-ICGO in Julia due to its native support for interval arithmetic via the IntervalArithmetic.jl package, which guarantees reliable bounds and correct rounding. In addition, Julia's multiple dispatch lets us define a single function definition that is automatically optimized at compile time for different argument types, for example, interval vectors or real vectors. Coupled with competitive runtime performance and efficient memory representation through the StaticArrays.jl package, Julia is a compelling choice for our purpose. The complete code and the scripts necessary to reproduce the experiments are available on GitHub [5].

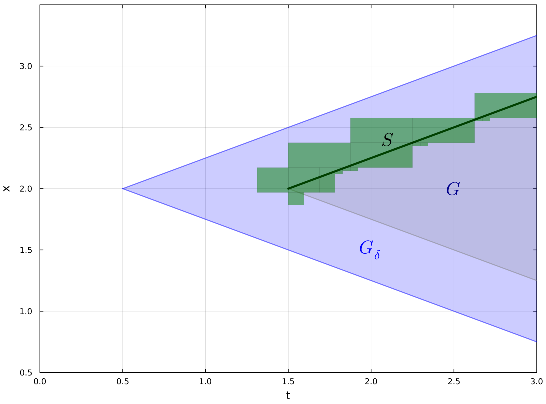

Figure 2: Output of P-ICGO for Example 4.1, $\varepsilon = \delta = 0.5$.

All experiments were conducted on a server running Linux Mint 22.3, equipped with an Intel Core i7-9700K processor (8 physical cores @ 3.60 GHz) and 31 GB of RAM. Julia (version 1.12.6) was executed in default single-threaded configuration.

Convergent upper and lower bounds on functions were calculated by interval-valued evaluation of the functions themselves, therefore interpreting all operations $+, -, \cdot, \ldots$ interval-valued. Using the IntervalArithmetic.jl package, e.g., the bound $\ell_f(\widehat{B}_x, B'_t)$ can be obtained by directly evaluating $f(\widehat{B}_x, B'_t)$ with the interval vectors defined by the boxes and then taking the lower bound of the resulting interval. This bounding procedure is very fast but leads to relatively coarse bounds.

In the following, we will analyze the output of P-ICGO. For convenience, we also save the list $\mathcal{O}^I$ of all initial elements that were moved to $\mathcal{O}$ (without further refinement steps). Clearly, we have

$$\bigcup_{B \in \mathcal{O}^I} B = \bigcup_{B \in \mathcal{O}} B.$$

For the plots in Section 4.1, it is convenient to use the description $\mathcal{O}^I$, which contains fewer elements. For calculating bounds on functions over the output boxes, as in Section 4.2, the refined list $\mathcal{O}$ is favorable.

4.1 Visual examples

In this section, we illustrate the output of five small parametric optimization problems with $n \in \{1, 2\}$ and $r = 1$. Since Assumption 2.2 is violated for some of these examples, we introduced an additional stopping criterion. Boxes with a width below $w_{\min}$ are not selected again in line 3. Thus, the execution of P-ICGO stops as soon as all the boxes in $\mathcal{W}$ are smaller than $w_{\min}$. We use the term ‘terminates’ for P-ICGO only if the defined criterion in line 2 is met, i.e., if $\mathcal{W}$ is empty. If, in one of the following examples, no parameter $w_{\min}$ is mentioned and we report that P-ICGO *terminated*, that means that $\mathcal{W}$ is empty.

Furthermore, we will always initialize $\mathcal{W} = \{B^0\}$ with $B^0 = X \times T$.

Example 4.1 (Ex. 2.4 cont.)**.** *We consider the optimization problem from Example 2.4 with $X = [-2.5, 4]$, $T = [0, 3]$, and execute P-ICGO with $\varepsilon = \delta = \frac{1}{2}$.*

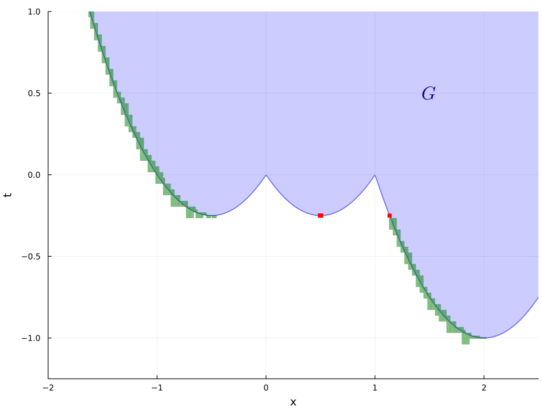

Figure 3: Output of P-ICGO for Example 4.2, $\varepsilon = \delta = 0.1$, $w_{\min} = 10^{-4}$.

For Example 4.1, P-ICGO terminates after 182 iterations with 18 boxes in $\mathcal{O}^I$ and 28 boxes in $\mathcal{O}$. These boxes are plotted in green in Figure 2. Since the relaxation $\mathcal{S}_{\varepsilon,\delta}$ was already determined analytically and is plotted in Figure 1, we can easily testify the correctness of the approximation.

The following two optimization problems with specific degeneracies are based on [6, Ex. 2.1 & 2.2].

Example 4.2. *Let $n = r = 1$, $X = [-2, 2.5]$, $T = [-\frac{5}{4}, 1]$,*

$$f(x,t) = x, \quad \omega(x,t) := \min\{a_1(x), a_2(x), a_3(x)\} - t,$$

with

$$a_1(x) := (x-2)^2 - 1,\ a_2(x) := (x - \tfrac{1}{2})^2 - \tfrac{1}{4},\ a_3(x) := (x + \tfrac{1}{2})^2 - \tfrac{1}{4}.$$

The feasible set is visualized in Figure 3. One can infer that the objective becomes minimal at the left-most feasible point. At $t = -\frac{1}{4}$, there is a degeneracy with

$$S(-\tfrac{1}{4}) = \{-\tfrac{1}{2}\} \neq \{-\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{5}{4}\} = \mathcal{R}(-\tfrac{1}{4}).$$

and by $v_<(-\frac{1}{4}) = \frac{5}{4} \neq -\frac{1}{2} = v(-\frac{1}{4})$, Assumption 2.2 is also violated. We execute P-ICGO with $\varepsilon = \delta = 0.1$ and $w_{\min} = 10^{-4}$.

