Numerical Algorithms (2022) 90:387–422
https://doi.org/10.1007/s11075-021-01192-9

ORIGINAL PAPER

A feasible proximal bundle algorithm with convexification for nonsmooth, nonconvex semi-infinite programming

Li-Ping Pang[1,2] · Qi Wu[3]

Received: 13 February 2021 / Accepted: 27 August 2021 / Published online: 22 September 2021

Abstract

We propose a new numerical method for semi-infinite optimization problems whose objective function is a nonsmooth function. Existing numerical methods for solving semi-infinite programming (SIP) problems make strong assumptions on the structure of the objective function, e.g., differentiability, or are not guaranteed to furnish a feasible point on finite termination. In this paper, we propose a feasible proximal bundle method with convexification for solving this class of problems. The main idea is to derive upper bounding problems for the SIP by replacing the infinite number of constraints with a finite number of convex relaxation constraints, introduce improvement functions for the upper bounding problems, construct cutting plane models of the improvement functions, and reformulate the cutting plane models as quadratic programming problems and solve them. The convex relaxation constraints are constructed with ideas from the αBB method of global optimization. Under mild conditions, we showed that every accumulation point of the iterative sequence is an ϵ-stationary point of the original SIP problem. Under slightly stronger assumptions, every accumulation point of the iterative sequence is a local solution of the original SIP problem. Preliminary computational results on solving nonconvex, nonsmooth constrained optimization problems and semi-infinite optimization problems are reported to demonstrate the good performance of the new algorithms.

✉ Qi Wu
wuqi@bisu.edu.cn

Li-Ping Pang
lppang@dlut.edu.cn

[1] School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China

[2] Key Laboratory for Computational Mathematics and Data Intelligence of Liaoning Province, Dalian, 116024, China

[3] Department of Basic Sciences, Beijing International Studies University, Beijing 100024, China

Keywords Nonsmooth nonconvex optimization · Semi-infinite programming · Bundle methods · Convexification approximation

1 Introduction

In this paper, we consider the following nonsmooth, nonconvex optimization problem

$$\begin{cases} \min\limits_{x \in X} \ f(x) \\ \ \text{s.t. } g(x,t) \leq 0, \ t \in T, \end{cases} \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is locally Lipschitz but potentially nonsmooth and nonconvex, $g : \mathbb{R}^n \times T \to \mathbb{R}$ is twice continuously differentiable on $X \times T$. In addition, T is a fixed, nonempty compact subset of $\mathbb{R}^p$ and X is a convex compact subset of $\mathbb{R}^n$. This problem is called the semi-infinite programming (SIP) problem. For simplicity of notation, we consider the semi-infinite problem with one semi-infinite constraint and a one-dimensional index set $T = [a, b]$, that is $p = 1$. However, the algorithm proposed in this paper might be helpful to deal with more general cases.

SIP has recently drawn a lot of attention due to its application in many different fields, including portfolio problem, robot trajectory planning, vibrating membrane problem, minimal cost control of air pollution, and optimization over probability measures (see [11, 15, 22]).

There exist a wide range of numerical methods for SIP problems. Traditional methods for SIP like discretization, exchange, and reduction-based [14, 21, 32, 39, 44] solve a sequence of finitely nonlinear programming subproblems, that is, subproblems with finitely many constraints chosen from the original constraints. In particular, in the reduction-based methods, the finite subset of constraints is generated from finding implicitly all the local maxima $t(x)$ of $g(x, \cdot)$ on T for each $x \in X$. Based on this reduction principle, a number of nonsmooth approaches have been presented [10, 24, 31]. These traditional methods suffer from the major drawback that they generally do not provide a feasible point in a finite number of iterations.

A breakthrough for SIP was achieved by Bhattacharjee et al. [4], who replaced the semi-infinite constraint by an overestimation obtained via interval extensions and presented a convergent upper bounding approach to SIP. It is the first algorithm with feasible iterates for SIP, to our knowledge. Based on this upper bounding approach, a number of algorithms for finding the global solution of SIP problems have been presented [5, 8, 27, 29, 30]. Under the assumptions that the functions f and $g(\cdot, t)$ are continuously differentiable on X for each $t \in T$, $g(x, \cdot)$ is continuous on T for each $x \in X$, there exist SIP Slater points arbitrarily close to all global minima, and the resulting finite nonlinear programs (NLPs) are solved to local optimality; the method using interval extensions with feasible iterates generates an ϵ-optimal estimate of the global solution of the SIP on finite termination [5]. The main drawback of this method is the rapid increase in the size of the lower and upper bounding problems as nodes of increasing depth are visited by the branch-and-bound procedure. Other methods to generate the global solution of the SIP problems were proposed by Mitsos and coworkers [8, 27, 30]. The upper bounding problems of these methods are not generated by interval extensions, but rather by a restriction of the constraints right-hand side by a negative value ($< -\epsilon^g$, for some $\epsilon^g > 0$). For fixed point x, these methods use three subproblems, including the lower bounding problem, the upper bounding problem, and the lower level problem. Under the assumptions that f is continuous on X, g is continuous on $X \times T$, there exists an ϵ^f-optimal SIP Slater point, and the resulting finite NLPs are solved to global optimality, the algorithms terminate finitely with a feasible point and solve the SIP approximately to global optimality. We point out that the authors in [8, 27, 29, 30] consider global solution of all subproblems, which is computationally expensive.

In contrast to these methods, the adaptive convexification method discussed in [9] does not focus on the global solution of the SIP problem, but global solutions of the lower level problem. The authors construct convex relaxations of the lower level problem using the αBB method [1, 2] of global optimization, replace the convex lower level problems with their Karush-Kuhn-Tucker (KKT) conditions, and solve the resulting mathematical programs with complementarity constraints. Under the assumptions that $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^m$ are twice continuously differentiable, the algorithm generates a feasible point and reaches a ϵ-stationary point of the original SIP problem on finite termination. Such method requires slightly stronger assumption on the constraints, but can give tighter bounds than interval extensions. Based on the method discussed in [9], a number of convexification algorithms have been presented [36, 38, 41]. We point out that the method discussed in [41] relies on constructing concave relaxations of the lower level problem and solving the resulting approximate problems with finitely many constraints.

Bundle methods are among the most efficient methods for solving nonsmooth optimization problems. See [18, 20, 25, 35] for a more detailed discussion and comparison of some bundle methods for nonsmooth convex constrained optimization. When the objective function is nonconvex, the recent results [12, 13] make assumptions on the structure of the objective function and propose a redistributed bundle method for unconstrained nonsmooth, nonconvex optimization problems. Based on this redistributed method, many nonconvex bundle approaches have been suggested. The authors [24, 43] propose proximal bundle methods for constrained nonsmooth, nonconvex optimization problems, where the objective and constraint functions are lower-C^2. Lv et al. [23] propose an inexact proximal bundle method for constrained nonsmooth, nonconvex optimization problems, where the objective and constraint functions are lower-C^1. Hoseini and Nobakhtian [16, 17] propose bundle methods for constrained nonsmooth nonconvex optimization problems, where the objective and constraint functions are regular. The methods presented in [23, 24, 31] can be used to solve the SIP problem, but the calculation of the lower level problem in these methods is costly since the lower level problem needs to be solved within a given fixed accuracy at each point. Moreover on finite termination, these methods are not guaranteed to furnish a feasible point.

In this paper, our aim is to design a feasible algorithm for nonsmooth, nonconvex SIP that combines some of the ideas of the proximal bundle methods and of the convexification methods. The main idea is to generate upper bounding problems for the SIP by replacing the lower level problem with its concave relaxations using the αBB method, and solve the upper bounding problems by a new feasible proximal bundle method. The relaxed lower level problem is a finite nonsmooth problem. Under certain assumptions given in Theorem 6 below, it can be shown that every accumulation point of the iterative sequence is an ϵ-stationary point of the original SIP problem. Under slightly stronger assumptions given in Theorem 7 below, every accumulation point of the iterative sequence is a local solution of the original SIP problem. These assumptions appeared in nonsmooth, nonconvex bundle methods cited above. Compared to the related works [23, 24, 31], our algorithm starts with a coarse subset of T and is advantageous to save computational time on constraint calculations. Moreover, we adopt an aggregation technique to control the size of subproblems flexibly. In particular, our algorithm can obtain a feasible point in finite iterations. Compared to [9, 36, 38, 41], the assumptions made on the structure of the objective function in this paper are weaker. We point out that the subproblems of our algorithm are solved by existing QP solver and do not depend on the number of the constraints of the upper bounding problem, whereas the subproblems of algorithm from [9, 36, 38, 41] are solved by the existing NLP solver and depend on the number of the constraints of the upper bounding problem. Thus, our algorithm is, naturally, significantly less time-consuming when the number of subintervals/nodes increases. The results of numerical experiment also show the good performance of the new method.

The remainder of this paper is organized as follows. The main concepts used throughout the paper are described in the next section, where we also state the methods to construct upper bounding functions of the constraints. Section 3 presents our feasible proximal bundle algorithm with convexification. Convergence of the proximal bundle method with convexification is addressed in Section 4. Section 5 presents some numerical results and Section 6 gives conclusions.

Our notation is fairly standard. The Euclidean inner product of two vectors $x, y \in \mathbb{R}^n$ is denoted by $\langle x, y\rangle := x^T y$, and the associated norm by $\|\cdot\|$. The positive-part function is denoted by $x^+ := \max\{x, 0\}$. For a set $X \in \mathbb{R}^n$, conv$X$ denotes its convex hull. The closed ball with center $x \in \mathbb{R}^n$ and radius $r > 0$ is denoted by $B(x; r)$. That is, $B(x; r) := \{y \in \mathbb{R}^n \mid \|y - x\| \le r\}$.

2 Preliminaries

In this section, we first recall some concepts and results related to nonsmooth analysis. Next, we give the necessary condition of the optimization measure. Then, we introduce the convexification method to construct the concave relaxation of the lower level problem.

2.1 Background

Definition 1 (Lipschitz continuity) A function $f : \mathbb{R}^n \to \mathbb{R}$ is *Lipschitz continuous* on a set $X \subset \mathbb{R}^n$ if there exists $L > 0$ such that

$$|f(y) - f(z)| \le L\|y - z\| \text{ for all } y, z \in X.$$

Then, L is called a *Lipschitz constant* for f on X.

Definition 2 The *directional derivative* of f at x in the direction $d \in \mathbb{R}^n$ is defined by

$$f'(x; d) := \lim_{h \to 0} \frac{f(x + hd) - f(x)}{h}.$$

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz continuous function at a point $x \in \mathbb{R}^n$. Then, the *generalized directional derivative* (Clarke) of f at x in the direction of $d \in \mathbb{R}^n$ is defined by

$$f^\circ(x; d) := \limsup_{\substack{y \to x \\ h \downarrow 0}} \frac{f(y + hd) - f(y)}{h}.$$

Definition 3 A function $f : \mathbb{R}^n \to \mathbb{R}$ is called *regular* at $x \in \mathbb{R}^n$ if it is locally Lipschitz continuous at x and for all $d \in \mathbb{R}^n$ the classical directional derivative $f'(x; d)$ exists and we have $f'(x; d) = f^\circ(x; d)$.

Definition 4 A function $f : \mathbb{R}^n \to \mathbb{R}$ is called *semismooth* at $x \in \mathbb{R}^n$ if f is Lipschitz on a ball about x and for each $d \in \mathbb{R}^n$ and for any sequences $\{h_k\} \subset \mathbb{R}^+$, $\{\theta_k\} \subset \mathbb{R}^n$ and $\{\xi_k\} \subset \mathbb{R}^n$ such that $\{h_k\} \downarrow 0$, $\{\theta_k / h_k\} \to 0 \in \mathbb{R}^n$ and $\xi_k \in \partial f(x + h_k d + \theta_k)$, the sequence $\{\langle \xi_k, d \rangle\}$ has exactly one accumulation point.

Definition 5 A locally Lipschitz function $f : O \to \mathbb{R}$, where O is an open subset of $\mathbb{R}^n$, is called *lower-*C^1 *(lower-*C^2*)* on O, if on some neighborhood V of each $x_0 \in O$ there is a representation $f(x) = \max_{\omega \in \Omega} f_\omega(x)$ in which the functions f_ω are of class (twice) continuously differentiable on V and the index set Ω is a compact space such that $f_\omega(x)$ and $\nabla f_\omega(x)$ $(\nabla_2 f_\omega(x))$ depend continuously not just on $x \in V$ but jointly on $(\omega, x) \in \Omega \times V$.

Definition 6 Let $f : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz continuous function at a point $x \in \mathbb{R}^n$. Then, the *subdifferential* of f at x is defined by

$$\partial f(x) := \{\xi \in \mathbb{R}^n \mid f^\circ(x; d) \geq \xi^T d \text{ for all } d \in \mathbb{R}^n\}.$$

Each element $\xi \in \partial f(x)$ is called a subgradient of f at x. If $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function, then the subdifferential of f at $x \in \mathbb{R}^n$ is defined by

$$\partial f(x) := \{\, \xi \mid f(y) \geq f(x) + \langle \xi, y - x \rangle \text{ for all } y \in \mathbb{R}^n \,\}.$$

Theorem 1 [3, Theorem 3.3] *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz continuous function at $x \in \mathbb{R}^n$ with a Lipschitz constant L. Then, the subdifferential $\partial f(x)$ is a nonempty, convex, and compact set such that*

$$\partial f(x) \subseteq B(0; L).$$

Theorem 2 [3, Theorem 3.23] *Let $f_i : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz continuous at x for all $i = 1, ..., m$. Then, the function*

$$f(x) = \max\{f_i(x) \mid i = 1, ..., m\}$$

is locally Lipschitz continuous at x and

$$\partial f(x) \subseteq \mathrm{conv}\{\partial f_i(x) \mid i \in \mathcal{I}(x)\}, \tag{2}$$

where

$$\mathcal{I}(x) := \{i \in \{1, ..., m\} \mid f_i(x) = f(x)\}.$$

In addition, if f_i is regular at x for all $i = 1, ..., m$, then f is also regular at x and equality holds in (2).

Now, we introduce the indicator function I_X of a set $X \subset \mathbb{R}^n$ as follows:

$$I_X(x) = \begin{cases} 0 & \text{if } x \in X, \\ \infty & \text{if } x \notin X. \end{cases}$$

The indicator function I_X of a set $X \subset \mathbb{R}^n$ is convex if and only if X is convex. Let the point x belong to X, then

$$\partial I_X(x) = N_X(x) := \{v \in \mathbb{R}^n \mid \langle v, y - x\rangle \leq 0, \text{ for all } y \in X\}.$$

From [26, Proposition 4] and [3, Theorem 3.13], we have the following:

Theorem 3 *If $f : X \to \mathbb{R}$ is continuously differentiable then f is semismooth and regular on X.*

From [37, Proposition 2.4] and [7, Theorem 2], we have the following:

Proposition 1 *Given an open set O containing X. The following statements are equivalent:*

1. *f is semismooth and regular on O.*
2. *For all $x \in O$, for all $\epsilon > 0$, there exists $\rho > 0$ such that for all $y \in B(x; \rho)$ and $\xi \in \partial f(y)$*

$$f(z) - f(y) \geq \langle \xi, z - y\rangle - \epsilon\|z - y\|, \qquad z \in B(x; \rho).$$

3. *f is lower-C^1 on O.*

Now we return to our basic problem (1), and give the following assumption.

Assumption 1 *The Slater constraint qualification holds for problem (1), i.e., there exists $\hat{x} \in \mathbb{R}^n$ such that $g(\hat{x}, t) < 0$ for all $t \in T$.*

Assumption 1 was used in [41]. In this paper, we always assume that the Slater constraint qualification holds.

2.2 Necessary conditions

To solve problem (1), we introduce the improvement function

$$H(y, x) := \max\{f(y) - f(x), G(y)\}, \quad y \in \mathbb{R}^n, \tag{3}$$

where

$$G(y) = \max_{t \in T} g(y, t).$$

Now, the subdifferential of G at a point x is defined by [6]

$$\partial G(x) = \text{conv}\left\{\nabla_x g(x, t) \mid t \in A(x)\right\}, \tag{4}$$

where $A(x) = \{t \in T \mid G(x) = g(x, t)\}$.

Definition 7 [42](EMFCQ) A feasible point x is said to satisfy the extended Mangasarian-Fromovitz constraint qualification for SIP problem (1) if there exists a feasible direction d of X at x such that

$$\nabla_x g(x, t)^T d < 0, \text{ for all } t \in T_{act}(x), \tag{5}$$

where $T_{act}(x) = \{t \in T \mid g(x, t) = 0\}$.

