The improvement function in branch-and-bound methods for complete global optimization

Stefan Schwarze* Oliver Stein# Peter Kirst‡
Marc Rodestock§

December 16, 2025

Abstract

We present a new spatial branch-and-bound approach for treating optimization problems with nonconvex inequality constraints. It is able to approximate the set of all global minimal points in case of solvability, and else to detect infeasibility. The new technique covers the nonconvex constraints by means of an improvement function which, although nonsmooth, can be treated by standard bounding operations.

The method is shown to be successful under a weak regularity condition, and we also give a transparent interpretation of the output in case that this condition is violated. Numerical tests illustrate the performance of the algorithm.

Key words: Nonconvex minimization; spatial branch-and-bound; complete global optimization; outer approximation; improvement function.

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

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

‡Operations Research and Logistics (ORL), Wageningen University & Research (WUR), The Netherlands, peter.kirst@wur.nl

§Institute for Operations Research (IOR), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, uisab@student.kit.edu

1 Introduction

With a nonempty bounded box $X \subseteq \mathbb{R}^n$ and continuous functions $f, \omega : X \to \mathbb{R}$ we consider the optimization problem

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

While we will introduce additional differentiability properties of f and ω as needed, we do not impose any convexity assumptions on f or ω. We shall denote the set of feasible points and the set of global minimal points of P by Ω and S, respectively, and the optimal value by v. We will occasionally refer to global minimal points as minimal points.

The aim of this paper is to present a new branch-and-bound approach that

- handles the nonconvex constraint ω in a deterministic manner which allows a convergence proof,
- approximates not only one, but all global minimal points of P in case of solvability ('complete global optimization'),
- else detects infeasibility.

The branch-and-bound paradigm was first introduced for discrete optimization problems in [32, 11]. Its extension to spatial branch-and-bound for continuous global optimization problems goes back to [19]. Various improvements and related methods have been proposed since then, such as branch-and-reduce [44, 45], symbolic branch-and-bound [47, 48], branch-and-contract [55] or branch-and-cut [52]. Several state-of-the-art global optimization solvers are based on implementations of branch-and-bound algorithms, for example ANTIGONE [36], BARON [46], COUENNE [9], LINDOGLOBAL [33], SCIP [1] and SHOT [34]. For extensive reviews on deterministic global optimization we refer to [20, 27].

Within a spatial branch-and-bound algorithm, lower bounds for the minimal value v of P are typically obtained by tailored bounding procedures. We will discuss some of them in Section 3.1. On the other hand, upper bounds can be generated by evaluating the objective function f at feasible points of P, either directly or indirectly by applying local solvers. However, as we do not assume the inequality constraint function ω to be convex, finding a feasible point may be as difficult as solving the problem itself. Therefore, there is no guarantee for the occurrence of feasible points within current branch-and-bound algorithms [30], let alone convergent valid upper bounds, which are required to prove finite termination. To cope with this, in the literature sometimes it is suggested to accept *almost* feasible points [20], but [30] shows that instead of upper bounds this may generate arbitrarily bad lower bounds.

As a remedy [30] suggests to enforce feasibility of infeasible iterates by their perturbation along Mangasarian-Fromovitz directions, which requires the solution of one auxiliary LP per iteration as well as the assumption of the Mangasarian-Fromovitz constraint qualification (Section 2.2) in all optimal points. Using different tools, the follow-up articles [21, 29] show how under suitable constraint qualifications valid upper bounds can be obtained for optimization problems with nonlinear equality constraints and nonconvex inequality constraints, respectively.

Most spatial branch-and-bound approaches have in common that, for solvable problems, they output the minimal value v up to some user-defined tolerance ε, together with *one* feasible point at which the objective function attains this approximate minimal value, hence an ε-minimal point. Some solvers, e.g. BARON, offer the option to compute multiple or all solutions and indicate that this functionality is reliable in continuous domains, as long as all solutions are *isolated*. The output then consists of multiple ε-minimal points that differ by a specified threshold. Although often this suffices for practical purposes, there exist important applications that require an assured approximation of *all exactly* minimal points, which we refer to as *complete global optimization* (not to be confused with complete solvers for global optimization [39]). This typically happens if the problem P is part of a larger model, for example as the lower level of a bilevel optimization problem [12], as a player problem in a (generalized) Nash equilibrium problem [18], or in the lexicographic optimization of multicriteria problems [16]. To facilitate the computation of the set of global minimal points, these problem classes are often treated under convexity assumptions, which possess, however, limited modeling power.

An available method that is capable of computing an enclosure of all minimal points of P is presented in the monograph [24]. It is referred to as an interval method because, in contrast to spatial branch-and-bound methods, every step of the algorithm is carried out using interval-valued quantities. In detail, it combines the branch-and-bound paradigm with discarding tests based on the interval-valued evaluation of necessary first- and second-order optimality conditions. It has been proven that the method returns a list of boxes in which each box has a width below a predefined tolerance and the objective function varies within a predefined tolerance. The highest objective value attained within the approximation is given by the maximum upper bound of the interval evaluation of f over all boxes. In particular, there is no guarantee that this upper bound will converge, because there is no deterministic procedure for computing arbitrarily good feasible points. The monograph does not provide a convergence proof, even when further assumptions on P are imposed. A solver-package that implements an interval method for complete global optimization is IBEX [40]. The procedures used to compute feasible points, described in [6], are classified as heuristic.

Complete global optimization is also considered in the field of Lipschitz optimization, see [25] for an introduction. These approaches construct a single bounding function by iteratively sampling points and their objective values, and, using the Lipschitz constant, form a so-called saw-tooth cover. For *box constrained* optimization problems, [25] states that Piyavskii's method [42] can, in principle, be adapted to perform complete global optimization. For generally constrained optimization problems, the Lipschitz optimization approach switches from sampling to the closely related branch-and-bound approach because of the necessity to remove infeasible partitions of the domain. An algorithm similar to that in [27] is introduced and also applied to the minimax risk evaluation with bounded parameters [22]. However, no extension to complete global optimization is mentioned and the authors argue that convergence is difficult to establish under the mere Lipschitz assumption, due to feasibility verification and that additional assumptions on the involved functions are not in the scope of Lipschitz optimization [25].

A spatial branch-and-bound method for the approximation of all global minimal points of nonconvex problems is presented in [17]. It handles twice differentiable *box-constrained* optimization problems and is based on the αBB technique from [4, 3, 5]. For predefined tolerances ε and δ, its output approximates the entire set of minimal points by a set of ε-minimal points A with the following property: Each minimal point has at most the distance of δ to some element of A. As there are only box constraints involved, the method terminates without any further assumptions. However, a generalization to optimization problems with nonconvex constraints is not straightforward since, as explained above, the reliable construction of feasible points for the generation of valid upper bounds can no longer be assumed. Furthermore, in the above mentioned applications, one might prefer an approximation that gives an enclosure of the set of minimal points.

We place our method in the broader class of spatial branch-and-bound techniques. Although we use interval arithmetic to compute rigorous lower and upper bounds, not all operations are carried out in that way. While in

Section 5, bounds are also computed via Lipschitz constants, we require differentiable functions, and thus our method deviates from a purely Lipschitz-optimization-based framework. Our approach takes the difficulty of finding arbitrarily good feasible points for nonconvex constraints into account. Under mild conditions, that allows us to outer approximate the set of all global minimal points deterministically with a predefined accuracy. It is based on the consideration of the auxiliary parametric optimization problem

$$Q(s): \qquad \min_x \psi(x,s) \quad \text{s.t.} \quad x \in X$$

with

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

and a parameter $s \in X$ which may be considered a 'parametric copy of x'.

Let $R(s)$ denote the set of minimal points of $Q(s)$ and

$$\mathcal{R} \; := \; \{s \in \Omega \mid s \in R(s)\} \tag{1}$$

the fixed-point set of the set-valued mapping $R : X \rightrightarrows X$ on Ω. We shall see that $\mathcal{R}$ coincides with the set S of global minimal points of P under mild assumptions. Note that they are both void in the case $\Omega = \emptyset$. For $S = \emptyset$ we follow the usual convention to set $v := +\infty$. On the other hand, for $\Omega \neq \emptyset$ the problem P is solvable by the Weierstrass theorem.

Constructions using the objective function ψ of $Q(s)$ have been used in the derivation of Karush-Kuhn-Tucker type necessary optimality conditions for nonsmooth constrained optimization problems [26, Th. VII.2.2.5], and in the derivation of first and second order optimality conditions for standard semi-infinite optimization problems in [10, Sec. 5.4]. In [53] it is used as one way to construct separators, which allow the global minimization of constrained problems using d.c. optimization. In [7] it is employed in an iterative linearization approach for nonsmooth constrained optimization problems (with the previous iterate in place of s), and there the function ψ is called improvement function. A related technique appears in [54] for the solution of semi-infinite optimization problems. Despite these different applications, we are not aware of any studies showing the explicit relation between the fixed point set $\mathcal{R}$ of R and the optimal point set S.

Compared to the problem P, in $Q(s)$ the improvement function moves functional constraints from the description of the feasible set Ω to the objective function without the need to specify appropriate penalty parameters as in penalty (or barrier) approaches, or multipliers as in (augmented) Lagrangian methods. Considering $Q(s)$ instead of P is particularly useful if X is a box and an algorithm, like a standard branch-and-bound method, requires a purely box-constrained feasible set.

This reformulation comes at the price of the introduction of a parametric optimization problem with nonsmooth objective function, and the fact that the validity of $S = \mathcal{R}$ requires some mild assumptions. The nonsmoothness of the max-type objective function in $Q(s)$ is, however, not an issue for branch-and-bound methods, since they do not rely on smoothness of the problem functions, but rather on the existence of convergent bounding procedures. The latter are available for the function ψ, if they are for f and ω (Section 3). The proposed branch-and-bound method can therefore approximate the set $\mathcal{R}$ without further assumptions, whereas we decouple the theoretical considerations on why mild regularity assumptions guarantee $S = \mathcal{R}$ from the algorithmic design of the method.

This article is structured as follows. In Section 2 we show that S is a subset of $\mathcal{R}$, and we provide mild conditions under which both sets coincide. Based on these results, in Section 3 we formulate the branch-and-bound method for the approximation of all global minimal points of P. Section 4 provides a convergence proof for the proposed method, and Section 5 presents our numerical experience on a test set. Section 6 concludes this article with some final remarks.

2 Reformulations by the improvement function

In the following, we investigate central theoretical properties of the introduced reformulation. In this section, we can reduce the assumptions on X to being a nonempty compact set from $\mathbb{R}^n$. However, for the later branch-and-bound approach in Section 3, we will assume X to be a box.

2.1 The relation of S and $\mathcal{R}$

The following result is the key to the algorithmic treatment of the parameter-dependence in the definition of $\mathcal{R}$ from (1) (Section 3).

Lemma 2.1. *The following assertions hold:*

$$\forall\, s \in \Omega : \min_{x \in X} \psi(x, s) \leq 0, \tag{2}$$

$$\mathcal{R} \;=\; \{s \in \Omega \mid \min_{x \in X} \psi(x, s) = 0\}. \tag{3}$$

Proof. Since each $s \in \Omega$ satisfies $s \in X$ and $\omega(s) \leq 0$, we obtain

$$\min_{x \in X} \psi(x, s) \;\leq\; \psi(s, s) \;=\; \max\{\omega(s), 0\} \;=\; 0,$$

showing (2). Assertion (3) follows by

$$\begin{aligned} \{s \in \Omega \mid \min_{x \in X} \psi(x, s) = 0\} \;&=\; \{s \in \Omega \mid \min_{x \in X} \psi(x, s) = \psi(s, s)\} \\ &=\; \{s \in \Omega \mid s \in R(s)\} \;=\; \mathcal{R}. \end{aligned}$$

□

Remark 2.2. *The construction from Lemma 2.1 is reminiscent of the Blum-Oettli type equilibrium formulation of P [41], namely to find some $s \in \Omega$ with $f(x) - f(s) \geq 0$ for all $x \in \Omega = \{x \in X \mid \omega(x) \leq 0\}$, where s corresponds to a global minimal point. From this point of view, the improvement function approach replaces this equilibrium problem by the task to find some $s \in \Omega$ with $\psi(x, s) = \max\{\omega(x), f(x) - f(s)\} \geq 0$ for all $x \in X$. As in general Blum-Oettli type equilibrium problems, also the function ψ satisfies $\psi(s, s) = 0$ for all $s \in \Omega$, but since Ω and X are different sets, the latter is not such an equilibrium problem. Nevertheless, the function $-\min_{x \in X} \psi(x, s)$ corresponds to the generalization of Auslender's gap function for variational inequalities to equilibrium problems from [28].*

To see how the set $\mathcal{R}$ may differ from S, and under which conditions both sets coincide, we make use of the 'strict feasible set' or 'set of Slater points' (recall, though, that we do not require any convexity assumptions)

$$\Omega_< \;:=\; \{x \in X \mid \omega(x) < 0\}$$

which satisfies $\operatorname{cl} \Omega_< \subseteq \Omega$. By $\operatorname{cl} A$ we denote the topological closure of a set A. We define the problem

$$P_< : \qquad \min_x \; f(x) \quad \text{s.t.} \quad x \in \operatorname{cl} \Omega_<$$

with the (possibly empty) compact feasible set $\operatorname{cl} \Omega_<$ and the minimal value $v_<$ (where $v_< := +\infty$ for $\operatorname{cl} \Omega_< = \emptyset$).

While the description of $\mathcal{R}$ by (3) is the basis for our algorithmic approach to the computation of S in Section 3, the subsequent description (4) is better suited for studying the relation between $\mathcal{R}$ and S.

Lemma 2.3. *The identity*

$$\mathcal{R} \;=\; \{s \in \Omega \mid f(s) \leq v_<\} \tag{4}$$

holds, where the formal case $v_< = +\infty$ corresponds to $\mathcal{R} = \Omega$.

Proof. Let $v_< < +\infty$. For each $s \in \mathcal{R}$ the condition $s \in \Omega$ is clear. In view of (3)

$$0 \;\leq\; \psi(x,s) \;=\; \max\{\omega(x),\; f(x) - f(s)\}$$

holds for every $x \in X$. From this the relation $f(x) - f(s) \geq 0$ follows in the case $\omega(x) < 0$, i.e., for all $x \in \Omega_<$. For every $x \in (\operatorname{cl}\Omega_<) \setminus \Omega_<$ the same relation follows from a continuity argument, since x is then the limit point of some sequence from $\Omega_<$. Thus, we have shown

$$f(s) \;\leq\; f(x) \quad \forall\, x \in \operatorname{cl}\Omega_<,$$

that is, $f(s) \leq v_<$.