For Example 4.2, P-ICGO stops after 2,725 iterations with 409 boxes smaller than $w_{\min}$ remaining in the list $\mathcal{W}$, 270 boxes in $\mathcal{O}$ and 123 boxes in $\mathcal{O}^I$. The boxes from $\mathcal{W}$ are plotted in red (as rectangular scatters; otherwise, they would not be visible due to their size). They are all located around $(\frac{1}{2}, -\frac{1}{4})$ and $(\frac{5}{4}, -\frac{1}{4})$. These points lie, as written above, in $\mathcal{R} \setminus \mathcal{S}$ and are therefore expected to be parts of the algorithm's approximation (see (9)). All other parts of the graph $\mathcal{R}$ are also included in $\mathcal{S}$ and are accurately approximated by 162 boxes from $\mathcal{O}^I$ (plotted in green).

Without the specification of $w_{\min}$, P-ICGO would fail to stop here. We remark that it is not possible to determine a clear threshold for $w_{\min}$ such that only boxes around the degeneracies remain in $\mathcal{W}$. Alternatively, one could stop P-ICGO after a predefined computing time. Then one could detect that all boxes that remained in $\mathcal{W}$ are located near $t = -\frac{1}{4}$, which suggests that there is a degeneracy.

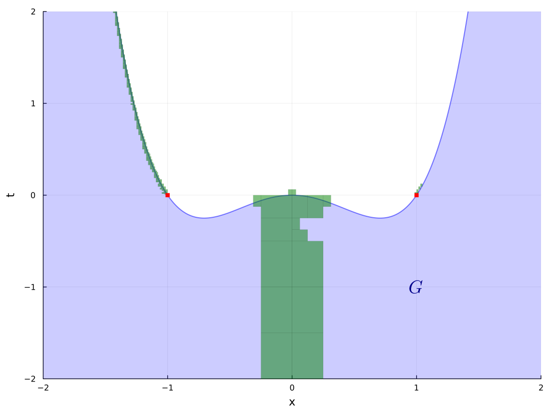

Figure 4: Output of P-ICGO for Example 4.3, $\varepsilon = \delta = 0.1$, $w_{\min} = 10^{-4}$.

Example 4.3. *Let $n = r = 1$, $X = T = [-2, 2]$,*

$$f(x, t) := x^2 - \max\{0, -1 - x\}, \quad \omega(x, t) := t - x^4 + x^2.$$

The feasible set is visualized in Figure 4. One can infer that $S(t) = \{0\}$ for $t \in [-2, 0]$ and for $t > 0$, the optimal x is as close as possible to 0 on the left side of the disconnected feasible set. For $t = 0$, there is a degeneracy with

$$S(0) = \{0\} \neq \{-1, 0, 1\} = \mathcal{R}(0)$$

and $v_<(0) = 1 \neq 0 = v(0)$, hence Assumption 2.2 is violated. We execute P-ICGO with $\varepsilon = \delta = 0.1$ and $w_{\min} = 10^{-4}$.

For Example 4.3, P-ICGO stops after 2,323 iterations with 41 boxes in $\mathcal{W}$, 493 boxes in $\mathcal{O}$ and 342 boxes in $\mathcal{O}^I$. These boxes are again marked in red and are located around the points $(-1, 0)$ and $(1, 0)$ from $\mathcal{R} \setminus \mathcal{S}$. The other areas of $\mathcal{S}$ are accurately approximated by 461 boxes of $\mathcal{O}^I$ (plotted in green).

One can observe that the larger gradient of f for $x < -1$ (compared to $x \approx 0$) results in smaller boxes in $\mathcal{O}^I$ necessary to meet the specified tolerances and be a subset of $\mathcal{R}_{\varepsilon,\delta}$.

Example 4.4 (see [9], Ex. 1.9.4)**.** *Let $n = 2$, $r = 1$, $X = [-2, 0] \times [0, 2]$, $T = [0, 2\pi]$,*

$$f(x, t) := \cos(t) \cdot x_1 + \sin(t) \cdot x_2, \quad \omega(x, t) := x_2 - x_1 - 1.$$

The graph of the optimal point mapping $\mathcal{S}$ is visualized in Figure 5 (left). We execute P-ICGO with $\varepsilon = \delta = 0.1$ and $w_{\min} = 10^{-3}$.

For Example 4.4, P-ICGO terminates after 26,941 iterations with an empty list $\mathcal{W}$, 3,742 boxes in $\mathcal{O}$ and 2,650 boxes in $\mathcal{O}^I$. The output is displayed in Figure 5 next to the graph $\mathcal{S}$. In particular, we can see that for $t \in \{\frac{\pi}{2}, \pi, \frac{7}{4}\pi\}$ the optimal point $S(t)$ is not unique and, in each case, a line through the entire cube has to be approximated.

Example 4.5 (see [9], Ex. 1.9.5)**.** *Let $n = 2$, $r = 1$, $X = [-2, 0] \times [0, 2]$, $T = [0, 2\pi]$,*

$$f(x, t) := x_1, \quad \omega(x, t) := \max\{a_1(x), a_2(x, t)\}$$

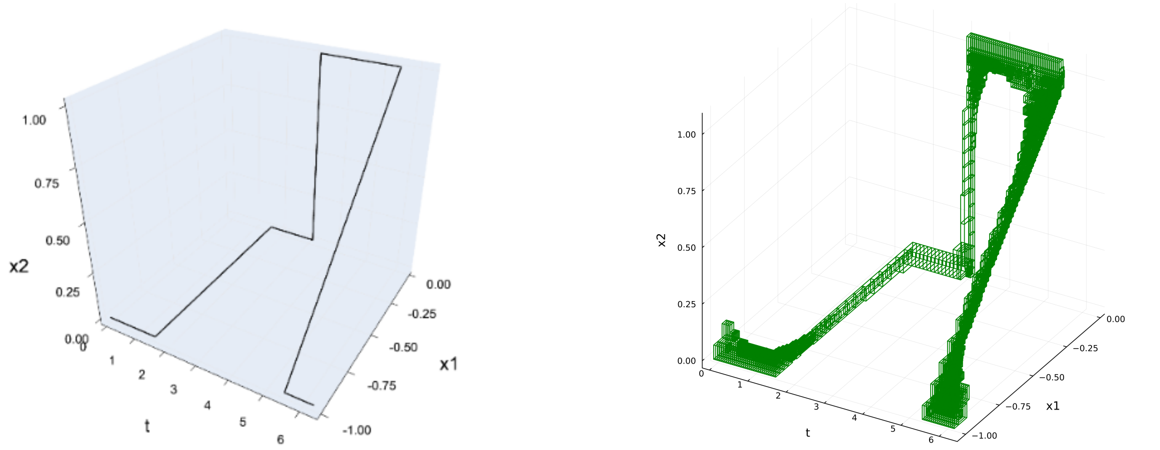

Figure 5: Plot of $\mathcal{S}$ for Example 4.4 (l) and the corresponding output of P-ICGO for $\varepsilon = \delta = 0.1$, $w_{\min} = 10^{-3}$.

