J Glob Optim (2015) 63:213–227
DOI 10.1007/s10898-015-0284-5

Convex mixed integer nonlinear programming problems and an outer approximation algorithm

Zhou Wei · M. Montaz Ali

Received: 2 November 2014 / Accepted: 10 February 2015 / Published online: 18 February 2015
© Springer Science+Business Media New York 2015

Abstract In this paper, we mainly study one class of convex mixed-integer nonlinear programming problems (MINLPs) with non-differentiable data. By dropping the differentiability assumption, we substitute gradients with subgradients obtained from KKT conditions, and use the outer approximation method to reformulate convex MINLP as one equivalent MILP master program. By solving a finite sequence of subproblems and relaxed MILP problems, we establish an outer approximation algorithm to find the optimal solution of this convex MINLP. The convergence of this algorithm is also presented. The work of this paper generalizes and extends the outer approximation method in the sense of dealing with convex MINLPs from differentiable case to non-differentiable one.

Keywords Convex MINLP · Outer approximation · Decomposition · Master program

Mathematics Subject Classification 90C11 · 90C25 · 90C30

1 Introduction

Many practical optimization problems are modelled as mixed-integer nonlinear programming problems (MINLPs) involving continuous and discrete variables and the study of solution algorithms for these optimization problems has been an active focus of research over the past

Z. Wei (✉)
Department of Mathematics, Yunnan University, Kunming 650091, People's Republic of China
e-mail: wzhou@ynu.edu.cn

M. M. Ali
School of Computational and Applied Mathematics, University of the Witwatersrand, Johannesburg 2050, South Africa

M. M. Ali
TCSE, Faculty of Engineering and Built Environment, University of the Witwatersrand, Johannesburg 2050, South Africa
e-mail: Montaz.Ali@wits.ac.za

decades (cf. [3,10,12–16,21,22,25] and references therein). Suppose $f, g_i : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}\ (i = 1, \ldots, m)$ are nonlinear functions, X is a nonempty compact convex set in $\mathbb{R}^n$ and Y is a set of discrete variables in $\mathbb{R}^p$. The general form for MINLPs is defined mathematically as follows:

$$\text{(P)} \begin{cases} \underset{x,\,y}{\text{minimize}} \ \ f(x, y) \\ \text{subject to } g_i(x, y) \leq 0, i = 1, \ldots, m, \\ \qquad\qquad x \in X, y \in Y \text{ discrete variable.} \end{cases} \tag{1.1}$$

This paper is devoted to one class of convex MINLPs in which objective and constraint functions f, g_i for $i = 1, \ldots, m$ are convex but not differentiable.

The class of convex MINLPs has been extensively studied by many authors and several methods for these MINLPs have been developed over past decades. These methods include branch-and-bound, generalized Benders decomposition, extended cutting-plane method, NLP/LP based branch and bound and outer approximation method (cf. [2,5–9,11,12,15,19,23–25] and references therein). Note that the extended cutting-plane method was proposed by Westerlund and Pettersson [24] for solving differentiable convex MINLPs. Subsequently, Westerlund and Pettersson [25] presented this method to deal with a more general case of MINLPs including pseudo-convex functions. It was shown in [25] that one MINLP with pseudo-convex functions and pseudo-convex constraints can be solved to global optimality by the cutting-plane techniques. In 2014, Eronen et al. [6] generalized the extended cutting-plane method for solving convex nonsmooth MINLPs and provided one ECP algorithm which was proved to converge to one global optimum. Recently they [7] further considered this extended cutting plane method to deal with nonsmooth MINLPs with pseudo-convexity assumptions.

It is known that Duran and Grossmann [5] introduced the outer approximation method to deal with a particular class of MINLPs which was restricted to contain separable convex differentiable functions and not general convex differentiable functions in all variables. These separable convex functions were composed of convex differentiable functions in continuous variables and linear functions in discrete variables separately. Afterwards Fletcher and Leyffer [8] further extended the outer approximation method for solving convex MINLPs with convex and continuously differentiable objective and constraint functions, and provided a linear outer approximation algorithm to attain the optimal solution of this MINLP by solving a finite sequence of relaxed subproblems. This extension is the pioneering work on outer approximation method in a sense of solving MINLPs where the discrete variables are considered as nonlinear. In 2008, Bonami et. al [2] also studied outer approximation algorithms for convex and continuously differentiable MINLPs. Recently the authors in [6] and [23] used the outer approximation method to study convex nonsmooth MINLPs and established the resulting algorithms. It is noted that differentiability of functions plays an important role in the construction of relaxation and is proved to be an important matter for allowing to solve these relaxed subproblems efficiently. Since nonsmooth optimization problems defined by non-differentiable functions appear in practice, from the theoretical viewpoint as well as for applications, it is interesting and significant to consider convex and non-differentiable MINLPs. Motivated by this, in this paper, we are inspired by [2,5,6,8,23] to continue studying one convex MINLP by dropping the differentiability assumption and aim to construct an outer approximation algorithm for solving this MINLP. The outer approximation method used herein is along the line given in [8,23] and consists of the use of KKT conditions to linearize the objective and constraint functions at different points for constructing an equivalent MILP relaxation of the problem.

The paper is organized as follows. In Sect. 2, we give some definitions and preliminaries used in this paper. Section 3 contains the equivalent reformulation of convex MINLP by the outer approximation method and one outer approximation algorithm for finding optimal solutions of this MINLP. The reformulation is mainly dependent on KKT conditions and projection techniques. For the algorithm construction, it is necessary to solve a finite sequence of nonlinear programs including feasible and infeasible subproblems and the relaxations of mixed-integer linear master program. The convergence theorem for the established algorithm is also presented therein. The conclusion of this paper is presented in Sect. 4. Section 5 is an "Appendix" which contains the proofs of the main results given for constructing the algorithm in the paper.

2 Preliminaries

Let $\|\cdot\|$ be the norm of $\mathbb{R}^n$ and denote the inner product between two elements of $\mathbb{R}^n$ by $\langle\cdot,\cdot\rangle$. Let Ω be a closed convex set of $\mathbb{R}^n$ and $x \in \Omega$. We denote $T(\Omega, x)$ the contingent cone of Ω at x; that is, $v \in T(\Omega, x)$ if and only if there exist a sequence $\{v_k\}$ in $\mathbb{R}^n$ converging to v and a sequence t_k in $(0, +\infty)$ decreasing to 0 such that $x + t_k v_k \in \Omega$ for all $k \in \mathbb{N}$, where $\mathbb{N}$ denotes the set of all natural numbers. It is known from [1] that

$$T(\Omega, x) = cl(\mathbb{R}_+(\Omega - x))$$

where cl denotes the closure.

Let $N(\Omega, x)$ denote the normal cone of Ω at x, that is

$$N(\Omega, x) := \{\gamma \in \mathbb{R}^n : \langle\gamma, z - x\rangle \le 0 \quad \text{for all } z \in \Omega\}. \tag{2.1}$$

It is easy to verify that normal cone $N(\Omega, x)$ and contingent cone $T(\Omega, x)$ are the polar dual; that is

$$N(\Omega, x) = \big(T(\Omega, x)\big)^\circ := \big\{\gamma \in \mathbb{R}^n : \langle\gamma, v\rangle \le 0 \quad \text{for all } v \in T(\Omega, x)\big\}.$$

Readers are invited to consult the book [1] for more details on contingent cone and normal cone.