Theorem 4 *Let x^* be a local minimizer of (1). Then the following statements hold:*

(a) $0 \in \partial H(x^*, x^*) + \partial I_X(x^*)$.
(b) There exist nonnegative multipliers λ_i, $i = 0, ..., n+1$, and indices $t_i \in A(x^*)$, $i = 1, ..., n+1$ such that $\sum_{i=0}^{n+1} \lambda_i = 1$, $\lambda_i g(x^*, t_i) = 0$, $i = 1, ..., n+1$ and

$$0 \in \lambda_0 \partial f(x^*) + \sum_{i=1}^{n+1} \lambda_i \nabla_x g(x^*, t_i) + \partial I_X(x^*).$$

(c) If EMFCQ holds at x^* for problem (1), then there exist nonnegative multipliers λ_i and indices $t_i \in T_{act}(x^*)$, $i = 1, ..., n+1$ such that

$$0 \in \partial f(x^*) + \sum_{i=1}^{n+1} \lambda_i \nabla_x g(x^*, t_i) + \partial I_X(x^*).$$

Proof (a) Let F be the feasible region of the problem (1), namely, $F = \{x \in X \mid G(x) \le 0\}$. Since $x^* \in \mathbb{R}^n$ is a local optimum of the problem (1), then $G(x^*) \le 0$ and there exists a neighborhood U of x^* such that $f(x^*) \le f(x)$ for all $x \in U \cap F$. Then, for all $x \in U \cap F \subseteq X$

$$H(x, x^*) = \max\{f(x) - f(x^*),\ G(x)\} \ge f(x) - f(x^*) \ge 0 = H(x^*, x^*).$$

This means that x^* is a local minimum of $H(\cdot, x^*)$ on X, which implies $0 \in \partial H(x^*, x^*) + \partial I_X(x^*)$.

(b) If $G(x^*) < 0$, we have $\partial H(x^*, x^*) \subseteq \partial f(x^*)$. Then, the assertion of the theorem is proved by choosing $\lambda_0 = 1$ and $\lambda_i = 0$ for $i = 1, ..., n+1$. If $G(x^*) = 0$, we have $\partial H(x^*, x^*) \subseteq \text{conv}\{\partial f(x^*) \cup \partial G(x^*)\}$, where $\partial G(x^*) = \text{conv}\{\nabla g(x^*, t_i),\ t_i \in A(x^*)\}$. Then due to the definition of convex hull and the Caratheodory's Theorem, there exist $\lambda_i \ge 0$, $i = 0, 1, \ldots, n+1$, such that

$$0 \in \lambda_0 \partial f(x^*) + \sum_{i=1}^{n+1} \lambda_i \nabla_x g(x^*, t_i) + \partial I_X(x^*).$$

(c) Suppose that $\lambda_0 = 0$, then we deduce that $G(x^*) = 0$ and $A(x^*) = T_{act}(x^*)$. Moreover, $-\sum_{i=1}^{n+1} \lambda_i \nabla_x g(x^*, t_i) \in \partial I_X(x^*)$. For any feasible direction d of X at x^*, we have

$$\left\langle -\sum_{i=1}^{n+1} \lambda_i \nabla_x g(x^*, t_i), d \right\rangle \leq 0, \quad t_i \in T_{act}(x^*),$$

which is an contradiction with (5). Thus, λ_0 is strictly positive, and the result item (c) holds by setting $\lambda_i = \lambda_i/\lambda_0$. □

Definition 8 x^* is called an ϵ-KKT point of (1) if there exist some multipliers λ_i, $i = 1, ..., n+1$ such that

$$0 \in \partial f(x^*) + \sum_{i=1}^{n+1} \lambda_i \nabla_x g(x^*, t_i) + \partial I_X(x^*), \;\; t_i \in T^{\epsilon}_{act}(x^*),$$

where $T^{\epsilon}_{act}(x^*) = \{t \in T \mid g(x^*, t) \in [-\epsilon, 0]\}$.

2.3 The concave relaxation of the lower level problem

In this subsection, we consider the following lower level problem

$$Q(x): \qquad \max_{t \in T} g(x, t).$$

We note that the lower level problem is actually a concave optimization problem for all $x \in X$ if $g(x, \cdot)$ is convex on T for these $x \in X$.

For $N \in \mathbb{N}$, let $a = t_0 < t_1 < ... < t_N = b$ define a subdivision E of T:

$$E := \{t_u \mid u \in U' = \{0, 1, ..., N\}\}.$$

Then, we have

$$\max_{t \in T} g(x, t) = \max_{u \in U} \max_{t \in T_u} g(x, t),$$

where $U = \{1, 2, ..., N\}$ and $T_u = [t_{u-1}, t_u]$. T_u is called a subinterval of T and satisfies

$$T = \bigcup_{u \in U} T_u.$$

The length of the subinterval T_u is defined by $|T_u| = |t_u - t_{u-1}|$ and the length of the subdivision E is defined by $|E| = \max_{u \in U} |T_u|$.

Following the approach of [41], for all $u \in U$, we generate the upper bound function g_u of g on T_u,

$$g_u(x, t) = g(x, t) + \frac{\alpha_u}{2}(t - \frac{t_{u-1} + t_u}{2})^2, \;\; \forall \; x \in X.$$

In this case, the function g_u is twice continuously differentiable with respect to t and the second-order derivative with respect to t of g_u is

$$\nabla_t^2 g_u(x, t) = \nabla_t^2 g(x, t) + \alpha_u.$$

It follows that $g_u(x,t)$ is convex with respect to t on T_u if the parameter α_u satisfies

$$\alpha_u \geq \max_{(x,t)\in X\times T_u} \left\{-\nabla_t^2 g(x,t)\right\}.$$

On the other hand, we require to construct an upper bound function of g. Thus,

$$\alpha_u \geq \max \left\{0, \max_{(x,t)\in X\times T_u} \left\{-\nabla_t^2 g(x,t)\right\}\right\}, \quad \forall\ u \in U. \tag{6}$$

Lemma 1 [41] *Let f be a convex function and $C = convS$, where S is an arbitrary set of points. Then*

$$\sup\{f(x) \mid x \in C\} = \sup\{f(x) \mid x \in S\},$$

where the first supremum is attained only when the second supremum is attained.

We know from Lemma 1 that the maximum of $g_u(x,t)(\forall\ u \in U)$ must be attained on the boundary of T_u, that is,

$$\begin{aligned}\max_{t\in T_u} g_u(x,t) = \max_{t\in\{t_{u-1},t_u\}} g_u(x,t) &= \max\left\{g(x,t_{u-1}), g(x,t_u)\right\} \\ &+\frac{\alpha_u}{8}(t_u - t_{u-1})^2, \quad \forall\ x \in X.\end{aligned}$$

Then, we have

$$\max_{t\in\{t_{u-1},t_u\}} g(x,t) \leq \max_{t\in T_u} g(x,t) \leq \max_{t\in\{t_{u-1},t_u\}} g_u(x,t), \quad \forall\ x \in X.$$

The convex upper bound function $G(\cdot,\alpha,E)$ of G on X is defined by

$$G(x,\alpha,E) = \max_{u\in U}\max_{t\in T_u} g_u(x,t).$$

It can be rewritten as

$$G(x,\alpha,E) = \max_{t_u\in E}\left\{g(x,t_u) + h(t_u,\alpha_u)\right\}, \tag{7}$$

where

$$h(t_u,\alpha_u) = \begin{cases} \frac{\alpha_{u+1}}{8}(t_{u+1}-t_u)^2, & \text{if } u = 0, \\ \max\left\{\frac{\alpha_u}{8}(t_u - t_{u-1})^2, \frac{\alpha_{u+1}}{8}(t_{u+1}-t_u)^2\right\}, & \text{if } 0 < u < N, \\ \frac{\alpha_u}{8}(t_u - t_{u-1})^2, & \text{if } u = N, \end{cases} \tag{8}$$

where the parameter α_u satisfies (6). We define

$$\bar{A}(x,\alpha,E) = \{t_u \in E \mid G(x,\alpha,E) = g(x,t_u) + h(t_u,\alpha_u)\}, \tag{9}$$

and

$$D(E,\alpha) = \max_{u\in U}\frac{\alpha_u}{8}(t_u - t_{u-1})^2. \tag{10}$$

Then, we can obtain that

$$G(x) \leq G(x,\alpha,E) \leq G(x) + D(E,\alpha). \tag{11}$$

Now the upper bounding problem of the original problem is defined by

$$\begin{cases} \min\limits_{x\in X} \ f(x) \\ \text{s.t.} \ \ G(x,\alpha,E) \le 0. \end{cases} \tag{12}$$

Using the improvement function mentioned in the preceding section, we define $H(\cdot, x, \alpha, E) : \mathbb{R}^n \to \mathbb{R}$ as follows:

$$H(y, x, \alpha, E) := \max\{\, f(y) - f(x)\,,\ G(y, \alpha, E)\,\}\,,$$

where α and E vary along the iterations.

Lemma 2 [41] *If EMFCQ holds for all feasible points of the original problem* (1)*, then the MFCQ holds for all feasible points of the approximation problem* (12) *when* $|E|$ *is sufficiently small.*

Theorem 5 *Suppose that* $f : X \to \mathbb{R}$ *and* $g : X \times T \to \mathbb{R}$ *are locally Lipschitz continuous, then the following statements hold:*

1. *If* x^* *is a local minimizer of* (12)*, then* $0 \in \partial H(x^*, x^*, \alpha, E) + \partial I_X(x^*)$.
2. *If* $0 \in \partial H(x^*, x^*, \alpha, E) + \partial I_X(x^*)$ *and* $G(x^*, \alpha, E) \le 0$*, then there exist nonnegative multipliers* λ_i*,* $i = 0, ..., n+1$*, and indices* $t_i \in \bar{A}(x^*, \alpha, E)$*,* $i = 1, ..., n+1$*, such that* $\sum_{i=0}^{n+1} \lambda_i = 1$*,* $\lambda_i(g(x^*, t_i) + h(t_i, \alpha_i)) = 0$*,* $i = 1, ..., n+1$*, and*

$$0 \in \lambda_0 \partial f(x^*) + \sum_{i=1}^{n+1} \lambda_i \nabla_x g(x^*, t_i) + \partial I_X(x^*).$$

If, in addition, MFCQ holds at x^* *for problem* (12)*, then there exist nonnegative multipliers* λ_i *and indices* $t_i \in \bar{T}_{act}(x^*, \alpha, E)$*,* $i = 1, ..., n+1$*, such that*

$$0 \in \partial f(x^*) + \sum_{i=1}^{n+1} \lambda_i \nabla_x g(x^*, t_i) + \partial I_X(x^*),$$

where $\bar{T}_{act}(x^*, \alpha, E) = \{t_u \in E \mid g(x^*, t_u) + h(t_u, \alpha_u) = 0\}$*. If, in addition,* $D(E, \alpha) \le \epsilon$*, then* x^* *is an* ϵ*-KKT point of the problem* (1).

Proof The proof of this theorem is analogous to Theorem 4; we only require to show that x^* is an ϵ-KKT point of the original problem if $D(E, \alpha) \le \epsilon$. It suffices to prove that if $t_i \in \bar{T}_{act}(x^*, \alpha, E)$, $i = 1, ..., n+1$, then $t_i \in T^{\epsilon}_{act}(x^*)$, $i = 1, ..., n+1$, provided that condition $D(E, \alpha) \le \epsilon$ holds. For each $t_i \in \bar{T}_{act}(x^*, \alpha, E)$, we have

$$g(x^*, t_i) + \frac{\alpha_{\bar{i}}}{8}|T_{\bar{i}}|^2 = 0,$$

where

$$\bar{i} = \begin{cases} i & \text{if } \dfrac{\alpha_i}{8}|T_i|^2 \ge \dfrac{\alpha_{i+1}}{8}|T_{i+1}|^2, \\ i+1 & \text{otherwise,} \end{cases}$$

where $T_0 = 0$ and $T_{N+1} = 0$. It implies that $g(x^*, t_i) \leq 0$ and

$$g(x^*, t_i) = -\frac{\alpha_{\bar{i}}}{8}|T_{\bar{i}}|^2 \geq -D(E, \alpha) \geq -\epsilon.$$

It follows that $t_i \in T^{\epsilon}_{act}(x^*)$, $i = 1, ..., n+1$, if the condition $D(E, \alpha) \leq \epsilon$ holds. □

3 The numerical approach

Let k be the current iteration index and x^k be the stability center (or serious iterate) at iteration k, where iterations generate a sequence of trial points $\{y^j\}_{j\in J_k^{ora}}$, with $J_k^{ora} \subseteq \{0, 1, ..., k\}$. In particular, $x^k \in \{y^j\}_{j\in J_k^{ora}}$. Let $E^k = \{t_u \mid u \in U'_k = \{0, 1, ..., N_k\}\}$ be the current subdivision of T and α^k be the corresponding convexification parameter.

Following the redistributed proximal approach [12, 13], we generate the augmented functions of f and $G(\cdot, \alpha^k, E^k)$ defined by

$$\begin{aligned} f^k(\cdot) &:= \quad f(\cdot) + \tfrac{\eta_1^k}{2}\|\cdot - x^k\|^2, \\ g^k(\cdot) &:= G(\cdot, \alpha^k, E^k) + \tfrac{\eta_2^k}{2}\|\cdot - x^k\|^2. \end{aligned}$$

The idea is to utilize the augmented functions in the model construction. We can first construct a convex piecewise linear approximation defined by

$$\begin{aligned} \hat{H}_k(y) := \max_{j\in J_k^{ora}} \Big\{ & f^k(y^j) + \left\langle \xi_f^j + \eta_1^k(y^j - x^k), y - y^j \right\rangle - f(x^k), \\ & g^k(y^j) + \left\langle \xi_g^{k,j} + \eta_2^k(y^j - x^k), y - y^j \right\rangle \Big\}, \end{aligned}$$

where $\xi_f^j \in \partial f(y^j)$ and $\xi_g^{k,j} \in \partial G(y^j, \alpha^k, E^k)$. It can be rewritten in an equivalent form

$$\hat{H}_k(y) = G^+(x^k, \alpha^k, E^k) + \max_{j\in J_k^{ora}} \left\{ -c_{k,j}^f + \langle s_f^{k,j}, y - x^k\rangle, -c_{k,j}^g + \langle s_g^{k,j}, y - x^k\rangle \right\}. \tag{13}$$

where

$$\begin{aligned} s_f^{k,j} &:= \xi_f^j + \eta_1^k \Delta_k^j \;\text{ and }\; c_{k,j}^f := e_{k,j}^f + \eta_1^k b_k^j + G^+(x^k, \alpha^k, E^k), \\ s_g^{k,j} &:= \xi_g^{k,j} + \eta_2^k \Delta_k^j \;\text{ and }\; c_{k,j}^g := e_{k,j}^g + \eta_2^k b_k^j + G^+(x^k, \alpha^k, E^k) - G(x^k, \alpha^k, E^k), \end{aligned} \tag{14}$$

with

$$\begin{aligned} e_{k,j}^f &:= f(x^k) - f(y^j) - \langle \xi_f^j, x^k - y^j\rangle, \\ e_{k,j}^g &:= G(x^k, \alpha^k, E^k) - G(y^j, \alpha^k, E^k) - \langle \xi_g^{k,j}, x^k - y^j\rangle, \\ b_k^j &:= \frac{1}{2}\|y^j - x^k\|^2, \;\; \Delta_k^j := y^j - x^k. \end{aligned} \tag{15}$$

Any choice for the parameters η_1^k, η_2^k that keeps $c_{k,j}^f, c_{k,j}^g$ in (13) nonnegative is acceptable. In our method, we take

$$\eta_1^k = \max\left\{ \max_{\substack{j \in J_k^{ora} \\ y^j \neq x^k}} \left\{ -\frac{e_{k,j}^f}{b_k^j} \right\}, 0 \right\} + \gamma, \quad \eta_2^k = \max\left\{ \max_{\substack{j \in J_k^{ora} \\ y^j \neq x^k}} \left\{ -\frac{e_{k,j}^g}{b_k^j} \right\}, 0 \right\} + \gamma, \tag{16}$$

where γ is a small positive parameter. In particular, since $x^k \in \{y^j\}_{j \in J_k^{ora}}$, we obtain that

$$\hat{H}_k(x^k) = G^+(x^k, \alpha^k, E^k) = H(x^k, x^k, \alpha^k, E^k). \tag{17}$$

Then, we seek for the new point y^{k+1} as a solution of

$$\begin{cases} \min \ \hat{H}_k(y) + \frac{1}{2}\mu_k \|y - x^k\|^2 \\ \text{s.t. } \ y \in X, \end{cases} \tag{18}$$

where $\mu_k > 0$ is a proximal parameter.

The following lemma introduces the characterization of the solution of the (18).