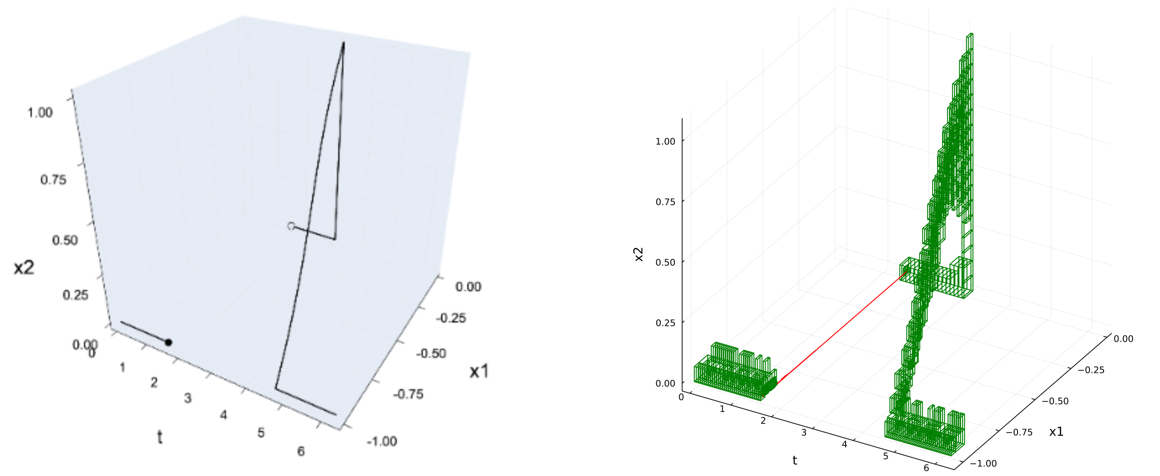

Figure 6: Plot of $\mathcal{S}$ for Example 4.5 (l) and the corresponding output of P-ICGO for $\varepsilon = \delta = 0.1$, $w_{\min} = 10^{-3}$.

with

$$a_1(x) := x_2 - x_1 - 1, \quad a_2(x,t) := \cos(t) \cdot x_1 + \sin(t) \cdot x_2.$$

The graph of the optimal point mapping $\mathcal{S}$ is visualized in Figure 6 (left). For $t = \frac{\pi}{2}$, there is a degeneracy with

$$\mathcal{S}(\tfrac{\pi}{2}) = \{(-1,0)^\intercal\} \neq \{[-1,0] \times \{0\}\} = \mathcal{R}(\tfrac{\pi}{2}).$$

We execute P-ICGO with $\varepsilon = \delta = 0.1$ and $w_{\min} = 10^{-3}$

For Example 4.5, P-ICGO stops after 19,459 iterations with 4,396 boxes in $\mathcal{W}$, 954 boxes in $\mathcal{O}$, and 497 boxes in $\mathcal{O}^I$. The output is displayed in Figure 6; the boxes from the approximation $\mathcal{O}^I$ are plotted in green and the boxes from $\mathcal{W}$ are plotted in red. The latter only occur around the degeneracy at $t = \frac{\pi}{2} \approx 1.571$ and approximate the line segment $\mathcal{R}(\frac{\pi}{2}) \setminus \mathcal{S}(\frac{\pi}{2})$ stated above. Indeed, the 't-component' of all boxes from $\mathcal{W}$ is included in the interval $[1.56926, 1.58461]$.

To conclude, we observed that degeneracies prohibit the termination of P-ICGO, but their effect is limited to the parameters at which they occur. Finally, we notice that the outer approximation of $\mathcal{S}$ is of high computational effort when the geometry of $\mathcal{S}$ is itself complex or large, e.g., due to non-unique optimal points.

4.2 Feasible sets of optimistic bilevel optimization problems

Definition and setup

In this section, we compute graphs of optimal point mappings of the follower's optimization problem in bilevel programs. In the following, we briefly recap the terminology required here; for an extensive introduction to bilevel optimization, we refer to [1, 2]. The considered programs from the bilevel optimization library (BOLIB) [10] take the form

$$\begin{aligned} BL: \qquad \min_{x,y} \quad F(x,y) \quad \text{s.t.} \quad & G(x,y) \leq 0 \\ & y \in S(x) := \operatorname{argmin}_y \{f(x,y) \mid g(x,y) \leq 0\}. \end{aligned}$$

The vector of variables x is assigned to the *leader*, who aims to optimize the function F. Hereby, the leader has to anticipate the reaction of the *follower*, who is in control of the variables y and optimizes her objective function f under constraints g, based on the leader's decision x, by solving the parametric lower-level problem

$$Q(x): \qquad \min_y f(x,y) \quad \text{s.t.} \quad g(x,y) \leq 0.$$

This relation is already captured by the above formulation. From the leader's point of view, in addition to the leader's inequality constraint $G(x,y) \leq 0$, for a given x, the vector y is constrained to the optimal point set $S(x)$ of the follower. We note that the above model corresponds to the so-called optimistic bilevel formulation, in which the leader may select a vector y from the followers' optimal point set that minimizes her objective. Therefore, it is beneficial to know *all* optimal points from the set $S(x)$ for specific choices of x.