To see the reverse inclusion, let $s \notin \mathcal{R}$. In the case $s \notin \Omega$ the assertion is trivially true. Otherwise, by Lemma 2.1 there exists some $x \in X$ with $0 > \psi(x,s) = \max\{\omega(x), f(x) - f(s)\}$. This implies $x \in \Omega_< \subseteq \operatorname{cl}\Omega_<$, $f(x) < f(s)$ and, therefore, $v_< < f(s)$.

The proof for the case $v_< = +\infty$ runs along the same lines. □

For $\operatorname{cl}\Omega_< \neq \emptyset$, due to $v \leq v_<$ the value $\sigma := v_< - v$ is nonnegative. Lemma 2.3 thus yields

$$\mathcal{R} \;=\; S_\sigma \;:=\; \{s \in \Omega \mid f(s) \leq v + \sigma\},$$

that is, $\mathcal{R}$ consists of the σ-optimal points of P with this special value σ.

In the case $\operatorname{cl}\Omega_< = \emptyset$ we have $v_< = +\infty$ and either $\Omega \neq \emptyset$ or $\Omega = \emptyset$. The subcase $\Omega \neq \emptyset$ results in $v \in \mathbb{R}$ and $\sigma = +\infty - v = +\infty$, while for the subcase $\Omega = \emptyset$ we define $\sigma = +\infty - (+\infty) := 0$.

In the next result $A \,\dot\cup\, B$ denotes the union of disjoint sets A and B.

Lemma 2.4. *For $\sigma = v_< - v$ the sets S and $\mathcal{R}$ satisfy*

$$\mathcal{R} \;=\; S \,\dot\cup\, \{s \in \Omega \mid v < f(s) \leq v + \sigma\}.$$

The formal case $v_< = +\infty$, $v \in \mathbb{R}$ leads to $\sigma = +\infty$ and $\mathcal{R} = S \,\dot\cup\, \{s \in \Omega \mid v < f(s)\}$, and in the formal case $v = +\infty$ all appearing sets are empty.

Proof. The assertion follows from Lemma 2.3 and $S = \{s \in \Omega \mid v = f(s)\}$. □

Example 2.5. *Fig. 1 illustrates the assertion of Lemma 2.4. Here $S = \{x^1\}$ and $\mathcal{R} = \{x^1, x^2, x^3\}$ hold.*

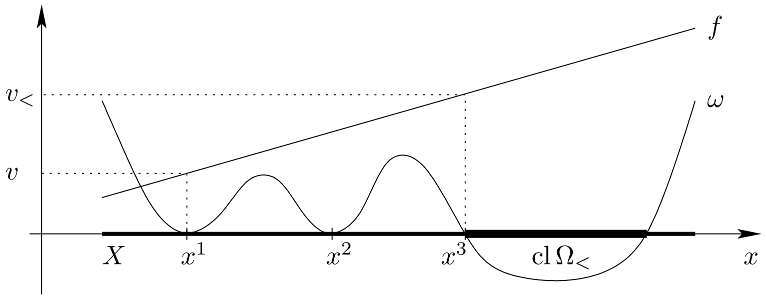

Figure 1: Optimal values $v_<$ and v in Example 2.5

Since we aim at characterizing S by its identity with $\mathcal{R}$, the set $\{s \in \Omega \mid v < f(s) \leq v + \sigma\}$ from Lemma 2.4 contains 'spurious' points. Example 2.5 illustrates that spurious points (here x^2, x^3) may possess any positive distance from S so that, for example, the Hausdorff distance between S and an outer approximation of $\mathcal{R}$ cannot be expected to tend to zero for refined approximations. This motivates the following assumption, under which Lemma 2.4 implies the subsequent Theorem 2.7.

Assumption 2.6. *The value $\sigma = v_< - v$ vanishes.*

Theorem 2.7. *Assumption 2.6 implies*

$$\mathcal{R} \;=\; S. \tag{5}$$

Assumption 2.6 acts as an abstract regularity condition. In the next section we will show that it is even weaker than a mild constraint qualification and, thus, mild itself. Still, it is only sufficient for (5), but not necessary, since (5) also holds in the case $S = \Omega$.

2.2 A regularity condition and a constraint qualification

In a first step we relate Assumption 2.6 to an abstract condition on feasible points.

Definition 2.8. *We say that the sequential Slater condition (SSC) holds at $x \in \Omega$ if $x \in \operatorname{cl} \Omega_<$ is true, that is, if x can be approximated by Slater points of Ω.*

We point out that in Definition 2.8 we call x a Slater point of $\Omega = \{x \in X \mid \omega(x) \leq 0\}$ if $x \in X$ and $\omega(x) < 0$ hold. In particular, with respect to X only feasibility of x is needed. As long as X is not assumed to possess a special structure, like being a box (Section 3), this admits to incorporate possible equality constraints from the description of Ω into the description of X, so that they do not prevent the SSC from holding.

Assumption 2.9. *At some $x \in S$ the SSC is satisfied.*

Example 2.10. *For $X = [-1,1]^2$, $f(x) = -x_1$, $\omega_1(x) = x_1^3 + x_2$ and $\omega_2(x) = -x_2$, Assumption 2.9 holds at the unique element of $S = \{0\}$,*

Lemma 2.11. *Assumption 2.9 implies Assumption 2.6.*

Proof. Under Assumption 2.9 we may choose some $x \in S \cap \operatorname{cl} \Omega_<$. Then there exists a sequence $(x^k) \subseteq \Omega_<$ with $\lim_k x^k = x$. Due to $f(x^k) \geq v_<$ for all k and the continuity of f we obtain $v = f(x) = \lim_k f(x^k) \geq v_<$ which, in view of $v \leq v_<$, yields Assumption 2.6. □

We remark that Assumption 2.9 is reminiscent of, but weaker than [29, Ass. 2], which there is crucial for the convergence proof.

Example 2.10 shows that the regularity condition SSC is not a constraint qualification, since it can be satisfied at minimal points which are not Karush-Kuhn-Tucker points. However, in the following we shall see how under a smooth functional description of Ω a mild constraint qualification can be employed to guarantee Assumption 2.9 and, thus, the absence of spurious points (5).

Indeed, we use the Mangasarian-Fromovitz constraint qualification (MFCQ) for continuously differentiable optimization problems. Therefore we assume

$$\omega(x) \;=\; \max_{j \in J} \omega_j(x)$$

with a finite index set J and on X continuously differentiable functions ω_j, $j \in J$. We also assume a functional description $X = \{x \in \mathbb{R}^n \mid \xi_k(x) \leq 0,\ k \in K\}$ of X with a finite index set K (satisfying $J \cap K = \emptyset$) and on $\mathbb{R}^n$ continuously differentiable functions ξ_k, $k \in K$. We thus may write

$$\Omega \;=\; \{x \in \mathbb{R}^n \mid \omega_j(x) \leq 0,\ j \in J,\ \xi_k(x) \leq 0,\ k \in K\}. \tag{6}$$

Note that the max-formulation in the auxiliary problem $Q(s)$ only affects the functions ω_j, but not the ξ_k, so that the nonempty and compact feasible set X guarantees the solvability of $Q(s)$.

In the following the active index sets of $x \in \Omega$ with respect to the two groups of inequalities are denoted by $J_0(x) = \{j \in J \mid \omega_j(x) = 0\}$ and

$K_0(x) = \{k \in K \mid \xi_k(x) = 0\}$, and $D\omega_j(x)$ stands for the row vector of partial derivatives of ω_j at x. The MFCQ is said to hold at $x \in \Omega$ if a vector $d \in \mathbb{R}^n$ with $D\omega_j(x)d < 0$, $j \in J_0(x)$, and $D\xi_k(x)d < 0$, $k \in K_0(x)$, exists. Occasionally we will refer to d as an MF vector.

Assumption 2.12. *The feasible set Ω is given in the form* (6) *with C^1-functions ω_j, $j \in J$, ξ_k, $k \in K$, and the MFCQ holds at some $x \in S$.*

Note that, in contrast to Assumption 2.9, Assumption 2.12 rules out the presence of equality constraints in the description of X.

Lemma 2.13. *At any $x \in \Omega$, given in the form* (6) *with C^1-functions ω_j, $j \in J$, ξ_k, $k \in K$, the MFCQ implies the SSC. In particular, Assumption 2.12 implies Assumption 2.9.*

Proof. Choose any $x \in \Omega$ at which the MFCQ holds with some corresponding MF vector d. Then for all sufficiently small $\alpha > 0$ one has $\omega_j(x + \alpha d) < 0$, $j \in J$, and $\xi_k(x + \alpha d) < 0$, $k \in K$. This implies $x + \alpha d \in \Omega_<$. In the limit for $\alpha \searrow 0$ one obtains $x \in \mathrm{cl}\,\Omega_<$, so that at x the SSC is satisfied. This shows the assertions. □

Although Assumption 2.12 is already mild, Example 2.10 shows that it is strictly stronger than Assumption 2.9.

We collect our findings from Theorem 2.7, Lemma 2.11 and Lemma 2.13 in the following result.

Theorem 2.14. *Under either of the Assumptions 2.9 and 2.12 the following assertions hold:*

a) $\sigma = v_< - v = 0$,

b) $\mathcal{R} = S$.

In Example 2.5 the MFCQ and even the weaker SSC are violated at the unique point $x^1 \in S$ and, indeed, none of the two assertions from Theorem 2.14 are true. On the other hand, if the set $X = [a, b]$ from Example 2.5 is replaced by $X' = [a', b]$ with some lower boundary point $a' \in (x^2, x^3)$, then Theorem 2.14 can be applied.

2.3 Generalization to δ-feasible and ε-minimal points

For numerical reasons it is often necessary to introduce tolerances in feasibility and optimality of points. In Section 3 we will need such tolerances in the termination criterion of the proposed branch-and-bound method.

The δ-feasible points of P for $\delta \geq 0$ form the set

$$\Omega_\delta \;:=\; \{x \in X \mid \omega(x) \leq \delta\}. \tag{7}$$

Moreover, for $\varepsilon \geq 0$ we consider the set of ε-minimal points

$$S_\varepsilon \;=\; \{x \in \Omega \mid f(x) \leq v + \varepsilon\}$$

as well as the set of strict ε-minimal points

$$S_\varepsilon^< \;=\; \{x \in \Omega \mid f(x) < v + \varepsilon\}$$

of P. The simultaneously δ-feasible and (strict) ε-minimal points of P thus form the sets

$$S_{\varepsilon,\delta} \;=\; \{x \in \Omega_\delta \mid f(x) \leq v + \varepsilon\}$$

and

$$S_{\varepsilon,\delta}^< \;=\; \{x \in \Omega_\delta \mid f(x) < v + \varepsilon\},$$

respectively. We emphasize that the latter sets are defined via the optimal value v of P, rather than by the optimal value v_δ of minimizing f over Ω_δ.

As an appropriate extension of the constructions from Section 2.1 to $\varepsilon > 0$, we employ the parametric problem

$$Q_\varepsilon(s): \qquad \min_x \; \psi_\varepsilon(x,s) \quad \text{s.t.} \quad x \in X$$

with

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

and parameter $s \in X$, and we define the sets

$$R_\varepsilon(s) := \{x \in X \mid \psi_\varepsilon(x,s) \leq \psi_\varepsilon(y,s) + \varepsilon \;\forall\, y \in X\}$$

and

$$R_\varepsilon^<(s) := \{x \in X \mid \psi_\varepsilon(x,s) < \psi_\varepsilon(y,s) + \varepsilon \;\forall\, y \in X\}$$

of all ε-minimal and strict ε-minimal points of $Q_\varepsilon(s)$, respectively. We also define the fixed point sets

$$\mathcal{R}_{\varepsilon,\delta} \;:=\; \{s \in \Omega_\delta \mid s \in R_\varepsilon(s)\}$$

and

$$\mathcal{R}^<_{\varepsilon,\delta} \;:=\; \{s \in \Omega_\delta \mid s \in R^<_\varepsilon(s)\},$$

of the set-valued mappings $R_\varepsilon, R^<_\varepsilon : X \rightrightarrows X$ on Ω_δ.

Lemma 2.15. *For all $\varepsilon \geq \delta \geq 0$ the following assertions hold:*

$$\forall\, s \in \Omega_\delta : \; \min_{x \in X} \psi_\varepsilon(x, s) \leq \varepsilon, \tag{8}$$

$$\mathcal{R}_{\varepsilon,\delta} = \{s \in \Omega_\delta \mid \min_{x \in X} \psi_\varepsilon(x, s) \in [0, \varepsilon]\}, \tag{9}$$

$$\mathcal{R}^<_{\varepsilon,\delta} = \{s \in \Omega_\delta \mid \min_{x \in X} \psi_\varepsilon(x, s) \in (0, \varepsilon]\}. \tag{10}$$

Proof. For $\varepsilon = 0$ the assertions (8) and (9) were shown in Lemma 2.1. Moreover, the set $\mathcal{R}^<_{0,0}$ is empty, in line with the fact that strictly 0-minimal points of $Q(s)$ do not exist.

For the proofs in the case $\varepsilon > 0$ we use that $\delta \leq \varepsilon$ implies

$$\forall\, s \in \Omega_\delta : \; \max\{\omega(s), \varepsilon\} \;=\; \varepsilon. \tag{11}$$

In analogy to the proof of (2), for the proof of (8) observe that, since each $s \in \Omega_\delta$ satisfies $s \in X$, from (11) we obtain

$$\min_{x \in X} \psi_\varepsilon(x, s) \;\leq\; \psi_\varepsilon(s, s) \;=\; \max\{\omega(s), \varepsilon\} \;=\; \varepsilon.$$

To see (9), observe that for every $s \in \Omega_\delta$, the point s lies in $R_\varepsilon(s)$ if and only if all $y \in X$ satisfy

$$\psi_\varepsilon(y, s) + \varepsilon \geq \psi_\varepsilon(s, s) = \max\{\omega(s), \varepsilon\} = \varepsilon,$$

where the last equality follows again from (11). The latter relation is equivalent to $\min_{y \in X} \psi_\varepsilon(y, s) \geq 0$ and, together with (8), shows (9). The proof of (10) is analogous. $\square$

For all $\varepsilon \geq \delta \geq 0$, in view of the continuity of $\min_{x \in X} \psi_\varepsilon(x, \cdot)$ and the closedness of Ω_δ, (9) yields the closedness of $\mathcal{R}_{\varepsilon,\delta}$. On the other hand, from (10) we see that $\mathcal{R}^<_{\varepsilon,\delta}$ with $\varepsilon > 0$ may neither be closed nor open.