Lemma 3 *Let y^{k+1} be the unique solution to (18) and assume $\mu_k > 0$. Then, we have that*

$$y^{k+1} = x^k - \frac{1}{\mu_k}(S^{k+1} + v^{k+1}), \ \textit{where } S^{k+1} \in \partial \hat{H}_k(y^{k+1}), \ v^{k+1} \in \partial I_X(y^{k+1}). \tag{19}$$

Lemma 3 implies that

$$\langle v^{k+1}, y^{k+1} - x^k \rangle \geq 0. \tag{20}$$

The linearization error of $\hat{H}_k(\cdot)$ is defined by

$$C^{k+1} := \hat{H}_k(x^k) - \hat{H}_k(y^{k+1}) - \langle S^{k+1}, x^k - y^{k+1} \rangle \geq 0. \tag{21}$$

Since the piecewise linear model $\hat{H}_k$ is convex, we can construct a convex linear approximation of $\hat{H}_k$ by

$$\hat{H}_k(y) \geq \hat{H}_k(y^{k+1}) + \langle S^{k+1}, y - y^{k+1} \rangle = G^+(x^k, \alpha^k, E^k) + \langle S^{k+1}, y - x^k \rangle - C^{k+1}.$$

Hence, the cutting plane model can be redefined in the form

$$\begin{aligned} \hat{H}_k(y) \ = \ G^+(x^k, \alpha^k, E^k) + \max\Bigg\{ & \max_{j \in J_k^{agg}} \left\{ -C^j + \langle S^j, y - x^k \rangle \right\}, \\ & \max_{j \in J_k^{ora}} \left\{ -c_{k,j}^f + \langle s_f^{k,j}, y - x^k \rangle, -c_{k,j}^g + \langle s_g^{k,j}, y - x^k \rangle \right\} \Bigg\}, \end{aligned} \tag{22}$$

where $J_k^{ora} \subseteq \{0, 1, ..., k\}$ and $J_k^{agg} \subseteq \{1, ..., k\}$. Then, (18) can be stated as a quadratic programming problem in $\mathbb{R}^1 \times X$:

$$\text{compute } (r, y) := \begin{cases} \arg\min \ r + \frac{\mu_k}{2}\|y - x^k\|^2 \\ \quad \text{s.t. } \ G^+(x^k, \alpha^k, E^k) - c_{k,j}^f + \langle s_f^{k,j}, y - x^k \rangle \leq r, \ \ j \in J_k^{ora}, \\ \quad\quad\quad G^+(x^k, \alpha^k, E^k) - c_{k,j}^g + \langle s_g^{k,j}, y - x^k \rangle \leq r, \ \ j \in J_k^{ora}, \\ \quad\quad\quad G^+(x^k, \alpha^k, E^k) - \ C^j \ + \langle \ S^j \ , y - x^k \rangle \leq r, \ \ j \in J_k^{agg}. \end{cases} \tag{23}$$

From the KKT conditions for (23), one easily obtains a representation for S^{k+1} and C^{k+1} (defined by (19) and (21)):

$$
\begin{aligned}
S^{k+1} &= \sum_{j\in J_k^{ora}} \left(\lambda_{1,k+1}^j s_f^{k,j} + \lambda_{2,k+1}^j s_g^{k,j}\right) + \sum_{j\in J_k^{agg}} \lambda_{3,k+1}^j S^j,\\
C^{k+1} &= \sum_{j\in J_k^{ora}} \left(\lambda_{1,k+1}^j c_{k,j}^f + \lambda_{2,k+1}^j c_{k,j}^g\right) + \sum_{j\in J_k^{agg}} \lambda_{3,k+1}^j C^j,
\end{aligned}
\tag{24}
$$

and $(\lambda_{1,k+1}, \lambda_{2,k+1}, \lambda_{3,k+1})$ is a solution to

$$
\begin{cases}
\min \ \dfrac{1}{2}\left\| \sum_{j\in J_k^{ora}} \left(\lambda_{1,k+1}^j s_f^{k,j} + \lambda_{2,k+1}^j s_g^{k,j}\right) + \sum_{j\in J_k^{agg}} \lambda_{3,k+1}^j S^j \right\|^2 \\
\qquad + \mu_k \left(\sum_{j\in J_k^{ora}} \left(\lambda_{1,k+1}^j c_{k,j}^f + \lambda_{2,k+1}^j c_{k,j}^g\right) + \sum_{j\in J_k^{agg}} \lambda_{3,k+1}^j C^j \right) \\
\text{s.t.} \ \sum_{j\in J_k^{ora}} \left(\lambda_{1,k+1}^j + \lambda_{2,k+1}^j\right) + \sum_{j\in J_k^{agg}} \lambda_{3,k+1}^j = 1, \\
\qquad \lambda_{1,k+1}^j,\ \lambda_{2,k+1}^j \geq 0 \text{ for all } j \in J_k^{ora} \text{ and } \lambda_{3,k+1}^j \geq 0 \text{ for all } j \in J_k^{agg}.
\end{cases}
\tag{25}
$$

To measure the quality of the candidate, the nominal decrease δ_{k+1} is defined by:

$$
\delta_{k+1} := C^{k+1} + \frac{1}{\mu_k}\|S^{k+1} + v^{k+1}\|^2 \geq 0. \tag{26}
$$

According to (20), (21), and Lemma 3, we can obtain that

$$
\begin{aligned}
\delta_{k+1} &= H(x^k, x^k, \alpha^k, E^k) - \hat{H}_k(y^{k+1}) - \langle v^{k+1}, y^{k+1} - x^k\rangle \\
&\leq H(x^k, x^k, \alpha^k, E^k) - \hat{H}_k(y^{k+1}).
\end{aligned}
\tag{27}
$$

We now present the criterion that determines whether a new stability center is generated:

$$
\max\left\{ f(y^{k+1}) - f(x^k)\, , G(y^{k+1}, \alpha^k, E^k) \right\} - G^+(x^k, \alpha^k, E^k) \leq -m\delta_{k+1}, \tag{28}
$$

where m is a positive descent parameter. If (28) holds, the trial point y^{k+1} brings a sufficient descent for the objective or constraint function. Then, the stability center will be moved to y^{k+1}, i.e., $x^{k+1} := y^{k+1}$. This case is called a serious step (or descent step). Otherwise, the trial point y^{k+1} is used to enrich the model and the stability center is not changed. This case is called a null step.

In the following, we introduce a subdivision strategy to refine the current subdivision and construct a refined approximation of the lower level problem by the following procedure.

Refinement procedure

1. For a given subset $T_u = [t_{u-1}, t_u]$, let α_u be the corresponding convexification parameter.
2. Define

$$
E^{\alpha BB}(T_u) = \left\{ t_{u,i} \mid t_{u,i} = t_{u-1} + \frac{i}{2s+1}(t_u - t_{u-1}), i = 1, ..., 2s,\ s \in \mathbb{N} \right\}. \tag{29}
$$

3. Put $T_{u,i} = [t_{u,i-1}, t_{u,i}]$, $i = 1, ..., 2s+1$, where $t_{u,0} = t_{u-1}$ and $t_{u,2s+1} = t_u$.

4. Determine the corresponding convexification parameters $\alpha_{u,i}$, $i = 1, ..., 2s + 1$, such that $\alpha_{u,i} \le \alpha_u$ and

$$\alpha_{u,i} \ge \max \left\{0, \max_{(x,t)\in X\times T_{u,i}} \left\{-\nabla_t^2 g(x,t)\right\}\right\}. \tag{30}$$

Remark 1 Note that the computation of $\alpha_{u,i}$ involves a global optimization problem. However, we can use an upper bound for the right-hand side in (30). In particular, we can choose $\alpha_{u,i} = \alpha_u$. Therefore, the computation of the globally optimal value of the lower level problem $\max_{t\in T}\{g(x,t)\}$ uses more computation efforts than the computation of $\alpha_{u,i}$.

Now, we show that our refinement procedure brings a sufficient descent for the new constraint function.

Lemma 4 *Assume that* $T_{u,i} = [t_{u,i-1}, t_{u,i}]$ *and* $\alpha_{u,i}$, $i = 1, ..., 2s + 1$, *are obtained by taking the refinement procedure on subset* $T_u = [t_{u-1}, t_u]$. *Then, we have*

$$g(x,t) \le g_{1,i}(x,t) \le g_2(x,t), \ \textit{for all } t \in [t_{u,i-1}, t_{u,i}], \ i = 1, ..., 2s + 1,$$

where

$$\begin{aligned} g_{1,i}(x,t) &= g(x,t) + \frac{\alpha_{u,i}}{2}\left(t - \frac{t_{u,i-1}+t_{u,i}}{2}\right)^2, \\ g_2(x,t) &= g(x,t) + \frac{\alpha_u}{2}\left(t - \frac{t_{u-1}+t_u}{2}\right)^2. \end{aligned}$$

Proof The first inequality can be obtained directly by the definitions of $g_{1,i}$, $i = 1, ..., 2s+1$. We set $\psi_i(t) = \frac{\alpha_{u,i}}{2}(t - \frac{t_{u,i-1}+t_{u,i}}{2})^2 - \frac{\alpha_u}{2}(t - \frac{t_{u-1}+t_u}{2})^2$. Then, we only need to prove that $\psi_i(t) \le 0$ for any $i = 1, ..., 2s + 1$. We note that $|t - \frac{t_{u,i-1}+t_{u,i}}{2}| \le \frac{t_u - t_{u-1}}{2(2s+1)} \le \frac{t_{u-1}+t_u}{2} - t$ for all $t \in [t_{u,i-1}, t_{u,i}]$, $i = 1, ..., s$. Then, for any $i \le s$, we have

$$\begin{aligned} \nabla_t \psi_i(t) &= \alpha_{u,i}\left(t - \frac{t_{u,i-1}+t_{u,i}}{2}\right) - \alpha_u\left(t - \frac{t_{u-1}+t_u}{2}\right) \\ &= \alpha_{u,i}\left(t - \frac{t_{u,i-1}+t_{u,i}}{2}\right) + \alpha_u\left(\frac{t_{u-1}+t_u}{2} - t\right) \\ &\ge 0. \end{aligned}$$

Therefore, we conclude that $\psi_i(t)$ is nondecreasing on $[t_{u,i-1}, t_{u,i}]$, $i = 1, ..., s$. It implies that $\psi_i(t) \le \psi_i(t_{u,i})$ for all $t \in [t_{u,i-1}, t_{u,i}]$, $i = 1, ..., s$. For any $i \le s$, we have

$$\begin{aligned} \psi_i(t_{u,i}) &= \frac{\alpha_{u,i}}{2}\left(\frac{t_{u,i} - t_{u,i-1}}{2}\right)^2 - \frac{\alpha_u}{2}\left(t_{u,i} - \frac{t_{u-1}+t_u}{2}\right)^2 \\ &\le \frac{\alpha_{u,i}}{2}\left(\left(\frac{t_{u,i} - t_{u,i-1}}{2}\right)^2 - \left(t_{u,i} - \frac{t_{u-1}+t_u}{2}\right)^2\right) \\ &= \frac{\alpha_{u,i}}{2}\left(\left(\frac{t_{u,i} - t_{u,i-1}}{2}\right)^2 - \left(\frac{t_u - t_{u-1}}{2(2s+1)}\right)^2\right) \\ &\le 0. \end{aligned}$$

Then, we have $\psi_i(t) \leq 0$ for any $[t_{u,i-1}, t_{u,i}]$, $i = 1, ..., s$. Similarly, we can prove that $\psi_i(t) \leq 0$ for any $[t_{u,i-1}, t_{u,i}]$, $i = s+2, ..., 2s+1$. For $i = s+1$, we have

$$\begin{aligned}\psi_i(t) &= \frac{\alpha_{u,i}}{2}\left(t - \frac{t_{u,i-1}+t_{u,i}}{2}\right)^2 - \frac{\alpha_u}{2}\left(t - \frac{t_{u-1}+t_u}{2}\right)^2 \\ &= \left(\frac{\alpha_{u,i}}{2} - \frac{\alpha_u}{2}\right)\left(t - \frac{t_{u-1}+t_u}{2}\right)^2 \\ &\leq 0.\end{aligned}$$

The proof is complete. □

Now, we present the algorithm in full detail.

Algorithm 1 (Feasible proximal bundle method with convexification for SIP problems)
Data. $m \in (0, 1)$, $tol_1 \geq 0$, $tol_2 > 0$, $tol_3 \geq 0$, $\mu_{\max} > \mu_{\min} > 0$, $\mu_0 \in [\mu_{\min}, \mu_{\max}]$, $\eta_1^0 \geq 0$, $\eta_2^0 \geq 0$, $\gamma > 0$, $s > 0$, $i_{n\max} > 0$, $\tilde{X} \subset \mathbb{R}$. Set $k = 0$, $N_0 = 2s+1$, and

$$E^0 = \left\{t_u \mid t_u = a + \frac{u}{N_0}(b-a),\ u = 0, 1, ..., N_0\right\} \subset T.$$

Determine the parameter α^0 with (6) on E^0.

Step 0 (Initialization) Solve the problem

$$\begin{cases} \min\limits_{(x,z)\in X\times\tilde{X}} & z \\ \text{s.t.} & G(x, \alpha^0, E^0) - z \leq 0. \end{cases} \tag{31}$$

and from the solution $\tilde{x}$ compute $G(\tilde{x}, \alpha^0, E^0)$. Set $y^0 = \tilde{x}$, $G(y^0, \alpha^0, E^0) = G(\tilde{x}, \alpha^0, E^0)$. If $z > 0$, set $i_z = 1$, $J_1^{ora} = \{0\}$, and go to **Step 7**. Otherwise, compute $f(y^0)$, $\xi_f^0 \in \partial f(y^0)$, $\xi_g^{0,0} \in \partial G(y^0, \alpha^0, E^0)$ and set $x^0 = y^0$, $f(x^0) = f(y^0)$, $G(x^0, \alpha^0, E^0) = G(y^0, \alpha^0, E^0)$. Define the associated $e^f_{0,0}$, $e^g_{0,0}$, b_0^0, Δ_0^0 by (15). Set $i_z = 0$, $i_x = 0$, $i_n = 0$, $J_0^{ora} = \{0\}$, $J_0^{agg} = \emptyset$ and go to **Step 1.**

Step 1 (Quadratic programming subproblem) Given the model $\hat{H}_k$ defined by (14) and (22), compute S^{k+1}, C^{k+1}, y^{k+1}, $(\lambda_{1,k+1}, \lambda_{2,k+1}, \lambda_{3,k+1})$ by solving the subproblem (23) and its dual (25).

Step 2 (Stopping test) Compute δ_{k+1} by (26). If $\delta_{k+1} \leq tol_1$ and $D(E^k, \alpha^k) \leq tol_2$, **STOP**.

Step 3 (Bundle management) Set

$$J_{k+1}^{ora} \subseteq J_k^{ora} \text{ and } J_{k+1}^{ora} \supseteq \{k+1\},$$
$$J_{k+1}^{agg} \subseteq J_k^{agg} \text{ and } J_{k+1}^{agg} \supseteq \{k+1\}.$$

Step 4 (Descent test) Compute

$$f(y^{k+1}),\ G(y^{k+1}, \alpha^k, E^k),\ \xi_f^{k+1} \in \partial f(y^{k+1}),\ \xi_g^{k,k+1} \in \partial G(y^{k+1}, \alpha^k, E^k).$$

Define the associated $e^f_{k,k+1}$, $e^g_{k,k+1}$, b_k^{k+1}, and Δ_k^{k+1} by (15). If the descent condition (28) holds, then set $i_n = 0$ and go to **Step 6**. Otherwise, set $i_n = i_n + 1$ and go to **Step 5**.

Step 5 (Null step update) Choose $\mu_{k+1} \in [\mu_k, \mu_{\max}]$ and set

$$\left(x^{k+1}, f(x^{k+1})\right) = \left(x^k, f(x^k)\right),$$
$$\left(e^f_{k+1,j}, b^j_{k+1}, \Delta^j_{k+1}\right) = \left(e^f_{k,j}, b^j_k, \Delta^j_k\right), \text{ for all } j \in J^{ora}_{k+1}.$$

If $i_n > i_{n\max}$ and $D(E^k, \alpha^k) > tol_2$, then go to **Step 7**. Otherwise, set

$$\left(E^{k+1}, \alpha^{k+1}, G(x^{k+1}, \alpha^{k+1}, E^{k+1})\right) = \left(E^k, \alpha^k, G(x^k, \alpha^k, E^k)\right),$$
$$\left(\xi_g^{k+1,j}, e^g_{k+1,j}\right) = \left(\xi_g^{k,j}, e^g_{k,j}\right), \text{ for all } j \in J^{ora}_{k+1}.$$

and go to **Step 8**.

