

Optimality-Based Discretization Methods for the Global Optimization of Nonconvex Semi-Infinite Programs

Evren M. Turan¹, Johannes Jäschke¹, and Rohit Kannan²

¹Department of Chemical Engineering, Norwegian University of Science and Technology (NTNU), Trondheim, Norway. E-mail: evren.m.turan@ntnu.no, johannes.jaschke@ntnu.no

²Center for Nonlinear Studies (T-CNLS) and Applied Mathematics & Plasma Physics (T-5), Los Alamos National Laboratory, Los Alamos, NM, USA. E-mail: rohitk@alum.mit.edu

February 28, 2023

Abstract

We use sensitivity analysis to design *optimality-based* discretization (cutting-plane) methods for the global optimization of nonconvex semi-infinite programs (SIPs). We begin by formulating the optimal discretization of SIPs as a max-min problem and propose variants that are more computationally tractable. We then use parametric sensitivity theory to design an efficient method for solving these max-min problems to local optimality and argue this yields valid discretizations without sacrificing global optimality guarantees. Finally, we formulate optimality-based *generalized* discretization of SIPs as max-min problems and design efficient local optimization algorithms to solve them approximately. Numerical experiments on test instances from the literature demonstrate that our new optimality-based discretization methods can significantly reduce the number of iterations for convergence relative to the classical feasibility-based method.

Key words: Semi-infinite programming, Robust optimization, Discretization, Global optimization, Cutting-planes, Sensitivity analysis

1 Introduction

Semi-infinite programs (SIPs) are mathematical optimization problems with a finite number of decision variables and an infinite number of constraints, with the infinite set of constraints in a SIP typically parametrized using a finite-dimensional parameter. They can be used to model several problems in science and engineering such as robust optimization, controller design, Chebyshev approximation, and design centering [13, 19, 25, 51]. There are many surveys and books on semi-infinite programming, see, e.g., [13, 18, 21, 25, 36, 38, 44, 45] for theory, algorithms, and applications. Our focus in this paper is on the *global* optimization of SIPs.

We consider the semi-infinite program

$$\begin{aligned} v^* &:= \min_{x \in X} f(x) && \text{(SIP)} \\ \text{s.t. } &g(x, y) \leq 0, \quad \forall y \in Y, && (1) \end{aligned}$$

where $X \subset \mathbb{R}^{d_x}$ and $Y \subset \mathbb{R}^{d_y}$ are nonempty compact sets, $|Y| = \infty$, and functions $f : \mathbb{R}^{d_x} \rightarrow \mathbb{R}$ and $g : \mathbb{R}^{d_x} \times \mathbb{R}^{d_y} \rightarrow \mathbb{R}$ are continuous. We assume (SIP) is feasible for simplicity, but do not necessarily assume X , Y , f , or g is convex. While we only consider a scalar semi-infinite constraint (1) for ease of exposition, Section 5 details a generalization to multiple such constraints. A key challenge in solving (SIP) is that even

Algorithm 1 The Blankenship and Falk algorithm [3]

```
1: Input feasibility tolerance  $\varepsilon_f \geq 0$ .
2: Initialize  $Y_d = \emptyset$ .
3: for  $k = 1, 2, \dots$  do
4:   Solve problem (LBP) globally to get solution  $x^k$ , lower bound  $LBD^k$ .
5:   Solve problem (LLP( $x$ )) with  $x = x^k$  globally to get solution  $y^{BF,k} \in Y$ .
6:   if  $G(x^k) \leq \varepsilon_f$  then
7:     Terminate with  $\varepsilon_f$ -feasible solution  $x^k$  to (SIP).
8:   else
9:     Set  $Y_d \leftarrow Y_d \cup \{y^{BF,k}\}$ .
10:  end if
11: end for
```

evaluating feasibility of a candidate solution $x \in X$ requires the *global* solution of the following lower-level problem:

$$G(x) := \max_{y \in Y} g(x, y). \quad (\text{LLP}(x))$$

Clearly, $x \in X$ is feasible for (SIP) if and only if $G(x) \leq 0$. Let $y^* : X \rightarrow Y$ be *any* mapping such that $y^*(x)$ is an optimal solution to (LLP(x)) for each $x \in X$. Then $G(\cdot) \equiv g(\cdot, y^*(\cdot))$ on X and (SIP) can be equivalently written as

$$\begin{aligned} \min_{x \in X} f(x) \\ \text{s.t. } g(x, y^*(x)) \leq 0. \end{aligned} \quad (2)$$

The above reformulation highlights the fact that semi-infinite programming is a special case of bilevel optimization [8, 44].

Several papers in the last two decades propose algorithms for the *global* minimization of nonconvex SIPs [2, 12, 17, 27, 28, 31, 46, 50] (see Djelassi et al. [11, 13] for a recent survey). Most of these approaches construct lower bounds for (SIP) by replacing the semi-infinite constraint (1) with a *finite* discretization

$$\begin{aligned} \min_{x \in X} f(x) \\ \text{s.t. } g(x, y) \leq 0, \quad \forall y \in Y_d, \end{aligned} \quad (\text{LBP})$$

where $Y_d \subsetneq Y$ with $|Y_d| < \infty$. The above lower bounding problem (LBP) is related to the classical cutting-plane method of Cheney and Goldstein [6] and Kelley [24] in convex optimization. A valid lower bound for the optimal value v^* of (SIP) can be obtained by solving this (nonconvex) relaxation to *global* optimality. Additionally, under mild conditions, the optimal value of a sequence of iteratively refined discretizations (LBP) converges to v^* [25].

The choice of discretization Y_d can greatly impact the tightness of the lower bound obtained by solving (LBP). Because naïve discretization approaches may require a large discretization Y_d for the optimal value of (LBP) to approximate v^* well [48], techniques for adaptively populating Y_d are of interest. Most approaches for the global minimization of (SIP) rely on the adaptive *feasibility-based* discretization method of Blankenship and Falk (BF, outlined in Algorithm 1) [3], which populates Y_d with points in Y corresponding to the largest violation of constraint (1) at incumbent solutions of (LBP).

The sequence of non-decreasing lower bounds $\{LBD^k\}$ generated by the BF algorithm converges to v^* under our assumptions on (SIP) (see [3, Theorem 2.1]). However, since the BF algorithm may not generate a feasible point for (SIP) finitely, global optimization algorithms supplement it with techniques for generating feasible points to guarantee finite convergence to a near-optimal solution of (SIP) [28]. Moreover, although the BF algorithm typically determines good lower bounds in a few iterations, it may still require an excessively

Iteration No.	1	2	3	4	5	10	15	20	25	28
Lower Bound	4	4.19	4.38	4.56	4.74	5.62	6.41	7.12	7.73	8

Table 1: Lower bounds generated by the BF Algorithm 1 on Example 1. The algorithm needs more than 20 iterations to approximate $v^* = 8$ to within 10%.

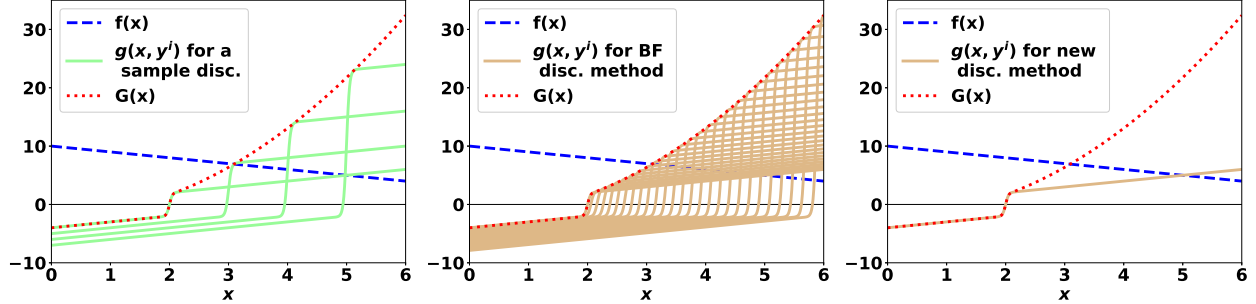


Figure 1: Benefits of optimality-based discretization: Left: objective $f(x)$ and constraint $G(x)$ for Example 1 along with discretization $Y_d = \{2, 3, 4, 5\}$ (cf. [11, Figure 3.1]). Middle: BF method needs 27 discretization points for its lower bounds to converge to v^* . Right: our proposed optimality-based discretization method in Section 3 only requires a single discretization point ($|Y_d| = 1$).

large discretization Y_d before the sequence $\{LBD^k\}$ converges to within a specified tolerance of the optimal value v^* . Example 1 below from Mitsos [29] illustrates the potentially slow convergence of the BF algorithm.

Example 1. [29, Ex. (DP)] Consider (SIP) with $d_x = 1$, $d_y = 1$, $X = [0, 6]$, $Y = [2, 6]$, $f(x) = 10 - x_1$, and $g(x, y) = \frac{y_1^2}{1 + \exp(-40(x_1 - y_1))} + x_1 - y_1 - 2$. The global solution is $x^* = 2$ with objective $v^* = 8$.

Table 1 notes the lower bounding iterates generated by the BF algorithm for Example 1, and Figure 1 plots its discretizations of the semi-infinite constraint for this instance. Clearly, solving (LBP) with the discretization $Y_d = \{2\}$ yields an optimal solution to (SIP) with lower bound equal to v^* . However, the feasibility-based BF algorithm needs more than 20 iterations for its sequence of lower bounds $\{LBD^k\}$ to converge to within 10% of v^* . This motivates our study of new optimality-based discretization methods for (SIP) that can mitigate this slow convergence of the BF algorithm (see Figure 1).

Our main idea in this paper is to populate the discretization Y_d with points in Y such that the lower bound from (LBP) is maximized. Since determining this *optimal discretization* may be challenging, we consider more tractable variants and design efficient solution methods. We also study how to optimally construct *generalized discretizations* and analyze their theoretical guarantees.

This paper is organized as follows. Section 2 reviews definitions and results from parametric sensitivity theory. Section 3 proposes new optimality-based discretization methods, designs efficient approaches for their solution, and presents theoretical guarantees. Section 4 introduces optimality-based generalized discretization methods for (SIP) and designs an efficient solution strategy. We briefly explore some extensions in Section 5. Section 6 presents detailed computational results demonstrating that our optimality-based discretization methods significantly reduce the number of iterations for convergence relative to the BF algorithm. We conclude with avenues for future work in Section 7.

Notation. Let $[n] := \{1, \dots, n\}$ and $\text{mid}(v^1, v^2, v^3)$ denote the componentwise median of vectors v^1, v^2, v^3 . Given $\delta > 0$, $v \in \mathbb{R}^n$, $S \subset \mathbb{R}^n$, let $\|v\|$ denote the Euclidean norm of v , $B_\delta(v)$ denote the open Euclidean ball of radius δ centered at v , $\text{proj}_S(v)$ denote (an element of) the Euclidean projection of v onto S , and $\text{diam}(S)$ denote the diameter of S with respect to the Euclidean norm.

2 Review of parametric sensitivity theory

We begin by reviewing standard definitions and results from parametric sensitivity theory [16, 49]. Consider the parametric nonlinear program (NLP)

$$\begin{aligned} \min_{z \in \mathbb{R}^n} \quad & F(z, p) \\ \text{s.t.} \quad & c_i(z, p) \leq 0, \quad \forall i \in \mathcal{I}, \\ & c_i(z, p) = 0, \quad \forall i \in \mathcal{E}, \end{aligned} \tag{3}$$

where $z \in \mathbb{R}^n$ are decision variables, $p \in \mathbb{R}^d$ are parameters, $F : \mathbb{R}^n \times \mathbb{R}^d \rightarrow \mathbb{R}$ is the objective function, $\mathcal{I} = \{1, \dots, m_I\}$ and $\mathcal{E} = \{m_I + 1, \dots, m_I + m_E\}$ are index sets for the inequality and equality constraints, and $c : \mathbb{R}^n \times \mathbb{R}^d \rightarrow \mathbb{R}^{m_I + m_E}$ are the constraint functions. We write $z^*(p)$ and $\nu^*(p)$ to denote a local minimum of problem (3) and the corresponding value function at this local minimum $\nu^*(p) := F(z^*(p), p)$, respectively. Throughout this section, we omit dependence on p whenever it is clear from the context.