To understand the relations between the sets $\mathcal{R}_{\varepsilon,\delta}$ and $\mathcal{R}^<_{\varepsilon,\delta}$ on the one hand, and $S_{\varepsilon,\delta}$ and $S^<_{\varepsilon,\delta}$ on the other, we consider again the problem $P_<$ with minimal value $v_<$ from Section 2.1.

Lemma 2.16. *For all $\varepsilon \geq \delta \geq 0$ the identities*

$$\mathcal{R}_{\varepsilon,\delta} = \{s \in \Omega_\delta \mid f(s) \leq v_< + \varepsilon\} \tag{12}$$

and

$$\mathcal{R}^<_{\varepsilon,\delta} = \{s \in \Omega_\delta \mid f(s) < v + \varepsilon\} = S^<_{\varepsilon,\delta} \tag{13}$$

hold. The formal cases $v_< = +\infty$ and $v = +\infty$ correspond to $\mathcal{R}_{\varepsilon,\delta} = \Omega_\delta$ and $\mathcal{R}^<_{\varepsilon,\delta} = S^<_{\varepsilon,\delta} = \Omega_\delta$, respectively.

Proof. We start with the proof of (13). To show its first identity, we first assume $v < +\infty$. Then, by (10), every $s \in \mathcal{R}^<_{\varepsilon,\delta}$ satisfies $s \in \Omega_\delta$ and

$$\begin{aligned} 0 &< \min_{x\in X} \psi_\varepsilon(x, s) = \min_{x\in X} \max\{\omega(x), f(x) - f(s) + \varepsilon\} \\ &\leq \min_{x\in \Omega} \max\{\omega(x), f(x) - f(s) + \varepsilon\} \\ &\leq \min_{x\in \Omega} \max\{0, f(x) - f(s) + \varepsilon\} \end{aligned}$$

In particular, for all $x \in \Omega$

$$0 \; < \; f(x) - f(s) + \varepsilon$$

is true, which yields

$$f(s) < v + \varepsilon.$$

To see the reverse inclusion, let $s \notin \mathcal{R}^<_{\varepsilon,\delta}$. For $s \notin \Omega_\delta$ we are done. Otherwise, by (8) and (10) there exists some $x \in X$ with $0 \geq \psi_\varepsilon(x, s)$. This implies $x \in \Omega$ and $f(s) \geq f(x) + \varepsilon \geq v + \varepsilon$, so that the first identity in (13) is shown. The second identity is just the definition of $S^<_{\varepsilon,\delta}$. The case $v = +\infty$ is equivalent to $\Omega = \emptyset$. Then all three sets coincide with the (not necessarily void) set Ω_δ.

The proof of (12) uses similar arguments, combined with those from the proof of (4). □

From (12) one sees that $S_{\varepsilon,\delta}$ coincides with $\mathcal{R}_{\varepsilon,\delta}$ under a mild regularity assumption, in analogy to the identity $S = \mathcal{R}$. Indeed, since we only need the identity $v_< = v$ from Assumption 2.6, we obtain the following generalization of Theorem 2.14.

Theorem 2.17. *Under either of the Assumptions 2.9 and 2.12*

$$\mathcal{R}_{\varepsilon,\delta} = S_{\varepsilon,\delta}$$

is true for all $\varepsilon \geq \delta \geq 0$.

On the other hand, from (13) we obtain that the sets $S^<_{\varepsilon,\delta}$ and $\mathcal{R}^<_{\varepsilon,\delta}$ coincide without any regularity assumptions, so that an approximation of $S^<_{\varepsilon,\delta}$ may alternatively be performed by approximating $\mathcal{R}^<_{\varepsilon,\delta}$. In this sense, the improvement function reformulation 'fits best' to the solution concept of strict ε-minimal (and δ-feasible) points of P. In this case one may even consider equality constraints in the description of Ω by rewriting them as a pair of inequality constraints, rather than moving them to the description of X, although this rules out the validity of MFCQ and SSC at any $x \in \Omega$ (so that, e.g., Th. 2.17 is not applicable). Unfortunately, the translation of this result to a branch-and-bound method is prevented, for example, by the fact that discarding tests in analogy to those from Section 3.2 must then be formulated with nonstrict instead of strict inequalities on certain bounds, which would be numerically infeasible.

3 A new branch-and-bound approach

The main idea of our new approach is to outer-approximate the set $\mathcal{R}_{\varepsilon,\delta}$ for some user-specified tolerances $\varepsilon \geq \delta \geq 0$. Under either of the mild Assumptions 2.9 or 2.12, by Theorem 2.17 $\mathcal{R}_{\varepsilon,\delta}$ coincides with the set $S_{\varepsilon,\delta}$ of δ-feasible and ε-minimal points of P. In particular, for the natural choice $\varepsilon = \delta = 0$ the proposed procedure then approximates the set of global minimal points of P. Additionally, two larger tolerances $\varepsilon_{\max} > \varepsilon$ and $\delta_{\max} > \delta$ are user-specified, and the algorithm returns an approximation which is sandwiched between $\mathcal{R}_{\varepsilon,\delta}$ and $\mathcal{R}_{\varepsilon_{\max},\delta_{\max}}$.

The algorithm is called improvement function based complete global optimization (ICGO). We briefly describe its structure, before the following subsections describe the decisive operations in detail. The framework is illustrated in Figure 2. ICGO maintains a working list $\mathcal{W}$ and an output list $\mathcal{O}$. The working list is initialized with the host set X, which is, as announced above, from now on assumed to be an n-dimensional bounded box. This is a prerequisite for branching by box partitioning and makes many prominent convergent bounding procedures accessible for our method. As long as the working list $\mathcal{W}$ is nonempty, some box X' is selected from it. For X', we firstly perform a discarding test, i.e. we try to prove $X' \cap \mathcal{R}_{\varepsilon,\delta} = \emptyset$. If this holds X' is discarded and hence removed from $\mathcal{W}$. If discarding was not successful, we perform an inclusion test and try to prove that X' lies completely inside the user-specified tolerances, i.e. $X' \subseteq \mathcal{R}_{\varepsilon_{\max},\delta_{\max}}$. If this holds, we move X' to the list $\mathcal{O}$. If not, we have to refine X' by partitioning it into smaller sub-boxes and append them to $\mathcal{W}$. After termination, ICGO returns

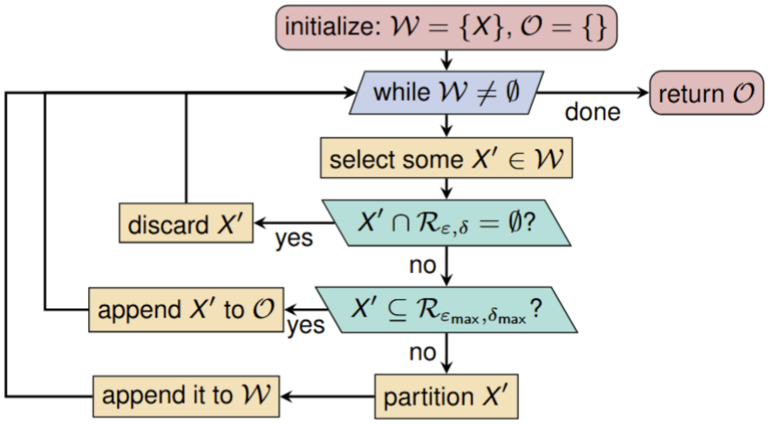

Figure 2: Algorithmic framework of ICGO

a list $\mathcal{O}$ with

$$\mathcal{R}_{\varepsilon,\delta} \subseteq \bigcup_{X' \in \mathcal{O}} X' \subseteq \mathcal{R}_{\varepsilon_{\max},\delta_{\max}}.$$

By Theorem 2.17, under Assumptions 2.9 or 2.12 the latter means that the set $\bigcup_{X' \in \mathcal{O}} X'$ is sandwiched between $S_{\varepsilon,\delta}$ and $S_{\varepsilon_{\max},\delta_{\max}}$.

In the following Section 3.1 we will define convergent bounding procedures. These are the basis for the computations of the discarding tests, described in Section 3.2 and the inclusion test, described in Section 3.3. Finally, we can write down the full algorithm in Section 3.4 and prove its convergence in Section 4.

3.1 Convergent bounding procedures

In the following, we use standard notation and terminology. In particular, we denote a box $X' \subseteq \mathbb{R}^n$ by $[\underline{x}', \overline{x}']$ with $\underline{x}', \overline{x}' \in \mathbb{R}^n$, $\underline{x}' \leq \overline{x}'$. By the width of a box X' we mean its diagonal length $\operatorname{diag}(X') = \|\overline{x}' - \underline{x}'\|_2$ with respect to the Euclidean distance, and we denote its midpoint by $\operatorname{mid}(X') = \frac{1}{2}(\underline{x}' + \overline{x}')$. Indeed, branch-and-bound methods rely on the following concept of convergent bounding procedures. Recall from, e.g., [27] that a sequence of boxes (X^k) is called exhaustive if firstly, $X^{k+1} \subseteq X^k$ is true for all $k \in \mathbb{N}$ and, secondly, $\lim_k \operatorname{diag}(X^k) = 0$ holds.

Definition 3.1. *Let $\mathcal{F}$ denote some subset of the set of lower semi-continuous functions on X.*

a) A function ℓ_f from the set of all sub-boxes X' of X to $\mathbb{R}$ and for $f \in$

$\mathcal{F}$ is called lower bounding procedure *on $\mathcal{F}$, if it satisfies $\ell_f(X') \leq \min_{x \in X'} f(x)$ for all sub-boxes $X' \subseteq X$ and any arbitrary $f \in \mathcal{F}$.*

b) A lower bounding procedure ℓ_f is called convergent *if*

$$\lim_k \ell_f(X^k) = \lim_k \min_{x \in X^k} f(x)$$

holds for any exhaustive sequence of boxes (X^k) and any arbitrary $f \in \mathcal{F}$.

c) A lower bounding procedure ℓ_f is called monotone, *if $\ell_f(X_2) \geq \ell_f(X_1)$ holds for all boxes $X_2 \subseteq X_1 \subseteq X$ and any arbitrary $f \in \mathcal{F}$.*

The concept of convergent upper bounding procedures u_f on $\mathcal{F}'$ is defined analogously for subsets $\mathcal{F}'$ of the set of upper semi-continuous functions, where one requires $u_f(X) \geq \max_{x \in X} f(x)$ in part a) and $u_f(X_2) \leq u_f(X_1)$ in part c).

We remark that for any exhaustive sequence of boxes (X^k) the set $\bigcap_{k \in \mathbb{N}} X^k$ is a singleton, say $\{\widetilde{x}\}$ with some $\widetilde{x} \in \mathbb{R}^n$. Then, in view of $\lim_k \operatorname{diag}(X^k) = 0$, the convergence of ℓ_f and lower semi-continuity of f yield

$$\lim_k \ell_f(X^k) = \lim_k \min_{x \in X^k} f(x) = f(\widetilde{x}). \tag{14}$$

Prominent examples of convergent bounding procedures arise in the so-called αBB relaxation as described in [4, 3, 5], in centered forms as described in [8, 31], in bounds based on duality concepts as examined in [14, 15] or in bounds based on linearization techniques (see [35, 51, 47, 48]). All of these bounding procedures can be used in a straightforward manner such that they are monotone as well. Many of these approaches take advantage of interval arithmetic as explained, for instance, in [38]. The direct application of interval arithmetic is also possible. For this reason, the defining functions must be assumed to be factorable, i.e. composed of a finite number of operations such as $+$, $-$, sin, cos, exp, etc.

In the algorithm, two infinite sequences of boxes with different purposes will appear in the same list (or branch-and-bound tree, respectively). This leads to some technicalities that require a new milder property for infinite sequences of boxes. We call a sequence (X^k) weakly exhaustive if there exists an exhaustive sequence (Z^k) such that $X^k \subseteq Z^k$ for all $k \in \mathbb{N}$. Note that a weakly exhaustive sequence of boxes, although it is not necessarily nested, also converges to a single point, as specified and exploited in the following lemma. Moreover, the lemma shows that we can replace 'exhaustive' by 'weakly exhaustive' in Definition 3.1, if we limit ourselves to continuous functions in $\mathcal{F}$.

Lemma 3.2. *Let f be a continuous function on X and (X^k) be a* weakly exhaustive sequence of boxes. *Then the following holds:*

a) Any sequence (x^k) with $x^k \in X^k$ converges to the same limit point, say $\widetilde{x} = \lim_k x^k$.

b) For every convergent lower bounding procedure, we have

$$\lim_k \ell_f(X^k) = \lim_k \min_{x \in X^k} f(x) = f(\widetilde{x}). \tag{15}$$

c) For every convergent upper bounding procedure, we have

$$\lim_k u_f(X^k) = \lim_k \max_{x \in X^k} f(x) = f(\widetilde{x}). \tag{16}$$

Proof. Since $X^k \subseteq Z^k$ holds for all $k \in \mathbb{N}$, $\lim_k \operatorname{diag}(Z^k) = 0$ implies $\lim_k \operatorname{diag}(X^k) = 0$. In addition with $\bigcap_{k\in\mathbb{N}} Z^k = \widetilde{x}$, we obtain that any sequence of points (x^k) with $x^k \in X^k$ converges to $\widetilde{x}$, which proves a). To prove b), we apply this to the sequence of minimal points and firstly obtain, together with the continuity of f, the identity

$$\lim_k \min_{x^k \in X^k} f(x) = f(\widetilde{x}).$$

It remains to show that the sequence $(\ell_f(X^k))$ converges to $f(\widetilde{x})$. Indeed, the monotonicity of the bounding procedure and $X^k \subseteq Z^k$ yield $\ell_f(X^k) \geq \ell_f(Z^k)$ for all k and, therefore

$$\liminf_k \ell_f(X^k) \geq \lim_k \ell_f(Z^k) = f(\widetilde{x}). \tag{17}$$

Furthermore, let us choose a subsequence with $\lim_l \ell_f(X^{k_l}) = \limsup_k \ell_f(X^k)$ as well as some sequence (x^{k_l}) with $x^{k_l} \in X^{k_l}$ for all l. By part a), it fulfills $x^{k_l} \to \widetilde{x}$ and thus $f(x^{k_l}) \to f(\widetilde{x})$. The lower bounding property of ℓ_f hence implies

$$f(\widetilde{x}) = \lim_l f(x^{k_l}) \geq \lim_l \ell_f(X^{k_l}) = \limsup_k \ell_f(X^k).$$

Together with (17) this shows $\lim_k \ell_f(X^k) = f(\widetilde{x})$, as required. The proof for c) works analogously. □

3.2 Discarding tests

For $\varepsilon \geq \delta \geq 0$, our aim is to outer-approximate $\mathcal{R}_{\varepsilon,\delta}$ by successively removing sub-boxes X' from X which lie completely in its set complement $X \setminus \mathcal{R}_{\varepsilon,\delta}$. The description

$$\mathcal{R}_{\varepsilon,\delta} \; = \; \{s \in \Omega_\delta \mid \min_{x \in X} \psi_\varepsilon(x,s) \geq 0\}$$

follows directly from Lemma 2.15 and yields

$$\begin{aligned} X \setminus \mathcal{R}_{\varepsilon,\delta} \; &= \; \{s \in X \mid \omega(s) > \delta\} \; \cup \; \{s \in \Omega_\delta \mid \min_{x \in X} \psi_\varepsilon(x,s) < 0\} \\ &= \; \{s \in X \mid \omega(s) > \delta\} \; \cup \; \{s \in X \mid \min_{x \in X} \psi_\varepsilon(x,s) < 0\}. \end{aligned}$$

For a sub-box X' of X we will thus design one discarding test which guarantees

$$\min_{s \in X'} \omega(s) > \delta \tag{18}$$

and one entailing

$$\max_{s \in X'} \min_{x \in X} \psi_\varepsilon(x,s) < 0 \tag{19}$$

(recall that the function $\min_{x \in X} \psi_\varepsilon(x,\cdot)$ is continuous on X).