Let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be a continuous convex function, $\bar{x} \in \mathbb{R}^n$ and $h \in \mathbb{R}^n$. Recall (cf. [20]) that $d^+\varphi(\bar{x})(h)$ denotes the directional derivative of φ at $\bar{x}$ along the direction h and is defined by

$$d^+\varphi(\bar{x})(h) := \lim_{t\to 0^+} \frac{\varphi(\bar{x} + th) - \varphi(\bar{x})}{t}.$$

We denote $\partial\varphi(\bar{x})$ the subdifferential of φ at $\bar{x}$ which is defined by

$$\partial\varphi(\bar{x}) := \{\alpha \in \mathbb{R}^n : \ \langle\alpha, x - \bar{x}\rangle \le \varphi(x) - \varphi(\bar{x}) \text{ for all } x \in \mathbb{R}^n\}.$$

Each vector in $\partial\varphi(\bar{x})$ is called a subgradient of φ at $\bar{x}$. It is known from [20] that $\alpha \in \partial\varphi(\bar{x})$ if and only if

$$\langle\alpha, h\rangle \le d^+\varphi(\bar{x})(h) \quad \text{for all } h \in \mathbb{R}^n.$$

Recall that φ is said to be Gâteaux differentiable at $\bar{x}$ if there exists $d\varphi(\bar{x}) \in \mathbb{R}^n$ such that

$$\lim_{t\to 0^+} \frac{\varphi(\bar{x} + th) - \varphi(\bar{x})}{t} = \langle d\varphi(\bar{x}), h\rangle \quad \text{for all } h \in \mathbb{R}^n \tag{2.2}$$

and φ is said to be Fréchet differentiable at $\bar{x}$ if φ is Gâteaux differentiable there and the limit in (2.2) exists uniformly for $\|h\| \leq 1$ as $t \to 0^+$.

It is known from [20] that φ is Gâteaux differentiable at $\bar{x}$ if and only if $\partial\varphi(\bar{x})$ is the singleton. Further, Gâteaux differentiability of φ is equivalent to the Fréchet differentiability of φ due to the local Lipschitzian property of φ and the compactness of unit closed ball in $\mathbb{R}^n$.

Given a continuous convex function $\phi : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}$ and $(\bar{x}, \bar{y}) \in \mathbb{R}^n \times \mathbb{R}^p$, one vector $(\alpha, \beta) \in \mathbb{R}^n \times \mathbb{R}^p$ is the subgradient of ϕ at $(\bar{x}, \bar{y})$ if and only if

$$\phi(x, y) \geq \phi(\bar{x}, \bar{y}) + (\alpha, \beta)^T \begin{pmatrix} x - \bar{x} \\ y - \bar{y} \end{pmatrix} \quad \text{for all } (x, y) \in \mathbb{R}^n \times \mathbb{R}^p, \tag{2.3}$$

where $(\alpha, \beta)^T$ is the transpose of matrix (α, β). When $\bar{y}$ is fixed (*resp.* $\bar{x}$ is fixed), the subdifferential of $\phi(\cdot, \bar{y})$ (*resp.* $\phi(\bar{x}, \cdot)$) at $\bar{x}$ (*resp.* $\bar{y}$) is the set defined by

$$\partial\phi(\cdot, \bar{y})(\bar{x}) := \left\{\alpha \in \mathbb{R}^n : \phi(x, \bar{y}) \geq \phi(\bar{x}, \bar{y}) + \langle \alpha, x - \bar{x}\rangle \quad \text{for all } x \in \mathbb{R}^n\right\}$$
$$\left(\textit{resp. } \partial\phi(\bar{x}, \cdot)(\bar{y}) := \left\{\beta \in \mathbb{R}^p : \phi(\bar{x}, y) \geq \phi(\bar{x}, \bar{y}) + \langle \beta, y - \bar{y}\rangle \quad \text{for all } y \in \mathbb{R}^p\right\}\right).$$

The following proposition on the subdifferential of convex functions is easy to verify from the definition.

Proposition 2.1 *Let $\phi : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}$ be a continuous convex function and $(\bar{x}, \bar{y}) \in \mathbb{R}^n \times \mathbb{R}^p$. Then for any $(\alpha, \beta) \in \partial\phi(\bar{x}, \bar{y})$, one has $\alpha \in \partial\phi(\cdot, \bar{y})(\bar{x})$ and $\beta \in \partial\phi(\bar{x}, \cdot)(\bar{y})$.*

It is an interesting question to consider the converse of Proposition 2.1. This question is also interesting even for smooth convex functions in mathematical analysis. The question is explicitly stated as follows:

Given one continuous convex function $\phi : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}$ and one vector $\bar{\alpha}$ from $\partial\phi(\cdot, \bar{y})(\bar{x})$, whether or not is there some vector $\bar{\beta} \in \mathbb{R}^p$ such that $(\bar{\alpha}, \bar{\beta}) \in \partial\phi(\bar{x}, \bar{y})$?

The following two propositions provided an affirmative answer to this question. These propositions will play a key role in construction of outer approximation algorithm in the sequel. The first proposition is on convex and Fréchet differentiable functions.

Proposition 2.2 *Let $\phi : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}$ be a continuous convex function and $(\bar{x}, \bar{y}) \in \mathbb{R}^n \times \mathbb{R}^p$. Suppose that $\phi(\cdot, \bar{y})$ is Fréchet differentiable at $\bar{x}$ and $\phi(\bar{x}, \cdot)$ is Fréchet differentiable at $\bar{y}$. Then ϕ is Fréchet differentiable at $(\bar{x}, \bar{y})$.*

Proof By the Fréchet differentiability of $\phi(\cdot, \bar{y})$ and $\phi(\bar{x}, \cdot)$, one has

$$\partial\phi(\cdot, \bar{y})(\bar{x}) = \{\nabla_x\phi(\bar{x}, \bar{y})\} \quad \text{and} \quad \partial\phi(\bar{x}, \cdot)(\bar{y}) = \{\nabla_y\phi(\bar{x}, \bar{y})\}.$$

This and Proposition 2.1 imply that $\partial\phi(\bar{x}, \bar{y})$ is the singleton and

$$\partial\phi(\bar{x}, \bar{y}) = \{(\nabla_x\phi(\bar{x}, \bar{y}), \nabla_y\phi(\bar{x}, \bar{y})\}.$$

Hence ϕ is Gâteaux differentiable at $(\bar{x}, \bar{y})$ and consequently Fréchet differentiable at $(\bar{x}, \bar{y})$. The proof is complete. □

Proposition 2.2 may not necessarily be true for non-convex functions. Consider function ϕ on $\mathbb{R} \times \mathbb{R}$ defined as: $\phi(x, y) = \frac{x^2y^2}{(x^2+y^2)^{3/2}}$ if $x^2 + y^2 \neq 0$ and $\phi(x, y) = 0$ if $x^2 + y^2 = 0$. Then ϕ is continuous on $\mathbb{R} \times \mathbb{R}$ and partial derivatives $\nabla_x\phi(0, 0)$ and $\nabla_y\phi(0, 0)$ exist ($\nabla_x\phi(0, 0) = \nabla_y\phi(0, 0) = 0$). However, one can verify that ϕ is not differentiable at $(0, 0)$.