The Lagrangian for problem (3) is $L(z, \lambda, p) := F(z, p) + \lambda^\top c(z, p)$, where $\lambda \in \mathbb{R}^{m_I + m_E}$ are Lagrange multipliers corresponding to the constraints in (3). We recall the following definitions for a given parameter $p \in \mathbb{R}^d$.

Definition 1. (KKT point) A point $z \in \mathbb{R}^n$ is called a Karush–Kuhn–Tucker (KKT) point for problem (3) if there exist Lagrange multipliers $\lambda \in \mathbb{R}^{m_I + m_E}$ such that

$$\begin{aligned} \nabla_z L(z, \lambda, p) &= 0, & c_i(z, p) &\leq 0, \quad \forall i \in \mathcal{I}, & c_i(z, p) &= 0, \quad \forall i \in \mathcal{E} \\ \lambda^\top c(z, p) &= 0, & \lambda_i &\geq 0, \quad \forall i \in \mathcal{I}. \end{aligned}$$

Let $\Lambda^*(p)$ denote the set of Lagrange multipliers λ satisfying the KKT conditions and $\mathcal{A}(z, p) := \{i \in \mathcal{I} : c_i(z, p) = 0\} \cup \mathcal{E}$ denote the index of active constraints at a feasible point z . We write $\lambda^*(p)$ to denote an element of $\Lambda^*(p)$.

Definition 2. (Strict complementarity) Suppose z^* is a KKT point for problem (3) with associated Lagrange multipliers λ^* . We say that strict complementarity (SC) holds at (z^*, λ^*) if $\lambda_i^* - c_i(z^*, p) > 0$, $\forall i \in \mathcal{I}$.

Definition 3. (LICQ) The linear independence constraint qualification (LICQ) holds at a feasible point z to problem (3) if $\nabla_z c_i(z, p)$, $i \in \mathcal{A}(z, p)$, are linearly independent.

Definition 4. (Strong second order sufficient conditions) Suppose z^* is a KKT point for problem (3) with associated multipliers λ^* . Let $\mathcal{A}_+^* := \{i \in \mathcal{I} : \lambda_i^* > 0\}$. The strong second order sufficient condition (SSOSC) holds at (z^*, λ^*) if

$$q^\top \nabla_z^2 L(z^*, \lambda^*, p) q > 0, \quad \forall q \text{ s.t. } q \neq 0 \text{ and } \nabla_z c_i(z^*, p)^\top q = 0, \quad \forall i \in \mathcal{E} \cup \mathcal{A}_+^*.$$

We now present sufficient conditions under which sensitivities of the value function $\nu^*(p)$ and the solution mapping $z^*(p)$ may be computed.

Theorem 1. [Parametric sensitivities] Let $z^*(p)$ be a KKT point for problem (3) with associated Lagrange multipliers $\lambda^*(p)$. Suppose for some $\bar{p} \in \mathbb{R}^d$, functions F and c are twice continuously differentiable in a neighborhood of $(z^*(\bar{p}), \bar{p})$. Assume that the LICQ and SC conditions hold at $(z^*(\bar{p}), \lambda^*(\bar{p}))$. Additionally, suppose either

- (a) $|\mathcal{A}(z^*(\bar{p}), \bar{p})| = n$, or
- (b) the SSOSC holds at $(z^*(\bar{p}), \lambda^*(\bar{p}))$.

Then, $\exists \delta > 0$ such that $\forall p \in B_\delta(\bar{p})$, we can choose the mappings $z^*(p)$ and $\lambda^*(p)$ to be continuously differentiable on $B_\delta(\bar{p})$ and $z^*(p)$ to be a strict local minimizer of (3). Additionally, for all $p \in B_\delta(\bar{p})$, the gradient of the value function ν^* is given by

$$\nabla_p \nu^*(p) = \nabla_p L(z^*(p), \lambda^*(p), p),$$

and the gradient of the solution mapping $z^*(p)$ may be computed for each $p \in B_\delta(\bar{p})$ as follows depending on whether condition (a) or (b) above holds:

- (a) Let $J_z(p) \in \mathbb{R}^{n \times n}$ and $J_p(p) \in \mathbb{R}^{n \times p}$ be matrices with rows $(\nabla_z c_i(z^*(p), p))^\top$, $i \in \mathcal{A}(z^*(p), p)$, and $(\nabla_p c_i(z^*(p), p))^\top$, $i \in \mathcal{A}(z^*(p), p)$, respectively. Then

$$\nabla_p z^*(p) = -[J_z(p)]^{-1} J_p(p).$$

- (b) Let $H_{z,\lambda}(p) := \begin{bmatrix} \nabla_z^2 L(z^*(p), \lambda^*(p), p) & J_z(p) \\ (J_z(p))^\top & 0 \end{bmatrix}$, where $J_z(p)$ is a $|\mathcal{A}(z^*(p), p)| \times n$ matrix with rows $(\nabla_z c_i(z^*(p), p))^\top$, $i \in \mathcal{A}(z^*(p), p)$. Then

$$\begin{bmatrix} \nabla_p z^*(p) \\ \nabla_p \lambda_{\mathcal{A}}^*(p) \end{bmatrix} = -[H_{z,\lambda}(p)]^{-1} \begin{bmatrix} \nabla_p L(z^*(p), \lambda^*(p), p) \\ (\nabla_p c_i(z^*(p), p))_{i \in \mathcal{A}(z^*(p), p)} \end{bmatrix},$$

where $\lambda_{\mathcal{A}}^*(p)$ denotes the Lagrange multipliers of the active constraints at $z^*(p)$.

Proof. See Chapter 3 of Fiacco [16], or the unified Theorem 4.4 in Still [49]. \square

Theorem 1 applies to parametric linear programs (LPs) whenever condition (a) therein holds. Next, we present sufficient conditions for the value function $\nu^*(p)$ to be locally Lipschitz continuous. We require the following definition.

Definition 5. (MFCQ) The Mangasarian-Fromovitz constraint qualification (MFCQ) holds at a feasible point z if $\nabla_z c_i(z, p)$, $i \in \mathcal{E}$, are linearly independent and $\exists q \in \mathbb{R}^n$ such that $\nabla_z c_i(z, p)^\top q = 0$, $\forall i \in \mathcal{E}$, and $\nabla_z c_i(z, p)^\top q < 0$, $\forall i \in \mathcal{A}(z, p) \setminus \mathcal{E}$.

Theorem 2. [Local Lipschitz continuity] Suppose functions F and c are continuously differentiable and $\bar{p} \in \mathbb{R}^d$.

- (a) Let $z^*(\bar{p})$ be a strict local minimizer of problem (3) at $p = \bar{p}$ and assume that the MFCQ holds at $z^*(\bar{p})$. Then for some constants $\varepsilon, \delta > 0$ and any $p \in B_\delta(\bar{p})$, there exists a local minimizer $z^*(p) \in B_\varepsilon(z^*(\bar{p}))$ such that the corresponding value function $\nu^*(p) = F(z^*(p), p)$ is locally Lipschitz continuous at \bar{p} .
- (b) Suppose the MFCQ holds at every global minimizer $z^*(\bar{p})$ of problem (3) at $p = \bar{p}$. Assume that $\cup_{p \in B_\varepsilon(\bar{p})} \{z \in \mathbb{R}^n : c_i(z, p) \leq 0, \forall i \in \mathcal{I}, c_i(z, p) = 0, \forall i \in \mathcal{E}\}$ is bounded for some $\varepsilon > 0$. Then the global minimum value function $\nu^*(p) = F(z^*(p), p)$ is locally Lipschitz continuous at \bar{p} .

Proof. See Lemma 6.2 of Still [49]. \square

Theorem 1.12 of Dempe [9] and its surrounding discussion provides estimates of generalized gradients of ν^* in the above setting. Additionally, conditions under which the solution mapping z^* is either Hölder continuous or locally Lipschitz continuous are presented in Theorems 6.2 to 6.5 of Still [49].

In the next two sections, we formulate optimality-based discretizations and generalized discretizations for (SIP) as max-min problems and use parametric sensitivity theory to design local optimization algorithms for solving them.

3 Optimality-based discretization methods

We propose new optimality-based discretization methods for (SIP) that can achieve faster convergence of lower bounds than the Blankenship and Falk algorithm. Similar to the BF algorithm, solving (LBP) with our new discretization strategies may not yield a feasible point for (SIP) finitely. Therefore, our lower bounding approaches must also be supplemented with techniques for generating feasible points to provably solve (SIP) to global optimality in finite time. We only focus on constructing tighter lower bounds in this work and

assume for simplicity that all subproblems solved to global optimality are solved *exactly* in finite time; our approaches may be adapted to the setting where these subproblems are only solved to ε -global optimality for some $\varepsilon > 0$ [11, 12, 20].

As introduced in Section 1, the BF algorithm is a feasibility-based approach that iteratively populates the discretization Y_d with points in Y corresponding to the largest violation of the semi-infinite constraint (1) at incumbent solutions x^k of (LBP). The key idea of our optimality-based discretization methods is to instead populate Y_d with points in Y that yield the highest lower bound. In the first iteration, instead of updating Y_d with a solution $y^{BF,1}$ of (LLP(x)) at $x = x^1$ as in Algorithm 1, we propose to solve the following max-min problem to determine a discretization $Y_d = \{\bar{y}^1\}$ that results in the highest lower bound:

$$\begin{aligned} \bar{y}^1 \in \arg \max_{y^1 \in Y} \min_{x \in X} f(x) \\ \text{s.t. } g(x, y^1) \leq 0. \end{aligned} \quad (4)$$

Techniques for solving problem (4) are discussed in Section 3.1. We consider two approaches for updating the discretization Y_d at iteration $k > 1$. The first approach discards the discretization from iteration $k - 1$ and determines a fresh discretization Y_d at iteration k by solving the max-min problem

$$\begin{aligned} (\bar{y}^1, \bar{y}^2, \dots, \bar{y}^k) \in \arg \max_{(y^1, y^2, \dots, y^k) \in Y^k} \min_{x \in X} f(x) \\ \text{s.t. } g(x, y^i) \leq 0, \quad \forall i \in [k]. \end{aligned} \quad (5)$$

The resulting discretization $Y_d := \{\bar{y}^1, \bar{y}^2, \dots, \bar{y}^k\}$ yields the highest lower bound among all possible discretizations (LBP) with at most k discretization points. However, the outer-maximization in problem (5) involves $k \times d_y$ variables compared to only d_y variables in problem (4). To mitigate this computational burden, our second approach updates the discretization $Y_d := \{\bar{y}^1, \bar{y}^2, \dots, \bar{y}^{k-1}\}$ at iteration $k - 1$ by adding a single best discretization point $\bar{y}^k \in Y$ for lower bound improvement. This can be formulated as

$$\begin{aligned} \bar{y}^k \in \arg \max_{y^k \in Y} \min_{x \in X} f(x) \\ \text{s.t. } g(x, \bar{y}^i) \leq 0, \quad \forall i \in [k - 1], \\ g(x, y^k) \leq 0. \end{aligned} \quad (6)$$

Problem (6) is a greedy approximation of problem (5). Formulations (5) and (6) are inspired by the idea of strong partitioning proposed by Kannan et al. [23]. We discuss two variants of these max-min problems in Section 3.2.

In contrast with Tsoukalas and Rustem [50] that treats the violation of the semi-infinite constraint (1) and the objective of (SIP) as two competing objectives, problems (5) and (6) directly optimize the discretization for the lower bound. Baltean-Lugojan et al. [1] propose optimality-based cuts with a similar flavor but tailored for outer-approximating semidefinite programs, which can be viewed as a special case of (SIP). Paulus et al. [35] also propose a similar formulation as problem (6) in the context of cutting-plane selection for mixed-integer linear programs (MILPs), where they use explicit enumeration to choose the best optimality-based cut from a *finite* list of candidate cuts.