Step 6 (Serious step update) Choose $\mu_{k+1} \in [\mu_{\min}, \mu_{\max}]$ and set

$$i_x = k+1, \ x^{k+1} = y^{k+1}, \ f(x^{k+1}) = f(y^{k+1}), \ J^{agg}_{k+1} = \emptyset.$$

For all $j \in J^{ora}_{l+1}$, we set

$$\begin{aligned} e^f_{k+1,j} &= e^f_{k,j} + f(x^{k+1}) - f(x^k) - \langle \xi^j_f, x^{k+1} - x^k \rangle, \\ b^j_{k+1} &= b^j_k + \|x^{k+1} - x^k\|^2/2 - \langle \Delta^j_k, x^{k+1} - x^k \rangle, \\ \Delta^j_{k+1} &= \Delta^j_k - x^{k+1} + x^k. \end{aligned}$$

If $D(E^k, \alpha^k) > tol_2$, go to **Step 7**. Otherwise, set

$$E^{k+1} = E^k, \ \alpha^{k+1} = \alpha^k, \ G(x^{k+1}, \alpha^{k+1}, E^{k+1}) = G(y^{k+1}, \alpha^k, E^k).$$

For all $j \in J^{ora}_{k+1}$, we set $\xi_g^{k+1,j} = \xi_g^{k,j}$ and

$$e^g_{k+1,j} = e^g_{k,j} + G(x^{k+1}, \alpha^{k+1}, E^{k+1}) - G(x^k, \alpha^k, E^k) - \langle \xi_g^{k,j}, x^{k+1} - x^k \rangle.$$

Then, go to **Step 8**.

Step 7 (Subdivision update) Select the index set W_k as follows:

$$W_k \supseteq \Big\{ u \,|\, G(y^j, \alpha^k, E^k) = g(y^j, t_u) + h(t_u, \alpha_u) \text{ and } h(t_u, \alpha_u) > tol_2, \\ \text{for all } j \in J^{ora}_{k+1}, \ t_u \in E^k \Big\},$$

and

$$W_k \subseteq \Big\{ u \,|\, h(t_u, \alpha_u) > tol_2, \ \text{for all } t_u \in E^k \Big\}.$$

Take the **Refinement procedure** on subsets $T_{\bar{u}}$, $u \in W_k$, where $\bar{u}$ is defined by

$$\bar{u} = \begin{cases} u & \text{if } \dfrac{\alpha_u}{8}|T_u|^2 \geq \dfrac{\alpha_{u+1}}{8}|T_{u+1}|^2, \\ u+1 & \text{otherwise,} \end{cases}$$

where $T_0 = 0$ and $T_{N_k+1} = 0$. Put

$$\begin{aligned} E^{k+1} &= \{t_u \,|\, u = 0, ..., N_{k+1}\} := E^k \bigcup \Big\{ E^{\alpha BB}(T_{\bar{u}}) \,|\, u \in W_k \Big\}, \\ \alpha^{k+1} &= \{\alpha_u \,|\, u = 1, ..., N_{k+1}\} := \alpha^k \bigcup \\ &\quad \Big\{ \alpha_{\bar{u},i} : u \in W_k, \ i = 1, ..., 2s+1 \Big\} \setminus \{\alpha_{\bar{u}} : u \in W_k\}. \end{aligned}$$

If $i_z = 1$, set $E^0 = E^{k+1}$, $\alpha^0 = \alpha^{k+1}$, $k = 0$ and go to **Step 0**. Otherwise, for all $j \in J^{ora}_{k+1}$, compute $G(y^j, \alpha^{k+1}, E^{k+1})$, $\xi_g^{k+1,j} \in \partial G(y^j, \alpha^{k+1}, E^{k+1})$ and define the associated $e^g_{k+1,j}$ by (15). Then, go to **Step 8**.

Step 8 (Parameter update) Compute the parameters η_1^{k+1} and η_2^{k+1} by (16). Set $k = k + 1$ and go to **Step 1**.

Remark 2 Under Assumption 1, there exist E, α, and a point x such that

$$G(x, \alpha, E) \leq 0 \text{ as } D(E, \alpha) \to 0.$$

Therefore, Algorithm 1 cannot pass infinitely many times through **step 0**, i.e., Algorithm 1 generates a feasible initial point x^0 such that $G(x^0, \alpha^0, E^0) \leq z \leq 0$ at **step 0**. According to (28), if the descent test is satisfied, then it holds that $G(x^{k+1}, \alpha^k, E^k) \leq G^+(x^k, \alpha^k, E^k) - m\delta_{k+1} \leq G^+(x^k, \alpha^k, E^k)$. By Lemma 4, we have $G(x^{k+1}, \alpha^{k+1}, E^{k+1}) \leq G^+(x^k, \alpha^k, E^k)$. If x^k is infeasible, then we can obtain that $G(x^0, \alpha^0, E^0) > 0$, which contradicts the fact that $G(x^0, \alpha^0, E^0) \leq z \leq 0$. Therefore, if a serious step is declared, then the following conditions hold:

$$f(x^{k+1}) - f(x^k) \leq -m\delta_{k+1} \text{ and } G(x^{k+1}, \alpha^{k+1}, E^{k+1}) \leq 0.$$

Lemma 5 Algorithm 1 *cannot pass infinitely many times through step 7, i.e., there exist* k_E, $\bar{E}$, $\bar{\alpha}$ *such that*

$$E^k = \bar{E},\ \alpha^k = \bar{\alpha},\ D(\bar{E}, \bar{\alpha}) \leq tol_2 \text{ for all } k \geq k_E.$$

4 Convergence

Firstly, we define the index set of serious step as follows:

$$\mathcal{K}_s := \{k \mid y^k \text{ is a serious step, i.e., } x^k = y^k\}.$$

We start with the following lemma.

Lemma 6 *Assume that* f *is bounded below on* X *and* Algorithm 1 *generates an infinite number of serious steps. Then* $C^k \to 0$ *and* $S^k + v^k \to 0$ *as* $k \to \infty$ *in* $\mathcal{K}_s$.

Proof We first show that

$$\sum_{k \in \mathcal{K}_s} \delta_k < +\infty. \tag{32}$$

Making use of (28) and Remark 2, we conclude that, for all $k \in \mathcal{K}_s$,

$$m\delta_k \leq f(x^{k-1}) - f(x^k).$$

Thus, the sequence $\{f(x^k)\}$ is nonincreasing and bounded below on X. Let $\bar{f} = \inf\{f(x) \mid x \in X\}$. Therefore, we obtain that

$$\sum_{k \in \mathcal{K}_s} \delta_k \leq \frac{1}{m}\left(f(x^0) - \bar{f}\right) < +\infty.$$

Now, we obtain that $\{\delta_k\}_{k \in \mathcal{K}_s} \to 0$. Making use of (26), we conclude that

$$C^k \leq \delta_k \text{ and } \frac{1}{\mu_{k-1}}\|S^k + v^k\|^2 \leq \delta_k.$$

It immediately follows that $C^k \to 0$ and $S^k + v^k \to 0$ as $k \to \infty$ in $\mathcal{K}_s$ (a consequence of $\mu_{k-1} \in [\mu_{\min}, \mu_{\max}]$). □

Let $\Psi_k(y^{k+1})$ denote the objective function optimal value of the subproblem (18), i.e., $\Psi_k(y^{k+1}) = \hat{H}_k(y^{k+1}) + \frac{\mu_k}{2}\|y^{k+1} - x^k\|^2$.

Lemma 7 *Assume that* Algorithm 1 *takes a finite number of serious steps and* $\bar{k}$ *is the last serious iteration, i.e., for all* $k \geq \bar{k}$ *we have* $x^k = x^{\bar{k}}$. *Then we can obtain that*

$$H(x^k, x^k, \alpha^k, E^k) \geq \Psi_k(y^{k+1}) \geq \Psi_{k-1}(y^k) + \frac{\mu_{k-1}}{2}\|y^{k+1} - y^k\|^2, \text{ for all } k > \max\{\bar{k}, k_E\}.$$

Proof In what follows, we consider $k > \max\{\bar{k}, k_E\}$. Using Lemma 3 and the convexity of $\hat{H}_k$, we obtain that

$$\begin{aligned} H(x^k, x^k, \alpha^k, E^k) &= \hat{H}_k(x^k) \\ &\geq \hat{H}_k(y^{k+1}) + \langle S^{k+1}, x^k - y^{k+1}\rangle \\ &\geq \hat{H}_k(y^{k+1}) + \langle S^{k+1}, x^k - y^{k+1}\rangle + \langle v^{k+1}, x^k - y^{k+1}\rangle \\ &\geq \hat{H}_k(y^{k+1}) + \frac{\mu_k}{2}\|y^{k+1} - x^k\|^2 \\ &= \Psi_k(y^{k+1}), \end{aligned}$$

where the first equality is by (17), the second inequality is due to (20). Since $k \in J_k^{agg}$, (21) and (22) imply that

$$\begin{aligned} \hat{H}_k(y^{k+1}) \ &\geq \ G^+(x^k, \alpha^k, E^k) - C^k + \langle S^k, y^{k+1} - x^k\rangle \\ &= \ \hat{H}_{k-1}(y^k) + \langle S^k, y^{k+1} - y^k\rangle. \end{aligned}$$

Moreover,

$$\begin{aligned} \Psi_k(y^{k+1}) &= \hat{H}_k(y^{k+1}) + \frac{\mu_k}{2}\|y^{k+1} - x^k\|^2 \\ &\geq \hat{H}_{k-1}(y^k) + \langle S^k, y^{k+1} - y^k\rangle + \frac{\mu_{k-1}}{2}\|y^{k+1} - x^k\|^2 \\ &\geq \hat{H}_{k-1}(y^k) + \langle S^k, y^{k+1} - y^k\rangle + \langle v^k, y^{k+1} - y^k\rangle + \frac{\mu_{k-1}}{2}\|y^{k+1} - x^k\|^2 \\ &\geq \hat{H}_{k-1}(y^k) - \mu_{k-1}\langle y^k - x^k, y^{k+1} - y^k\rangle + \frac{\mu_{k-1}}{2}\|y^k - x^k + y^{k+1} - y^k\|^2 \\ &= \hat{H}_{k-1}(y^k) + \frac{\mu_{k-1}}{2}\|y^k - x^k\|^2 + \frac{\mu_{k-1}}{2}\|y^{k+1} - y^k\|^2 \\ &= \Psi_{k-1}(y^k) + \frac{\mu_{k-1}}{2}\|y^{k+1} - y^k\|^2, \end{aligned}$$

where the second inequality is by the fact that $\langle v^k, y^{k+1} - y^k\rangle \leq 0$. □

Lemma 8 *Assume that* Algorithm 1 *takes a finite number of serious steps. Suppose that* $\{\eta_1^k\}$ *and* $\{\eta_2^k\}$ *are bounded above. Then,* $C^k \to 0$ *and* $S^k + v^k \to 0$ *as* $k \to \infty$.

Proof Let $\bar{k}$ be the last serious iteration. In what follows, we consider $k > \{\bar{k}, k_E\}$. By the definition (22) of $\hat{H}_k$ and the fact $k \in J_k^{ora}$, we obtain that

$$\hat{H}_k(y^{k+1}) \geq G^+(x^k, \alpha^k, E^k) + \left\{-c_{k,k}^f + \langle s_f^{k,k}, y^{k+1} - x^k\rangle, -c_{k,k}^g + \langle s_g^{k,k}, y^{k+1} - x^k\rangle\right\}.$$

It can be rewritten as

$$\begin{aligned}\hat{H}_k(y^{k+1}) \geq\ & \max\left\{f^k(y^k)+\left\langle \xi_f^k + \eta_1^k(y^k-x^k),\, y^{k+1}-y^k\right\rangle - f(x^k),\right.\\ & \left. g^k(y^k)+\left\langle \xi_g^{k,k}+\eta_2^k(y^k-x^k),\, y^{k+1}-y^k\right\rangle\right\}\\ \geq\ & \max\left\{\left\langle \xi_f^k + \eta_1^k(y^k-x^k),\, y^{k+1}-y^k\right\rangle + f(y^k)-f(x^k),\right.\\ & \left.\left\langle \xi_g^{k,k}+\eta_2^k(y^k-x^k),\, y^{k+1}-y^k\right\rangle + G(y^k,\alpha^k,E^k)\right\}.\end{aligned}$$

Since $\{y^k\} \in X$, the functions f and g are Lipschitz continuous on X and $X \times T$ respectively and X is compact, Theorem 1 implies that $\{\xi_f^k\}$ and $\{\xi_g^{k,k}\}$ are bounded on that set. Therefore by the boundedness of $\{\eta_1^k\}$ and $\{\eta_2^k\}$, we get $\{\xi_f^k + \eta_1^k(y^k-x^k)\}$ and $\{\xi_g^{k,k}+\eta_2^k(y^k-x^k)\}$ are bounded. We set

$$L' = \max\left\{\|\xi_f^k + \eta_1^k(y^k-x^k)\|,\, \|\xi_g^{k,k}+\eta_2^k(y^k-x^k)\|\right\}.$$

Then, we can obtain

$$\hat{H}_k(y^{k+1}) \geq \max\left\{f(y^k)-f(x^k),\, G(y^k,\alpha^k,E^k)\right\} - L'\|y^{k+1}-y^k\|.$$

According to (27) and (28), we can obtain that

$$\begin{aligned}\delta_{k+1} &\leq H(x^k,x^k,\alpha^k,E^k) - \hat{H}_k(y^{k+1})\\ &\leq G^+(x^k,\alpha^k,E^k) - \max\left\{f(y^k)-f(x^k),\, G(y^k,\alpha^k,E^k)\right\} + L'\|y^{k+1}-y^k\|\\ &\leq m\delta_k + L'\|y^{k+1}-y^k\|,\end{aligned}$$

where the third inequality holds by $k > \bar{k}$. Combining this relation with (26) yields

$$0 \leq \delta_{k+1} \leq\ m\delta_k + L'\|y^{k+1}-y^k\|. \tag{33}$$

According to Lemma 7, we infer that

$$\{y^{k+1}-y^k\} \to 0, \quad \text{as } \ k \to \infty,$$

which implies $\lim_{k\to\infty} \delta_k = 0$ [33, Lemma 3, p. 45]. Making use of (26), we conclude that

$$C^k \leq \delta_k \ \text{ and } \ \frac{1}{\mu_{k-1}}\|S^k + v^k\|^2 \leq \delta_k.$$

It immediately follows that $C^k \to 0$ and $S^k + v^k \to 0$ as $k \to \infty$. □

Let

$$\eta_{\min}^{1,j} = \max\{0, -\frac{e_{k,j}^f}{b_k^j}\} + \gamma,\ \eta_{\min}^{2,j} = \max\{0, -\frac{e_{k,j}^g}{b_k^j}\} + \gamma,\ \forall\ j \in \{0, ..., 1\},$$

$$\eta_{\max}^i = \max_{j\in\{0,...,k\}} \eta_i^j,\ i \in \{1,2\},\ J_k^C = \{j \in J_k^{ora} \mid \lambda_{1,k+1}^j = \lambda_{2,k+1}^j = 0\}.$$

Lemma 9 *Let x^k $(k \geq k_E)$ be the current stability center. Suppose that $\{\eta_1^k\}$ and $\{\eta_2^k\}$ are bounded above, then there exist $\hat{\lambda}_{1,k+1}, \hat{\lambda}_{2,k+1} \in \mathbb{R}^{k+1}$ and $\hat{J}_k$ with*

$$\begin{cases} \hat{J}_k \subseteq \{0,1,...,k\},\\ \hat{\lambda}_{1,k+1}^j, \hat{\lambda}_{2,k+1}^j \geq 0,\ \hat{\lambda}_{1,k+1}^j + \hat{\lambda}_{2,k+1}^j > 0,\ \textit{for all } j \in \hat{J}_k,\\ \sum_{j\in\hat{J}_k}(\hat{\lambda}_{1,k+1}^j + \hat{\lambda}_{2,k+1}^j) = 1,\end{cases} \tag{34}$$

such that

$$S^{k+1} = \sum_{j \in \hat{J}_k} \left(\hat{\lambda}_{1,k+1}^j \left(\xi_f^j + \hat{\eta}_1^j \Delta_k^j \right) + \hat{\lambda}_{2,k+1}^j \left(\xi_g^{k,j} + \hat{\eta}_2^j \Delta_k^j \right) \right), \tag{35}$$

and

$$\begin{aligned} C^{k+1} =& \sum_{j \in \hat{J}_k} \left(\hat{\lambda}_{1,k+1}^j \left(e_{k,j}^f + \hat{\eta}_1^j b_k^j + G^+(x^k, \bar{\alpha}, \bar{E}) \right) \right. \\ &\left. + \hat{\lambda}_{2,k+1}^j \left(e_{k,j}^g + \hat{\eta}_2^j b_k^j + G^+(x^k, \bar{\alpha}, \bar{E}) - G(x^k, \bar{\alpha}, \bar{E}) \right) \right), \end{aligned} \tag{36}$$

where $\hat{\eta}_i^j \in [\eta_{\min}^{i,j}, \eta_{\max}^i],\ i \in \{1, 2\}$.