Proposition 2.3 *Let $\phi : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}$ be a continuous convex function and $(\bar{x}, \bar{y}) \in \mathbb{R}^n \times \mathbb{R}^p$. Then for any $\bar{\alpha} \in \partial\phi(\cdot, \bar{y})(\bar{x})$, there exists $\bar{\beta} \in \mathbb{R}^p$ such that $(\bar{\alpha}, \bar{\beta}) \in \partial\phi(\bar{x}, \bar{y})$.*

Proof Let $F_{\bar{y}} : \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^p$ be defined by $F_{\bar{y}}(x) := (x, \bar{y})$. Then $\phi(\cdot, \bar{y}) = \phi \circ F_{\bar{y}}$, and it is easy to verify that $F_{\bar{y}}$ is differentiable at $\bar{x}$ and

$$\nabla F_{\bar{y}}(\bar{x})(h) = (h, 0) \in \mathbb{R}^n \times \mathbb{R}^p \tag{2.4}$$

holds for all $h \in \mathbb{R}^n$. Let $\bar{\alpha} \in \partial\phi(\cdot, \bar{y})(\bar{x})$. We first prove that

$$\bar{\alpha} \in \nabla F_{\bar{y}}(\bar{x})^*(\partial\phi(\bar{x}, \bar{y})) \tag{2.5}$$

where $\nabla F_{\bar{y}}(\bar{x})^*$ is the conjugate operator of $\nabla F_{\bar{y}}(\bar{x})$.

Since ϕ is continuous at $(\bar{x}, \bar{y})$, it follows that $\partial\phi(\bar{x}, \bar{y})$ is a nonempty, convex and compact subset by [20, Proposition1.11] and then $\nabla F_{\bar{y}}(\bar{x})^*(\partial\phi(\bar{x}, \bar{y}))$ is convex and compact as $\nabla F_{\bar{y}}(\bar{x})^*$ is continuous.

Suppose to the contrary that $\bar{\alpha} \notin \nabla F_{\bar{y}}(\bar{x})^*(\partial\phi(\bar{x}, \bar{y}))$. By the seperation theorem, there exists $\bar{u} \in \mathbb{R}^n$ with $\|\bar{u}\| = 1$ such that

$$\begin{aligned}\langle \bar{\alpha}, \bar{u}\rangle &> \max\{\langle \nabla F_{\bar{y}}(\bar{x})^*(\alpha, \beta), \bar{u}\rangle : (\alpha, \beta) \in \partial\phi(\bar{x}, \bar{y})\}\\ &= \max\{\langle (\alpha, \beta), \nabla F_{\bar{y}}(\bar{x})(\bar{u})\rangle : (\alpha, \beta) \in \partial\phi(\bar{x}, \bar{y})\}.\end{aligned}$$

This and (2.4) imply that

$$\langle \bar{\alpha}, \bar{u}\rangle > \max\{\langle (\alpha, \beta), (\bar{u}, 0)\rangle : (\alpha, \beta) \in \partial\phi(\bar{x}, \bar{y})\}. \tag{2.6}$$

Noting that $\bar{\alpha} \in \partial\phi(\cdot, \bar{y})(\bar{x})$ and ϕ is a continuous convex function on $\mathbb{R}^n \times \mathbb{R}^p$, it follows from [20, Proposition 2.24] and (2.6) that

$$d^+\phi(\cdot, \bar{y})(\bar{x})(\bar{u}) \geq \langle \bar{\alpha}, \bar{u}\rangle > d^+\phi(\bar{x}, \bar{y})(\bar{u}, 0) = d^+\phi(\cdot, \bar{y})(\bar{x})(\bar{u}),$$

which is contradiction. Thus (2.5) holds.

By virtue of (2.5), there exists $(\hat{\alpha}, \bar{\beta}) \in \partial\phi(\bar{x}, \bar{y})$ such that $\bar{\alpha} = \nabla F_{\bar{y}}(\bar{x})^*(\hat{\alpha}, \bar{\beta})$. It suffices to prove that $\bar{\alpha} = \hat{\alpha}$.

For any $h \in \mathbb{R}^n$, by using (2.4), one has

$$\langle \bar{\alpha}, h\rangle = \langle \nabla F_{\bar{y}}(\bar{x})^*(\hat{\alpha}, \bar{\beta}), h\rangle = \langle (\hat{\alpha}, \bar{\beta}), \nabla F_{\bar{y}}(\bar{x})(h)\rangle = \langle (\hat{\alpha}, \bar{\beta}), (h, 0)\rangle = \langle \hat{\alpha}, h\rangle.$$

This means that $\bar{\alpha} = \hat{\alpha}$. The proof is complete. □

The following proposition is on the subdifferential of maximum function of two convex functions which is from [26, Theorem 2.4.18]. This result will be used later in our analysis.

Proposition 2.4 *Let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be a convex and continuous function. Define $\varphi_+(x) := \max\{\varphi(x), 0\}$ for all $x \in \mathbb{R}^n$. Then φ_+ is a convex continuous function and*

$$\partial\varphi_+(x) = [0, 1]\partial\varphi(x) \tag{2.7}$$

holds for all $x \in \mathbb{R}^n$ with $\varphi(x) = 0$, where $[0, 1]\partial\varphi(x) := \{t\gamma : t \in [0, 1]$ and $\gamma \in \partial\varphi(x)\}$ for any $x \in \mathbb{R}^n$.

3 Main results

In this section, we mainly study convex MINLP problem of (1.1) by dropping the differentiability assumption and aim to establish one outer approximation algorithm for solving such problem.

Let convex MINLP be defined as (1.1) and set $g := (g_1, \ldots, g_m)$. For the case that $f, g_i (i = 1, \ldots, m)$ in (1.1) are convex and smooth, it is known from [2,6,8] that main idea of outer approximation algorithm for convex smooth MINLPs is using linearization of the objective function and the constraints at different points to build a mixed-integer linear program (MILP) relaxation of the problem; that is, given some set K with optimal solutions of several optimization problems, it is possible to build a relaxation of problem (P) in (1.1):

$$\begin{cases} \text{minimize } \theta \\ \text{subject to } f(x_j, y_j) + \nabla f(x_j, y_j)^T \begin{pmatrix} x - x_j \\ y - y_j \end{pmatrix} \leq \theta, \\ \qquad\qquad g(x_j, y_j) + \nabla g(x_j, y_j)^T \begin{pmatrix} x - x_j \\ y - y_j \end{pmatrix} \leq 0, \\ \qquad\qquad x \in X, y \in Y \text{ discrete variable.} \end{cases} \quad \forall (x_j, y_j) \in K \tag{3.1}$$

When dealing with problem (P) in (1.1), the concept of subgradient is the substitute of the gradient in relaxation of (P). Note that arbitrary subgradients substituting gradients in (3.1) is not sufficient to equivalently reformulate problem (P) (see Example 3.1 below). As in [3,6], with the help of KKT conditions, we obtain several special subgradients, which we then use to reformulate problem (P) as one equivalent MILP master program such as (3.1).