3.1 Solving the max-min problems

While solving the max-min problems (5) or (6) to global optimality is clearly desirable, this may be as difficult as solving (SIP) itself. Hence, we exploit the fact that solving these problems to local optimality is sufficient to obtain a valid discretization Y_d . In Section 3.3, we outline how the sequences of discretizations (LBP) determined by solving (5) and (6) can be used to obtain sequences of lower bounds converging to v^* . Numerical experiments in Section 6 demonstrate that solving problems (5) or (6) to local optimality is almost always enough to determine good discretizations with faster convergence than the BF algorithm. We begin with some useful notation.

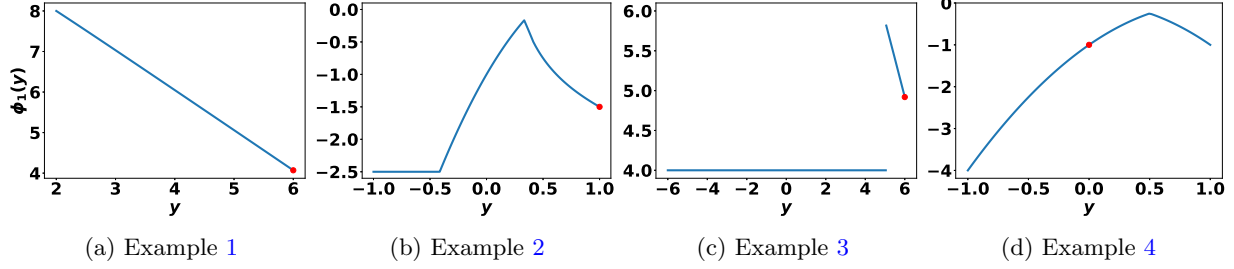


Figure 2: Value functions ϕ_1 (note the discontinuity for Example 3). The red dot indicates ϕ_1 at the point $y^{BF,1}$ determined by the BF algorithm at iteration 1.

Let $\phi_k : Y^k \rightarrow \mathbb{R}$ denote the value function of the inner-minimization in problem (5) at iteration $k \in \mathbb{N}$:

$$\begin{aligned} \phi_k(y^1, y^2, \dots, y^k) &:= \min_{x \in X} f(x) \\ &\text{s.t. } g(x, y^i) \leq 0, \quad \forall i \in [k]. \end{aligned} \quad (7)$$

Assuming the maximum is attained¹, problem (5) can be rewritten as

$$(\bar{y}^1, \bar{y}^2, \dots, \bar{y}^k) \in \arg \max_{(y^1, y^2, \dots, y^k) \in Y^k} \phi_k(y^1, y^2, \dots, y^k).$$

Similarly, given a sequence $\{\tilde{y}^i\}_{i=1}^{|\mathcal{I}_{k-1}|} \subset Y$ for some index set $\mathcal{I}_{k-1} \subset \mathbb{N}$, let $\psi_k : Y \rightarrow \mathbb{R}$ denote the value function

$$\begin{aligned} \psi_k(y^k; \{\tilde{y}^i\}_{i=1}^{|\mathcal{I}_{k-1}|}) &:= \min_{x \in X} f(x) \\ &\text{s.t. } g(x, \tilde{y}^i) \leq 0, \quad \forall i \in \mathcal{I}_{k-1}, \\ &\quad g(x, y^k) \leq 0, \end{aligned} \quad (8)$$

where $\{\tilde{y}^i\}_{i=1}^{|\mathcal{I}_{k-1}|}$ are parameters in the above optimization. Assuming again that the maximum is attained, problem (6) can be equivalently written as

$$\bar{y}^k \in \arg \max_{y^k \in Y} \psi_k(y^k; \{\bar{y}^i\}_{i=1}^{k-1}).$$

Note that the function $\phi_1(\cdot)$ is equivalent to $\psi_1(\cdot; \emptyset)$ on Y .

Before we outline our approach for solving problems (5) and (6), we plot ϕ_1 for some examples from the literature. We consider the following three examples in addition to Example 1 in the introduction.

Example 2. [40, Ex. 2.1] Consider (SIP) with $d_x = 2$, $d_y = 1$, $X = [-1, 1]^2$, $Y = [-1, 1]$, $f(x) = -x_1 + 1.5x_2$, and $g(x, y) = -y_1^2 + 2y_1x_1 - x_2$. The global solution is $x^* = (\frac{1}{3}, \frac{1}{9})$ with $v^* = -\frac{1}{6}$.

Example 3. [50, Ex. 2.1] Consider (SIP) with $d_x = 1$, $d_y = 1$, $X = [-6, 6]$, $Y = [-6, 6]$, $f(x) = 10 - x_1$, and $g(x, y) = -x_1^4 + x_1^2 - x_1^2y_1^2 + 2x_1^3y_1 - 4$ (note that [50, Ex. 2.1] includes a typo in the description of g). The global solution is $x^* = 2$ with $v^* = 8$.

Example 4. [29, Ex. (H)] Consider (SIP) with $d_x = 2$, $d_y = 1$, $X = [0, 1] \times [-10^3, 10^3]$, $Y = [-1, 1]$, $f(x) = x_2$, and $g(x, y) = -(x_1 - y_1)^2 - x_2$. The global solution is $x^* = (\bar{x}_1, 0)$ for any $\bar{x}_1 \in [0, 1]$ with $v^* = 0$.

Figure 2 plots the (global) value function ϕ_1 for Examples 1 to 4. It illustrates that ϕ_1 may be nonconcave, nondifferentiable, or even discontinuous with large flat regions. Additionally, the supremum in problem (4)

¹If the maximum is not attained, then we pick any ε -optimal solution for some $\varepsilon > 0$.

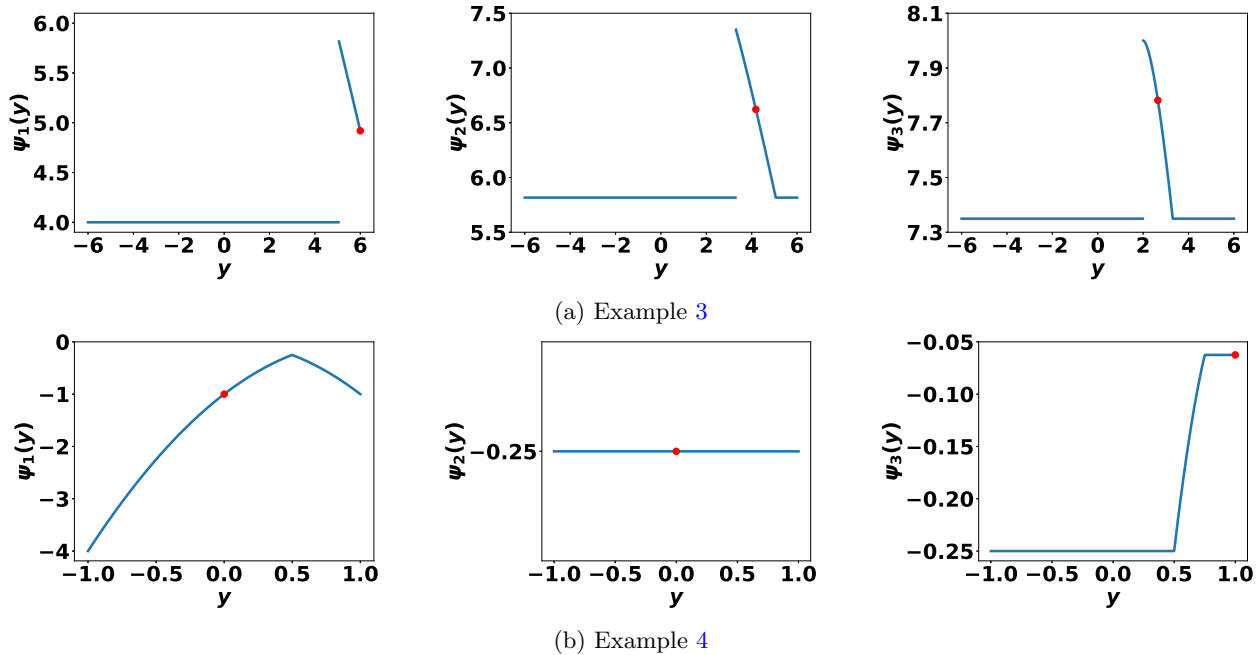


Figure 3: Value functions ψ_k , $k \in [3]$, in the first three iterations of problem (6) for Examples 3 and 4 (all three functions are discontinuous for Example 3). The red dot indicates the initial guess \hat{y}^k (see line 5 of Algorithm 2) for maximizing ψ_k over Y when the discretization is set using Algorithm GREEDY in Section 3.2.

is unattainable for Example 3. This is because under our assumptions on (SIP), Lemma 5.3 of Still [49] only guarantees that the functions ϕ_k and ψ_k are lower semicontinuous on their domains for each $k \in \mathbb{N}$ and does not eliminate the possibility that they may be discontinuous. However, Theorem 2 implies that ϕ_k and ψ_k are locally Lipschitz continuous under relatively mild assumptions, which along with Rademacher’s theorem then implies that these value functions are differentiable almost everywhere. Figure 2 also illustrates that ϕ_1 is usually continuously differentiable at almost every point in its domain and that the BF point $y^{BF,1}$ provides a good initial guess for approximately solving (4).

Figure 3 shows that the value functions ψ_k may become increasingly challenging to optimize over as k increases²; however, solving the lower-level problem (LLP(x)) at incumbent (lower bounding) solutions continues to yield a good initial guess for Examples 3 and 4. These observations open up the possibility of using (generalized) gradient information for solving the max-min problems (5) and (6) to local optimality.

Due to the potentially nonsmooth and discontinuous nature of the functions ϕ_k and ψ_k , we propose to solve problems (5) and (6) using gradients (whenever they exist) of ϕ_k and ψ_k within a bundle solver for nonsmooth nonconvex optimization [26]. Each iteration of the bundle method requires function and generalized gradient evaluations. We calculate function values by solving problem (7) and problem (8) to local optimality. We then use Theorem 1 to compute gradients of the local minimum value function ϕ_k or ψ_k (this involves a linear solve) whenever its assumptions hold.

Before we outline our optimality-based discretization methods, we mention some alternatives for solving problems (5) and (6). If the inner-minimization problems (7) and (8) are convex and satisfy a constraint qualification (e.g., Slater’s condition), then we may be able to use strong duality to reformulate problems (5) and (6) into single-level maximization problems. The resulting problems can then be solved to local optimality to update the discretization Y_d . Alternatively, suppose problems (7) and (8) satisfy a constraint qualification for each feasible point of the outer-maximization problems (5) and (6) and the functions in problems (7) and (8) are continuously differentiable. Then we can reformulate problems (5) and (6) into maximization

²We do not plot ψ_2 and ψ_3 for Examples 1 and 2 as $\max_{y \in Y} \phi_1(y) = v^*$ for these instances.

Algorithm 2 Prototype optimality-based discretization algorithm

```
1: Input: feasibility tolerance  $\varepsilon_f \geq 0$ , minimum bound improvement  $\delta \geq 0$ .
2: Initialize  $Y_d = \emptyset$ .
3: for  $k = 1, 2, \dots$  do
4:   Solve problem (LBP) globally to get solution  $x^k$ , lower bound  $LBD^k$ .
5:   Solve problem (LLP( $x$ )) with  $x = x^k$  globally to get solution  $\hat{y}^k \in Y$ .
6:   if  $G(x^k) \leq \varepsilon_f$  then
7:     Terminate with  $\varepsilon_f$ -feasible solution  $x^k$  to (SIP).
8:   else
9:     Solve a max-min problem to determine points  $\{\bar{y}^{k,i}\}_{i=1}^{|\mathcal{I}_k|}$ .
10:    if optimal value of max-min problem  $\geq LBD^k + \delta$  then
11:      Update the discretization  $Y_d$  using  $\{\bar{y}^{k,i}\}_{i=1}^{|\mathcal{I}_k|}$ .
12:    else
13:      Set  $Y_d \leftarrow Y_d \cup \{\hat{y}^k\}$ .
14:    end if
15:  end if
16: end for
```

problems with complementarity constraints using the KKT conditions for the inner-minimization problems, which can then be solved locally to update Y_d (cf. [47]).