Proof In what follows, we consider $k \geq k_E$. If $k \in \mathcal{K}_s$ or $k = k_E$, then we have $J_k^{agg} = \emptyset$ and

$$\begin{aligned} S^{k+1} &= \sum_{j \in J_k^{ora}} \left(\lambda_{1,k+1}^j \left(\xi_f^j + \eta_1^k \Delta_k^j \right) + \lambda_{2,k+1}^j \left(\xi_g^{k,j} + \eta_2^k \Delta_k^j \right) \right), \\ C^{k+1} &= \sum_{j \in J_k^{ora}} \left(\lambda_{1,k+1}^j \left(e_{k,j}^f + \eta_1^k b_k^j + G^+(x^k, \bar{\alpha}, \bar{E}) \right) \right. \\ &\quad \left. + \lambda_{2,k+1}^j \left(e_{k,j}^g + \eta_2^k b_k^j + G^+(x^k, \bar{\alpha}, \bar{E}) - G(x^k, \bar{\alpha}, \bar{E}) \right) \right), \end{aligned}$$

where $\lambda_{1,k+1}^j, \lambda_{2,k+1}^j \geq 0$ for all $j \in J_k^{ora}$ and $\sum_{j \in J_k^{ora}} (\lambda_{1,k+1}^j + \lambda_{2,k+1}^j) = 1$. Then, the result holds by setting $\hat{J}_k = J_k^{ora} \setminus J_k^C$, $\hat{\lambda}_{1,k+1}^j = \lambda_{1,k+1}^j$, $\hat{\lambda}_{2,k+1}^j = \lambda_{2,k+1}^j$, $\hat{\eta}_1^j = \eta_1^k$ and $\hat{\eta}_2^j = \eta_2^k$.

If $k \notin \mathcal{K}_s$ and $k > k_E$, then there exists $\tilde{k} \in \{k_E + 1, ..., k - 1\}$ such that $x^k = x^{\tilde{k}}$ and $J_{\tilde{k}}^{agg} = \emptyset$. It holds that $\Delta_{\tilde{k}}^j = \Delta_k^j$, $b_{\tilde{k}}^j = b_k^j$, $\xi_g^{\tilde{k},j} = \xi_g^{k,j}$ for all $j \in J_k^{ora}$ and $G(x^{\tilde{k}}, \bar{\alpha}, \bar{E}) = G(x^k, \bar{\alpha}, \bar{E})$. Hence,

$$\begin{aligned} S^{\tilde{k}+1} &= \sum_{j \in J_{\tilde{k}}^{ora}} \left(\lambda_{1,\tilde{k}+1}^j \left(\xi_f^j + \eta_1^{\tilde{k}} \Delta_k^j \right) + \lambda_{2,\tilde{k}+1}^j \left(\xi_g^{k,j} + \eta_2^{\tilde{k}} \Delta_k^j \right) \right), \\ C^{\tilde{k}+1} &= \sum_{j \in J_{\tilde{k}}^{ora}} \left(\lambda_{1,\tilde{k}+1}^j \left(e_{k,j}^f + \eta_1^{\tilde{k}} b_k^j + G^+(x^k, \bar{\alpha}, \bar{E}) \right) + \right. \\ &\quad \left. \lambda_{2,\tilde{k}+1}^j \left(e_{k,j}^g + \eta_2^{\tilde{k}} b_k^j + G^+(x^k, \bar{\alpha}, \bar{E}) - G(x^k, \bar{\alpha}, \bar{E}) \right) \right), \end{aligned}$$

where $\sum_{j \in J_{\tilde{k}}^{ora}} \left(\lambda_{1,\tilde{k}+1}^j + \lambda_{2,\tilde{k}+1}^j \right) = 1$. Since $J_{\tilde{k}}^{agg} = \emptyset$, then $J_{\tilde{k}+1}^{agg} = \tilde{k} + 1$. Hence,

$$\begin{aligned} S^{\tilde{k}+2} &= \sum_{j \in J_{\tilde{k}+1}^{ora}} \left(\lambda_{1,\tilde{k}+2}^j s_f^{\tilde{k}+1,j} + \lambda_{2,\tilde{k}+2}^j s_g^{\tilde{k}+1,j} \right) + \sum_{j \in J_{\tilde{k}+1}^{agg}} \lambda_{3,\tilde{k}+2}^j \cdot S^j \\ &= \sum_{j \in J_{\tilde{k}+1}^{ora}} \left(\lambda_{1,\tilde{k}+2}^j \left(\xi_f^j + \eta_1^{\tilde{k}+1} \Delta_k^j \right) + \lambda_{2,\tilde{k}+2}^j \left(\xi_g^{k,j} + \eta_2^{\tilde{k}+1} \Delta_k^j \right) \right) \\ &\quad + \sum_{j \in J_{\tilde{k}}^{ora}} \left(\lambda_{1,\tilde{k}+1}^j \lambda_{3,\tilde{k}+2}^{k+1} \left(\xi_f^j + \eta_1^{\tilde{k}} \Delta_k^j \right) + \lambda_{2,\tilde{k}+1}^j \lambda_{3,\tilde{k}+2}^{k+1} \left(\xi_g^{\tilde{k},j} + \eta_2^{\tilde{k}} \Delta_k^j \right) \right) \end{aligned}$$

and

$$
\begin{aligned}
C^{\tilde{k}+2} &= \sum_{j\in J_{\tilde{k}+1}^{ora}} \left(\lambda_{1,\tilde{k}+2}^{j} c_{\tilde{k}+1,j}^{f} + \lambda_{2,\tilde{k}+2}^{j} c_{\tilde{k}+1,j}^{g}\right) + \sum_{j\in J_{\tilde{k}+1}^{agg}} \lambda_{3,\tilde{k}+2}^{j} \cdot C^{j} \\
&= \sum_{j\in J_{\tilde{k}+1}^{ora}} \left(\lambda_{1,\tilde{k}+2}^{j} \left(e_{k,j}^{f} + \eta_1^{\tilde{k}+1} b_k^j + G^{+}(x^k,\bar{\alpha},\bar{E})\right)\right. \\
&\quad \left. + \lambda_{2,\tilde{k}+2}^{j} \left(e_{k,j}^{g} + \eta_2^{\tilde{k}+1} b_k^j + G^{+}(x^k,\bar{\alpha},\bar{E}) - G(x^k,\bar{\alpha},\bar{E})\right)\right) \\
&\quad + \sum_{j\in J_{\tilde{k}}^{ora}} \left(\lambda_{1,\tilde{k}+1}^{j} \lambda_{3,\tilde{k}+2}^{\tilde{k}+1} \left(e_{k,j}^{f} + \eta_1^{\tilde{k}} b_k^j + G^{+}(x^k,\bar{\alpha},\bar{E})\right)\right. \\
&\quad \left. + \lambda_{2,\tilde{k}+1}^{j} \lambda_{3,\tilde{k}+2}^{\tilde{k}+1} \left(e_{k,j}^{g} + \eta_2^{\tilde{k}} b_k^j + G^{+}(x^k,\bar{\alpha},\bar{E}) - G(x^k,\bar{\alpha},\bar{E})\right)\right).
\end{aligned}
$$

It is easily seen that

$$
\sum_{j\in J_{\tilde{k}+1}^{ora}} \left(\lambda_{1,\tilde{k}+2}^{j} + \lambda_{2,\tilde{k}+2}^{j}\right) + \sum_{j\in J_{\tilde{k}}^{ora}} \left(\lambda_{1,\tilde{k}+1}^{j} \lambda_{3,\tilde{k}+2}^{\tilde{k}+1} + \lambda_{2,\tilde{k}+1}^{j} \lambda_{3,\tilde{k}+2}^{\tilde{k}+1}\right) = 1.
$$

We set

$$
\hat{J}_{\tilde{k}+1} = \begin{cases} J_{\tilde{k}+1}^{ora} \setminus J_{\tilde{k}+1}^{C}, & \text{if } \lambda_{3,\tilde{k}+2}^{\tilde{k}+1} = 0, \\ (J_{\tilde{k}+1}^{ora} \setminus J_{\tilde{k}+1}^{C}) \cup (J_{\tilde{k}}^{ora} \setminus J_{\tilde{k}}^{C}), & \text{if } \lambda_{3,\tilde{k}+2}^{\tilde{k}+1} \neq 0. \end{cases}
$$

For any $j \in \hat{J}_{\tilde{k}+1}$ and $i \in \{1, 2\}$, we set

$$
\begin{cases} \hat{\lambda}_{i,\tilde{k}+2}^{j} = \lambda_{i,\tilde{k}+1}^{j} \lambda_{3,\tilde{k}+2}^{\tilde{k}+1} & \text{and} \quad \hat{\eta}_i^j = \eta_i^{\tilde{k}}, \quad \text{if } j \in J_{\tilde{k}}^{ora} \setminus J_{\tilde{k}+1}^{ora}, \\ \hat{\lambda}_{i,\tilde{k}+2}^{j} = \lambda_{i,\tilde{k}+2}^{j} & \text{and} \quad \hat{\eta}_i^j = \eta_i^{\tilde{k}+1}, \quad \text{if } j \in J_{\tilde{k}+1}^{ora} \setminus J_{\tilde{k}}^{ora}, \\ \hat{\lambda}_{i,\tilde{k}+2}^{j} = \lambda_{i,\tilde{k}+2}^{j} + \lambda_{i,\tilde{k}+1}^{j} \lambda_{3,\tilde{k}+2}^{\tilde{k}+1} & \text{and} \quad \hat{\eta}_i^j = \ddot{\eta}_i^j, \quad \text{if } j \in J_{\tilde{k}}^{ora} \cap J_{\tilde{k}+1}^{ora}, \end{cases}
$$

where

$$
\ddot{\eta}_i^j = \begin{cases} \dfrac{\eta_i^{\tilde{k}} + \eta_i^{\tilde{k}+1}}{2}, & \text{if } \hat{\lambda}_{i,\tilde{k}+2}^{j} = 0, \\[2ex] \dfrac{\lambda_{i,\tilde{k}+2}^{j} \eta_i^{\tilde{k}+1} + \lambda_{i,\tilde{k}+1}^{j} \lambda_{3,\tilde{k}+2}^{\tilde{k}+1} \eta_i^{\tilde{k}}}{\hat{\lambda}_{i,\tilde{k}+2}^{j}}, & \text{if } \hat{\lambda}_{i,\tilde{k}+2}^{j} > 0. \end{cases}
$$

Combining these, we have

$$
\eta_{\min}^{i,j} \leq \hat{\eta}_i^j \leq \eta_{\max}^{i}, \ \forall\, i = \{1, 2\},\ j \in \hat{J}_{\tilde{k}+1},
$$

and

$$
\begin{aligned}
S^{\tilde{k}+2} &= \sum_{j\in \hat{J}_{\tilde{k}+1}} \left(\hat{\lambda}_{1,\tilde{k}+2}^{j} \left(\xi_f^j + \hat{\eta}_1^j \Delta_k^j\right) + \hat{\lambda}_{2,\tilde{k}+2}^{j} \left(\xi_g^{\tilde{k}+1,j} + \hat{\eta}_2^j \Delta_k^j\right)\right), \\
C^{\tilde{k}+2} &= \sum_{j\in \hat{J}_{\tilde{k}+1}} \left(\hat{\lambda}_{1,\tilde{k}+2}^{j} \left(e_{k,j}^{f} + \hat{\eta}_1^j b_k^j + G^{+}(x^k,\bar{\alpha},\bar{E})\right)\right. \\
&\quad \left. + \hat{\lambda}_{2,\tilde{k}+2}^{j} \left(e_{k,j}^{g} + \hat{\eta}_2^j b_k^j + G^{+}(x^k,\bar{\alpha},\bar{E}) - G(x^k,\bar{\alpha},\bar{E})\right)\right).
\end{aligned}
$$

By the same argument, we can obtain that there exist $\hat{\lambda}_{1,k+1}, \hat{\lambda}_{2,k+1} \in \mathbb{R}^{k+1}$ and $\hat{J}_k$ satisfying (34) such that (35) and (36) hold. □

From Theorems 2 and 3, we can obtain that $G(\cdot, \alpha^k, E^k)$ is regular on X. The authors [17] state that the following lemma holds. We give the proof since the cutting plane model and the bundle update are different from [17]. In addition, the proof of [17] does not consider the aggregate information of the cutting plane model.

Lemma 10 *Assume that f is regular on X. Suppose that $\{\eta_1^k\}$ and $\{\eta_2^k\}$ are bounded above. Let x^k ($k \geq k_E$) be the current stability center and $\bar{\eta} = \max_{j\in\{0,1,\ldots,k\}}\{\eta_1^j, \eta_2^j\}$. The upper envelope model is defined as*

$$\Phi(y, x^k, \alpha^k, E^k) = G^+(x^k, \alpha^k, E^k) + \sup_{i=1,2} \left\{ m^i_{y^+,\xi_i,\eta} \mid y^+ \in B(x^k; R), \xi_1 \in \partial f(y^+), \xi_2 \in \partial G(y^+, \alpha^k, E^k), \eta \in [\gamma, \bar{\eta}] \right\},$$

where

$$m^i_{y^+,\xi_i,\eta} := -\frac{\gamma}{2}\|y^+ - x^k\|^2 + \langle \xi_i + \eta(y^+ - x^k), y - x^k \rangle, \ i = 1, 2. \tag{37}$$

$B(x^k; R)$ is a fixed ball large enough to contain all possible trial steps during the iteration k. Then, the following statements hold.

1. *$\Phi(\cdot, x^k, \alpha^k, E^k)$ is a convex function and $\hat{H}_k(\cdot) \leq \Phi(\cdot, x^k, \alpha^k, E^k)$.*
2. *$\Phi(x^k, x^k, \alpha^k, E^k) = G^+(x^k, \alpha^k, E^k)$.*
3. *$\partial\Phi(x^k, x^k, \alpha^k, E^k) \subseteq \partial H(x^k, x^k, \alpha^k, E^k)$.*

Proof We only show that $\hat{H}_k(\cdot) \leq \Phi(\cdot, x^k, \alpha^k, E^k)$. For the remainder of the proofs, we refer the reader to [17].

Using (14) together with (15), for $j \in J_k^{ora}$ we obtain that

$$\begin{aligned} -c^f_{k,j} + \langle s_f^{k,j}, y - x^k \rangle &\leq -\gamma b_k^j + \langle \xi_f^j + \eta_1^k(y^j - x^k), y - x^k \rangle = m^1_{y^j,\xi_f^j,\eta_1^k}, \\ -c^g_{k,j} + \langle s_g^{k,j}, y - x^k \rangle &\leq -\gamma b_k^j + \langle \xi_g^{k,j} + \eta_2^k(y^j - x^k), y - x^k \rangle = m^2_{y^j,\xi_g^{k,j},\eta_2^k}. \end{aligned}$$

On the other hand, from Lemma 9, we can obtain that for $j \in J_k^{agg}$ there exist $\hat{\lambda}_{1,j}, \hat{\lambda}_{2,j} \in \mathbb{R}^j$, and $\hat{J}_{j-1}$ satisfying (34) such that

$$\begin{aligned} S^j &= \sum_{l\in\hat{J}_{j-1}} \left(\hat{\lambda}^l_{1,j} \left(\xi_f^l + \hat{\eta}_1^l \Delta_k^l \right) + \hat{\lambda}^l_{2,j} \left(\xi_g^{k,l} + \hat{\eta}_2^l \Delta_k^l \right) \right), \\ C^j &= \sum_{l\in\hat{J}_{j-1}} \left(\hat{\lambda}^l_{1,j} \left(e^f_{k,l} + \hat{\eta}_1^l b_k^l + G^+(x^k, \bar{\alpha}, \bar{E}) \right) \right. \\ &\quad \left. + \hat{\lambda}^l_{2,j} \left(e^g_{k,l} + \hat{\eta}_2^l b_k^l + G^+(x^k, \bar{\alpha}, \bar{E}) - G(x^k, \bar{\alpha}, \bar{E}) \right) \right), \end{aligned}$$

where $\hat{\eta}_i^l \geq \eta_{\min}^{i,l}$, $i = 1, 2$. According to (16) and the definition of $\eta_{\min}^{i,l}$, we can obtain that

$$e^f_{k,l} + \hat{\eta}_1^l b_k^l \geq \gamma b_k^l, \ \ e^g_{k,l} + \hat{\eta}_2^l b_k^l \geq \gamma b_k^l.$$

Hence,

$$\begin{aligned} & -C^j + \langle S^j, y - x^k \rangle \\ \leq & \sum_{l\in\hat{J}_{j-1}} \left(\hat{\lambda}^l_{1,j} \left(-\gamma b_k^l + \langle \xi_f^l + \hat{\eta}_1^l \Delta_k^l, y - x^k \rangle \right) \right. \\ & \left. + \hat{\lambda}^l_{2,j} \left(-\gamma b_k^l + \langle \xi_g^{k,l} + \hat{\eta}_2^l \Delta_k^l, y - x^k \rangle \right) \right) \\ \leq & \max\{-\gamma b_k^l + \langle \xi_f^l + \hat{\eta}_1^l \Delta_k^l, y - x^k \rangle, -\gamma b_k^l + \langle \xi_g^{k,l} + \hat{\eta}_2^l \Delta_k^l, y - x^k \rangle\} \\ \leq & \sup_{\eta\in[\gamma,\bar{\eta}]} \{m^1_{y^j,\xi_f^j,\eta}, m^2_{y^j,\xi_g^{k,j},\eta}\}. \end{aligned}$$

By the definition (22) of $\hat{H}_k(\cdot)$, we can obtain that $\hat{H}_k(\cdot) \leq \Phi(\cdot, x^k, \alpha^k, E^k)$. □

Theorem 6 *Assume that f is bounded below and regular on X. Suppose that $\{\eta_1^k\}$ and $\{\eta_2^k\}$ are bounded above. Consider the stopping parameter $tol_1 = 0$ and suppose there is no termination. Then, the following mutually exclusive situations hold:*

(i) Algorithm 1 *generates an infinite sequence of serious steps. Let $\bar{x}$ be the accumulation point of $\{x^k\}$, then $0 \in \partial H(\bar{x}, \bar{x}, \bar{\alpha}, \bar{E}) + \partial I_X(\bar{x})$.*

(ii) Algorithm 1 *generates a finite sequence of serious steps. Let $\bar{x}$ be the last stability center, then $0 \in \partial H(\bar{x}, \bar{x}, \bar{\alpha}, \bar{E}) + \partial I_X(\bar{x})$.*

In addition, if MFCQ holds at $\bar{x}$ for problem (12) with $E = \bar{E}$, then $\bar{x}$ is a tol_2-KKT point of the problem (1).