The lower-level problem $Q(x)$ almost possesses the format of the parametric problem $P(t)$ from Section 2.1, where x plays the role of the parameter t and y that of the decision variable x. The only difference is the lack of compact sets X, Y with $x \in X$ and $y \in Y$, analogous to X and T for $P(t)$. With $\mathcal{S} = \operatorname{gph} S$, the feasible set of BL may be written as

$$M_{BL} = \{(x, y) \in \mathcal{S} \mid G(x, y) \leq 0\}.$$

If we can specify boxes X and Y, such that $M_{BL} \subseteq X \times Y$, we may employ Algorithm 1 to generate a rigorous outer approximation of M_{BL} with guaranteed tolerances. This results from the chain of inclusions

$$\mathcal{S} \subseteq \mathcal{R} \subseteq \cup_{B \in \mathcal{O}} B \subseteq \mathcal{R}_{\varepsilon,\delta},$$

where the first inclusion becomes an equality under Assumption 2.2 (cf. Theorem 2.3), and the remaining inclusions are given by the output (9) of Algorithm 1.

We emphasize that our focus in this section is not on the algorithmic solution of problems BL from the BOLIB but on studying the approximation behavior of Algorithm 1 for their feasible sets.

Preparation of the testbed and parameter choice

From the nonlinear testbed of the library, we selected instances with 2-4 variables and parsed the defining functions into the Julia language. By doing this, 6 instances were excluded because parsing to Julia was not successful or due to irregular definitions, e.g. the call of external functions in the .m files. The resulting testbed contains 99 instances.

In order to obtain bounded boxes X and Y for the initialization of P-ICGO, bound constraints on x and y were extracted from g and G. If no such bounds were available, we artificially bounded variable i based on the best-known point $(x^\star, y^\star)$, denoted in [10] and set $x_i^\star + 5$ (resp. $y_i^\star + 5$) as upper and $x_i^\star - 5$ (resp. $y_i^\star - 5$) as lower bound. If no best-known point was specified, we set 5 as upper and -5 as lower bound.

For the experiment, we choose the parameters $\varepsilon = \delta = 0.1$. In contrast to the experiments in Section 4.1, we did not specify a minimal box size and instead employed a time limit of 5200 seconds per instance to ensure the completion in a reasonable time.

Discussion of the results

After completion, we report the following statistics in Tables B1 and B2 for each run: The *name* of the instance, the number of variables n_x, n_y, the number of constraints n_g, the volume of the initial box $B_0 := X \times Y$ that was given as input, the run time t, and the number of *iterations*. Furthermore, the *number of boxes* in the lists $\mathcal{W}$, $\mathcal{O}$, and $\mathcal{O}^I$ is denoted as the cardinality of the list. Moreover, to demonstrate the correctness of the discarding step, we compute a lower bound F_l on the leader's objective as follows. First, we identify all boxes that may contain feasible points for the leader

$$\mathcal{B}_{feas} := \{B \in \mathcal{O} \cup \mathcal{W} \mid \ell(G(B)) \leq 0\}.$$

Then, we compute the smallest lower bound of F over these boxes

$$F_l := \min_{B \in \mathcal{B}_{feas}} \ell(F(B)).$$

Since, by construction, the best-known point $(x^\star, y^\star)$ is inside the initial box B_0, we should obtain $F_b := F(x^\star, y^\star) \geq F_l$ for all instances. Hence, the relative gap

$$(F_b - F_l)_r = (F_b - F_l)/(\max\{|F_b|, |F_l|, 10^{-6}\})$$

should be nonnegative. Otherwise, the point $(x^\star, y^\star)$ or the discarding step is wrong. Note that, apart from the sign, for $F_b = 0$ the relative gap is not meaningful. For $F_b \neq 0$ and if the best-known point is optimal, the relative gap gives us an idea of how close the outer approximation is to M_{BL}.

Termination behavior We report that 52 of 99 instances terminated within the time limit, 33 of them in less than one second. A clear dimensionality threshold emerged: *all but one* of the terminated instances satisfied $n_x + n_y \leq 3$. The only exception (DempeFranke2011Ex41) had one variable fixed by bounds, effectively reducing it to three dimensions. Among the 47 non-terminated instances, 24 were four-dimensional, indicating that the algorithm struggles significantly beyond three effective variables.

Runtime and vol(B_0) We observe a strong correlation between the volume of the initial box B_0 and runtime:

	vol(B_0)	
	mean	median
terminated in < 1 s	25	4
terminated in ≥ 1 s	122	50
non-terminated in 5200 s	219,672	505

Clearly, vol(B_0) is also strongly correlated with dimensionality. However, for example, DempeFranke2011Ex41 and Ex42 share identical objective functions and effective dimensionality, yet Ex42 has an initial box volume approximately ten times larger and failed to terminate in $87,000$ iterations, whereas Ex41 terminated in $7,898$ iterations. Moreover, for parameters with non-unique optimal points, we observed in Example 4.4 that line segments through the whole X-part of B_0 had to be approximated, which is increasingly difficult for larger initial boxes.

Solution quality Across all instances, the relative gap $(F_b - F_l)_r$ was nonnegative, validating the correctness of the discarding step and the best-known reference points. However, the following analysis reveals that the approximation quality degrades for non-terminated instances:

	$(F_b - F_l)_r$		
	Q1	Median	Q3
terminated ($F_b \neq 0$)	0.02	0.15	0.49
non-terminated ($F_b \neq 0$)	0.12	0.33	0.77

Although the bounds remain valid, the outer approximation becomes notably looser when the algorithm cannot complete. This was expected since the boxes of the nonempty work list $\mathcal{W}$, for which optimality and feasibility tolerance could not yet be proven, are still part of the outer approximation.

Length of lists $\mathcal{O}$ and $\mathcal{O}^I$ For the terminated instances, the ratio $|\mathcal{O}|/|\mathcal{O}^I|$ had a median of 1.17 (Q1: 1.01, Q3: 1.37), indicating that P-ICGO performs relatively few refinement steps inside $\mathcal{O}$. Within the non-terminated instances, there are 21 with $|\mathcal{O}| > 0$, which means that there are parts of the outer approximation already proven to lie inside $\mathcal{R}_{\varepsilon,\delta}$. Thereof, 6 instances have $|\mathcal{O}|/|\mathcal{W}| \geq 0.3$. A high ratio $|\mathcal{O}|/|\mathcal{W}|$ suggests that P-ICGO is well advanced in its convergence. This implies that time limits, rather than algorithmic failure, prevented completion for these instances. We will verify this assumption in a second experiment.