3.1 An overview of the method

For the equivalent reformulation of problem (P) in (1.1) and by using techniques in (3.1), we appeal to the concept of projection for expressing problem (P) onto y variables. For any fixed $y \in Y$, we consider the following subproblem P^y:

$$P^y \begin{cases} \underset{x}{\text{minimize}} \ f(x, y) \\ \text{subject to } g(x, y) \leq 0, \\ \qquad\qquad x \in X. \end{cases} \tag{3.2}$$

If there exists some $x \in X$ such that $g(x, y) \leq 0$, the subproblem P^y is said to be feasible; otherwise, P^y is said to be infeasible.

For the validness of KKT conditions, we assume that the following Slater constraint qualification holds:

Assumption (A1) *For any $y \in Y$ satisfying that the subproblem P^y is feasible, the following Slater constraint qualification holds:*

$$g(\hat{x}, y) < 0 \ \textit{for some } \hat{x} \in X. \tag{Slater CQ}$$

Let

$$\Sigma := \{y \in Y : g(x, y) \leq 0 \quad \text{for some } x \in X\} \tag{3.3}$$

denote the set of all discrete variables y that produce feasible subproblems. Then the projection of problem (P) onto variable y can be given as follows:

$$\underset{y_j \in \Sigma}{\text{minimize}} \begin{cases} \underset{x}{\text{minimize}} \ f(x, y_j) \\ \text{subject to } g(x, y_j) \leq 0, \\ \qquad\qquad x \in X. \end{cases} \tag{3.4}$$

Now let $y_j \in \Sigma$ be fixed. Since X is compact and f, g_i are continuous, it follows that the optimal solution to subproblem P^{y_j} exists. Thus we can suppose that x_j is one optimal solution to P^{y_j}. By the assumption (A1) and KKT conditions, there exist $(\lambda_{j,1}, \ldots, \lambda_{j,m}) \in \mathbb{R}^m$ such that

$$\begin{cases} 0 \in \partial f(\cdot, y_j)(x_j) + \sum\limits_{i \in I(x_j)} \lambda_{j,i} \partial g_i(\cdot, y_j)(x_j) + N(X, x_j), \\ \lambda_{j,i} g_i(x_j, y_j) = 0, \quad i = 1, \ldots, m, \\ \lambda_{j,i} \geq 0, \quad i = 1, \ldots, m, \end{cases} \tag{3.5}$$

where

$$I(x_j) := \{i \in \{1, \ldots, m\} : g_i(x_j, y_j) = 0\} \tag{3.6}$$

is the active constraint set. This means that we can take $\alpha_j \in \partial f(\cdot, y_j)(x_j)$ and $\xi_{j,i} \in \partial g_i(\cdot, y_j)(x_j)(i = 1, \ldots, m)$ such that

$$-\alpha_j - \sum_{i \in I(x_j)} \lambda_{j,i} \xi_{j,i} \in N(X, x_j). \tag{3.7}$$

By Proposition 2.3, there exist $\beta_j \in \mathbb{R}^p$ and $\eta_{j,i} \in \mathbb{R}^p (i = 1, \ldots, m)$ such that

$$(\alpha_j, \beta_j) \in \partial f(x_j, y_j) \quad \text{and} \quad (\xi_{j,i}, \eta_{j,i}) \in \partial g_i(x_j, y_j), \forall i \in \{1, \ldots, m\}. \tag{3.8}$$

Set $\xi_j := (\xi_{j,1}, \ldots, \xi_{j,m})$ and $\eta_j := (\eta_{j,1}, \ldots, \eta_{j,m})$. We consider the following linear problem:

$$LP(x_j, y_j) \begin{cases} \underset{x}{\text{minimize}} \ f(x_j, y_j) + (\alpha_j, \beta_j)^T \begin{pmatrix} x - x_j \\ 0 \end{pmatrix} \\ \text{subject to } g(x_j, y_j) + (\xi_j, \eta_j)^T \begin{pmatrix} x - x_j \\ 0 \end{pmatrix} \leq 0, \\ \qquad x \in X. \end{cases} \tag{3.9}$$

The following theorem establishes the equivalence between subproblem P^{y_j} and linear program $LP(x_j, y_j)$ of (3.9). The proof of this theorem will be given in Sect. 5.

Theorem 3.1 *Let $LP(x_j, y_j)$ be defined as (3.9). Then x_j is one optimal solution for $LP(x_j, y_j)$ in (3.9) and $f(x_j, y_j)$ is the optimal value of $LP(x_j, y_j)$ in (3.9).*

We denote

$$T := \left\{ j : P^{y_j} \text{ is feasible and } x_j \text{ is an optimal solution to } P^{y_j} \right\}. \tag{3.10}$$

Let $j \in T$. By assumption (A1), we can take $(\lambda_{j,1}, \ldots, \lambda_{j,m}) \in \mathbb{R}^m_+$, $(\alpha_j, \beta_j) \in \partial f(x_j, y_j)$ and $(\xi_{j,i}, \eta_{j,i}) \in \partial g_i(x_j, y_j)$ $(i = 1, \ldots, m)$ such that (3.7) holds. Applying Proposition 2.3, there exist $\beta_j \in \mathbb{R}^p$ and $\eta_{j,i} \in \mathbb{R}^p$ $(i = 1, \ldots, m)$ such that (3.8) holds. Then we set

$$\xi_j := (\xi_{j,1}, \ldots, \xi_{j,m}) \quad \text{and} \quad \eta_j := (\eta_{j,1}, \ldots, \eta_{j,m}).$$

We consider the following MILP:

$$(M_\Sigma) \begin{cases} \underset{x, y, \theta}{\text{minimize}} \ \theta \\ \text{subject to } f(x_j, y_j) + (\alpha_j, \beta_j)^T \begin{pmatrix} x - x_j \\ y - y_j \end{pmatrix} \leq \theta \quad \forall j \in T, \\ \qquad g(x_j, y_j) + (\xi_j, \eta_j)^T \begin{pmatrix} x - x_j \\ y - y_j \end{pmatrix} \leq 0 \quad \forall j \in T, \\ \qquad x \in X, y \in \Sigma \text{ discrete variable.} \end{cases} \tag{3.11}$$

By virtue of Theorem 3.1, we obtain the following theorem on the equivalence of problem (P) of (1.1) and MILP (M_Σ) of (3.11).