When the assumptions of Theorem 1 fail to hold, we may be able to either use (estimates of) generalized gradients [9, 32] of ϕ_k and ψ_k , or directional derivatives [37] of ϕ_k and ψ_k and their generalizations [42, 43] to solve problems (5) and (6) to local optimality [4]. Finally, we mention that techniques for minimizing discontinuous functions (see, e.g., Ermoliev et al. [15]) may also be used to maximize ϕ_k and ψ_k over their domains when none of the aforementioned approaches are applicable.

3.2 Outline of optimality-based discretization methods

Algorithm 2 outlines a prototype optimality-based discretization method for (SIP). Similar to the BF algorithm 1, this prototype algorithm begins with an empty discretization Y_d . At the start of each iteration, it solves (LBP) to *global* optimality to determine a candidate solution x^k and a corresponding lower bound LBD^k . It then solves the lower-level problem (LLP(x)) with $x = x^k$ to *global* optimality to determine a point \hat{y}^k to certify whether x^k is ε_f -feasible for (SIP). The key difference with the BF algorithm (lines 9 to 11 of Algorithm 2) is that if x^k is not ε_f -feasible, Algorithm 2 solves a max-min problem (locally) to identify new points that may be used to update the discretization Y_d if a sufficient bound increase condition holds, instead of simply adding \hat{y}^k to the discretization as done in the BF algorithm. Checking this sufficient bound increase condition requires the global solution of the inner-minimization problems (7) and (8) (i.e., the global solution of (LBP) with the new candidate discretization). If problem (LBP) is not convex, we assume this additional global solve is performed *only* to establish our theoretical results³.

We consider four realizations that only vary in lines 9 and 11 of Algorithm 2: OPT, GREEDY, 2GREEDY, and HYBRID.

- Algorithm OPT seeks to discard the discretization Y_d at iteration $k - 1$ and replace it with a fresh discretization that is obtained by solving problem (5) to local optimality⁴ (which in turn involves solving a sequence of inner problems (7) to local optimality). It initializes the solution of this max-min problem using \hat{y}^k and the discretization Y_d at iteration $k - 1$.

³An extra global solve is needed only if sufficient bound increase is violated *after* running line 4.

⁴We use the term “local optimality” for the outer-maximization problems loosely because the inner-minimization problems are only solved locally.

- Algorithm **GREEDY** looks to add a single point to the discretization $Y_d := \{\bar{y}^1, \bar{y}^2, \dots, \bar{y}^{k-1}\}$ at iteration $k - 1$ by solving problem (6) to local optimality (which in turn involves solving a sequence of inner problems (8) to local optimality) with the initialization \hat{y}^k .
- Algorithm **2GREEDY** first adds the point \hat{y}^k to the discretization Y_d at iteration $k - 1$. It then solves problem (6) to local optimality with parameters $Y_d \cup \{\hat{y}^k\}$ to try and find a second point to add to the discretization that increases the lower bound by at least δ (this max-min problem is initialized with a random point in Y so that **LICQ** may hold).
- Algorithm **HYBRID** solves problem (5) locally to try and determine a fresh discretization with sufficient lower bound improvement (similar to **OPT**) at each of the first K iterations. From iteration $K + 1$, it then switches to the **GREEDY** strategy and solves problem (6) to local optimality to try and find a single best point to add to the previous discretization Y_d .

All four realizations of Algorithm 2 use the freely available point \hat{y}^k to either construct an initial guess, or to add to the discretization Y_d . The motivation behind the **2GREEDY** scheme is that adding two discretization points per iteration (including the point \hat{y}^k) may reduce the number of (**LBP**) and (**LLP**(x)) subproblems that are solved to global optimality before the sequence of lower bounds $\{LBD^k\}$ converges to v^* . This may in turn reduce the overall time spent on computing lower bounds. Algorithm **HYBRID** relieves the increasing computational burden of solving problem (5) as k increases.

Enhancements. There are several ways in which Algorithms **OPT**, **GREEDY**, **2GREEDY**, and **HYBRID** may be enhanced in practice. First, the local solution of the sequence of inner-minimization problems (7) or (8) can be effectively warm-started using active set methods. Second, solving problems (7) or (8) using multi-start techniques can greatly increase the likelihood that we solve problems (5) and (6) with the global value functions ϕ_k and ψ_k . Finally, as we saw in Examples 2 to 4, the assumptions of Theorem 1 may not hold at each point in the domain of the value functions ϕ_k and ψ_k . In such situations, we may either return the best found solution to problems (5) and (6), or randomly perturb the current iterate in an attempt to avoid points of nondifferentiability (cf. Theorem 2).

3.3 Convergence guarantees

We establish convergence of the sequence $\{LBD^k\}$ of lower bounds determined by Algorithms **OPT**, **GREEDY**, **2GREEDY**, and **HYBRID**. We emphasize that Theorem 3 below does *not* assume the outer-maximization and inner-minimization problems in the above realizations of Algorithm 2 are solved to global optimality. Instead, we assume *only for its proof* that the sufficient bound increase check on line 10 of Algorithm 2 involves the *global* solution of problem (**LBP**) with the candidate discretization.

Theorem 3. Consider Algorithm 2 with $\varepsilon_f = 0$, $\delta > 0$. Suppose the discretization Y_d is updated using Algorithm **OPT**, **GREEDY**, **2GREEDY**, or **HYBRID**. Then $\lim_{k \rightarrow \infty} LBD^k = v^*$.

Proof. Since f and g are continuous, X and Y are compact, and (**SIP**) is assumed to be feasible, the optimal value v^* is finite and bounded below by $\min_{x \in X} f(x) > -\infty$. Line 11 of Algorithm 2 updates the discretization Y_d using the points $\{\bar{y}^{k,i}\}_{i=1}^{|\mathcal{I}_k|}$ only if this candidate discretization increases the lower bound in iteration $k + 1$ by at least δ . Since $v^* - LBD^0 := v^* - \min_{x \in X} f(x) < \infty$, line 11 of Algorithm 2 can be executed only a finite number of times before the lower bound converges to v^* . Therefore, line 13 is executed for all k large enough and the asymptotic behavior of Algorithm 2 is thus the same as that of the **BF** Algorithm 1. The result that $LBD^k \rightarrow v^*$ then follows from Lemma 2.2 of Mitsos [28] (cf. Theorem 3.1 of Harwood et al. [20]). \square

Our next result identifies favorable properties of Algorithm **OPT** when the problem (5) is solved to *global* optimality. We say that (**SIP**) is convex if the set X is convex and functions f and g are convex with respect to x (on X).

Proposition 4. Consider Algorithm OPT with parameters $\varepsilon_f = \delta = 0$ and assume that problem (5) is solved to *global* optimality.

1. The sequence of lower bounds $\{LBD^k\}$ generated by Algorithm OPT dominates any sequence of lower bounds generated by solving (LBP) with at most $k - 1$ discretization points in iteration k .
2. Suppose (SIP) is convex and $\exists \bar{x} \in X$ such that $G(\bar{x}) < 0$. Then, there exists a discretization $Y_d := \{\hat{y}^1, \hat{y}^2, \dots, \hat{y}^m\} \subset Y$, with $m \leq d_x$, such that the optimal value of (LBP) equals v^* , i.e., Algorithm OPT converges in at most d_x iterations.

Proof. The first part holds by the definition of problem (5), whereas the second part follows from Theorem 3.2 of Shapiro [41]. \square

Our next results establish rates of convergence of the BF and OPT lower bounds.

Theorem 5. Consider the BF Algorithm 1 with $\varepsilon_f > 0$. Suppose $\{g(\cdot, y)\}_{y \in Y}$ is uniformly Lipschitz continuous on X with Lipschitz constant $L_{g,x} > 0$, i.e.,

$$|g(x, y) - g(\bar{x}, y)| \leq L_{g,x} \|x - \bar{x}\|, \quad \forall x, \bar{x} \in X, y \in Y.$$

Then Algorithm 1 terminates in at most $\left(\frac{\text{diam}(X)L_{g,x}}{2\varepsilon_f} + 1\right)^{d_x}$ iterations.

Proof. Follows, e.g., from Section 5.2 of Mutapic and Boyd [33]. \square

Theorem 6. Consider Algorithm OPT with $\varepsilon_f > 0$ and $\delta = 0$. Suppose $\{g(x, \cdot)\}_{x \in X}$ is uniformly Lipschitz continuous on Y with Lipschitz constant $L_{g,y} > 0$, i.e.,

$$|g(x, y) - g(x, \bar{y})| \leq L_{g,y} \|y - \bar{y}\|, \quad \forall y, \bar{y} \in Y, x \in X.$$

If problem (5) is solved to *global* optimality, then Algorithm OPT terminates with an ε_f -feasible point in at most $\left(\frac{\text{diam}(Y)L_{g,y}}{2\varepsilon_f}\right)^{d_y}$ iterations. Furthermore, if the assumptions of Theorem 5 hold, then Algorithm OPT terminates in at most $\min\left\{\left(\frac{\text{diam}(Y)L_{g,y}}{2\varepsilon_f}\right)^{d_y}, \left(\frac{\text{diam}(X)L_{g,x}}{2\varepsilon_f} + 1\right)^{d_x}\right\}$ iterations.

Proof. Suppose Algorithm OPT has not converged by iteration $k > 1$. The candidate solution x^k at iteration k of Algorithm OPT satisfies for each $1 \leq j < k$:

$$\begin{aligned} g(x^k, \hat{y}^k) > \varepsilon_f \text{ and } g(x^k, \bar{y}^j) \leq 0 &\implies g(x^k, \hat{y}^k) - g(x^k, \bar{y}^j) > \varepsilon_f \\ &\implies \|\hat{y}^k - \bar{y}^j\| > \frac{\varepsilon_f}{L_{gy}}, \end{aligned}$$

where $Y_d = \{\bar{y}^i\}_{i=1}^{k-1}$ denotes the discretization at the start of iteration k . Therefore, an upper bound on the number of iterations for convergence can be obtained by calculating the number of Euclidean balls of radius $\frac{\varepsilon_f}{L_{gy}}$ needed to cover Y (cf. [33]). The second assertion follows from Theorem 5 and part 1 of Proposition 4. \square

When $d_y \ll d_x$, as in many applications, the bound on the number of iterations for Algorithm OPT can be much smaller than the bound on the BF algorithm (cf. [48, Theorem 2]). Theorem 6 can also be sharpened to estimate the number of balls needed to cover $\{y^*(x) : x \in X\}$ instead of Y . We now link Theorems 5 and 6 to the rate of convergence of their sequence of lower bounds.

Proposition 7. Suppose the value function $V(z) := \min_{\{x \in X : G(x) \leq z\}} f(x)$ is Lipschitz continuous on $[0, \bar{\varepsilon}]$ with a Lipschitz constant $L_V > 0$. Additionally, suppose the assumptions of Theorems 5 and 6 hold. Then for any $\varepsilon \in (0, \bar{\varepsilon})$:

1. Whenever $k \geq \left(\frac{\text{diam}(X)L_{g,x}L_V}{2\varepsilon} + 1 \right)^{d_x}$, the BF lower bound $LBD^k \geq v^* - \varepsilon$.
2. Whenever $k \geq \min \left\{ \left(\frac{\text{diam}(Y)L_{g,y}L_V}{2\varepsilon_f} \right)^{d_y}, \left(\frac{\text{diam}(X)L_{g,x}L_V}{2\varepsilon_f} + 1 \right)^{d_x} \right\}$, the OPT lower bound $LBD^k \geq v^* - \varepsilon$.

Proof. Readily follows from Theorems 5 and 6 upon noting $V(\varepsilon) \geq v^* - L_V\varepsilon$. \square

The Lipschitz assumption in Proposition 7 is satisfied by convex SIPs when the objective f is Lipschitz and Slater's condition holds (see Corollary 2 to Theorem 6.3.2 in Clarke [7]). Chapter 6 of Clarke [7] also details other constraint qualifications under which this Lipschitz assumption holds.

The example below shows discretization-based lower bounding methods involving (LBP) (such as the BF algorithm and *any* realization of Algorithm 2) may take an exponential number of iterations in the problem dimensions to converge. Although this behavior is expected, we are not aware of such an example in the SIP literature (similar examples are known for the classical cutting-plane method in convex optimization, see [34, Ex. 3.3.1], [22, Ex. 1]).