A second experiment

To demonstrate that the algorithm is generally capable of processing four-dimensional instances and, in particular, multidimensional parameter sets, we perform an additional experiment. Here, we shrink the initial boxes to a box with $\mathbf{vol}(B_0) \leq 1$ around the best-known point $(x^\star, y^\star)$ and rerun P-ICGO on 19 four-dimensional instances for which a best-known point is available. For details on the experimental setup and the results, see Appendix C and Table C1 therein. We report that for 7 of 19 instances, P-ICGO terminated within 5400 seconds. We conclude that the algorithm is generally able to handle multidimensional parameter sets, but the exponentially growing complexity prohibited the termination of several four-dimensional instances in the preceding experiment. For the 12 non-terminated instances, we cannot determine with certainty whether growing computational complexity or degeneracies prevent termination. However, in Section 4.1 we observed that degeneracies typically occur at isolated parameter values, while P-ICGO continues to approximate the optimal point mapping accurately for the remaining parameter values. Since the list $\mathcal{O}$ remained empty for 9 of these 12 instances, we conjecture that complexity, rather than degeneracy, is the primary cause of non-termination.

Recommendations for higher-dimensional applications

Based on these findings, we propose the following strategies to apply P-ICGO to more challenging instances. Firstly, one should cut down the search effort in lines 7 and 13, which increases with the size of the lists $\mathcal{W}$ and $\mathcal{O}$. This could be done by

- **Pre-partitioning T:** Instead of initializing with (X, T), partition T into sub-boxes $T_1, \ldots, T_q$ and run P-ICGO with $(X, T_1), \ldots, (X, T_q)$ in parallel, or
- **Exploiting box hierarchies:** Maintain links between parent and child nodes so that sub-lists $\mathcal{W}(B'_t)$ and $\mathcal{O}(B'_t)$ inherit information from predecessors, rather than recomputing from scratch.

Furthermore, the initial boxes should be kept as small as possible for the target. Where possible, extract or compute tighter bound constraints from g and G before initialization. Artificial bounds (± 5 around reference points) should be avoided when problem structure permits stronger restrictions. Finally, one should reduce refinement effort in $\mathcal{W}$ and $\mathcal{O}$ by computing tighter bounds. Instead of relying solely on generic interval arithmetic, the usage of tailored bound computations to specific function properties (e.g., using monotonicity, convexity, or separability) should be preferred.

5 Conclusions

This paper presented P-ICGO, a novel branch-and-bound algorithm for constructing rigorous outer approximations of optimal point mappings for parametric optimization problems. The method applies a branch-and-bound method to the improvement function reformulation to generate approximations with guaranteed feasibility and optimality tolerances. We proved finite termination under the same regularity conditions that ensure exactness of the underlying reformulation.

Our numerical evaluation demonstrated both the capabilities and limitations. Visual examples with up to three variables illustrated the approximation behavior and also showed the impact of specific degeneracies where the regularity conditions are violated. On the full testbed, P-ICGO successfully terminated on more than half of the tested bilevel optimization instances, with many completing in under one second. The discarding step was validated across all instances through nonnegative relative gaps between computed lower bounds and best-known reference points. However, a clear dimensionality threshold emerged, since all terminated instances had at most three effective variables. The computational effort grows exponentially with both problem dimension and initial search volume. A follow-up experiment confirmed that several four-dimensional instances become tractable when the initial search volume is reduced sufficiently.

These findings suggest that P-ICGO is well-suited for low-dimensional parametric problems where rigorous guarantees are essential. Possible applications include computing feasible sets in bilevel optimization, characterizing equilibria in generalized Nash games, or analyzing solution stability in parametric programming.

However, the *full* approximation of optimal point mappings is not the only potential application of the presented approach. For instance, in bilevel optimization, the ultimate goal is not the full geometric approximation of the follower's optimal point mapping but rather rigorous bounds on the leader's objective value. Integrating the leader's objective directly into the discarding tests could allow the algorithm to prune regions of the parameter space that are suboptimal for the leader, even if they contain valid follower solutions. This targeted approach would drastically reduce the computational effort by only approximating relevant parts of the optimal point mapping. This is just one example of how the method could be used as a building block in an application.

To further improve efficiency, we recommend the integration of parallel computing and a suitable data structure for linking the boxes. An open-source Julia implementation is available to facilitate further development and application [5].

References

[1] Beck, Y., Ljubić, I., Schmidt, M.: Linear and Mixed-Integer Bilevel Optimization: Theory and Algorithms (2026). URL `https://yasminebeck.github.io/files/bilevel-optimization-cup.pdf`. Accessed: 2026-09-22

[2] Dempe, S.: Foundations of Bilevel Programming. Kluwer Academic Publishers, Boston (2002). doi:10.1007/b101970

[3] Diestel, R.: Graph Theory, *Graduate Texts in Mathematics*, vol. 173. Springer Berlin Heidelberg (2025). doi:10.1007/978-3-662-70107-2

[4] Kirst, P., Schwarze, S., Stein, O.: A Branch-and-Bound Algorithm for Nonconvex Nash Equilibrium Problems. SIAM Journal on Optimization **34**(4), 3371–3398 (2024). doi:10.1137/23M1548189

[5] Schwarze, S.: Julicgo: Parametric improvement function based complete global optimization in Julia. https://github.com/schwarze-st/julicgo. Accessed: 2026-08-12

[6] Schwarze, S., Stein, O.: The Improvement Function Reformulation for Graphs of Minimal Point Mappings. Journal of Optimization Theory and Applications **208**(3), 99 (2026). doi:10.1007/s10957-025-02911-1

[7] Schwarze, S., Stein, O., Kirst, P., Rodestock, M.: The improvement function in branch-and-bound methods for complete global optimization. Optimization Online (2025). URL `https://optimization-online.org/?p=30132`

[8] Stein, O.: Bi-Level Strategies in Semi-infinite Programming. Springer Science & Business Media (2003). doi:10.1007/978-1-4419-9164-5

[9] Stein, O.: Grundzüge der Parametrischen Optimierung. Springer Berlin Heidelberg (2021). doi:10.1007/978-3-662-61990-2

[10] Zhou, S., Zemkoho, A.B., Tin, A.: BOLIB: Bilevel Optimization LIBrary of test problems (2020). doi:10.48550/arXiv.1812.00230

A Existence of weakly exhaustive sequences of boxes

The following Lemma can also be found in [7]. For the sake of completeness, it is adapted to the notation of P-ICGO and stated below.