Proof In what follows, we consider $k \geq k_E$. From Theorem 1 and Lemma 9, we can obtain that $\{S^k\}$ is bounded; thus, we can assume that $S^{k+1} \to \bar{S}$. Hence, $v^{k+1} \to -\bar{S}$ as $k \to \infty$. Using (21) together with Lemma 10, we can obtain that

$$\begin{aligned}\Phi(y, x^k, \bar{\alpha}, \bar{E}) &\geq \hat{H}_k(y) \\ &\geq \hat{H}_k(y^{k+1}) + \langle S^{k+1}, y - y^{k+1}\rangle \\ &= G^+(x^k, \bar{\alpha}, \bar{E}) - C^{k+1} + \langle S^{k+1}, y - x^k\rangle.\end{aligned}$$

By (19) and the convexity of I_X, $I_X(y) \geq I_X(y^{k+1}) + \langle v^{k+1}, y - y^{k+1}\rangle$. As a result, it holds that

$$\Phi(y, x^k, \bar{\alpha}, \bar{E}) + I_X(y) \geq G^+(x^k, \bar{\alpha}, \bar{E}) - C^{k+1} + \langle S^{k+1} + v^{k+1}, y - x^k\rangle + \langle v^{k+1}, x^k - y^{k+1}\rangle. \tag{38}$$

(i) First, note that $\{x^k\}$ is in the compact set X, so it has an accumulation point, say for some infinite set $K \subseteq \mathcal{K}_s$, $x^k \to \bar{x}$ as $K \ni k \to \infty$. For all $k+1 \in K$, we have $x^k - y^{k+1} = \frac{1}{\mu_k}(S^{k+1} + v^{k+1})$. By Lemma 6, $C^{k+1} \to 0$, $S^{k+1} + v^{k+1} \to 0$, $x^k - y^{k+1} \to 0$, $x^k \to \bar{x}$ as $K \ni k+1 \to \infty$. Passing onto the limit in (38) as $K \ni k+1 \to \infty$, we obtain that

$$\Phi(y, \bar{x}, \bar{\alpha}, \bar{E}) + I_X(y) \geq G^+(\bar{x}, \bar{\alpha}, \bar{E}) + \langle 0, y - \bar{x}\rangle = \Phi(\bar{x}, \bar{x}, \bar{\alpha}, \bar{E}) + \langle 0, y - \bar{x}\rangle.$$

Using the convexity of $\Phi(\cdot, \bar{x}, \bar{\alpha}, \bar{E}) + I_X(\cdot)$, we have $0 \in \partial\Phi(\bar{x}, \bar{x}, \bar{\alpha}, \bar{E}) + \partial I_X(\bar{x})$. According to Lemma 10, we can conclude that $0 \in \partial H(\bar{x}, \bar{x}, \bar{\alpha}, \bar{E}) + \partial I_X(\bar{x})$.

(ii) Let $\bar{k}$ be the last serious iteration, i.e., $x^k = \bar{x}$ for all $k \geq \bar{k}$. For $k \geq \bar{k}$, we have that $\bar{x} - y^{k+1} = x^k - y^{k+1} = \frac{1}{\mu_k}(S^{k+1} + v^{k+1})$. By Lemma 8, $C^{k+1} \to 0$, $S^{k+1} + v^{k+1} \to 0$, $y^{k+1} - x^k \to 0$ as $k \to \infty$. Passing onto the limit in (38) as $k \to \infty$, we obtain that

$$\Phi(y, \bar{x}, \bar{\alpha}, \bar{E}) + I_X(y) \geq G^+(\bar{x}, \bar{\alpha}, \bar{E}) + \langle 0, y - \bar{x}\rangle = \Phi(\bar{x}, \bar{x}, \bar{\alpha}, \bar{E}) + \langle 0, y - \bar{x}\rangle.$$

By the same discussion, we deduce that $0 \in \partial H(\bar{x}, \bar{x}, \bar{\alpha}, \bar{E}) + \partial I_X(\bar{x})$.

From Theorem 5 and Remark 2, we obtain that $\bar{x}$ is a tol_2-KKT point of the original problem (1). The proof is complete. □

Theorem 7 *Assume that f is bounded below and regular on X. Suppose that $\{\eta_1^k\}$ and $\{\eta_2^k\}$ are bounded above. Consider the stopping parameter $tol_1 = 0$ and suppose that algorithm loops forever. If f is semismooth, then for any accumulation point $\bar{x}$ of the sequence $\{x^k\}$, for each $\epsilon > 0$ there exists $\rho > 0$ such that*

$$\max\left\{f(y) - f(\bar{x}), G(y, \bar{\alpha}, \bar{E})\right\} \geq -\epsilon\|y - \bar{x}\|, \textit{ for all } y \in X \cap B(\bar{x}; \rho).$$

If, in addition, the set $X_{\rho,\epsilon,tol_2} = \{y \in X \cap B(\bar{x};\rho) : G(y) < -(\rho\epsilon + tol_2)\}$ *is not empty, then*

$$f(y) \geq f(\bar{x}) - \epsilon\|y - \bar{x}\| \text{ for all } y \in X_{\rho,\epsilon,tol_2}.$$

Proof Let $\hat{\lambda}^j = \hat{\lambda}^j_{1,k+1} + \hat{\lambda}^j_{2,k+1}$, then we can obtain that

$$C^{k+1} \geq \gamma \sum_{j\in\hat{J}_k} \hat{\lambda}^j b^j_k \geq \gamma \sum_{j\in\hat{J}_k} (\hat{\lambda}^j)^2 b^j_k = \frac{\gamma}{2} \sum_{j\in\hat{J}_k} \left(\hat{\lambda}^j \|y^j - x^k\|\right)^2. \tag{39}$$

Equation (39) together with Lemma 6 and 8 implies that $\|y^j - x^k\| \to 0$ for all $j \in \hat{J}_k$. Moreover, $\|y^j - \bar{x}\| \leq \|y^j - x^k\| + \|x^k - \bar{x}\| \to 0$ for all $j \in \hat{J}_k$. From Proposition 1 and Theorem 3, we can obtain that to fix any $\epsilon > 0$ there exist $\rho > 0$, $\hat{\lambda}_{1,k+1}, \hat{\lambda}_{2,k+1} \in \mathbb{R}^{k+1}$ and $\hat{J}_k$ satisfying (34) such that $y^j \in B(\bar{x};\rho)$, for all $j \in \hat{J}_k$ and

$$\begin{aligned} f(y) &\geq f(y^j) + \langle \xi^j_f, y - y^j\rangle - \epsilon\|y - y^j\|, \ y \in X \cap B(\bar{x};\rho), \\ G(y,\bar{\alpha},\bar{E}) &\geq G(y^j,\bar{\alpha},\bar{E}) + \langle \xi^{k,j}_g, y - y^j\rangle - \epsilon\|y - y^j\|, \ y \in X \cap B(\bar{x};\rho). \end{aligned}$$

Using (35) together with (36), for $j \in \hat{J}_k$ we see that

$$\begin{aligned} &\max\left\{f(y) - f(x^k), G(y,\bar{\alpha},\bar{E})\right\} \\ \geq &\sum_{j\in\hat{J}_k} \left(\hat{\lambda}^j_{1,k+1}\left(f(y) - f(x^k)\right) + \hat{\lambda}^j_{2,k+1} G(y,\bar{\alpha},\bar{E})\right) \\ \geq &\sum_{j\in\hat{J}_k} \left(\hat{\lambda}^j_{1,k+1}\left(f(y^j) + \langle \xi^j_f, y - y^j\rangle - \epsilon\|y - y^j\| - f(x^k)\right)\right. \\ &\left. + \hat{\lambda}^j_{2,k+1}\left(G(y^j,\bar{\alpha},\bar{E}) + \langle \xi^{k,j}_g, y - y^j\rangle - \epsilon\|y - y^j\|\right)\right) \\ = &\sum_{j\in\hat{J}_k} \left(-\epsilon\hat{\lambda}^j\|y - y^j\| + \hat{\lambda}^j_{1,k+1}\left(-e^f_{k,j} + \langle \xi^j_f, y - x^k\rangle\right)\right. \\ &\left. + \hat{\lambda}^j_{2,k+1}\left(-e^g_{k,j} + \langle \xi^{k,j}_g, y - x^k\rangle + G(x^k,\bar{\alpha},\bar{E})\right)\right) \\ \geq &-C^{k+1} + G^+(x^k,\bar{\alpha},\bar{E}) + \langle S^{k+1}, y - x^k\rangle - \left\langle \sum_{j\in\hat{J}_k} \hat{\eta}^j (y^j - x^k), y - x^k \right\rangle \\ &-\epsilon \sum_{j\in\hat{J}_k} \hat{\lambda}^j \|y - y^j\| \\ \geq &-C^{k+1} + \langle S^{k+1}, y - x^k\rangle - \left\langle \sum_{j\in\hat{J}_k} \hat{\eta}^j (y^j - x^k), y - x^k \right\rangle \\ &-\epsilon \sum_{j\in\hat{J}_k} \hat{\lambda}^j \left(\|y - x^k\| + \|y^j - x^k\|\right). \end{aligned}$$

From Theorem 1 and Lemma 9, we can obtain that $\{S^k\}$ is bounded; thus, we can assume that $S^{k+1} \to \bar{S}$. By noting Lemma 6 and Lemma 8, and passing to the limit as $k \to \infty$, we obtain that

$$\max\left\{f(y) - f(\bar{x}), G(y,\bar{\alpha},\bar{E})\right\} \geq \langle \bar{S}, y - \bar{x}\rangle - \epsilon\|y - \bar{x}\|.$$

As already seen, $S^{k+1} + v^{k+1} \to 0$ implies that $-\bar{S} \in \partial I_X(\bar{x})$, thus $\langle -\bar{S}, y - \bar{x}\rangle \leq 0$. Hence

$$\max\left\{f(y) - f(\bar{x}), G(y,\bar{\alpha},\bar{E})\right\} \geq -\epsilon\|y - \bar{x}\|.$$

By (11), we have $G(y, \bar{\alpha}, \bar{E}) \leq G(y) + D(\bar{E}, \bar{\alpha}) \leq G(y) + tol_2$. Since also $y \in B(\bar{x}; \rho)$, then $f(y) \geq f(\bar{x}) - \epsilon\|y - \bar{x}\|$ for all $y \in X_{\rho,\epsilon,tol_2} = \{y \in X \cap B(\bar{x}; \rho) : G(y) < -(\rho\epsilon + tol_2)\}$. The proof is complete. □

5 Computational results

From Section 2.3, we find that the SIP can be rewritten as a finite constrained optimization problem when the lower level problem is a concave optimization problem. Hence, we consider the nonconvex nonsmooth constrained optimization problems and the nonconvex nonsmooth SIP problems.

In this section, we report some numerical results to verify the practical efficiency of our approach. We consider two classes of test problems and those test problems are introduced in Appendix 1.

Solvers. For comparison purposes, we use the following solvers:

- PBM-inexact, the Proximal Bundle Method from [23].
- ICPBM, the Infeasible Constrained Proximal Bundle Method from [17].
- SolvOpt, the pubic software for local nonlinear optimization problems. The code is available at https://imsc.uni-graz.at/kuntsevich/solvopt/index.html.
- ASA, the Adaptive Subdivision Algorithm from [41].
- *fseminf*, Matlab Optimization Toolbox for finding minimum of semi-infinitely constrained multivariable nonlinear function.

Solvers ICPBM and SolvOpt are designed for nonsmooth nonconvex constrained problems; these methods are only applicable to Problems 1–8. Solvers ASA and *fseminf* are designed for semi-infinite programming problems; these methods are only applicable to Problems 9–23. All solvers were implemented in MATLAB R2018a (Windows 10 64-bit). The experiments were performed on a PC with Intel Core i5, 2.4 GHz. We solve the quadratic subproblems using *quadprog*, which is available in the Matlab Optimization Toolbox. For solving the NLP subproblems, ASA and Algorithm 1 use *fmincon* of Matlab Optimization Toolbox.

Parameters. Parameters of Algorithm 1 are set as follows: $tol_1 = 10^{-6}$, $tol_2 = 10^{-6}$, $s = 2$, $\gamma = 2$, $i_{n\max} = 100$. We choose the descent parameter as follows:

$$m = \begin{cases} 0.01, & \text{if } n \leq 20 \\ 0.25, & \text{if } 20 < n \leq 200 \\ 0.55, & \text{if } n > 200. \end{cases}$$

The proximal parameter is only updated at serious steps. We set $\mu_0 = 5$ and

$$\mu_{k+1} := \max\left(\mu_{\min}, \min\left(\bar{\mu}_{k+1}, \mu_{\max}\right)\right),$$

where

$$\bar{\mu}_{k+1} := \frac{\|S^{k+1} - S^k\|^2}{\langle S^{k+1} - S^k, y^{k+1} - y^k\rangle},$$
$$\mu_{\min} = 1, \ \mu_{\max} = 10^5.$$

In ICPBM, the stopping tolerance tol has been set to 10^{-6}. For all the other parameters, we have used the default values suggested by the respective codes [17]. Furthermore, in the parameters selection of PBM-inexact and ASA we have used the default values [23, 41]. In

Table 1 Summary of numerical results with PBM-inexact, ICPBM, and SolvOpt

P	Solvers	f^*	g^*	NF	Time
1	Algorithm 1	0.0858	−1.3148E−06	60	0.0938
	PBM-inexact	0.0858	−1.5580E−06	44	0.0794
	ICPBM	0.0858	−1.2481E−06	48	0.0969
	SolvOpt	0.0858	1.0000E−08	612	0.0338
2	Algorithm 1	1.9522	−0.3542	40	0.0578
	PBM-inexact	1.9522	−0.3571	48	0.0731
	ICPBM	1.9522	−0.3574	44	0.0781
	SolvOpt	1.9522	−0.3537	140	0.0141
3	Algorithm 1	0.6164	−2.6861	46	0.0641
	PBM-inexact	0.6165	−2.6900	286	2.5147
	ICPBM	0.6164	−2.6847	210	0.8938
	SolvOpt	0.6164	−2.6860	266	0.0219
4	Algorithm 1	2.0000	−1.0670E−06	50	0.0719
	PBM-inexact	2.0000	−1.1611E−06	56	0.0913
	ICPBM	2.0000	−1.5383E−06	52	0.1078
	SolvOpt	2.0000	−3.4893E−07	282	0.0200
5	Algorithm 1	2.2500	−1.0539E−06	164	0.2203
	PBM-inexact	2.2500	−1.1364E−06	158	0.4903
	ICPBM	2.2500	1.1914E−06	84	0.2219
	SolvOpt	2.2500	−3.1385E−09	306	0.0178
6	Algorithm 1	1.1095E−08	−1.7304E−06	52	0.0922
	PBM-inexact	7.2299E−07	−1.1266E−04	58	0.0991
	ICPBM	1.5165E−07	−1.4008E−06	60	0.1219
	SolvOpt	1.2568E−13	2.5665E−11	808	0.0516
7 (n=20)	Algorithm 1	7.3914	−1.5779E−06	86	0.1672
	PBM-inexact	7.3914	−1.2867E−06	260	1.6594
	ICPBM	7.3915	−1.3064E−06	338	1.4578
	SolvOpt	7.3919	−3.2938E−09	12,024	0.4176
7 (n=50)	Algorithm 1	21.3548	−8.8956E−07	168	0.3328
	PBM-ineaxct	21.3548	−1.3349E−06	498	10.8950
	ICPBM	21.3548	−1.3313E−06	558	2.5906
	SolvOpt	21.3587	−8.5937E−09	72,294	2.5394
7 (n=100)	Algorithm 1	45.3797	−8.4573E−07	378	0.6344
	PBM-inexact	45.3798	−1.0473E−06	730	39.6781
	ICPBM	45.3798	−1.2934E−06	854	4.2422
	SolvOpt	46.3036	−5.8039E−08	302,652	44.0375
7 (n=200)	Algorithm 1	94.1840	−1.1525E−06	534	0.9438
	PBM-inexact	94.1840	−1.1954E−06	1122	180.6016
	ICPBM	94.1841	−1.3034E−06	1298	7.0609