Theorem 3.2 *Assmue that MINLP problem (P) of* (1.1) *satisfies assumption (A1). Then MILP* (M_Σ) *of* (3.11) *are equivalent to problem (P) in the sense that both have the same optimal value and that the optimal solution* $(\bar{x}, \bar{y})$ *to problem (P) corresponds to the optimal solution* $(\bar{x}, \bar{y}, \bar{\theta})$ *to* (M_Σ) *of* (3.11) *with* $\bar{\theta} = f(\bar{x}, \bar{y})$.

For completely reformulating the problem (P), it remains to provide an appropriate representation of constraint $y \in Y \backslash \Sigma$ by supporting hyperplanes. Along the lines in [2,8], we are inspired to study infeasible subproblems so as to eliminate those discrete variables that give rise to infeasibility.

Let $y_l \in Y \backslash \Sigma$. Then subproblem P^{y_l} is infeasible; that is,

$$\nexists x \in X \quad \text{satisfying } g_i(x, y_l) \leq 0 \quad \text{for all } i = 1, \ldots, m.$$

Let J_l be one subset of $\{1, \ldots, m\}$ such that there is some $\hat{x} \in X$ satisfying

$$g_i(\hat{x}, y_l) < 0, \quad \forall i \in J_l. \tag{3.12}$$

Denote $J_l^\perp := \{1, \ldots, m\} \backslash J_l$ the complement of J_l. To detect the infeasibility, we study the following subproblem F^{y_l}:

$$F^{y_l} \begin{cases} \underset{x}{\text{minimize}} \displaystyle\sum_{i \in J_l^\perp} [g_i(x, y_l)]_+ \\ \text{subject to } g_i(x, y_l) \leq 0 \quad \forall i \in J_l, \\ \qquad\qquad\quad x \in X, \end{cases} \tag{3.13}$$

where $[g_i(x, y_l)]_+ := \max\{g_i(x, y_l), 0\}$.

Since X is compact and g_i for $i = 1, \ldots, m$ are continuous, then the optimal solution to subproblem F^{y_l} exists. Thus we can assume that x_l is one optimal solution to subproblem F^{y_l}. For convenience to state the process, we divide the set $J_l^\perp$ into three disjoint subsets which are denoted by J_l^1, J_l^2 and J_l^3. These three subsets are defined as

$$\begin{cases} J_l^1 := \{i \in J_l^\perp : g_i(x_l, y_l) = 0\}, \\ J_l^2 := \{i \in J_l^\perp : g_i(x_l, y_l) > 0\}, \\ J_l^3 := \{i \in J_l^\perp : g_i(x_l, y_l) < 0\}. \end{cases} \tag{3.14}$$

This means that $J_l^\perp = J_l^1 \cup J_l^2 \cup J_l^3$ and by using continuity of g_i, one has

$$\partial[g_i(\cdot, y_l)]_+(x_l) = \partial g_i(\cdot, y_l)(x_l), \ \forall i \in J_l^2 \quad \text{and} \quad \partial[g_i(\cdot, y_l)]_+(x_l) = \{0\}, \ \forall i \in J_l^3.$$

By (3.12) and KKT condition, there exist $\lambda_{l,i} \in \mathbb{R}$ for all $i \in J_l$ such that

$$\begin{cases} 0 \in \displaystyle\sum_{i \in J_l^\perp} \partial[g_i(\cdot, y_l)]_+(x_l) + \sum_{i \in J_l} \lambda_{l,i} \partial g_i(\cdot, y_l)(x_l) + N(X, x_l), \\ \lambda_{l,i} g_i(x_l, y_l) = 0, \ \ \forall i \in J_l, \\ \lambda_{l,i} \geq 0, \ \ \forall i \in J_l. \end{cases} \tag{3.15}$$

Denote $\lambda_{l,i} \equiv 1$ for all $i \in J_l^2$ and $\lambda_{l,i} \equiv 0$ for all $i \in J_l^3$. Using Proposition 2.4, there exist $\lambda_{l,i} \in [0, 1] (\forall i \in J_l^1)$ and $\xi_{l,i} \in \partial g_i(\cdot, y_l)(x_l) (\forall i \in J_l^\perp \cup J_l)$ such that

$$-\sum_{i \in J_l^\perp \cup J_l} \lambda_{l,i} \xi_{l,i} \in N(X, x_l). \tag{3.16}$$

By virtue of Proposition 2.3, there exist $\eta_{l,i} \in \mathbb{R}^p$ such that $(\xi_{l,i}, \eta_{l,i}) \in \partial g_i(x_l, y_l)$ for all $i \in J_l^{\perp} \cup J_l$.

Since subproblem P^{y_l} is infeasible, then there exists one optimal solution x_l to subproblem F^{y_l} such that $\sum_{i \in J_l^{\perp}} [g_i(x_l, y_l)]_+ > 0$, by the continuity of g_i and compactness of X. This gives the following theorem on subproblem F^{y_l}. The proof is also given in Section 5.

Theorem 3.3 *The discrete variable $y_l \in Y \backslash \Sigma$ is infeasible to the following constraint:*

$$\begin{cases} g_i(x_l, y_l) + (\xi_{l,i}, \eta_{l,i})^T \begin{pmatrix} x - x_l \\ y - y_l \end{pmatrix} \leq 0, \quad \forall i \in J_l^{\perp} \cup J_l, \\ x \in X, y \in Y. \end{cases} \tag{3.17}$$

It is necessary to ensure that discrete variables that produce infeasible subproblems are also infeasible in the reformulated master program. We denote

$$S := \left\{ l : P^{y_l} \text{ is infeasible and } x_l \text{ solves } F^{y_l} \right\}. \tag{3.18}$$

For any $l \in S$, take $\lambda_{l,i} \geq 0$ and $\xi_{l,i} \in \partial g_i(\cdot, y_l)(x_l)(i = 1, \ldots, m)$ such that (3.16) holds. Take $\eta_{l,i} \in \mathbb{R}^p$ such that $(\xi_{l,i}, \eta_{l,i}) \in \partial g_i(x_l, y_l)$ for any $i \in \{1, \ldots, m\}$ by Proposition 2.3. We set $\xi_l := (\xi_{l,1}, \ldots, \xi_{l,m})$ and $\eta_l := (\eta_{l,1}, \ldots, \eta_{l,m})$. Then by using Theorem 3.3, we have the following theorem which shows how to eliminate those discrete variables giving rise to infeasible subproblems.

Theorem 3.4 *For any $l \in S$, let (ξ_l, η_l) be defined as above. Then the following constraints*

$$\begin{cases} g(x_l, y_l) + (\xi_l, \eta_l)^T \begin{pmatrix} x - x_l \\ y - y_l \end{pmatrix} \leq 0, \quad \forall l \in S, \\ x \in X, y \in Y \end{cases} \tag{3.19}$$

exclude all discrete variables $y_l \in Y$ for which subproblem P^{y_l} is infeasible.

It is known from Theorem 3.4 that we can add linearization from F^{y_l} when subproblem P^{y_l} is infeasible so as to correctly represent the constraints $y \in \Sigma$ in (3.3). This gives rise to the MILP master program (MP) which is equivalent to MINLP problem (P) in (1.1) and used to reformulate problem (P).