Using a lower bounding procedure ℓ_ω, a discarding test guaranteeing (18) is $\ell_\omega(X') > \delta$. Our proposal for a discarding test entailing (19) requires some further considerations.

First observe that, with a lower bounding procedure ℓ_f, for any $\widetilde{y} \in X$ the term

$$u_{\psi_\varepsilon}(\{\widetilde{y}\}, X') \; := \; \max\{\omega(\widetilde{y}), f(\widetilde{y}) - \ell_f(X') + \varepsilon\}$$

satisfies

$$\begin{aligned} u_{\psi_\varepsilon}(\{\widetilde{y}\}, X') \; &\geq \; \max\{\omega(\widetilde{y}), f(\widetilde{y}) - \min_{s \in X'} f(s) + \varepsilon\} \\ &= \; \max_{s \in X'} \max\{\omega(\widetilde{y}), f(\widetilde{y}) - f(s) + \varepsilon\} \\ &= \; \max_{s \in X'} \psi_\varepsilon(\widetilde{y}, s) \\ &\geq \; \max_{s \in X'} \min_{x \in X} \psi_\varepsilon(x,s). \end{aligned}$$

Therefore $u_{\psi_\varepsilon}(\{\widetilde{y}\}, X') < 0$ may be used as a discarding test entailing (19). This discarding test is equivalent to $\omega(\widetilde{y}) < 0$ and $f(\widetilde{y}) < \ell_f(X') - \varepsilon$. Thus, to discard successfully, we need to compute strictly feasible points with increasingly small objective values. We call the best known point $\widetilde{y} \in \Omega_<$ the *incumbent* and keep track of it in the course of the algorithm. However, for a nonconvex function ω, the construction of some point y with $\omega(y) < 0$ is already a nontrivial task, even if it exists.

We propose to address this issue by considering the term

$$\ell_{\psi_\varepsilon}(Y, X') \ := \ \max\{\ell_\omega(Y), \ell_f(Y) - u_f(X') + \varepsilon\}$$

for a sub-box Y of X and computing

$$\min_{Y \in \mathcal{W} \cup \mathcal{O}} \ell_{\psi_\varepsilon}(Y, X').$$

Our aim is to improve the current incumbent $\widetilde{y}$ with a point from the box

$$\widehat{Y} \in \operatorname{argmin}_{Y \in \mathcal{W} \cup \mathcal{O}} \ell_{\psi_\varepsilon}(Y, X')$$

and also refine $\widehat{Y}$ further if the discarding was not successful. This can be motivated as follows. We can only find some point $\widetilde{y}$ with $u_{\psi_\varepsilon}(\{\widetilde{y}\}, X') < 0$ inside the box $\widehat{Y}$, if $\ell_{\psi_\varepsilon}(\widehat{Y}, X') < 0$ holds. This can be seen from

$$\begin{aligned}
\ell_{\psi_\varepsilon}(\widehat{Y}, X') \ &\leq \ \max\{\min_{y \in \widehat{Y}} \omega(y), \min_{y \in \widehat{Y}} f(y) - \max_{s \in X'} f(s) + \varepsilon\} \\
&\leq \ \min_{y \in \widehat{Y}} \max\{\omega(y), f(y) - \max_{s \in X'} f(s) + \varepsilon\} \\
&\leq \ \max\{\omega(\widetilde{y}), f(\widetilde{y}) - \min_{s \in X'} f(s) + \varepsilon\} \\
&\leq \ \max\{\omega(\widetilde{y}), f(\widetilde{y}) - \ell_f(X') + \varepsilon\} \\
&= \ u_{\psi_\varepsilon}(\{\widetilde{y}\}, X'),
\end{aligned}$$

where the second inequality follows with standard arguments and the third exploits the assumption $\widetilde{y} \in \widehat{Y}$. If convergent bounding procedures are used, then in the case $\ell_{\psi_\varepsilon}(\widehat{Y}, X') < 0$, and for sufficiently small boxes $\widehat{Y}$, X', it should be possible to prove that for any $\widetilde{y} \in \widehat{Y}$ also $u_{\psi_\varepsilon}(\{\widetilde{y}\}, X') < 0$ holds, so that X' can be discarded. That this is indeed the case lies at the heart of the convergence proof in Section 4.

In fact, the method converges if we just select any point from $\widehat{Y}$, e.g. the midpoint, and try if it improves the incumbent. We stress that more elaborate strategies may improve efficiency. Naturally, in addition to the computation of points from $\widehat{Y}$, the incumbent $\widetilde{y}$ could also be improved by local solvers or heuristics.

3.3 Inclusion test

To prove the inclusion of a sub-box X' from X inside the relaxed set

$$\mathcal{R}_{\varepsilon_{\max},\delta_{\max}} = \{s \in \Omega_{\delta_{\max}} \mid f(s) \leq v_< + \varepsilon_{\max}\},$$

we use the following lemma. For $\varepsilon_{\max} > \varepsilon$ and $\delta_{\max} > \delta$ the latter set is a relaxation of $\mathcal{R}_{\varepsilon,\delta}$ in its description (12), and we also assume $\varepsilon_{\max} \geq \delta_{\max}$.

Lemma 3.3. *Every sub-box X' from X with*

$$0 \;\leq\; \min_{Y \in \mathcal{W} \cup \mathcal{O}} \ell_{\psi_{\varepsilon_{\max}}}(Y, X') \tag{20}$$

satisfies $X' \cap \Omega_{\delta_{\max}} \subseteq \mathcal{R}_{\varepsilon_{\max},\delta_{\max}}$.

Proof. In the case $S_< = \emptyset$, Lemma 2.16 yields $\mathcal{R}_{\varepsilon_{\max},\delta_{\max}} = \Omega_{\delta_{\max}}$, so the assertion holds. It remains to treat the case $S_< \neq \emptyset$.

From $0 \leq \min_{Y \in \mathcal{W} \cup \mathcal{O}} \ell_{\psi_{\varepsilon_{\max}}}(Y, X')$ we obtain for all $Y \in \mathcal{W} \cup \mathcal{O}$

$$\begin{aligned} 0 \;&\leq\; \ell_{\psi_{\varepsilon_{\max}}}(Y, X') \\ &\leq\; \max\{\min_{y \in Y} \omega(y), \min_{y \in Y} f(y) - \max_{s \in X'} f(s) + \varepsilon_{\max}\} \\ &\leq\; \min_{y \in Y} \max\{\omega(y), f(y) - \max_{s \in X'} f(s) + \varepsilon_{\max}\}. \end{aligned} \tag{21}$$

This means that all $y \in Y$ fulfill $\omega(y) \geq 0$ or $f(y) \geq \max_{s \in X'} f(s) - \varepsilon_{\max}$ and, hence, all $y \in Y \cap \Omega_<$ satisfy $f(y) \geq \max_{s \in X'} f(s) - \varepsilon_{\max}$. By the continuity of f, the latter inequality also holds for all $y \in Y \cap \operatorname{cl} \Omega_<$.

By $S_< \neq \emptyset$, we may consider an optimal point $y_<$ of $P_<$. In view of Lemma 2.16, identity (12), the point $y_<$ lies in $\mathcal{R}_{\varepsilon,\delta}$ and is therefore never discarded by the tests designed in Section 3.2. We can thus choose some box $Y \in \mathcal{W} \cup \mathcal{O}$ that contains $y_<$.

Due to $y_< \in Y \cap \operatorname{cl} \Omega_<$, the above considerations yield $v_< = f(y_<) \geq \max_{s \in X'} f(s) - \varepsilon_{\max}$. Hence, by (12), all the points $s \in X' \cap \Omega_{\delta_{\max}}$ lie in $\mathcal{R}_{\varepsilon_{\max},\delta_{\max}}$. □

Condition (20) from Lemma 3.3 and the additional condition

$$u_\omega(X') \;\leq\; \delta_{\max},$$

which implies $X' \subseteq \Omega_{\delta_{\max}}$, ensure together $X' \subseteq \mathcal{R}_{\varepsilon_{\max},\delta_{\max}}$.

We note that the computation of (20) is quite expensive, as it involves a search through the entire list $\mathcal{W} \cup \mathcal{O}$. However, the bounds $\ell_\omega(Y)$, $\ell_f(Y)$, and $u_f(X')$ needed for the computation of $\ell_{\psi_{\varepsilon_{\max}}}(Y,X')$ also appear in the previously computed term $\ell_{\psi_\varepsilon}(Y,X')$ and do not have to be computed twice. Furthermore, we compute only (20) in the algorithm if $u_\omega(X') \leq \delta_{\max}$ already holds. Moreover, by Lemma A.1, we can save computational effort as follows. If the left part of the maximum within $\ell_{\psi_\varepsilon}(\widehat{Y},X')$ exceeds the right part by $\varepsilon_{\max} - \varepsilon$, we know that the minimum (20) is also attained at $\widehat{Y}$ and there is no explicit computation needed.

3.4 The complete algorithm

A formal description of the improvement function based complete global optimization algorithm (ICGO) is stated in Algorithm 1.

We remark that the list $\mathcal{O}$ may be empty upon termination of Algorithm 1. This implies

$$\emptyset \;=\; \bigcup_{X' \in \mathcal{O}} X' \;\supseteq\; \mathcal{R}_{\varepsilon,\delta} \;\supseteq\; S_{\varepsilon,\delta} \;\supseteq\; S_{\varepsilon,0} \;=\; S_\varepsilon$$

Since, under our standing assumptions, S_ε is nonempty in the case $\Omega \neq \emptyset$, an empty list $\mathcal{O}$ implies $\Omega = \emptyset$.

Vice versa, an empty set Ω leads to an empty list $\mathcal{O}$ upon termination, if $\delta_{\max}$ is chosen sufficiently small. Indeed, Ω is empty if and only if $\min_{x\in X}\omega(x) > 0$ holds. Any choice $\delta_{\max} \in (0, \min_{x\in X}\omega(x))$ thus leads to

$$\bigcup_{X' \in \mathcal{O}} X' \;\subseteq\; \mathcal{R}_{\varepsilon_{\max},\delta_{\max}} \;\subseteq\; \Omega_{\delta_{\max}} \;=\; \emptyset \tag{22}$$

and, hence, an empty list $\mathcal{O}$.

4 Proof of convergence

In this section, we present the convergence proof for ICGO. It is stated under milder assumptions than the constraint qualifications from Section 2.2. The termination is ensured for finite values of $v_<$ and v, which imply non-negative finite values of σ (cf. Section 2.1), given that the tolerance $\varepsilon_{\max}$ is chosen appropriately. We will analyze and illustrate its output in Section 5.

In order to exploit the convergence property of bounding procedures, we need to ensure that whenever ICGO generates an infinite sequence of successively

Algorithm 1: (ICGO) Improvement function based complete global optimization

Input: Problem P; tolerances $\varepsilon_{\max} > \varepsilon \geq \delta \geq 0$, $\delta_{\max} \in (\delta, \varepsilon_{\max}]$;

Output: List $\mathcal{O}$: $\mathcal{R}_{\varepsilon,\delta} \subseteq \bigcup_{X' \in \mathcal{O}} X' \subseteq \mathcal{R}_{\varepsilon_{\max},\delta_{\max}}$; incumbent $\widetilde{y}$

initialize $\mathcal{W} = \{X\}$, $\mathcal{O} = \{\}$, $\widetilde{y} = \text{null}$, $\omega(\widetilde{y}) = f(\widetilde{y}) = +\infty$;

while $\mathcal{W} \neq \emptyset$ **do**

select an element $X' \in \mathcal{W}$;

if $\ell_\omega(X') > \delta$ **then**

discard X';

else if $u_{\psi_\varepsilon}(\{\widetilde{y}\}, X') < 0$ **then**

discard X';

else

find a box $\widehat{Y} \in \operatorname{argmin}_{Y \in \mathcal{W} \cup \mathcal{O}} \ell_{\psi_\varepsilon}(Y, X')$;

select a point $\widehat{y} \in \widehat{Y}$;

if $\psi(\widehat{y}, \widetilde{y}) < 0$ **then**

if $u_{\psi_\varepsilon}(\{\widehat{y}\}, X') < 0$ **then**

discard X'

end

update $\widetilde{y} = \widehat{y}$;

end

end

if $X' \in \mathcal{W}$ **then** // X' not discarded

if $u_\omega(X') \leq \delta_{\max}$ **then**

find a box $\check{Y} \in \operatorname{argmin}_{Y \in \mathcal{W} \cup \mathcal{O}} \ell_{\psi_{\varepsilon_{\max}}}(Y, X')$;

if $\ell_{\psi_{\varepsilon_{\max}}}(\check{Y}, X') \geq 0$ **then**

remove X' from $\mathcal{W}$;

append X' to $\mathcal{O}$;

end

end

if $X' \in \mathcal{W}$ **then** // X' not appended to $\mathcal{O}$

remove X' from $\mathcal{W}$;

partition X' into sub-boxes and append them to $\mathcal{W}$;

if $\widehat{Y} \neq X'$ **then**

remove $\widehat{Y}$ from $\mathcal{W}$ or $\mathcal{O}$, respectively;

partition $\widehat{Y}$ into sub-boxes and append them to the list, from which $\widehat{Y}$ has been taken;

end

end

end

end

refined boxes, the diameter of the boxes converges to zero (cf. Section 3.1). We do this by employing a simple rule.