Example 5. (Based on [22, Ex. 1]) Consider the instance of (SIP) with $X = [-1, 1]^{d_x}$, $Y = \{y \in \mathbb{R}^{d_y} : \|y\|^2 = d_y - 1\}$, $f(x) = -\|x\|^2$, and $g(x, y) = \sum_{i=1}^{d_x} (x_i - y_i)y_i$. Note that this semi-infinite constraint (1) may be reformulated as the convex constraint $\|x\| \leq \sqrt{d_x - 1}$. Any $x \in X$ with $\|x\| = \sqrt{d_x - 1}$ solves (SIP) with $v^* = 1 - d_x$.

Lemma 2.1 of Hijazi et al. [22] implies any discretization point $y \in Y$ can exclude at most one vertex of the cube X . Because every vertex of X is a solution to (LBP) with the discretization $Y_d = \emptyset$, *any* discretization-based lower bounding algorithm that solves (LBP) requires exponentially many discretization points in the dimension d_x for its sequence of lower bounds $\{LBD^k\}$ to converge to within 0.5 of v^* .

4 Generalized optimality-based discretizations

In this section, we propose generalized optimality-based discretization methods for (SIP) that can achieve faster convergence than *any* traditional discretization method relying on (LBP). We assume in our proofs that all subproblems that are solved to global optimality are solved *exactly*. Numerical results in Section 6 show that our proposed algorithms work well when our subproblems are solved with finite precision and to a specified optimality tolerance.

Consider the reformulation (2) of (SIP), where the semi-infinite constraint (1) is replaced with the constraint $g(x, y^*(x)) \leq 0$. The BF Algorithm 1 approximates $y^*(x)$ with solutions $\{y^*(x^k)\}$ of the lower-level problem (LLP(x)) at incumbent solutions $\{x^k\}$ of problem (LBP). This zeroth-order approximation of y^* can be crude as noted in Section 1. Assuming that the mapping y^* is differentiable (cf. Theorem 1), Seidel and Küfer [40] and Djelassi [11] instead propose to use the first-order approximation $y^*(x) \approx y^*(x^k) + J_y^*(x^k)(x - x^k)$ at incumbent solutions $\{x^k\}$, where $J_y^*(x^k)$ is the $d_y \times d_x$ Jacobian matrix with rows $(\nabla_x y_1^*(x^k))^T, \dots, (\nabla_x y_{d_y}^*(x^k))^T$. In particular, Section 3.4 of Djelassi [11] suggests the following generalization of the BF lower bounding problem (LBP):

$$\begin{aligned} \min_{x \in X} \quad & f(x) \\ \text{s.t.} \quad & g(x, \text{proj}_Y(\bar{A}^k x + \bar{b}^k)) \leq 0, \quad \forall k \in [n], \end{aligned} \tag{G-LBP}$$

where constants $\bar{A}^k \in \mathbb{R}^{d_y \times d_x}$ and $\bar{b}^k \in \mathbb{R}^{d_y}$ are specified as $\bar{A}^k = J_y^*(x^k)$, $\bar{b}^k = y^*(x^k) - J_y^*(x^k)x^k$ at a sequence of candidate solutions $\{x^k\} \subset X$ to (SIP). The projection step ensures problem (G-LBP) is a relaxation of (SIP), which implies solving (G-LBP) to *global* optimality yields a valid lower bound on the optimal value v^* . However, it also makes (G-LBP) nonsmooth and more challenging to solve than (LBP). Setting $\bar{A}^k = 0$ and $\bar{b}^k = y^{BF,k} = y^*(x^k)$ in problem (G-LBP) recovers the BF lower bounding problem.

In this work, we primarily consider the setting where $Y = [y^L, y^U]$ for some vectors $y^L, y^U \in \mathbb{R}^{d_y}$. The function $\text{proj}_Y(\cdot)$ can then be reformulated as the MILP-representable function $\text{mid}(y^L, \cdot, y^U)$, and

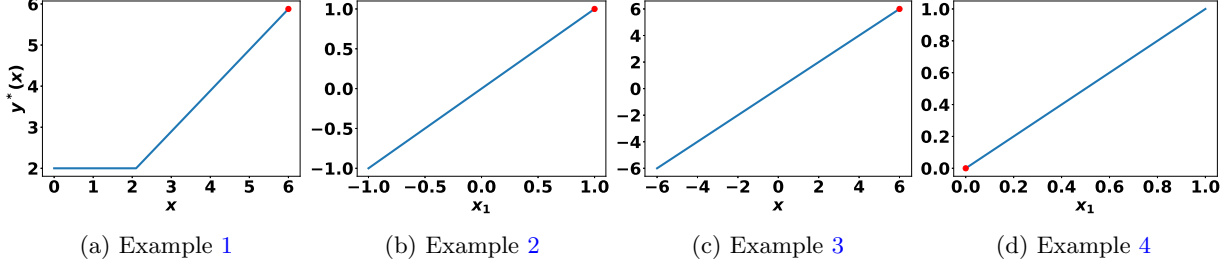


Figure 4: Optimal solution mapping y^* (it only depends on x_1 for Examples 2 and 4). The red dot indicates $y^*(x^1)$ at $x^1 \in \arg \min_{x \in X} f(x)$.

problem (G-LBP) can in turn be reformulated as a mixed-integer nonlinear program (MINLP). Section 3.4 of Djelassi [11] also considers more general settings for Y .

We construct optimality-based generalized discretizations of (SIP) in the form of (G-LBP) that can achieve faster convergence of lower bounds than the discretization methods in Section 3. As in Section 3, our idea is to construct generalized discretizations (G-LBP) that yield the highest lower bound. At the first iteration, we propose to solve the following max-min problem to determine constants A^1 and b^1 that result in the highest lower bound:

$$(\bar{A}^1, \bar{b}^1) \in \arg \max_{A^1 \in \mathbb{R}^{d_y \times d_x}, b^1 \in \mathbb{R}^{d_y}} \min_{x \in X} f(x) \quad (9)$$

$$\text{s.t. } g(x, \text{proj}_Y(A^1 x + b^1)) \leq 0.$$

We consider the following analogues of problems (5) and (6) for updating the generalized discretizations at iteration $k > 1$. The first approach determines a fresh generalized discretization at iteration k by solving the max-min problem

$$(\bar{A}^1, \bar{b}^1, \dots, \bar{A}^k, \bar{b}^k) \in \arg \max_{\substack{A^1, \dots, A^k \in \mathbb{R}^{d_y \times d_x} \\ b^1, \dots, b^k \in \mathbb{R}^{d_y}}} \min_{x \in X} f(x) \quad (10)$$

$$\text{s.t. } g(x, \text{proj}_Y(A^i x + b^i)) \leq 0, \quad \forall i \in [k].$$

The resulting generalized discretization yields the highest lower bound among all possible generalized discretizations (G-LBP) with k generalized discretization cuts. To mitigate the computational burden of solving problem (10), our second approach updates the generalized discretization at iteration $k - 1$ by adding a single best generalized discretization cut with parameters $\bar{A}^k \in \mathbb{R}^{d_y \times d_x}$ and $\bar{b}^k \in \mathbb{R}^{d_y}$. This can be formulated as the max-min problem

$$(\bar{A}^k, \bar{b}^k) \in \arg \max_{A^k \in \mathbb{R}^{d_y \times d_x}, b^k \in \mathbb{R}^{d_y}} \min_{x \in X} f(x) \quad (11)$$

$$\text{s.t. } g(x, \text{proj}_Y(\bar{A}^i x + \bar{b}^i)) \leq 0, \quad \forall i \in [k - 1],$$

$$g(x, \text{proj}_Y(A^k x + b^k)) \leq 0.$$

Before outlining our proposal for solving problems (10) and (11), we provide empirical motivation for why our optimality-based generalized discretization methods may perform well. Figure 4 plots the optimal solution mapping y^* for Examples 1 to 4. Interestingly, this mapping is well-behaved for all four examples (it is piecewise-linear for Example 1 and linear for Examples 2 to 4). Moreover, using $\bar{A}^1 = J_y^*(x^1)$ and $\bar{b}^1 := y^*(x^1) - J_y^*(x^1)x^1$ in (G-LBP) at the point $(x^1, y^*(x^1))$ highlighted in these plots yields a lower bound equal to v^* for all four examples. However, this favorable situation may not always be the case, and the mapping y^* may be nonconvex, nonsmooth, and even discontinuous in general. For example, any (SIP) with $X = Y = [0, 1]$ and $g(x, y) = (x - 0.5)y$ results in the optimal solution mapping $y^*(x) = \mathbf{1}(x - 0.5)$, where

$\mathbb{1}(z) = 1$ if $z \geq 0$ and zero otherwise, which is discontinuous at $x = 0.5$. Example 5 from Section 3.3 provides another instance where y^* is discontinuous.

Example 5: Consider the optimal solution mapping $y^*(x) := \frac{x}{\|x\|} \sqrt{d_x - 1}$ if $x \neq 0$ and $y^*(x) := \bar{y}$ if $x = 0$. It is discontinuous at $x = 0$ irrespective of $\bar{y} \in Y$. However, setting $\bar{A}^1 = I$ (the identity matrix) and $\bar{b}^1 = 0$ in (G-LBP) with $k = 1$ yields a lower bound equal to v^* since $\text{proj}_Y(x) = y^*(x)$, $\forall x \in X$. Therefore, a single generalized discretization cut is sufficient for convergence (this is in stark contrast with the discretization methods in Section 3.3 that require exponentially many iterations in the dimension d_x to converge).

4.1 Solving the max-min problems

In addition to the challenges outlined in Section 3, solving problems (10) and (11) to global optimality may also be challenging due to the nonsmooth projection step. Therefore, we design local optimization methods for solving a smooth approximation of these max-min problems and empirically show in Section 6 that this may be sufficient to determine good generalized discretizations that yield fast convergence of lower bounds.

Let $\phi_k^G : (\mathbb{R}^{d_y \times d_x} \times \mathbb{R}^{d_y})^k \rightarrow \mathbb{R}$ denote the value function of the inner-minimization in problem (10) at iteration $k \in \mathbb{N}$, i.e.,

$$\begin{aligned} \phi_k^G(A^1, b^1, \dots, A^k, b^k) &:= \min_{x \in X} f(x) \\ \text{s.t. } g(x, \text{proj}_Y(A^i x + b^i)) &\leq 0, \quad \forall i \in [k]. \end{aligned} \quad (12)$$

Assuming the maximum is attained, problem (10) can be rewritten as

$$(\bar{A}^1, \bar{b}^2, \dots, \bar{A}^k, \bar{b}^k) \in \arg \max_{\substack{A^1, \dots, A^k \in \mathbb{R}^{d_y \times d_x} \\ b^1, \dots, b^k \in \mathbb{R}^{d_y}}} \phi_k^G(A^1, b^1, \dots, A^k, b^k).$$

Similarly, given a sequence $\{(\tilde{A}^i, \tilde{b}^i)\}_{i=1}^{|\mathcal{I}_{k-1}|} \subset Y$ for some index set $\mathcal{I}_{k-1} \subset \mathbb{N}$, let $\psi_k^G : Y \rightarrow \mathbb{R}$ denote the value function

$$\begin{aligned} \psi_k^G(A^k, b^k; \{(\tilde{A}^i, \tilde{b}^i)\}_{i=1}^{|\mathcal{I}_{k-1}|}) &:= \min_{x \in X} f(x) \\ \text{s.t. } g(x, \text{proj}_Y(\tilde{A}^i x + \tilde{b}^i)) &\leq 0, \quad \forall i \in [k-1], \\ g(x, \text{proj}_Y(A^k x + b^k)) &\leq 0, \end{aligned} \quad (13)$$

where $\{(\tilde{A}^i, \tilde{b}^i)\}_{i=1}^{|\mathcal{I}_{k-1}|}$ are parameters in the above problem. Assuming again that the maximum is attained, problem (11) can be equivalently written as

$$(\bar{A}^k, \bar{b}^k) \in \arg \max_{A^k \in \mathbb{R}^{d_y \times d_x}, b^k \in \mathbb{R}^{d_y}} \psi_k^G(A^k, b^k; \{(\bar{A}^i, \bar{b}^i)\}_{i=1}^{k-1}).$$

Note that the function $\phi_1^G(\cdot)$ is equivalent to $\psi_1^G(\cdot; \emptyset)$ on $\mathbb{R}^{d_y \times d_x} \times \mathbb{R}^{d_y}$.