Let T and S be defined as (3.10) and (3.18), respectively. For any $j \in T$, by assumption (A1), we can take $\lambda_{j,i} \geq 0$ $(i = 1, \ldots, m)$, $\alpha_j \in \partial f(\cdot, y_j)(x_j)$ and $\xi_{j,i} \in \partial g_i(\cdot, y_j)(x_j)(i = 1, \ldots, m)$ such that (3.7) holds, and by Proposition 2.3, we take $\beta_j \in \mathbb{R}^p$ and $\eta_{j,i} \in \mathbb{R}^p(i = 1, \ldots, m)$ such that (3.8) holds. We set

$$\xi_j := (\xi_{j,1}, \ldots, \xi_{j,m}) \quad \text{and} \quad \eta_j := (\eta_{j,1}, \ldots, \eta_{j,m}).$$

For any $l \in S$, we take $\lambda_{l,i} \geq 0$ and $\xi_{l,i} \in \partial g_i(\cdot, y_l)(x_l)(i = 1, \ldots, m)$ such that (3.16) holds and by Proposition 2.3, we take $\eta_{l,i} \in \mathbb{R}^p$ such that $(\xi_{l,i}, \eta_{l,i}) \in \partial g_i(x_l, y_l)$. Set

$$\xi_l := (\xi_{l,1}, \ldots, \xi_{l,m}) \quad \text{and} \quad \eta_l := (\eta_{l,1}, \ldots, \eta_{l,m}).$$

The MILP master problem (MP) is given as follows:

$$
(\mathrm{MP})\begin{cases}\underset{x,y,\theta}{\text{minimize}}\ \theta\\ \text{subject to}\ f(x_j,y_j)+(\alpha_j,\beta_j)^T\begin{pmatrix}x-x_j\\ y-y_j\end{pmatrix}\le\theta\quad\forall j\in T,\\ \qquad g(x_j,y_j)+(\xi_j,\eta_j)^T\begin{pmatrix}x-x_j\\ y-y_j\end{pmatrix}\le 0\quad\forall j\in T,\\ \qquad g(x_l,y_l)+(\xi_l,\eta_l)^T\begin{pmatrix}x-x_l\\ y-y_l\end{pmatrix}\le 0\quad\forall l\in S,\\ \qquad x\in X,\ y\in Y\ \text{discrete variable.}\end{cases}\tag{3.20}
$$

The following theorem, immediate from Theorems 3.3 and 3.4, is one main result in the procedure of reformulating MINLP problem (P) of (1.1) as the equivalent MILP master program (MP).

Theorem 3.5 *Assume that MINLP problem (P) of* (1.1) *satisfies assumption (A1). Then master program (MP) of* (3.20) *is equivalent to problem (P) in the sense that both problems have the same optimal value and that the optimal solution* $(\bar{x},\bar{y})$ *to problem (P) corresponds to the optimal solution* $(\bar{x},\bar{y},\bar{\theta})$ *to (MP) of* (3.20) *with* $\bar{\theta}=f(\bar{x},\bar{y})$.

Remark 3.1 Theorem 3.5 is one extension of main results given in [8,23], and it generalizes the outer approximation method in the sense of equivalently reformulating convex MINLP problem (P) from differentiable case to the non-differentiable one. Further, it is known from Theorem 3.5 that all optimal solutions of problem (P) are optimal solutions to master program (MP). However, the converse is not necessarily true since some optimal solutions of (MP) may be infeasible to problem (P). We refer the reader to [2, Example 1] and [23, Remark3.1] for the detail.

Theorem 3.5 shows that some subgradients obtained from the KKT conditions enable to reformulate MINLP problem (P) as an equivalent MILP master program by outer approximation method. However, this procedure is not valid if arbitrary subgradients are chosen to replace gradients. The following example demonstrates that the substitution of the gradient by an arbitrary subgradient in the outer approximation method is insufficient for the equivalent reformulation.

Example 3.1 We consider the following convex MINLP problem:

$$
\begin{cases}\underset{x,y}{\text{minimize}}\ f(x,y):=x+y\\ \text{subject to}\ g_1(x,y):=\max\{-x+y+1,x-y+1\}\le 0,\\ \qquad g_2(x,y):=x-y\le 0,\\ \qquad x\in[0,2],\ \ y\in\{1,2,3\}.\end{cases}\tag{3.21}
$$

One can verify that this convex MINLP in (3.21) is infeasible. However, let the initial point $y_0=1$. Then subproblem P^{y_0} is infeasible and we can consider the following subproblem F^{y_0}:

$$
\begin{cases}\underset{x}{\text{minimize}}\ [g_1(x,y_0)]_+=g_1(x,y_0)\\ \text{subject to}\ g_2(x,y_0)\le 0,\\ \qquad x\in[0,2].\end{cases}\tag{3.22}
$$

It is easy to verify that $x_0=1$ is the optimal solution to subproblem F^{y_0} and

$$
[-1,1]\times[-1,1]\subset\partial g_1(x_0,y_0).
$$

Now, if we take $(\xi_{0,1}, \eta_{0,1}) = (1, 1) \in \partial g_1(x_0, y_0)$, $(\xi_{0,2}, \eta_{0,2}) = \nabla g_2(x_0, y_0)$ and $(\alpha_0, \beta_0) = \nabla f(x_0, y_0)$, then the MILP is defined as

$$\begin{cases} \underset{x,\, y,\, \theta}{\text{minimize}}\ \theta \\ \text{subject to } x + y \le \theta, \\ \qquad\qquad x + y - 1 \le 0, \\ \qquad\qquad x - y \le 0, \\ \qquad\qquad x \in [0, 2],\ \ y \in \{1, 2, 3\}. \end{cases} \tag{3.23}$$

The optimal solution to MILP in (3.23) is $(x, y, \theta) = (0, 1, 1)$. This means that the outer approximation method for this MINLP may generate an infinite loop between points (x_0, y_0) and $(0, 1)$. Thus the outer approximation method is invalid for this MINLP problem in (3.21). Further, when tracking down why this method is not valid here, the reason noticed is that the KKT conditions at (x_0, y_0) for $(\xi_{0,1}, \eta_{0,1})$ does not hold; that is,

$$\nexists(\lambda_{0,1}, \lambda_{0,2}) \in \mathbb{R}^2_+ \ \text{ satisfying } \ \nabla f(x_0, y_0) + \lambda_{0,1}(\xi_{0,1}, \eta_{0,1}) + \lambda_{0,2}\nabla g_2(x_0, y_0) = 0.$$

3.2 The algorithm

In this subsection, based on the solution of MILP master program (MP) in (3.20), we pay main attention to one outer approximation algorithm for finding the optimal solution of problem (P) in (1.1) along the line in [5,8,23].