Partitioning rule (PR): We partition a box B by dividing it along the midpoint of a longest edge into two boxes B^1 and B^2.

Theorem 4.1. *Let the bounding procedures ℓ_ω, u_ω, ℓ_f and u_f be convergent and monotone, let $v_<$ and v be finite, let $\varepsilon_{\max} > \varepsilon + \sigma$, and let all boxes be partitioned according to (PR). Then ICGO terminates after finitely many iterations.*

Proof. Assume that the assertion is wrong. Then $\mathcal{W}$ never becomes empty and in infinitely many iterations, an element $X' \in \mathcal{W}$ can be selected. Inside the while-loop, a selected element is discarded (i.e. removed) from $\mathcal{W}$ or further partitioned into two sub-boxes that are appended to $\mathcal{W}$. An infinite number of iterations is possible only if the box X' is further partitioned in infinitely many iterations. Otherwise, the list $\mathcal{W}$ would become empty at some point. Thus, we may consider an infinite sequence of boxes (X^k), whose elements are neither discarded in lines 4 to 16, nor moved from $\mathcal{W}$ to $\mathcal{O}$ in lines 22 to 23. As a result, the following statements must hold for each element X^k of this sequence

$$\ell_\omega(X^k) \leq \delta, \tag{23}$$

$$0 \;\leq\; u_{\psi_\varepsilon}(\{\widetilde{y}^k\}, X^k), \tag{24}$$

where $\widetilde{y}^k$ denotes the incumbent $\widetilde{y}$ available in the iteration where X^k is selected,

$$\psi(\widehat{y}^k, \widetilde{y}^k) \geq 0 \quad \text{or} \quad 0 \;\leq\; u_{\psi_\varepsilon}(\{\widehat{y}^k\}, X^k), \tag{25}$$

where $\widehat{y}^k$ denotes the point $\widehat{y}$ computed in the iteration where X^k is selected, and

$$u_\omega(X^k) > \delta_{\max} \quad \text{or} \quad \ell_{\psi_{\varepsilon_{\max}}}(\check{Y}^k, X^k) < 0, \tag{26}$$

where $\check{Y}^k$ denotes the box $\check{Y}$ computed in correspondence to X^k in this iteration. If one of the statements does not hold, X^k is removed from $\mathcal{W}$ without being further refined and re-appended to $\mathcal{W}$ in line 27. This contradicts the generation of an infinite sequence of boxes (X^k). We will now focus on the statement (26) and only come back to (23), (24) and (25) if necessary.

At least one of the expressions in (26) must hold in infinitely many iterations.

Case 1: $u_\omega(X^k) > \delta_{\max}$ holds in infinitely many iterations.

Since each element of (X^k) is removed from $\mathcal{W}$, before a partition of it is appended to $\mathcal{W}$, the elements of (X^k) are pairwise distinct. Therefore, Lemma A.2 implies that some subsequence of (X^k) is weakly exhaustive, which we will consider from now on. We obtain

$$\delta_{\max} \;\leq\; \lim_k u_\omega(X^k) \;=\; \omega(\widetilde{x}) \;=\; \lim_k \ell_\omega(X^k).$$

from Lemma 3.2 with some $\widetilde{x} \in X$. For any sufficiently large k this implies $\ell_\omega(X^k) \geq (\delta_{\max} + \delta)/2 > \delta$, which contradicts (23).

Case 2: $\ell_{\psi_{\varepsilon_{\max}}}(\check{Y}^k, X^k) < 0$ holds in infinitely many iterations.

We first consider the infinite sequence $(\widehat{Y}^k)$ produced in line 9 of the algorithm in correspondence to (X^k). Since the boxes $\widehat{Y}^k$ also get refined in the corresponding iterations, by Lemma A.2, we may also pass to a weakly exhaustive subsequence $(\widehat{Y}^k)$. For the sequence of points $\widehat{y}^k$ in $\widehat{Y}^k$, $k \in \mathbb{N}$, Lemma 3.2.a) implies $\lim_k \widehat{y}^k = \widehat{y}$ with some $\widehat{y} \in X$.

Part 2.1: Construct a lower bound on $f(\widetilde{x})$

Since (24) must hold for all elements of the infinite sequence (X^k), we obtain

$$0 \;\leq\; u_{\psi_\varepsilon}(\{\widetilde{y}^k\}, X^k) \;=\; \max\{\omega(\widetilde{y}^k), f(\widetilde{y}^k) - \ell_f(X^k) + \varepsilon\}. \tag{27}$$

We now show that this relation also holds if we replace $\widetilde{y}^k$ by $\widehat{y}^k$. In iterations where the if-statement in line 11 is true, it is clearly established by (25). If otherwise the if-statement in line 11 is false, i.e. $\psi(\widehat{y}^k, \widetilde{y}^k) \geq 0$, we obtain $\omega(\widehat{y}^k) \geq 0$ or $f(\widehat{y}^k) \geq f(\widetilde{y}^k)$. As we only update $\widetilde{y}^k$ with strictly feasible points, the second term of the maximum in (27) is non-negative, and the case $f(\widehat{y}^k) \geq f(\widetilde{y}^k)$ implies $f(\widehat{y}^k) - \ell_f(X^k) + \varepsilon \geq 0$. Consequently,

$$0 \;\leq\; u_{\psi_\varepsilon}(\{\widehat{y}^k\}, X^k) \;=\; \max\{\omega(\widehat{y}^k), f(\widehat{y}^k) - \ell_f(X^k) + \varepsilon\}$$

holds for all $k \in \mathbb{N}$. The convergence of the lower bounding procedure ℓ_f (see (15)) implies

$$0 \leq \lim_k u_{\psi_\varepsilon}(\{\widehat{y}^k\}, X^k) = \psi_\varepsilon(\widehat{y}, \widetilde{x}). \tag{28}$$

We now consider the sequence $(\check{Y}^k)$ generated in line 20. In this sub-case, as the statement holds in infinitely many iterations, we may pass to an infinite subsequence such that

$$\ell_{\psi_{\varepsilon_{\max}}}(\check{Y}^k, X^k) < 0 \tag{29}$$

holds for all $k \in \mathbb{N}$. Then, by $\varepsilon_{\max} > \varepsilon$, the following inequalities hold

$$\min_{Y \in \mathcal{W} \cup \mathcal{O}} \ell_{\psi_\varepsilon}(Y, X^k) = \ell_{\psi_\varepsilon}(\widehat{Y}^k, X^k) \leq \ell_{\psi_\varepsilon}(\check{Y}^k, X^k) \leq \ell_{\psi_{\varepsilon_{\max}}}(\check{Y}^k, X^k) < 0.$$

The convergence property (15) and (28) imply

$$\lim_k \min_{Y \in \mathcal{W} \cup \mathcal{O}} \ell_{\psi_\varepsilon}(Y, X^k) = \lim_k \ell_{\psi_\varepsilon}(\widehat{Y}^k, X^k) = \psi_\varepsilon(\widehat{y}, \widetilde{x}) \geq 0.$$

Together, the sandwich theorem yields

$$\lim_k \ell_{\psi_{\varepsilon_{\max}}}(\check{Y}^k, X^k) = 0, \tag{30}$$

which implies that $(\check{Y}^k)$ also contains a weakly exhaustive sequence of boxes. To see this, note that (29) holds for all $k \in \mathbb{N}$. This value can only converge to zero for infinitely many pairwise distinct boxes. Together with Lemma A.2 follows the weak exhaustiveness of $(\check{Y}^k)$, after possibly passing to a subsequence. Furthermore, with $\check{y} \in X$, (30), (15), and (16) imply

$$\begin{aligned} 0 &= \max\{\lim_k \ell_\omega(\check{Y}^k), \lim_k \ell_f(\check{Y}^k) - \lim_k u_f(X^k) + \varepsilon_{\max}\} \\ &= \max\{\omega(\check{y}), f(\check{y}) - f(\widetilde{x}) + \varepsilon_{\max}\}. \end{aligned}$$

Due to $0 \geq \omega(\check{y})$ this yields the feasibility of $\check{y}$ for P and, thus, $f(\check{y}) \geq v$. Therefore, we obtain the lower bound

$$f(\widetilde{x}) \;\geq\; v + \varepsilon_{\max}. \tag{31}$$

Part 2.2: Construct an upper bound on $f(\widetilde{x})$.

For the sequence of incumbents, the update condition in line 11 implies $f(\widetilde{y}^k) \geq f(\widetilde{y}^{k+1})$ for all $k \in \mathbb{N}$. Furthermore, $\widetilde{y}^k \subseteq \Omega_<$ implies $f(\widetilde{y}^k) \geq v_<$. The monotonicity and boundedness of the sequence of values $f(\widetilde{y}^k)$, $k \in \mathbb{N}$ implies its convergence.

Case 2.2.1: The sequence $f(\widetilde{y}^k)$ does not converge to $v_<$.

Firstly, we obtain $\lim_k f(\widetilde{y}^k) > v_<$. We now construct a point to obtain an upper bound for $f(\widetilde{x})$. Let $y_<$ be an arbitrary solution of $P_<$, which means $f(y_<) = v_<$ and $y_< \in \operatorname{cl}\Omega_<$. For any given tolerance $\rho > 0$, we may choose some $y_<^\rho \in \Omega_<$ near $y_<$ such that $f(y_<^\rho) \leq v_< + \rho$ and $f(y_<^\rho) \leq f(\widetilde{y}^k)$ hold for all $k \in \mathbb{N}$. The box containing the point $y_<^\rho$ can never be discarded, because it contains a strictly feasible point that cannot be improved by any $\widetilde{y}^k$.

Furthermore, the relation

$$\psi_\varepsilon(\widehat{y}, \widetilde{x}) \;\leq\; \psi_\varepsilon(y_<^\rho, \widetilde{x}) \tag{32}$$

holds. To see this, assume that $\psi_\varepsilon(\widehat{y}, \widetilde{x}) > \psi_\varepsilon(y_<^\rho, \widetilde{x})$ is true. The convergence of the bounding procedures yields

$$\ell_{\psi_\varepsilon}(\widehat{Y}^k, X^k) = \max\{\ell_\omega(\widehat{Y}^k), \ell_f(\widehat{Y}^k) - u_f(X^k) + \varepsilon\} \;>\; \psi_\varepsilon(y_<^\rho, \widetilde{x}). \tag{33}$$

for sufficiently large k. We now consider the box $Y^\rho \in \mathcal{W} \cup \mathcal{O}$ with $y_<^\rho \in Y^\rho$. Then we have

$$\begin{aligned} \psi_\varepsilon(\widehat{y}, \widetilde{x}) > \psi_\varepsilon(y_<^\rho, \widetilde{x}) &= \max\{\omega(y_<^\rho), f(y_<^\rho) - f(\widetilde{x}) + \varepsilon\} \\ &\geq \max\{\ell_\omega(Y^\rho), \ell_f(Y^\rho) - u_f(X^k) + \varepsilon\} = \ell_{\psi_\varepsilon}(Y^\rho, X^k). \end{aligned}$$

Together with (33), we obtain a contradiction since, for sufficiently large k, the box $\widehat{Y}^k$ would not have been chosen in line 9, which proves (32).

Consequently, together with (28), we can state

$$0 \;\leq\; \psi_\varepsilon(\widehat{y}, \widetilde{x}) \;\leq\; \psi_\varepsilon(y_<^\rho, \widetilde{x}) \;=\; \max\{\omega(y_<^\rho), f(y_<^\rho) - f(\widetilde{x}) + \varepsilon\}. \qquad (34)$$

From $\omega(y_<^\rho) < 0$ we obtain

$$f(\widetilde{x}) \;\leq\; f(y_<^\rho) + \varepsilon \;\leq\; v_< + \rho + \varepsilon$$

and, since $\rho > 0$ was arbitrary, the upper bound is

$$f(\widetilde{x}) \;\leq\; v_< + \varepsilon. \qquad (35)$$

Case 2.2.2: The sequence $f(\widetilde{y}^k)$ converges to $v_<$.

This implies that each cluster point of $\widetilde{y}^k$ is an optimal point of $P_<$. Thus, we obtain $\lim_k \widetilde{y}^k = y_<$, after possibly passing to a subsequence. In particular, for this subsequence (24) must hold for all $k \in \mathbb{N}$, i.e.

$$0 \leq u_{\psi_\varepsilon}(\widetilde{y}^k, X^k) \;=\; \max\{\omega(\widetilde{y}^k), f(\widetilde{y}^k) - \ell_f(X^k) + \varepsilon\}.$$

By $\widetilde{y}^k \in \Omega_<$,

$$0 \leq f(\widetilde{y}^k) - \ell_f(X^k) + \varepsilon$$

holds for all $k \in \mathbb{N}$. Since in this case $\lim_k f(\widetilde{y}^k) = v_<$ holds, we obtain in the limit

$$0 \leq v_< - f(\widetilde{x}) + \varepsilon.$$

This yields the same upper bound as (35) in Case 2.2.1.

Part 2.3: Obtain the contradiction with both bounds.

Combining the lower bound (31) and the upper bound (35) yields

$$v + \varepsilon_{\max} \;\leq\; f(\widetilde{x}) \;\leq\; v_< + \varepsilon$$

and therefore $\varepsilon_{\max} \leq \varepsilon + v_< - v = \varepsilon + \sigma$ must hold. This contradicts the assumption $\varepsilon_{\max} > \varepsilon + \sigma$, so $\ell_{\psi_{\varepsilon_{\max}}}(\check{Y}^k, X^k) < 0$ cannot hold in infinitely many iterations. □

As mentioned in the previous sections, v is finite exactly for $\Omega \neq \emptyset$, and $v_<$ is finite exactly for $\operatorname{cl}\Omega_< \neq \emptyset$. Obviously, the finiteness of $v_<$ implies the finiteness of v. In the following remark, we will briefly discuss the convergence behavior for inconsistency of $\operatorname{cl}\Omega_<$, i.e. for $v_< = +\infty$.