Lemma A.1 (see also [7], Lemma A.2)**.** *Let (B^k) be an infinite sequence of pairwise distinct boxes generated by the following process: The tree is initialized with an 'artificial' root node B^0, then a list of boxes $\{B^{0,i}, \dots, B^{0,n}\}$, $n \geq 1$ with $\operatorname{int} B^{0,i} \neq \emptyset$, $i = 1, \dots, n$ are added as child nodes to B^0. At each iteration $k \in \mathbb{N}$:*

1. *one leaf node B^k and at most one other distinct leaf node $\widehat{B}^k$ are processed as follows,*
2. *if specific conditions hold, finitely many child nodes are added to B^k (a partition of it),*
3. *if some $\widehat{B}^k \neq B^k$ was selected and specific conditions hold, finitely many child nodes are added to $\widehat{B}^k$ (a partition of it).*

Moreover, the partition fulfills Assumption 3.5. Then there exists an infinite weakly exhaustive subsequence of (B^k).

Proof. Naturally, any node has only finitely many child nodes. The tree is a locally finite graph [3] and the infinite sequence of distinct boxes corresponds to an infinite set of nodes in the graph. Using the Star-Comb Lemma [3, Lem. 8.2.2] we may define a comb in the graph such that an infinite subsequence of (B^k) lies on that comb. By definition of a comb (see [3, Sec. 8.1]), this means that there exists an infinite nested sequence (R^k) with $R^{k+1} \subseteq R^k$ and $B^k \subseteq R^k$ for all $k \in \mathbb{N}$ (a so-called 'ray'). In view of the Assumption 3.5 on partitioning of boxes, we have $\lim_k \operatorname{diag}(R^k) = 0$. Thus, (R^k) is an exhaustive sequence of boxes and, therefore, the infinite subsequence (B^k) is weakly exhaustive. $\square$

B Tables for experiment 1 on BOLIB

The following tables contain all results of the first experiment documented in Section 4.2.

Table B1: Results of P-ICGO with $\varepsilon = \delta = 0.1$ on nonlinear BOLIB testbed: 52 instances that terminated.

name	n_x	n_y	n_g	**vol**(B_0)	**t**	**iterations**	$\lvert\mathcal{W}\rvert$	$\lvert\mathcal{O}\rvert$	$\lvert\mathcal{O}^I\rvert$	F_b	F_l	$(F_b - F_l)_r$[1]
Bard1988Ex1	1	1	3	36.4	331.67	169280	0	48959	36663	17.000	16.956	0.003
CalamaiVicente1994a	1	1	3	100	0.20	3490	0	865	749	0.000	-0.000	0.000
ClarkWesterberg1990a	1	1	3	80	0.72	10908	0	1520	1164	5.000	4.952	0.010
Colson2002BIPA3	1	1	1	45	1.50	17113	0	3476	3476	2.000	1.995	0.002
Colson2002BIPA5	1	2	2	55.52	2859.75	390383	0	137396	55146	2.750	2.695	0.020
Dempe1992b	1	1	1	100	1.72	15523	0	1634	1094	31.250	27.631	0.116
DempeDutta2012Ex24	1	1	1	100	0.03	183	0	32	32	0.000	-0.000	0.000
DempeEtal2012	1	1	0	2	0.04	497	0	183	183	-1.000	-1.000	0.000
DempeFranke2011Ex41	2	2	1	24	21.19	15184	0	2591	2202	5.000	4.648	0.070
GumusFloudas2001Ex4	1	1	0	80	0.03	183	0	32	32	9.000	7.223	0.197
HatzEtal2013	1	2	0	250	436.22	246392	0	43220	38496	0.000	-0.234	1.000
HenrionSurowiec2011	1	1	0	100	489.26	142685	0	80373	50362	0.000	-0.000	0.000
IshizukaAiyoshi1992a	1	2	5	1000	108.07	88082	0	16199	13735	0.000	-0.153	1.000
KleniatiAdjiman2014Ex3	1	1	0	4	0.07	720	0	235	234	-1.000	-1.016	0.015
LamparielloSagratella2017Ex23	1	2	1	100	142.32	195628	0	9204	4633	-1.000	-1.000	0.000
LamparielloSagratella2017Ex31	1	1	1	50	0.05	1734	0	249	195	1.000	1.000	0.000
LamparielloSagratella2017Ex32	1	1	0	100	0.05	1248	0	161	155	0.500	0.236	0.527
LamparielloSagratella2017Ex33	1	2	1	137.5	1.29	9645	0	2079	2070	0.500	0.413	0.173
LamparielloSagratella2017Ex35	1	1	1	2	0.04	412	0	82	67	0.800	0.728	0.091
LuDebSinha2016b	1	1	0	10	0.05	1	0	0	0	0.000	Inf	NaN
LuDebSinha2016e	1	2	3	100	0.80	8072	0	1450	1450	NaN	2.250	NaN
LucchettiEtal1987	1	1	0	1	0.04	438	0	173	173	0.000	-0.000	0.000
Mirrlees1999	1	1	0	40	0.22	3088	0	1494	990	1.002	0.850	0.152
MitsosBarton2006Ex310	1	1	0	1.8	1.29	6717	0	3302	2172	0.500	0.344	0.312
MitsosBarton2006Ex311	1	1	0	3.6	3.63	15145	0	5944	4514	-0.800	-0.800	0.000