Similar to the value functions ϕ_k and ψ_k in (7) and (8), the functions ϕ_k^G and ψ_k^G above may be nonconcave, nonsmooth, or discontinuous with large flat regions. Furthermore, Theorems 1 and 2 do not apply because of the nonsmooth projection step in problems (12) and (13). In order to be able to use sensitivity information, we replace the projection operator with a smooth approximation. For instance, when $Y := [y^L, y^U]$ is a compact interval, we can replace $\text{proj}_Y(\cdot) = \text{mid}(y^L, \cdot, y^U)$ with a sigmoid whose range equals (y^L, y^U) . We then use gradients (whenever they exist) of this smooth approximation of the value functions ϕ_k^G and ψ_k^G within a bundle solver for nonsmooth nonconvex optimization [26] to solve problems (10) and (11) to local optimality. Function values for the bundle solver are obtained by solving the smooth approximation of problem (12) or (13) to local optimality, and gradients of the smooth approximation of ϕ_k^G and ψ_k^G may be computed using Theorem 1 whenever its assumptions hold. We initialize the variables A^1 and b^1 in problem (9) with $\bar{A}^1 = J_y^*(x^1)$ and $\bar{b}^1 = y^*(x^1) - J_y^*(x^1)x^1$ whenever y^* is differentiable at the initial point $x^1 \in \arg \min_{x \in X} f(x)$ (we consider similar initializations at later iterations). Section 3.1 outlines alternative approaches for solving problems (10) and (11), and Section 3.2 lists possible enhancements.

Algorithm 3 Prototype optimality-based generalized discretization algorithm

```

1: Input: feasibility tolerance  $\varepsilon_f \geq 0$ , minimum bound improvement  $\delta \geq 0$ .
2: Initialize  $P_d^G = \emptyset$ .
3: for  $k = 1, 2, \dots$  do
4:   Solve problem (G-LBP) globally to get solution  $x^k$ , lower bound  $LBD^k$ .
5:   Solve problem (LLP( $x$ )) with  $x = x^k$  globally to get solution  $y^*(x^k)$ .
6:   If assumptions of Theorem 1 hold for (LLP( $x$ )), compute  $J_y^*(x^k)$ .
7:   if  $G(x^k) \leq \varepsilon_f$  then
8:     Terminate with  $\varepsilon_f$ -feasible solution  $x^k$  to (SIP).
9:   else
10:    Solve a max-min problem to determine  $\{(\bar{A}^{k,i}, \bar{b}^{k,i})_{i=1}^{|Z_k|}\}$ .
11:    if optimal value of max-min problem  $\geq LBD^k + \delta$  then
12:      Update the discretization  $P_d^G$  using  $\{(\bar{A}^{k,i}, \bar{b}^{k,i})_{i=1}^{|Z_k|}\}$ .
13:    else
14:      Set  $P_d^G \leftarrow P_d^G \cup \{(J_y^*(x^k), y^*(x^k) - J_y^*(x^k)x^k)\}$  if assumptions of Theorem 1 hold, and
         $P_d^G \leftarrow P_d^G \cup \{(0, y^*(x^k))\}$  otherwise
15:    end if
16:  end if
17: end for

```

4.2 Outline of the generalized discretization methods

Algorithm 3 outlines a prototype optimality-based generalized discretization method for (SIP). We consider four realizations of Algorithm 3 that only vary in lines 10 and 12: G-OPT, G-GREEDY, G-2GREEDY, and G-HYBRID. Algorithms G-OPT, G-GREEDY, and G-HYBRID are direct analogues of OPT, GREEDY, and HYBRID that rely on problems (10) and (11) instead of problems (5) and (6). Algorithm G-2GREEDY first adds $(\bar{A}^k, \bar{b}^k) = (J_y^*(x^k), y^*(x^k) - J_y^*(x^k)x^k)$ or $(\bar{A}^k, \bar{b}^k) = (0, y^*(x^k))$ to P_d^G and then solves problem (11) to try and find another generalized discretization cut to add to the formulation.

4.3 Convergence guarantees

We begin by establishing convergence of the sequence of lower bounds generated by Algorithms G-OPT, G-GREEDY, G-2GREEDY, and G-HYBRID to v^* . Like Theorem 3, this result also does *not* assume the max-min problems (10) and (11) are solved to global optimality.

Theorem 8. Consider Algorithm 3 with $\varepsilon_f = 0$, $\delta > 0$. Suppose Y is convex and the generalized discretization P_d^G is updated using Algorithm G-OPT, G-GREEDY, G-2GREEDY, or G-HYBRID. Then $\lim_{k \rightarrow \infty} LBD^k = v^*$.

Proof. The proof follows a similar outline as the proof of Theorem 3 and Lemma 2.2 of Mitsos [28] (cf. Theorem 3.1 of Harwood et al. [20]).

Suppose each $x^k \in X$ is infeasible to (SIP) (otherwise, $LBD^k = v^*$ for all k large enough). Since X is compact, we can assume (by moving to a subsequence) that $x^k \rightarrow x^* \in X$. We show that x^* is feasible to (SIP), which implies $LBD^k \rightarrow v^*$.

Following arguments in Theorem 3, line 14 of Algorithm 3 must be run infinitely often with $P_d^G \leftarrow P_d^G \cup \{(J_y^*(x^k), y^*(x^k) - J_y^*(x^k)x^k)\}$ or $P_d^G \leftarrow P_d^G \cup \{(0, y^*(x^k))\}$. Therefore, the asymptotic behavior of Algorithm 3 is the same as the algorithm that adds at each iteration either $(J_y^*(x^k), y^*(x^k) - J_y^*(x^k)x^k)$ to P_d^G if Theorem 1 holds, or $(0, y^*(x^k))$ to P_d^G otherwise. We show that $LBD^k \rightarrow v^*$ for the above algorithm. Define the index sets $\mathcal{J}_k := \{j \in [k] : \text{Theorem 1 holds}\}$ and $\mathcal{L}_k := \{1, \dots, k\} \setminus \mathcal{J}_k$.

By construction of the above problem, we have $\forall l, k$ such that $l > k$:

$$g(x^l, \text{proj}_Y(\tilde{A}^k x^l + \tilde{b}^k)) \leq 0, \quad \text{if } k \in \mathcal{J}_k, \quad \text{and} \quad g(x^l, y^*(x^k)) \leq 0, \quad \text{if } k \in \mathcal{L}_k,$$

where $\tilde{A}^k := J_y^*(x^k)$, $\tilde{b}^k := y^*(x^k) - J_y^*(x^k)x^k$ if $k \in \mathcal{J}_k$. Continuity of g , $\text{proj}_Y(\cdot)$ and compactness of X, Y ensure uniform continuity, which implies that for any $\varepsilon > 0$, there exists $\kappa > 0$ such that for all $x \in X$ with $\|x - x^l\| < \kappa$ and $\forall l, k$ with $l > k$:

$$g(x, \text{proj}_Y(\tilde{A}^k x + \tilde{b}^k)) < \varepsilon, \quad \text{if } k \in \mathcal{J}_k, \quad \text{and} \quad g(x, y^*(x^k)) < \varepsilon, \quad \text{if } k \in \mathcal{L}_k. \quad (14)$$

Since $x^k \rightarrow x^*$, we have $\|x^l - x^k\| < \kappa$, $\forall l, k$ with $l > k \geq \bar{K}$. Plugging $x = x^k$ in (14) and noting $\tilde{A}^k x^k + \tilde{b}^k = y^*(x^k)$ if $k \in \mathcal{J}_k$ yields $0 < g(x^k, y^*(x^k)) < \varepsilon$, $\forall k \geq \bar{K}$. Therefore, $g(x^k, y^*(x^k)) = G(x^k) \rightarrow 0$ and continuity of G implies $G(x^*) = 0$. \square

By construction, Algorithms **G-OPT** and **G-GREEDY** generate tighter lower bounds than Algorithms **OPT** and **GREEDY**, respectively, since we can set $\tilde{A}^k = 0$. Our next result establishes rate of convergence of the **G-OPT** lower bounds when problem (10) is solved to *global* optimality. We require the following lemma, which is sharp for affine functions (cf. Figure 4).

Lemma 9. Suppose $Z \subset \mathbb{R}^N$ is a compact convex set and $F : Z \rightarrow \mathbb{R}^M$ is continuously differentiable with a Lipschitz continuous gradient. Let $L_{\nabla F}$ denote the Lipschitz constant of ∇F on Z . Then $\forall \varepsilon > 0$, there exist $J = 1 + \left(\frac{\text{diam}(Z)}{2} \sqrt{\frac{L_{\nabla F}}{2\varepsilon}} \right)^N$ affine functions $\{(\alpha^j)^\top z + \beta^j\}_{j=1}^J$ such that $\sup_{z \in Z} \min_{j \in [J]} \|F(z) - ((\alpha^j)^\top z + \beta^j)\| < \varepsilon$.

Proof. The integral form of Taylor's theorem implies for any $z, \bar{z} \in Z$:

$$\|F(z) - F(\bar{z}) - \nabla F(\bar{z})^\top(z - \bar{z})\| \leq \frac{L_{\nabla F}}{2} \|z - \bar{z}\|^2.$$

Therefore, $\|F(z) - F(\bar{z}) - \nabla F(\bar{z})^\top(z - \bar{z})\| < \varepsilon$ whenever $\left\{ z \in Z : \|z - \bar{z}\| < \sqrt{\frac{2\varepsilon}{L_{\nabla F}}} \right\}$. The stated result follows by covering Z using balls of radius $\sqrt{\frac{2\varepsilon}{L_{\nabla F}}}$, setting z^j to be the center of the j th ball, and setting $\alpha^j = \nabla F(z^j)$, $\beta^j = F(z^j) - \nabla F(z^j)^\top z^j$. \square

Theorem 10. Consider Algorithm **G-OPT** with $\varepsilon_f > 0$ and $\delta = 0$. Suppose X and Y are convex sets and $\{g(x, \cdot)\}_{x \in X}$ is uniformly Lipschitz continuous on Y with Lipschitz constant $L_{g,y} > 0$. Assume additionally that y^* is continuously differentiable on X with a Lipschitz continuous gradient. Let $L_{\nabla y}$ denote the Lipschitz constant of ∇y^* on X . If problem (10) is solved to *global* optimality, then Algorithm **G-OPT** terminates with an ε_f -feasible point in at most $1 + \left(\frac{\text{diam}(X)}{2} \sqrt{\frac{L_{\nabla y} L_{gy}}{2\varepsilon_f}} \right)^{d_x}$ iterations. Furthermore, if the assumptions of Theorem 6 also hold, then Algorithm **G-OPT** terminates with an ε_f -feasible point in at most $\min \left\{ 1 + \left(\frac{\text{diam}(X)}{2} \sqrt{\frac{L_{\nabla y} L_{gy}}{2\varepsilon_f}} \right)^{d_x}, \left(\frac{\text{diam}(Y) L_{g,y}}{2\varepsilon_f} \right)^{d_y} \right\}$ iterations.