Remark 4.2. *Firstly, if* $\Omega_< = \emptyset$ *and* $\Omega = \emptyset$ *holds, we have* $\sigma = \infty - \infty = 0$ *and* $\varepsilon_{\max} > \varepsilon + \sigma$ *can easily be accomplished. The arguments from Case 1 of Theorem 4.1 show that for sufficiently small* $\delta_{\max}$*, also* $\Omega_{\delta_{\max}} = \emptyset$ *holds and Algorithm 1 returns an empty list* $\mathcal{O}$ *after finitely many steps and thus terminates (see* (22)*). Otherwise, if* $\Omega_{\delta_{\max}} \neq \emptyset$*, there is no statement possible, as not all boxes get discarded by Case 1, and Case 2 requires an optimal point of* $P_<$*, which does not exist.*

Secondly, if $\Omega_< = \emptyset$ *and* $\Omega \neq \emptyset$ *holds, we have* $\sigma = +\infty$ *and* $\varepsilon_{max} > \varepsilon + \sigma$ *is always violated. However, the arguments in Case 1 of this theorem again imply that* $\delta_{\max}$*-feasibility is enforced by Algorithm 1. Thus*

$$\mathcal{R}_{\varepsilon,\delta} = \Omega_\delta \subseteq \bigcup_{X' \in \mathcal{O}} X' \subseteq \Omega_{\delta_{\max}} = \mathcal{R}_{\varepsilon_{\max},\delta_{\max}}$$

must hold after finitely many steps. Note that the identities follow from Lemma 2.16. Despite the correctness of the approximation $\bigcup_{X' \in \mathcal{O}} X'$*, Case 2 could still hold infinitely many times so that the convergence is not ensured.*

5 Numerical tests

In this section, we discuss the properties of ICGO along with eight illustrative examples. The algorithm is implemented in Python 3.9 and the code is available on GitHub and is listed in the Python Package Index [43]. The experiments were conducted on an Intel Core i7-9700K CPU @ 3.60GHz with Linux Mint 20 and 32 GB RAM.

Implementational details

The pseudo-code of ICGO leaves some freedom of design in the concrete implementation. We will address some important topics and refer to [43] for additional technical details.

Firstly, the selection in line 3 from $\mathcal{Z}$ can be performed in various ways and does not affect the termination after a finite number of iterations, as proven in Theorem 4.1. We employ a breadth-first (first in first out) selection, which was found to be superior to a depth-first (last in first out) selection in previous tests. More advanced branching strategies such as strong branching, pseudo-cost branching, or reliability branching could, if adapted to our framework, have a positive influence on efficiency [2].

For the bound computations, the package [43] provides convergent bounding procedures based on direct implementation of interval arithmetic on the function, centered forms, optimal centered forms and convex relaxations with αBB. Therein, it uses the implementation of the interval arithmetic of [50]. The optimal centered forms are used consistently in the experiments. Moreover, we will avoid reoccurring bound computations as much as possible and thus save, e.g., the bounds $\ell_\omega(Y')$ and $\ell_f(Y')$ together with $Y' \in \mathcal{W} \cup \mathcal{O}$. To make the code more efficient, we replaced the checks $X' \in \mathcal{W}$ with a more explicit control flow that intervenes directly at the points where X' is removed from $\mathcal{W}$. In addition, we avoided the use of boolean flags whenever possible, in accordance with Python programming principles.

Introduction of the test set

As in Section 2.2, we set $\omega(x) = \max_{j \in J} \omega_j(x)$ with finitely many differentiable functions. The test problems (TP1-6) are defined as follows, with $X = [0, 4.5]^2$ throughout:

$$
\begin{aligned}
TP1: f(x) &= x_1 + x_2,\\
\omega_1(x) &= -(x_1^2 + x_2^2) + 6.5,\\
\omega_2(x) &= -x_1 + x_2 - 2,\\
\omega_3(x) &= x_1 - x_2 - 2,\\
\omega_4(x) &= x_1^2 + x_2^2 - 16.
\end{aligned}
$$

$$
\begin{aligned}
TP2: f(x) &= \tfrac{8}{10}((x_1-2)^2 + (x_2-2)^2)\\
&\quad - \tfrac{1}{20}((x_1-2)^2 + (x_2-2)^2)^3,\\
\omega_1(x) &= (x_1-3)^3 + x_2 - 3,\\
\omega_2(x) &= -x_1 + x_2 - 2,\\
\omega_3(x) &= x_1 - x_2 - 2,\\
\omega_4(x) &= 1 - \log((x_1 + \tfrac{1}{2})(x_2 + \tfrac{1}{2})).
\end{aligned}
$$

$$
\begin{aligned}
TP3: f(x) &= (x_1-2)^2 + (x_2-1)^2,\\
\omega_1(x) &= -(x_1+1)^2 + x_2,\\
\omega_2(x) &= -(x_1-2)^2 + x_2,\\
\omega_3(x) &= -(x_1-5)^2 + x_2,\\
\omega_4(x) &= 1 - x_2.
\end{aligned}
$$

$$
\begin{aligned}
TP4.1: f(x) &= -\tfrac{1}{2}(x_1^2 + x_2^2),\\
\omega_1(x) &= (x_1-2)^2 + (x_2-1)^2 - 4,\\
\omega_2(x) &= -\tfrac{1}{3}(x_2-4)^2 + x_1,\\
\omega_3(x) &= -x_2 + 1.
\end{aligned}
$$

$$
\begin{aligned}
TP4.2: f(x) &= -\tfrac{1}{2}(x_1^2 + x_2^2),\\
\omega_1(x) &= (x_1-2)^2 + (x_2-1)^2 - 4,\\
\omega_2(x) &= \tfrac{1}{9}(x_1-3)^3 - 1 + x_2,\\
\omega_3(x) &= -x_2 + 1.
\end{aligned}
$$

$$
\begin{aligned}
TP5: f(x) &= x_1 + x_2,\\
\omega_1(x) &= -(x_1-1)^2 + x_2 - 1,\\
\omega_2(x) &= x_1 - x_2,\\
\omega_3(x) &= (x_1-2)^2 + (x_2-2)^2 - 2.
\end{aligned}
$$

$$
\begin{aligned}
TP6.1: f(x) &= x_1 + x_2 - 3,\\
\omega_1(x) &= 1 - (x_1-2)^2 - \tfrac{1}{3}(x_2-1)^2\\
\omega_2(x) &= 1 - f(x)\\
\omega_3(x) &= \tfrac{1}{2} - x_1,\ \omega_4(x) = 1 - x_2.
\end{aligned}
$$

$$
\begin{aligned}
TP6.2: f(x) &= (x_1-2)^2 + \tfrac{1}{3}(x_2-1)^2,\\
\omega_1(x) &= 1 - f(x),\\
\omega_2(x) &= 4 - x_1 - x_2,\\
\omega_3(x) &= \tfrac{1}{2} - x_1,\ \omega_4(x) = 1 - x_2.
\end{aligned}
$$

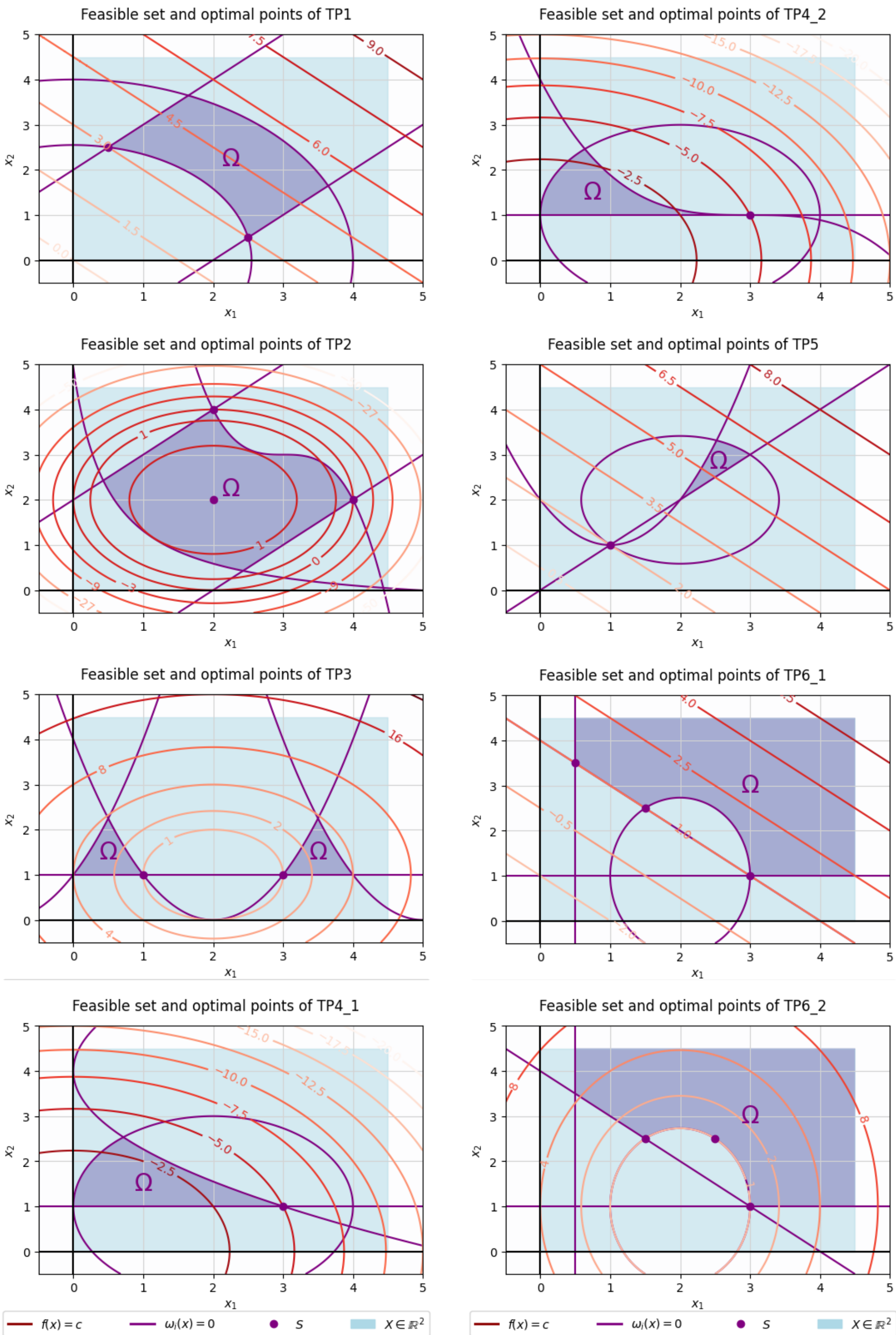

Figure 3: Feasible set and optimal points for TP1-6.

The feasible set and the optimal points are shown in Figure 3. We can observe that each of these problems has a non-convex feasible set Ω, the one of TP3 is even disconnected. Moreover, TP2-5 also have non-convex objective functions. As the aim of this paper is to approximate the *all global* minimal points, we defined test problems with non-unique optimal points. TP1 and TP3 have two optimal points at the boundary of Ω and TP2 has three optimal points, where one is in the interior and two at the boundary. Finally, TP6.1 and 6.2 have infinitely many optimal points, which are located on the circular or diagonal constraint, respectively (through the three plotted dots).

Furthermore, we also study two forms of degeneracy in our test problems. For TP4.2, the MFCQ is violated at the unique optimal point, thus Assumption 2.12 is violated. On the other hand, the SSC holds for this instance and the Assumption 2.9 holds. Therefore, we expect ICGO to approximate the solution set correctly. For TP5, even the SSC is violated at the unique optimal point, thus Assumption 2.9 is violated. Here, the optimal point of the interior problem $TP5_<$ is $(2, 2)$ and $v_< = 4 > 2 = v$, which implies $\sigma = 2$ and $\{(1,1)^\intercal, (2,2)^\intercal\} = \mathcal{R} \neq \mathcal{S}$. By Theorem 4.1, we can only expect a termination for $\varepsilon_{\max} > \varepsilon + 2$.

Discussion of the results

In the experiment, we solved TP1-TP6 with tolerances $\varepsilon, \delta = 0$, $\varepsilon_{\max} \in \{0.1 + \sigma, 0.5 + \sigma\}$, $\delta_{\max} \in \{0.1, 0.5\}$ and a time limit of 240 seconds. Figures 4, 5 and 6 illustrate the output of $\varepsilon_{\max} = \sigma + 0.5$ and $\delta_{\max} = 0.5$, for which ICGO always terminated within the iteration limit. This coarse tolerance allows us to identify the sets $\mathcal{R}_{\varepsilon_{\max},\delta_{\max}}$ and discuss the properties of this form of approximation. In these figures, we can also observe the approximation process. All boxes contained in $\mathcal{O} \cup \mathcal{W}$ are plotted after approximately one third (left) and two thirds (middle) of the iterations and after the termination (right). In order to discuss the correctness of the output, we firstly concentrate on the right images. With help of the level lines, we observe that the output meets the tolerance for optimality and therefore only contains points that are at most '0.5-suboptimal' with respect to $v_<$. The tolerance on feasibility is also fulfilled throughout, we exemplary plotted the level lines of $\omega_i(x) - \delta_{\max} = 0$ (dashed) for TP3 but we refrained from doing so in the remaining graphics for the sake of clarity.

Nevertheless, we note that the tolerance $\delta_{\max}$ can result in a significant portion of the boxes in $\mathcal{O}$ containing infeasible points. The approximations of

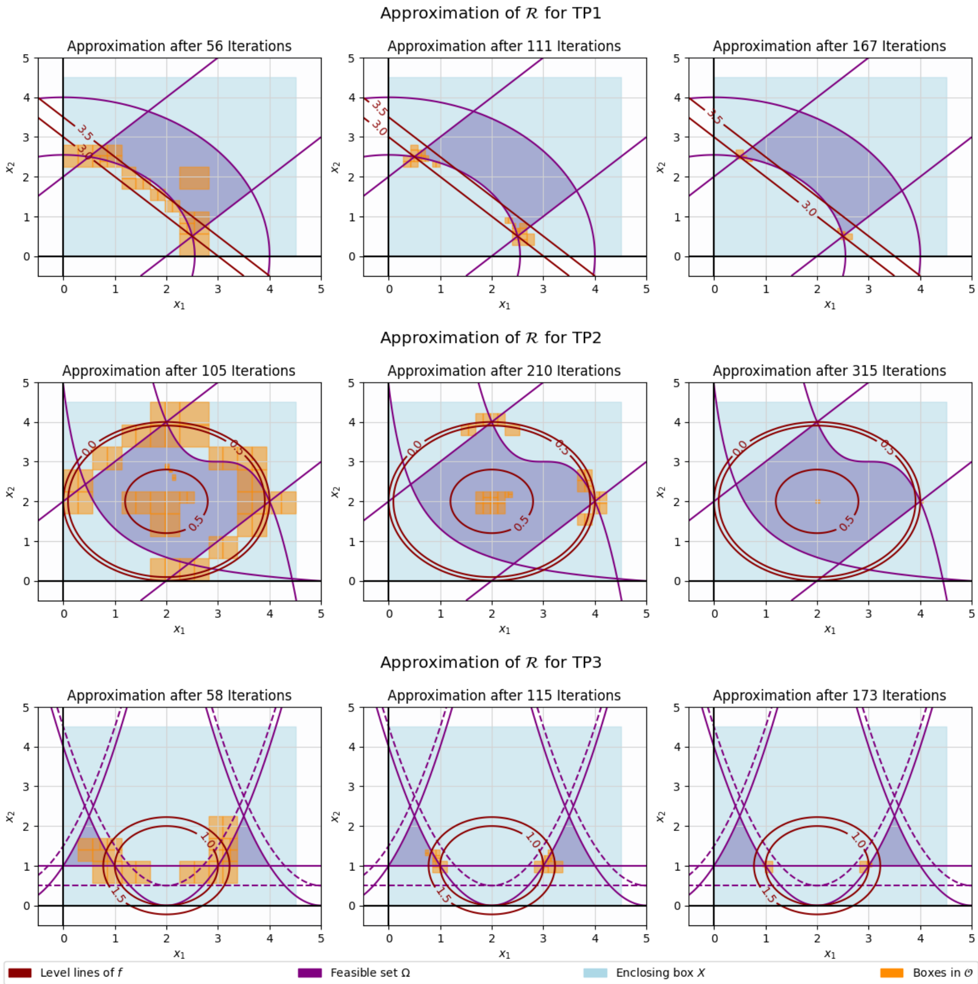

Figure 4: Approximation process of ICGO for TP1, TP2 and TP3.