... continued on the next page

name	n_x	n_y	n_g	**vol**(B_0)	**t**	**iterations**	$\lvert\mathcal{W}\rvert$	$\lvert\mathcal{O}\rvert$	$\lvert\mathcal{O}^I\rvert$	F_b	F_l	$(F_b - F_l)_r$ [1]
MitsosBarton2006Ex312	1	1	0	4	0.18	1493	0	778	637	0.000	-0.312	1.000
MitsosBarton2006Ex313	1	1	0	4	0.07	1098	0	485	419	-1.000	-1.250	0.200
MitsosBarton2006Ex314	1	1	0	4	0.06	981	0	432	368	0.250	0.062	0.750
MitsosBarton2006Ex315	1	1	0	4	0.04	332	0	87	84	0.000	-0.031	1.000
MitsosBarton2006Ex316	1	1	0	4	0.05	492	0	110	107	-2.000	-2.250	0.111
MitsosBarton2006Ex317	1	1	0	4	0.06	499	0	255	224	0.188	0.031	0.833
MitsosBarton2006Ex318	1	1	0	4	0.06	970	0	192	188	-0.250	-1.000	0.750
MitsosBarton2006Ex319	1	1	0	4	0.11	814	0	395	340	-0.258	-0.500	0.484
MitsosBarton2006Ex320	1	1	0	4	0.12	1633	0	805	589	0.312	0.125	0.600
MitsosBarton2006Ex321	1	1	0	4	0.12	1015	0	399	292	0.209	0.026	0.878
MitsosBarton2006Ex322	1	1	1	4	0.14	1106	0	390	296	0.209	0.062	0.702
MitsosBarton2006Ex324	1	1	0	3	0.04	123	0	31	31	-1.755	-1.906	0.079
MitsosBarton2006Ex38	1	1	0	4	0.20	2603	0	1119	855	0.000	-0.000	0.000
MitsosBarton2006Ex39	1	1	0	20	0.04	1519	0	256	256	-1.000	-1.039	0.038
MorganPatrone2006a	1	1	0	2	0.06	861	0	335	330	-1.000	-1.016	0.015
MuuQuy2003Ex1	1	2	1	50	1202.11	242739	0	115219	64447	-2.077	-2.380	0.127
PaulaviciusAdjiman2017a	1	1	0	4	0.06	970	0	192	188	0.250	0.191	0.234
PaulaviciusAdjiman2017b	1	1	0	4	0.16	1870	0	911	760	-2.000	-2.000	0.000
SahinCiric1998Ex2	1	1	3	80	0.81	10910	0	1502	1148	5.000	4.955	0.009
ShimizuEtal1997a	1	1	3	100	701.63	266600	0	57241	39025	NaN	16.832	NaN
SinhaMaloDeb2014TP6	1	2	4	125	266.09	157172	0	45494	27444	-1.209	-1.224	0.012
TuyEtal2007	1	1	3	50	1.46	2465	0	460	374	22.500	21.808	0.031
Vogel2012	1	1	1	50	299.21	347936	0	19222	13730	1.000	0.994	0.006
YeZhu2010Ex42	1	1	1	50	303.69	347936	0	19222	13730	1.000	0.830	0.170
Yezza1996Ex31	1	1	0	1	0.47	3629	0	1145	1091	1.500	1.469	0.021
Yezza1996Ex41	1	1	1	50	3.35	12787	0	7153	4100	0.500	0.253	0.495
Zlobec2001b	1	1	2	1	0.05	63	0	13	13	NaN	-0.000	NaN

[1] $(F_b - F_l)_r = (F_b - F_l)/(\max\{|F_b|, |F_l|, 10^{-6}\})$

Table B2: Results of P-ICGO with $\varepsilon = \delta = 0.1$ on nonlinear BOLIB testbed: 47 instances that did not terminate.

name	n_x	n_y	n_g	**vol**(B_0)	**t**	**iterations**	$\lvert\mathcal{W}\rvert$	$\lvert\mathcal{O}\rvert$	$\lvert\mathcal{O}^I\rvert$	F_b	F_l	$(F_b - F_l)_r$[1]
AiyoshiShimizu1984Ex2	2	2	2	$2.25 \cdot 10^6$	5400.49	213784	362591	0	0	5.000	-33.125	1.151
AllendeStill2013	2	2	2	1100	5400.04	221410	414783	0	0	1.000	0.070	0.930
Bard1988Ex3	2	2	2	625	5400.03	201877	232649	0	0	-12.680	-20.098	0.369
Bard1991Ex1	1	2	1	110	5400.32	686271	126697	54170	24031	2.000	2.000	0.000
BardBook1998	2	2	3	$2.25 \cdot 10^6$	5400.02	190779	260236	0	0	0.000	-0.000	0.000
Colson2002BIPA1	1	1	1	100	5400.00	545321	862863	62	61	250.000	238.644	0.045
Colson2002BIPA2	1	1	3	30	5400.01	541346	202285	191482	106303	17.000	16.938	0.004
Colson2002BIPA4	1	1	1	28.02	5400.01	451514	187159	278691	162686	88.790	86.782	0.023
DeSilva1978	2	2	0	100	5400.02	416013	426258	370	370	-1.000	-1.432	0.302
Dempe1992a	2	2	1	5000	5400.00	303470	238359	2491	2453	NaN	-0.781	NaN
DempeDutta2012Ex31	2	2	2	3260.4	5400.01	363917	292444	1179	926	-1.000	-2.367	0.578
DempeFranke2011Ex42	2	2	2	120	5400.03	220356	60183	22355	6212	2.130	1.866	0.124
DempeFranke2014Ex38	2	2	1	5	5400.02	253116	223665	34618	34618	-1.000	-2.875	0.652
DempeLohse2011Ex31a	2	2	2	3600	5400.02	319903	395409	10	7	-5.500	-6.750	0.185
FalkLiu1995	2	2	0	100	5400.01	402882	413452	340	340	-2.196	-3.475	0.368
FloudasEtal2013	2	2	3	$2.25 \cdot 10^6$	5400.04	212213	298625	0	0	0.000	-37.031	1.000
FloudasZlobec1998	1	2	2	200	5400.05	219460	145882	393	335	1.000	0.664	0.336
GumusFloudas2001Ex1	1	1	1	625	5400.01	502415	960112	149	147	2250.000	-0.000	1.000
GumusFloudas2001Ex5	1	2	2	970.3	5400.01	243886	128326	361	361	0.194	0.100	0.485
HendersonQuandt1958	1	1	0	316.67	5400.01	510166	906497	0	0	-3266.670	-3314.381	0.014
LuDebSinha2016d	2	2	3	$4.2 \cdot 10^5$	5400.01	473446	258940	0	0	NaN	-192.969	NaN
LuDebSinha2016f	2	1	0	$8.4 \cdot 10^5$	5400.30	751194	421467	0	0	NaN	-200.000	NaN
MacalHurter1997	1	1	0	100	5400.01	584572	834933	0	0	81.327	42.947	0.472
MitsosBarton2006Ex323	1	1	1	4	5400.04	232742	93389	214	82	0.176	0.141	0.201
NieWangYe2017Ex34	1	2	1	50	5400.01	702124	488965	3466	3464	2.000	1.980	0.010
NieWangYe2017Ex61	2	2	1	400	5400.03	331921	262916	0	0	-1.022	-1.862	0.451
Outrata1990Ex1a	2	2	2	5168	5400.01	172553	229241	0	0	-8.920	-12.236	0.271
Outrata1990Ex1b	2	2	2	4426	5400.04	182844	235543	7	7	-7.560	-11.734	0.356
Outrata1990Ex1c	2	2	2	6400	5400.08	184497	361550	0	0	-12.000	-12.375	0.030