Table 1 (continued)

P	Solvers	f^*	g^*	NF	time
	SolvOpt	96.7131	−5.2519E−08	472,148	80.6516
8 (n=10)	Algorithm 1	18.2390	−1.0595E−06	686	0.7938
	PBM-inexact	18.2390	−9.8375E−07	262	1.9313
	ICPBM	18.2390	−1.1261E−06	208	0.9578
	SolvOpt	18.2390	−2.7993E−08	2014	0.0766
8 (n=50)	Algorithm 1	189.7162	−9.6575E−07	180	0.3313
	PBM-inexact	189.7162	−1.1590E−06	318	3.2272
	ICPBM	189.7162	−1.1826E−06	526	2.5344
	SolvOpt	189.7166	−6.6921E−07	19,800	0.6784
8 (n=100)	Algorithm 1	413.0583	−8.8474E−07	154	0.3000
	PBM-inexact	413.0583	−1.1991E−06	160	0.6594
	ICPBM	413.0584	−1.1658E−06	608	3.1328
	SolvOpt	413.0635	−1.2452E−06	34,130	4.9025
8 (n=200)	Algorithm 1	862.5086	−9.9902E−07	158	0.3391
	PBM-inexact	862.5086	−1.1208E−06	166	0.7459
	ICPBM	862.5087	−1.1677E−06	654	3.5688
	SolvOpt	862.5118	−9.7218E−06	83,802	14.5703
8 (n=500)	Algorithm 1	2213.7389	−1.0158E−06	156	0.7750
	PBM-inexact	2213.7389	−1.0260E−06	172	1.2391
	ICPBM	2213.7389	−1.1241E−06	694	4.9156
	SolvOpt	2213.8271	−4.6555E−06	82,590	64.3891
8 (n=1000)	Algorithm 1	4466.9124	−7.0467E−07	160	1.9719
	PBM-inexact	4466.9124	−1.1771E−06	174	2.8219
	ICPBM	4466.9124	−1.1522E−06	704	12.2547
	SolvOpt	4467.0549	−1.0938E−05	134,142	347.2688

ASA, the maximum number of iterations is fixed to 10,000. In *fseminf*, the initial sampling interval is set as follows:

$$\begin{cases} a : \dfrac{b-a}{10^3} : b, & p = 1, \\ a_i : \dfrac{b_i - a_i}{10^2} : b_i, i = 1, 2, & p = 2. \end{cases}$$

The stopping criterions of *fseminf* are set as follows:

$$TOLFun = 10^{-6},\ TOLCon = 10^{-6},\ TOLX = 10^{-6}.$$

Results. We have summarized the results of our numerical experiments in Tables 1 and 2, where we have used the following notations:

- P is the number of problems,
- NF is the number of function evaluations,
- NT is the number of nodes in the final subdivision,
- $time$ is the CPU time in seconds,

Table 2 Summary of numerical results with PBM-inexact, ASA, and fseminf

P	Solvers	f^*	G^*	NF	NT	time
9	Algorithm 1	0.1945	−2.1740E−06	62	90	0.0875
	PBM-inexact	0.1945	−3.4578E−06	28	–	0.0447
	ASA	0.1945	−5.8076E−08	–	48	0.3016
	fseminf	0.1945	−2.1805E−13	–	–	0.0354
10	Algorithm 1	4.3012	−1.2546E−06	140	206	0.2094
	PBM-inexact	4.3012	−1.6448E−06	104	–	0.2178
	ASA	4.3012	−1.5595E−07	–	56	0.2750
	fseminf	4.3010	2.1838E−04	–	–	0.1505
11	Algorithm 1	0.6351	−4.1253E−06	166	586	0.2188
	PBM-inexact	0.6351	−1.5555E−06	94	–	0.2059
	ASA	0.6351	−1.0454E−07	–	66	0.3519
	fseminf	0.6350	5.4727E−05	–	–	0.0500
12	Algorithm 1	5.3347	−2.5102E−06	59	46	0.1406
	PBM-inexact	5.3347	−7.9148E−07	60	–	2.5641
	ASA	5.3347	−4.3557E−07	–	22	0.1775
	fseminf	5.1722	0.4332	–	–	0.1641
13	Algorithm 1	1.6701E−06	−1.6700E−06	75	130	0.1609
	PBM-inexact	2.1311E−06	−2.1310E−06	62	–	0.1103
	ASA	Fail	Fail	–	Fail	Fail
	fseminf	6.7670E−07	−7.3619E−09	–	–	0.0911
14	Algorithm 1	2.4172E−06	−2.4148E−06	63	1804	0.1094
	PBM-inexact	5.7657E−06	−5.7656E−06	76	–	4.7275
	ASA	Fail	Fail	–	Fail	Fail
	fseminf	0	−6.7463E−12	–	–	0.1339
15	Algorithm 1	0.4207	−3.1679E−06	52	900	0.0828
	PBM-inexact	0.4207	−5.3241E−06	34	–	0.0863
	ASA	0.4207	−5.2269E−08	–	324	0.2891
	fseminf	0.2787	0.2841	–	–	0.1583
16	Algorithm 1	−3.6922	−0.8149	213	484	0.2672
	PBM-inexact	−3.6922	−0.8152	294	–	2.4753
	ASA	Fail	Fail	–	Fail	Fail
	fseminf	−3.6914	−0.8289	–	–	0.1906
17	Algorithm 1	50.3875	−1.2187E−06	638	1444	1.2734
	PBM-inexact	50.3875	−3.0114E−06	444	–	5.6722
	ASA	50.3875	−2.6130E−09	–	784	7.7278
	fseminf	1.8800E+05	5.4906E+04	–	–	0.0620
18	Algorithm 1	7.9221	−1.1413E−06	152	1444	0.2281
	PBM-inexact	7.9221	−3.0445E−06	110	–	0.2594
	ASA	Fail	Fail	–	Fail	Fail
	fseminf	7.9238	5.8671E−09	–	–	0.3349

Table 2 (continued)

P	Solvers	f^*	G^*	NF	NT	time
19	Algorithm 1	0.2305	−2.6835E−06	172	5580	0.2031
	PBM-inexact	0.2305	−5.6445E−06	228	–	1.4281
	ASA	Fail	Fail	–	Fail	Fail
	fseminf	0.2187	0.0247	–	–	0.2625
20	Algorithm 1	13.3125	−1.4231E−06	194	1444	0.3453
	PBM-inexact	13.3125	−3.2437E−06	134	–	0.3563
	ASA	13.3125	−4.1815E−07	–	256	0.1875
	fseminf	4.9898E+04	5.4906E+04	–	–	0.0833
21	Algorithm 1	0.2079	−2.6777E−06	68	3132	0.1250
	PBM-inexact	0.2079	−5.3948E−06	168	–	0.8125
	ASA	Fail	Fail	–	Fail	Fail
	fseminf	0.1018	0.3082	–	–	0.0969
22	Algorithm 1	13.3125	−1.4231E−06	194	1444	0.2641
	PBM-inexact	13.3125	−3.2437E−06	134	–	0.3353
	ASA	13.3125	−4.1815E−07	–	256	0.1891
	fseminf	4.9898E+04	5.4906E+04	–	–	0.0609
23	Algorithm 1	0.6106	−1.8779E−06	72	2116	0.1141
	PBM-inexact	0.6106	−6.8263E−06	44	–	0.1375
	ASA	0.6106	−6.4529E−10	–	1024	0.5359
	fseminf	0.2227	0.5955	–	–	0.6224

– f^* is the objective function value when the algorithm terminates,
– g^* is the constraint function value when the algorithm terminates for constrained optimization problems,
– G^* measures the feasibility of x^* when the algorithm terminates for SIP problems, i.e.,

$$G^* = \max_{t\in\bar{E}} g(x^*, t),$$

where x^* is the obtained solution of the problem when the algorithm terminates and

$$\bar{E} = \begin{cases} a : \dfrac{b-a}{10^6} : b, & p = 1, \\ a_i : \dfrac{b_i - a_i}{10^4} : b_i, i = 1, 2, & p = 2. \end{cases}$$

The results presented in Table 1 show that the new solver Algorithm 1 is more reliable to find an optimal solution to each problem than the other solvers. The solvers Algorithm 1 and PBM-inexact reach the same optimal value in most problems, only in some of the problems (Problem 3, 7 (n=100)) Algorithm 1 reaches lower objective value. Despite being more unreliable than Algorithm 1, the solver PBM-inexact has also quite a good performance and yields feasible points in all problems. In some of the problems (Problem 7 (n=20,100,200), 8 (n=100,200)) Algorithm 1 reaches a lower objective value than the solver ICPBM. The solver SolvOpt fails to find the optimal value in Problems 7 and 8 ($n\geq$20).

Figure 1 depicts the performance plot comparing the solvers Algorithm 1, PBM-inexact, ICPBM and SolvOpt on Problems 1–8. It can be observed that Algorithm 1 performs much

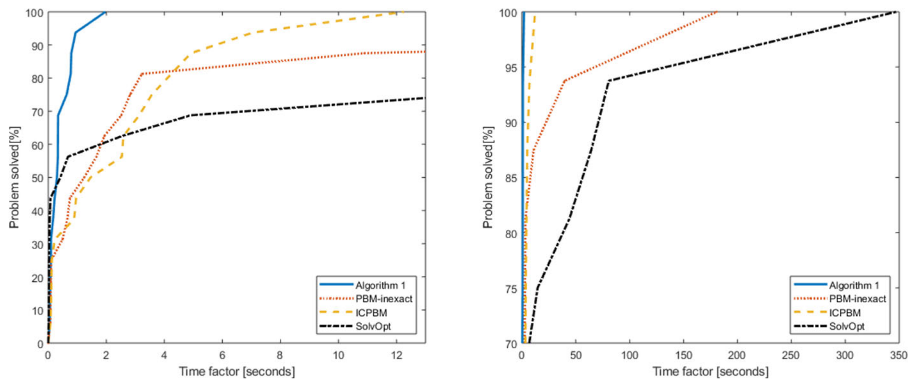

Fig. 1 Performance plot for Algorithm 1, PBM-inexact, ICPBM and SolvOpt

better in terms of CPU time, compared to the other solvers. The solver SolvOpt is the faster one for $n \leq 10$. However, it uses more computational efforts than the other solvers in Problems 7 and 8 ($n > 10$).

All in all, we can conclude that the new solver Algorithm 1 is really efficient to solve the nonconvex, nonsmooth constrained optimization problems.

The results presented in Table 2 show that the solver Algorithm 1 is efficient to find an optimal solution to each SIP problem. The solvers Algorithm 1 and PBM-inexact find a feasible point to each SIP problem. The solver ASA fails to find the optimal solution approximately in 40% of the cases while *fseminf* fails in 80% of the cases. From Table 2, it observes that Algorithm 1 requires a lot more nodes than ASA. However, Algorithm 1 is much faster than ASA. This is expected because of the subproblems of ASA use more computational efforts than the subproblems of Algorithm 1 and the subproblems of our method do not depend on the number of the nodes in the subdivision.

Figure 2 depicts the performance plot comparing the solvers Algorithm 1, PBM-inexact, ASA, and *fseminf* on Problems 9–23. Algorithm 1 performs much better in terms of CPU time, compared to the solvers PBM-inexact and ASA. The solver *fseminf* is the faster one in some cases. However, it cannot find the optimal value in most problems.

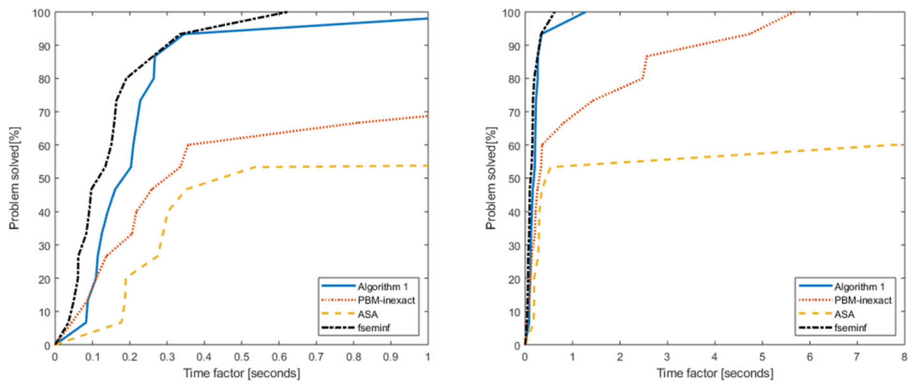

Fig. 2 Performance plot for Algorithm 1, PBM-inexact, ASA, and *fseminf*

All in all, we can conclude that Algorithm 1 is really efficient to solve the nonconvex, nonsmooth SIP problems.

6 Conclusions

In this paper, we design a feasible proximal bundle method with convexification for SIP problems. We consider SIP problems that involve nonsmooth, nonconvex functions and present upper bounding problem. The upper bounding problem is constructed based on a concave relaxation of the lower level problem which results in a restriction of the SIP. The relaxed lower level problem is a finite nonsmooth problem and constructed with ideas from the αBB method of global optimization. Then, we introduce an improvement function of the upper bounding problem, construct a cutting plane model of the improvement function, and reformulate the cutting plane model as a quadratic programming problem and solve it. In contrast to the traditional bundle method, the cutting plane model involves new parameters $E,\ \alpha$ which vary along the iterations. If $D(E,\ \alpha) \to 0$, the upper bounding problems converge to the original SIP problem. We perform an initialization to generate a feasible initial point of our proximal bundle method. Under our refinement procedure and descent condition, we can obtain that each iteration point is feasible. The reasonable convergence properties of the our algorithms are obtained under mild assumptions. The presented results of numerical experiments confirm that Algorithm 1 is efficient for solving nonsmooth, nonconvex SIP problems. The feasibility of Algorithm 1 is demonstrated through theoretical analysis and numerical experiments.

Appendix 1: Test problems

Problem 1 [23, 40] Dimension: n=2,
$f(x) = 8|x_1^2 - x_2| + (1 - x_1)^2$,
$g(x) = \max\left\{\sqrt{2}x_1,\ 2x_2\right\} - 1$,
$X = [-2, 2]^2$, $x^0 = (1, 1)^T$.

Problem 2 [34] Dimension: n=2,
$f(x) = \max\left\{x_1^2 + x_2^4,\ (2 - x_1)^2 + (2 - x_2)^2,\ 2\exp(x_2 - x_1)\right\}$,
$g(x) = \max\left\{-x_1^4 - 2x_2^2 - 1,\ 2x_1^3 - x_2^2 - 2.5\right\}$,
$X = [-4, 4]^2$, $x^0 = (2, 2)^T$.

Problem 3 [34] Dimension: n=2,
$f(x) = \max\left\{x_1^2 + x_2^2 + x_1x_2,\ -x_1^2 - x_2^2 - x_1x_2,\ \sin x_1, -\sin x_1,\ \cos x_2,\ -\cos x_2\right\}$,
$g(x) = \max\left\{-x_1^4 - 2x_2^2 - 1,\ 2x_1^3 - x_2^2 - 2.5\right\}$,
$X = [-4, 4]^2$, $x^0 = (3, 1)^T$.

Problem 4 [34] Dimension: n=2,
$f(x) = \max\left\{x_1^4 + x_2^2,\ (2 - x_1)^2 + (2 - x_2)^2,\ 2\exp(x_2 - x_1)\right\}$,
$g(x) = \max\left\{x_1^2 - x_2^2,\ -2x_1^3 - x_2^2\right\}$,
$X = [-4, 4]^2$, $x^0 = (0, 1)^T$.

Problem 5 [34] Dimension: n=2,
$f(x) = \max\left\{x_1^2 + x_2^2,\ (2 - x_1)^2 + (2 - x_2)^2,\ 2\exp(x_2 - x_1)\right\}$,
$g(x) = \max\left\{x_1 + x_2 - 2,\ -x_1^2 - x_2^2 + 2.25\right\}$,
$X = [-4, 4]^2$, $x^0 = (2.1, 1.9)^T$.

Problem 6 [34] Dimension: n=2,
$f(x) = \max\left\{10(x_2 - x_1^2),\ 10(x_1^2 - x_2),\ 1 - x_1,\ x_1 - 1\right\}$,
$g(x) = \max\left\{100x_1^2 + x_2^2 - 101,\ 80x_1^2 - x_2^2 - 79\right\}$,
$X = [-4, 4]^2$, $x^0 = (-1.2, 1)^T$.