At iteration k, the sets T and S in master program (MP) of (3.20) are substituted by the sets T^k and S^k respectively which are defined as

$$\begin{cases} T^k := \{j \le k : P^{y_j} \text{ is feasible and } x_j \text{ solves } \ P^{y_j}\}, \\ S^k := \{l \le k : P^{y_l} \text{ is infeasible and } x_l \text{ solves } \ F^{y_l}\}. \end{cases} \tag{3.24}$$

If $k \in T^k$ then x_k solves P^{y_k} and there exist $(\alpha_k, \beta_k) \in \mathbb{R}^n \times \mathbb{R}^p$ and $(\xi_{k,i}, \eta_{k,i}) \in \mathbb{R}^n \times \mathbb{R}^p$ for all $i = 1, \ldots, m$ such that

$$\begin{cases} -\alpha_k - \sum\limits_{i=1}^{m} \lambda_{k,i}\xi_{k,i} \in N(X, x_k) \ \text{ for some } (\lambda_{k,1} \ldots, \lambda_{k,m}) \in \mathbb{R}^m_+, \\ (\alpha_k, \beta_k) \in \partial f(x_k, y_k), \\ (\xi_{k,i}, \eta_{k,i}) \in \partial g_i(x_k, y_k),\ \forall i = 1, \ldots, m. \end{cases} \tag{3.25}$$

If $k \in S^k$ then x_k solves F^{y_k} and there exist $(\xi_{k,i}, \eta_{k,i}) \in \mathbb{R}^n \times \mathbb{R}^p$ for all $i = 1, \ldots, m$ such that

$$\begin{cases} -\sum\limits_{i=1}^{m} \lambda_{k,i}\xi_{k,i} \in N(X, x_k) \ \text{ for some } (\lambda_{k,1} \ldots, \lambda_{k,m}) \in \mathbb{R}^m_+, \\ (\xi_{k,i}, \eta_{k,i}) \in \partial g_i(x_k, y_k),\ \forall i = 1, \ldots, m. \end{cases} \tag{3.26}$$

Set

$$\xi_k := (\xi_{k,1}, \ldots, \xi_{k,m}) \quad \text{and} \quad \eta_k := (\eta_{k,1}, \ldots, \eta_{k,m}).$$

To prevent discrete variable assignment y_j (for any $j \in T^k$) from being the solution to the relaxed master program, it is necessary to define $UBD^k := \min\{f(x_j, y_j) : j \in T^k\}$ and add a constraint $\theta < UBD^k$ to the master program. This gives rise to the following relaxed

master program MP^k:

$$MP^k \begin{cases} \underset{x,\,y,\,\theta}{\text{minimize}}\ \theta \\ \text{subject to } \theta < UBD^k \\ \qquad f(x_j, y_j) + (\alpha_j, \beta_j)^T \begin{pmatrix} x - x_j \\ y - y_j \end{pmatrix} \le \theta \quad \forall j \in T^k, \\ \qquad g(x_j, y_j) + (\xi_j, \eta_j)^T \begin{pmatrix} x - x_j \\ y - y_j \end{pmatrix} \le 0 \quad \forall j \in T^k, \\ \qquad g(x_l, y_l) + (\xi_l, \eta_l)^T \begin{pmatrix} x - x_l \\ y - y_l \end{pmatrix} \le 0 \quad \forall l \in S^k, \\ \qquad x \in X,\ y \in Y \text{ discrete variable.} \end{cases} \tag{3.27}$$

The new discrete variable assignment y_{k+1} can be obtain by solving MP^k and the whole process is repeated iteratively until the relaxed master program is infeasible.

We are now in a position to state the outer approximation algorithm for solving problem (P) in detail.

Algorithm 1 (Outer Approximation Algorithm)

1: Initialization. Given an initial $y_0 \in Y$, set $T^0 = S^0 := \emptyset$, $UBD^0 := \infty$ and let $k := 1$
2: **for** $k = 1, 2, \ldots,$ **do**
3: Solve subproblem P^{y_k}
4: **if** P^{y_k} is feasible **then**
5: Choose one solution x_k of P^{y_k}
Choose (α_k, β_k) and $\xi_k := (\xi_{k,1}, \ldots, \xi_{k,m})$, $\eta_k := (\eta_{k,1}, \ldots, \eta_{k,m})$ as in (3.25)
Set $T^k := T^{k-1} \cup \{k\}$, $S^k := S^{k-1}$ and $UBD^k := \min\{UBD^{k-1}, f(x_k, y_k)\}$
6: **else**
7: Solve subproblem F^{y_k} and choose one solution x_k of F^{y_k}
Choose $\xi_k := (\xi_{k,1}, \ldots, \xi_{k,m})$, $\eta_k := (\eta_{k,1}, \ldots, \eta_{k,m})$ as in (3.26)
Set $S^k := S^{k-1} \cup \{k\}$, $T^k := T^{k-1}$ and $UBD^k := UBD^{k-1}$
8: **end if**
9: Solve the relaxation MP^k and obtain a new discrete variable y_{k+1}
Set $k := k + 1$ and go back to line 3
10: **end for**

Under the assumption of finite cardinality of discrete variable subset Y, the following theorem shows Algorithm 1 can detect feasibility or infeasibility of problem (P) in (1.1) and the procedure in Algorithm 1 terminates after a finite number of steps. The proof is also given in Sect. 5.

Theorem 3.6 *Suppose that MINLP problem (P) in* (1.1) *satisfies assumption (A1) and the cardinality of Y is finite. Then either problem (P) is infeasible or Algorithm 1 terminates in a finite number of steps at an optimal value of problem (P).*

4 Conclusions

This paper is mainly devoted to the study of one convex MINLP in which objective and constraint functions are continuous and non-differentiable. With no differentiability assumption, subgradients of objective and constraint functions, the substitute of gradients in convex and smooth MINLP, are chosen from the KKT conditions and used to reformulate the MINLP

problem as one equivalent mixed-integer linear program. A counterexample shows that the chosen subgradients, if not satisfying KKT conditions, may be invalid for the MILP reformulation, which demonstrates the necessity of KKT conditions in the equivalent reformulation. By solving a finite sequence of subproblems and relaxed MILP problems, one outer approximation algorithm for this convex MINLP is presented to find the optimal solution of the problem. The finite convergence of the algorithm is also proved. The work of this paper is the extension of references [5,8,23] and also generalizes outer approximation method in the sense of dealing with convex MINLP from differentiable case to the non-differentiable one.

Acknowledgments We are grateful to the referee for careful reading this paper and valuable comments which help us to improve the original version. This research was supported by the National Natural Science Foundations of P. R. China (Grant No. 11401518 and No. 11261067) and IRTSTYN, and by the Claude Leon Foundation of South Africa.

5 Appendix: Proofs of Theorems 3.1, 3.3 and 3.6

In this section, we present the proofs of several key results in Section 3.