... continued on the next page

name	n_x	n_y	n_g	**vol**(B_0)	**t**	**iterations**	$\lvert\mathcal{W}\rvert$	$\lvert\mathcal{O}\rvert$	$\lvert\mathcal{O}^I\rvert$	F_b	F_l	$(F_b - F_l)_r$[1]
Outrata1990Ex1d	2	2	2	3500	5400.01	195239	246955	0	0	-3.600	-12.344	0.708
Outrata1990Ex1e	2	2	2	3290	5400.05	143954	239741	0	0	-3.150	-9.678	0.675
Outrata1990Ex2a	1	2	2	452.48	5400.01	571185	386252	0	0	0.500	0.439	0.121
Outrata1990Ex2b	1	2	2	320	5400.01	572346	379469	0	0	0.500	0.439	0.121
Outrata1990Ex2c	1	2	2	429.27	5400.02	521545	440771	0	0	1.860	0.428	0.770
Outrata1990Ex2d	1	2	2	504.84	5400.00	492985	340151	0	0	0.920	-0.000	1.000
Outrata1990Ex2e	1	2	2	533.03	5400.03	511011	388310	0	0	0.900	-0.000	1.000
Outrata1993Ex31	1	2	2	398.64	5400.00	555085	389114	0	0	1.560	0.626	0.599
Outrata1993Ex32	1	2	2	451.56	5400.01	429791	560167	9510	5256	3.208	2.481	0.227
Outrata1994Ex31	1	2	2	250	5400.02	452943	536600	8974	6804	3.208	2.552	0.205
OutrataCervinka2009	2	2	3	10000	5400.07	136081	198631	0	0	0.000	-4.375	1.000
ShimizuAiyoshi1981Ex1	1	1	1	300	5400.01	1398730	4735	183265	100961	100.000	98.056	0.019
ShimizuAiyoshi1981Ex2	2	2	0	10000	5400.03	358134	334517	0	0	225.000	129.395	0.425
ShimizuEtal1997b	1	1	1	25	5400.01	485735	969075	0	0	2250.000	-0.000	1.000
SinhaMaloDeb2014TP3	2	2	2	625	5400.04	158684	187832	0	0	-18.679	-26.519	0.296
SinhaMaloDeb2014TP7	2	2	2	625	5400.00	249486	285609	0	0	-1.961	-25.787	0.924
SinhaMaloDeb2014TP8	2	2	2	$2.25 \cdot 10^6$	5400.02	202480	343163	0	0	0.000	-0.000	0.000
YeZhu2010Ex43	1	1	1	40	5400.03	756275	274267	165591	126099	1.250	1.062	0.151

[1] $(F_b - F_l)_r = (F_b - F_l)/(\max\{|F_b|, |F_l|, 10^{-6}\})$

C Table for experiment 2 on BOLIB

The smaller boxes B_0 in the second experiment were generated as follows: For a given instance with the best-known point $(x^\star, y^\star)$, we set $\max\{x_i^\star - 0.5, x_i^\ell\}$ as the lower bound of variable i and $\min\{x_i^\star + 0.5, x_i^u\}$ as the upper bound, where x_i^ℓ and x_i^u denote the existing lower and upper bounds in the model. Consequently, $\mathrm{vol}(B_0) \leq 1$ holds for all instances in the experiment.

Table C1: Results of P-ICGO with $\varepsilon = \delta = 0.1$ on four-dimensional nonlinear BOLIB instances with available best-known point.

name	n_g	**vol**(B_0)	**t**	**iterations**	$\lvert W\rvert$	$\lvert O\rvert$	$\lvert O^I\rvert$	F_b	F_l	$(F_b - F_l)_r$
AiyoshiShimizu1984Ex2	2	1.00	94.57	45619	0	8620	7758	5.000	4.188	0.163
AllendeStill2013	2	1.00	5400.00	226133	298997	0	0	1.000	0.387	0.613
BardBook1998	3	1.00	5400.03	173505	200805	0	0	0.000	-0.000	0.000
DeSilva1978	0	0.25	65.89	38565	0	8985	8981	-1.000	-1.309	0.236
Dempe1992a	1	0.50	5400.05	347862	79171	37138	20952	NaN	-0.375	NaN
DempeDutta2012Ex31	2	1.00	5400.00	364022	212157	37341	20730	-1.000	-1.500	0.333
DempeFranke2011Ex41	1	0.50	4.29	6003	0	1311	1236	5.000	4.633	0.073
DempeFranke2011Ex42	2	0.50	12.81	11116	0	1222	735	3.000	2.762	0.079
DempeFranke2014Ex38	1	0.03	2600.07	247620	0	97915	97915	-1.000	-1.078	0.072
DempeLohse2011Ex31a	2	1.00	5400.01	260544	226309	19949	18870	-5.500	-6.148	0.105
FalkLiu1995	0	0.75	2.63	6361	0	1057	1041	-2.196	-2.819	0.221
FloudasEtal2013	3	0.06	5400.05	240508	416025	0	0	0.000	-0.188	1.000
LuDebSinha2016d	3	0.50	1731.36	340209	0	83764	83764	NaN	Inf	NaN
Outrata1990Ex1a	2	1.00	5400.06	139037	190275	0	0	-8.920	-10.313	0.135
Outrata1990Ex1b	2	1.00	5400.10	147691	151898	0	0	-7.560	-9.341	0.191
Outrata1990Ex1c	2	1.00	5400.05	194089	368659	0	0	-12.000	-12.061	0.005
Outrata1990Ex1d	2	0.50	5400.01	161242	233170	0	0	-3.600	-5.820	0.381
Outrata1990Ex1e	2	0.50	5400.02	102624	176234	0	0	-3.150	-6.171	0.490
ShimizuAiyoshi1981Ex2	0	0.50	5400.01	265181	354945	0	0	225.000	209.518	0.069