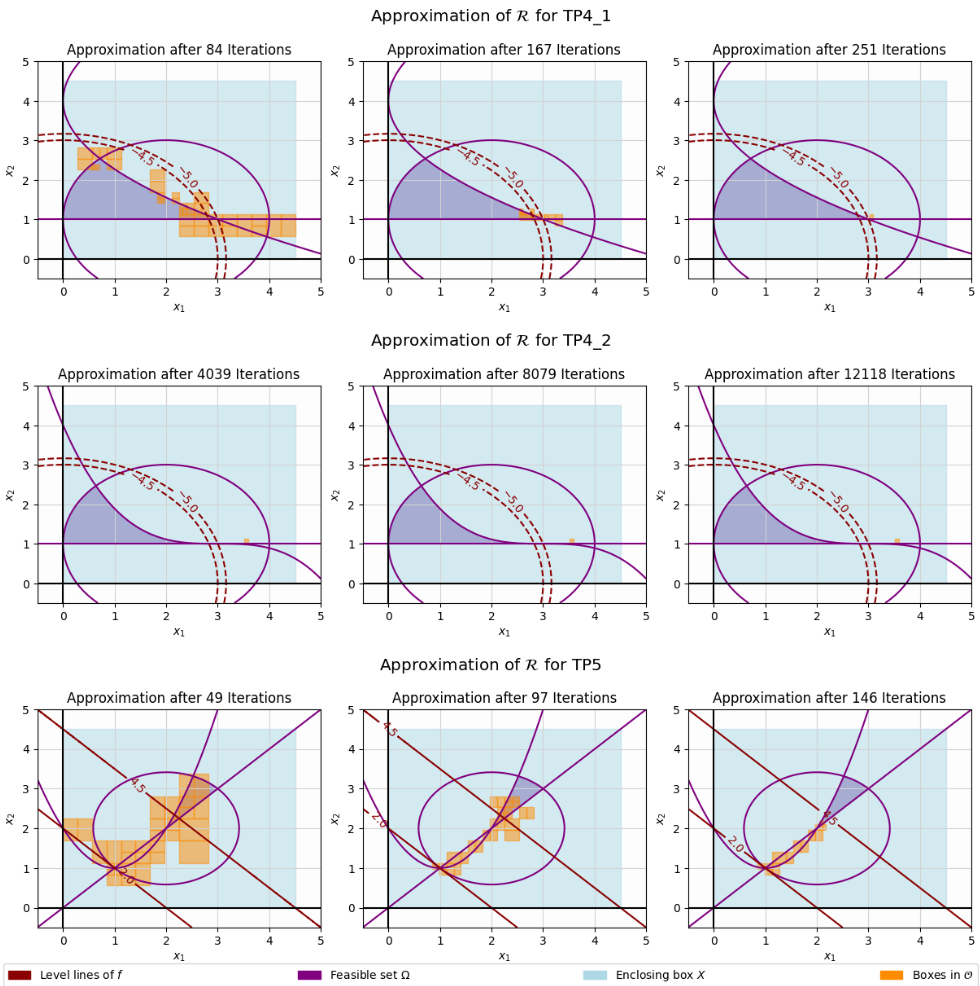

Figure 5: Approximation process of ICGO for TP4.1, TP4.2 and TP5.

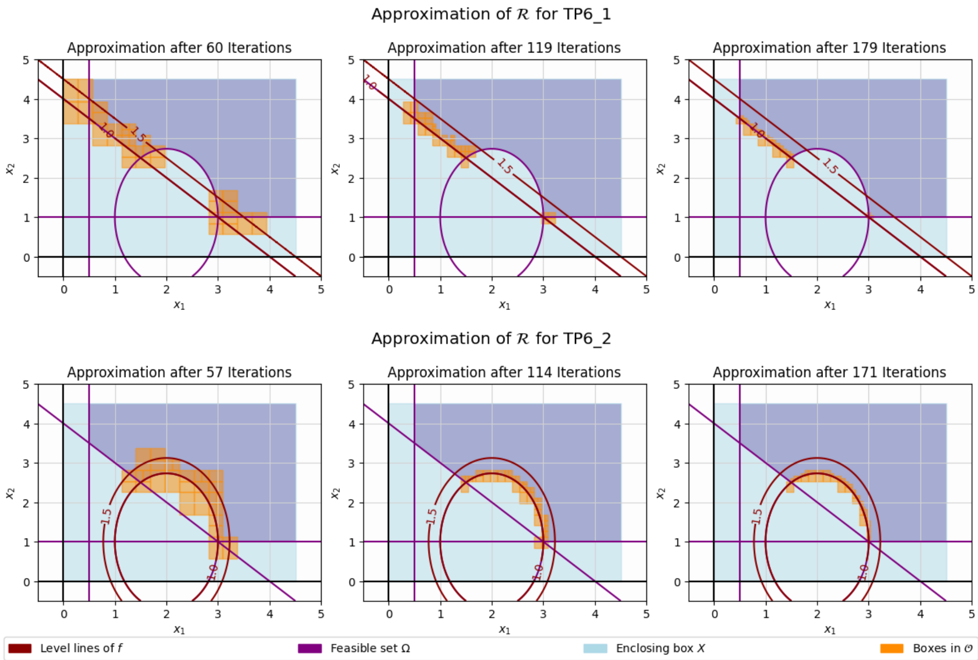

Figure 6: Approximation process of ICGO for TP6.1 and TP6.2.

TP5 and TP6.2 highlight this consequence and underscore the importance of carefully selecting $\delta_{\max}$. Note that, by definition, these points from $\Omega_{\delta_{\max}} \setminus \Omega$ may also have objective values smaller than v. For TP5, ICGO provides an accurate outer approximation of $\mathcal{R}$. We remark that, while the spurious point $(2,2)^\intercal$ is not the correct minimal point in this example, a practitioner may prefer it over the correct one, since it is stable in the sense that the MFCQ is satisfied there.

	$\varepsilon_{\max} = \sigma + 0.5, \delta_{\max} = 0.5$					$\varepsilon_{\max} = \sigma + 0.5, \delta_{\max} = 0.1$					$\varepsilon_{\max} = \sigma + 0.1, \delta_{\max} = 0.1$				
Name	$\varepsilon_{\max}$	t	k	$\lvert\mathcal{O}\rvert$	$\widetilde{v}_< - v_<$	$\varepsilon_{\max}$	t	k	$\lvert\mathcal{O}\rvert$	$\widetilde{v}_< - v_<$	$\varepsilon_{\max}$	t	k	$\lvert\mathcal{O}\rvert$	$\widetilde{v}_< - v_<$
TP1	0.5	0.09	167	11	2.3e-02	0.5	0.13	231	9	1.5e-02	0.1	0.13	244	10	1.5e-02
TP2	0.5	0.37	315	10	2.4e-05	0.5	0.37	320	10	2.4e-05	0.1	0.42	365	11	3.8e-07
TP3	0.5	0.13	173	9	8.9e-02	0.5	0.11	167	6	1.1e-01	0.1	0.15	226	9	4.1e-02
TP4.1	0.5	0.13	251	7	5.9e-02	0.5	0.11	217	7	1.7e-01	0.1	0.19	366	9	2.7e-03
TP4.2	0.5	88.57	35074	157	9.6e-02	0.5	88.67	35064	159	9.6e-02	0.1	240.01	78079	33	4.9e-02
TP5	2.5	0.08	146	14	2.2e-01	2.5	0.13	262	18	7.8e-02	2.1	0.16	282	14	3.4e-02
TP6.1	0.5	0.11	179	25	7.8e-03	0.5	0.16	247	53	7.8e-03	0.1	0.49	805	160	7.8e-03
TP6.2	0.5	0.12	171	33	8.1e-03	0.5	0.17	241	50	6.9e-03	0.1	2.02	2454	625	8.2e-06

Table 1: Solving statistics of TP1-6 with $\delta_{\max}, \varepsilon_{\max} \in \{0.5, 0.1\}$.

In the left images of Figure 4, 5 and 6, we observe that δ-infeasible boxes are discarded relatively quickly by the if-statement in line 5. In contrast, the images in the middle show which areas of the feasible set pose difficulties and are discarded relatively late. Let us explore the potential explanations for this. Firstly, some boxes are on the boundary of the feasible set and realize low objective function values outside the feasible set, like, e.g. the boxes in the center for TP4.1. Secondly, a steep ascent of the objective function towards the optimum points requires very small boxes to fulfill the inclusion property. This effect can, for example, be observed in the two optimal points at the boundary for TP2, where the level lines are very close and the approximation becomes invisibly small in the right picture. Lastly, the geometry of Ω also has a strong influence. If the optimum point lies in a very narrow peak, the boxes must also be refined in many steps. This is illustrated in TP4.1 and TP4.2, with TP4.2 presenting the worst case where the MFCQ is violated.

In Table 1 the following solving statistics are displayed. The runtime t, the total number of iterations k, and the length of the list $\mathcal{O}$ after termination. In addition, $\widetilde{v}_< - v_<$ indicates the accuracy of the incumbent $\widetilde{y}$. The columns on the left with $\varepsilon_{\max}, \delta_{\max} = 0.5, 0.5$ show the statistics for the run with which Figures 4, 5 and 6 were generated. In the other columns, we see how the data changes, when first only $\delta_{\max}$ is scaled down by factor five and second also $\varepsilon_{\max}$ is scaled down by factor five. It is worth noticing that $\delta_{\max} \leq \varepsilon_{\max}$ must apply.

Across all runs, ICGO almost always terminates within a half second, with TP4.2 and TP6.2 being the only exceptions. Instance TP4.2 is the most computationally demanding and does not finish within the time limit when $\varepsilon_{\max}$ and $\delta_{\max}$ are both set to 0.1. Considering the length of $\mathcal{O}$, we observe that for TP1-4.1 and TP5 it is less than 14 and does not increase remarkably when tolerances are reduced. On the other hand, for TP4.2 and TP6.1-6.2 the list $\mathcal{O}$ is longer, and, for the latter ones, it increases noticeably when the tolerance $\varepsilon_{\max}$ is reduced (for TP4.2 we cannot draw conclusions as ICGO did not terminate). This difference seems to depend on whether $\mathcal{R}$ or $\mathcal{R}_{\varepsilon_{\max},\delta_{\max}}$ collapse to single points or describe a shape that needs to be approximated by $\mathcal{O}$. In the case of TP4.2, $\mathcal{R}_{\varepsilon_{\max},\delta_{\max}}$ covers a line and in the case of TP6.1/.2, already $\mathcal{R}$ is linear/circular shaped.

When the tolerance $\delta_{\max}$ is reduced by factor five, the run time increases (in the mean) by 20% and the time per iteration remains nearly constant with a mean of 0.9 milliseconds per iteration. When the tolerance $\varepsilon_{\max}$ is then additionally reduced by factor five, the run time increases (in the mean) by

206%, neglecting the outlier TP4.2 that reached the time limit. For instances TP4.2, TP5, and TP6.2, the time per iteration increased significantly, ranging from 14% to 22%, while for other instances it remained nearly constant.

For all instances, the precision of $\widetilde{v}_<$ is always below 0.22, given $\varepsilon_{\max} = 0.5$ and below 0.05, given $\varepsilon_{\max} = 0.1$. Naturally, incumbents must be better than this tolerance to discard all '$\varepsilon_{\max}$-suboptimal' areas in line 13. However, for TP2 and TP6, ICGO delivers incumbents with very high precision, compared to $\varepsilon_{\max}$.

To conclude, we suspect that the search operations inside $\mathcal{W} \cup \mathcal{O}$ account for a large portion of the complexity. The search operations fulfill two functions. Firstly, find a good incumbent in line 10 and secondly, compute a lower bound in line 21 to prove the inclusion inside $\mathcal{R}_{\varepsilon_{\max},\delta_{\max}}$. The first could be enhanced by incorporating local optimization techniques or heuristics. The second might be improved by advanced list structures that enable a more efficient search.

6 Final remarks

The presented spatial branch-and-bound approach for complete nonconvex optimization relies on a reformulation with the improvement function that preserves the complete set of optimal points under very mild conditions. Furthermore, the effect of degeneracies, the additional approximation of so-called spurious points, was described and discussed in detail. Loosely speaking, the potential existence of spurious points in $\mathcal{R}$ is caused by the fact that the improvement of some $s \in \Omega$ by $x \in X$ requires both the nonstrict inequality $\omega(x) \le 0$ and the strict inequality $f(x) < f(s)$, which the improvement function can neither model by $\psi(x,s) \le 0$ nor by $\psi(x,s) < 0$.