Proof. Suppose Algorithm **G-OPT** has not converged by iteration $k > 1$. The candidate solution x^k at iteration k of Algorithm **G-OPT** satisfies for each $1 \leq j < k$:

$$\begin{aligned} & g(x^k, y^*(x^k)) > \varepsilon_f \quad \text{and} \quad g(x^k, \text{proj}_Y(\bar{A}^j x^k + \bar{b}^j)) \leq 0 \\ \implies & g(x^k, y^*(x^k)) - g(x^k, \text{proj}_Y(\bar{A}^j x^k + \bar{b}^j)) > \varepsilon_f \\ \implies & L_{gy} \|y^*(x^k) - \text{proj}_Y(\bar{A}^j x^k + \bar{b}^j)\| > \varepsilon_f, \\ \implies & \|y^*(x^k) - (\bar{A}^j x^k + \bar{b}^j)\| > \frac{\varepsilon_f}{L_{gy}}, \end{aligned}$$

where $P_d^G = \{(\bar{A}^i, \bar{b}^i)\}_{i=1}^{k-1}$ denotes the generalized discretization at the start of iteration k . Therefore, an upper bound on the number of iterations for **G-OPT** to converge can be obtained by estimating the minimal number k of generalized discretization cuts required for $\sup_{x \in X} \min_{j \in [k]} \|y^*(x) - (\bar{A}^j x + \bar{b}^j)\| \leq \frac{\varepsilon_f}{L_{gy}}$. The first result then follows from Lemma 9, whereas the second result follows from Theorem 6. \square

Chapter 3 of Fiacco [16] presents conditions when the assumption on y^* holds. The convexity assumption on the set X may be relaxed by considering any convex superset of X . Note that the bound on the number of iterations in Theorem 10 scales like $\varepsilon_f^{-0.5d_x}$ compared to the $\varepsilon_f^{-d_x}$ scaling in Theorem 5. The rate at which $LBD^k \rightarrow v^*$ for G-OPT can be derived similar to Proposition 7.

5 Generalizations

We discuss some extensions of our optimality-based discretization methods.

Multiple semi-infinite constraints. Suppose (SIP) includes $|\mathcal{I}|$ semi-infinite constraints $g_i(x, y) \leq 0$, $\forall y \in Y, i \in \mathcal{I}$. The formulation below generalizes the max-min problem (4) for constructing optimality-based discretizations at the first iteration of Algorithm 2.

$$\begin{aligned} (\bar{y}^1, \dots, \bar{y}^{|\mathcal{I}|}) \in \arg \max_{(y^1, \dots, y^{|\mathcal{I}|}) \in Y^{|\mathcal{I}|}} \min_{x \in X} f(x) \\ \text{s.t. } g_i(x, y^i) \leq 0, \quad \forall i \in \mathcal{I}. \end{aligned}$$

Extensions of problems (5) and (6) and the generalized discretization methods in Section 4 readily follow.

Generalized discretizations based on higher-order cuts. Instead of restricting ourselves to optimality-based linear approximations of $y^*(x)$ as in Section 4, we can construct optimality-based *nonlinear* approximations of y^* for potentially faster convergence. Let $\gamma : X \times \Theta \rightarrow \mathbb{R}^{d_y}$ be any family of functions. We propose the following generalization of (G-LBP):

$$\begin{aligned} \min_{x \in X} f(x) \\ \text{s.t. } g(x, \text{proj}_Y(\gamma(x, \theta^k))) \leq 0, \quad \forall k \in [n]. \end{aligned} \tag{15}$$

Clearly, (G-LBP) is a special case of problem (15) where γ is the family of parametric affine (in x) functions. Extensions of the max-min problems (10) and (11) to determine an optimal sequence of parameters $\{\theta^k\}$ readily follow.

Mixed-integer SIPs. The presence of integer variables in (SIP) precludes the use of the sensitivity theory in Section 2 for solving the max-min problems (5), (6), (10), and (11). Because we can heuristics to solve these max-min problems *without* sacrificing convergence of our (generalized) discretization methods, one heuristic is to use sensitivities of the value functions of these inner-minimization problems with the integer variables fixed to an optimal solution. An alternative is to use smoothing-based approaches [15] for approximating sensitivity information.

6 Numerical results

We compare Algorithms GREEDY, 2GREEDY, HYBRID, and OPT in Section 3.2 and G-GREEDY, G-2GREEDY, G-HYBRID, and G-OPT in Section 4.2 with the BF algorithm on instances from the literature. We only consider scalar semi-infinite constraints and $Y = [y^L, y^U]$, and omit instances with trigonometric functions.

6.1 Implementational details

Our codes are compiled using Julia 1.7.3, JuMP 1.3.1 [14], BARON 23.1.5 [39] or Gurobi 9.1.2 as the global solver, Knitro 13.1.0 as the local NLP solver, Gurobi 9.1.2 as the LP solver, and the bundle solver MPBNGC 2.0 [26] for the max-min problems. They will be made available at <https://github.com/Process-Optimization-and-Control/Optimal-SIP-Discretizations>.

Instance	d_x	d_y	BF	GREEDY	2GREEDY	HYBRID	OPT
number of iterations for convergence							
Watson 2	2	1	2	2	2	2	2
Watson 5	3	1	5	4	2	3	3
Watson 6	2	1	3	2	2	2	2
Watson 7	3	2	2	2	2	2	2
Watson 8	6	2	15	6	4	9	9
Watson 9	6	2	9	8	5	5	13
Watson h	2	1	18	29	19	29	26
Watson n	2	1	3	3	2	3	3
Seidel & Küfer 2.1	2	1	8	2	2	2	2
Tsoukalas & Rustem 2.1	1	1	8	5	6	5	5
Mitsos 4.3	3	1	5	4	2	6	5
Mitsos 4.6	6	1	7	8	6	5	5
Mitsos DP	1	1	28	2	2	2	2
Cerulli et al. PSD 1	21	5	2	2	2	2	2
Cerulli et al. PSD 2	21	5	2	2	2	2	2
Cerulli et al. PSD 3	21	5	2	2	2	2	2
Cerulli et al. PSD 4	21	5	2	2	2	2	2
Cerulli et al. PSD 5	66	10	5	2	5	2	2
Cerulli et al. PSD 6	66	10	6	2	2	2	2
Cerulli et al. PSD 7	105	13	7	2	5	2	2
Cerulli et al. PSD 8	105	13	5	2	5	2	2

Table 2: Comparison of the BF, GREEDY, 2GREEDY, HYBRID, and OPT algorithms. Bold entries correspond to the minimum number of iterations for each instance.

Handling weakly active constraints. Theorem 1 requires strict complementarity. When *only* this condition does *not* hold, the value functions ϕ_k , ψ_k , ϕ_k^G , and ψ_k^G are piecewise-differentiable with a kink at the evaluation point. We can compute a subgradient in this case by excluding a subset of the weakly active constraints from the set of active constraints before solving the linear system in Theorem 1. In general, we exclude weakly active bound constraints as including them causes the derivative of the variable at the bound to be zero. All other weakly active constraints are included for the discretization methods, but excluded for the generalized discretization methods due to observed numerical issues (these choices were found to perform well in our experiments). Alternatively, one may consider using generalized derivatives [43], which come at a higher computational expense.

General algorithmic parameters. Since (LBP) and (G-LBP) may not yield a feasible point finitely, we run our algorithms until the lower bound converges absolutely or relatively to within 10^{-3} of v^* (which is computed offline). We use $\varepsilon_f = \delta = 10^{-8}$ in Algorithms 1, 2, and 3. The feasibility and optimality tolerance of Gurobi is set to 10^{-8} . All of BARON’s parameters, except `MaxTime` = 7200s, are kept to default. Knitro is used with the following parameters: `algorithm` = 5, `ftol` = 10^{-8} , and `feastol` = 10^{-8} .

Algorithm-specific parameters. The starting point for the max-min problem solved by the 2GREEDY method is specified as $0.99\hat{y}^k + 0.005(y^L + y^U)$. Similarly, for the G-2GREEDY method, the elements of A^k are initialized randomly from $[0, 1)$ and b^k is set to $0.99\hat{y}^k + 0.005(y^L + y^U) - A^k x^k$. For the generalized discretization methods, the projection function $\text{mid}\{y^L, y, y^U\}$ is reformulated using the “lambda/convex combination formulation” [52] to derive an MINLP reformulation of (G-LBP), whereas the smooth approximation $t^{-1} \log((\exp(ty^L) + \exp(ty))^{-1} + \exp(-ty^U))$ with smoothing parameter $t = 100$ is used in the max-min problems. If the solution of these smooth approximations terminate after a single iteration of the

Instance	d_x	d_y	BF	GREEDY	2GREEDY	HYBRID	OPT
number of iterations for convergence							
Watson 2	2	1	2	2	2	2	2
Watson 5	3	1	5	6	2	3	3
Watson 6	2	1	3	3	3	3	3
Watson 7	3	2	2	2	2	2	2
Watson 8	6	2	15	20	14	16	15
Watson 9	6	2	9	5	4	19	7
Watson h	2	1	18	18	18	18	18
Watson n	2	1	3	2	2	2	2
Seidel & Küfer 2.1	2	1	8	2	6	2	2
Tsoukalas & Rustem 2.1	1	1	8	7	6	12	> 100
Mitsos 4_3	3	1	5	3	2	3	3
Mitsos 4_6	6	1	7	6	5	4	4
Mitsos DP	1	1	28	2	2	2	2

Table 3: Results with the MPCC reformulation of the max-min problems. Bold entries correspond to the minimum number of iterations for each instance.

bundle method, we restart their solution with a random initialization for (A^k, b^k) . We set the parameter $K = 3$ for Algorithms HYBRID and G-HYBRID.

6.2 Optimality-based discretization methods

We test our discretization methods on small-scale SIPs ($d_x \in [1, 6]$ and $d_y \in \{1, 2\}$) from Watson [53], Seidel and Küfer [40], Tsoukalas and Rustem [50], and Mitsos [29]. We also test our discretization methods on larger SIPs ($d_x \in [21, 105]$ and $d_y \in [5, 13]$) from Cerulli et al. [5] where (LBP) is a quadratically constrained quadratic program (QCQP) that is solved globally using Gurobi.

Table 2 details the instance, dimensions d_x and d_y , and the number of iterations k taken by the different methods for their lower bounds to converge to v^* (note that the number of discretization points at termination is $2k - 2$ for 2GREEDY and $k - 1$ for the other methods). We do not report computational times since all methods take similar time (seconds) for their lower bounds to converge (any gains from a decrease in the number of global solves is offset by the time taken for solving the max-min problems).

On almost all instances, our proposed optimality-based discretization methods require fewer iterations for convergence than the BF algorithm (the main exception is “Watson h”, where our proposed methods face numerical issues). Notably, our new discretization methods require only two iterations (a single discretization point for GREEDY, HYBRID, and OPT, and two discretization points for 2GREEDY) to converge for most instances. Although OPT is theoretically expected to perform at least as well as the other discretization methods, Table 2 shows that this is not always the case in practice. This is because the solution of the max-min problem (5) can get stuck at poor local maxima. Overall, Algorithm 2GREEDY performs the best on the small-scale instances whereas GREEDY, HYBRID, and OPT all perform extremely well on the larger instances.

Table 3 notes the performance of our discretization methods on the small-scale instances when max-min problems (5) and (6) are reformulated and solved as mathematical programs with complementarity constraints (MPCCs), see Section 3.1. These MPCCs are solved using Knitro’s tailored algorithms. Comparing Table 3 with Table 2 reveals that the discretizations generated by the MPCC formulations sometimes converge in fewer iterations, but require more iterations in other cases. The MPCC formulation performs poorly on “Tsoukalas & Rustem 2.1” possibly because it does not satisfy MFCQ at x^* .

6.3 Optimality-based generalized discretization methods

We test our generalized discretization methods on small-scale SIP instances.

Instance	d_x	d_y	BF	G-GREEDY	G-2GREEDY	G-HYBRID	G-OPT
number of iterations for convergence							
Watson 2	2	1	2	2	2	2	2
Watson 5	3	1	5	4	3	4	4
Watson 6	2	1	3	2	2	2	2
Watson 7	3	2	2	2	2	2	2
Watson 8	6	2	15	TLE	9	TLE	TLE
Watson 9	6	2	9	TLE	TLE	TLE	8
Watson h	2	1	18	2	2	2	2
Watson n	2	1	3	3	3	3	3
Seidel & Küfer 2.1	2	1	8	3	3	3	3
Tsoukalas & Rustem 2.1	1	1	8	2	3	2	3
Mitsos DP	1	1	28	2	2	2	2
Mitsos 4.3	3	1	5	5	5	5	4
Mitsos 4.6	6	1	7	6	TLE	6	TLE

Table 4: Comparison of the BF, G-GREEDY, G-2GREEDY, G-HYBRID, and G-OPT algorithms. Bold entries correspond to the minimum number of iterations for each instance, and TLE denotes the time limit of 2 hours was exceeded.