Problem 7 [19] Dimension: n=20, 50, 100, 200,
$$f(x) = \max\left\{\sum_{i=1}^{n-1}(x_i^2 + (x_{i+1} - 1)^2 + x_{i+1} - 1),\ \sum_{i=1}^{n-1}(-x_i^2 - (x_{i+1} - 1)^2 + x_{i+1} + 1)\right\},$$
$g(x) = \sum_{i=1}^{n-1}(x_i^2 + x_{i+1}^2 + x_i x_{i+1} - 2x_i - 2x_{i+1} + 1.0)$,
$X = [-10, 10]^n$, $x^0 = \text{ones}(n, 1)$.

Problem 8 [19, 40] Dimension: n=10, 50, 100, 200, 500, 1000,
$f(x) = \sum_{i=1}^{n-1}(-x_i + 2(x_i^2 + x_{i+1}^2 - 1) + 1.75|x_i^2 + x_{i+1}^2 - 1|)$,
$g(x) = \sum_{i=1}^{n-2}((3 - 2x_{i+1})x_{i+1} - x_i - 2x_{i+2} + 2.5)$,
$X = [-10, 10]^n$, $x^0 = \text{ones}(n, 1)$.

Problem 9 [41] Dimension: $n = 2,\ p = 1$,
$f(x) = \frac{1}{3}x_1^2 + x_2^2 + \frac{1}{2}x_1$,
$g(x, t) = (1 - x_1^2 t^2)^2 - x_1 t^2 - x_2^2 + x_2$,
$X = [-2, 2]^2$, $T = [0, 1]$, $x^0 = (-1, -1)^T$.

Problem 10 [28] Dimension: $n = 3,\ p = 1$,
$f(x) = \exp(x_1) + \exp(x_2) + \exp(x_3)$,
$g(x, t) = 1/(1 + t^2) - x_1 - x_2 t - x_3 t^2$,
$X = [-2, 2]^3$, $T = [0, 1]$, $x^0 = (1, 1, 1)^T$.

Problem 11 [28] Dimension: $n = 3,\ p = 1$,
$f(x) = x_1 + x_2/2 + x_3/3$,
$g(x, t) = \exp(t - 1) - x_1 - x_2 t - x_3 t^2$,
$X = [-2, 2]^3$, $T = [0, 1]$, $x^0 = (1, 1, 1)^T$.

Problem 12 [41] Dimension: $n = 3,\ p = 1$,
$f(x) = x_1^2 + x_2^2 + x_3^2$,
$g(x, t) = x_1 + x_2\exp(x_3 t) + \exp(2t) - 2\sin(4t)$,
$X = [-4, 2]^3$, $T = [0, 1]$, $x^0 = (1, 1, 1)^T$.

Problem 13 [31, 32] Dimension: $n = 3,\ p = 2$,
$f(x) = |x_1| + |x_2| + |x_3|$,

$g(x,t) = x_1 + x_2\exp(x_3t) - \exp(2x_1t) + \sin(4t)$,
$X = [-1, 1]^3$, $T = [0, 1]$, $x^0 = \text{ones}(3, 1)$.

Problem 14 [31, 32] Dimension: $n = 3$, $p = 2$,
$f(x) = |x_1| + |x_2| + |x_3|$,
$g(x,t) = x_1 + x_2\exp(x_3t_1) - \exp(2t_2) + \sin(4t_1)$,
$X = [-1, 1]^3$, $T = [0, 1]^2$, $x^0 = \text{ones}(3, 1)$.

Problem 15 [31, 32] Dimension: $n = 4$, $p = 2$,
$f(x) = 1/2(|x_1| + |x_2| + |x_3| + |x_4|)$,
$g(x,t) = \sin(t_1t_2) - x_1 - x_2t_1 - x_3t_2 - x_4t_1t_2$,
$X = [-4, 4]^6$, $T = [0, 1]^2$, $x^0 = \text{ones}(4, 1)$.

Problem 16 Dimension: $n = 6$, $p = 2$,
$f(x) = \sum_{i=1}^{n-1}(-x_i + 2(x_i^2 + x_{i+1}^2 - 1) + 1.75|x_i^2 + x_{i+1}^2 - 1|)$,
$g(x,t) = \sin(t_1t_2) - x_1 - x_2t_1 - x_3t_2 - x_4t_1^2 - x_5t_1t_2 - x_6t_2^2$,
$X = [-4, 4]^6$, $T = [0, 1]^2$, $x^0 = \text{ones}(6, 1)$.

Problem 17 Dimension: $n = 6$, $p = 2$,
$f(x) = \sum_{i=1}^{n-1}(-x_i + 2(x_i^2 + x_{i+1}^2 - 1) + 1.75|x_i^2 + x_{i+1}^2 - 1|)$,
$g(x,t) = (1 + t_1^2 + t_2^2)^2 - x_1 - x_2t_1 - x_3t_2 - x_4t_1^2 - x_5t_1t_2 - x_6t_2^2$,
$X = [-4, 4]^6$, $T = [0, 1]^2$, $x^0 = \text{ones}(6, 1)$.

Problem 18 Dimension: $n = 6$, $p = 2$,
$f(x) = \sum_{i=1}^{n-1}\max\left\{x_i^2 + (x_{i+1} - 1)^2 + x_{i+1} - 1,\ -x_i^2 - (x_{i+1} - 1)^2 + x_{i+1} + 1\right\}$,
$g(x,t) = \exp(t_1^2 + t_2^2) - x_1 - x_2t_1 - x_3t_2 - x_4t_1^2 - x_5t_1t_2 - x_6t_2^2$,
$X = [-4, 4]^6$, $T = [0, 1]^2$, $x^0 = \text{ones}(6, 1)$.

Problem 19 Dimension: $n = 6$, $p = 2$,
$f(x) = \sum_{i=1}^{n-1}\max\left\{x_i^2 + (x_{i+1} - 1)^2 + x_{i+1} - 1,\ -x_i^2 - (x_{i+1} - 1)^2 + x_{i+1} + 1\right\}$,
$g(x,t) = \sin(t_1t_2) - x_1 - x_2t_1 - x_3t_2 - x_4t_1^2 - x_5t_1t_2 - x_6t_2^2$,
$X = [-4, 4]^6$, $T = [0, 1]^2$, $x^0 = \text{ones}(6, 1)$.

Problem 20 Dimension: $n = 6$, $p = 2$,
$f(x) = \sum_{i=1}^{n-1}\max\left\{x_i^2 + (x_{i+1} - 1)^2 + x_{i+1} - 1,\ -x_i^2 - (x_{i+1} - 1)^2 + x_{i+1} + 1\right\}$,
$g(x,t) = (1 + t_1^2 + t_2^2)^2 - x_1 - x_2t_1 - x_3t_2 - x_4t_1^2 - x_5t_1t_2 - x_6t_2^2$,
$X = [-4, 4]^6$, $T = [0, 1]^2$, $x^0 = \text{ones}(6, 1)$.

Problem 21 Dimension: $n = 6$, $p = 2$,

$$f(x) = \max\left\{\sum_{i=1}^{n-1}(x_i^2 + (x_{i+1} - 1)^2 + x_{i+1} - 1),\ \sum_{i=1}^{n-1}(-x_i^2 - (x_{i+1} - 1)^2 + x_{i+1} + 1)\right\},$$

$g(x,t) = \sin(t_1t_2) - x_1 - x_2t_1 - x_3t_2 - x_4t_1^2 - x_5t_1t_2 - x_6t_2^2$,
$X = [-4, 4]^6$, $T = [0, 1]^2$, $x^0 = \text{ones}(6, 1)$.

Problem 22 Dimension: $n = 6,\ \ p = 2$,

$$f(x) = \max\left\{\sum_{i=1}^{n-1}(x_i^2 + (x_{i+1}-1)^2 + x_{i+1} - 1),\right.$$
$$\left.\sum_{i=1}^{n-1}(-x_i^2 - (x_{i+1}-1)^2 + x_{i+1} + 1)\right\},$$
$$g(x,t) = (1 + t_1^2 + t_2^2)^2 - x_1 - x_2t_1 - x_3t_2 - x_4t_1^2 - x_5t_1t_2 - x_6t_2^2,$$
$$X = [-4, 4]^6,\ T = [0, 1]^2,\ x^0 = \text{ones}(6, 1).$$

Problem 23 Dimension: $n = 6,\ \ p = 2$,

$$f(x) = \max_{1\le i\le n}\left\{h(-\textstyle\sum_{i=1}^n x_i),\ h(x_i)\right\}, \text{ where } h(y) = \ln(|y|+1),\ \forall y \in \mathbb{R},$$
$$g(x,t) = \sin(t_1t_2) - x_1 - x_2t_1 - x_3t_2 - x_4t_1^2 - x_5t_1t_2 - x_6t_2^2,$$
$$X = [-4, 4]^6,\ T = [0, 1]^2,\ x^0 = \text{ones}(6, 1).$$

Funding This work was supported by the Key Research and Development Projects of Shandong Province (NO. 2019GGX104089), and the Natural Science Foundation of Shandong Province (NO. ZR2019BA014).

References

1. Adjiman, C.S., Androulakis, I.P., Floudas, C.A.: A global optimization method, αBB, for general twice-differentiable constrained NLPs-I: theoretical advances. Comput. Chem. Eng. **22**, 1137–1158 (1998)
2. 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. Comput. Chem. Eng. **22**, 1159–1179 (1998)
3. Bagirov, A., Karmitsa, N., Mäkelä, M.M.: Introduction to Nonsmooth Optimization: Theory, Practice and Software. Springer, Cham (2014)
4. Bhattacharjee, B., Green, Jr.W.H., Barton, P.I.: Interval methods for semi-infinite programs. Comput. Optim. Appl. **30**, 63–93 (2005)
5. Bhattacharjee, B., Lemonidis, P., Green, W.H. Jr.., Barton, P.I.: Global solution of semi-infinite programs. Math. Program. **103**, 283–307 (2005)
6. Clarke, F.H., Ledyaev, Y.S., Stern, R.J., Wolenski, P.R.: Nonsmooth Analysis and Control Theory. Springer, New York (1998)
7. Daniilidis, A., Georgiev, P.: Approximate convexity and submonotonicity. J. Math. Anal. Appl. **291**(1), 292–301 (2004)
8. Djelassi, H., Mitsos, A.: A hybrid discretization algorithm with guaranteed feasibility for the global solution of semi-infinite programs. J Glob Optim. **68**, 227–253 (2017)
9. Floudas, C.A., Stein, O.: The adaptive convexification algorithm: a feasible point method for semi-infinite programming. SIAM J. Optim. **18**, 1187–1208 (2007)
10. Fuduli, A., Gaudioso, M., Giallombardo, G., Miglionico, G.: A partially inexact bundle method for convex semi-infinite minmax problems. Commun. Nonlinear Sci. Numer. Simul. **21**, 172–180 (2015)
11. Gustafson, S.-Å., Kortanek, K.: Semi-infinite programming and applications. In: Mathematical Programming The State of the Art, pp. 132–157. Springer, Berlin (1983)
12. Hare, W., Sagastizábal, C.: A redistributed proximal bundle method for nonconvex optimization. SIAM J. Optim. **20**(5), 2442–2473 (2010)
13. Hare, W., Sagastizábal, C., Solodov, M.: A proximal bundle method for nonsmooth nonconvex functions with inexact information. Comput. Optim. Appl. **63**, 1–28 (2016)

14. Hettich, R.: An implementation of a discretization method for semi-infinite programming. Math. Program. **34**(3), 354–361 (1986)
15. Hettich, R., Kortanek, K.O.: Semi-infinite programming: theory, methods, and applications. Soc. Ind. Appl. Math. **35**(3), 380–429 (1993)
16. Hoseini Monjezi, N., Nobakhtian, S.: A filter proximal bundle method for nonsmooth nonconvex constrained optimization. J. Glob. Optim. (2020)
17. Hoseini Monjezi, N., Nobakhtian, S.: A new infeasible proximal bundle algorithm for nonsmooth nonconvex constrained optimization. Comput Optim Appl. **74**, 443–480 (2019)
18. Karas, E., Ribeiro, A., Sagastizábal, C., Solodov, M.: A bundle-filter method for nonsmooth convex constrained optimization. Math. Program. **116**, 297–320 (2009)
19. Karmitsa, N.: Test problems for large-scale nonsmooth minimization. Reports of the Department of Mathematical Information Technology, Series B, Scientific computing, No. B 4/2007 (2007)
20. Kiwiel, K.C.: Methods of Descent for Nondifferentiable Optimization, vol. 1133. Springer, Berlin, Heidelberg (2006)
21. Kortanek, K.O., No, H.: A central cutting plane algorithm for convex semi-infinite programming problems. SIAM J. Optim. **3**(4), 901–918 (1993)
22. López, M., Still, G.: Semi-infinite programming. Eur. J. Oper. Res. **180**(2), 491–518 (2007)
23. Lv, J., Pang, L.P., Meng, F.Y.: A proximal bundle method for constrained nonsmooth nonconvex optimization with inexact information. J. Global Optim. **70**, 517–549 (2018)
24. Lv, J., Pang, L.P., Xu, N., Xiao, Z.H.: An infeasible bundle method for nonconvex constrained optimization with application to semi-infinite programming problems. Numer. Algorithms **80**, 397–427 (2019)
25. Meng, F.Y., Pang, L.P., Lv, J., Wang, J.H.: An approximate bundle method for solving nonsmooth equilibrium problems. J. Global Optim. **68**, 537–562 (2017)
26. Mifflin, R.: Semismooth and semiconvex functions in constrained optimization. SAIM J. Control Optim. **15**(6), 959–972 (1977)
27. Mitsos, A.: Global optimization of semi-infinite programs via restriction of the right-hand side. Optimization **60**(10-11), 1291–1308 (2011)
28. Mitsos, A., Djelassi, H.: A Test Set of Semi-Infinite Programs. Process Systems Engineering (AVT.SVT), RWTH Aachen University, Aachen, Germany. http://web.mit.edu/mitsos/www/pubs/siptestset.pdf (2016)
29. Mitsos, A., Lemonidis, P., Lee, C.K., Barton, P.I.: Relaxation-based bounds for semi-infinite programs. SIAM J. Optim. **19**, 77–113 (2007)
30. Mitsos, A., Tsoukalas, A.: Global optimization of generalized semi-infinite programs via restriction of the right hand side. J. Glob. Optim. **61**(1), 1–17 (2015)
31. Pang, L.P., Lv, J., Wang, J.H.: Constrained incremental bundle method with partial inexact oracle for nonsmooth convex semi-infinite programming problems. Comput. Optim. Appl. **64**(2), 433–465 (2016)
32. Pang, L., Wu, Q., Wang, J., et al.: A discretization algorithm for nonsmooth convex semi-infinite programming problems based on bundle methods. Comput Optim Appl. **76**, 125–153 (2020)
33. Polyak, B.T.: Introduction to Optimization. Optimization Software, Inc., Publications Division, New York (1987)
34. Rustem, B., Nguyen, Q.: An algorithm for the inequality-constrained discrete min-max problem. SIAM J. Optim. **8**, 265–283 (1998)
35. Sagastizábal, C., Solodov, M.: An infeasible bundle method for nonsmooth convex constrained optimization without a penalty function or a filter. SIAM J. Optim. **16**(1), 146–169 (2005)
36. Shiu, T.J., Wu, S.Y.: Relaxed cutting plane method with convexification for solving nonlinear semi-infinite programming problems. Comput Optim Appl. **53**, 91–113 (2012)
37. Spingarn, J.E.: Submonotone subdifferentials of Lipschitz functions. Trans. Am. Math. Soc. **264**, 77–89 (1981)

38. Stein, O., Steuermann, P.: The adaptive convexification algorithm for semi-infinite programming with arbitrary index sets. Math. Program. **136**, 183–207 (2012)
39. Still, G.: Discretization in semi-infinite programming: the rate of convergence. Math. Programm. **91**(1), 53–69 (2001)
40. Tang, C., Liu, S., Jian, J., et al.: A feasible SQP-GS algorithm for nonconvex, nonsmooth constrained optimization. Numer Algor. **65**, 1–22 (2014)
41. Wang, S.X., Yuan, Y.X.: Feasible method for semi-infinite programs. SIAM J. Optim. **25**(4), 2537–2560 (2015)
42. Xu, M.W., Wu, S.Y., Ye, J.J.: Solving semi-infinite programs by smoothing projected gradient method. Comput. Optim. Appl. **59**, 591–616 (2014)
43. Yang, Y., Pang, L., Ma, X., et al.: Constrained Nonconvex Nonsmooth Optimization via Proximal Bundle Method. J Optim Theory Appl. **163**, 900–925 (2014)
44. Zhang, L.P., Wu, S.Y., López, M.A.: A new exchange method for convex semi-infinite programming. SIAM J. Optim. **20**(6), 2959–2977 (2010)