A similar effect occurs in the approach from [23, 37, 49], which reformulates generalized semi-infinite constraints as standard semi-infinite constraints. In fact, for continuous functions f and ω as well as a compact set X and $\Omega(s) = \{x \in X \mid \omega(x,s) \le 0\}$ the constraint

$$\min_{x \in \Omega(s)} f(x,s) \ge 0 \tag{36}$$

is called generalized semi-infinite. A point s satisfies this constraint if and only if for all $x \in X$ the strict constraint $\omega(x,s) > 0$ or the nonstrict constraint $f(x,s) \ge 0$ holds. With $\psi(x,s) := \max\{\omega(x,s), f(x,s)\}$ the standard semi-infinite constraint

$$\min_{x \in X} \psi(x, s) \geq 0 \tag{37}$$

thus relaxes (36). Therefore the sets $F = \{s \text{ with } (36)\}$ and $\Psi = \{s \text{ with } (37)\}$ satisfy $F \subseteq \Psi$. While Ψ is closed, F may not be closed, but [23] shows that for generic functions f and ω at least the identity $\mathrm{cl}\, F = \Psi$ is true. In [37] this fact is used to solve a generalized semi-infinite optimization problem with feasible set F by the reformulation to the standard semi-infinite problem with the same objective function, but feasible set Ψ. We conjecture that, in absence of the genericity assumption, it is possible to use the techniques of the present paper to also characterize the set of spurious points $\Psi \setminus \mathrm{cl}\, F$. We leave this question for future research.

7 Data availability

All data used in the numerical experiments are explicitly provided in the manuscript. The code used to generate the results and figures is available at [43].

References

[1] Achterberg, T.: SCIP: Solving constraint integer programs. Mathematical Programming Computation **1**(1), 1–41 (2009). doi:10.1007/s12532-008-0001-1

[2] Achterberg, T., Koch, T., Martin, A.: Branching rules revisited. Operations Research Letters **33**(1), 42–54 (2005). doi:10.1016/j.orl.2004.04.002

[3] Adjiman, C.S., Androulakis, I.P., Floudas, C.A.: A global optimization method, αBB, for general twice-differentiable constrained NLPs—II. Implementation and computational results. Computers & Chemical Engineering **22**(9), 1159–1179 (1998). doi:10.1016/S0098-1354(98)00218-X

[4] Adjiman, C.S., Dallwig, S., Floudas, C.A., Neumaier, A.: A global optimization method, αBB, for general twice-differentiable constrained NLPs — I. Theoretical advances. Computers & Chemical Engineering **22**(9), 1137–1158 (1998). doi:10.1016/S0098-1354(98)00027-1

[5] Androulakis, I.P., Maranas, C.D., Floudas, C.A.: αBB: A global optimization method for general constrained nonconvex problems. Journal of Global Optimization **7**(4), 337–363 (1995). doi:10.1007/BF01099647

[6] Araya, I., Trombettoni, G., Neveu, B., Chabert, G.: Upper bounding in inner regions for global optimization under inequality constraints. Journal of Global Optimization **60**(2), 145–164 (2014). doi:10.1007/s10898-014-0145-7

[7] Bagirov, A., Karmitsa, N., Mäkelä, M.M.: Introduction to Nonsmooth Optimization: Theory, Practice and Software. Springer International Publishing, Cham (2014). doi:10.1007/978-3-319-08114-4

[8] Baumann, E.: Optimal centered forms. BIT Numerical Mathematics **28**(1), 80–87 (1988). doi:10.1007/BF01934696

[9] Belotti, P., Lee, J., Liberti, L., Margot, F., Wächter, A.: Branching and bounds tightening techniques for non-convex MINLP. Optimization Methods & Software **24**(4-5), 597–634 (2009)

[10] Bonnans, J.F., Shapiro, A.: Perturbation Analysis of Optimization Problems. Springer Science & Business Media (2000)

[11] Dakin, R.J.: A tree-search algorithm for mixed integer programming problems. The Computer Journal **8**(3), 250–255 (1965). doi:10.1093/comjnl/8.3.250

[12] Dempe, S.: Foundations of Bilevel Programming, *Nonconvex Optimization and Its Applications*, vol. 61. Kluwer Academic Publishers, Boston (2002). doi:10.1007/b101970

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

[14] Dür, M.: Dual bounding procedures lead to convergent Branch–and–Bound algorithms. Mathematical Programming **91**(1), 117–125 (2001). doi:10.1007/s101070100236

[15] Dür, M.: A class of problems where dual bounds beat underestimation bounds. Journal of Global Optimization **22**(1), 49–57 (2002). doi:10.1023/A:1013890609372

[16] Ehrgott, M.: Multicriteria optimization, vol. 491. Springer Science & Business Media (2005)

[17] Eichfelder, G., Gerlach, T., Sumi, S.: A modification of the αBB method for box-constrained optimization and an application to inverse kinematics. EURO Journal on Computational Optimization **4**(1), 93–121 (2016). doi:10.1007/s13675-015-0056-5

[18] Facchinei, F., Kanzow, C.: Generalized Nash equilibrium problems. Annals of Operations Research **175**(1), 177–211 (2010). doi:10.1007/s10479-009-0653-x

[19] Falk, J.E., Soland, R.M.: An algorithm for separable nonconvex programming problems. Management Science **15**(9), 550–569 (1969)

[20] Floudas, C.A.: Deterministic Global Optimization, *Nonconvex Optimization and Its Applications*, vol. 37. Springer US, Boston, MA (2000). doi:10.1007/978-1-4757-4949-6

[21] Füllner, C., Kirst, P., Stein, O.: Convergent upper bounds in global minimization with nonlinear equality constraints. Mathematical Programming **187**(1-2), 617–651 (2021). doi:10.1007/s10107-020-01493-2

[22] Gourdin, E., Jaumard, B., MacGibbon, B.: Global Optimization Decomposition Methods for Bounded Parameter Minimax Risk Evaluation. SIAM Journal on Scientific Computing **15**(1), 16–35 (1994). doi:10.1137/0915002

[23] Günzel, H., Jongen, H.T., Stein, O.: On the closure of the feasible set in generalized semi-infinite programming. Central European Journal of Operations Research **15**(3), 271–280 (2007). doi:10.1007/s10100-007-0030-2

[24] Hansen, E.R., Walster, G.W.: Global Optimization Using Interval Analysis, 2nd ed, rev. and expanded edn. No. 264 in Pure and Applied Mathematics. Dekker, New York (2004)

[25] Hansen, P., Jaumard, B.: Lipschitz Optimization. In: R. Horst, P.M. Pardalos (eds.) Handbook of Global Optimization, pp. 407–493. Springer US, Boston, MA (1995). doi:10.1007/978-1-4615-2025-2_9

[26] Hiriart-Urruty, J.B., Lemaréchal, C.: Convex Analysis and Minimization Algorithms I: Fundamentals, vol. 305. Springer science & business media (1996)

[27] Horst, R., Tuy, H.: Global Optimization: Deterministic Approaches, 3rd ed., rev. and enl edn. Springer, Berlin ; New York (1996)

[28] Iusem, A.N., Sosa, W.: New existence results for equilibrium problems. Nonlinear Analysis: Theory, Methods & Applications **52**(2), 621–635 (2003)

[29] Kirst, P., Füllner, C.: On the use of restriction of the right-hand side in spatial branch-and-bound algorithms to ensure termination. Computational Optimization and Applications (2025). doi:10.1007/s10589-025-00652-5

[30] Kirst, P., Stein, O., Steuermann, P.: Deterministic upper bounds for spatial branch-and-bound methods in global minimization with nonconvex constraints. TOP **23**(2), 591–616 (2015). doi:10.1007/s11750-015-0387-7

[31] Krawczyk, R., Nickel, K.: Die zentrische Form in der Intervallarithmetik, ihre quadratische Konvergenz und ihre Inklusionsisotonie. Computing **28**(2), 117–137 (1982). doi:10.1007/BF02241818

[32] Land, A.H., Doig, A.G.: An automatic method of solving discrete programming problems. Econometrica **28**(3), 497–520 (1960)

[33] Lin, Y., Schrage, L.: The global solver in the LINDO API. Optimization Methods & Software **24**(4-5), 657–668 (2009)

[34] Lundell, A., Kronqvist, J., Westerlund, T.: The supporting hyperplane optimization toolkit for convex MINLP. Journal of Global Optimization **84**(1), 1–41 (2022). doi:10.1007/s10898-022-01128-0

[35] McCormick, G.P.: Computability of global solutions to factorable nonconvex programs: Part I—Convex underestimating problems. Mathematical Programming **10**(1), 147–175 (1976)

[36] Misener, R., Floudas, C.A.: ANTIGONE: algorithms for continuous/integer global optimization of nonlinear equations. Journal of Global Optimization **59**(2), 503–526 (2014)

[37] Mitsos, A., Tsoukalas, A.: Global optimization of generalized semi-infinite programs via restriction of the right hand side. Journal of Global Optimization **61**(1), 1–17 (2015). doi:10.1007/s10898-014-0146-6

[38] Neumaier, A.: Interval Methods for Systems of Equations. Cambridge university press (1990)

[39] Neumaier, A., Shcherbina, O., Huyer, W., Vinkó, T.: A comparison of complete global optimization solvers. Mathematical Programming **103**, 335–356 (2005)

[40] Ninin, J.: Global optimization based on contractor programming: An overview of the IBEX library. In: I.S. Kotsireas, S.M. Rump, C.K. Yap (eds.) Mathematical Aspects of Computer and Information Sciences, *Lecture Notes in Computer Science*, vol. 9582, pp. 555–559. Springer, Cham (2016). doi:10.1007/978-3-319-32859-1_47

[41] Oettli, W., Blum, E.: From optimization and variational inequality to equilibrium problems. Mathematics Student **63**, 123–145 (1994)

[42] Piyavskii, S.A.: An algorithm for finding the absolute extremum of a function. USSR Computational Mathematics and Mathematical Physics **12**(4), 57–67 (1972). doi:10.1016/0041-5553(72)90115-2

[43] Rodestock, M.: pyimpBB - A branch-and-bound method using the improvement function in Python. `https://github.com/uisab/pyimpBB`. The Python Package Index (PyPI), accessed: 2025-12-16

[44] Ryoo, H.S., Sahinidis, N.V.: Global optimization of nonconvex NLPs and MINLPs with applications in process design. Computers & Chemical Engineering **19**(5), 551–566 (1995)

[45] Ryoo, H.S., Sahinidis, N.V.: A branch-and-reduce approach to global optimization. Journal of Global Optimization **8**, 107–138 (1996)

[46] Sahinidis, N.V.: BARON: A general purpose global optimization software package. Journal of Global Optimization **8**, 201–205 (1996)

[47] Smith, E.M., Pantelides, C.C.: Global optimisation of nonconvex MINLPs. Computers & Chemical Engineering **21**, S791–S796 (1997)

[48] Smith, E.M., Pantelides, C.C.: A symbolic reformulation/spatial branch-and-bound algorithm for the global optimisation of nonconvex MINLPs. Computers & Chemical Engineering **23**(4-5), 457–478 (1999)

[49] Stein, O.: A semi-infinite approach to design centering. In: S. Dempe, V. Kalashnikov (eds.) Optimization with Multivalued Mappings: Theory, Applications, and Algorithms, pp. 209–228. Springer US, Boston, MA (2006). doi:10.1007/0-387-34221-4_10

[50] Taschini, S.: PyInterval — interval arithmetic in Python. https://github.com/taschini/pyinterval. The Python Package Index (PyPI), accessed: 2025-04-01

[51] Tawarmalani, M., Sahinidis, N.V.: Convexification and Global Optimization in Continuous and Mixed-Integer Nonlinear Programming, *Nonconvex Optimization and Its Applications*, vol. 65. Springer US, Boston, MA (2002). doi:10.1007/978-1-4757-3532-1

[52] Tawarmalani, M., Sahinidis, N.V.: A polyhedral branch-and-cut approach to global optimization. Mathematical Programming **103**(2), 225–249 (2005)

[53] Thach, P.T., Tuy, H.: The relief indicator method for constrained global optimization. Naval Research Logistics (NRL) **37**(4), 473–497 (1990)

[54] Tsoukalas, A., Rustem, B.: A feasible point adaptation of the Blankenship and Falk algorithm for semi-infinite programming. Optimization Letters **5**(4), 705–716 (2011). doi:10.1007/s11590-010-0236-4

[55] Zamora, J.M., Grossmann, I.E.: A branch and contract algorithm for problems with concave univariate, bilinear and linear fractional terms. Journal of Global Optimization **14**, 217–249 (1999)

A Additional Lemmata

Lemma A.1. *Given a set $A \subseteq \mathbb{R}^2$, let $a^\star \in A$ solve*

$$\min_{a \in A} \max\{a_1, a_2\}$$

and let $a_1^\star - a_2^\star \geq \varepsilon > 0$. Then $a^\star$ also solves

$$\min_{a \in A} \max\{a_1, a_2 + \varepsilon\}.$$

Proof. We prove the above statement by contradiction. Assume there exists a point $a' \in A$ with

$$\max\{a'_1, a'_2 + \varepsilon\} < \max\{a^\star_1, a^\star_2 + \varepsilon\}.$$

By optimality of $a^\star$, $a_1^\star \leq \max\{a_1^\star, a_2^\star\} \leq \max\{a'_1, a'_2\}$ must also hold. As $\varepsilon > 0$, additionally $\max\{a'_1, a'_2\} \leq \max\{a'_1, a'_2 + \varepsilon\}$ holds. Together we obtain $a_1^\star < \max\{a_1^\star, a_2^\star + \varepsilon\}$. If the maximum is attained by the first argument, we obtain the contradiction $a_1^\star < a_1^\star$. If it is attained by the second argument, we obtain $a_1^\star < a_2^\star + \varepsilon$, which is in contradiction to $a_1^\star - a_2^\star \geq \varepsilon$. □

Lemma A.2. *Let* (X^k) *be an infinite sequence of pairwise distinct boxes generated by ICGO. Then there exists an infinite weakly exhaustive subsequence of* (X^k).

Proof. Consider the branch-and-bound tree generated by ICGO in the following process: After initializing the tree with one root node, at each iteration $k \in \mathbb{N}$:

1. one leaf node X^k and at most one other distinct leaf node Y^k are selected
2. if specific conditions hold, finitely many child nodes are added to X^k (a partition of it)
3. if some $Y^k \neq X^k$ was selected and specific conditions hold, finitely many child nodes are added to Y^k (a partition of it)

Naturally, any node has only finitely many child nodes. The tree is a locally finite graph [13] and the infinite sequence of distinct boxes corresponds to an infinite set of nodes in the graph. Using the Star-Comb Lemma [13, Lem. 8.2.2] we may define a comb in the graph such that an infinite subsequence of (X^k) lies on that comb. By definition of a comb (see [13, Sec. 8.1]), this means that there exists an infinite nested sequence (R^k) with $R^{k+1} \subseteq R^k$ and $X^k \subseteq R^k$ for all $k \in \mathbb{N}$ (a so-called 'ray'). In view of the partitioning rule **(PR)**, we have $\lim_k \operatorname{diag}(R^k) = 0$. Thus, (R^k) is an exhaustive sequence of boxes and, therefore, the infinite subsequence (X^k) is weakly exhaustive. $\square$