Table 4 summarizes the performance of our generalized discretization methods relative to the BF algorithm. These new methods perform well on most instances, but BARON times out during the solution of the nonconvex MINLP (G-LBP) for “Watson 8”, “Watson 9”, and “Mitsos 4_6”. BARON appears to “stall” while solving these MINLPs—this could potentially be due to weak relaxations of (G-LBP). In most cases, the bundle method also terminates after one iteration during the solution of problems (5) and (6), which suggests our initial guess is either locally optimal or does not provide a clear direction for improvement (which could be due to the smooth approximation of the projection operator). Apart from “Watson h” and “Tsoukalas & Rustem 2.1”, the generalized discretization methods do not offer a significant advantage over the optimality-based discretization methods in Section 3.

Summary. Tables 2 to 4 show that our optimality-based (generalized) discretization methods have the potential to significantly reduce the number of iterations for convergence relative to the classical feasibility-based approach. We expect our new discretization methods to be particularly advantageous when solving (LBP) to global optimality is expensive. While our preliminary results are encouraging, designing more efficient and reliable algorithms for solving our max-min formulations (in particular, our generalized discretization formulations (10) and (11)) merits further investigation.

7 Future work

There are many interesting avenues for future work. First, we would like to extend our optimality-based (generalized) discretization methods to generalized semi-infinite programs [13, 30]. Second, our optimality-based discretization methods could be modified (cf. [28]) to generate feasible points to (SIP). Third, extensions of our max-min formulations can enable the design of more efficient cutting-plane methods for a broader class of optimization problems (cf. [35]). Finally, using machine learning to learn a sequence of optimal discretizations (cf. [10, 23, 35]) can mitigate the computational burden of solving our max-min problems for larger dimensions.

Acknowledgments

E.M.T. and J.J. acknowledge the support of the Norwegian Research Council through the AutoPRO project (RN: 309628). R.K. acknowledges funding from the Center for Nonlinear Studies at LANL and the U.S. Department of Energy’s Laboratory Directed Research and Development program (projects 20210078DR and 20230091ER). R.K. also thanks Dr. Harsha Nagarajan (LANL) and Prof. Qi Zhang (UMN) for helpful discussions.

References

- [1] R. Baltean-Lugojan, P. Bonami, R. Misener, and A. Tramontani. Scoring positive semidefinite cutting planes for quadratic optimization via trained neural networks. *Optimization Online*. URL: http://www.optimization-online.org/DB_HTML/2018/11/6943.html, 2019.
- [2] B. Bhattacharjee, P. Lemonidis, W. H. Green Jr, and P. I. Barton. Global solution of semi-infinite programs. *Mathematical Programming*, 103(2):283–307, 2005.
- [3] J. W. Blankenship and J. E. Falk. Infinitely constrained optimization problems. *Journal of Optimization Theory and Applications*, 19(2):261–281, 1976.
- [4] J. V. Burke, F. E. Curtis, A. S. Lewis, M. L. Overton, and L. E. Simões. *Gradient sampling methods for nonsmooth optimization*, pages 201–225. Springer, Cham, 2020. doi: 10.1007/978-3-030-34910-3_6.
- [5] M. Cerulli, A. Oustry, C. d’Ambrosio, and L. Liberti. Convergent algorithms for a class of convex semi-infinite programs. *SIAM Journal on Optimization*, 32(4):2493–2526, 2022.
- [6] E. W. Cheney and A. A. Goldstein. Newton’s method for convex programming and Tchebycheff approximation. *Numerische Mathematik*, 1(1):253–268, 1959.
- [7] F. H. Clarke. *Optimization and nonsmooth analysis*. SIAM, Philadelphia, 1990.
- [8] S. Dempe. *Foundations of bilevel programming*. Springer Science & Business Media, New York, 2002.
- [9] S. Dempe. *Bilevel Optimization: Reformulation and First Optimality Conditions*, pages 1–20. Springer, Singapore, 2017. doi: 10.1007/978-981-10-4774-9_1.
- [10] A. Deza and E. B. Khalil. Machine learning for cutting planes in integer programming: A survey. *arXiv preprint arXiv:2302.09166*, 2023.
- [11] H. Djelassi. *Discretization-Based Algorithms for the Global Solution of Hierarchical Programs*. PhD thesis, RWTH Aachen, 2020.
- [12] H. Djelassi and A. Mitsos. A hybrid discretization algorithm with guaranteed feasibility for the global solution of semi-infinite programs. *Journal of Global Optimization*, 68(2):227–253, 2017.
- [13] H. Djelassi, A. Mitsos, and O. Stein. Recent advances in nonconvex semi-infinite programming: Applications and algorithms. *EURO Journal on Computational Optimization*, 9:100006, 2021.
- [14] I. Dunning, J. Huchette, and M. Lubin. JuMP: A modeling language for mathematical optimization. *SIAM Review*, 59(2):295–320, 2017.
- [15] Y. M. Ermoliev, V. I. Norikin, and R. J. Wets. The minimization of semicontinuous functions: mollifier subgradients. *SIAM Journal on Control and Optimization*, 33(1):149–167, 1995.
- [16] A. V. Fiacco. *Introduction to sensitivity and stability analysis in nonlinear programming*. Academic Press, New York, 1983.
- [17] C. A. Floudas and O. Stein. The adaptive convexification algorithm: a feasible point method for semi-infinite programming. *SIAM Journal on Optimization*, 18(4):1187–1208, 2008.
- [18] M. A. Goberna and M. A. Lopez. Linear semi-infinite programming theory: An updated survey. *European Journal of Operational Research*, 143(2):390–405, 2002.
- [19] I. E. Grossmann, K. P. Halemane, and R. E. Swaney. Optimization strategies for flexible chemical processes. *Computers & Chemical Engineering*, 7(4):439–462, 1983.
- [20] S. M. Harwood, D. J. Papageorgiou, and F. Trespalacios. A note on semi-infinite program bounding methods. *Optimization Letters*, 15(4):1485–1490, 2021.
- [21] R. Hettich and K. O. Kortanek. Semi-infinite programming: theory, methods, and applications. *SIAM Review*, 35(3):380–429, 1993.
- [22] H. Hijazi, P. Bonami, and A. Ouorou. An outer-inner approximation for separable mixed-integer nonlinear programs. *INFORMS Journal on Computing*, 26(1):31–44, 2014.
- [23] R. Kannan, H. Nagarajan, and D. Deka. Learning to accelerate the global optimization of quadratically-constrained quadratic programs. *arXiv preprint arXiv:2301.00306*, 2022. doi: <https://doi.org/10.48550/arXiv.2301.00306>.

- [24] J. E. Kelley, Jr. The cutting-plane method for solving convex programs. *Journal of the Society for Industrial and Applied Mathematics*, 8(4):703–712, 1960.
- [25] M. López and G. Still. Semi-infinite programming. *European Journal of Operational Research*, 180(2):491–518, 2007.
- [26] M. M. Mäkelä. Multiobjective proximal bundle method for nonconvex nonsmooth optimization: Fortran subroutine MPB-NGC 2.0. url: http://napsu.karmita.fi/publications/pbncgc_report.pdf. *Reports of the Department of Mathematical Information Technology, Series B. Scientific Computing, B*, 13, 2003.
- [27] A. Marendet, A. Goldsztejn, G. Chabert, and C. Jermann. A standard branch-and-bound approach for nonlinear semi-infinite problems. *European Journal of Operational Research*, 282(2):438–452, 2020.
- [28] A. Mitsos. Global optimization of semi-infinite programs via restriction of the right-hand side. *Optimization*, 60(10-11):1291–1308, 2011.
- [29] A. Mitsos. A test set of semi-infinite programs (revised by Hatim Djelassi). Last Accessed on January 11, 2023. <https://www.avt.rwth-aachen.de/cms/AVT/Forschung/Systemverfahrenstechnik/~kpdo/A-Test-Set-of-Semi-Infinite-Programs/>, 2016.
- [30] A. Mitsos and A. Tsoukalas. Global optimization of generalized semi-infinite programs via restriction of the right hand side. *Journal of Global Optimization*, 61(1):1–17, 2015.
- [31] A. Mitsos, P. Lemonidis, C. K. Lee, and P. I. Barton. Relaxation-based bounds for semi-infinite programs. *SIAM Journal on Optimization*, 19(1):77–113, 2008.
- [32] B. S. Mordukhovich, N. M. Nam, and N. D. Yen. Subgradients of marginal functions in parametric mathematical programming. *Mathematical Programming*, 116(1-2):369–396, 2009.
- [33] A. Mutapic and S. Boyd. Cutting-set methods for robust convex optimization with pessimizing oracles. *Optimization Methods & Software*, 24(3):381–406, 2009.
- [34] Y. Nesterov. *Lectures on convex optimization*, volume 137. Springer, Switzerland, 2018.
- [35] M. B. Paulus, G. Zarpellon, A. Krause, L. Charlin, and C. Maddison. Learning to cut by looking ahead: Cutting plane selection via imitation learning. In *International Conference on Machine Learning*, pages 17584–17600, 2022.
- [36] E. Polak. *Optimization: algorithms and consistent approximations*. Springer New York, NY, 1997.
- [37] D. Ralph and S. Dempe. Directional derivatives of the solution of a parametric nonlinear program. *Mathematical Programming*, 70(1-3):159–172, 1995.
- [38] R. Reemtsen and J.-J. Rückmann. *Semi-infinite programming*. Springer Science & Business Media, New York, 1998.
- [39] N. V. Sahinidis. BARON: A general purpose global optimization software package. *Journal of Global Optimization*, 8(2):201–205, 1996.
- [40] T. Seidel and K.-H. Küfer. An adaptive discretization method solving semi-infinite optimization problems with quadratic rate of convergence. *Optimization*, 71(8):2211–2239, 2022.
- [41] A. Shapiro. Semi-infinite programming, duality, discretization and optimality conditions. *Optimization*, 58(2):133–161, 2009.
- [42] P. Stechliniski, K. A. Khan, and P. I. Barton. Generalized sensitivity analysis of nonlinear programs. *SIAM Journal on Optimization*, 28(1):272–301, 2018.
- [43] P. Stechliniski, J. Jäschke, and P. I. Barton. Generalized sensitivity analysis of nonlinear programs using a sequence of quadratic programs. *Optimization*, 68(2-3):485–508, 2019.
- [44] O. Stein. *Bi-level strategies in semi-infinite programming*. Springer Science & Business Media, New York, 2003.
- [45] O. Stein. How to solve a semi-infinite optimization problem. *European Journal of Operational Research*, 223(2):312–320, 2012.
- [46] O. Stein and P. Steuermann. The adaptive convexification algorithm for semi-infinite programming with arbitrary index sets. *Mathematical Programming*, 136(1):183–207, 2012.
- [47] O. Stein and G. Still. Solving semi-infinite optimization problems with interior point techniques. *SIAM Journal on Control and Optimization*, 42(3):769–788, 2003.
- [48] G. Still. Discretization in semi-infinite programming: the rate of convergence. *Mathematical Programming*, 91(1):53–69, 2001.
- [49] G. Still. Lectures on parametric optimization: An introduction. *Optimization Online*. URL: <https://optimization-online.org/2018/04/6587/>, 2018.
- [50] A. Tsoukalas and B. Rustem. A feasible point adaptation of the Blankenship and Falk algorithm for semi-infinite programming. *Optimization Letters*, 5(4):705–716, 2011.
- [51] E. M. Turan, R. Kannan, and J. Jäschke. Design of PID controllers using semi-infinite programming. volume 49 of *Computer Aided Chemical Engineering*, pages 439–444. Elsevier, 2022.
- [52] J. P. Vielma. Mixed integer linear programming formulation techniques. *SIAM Review*, 57(1):3–57, 2015.
- [53] G. A. Watson. Numerical experiments with globally convergent methods for semi-infinite programming problems. In *Semi-infinite programming and applications*, pages 193–205. Springer, Berlin, Heidelberg, 1983.