Proof of Theorem 3.1 To prove Theorem 3.1, it suffices to show that

$$(\alpha_j, \beta_j)^T \begin{pmatrix} x - x_j \\ 0 \end{pmatrix} \geq 0, \quad \forall x \in X \text{ with } g(x_j, y_j) + (\xi_j, \eta_j)^T \begin{pmatrix} x - x_j \\ 0 \end{pmatrix} \leq 0. \tag{5.1}$$

Let $x \in X$ be such that

$$g(x_j, y_j) + (\xi_j, \eta_j)^T \begin{pmatrix} x - x_j \\ 0 \end{pmatrix} \leq 0$$

and let $I(x_j)$ be defined as (3.6). Then

$$\langle \xi_{j,i}, x - x_j \rangle \leq 0, \forall i \in I(x_j). \tag{5.2}$$

By using (3.7), there exists $\gamma \in N(X, x_j)$ such that

$$\alpha_j + \sum_{i \in I(x_j)} \lambda_{j,i} \xi_{j,i} + \gamma = 0. \tag{5.3}$$

Noting that X is convex, it follows that $x - x_j \in T(X, x_j)$. This together with (5.2) and (5.3) implies that

$$(\alpha_j, \beta_j)^T \begin{pmatrix} x - x_j \\ 0 \end{pmatrix} = \langle \alpha_j, x - x_j \rangle = -\Big\langle \sum_{i \in I(x_j)} \lambda_{j,i} \xi_{j,i} + \gamma, x - x_j \Big\rangle \geq 0.$$

Hence (5.1) holds. The proof is complete. □

Proof of Theorem 3.3 Since X is compact and g is continuous, then one has

$$\sum_{i \in J_l^{\perp}} [g_i(x_l, y_l)]_+ > 0. \tag{5.4}$$

Suppose to the contrary that there exists $\hat{x} \in X$ such that $(\hat{x}, y_l)$ is feasible to the constraint of (3.17). Then

$$g_i(x_l, y_l) + \langle \xi_{l,i}, \hat{x} - x_l \rangle \leq 0, \quad \forall i \in J_l^{\perp} \cup J_l. \tag{5.5}$$

Noting that $\hat{x} - x_l \in T(X, x_l)$ by the convexity of X, it follows from (3.16) that there exists $\gamma \in N(X, x_l)$ such that

$$\sum_{i \in J_l^1 \cup J_l^2 \cup J_l^3} \lambda_{l,i} \xi_{l,i} + \gamma = 0, \tag{5.6}$$

where $\lambda_{l,i} \equiv 1$ for all $i \in J_l^2$ and $\lambda_{l,i} \equiv 0$ for all $i \in J_l^3$. By multiplying (5.5) by $\lambda_{l,i}$ for any $i \in J_l^1 \cup J_l^2 \cup J_l^3$, it follows from (5.6) that

$$\begin{aligned} 0 &\geq \sum_{i \in J_l^1 \cup J_l^2} \lambda_{l,i} g_i(x_l, y_l) + \Big\langle \sum_{i \in J_l^1 \cup J_l^2 \cup J_l} \lambda_{l,i} \xi_{l,i} + \gamma, \hat{x} - x_l \Big\rangle \\ &\geq \sum_{i \in J_l^2} g_i(x_l, y_l) = \sum_{i \in J_l^{\perp}} [g_i(x_l, y_l)]_+ \end{aligned}$$

as $\hat{x} - x_l \in T(X, x_l)$ and $\lambda_{l,i} g_i(x_l, y_l) = 0 (\forall i \in J_l)$, which contradicts (5.4). The proof is complete. □

Proof of Theorem 3.6 (This is similar to the proof for [23, Theorem 3.6]. For the sake of completeness, we provide the proof in brief.)

For the proof of Theorem 3.6, we first prove that there is no discrete variable in Y generated more than once by Algorithm 1. Granting this, it follows from the finite cardinality of Y that the termination of Algorithm 1 holds after a finite number steps.

At iteration k, let $(\hat{x}, \hat{y}, \hat{\theta})$ be an optimal solution to the relaxed master program MP^k. By virtue of Theorem 3.4, one can verify that $\hat{y} \neq y_l$ for all $l \in S^k$. Suppose to the contrary that $\hat{y} = y_{j_k}$ for some $j_k \in T^k$. Then $(\hat{x}, y_{j_k}, \hat{\theta})$ solves the relaxed master program MP^k and

$$\begin{cases} \hat{\theta} < UBD^k \leq f(x_{j_k}, y_{j_k}), \\ f(x_{j_k}, y_{j_k}) + (\alpha_{j_k}, \beta_{j_k})^T \begin{pmatrix} \hat{x} - x_{j_k} \\ 0 \end{pmatrix} \leq \hat{\theta}, \\ g(x_{j_k}, y_{j_k}) + (\xi_{j_k}, \eta_{j_k})^T \begin{pmatrix} \hat{x} - x_{j_k} \\ 0 \end{pmatrix} \leq 0. \end{cases} \tag{5.7}$$

By using (5.1) in the proof of Theorem 3.1, one has

$$(\alpha_{j_k}, \beta_{j_k})^T \begin{pmatrix} \hat{x} - x_l \\ 0 \end{pmatrix} \geq 0.$$

This and (5.7) imply that $f(x_{j_k}, y_{j_k}) \leq \hat{\theta}$, which contradicts $\hat{\theta} < f(x_{j_k}, y_{j_k})$ in (5.7). Hence $\hat{y} \neq y_j$ for all $j \in T_k$. This means that $\hat{y}$ is distinct from any y_j for all $j \in T^k \cup S^k$.

Now suppose that the relaxed master program MP^k is infeasible for some k. Then Algorithm 1 terminate at k-th step. Let ρ denote the optimal value of MINLP problem (P). If there is some $j \in T^{k-1}$ such that $f(x_j, y_j) = \rho$, then the conclusion holds. Next, we assume that $f(x_j, y_j) > \rho$ for all $j \in T^{k-1}$. Then $UBD^{k-1} > \rho$ and $k \in T^k$ (otherwise, $k \in S^k$, $UBD^k = UBD^{k-1}$ by Algorithm 1 and consequently MP^k is feasible, a contradiction). Thus subproblem P^{y_k} is feasible and $f(x_k, y_k) \geq \rho$.

Suppose to the contrary that $f(x_k, y_k) > \rho$. If $f(x_k, y_k) \leq UBD^{k-1}$, then $\rho < f(x_k, y_k) = UBD^k$ and MP^k is feasible, a contradiction. If $f(x_k, y_k) > UBD^{k-1}$ then $\rho < UBD^{k-1} = UBD^k$ and thus MP^k is feasible, a contradiction. This means $f(x_k, y_k) = \rho$. The proof is complete. □

References

1. Aubin, J.P., Frankowska, H.: Set-valued Analysis. Birkhäuser, Boston (1990)
2. Bonami, P., Biegler, L., Conn, A.R., Cornuéjols, G., Grossmann, I.E., Laird, C., Lee, J., Lodi, A., Margot, F., Sawaya, N., Wächter, A.: An algorithmic framework for convex mixed integer nonlinear programs. Discrete. Optim. **5**(2), 186–204 (2008)
3. Castillo, I., Westerlund, J., Emet, S., Westerlund, T.: Optimization of block layout design problems with unequal areas: a comparison of MILP and MINLP optimization methods. Comput. Chem. Eng. **30**, 54–69 (2005)
4. Drewes, S., Ulbrich, S.: Subgradient based outer approximation for mixed integer second order cone pogramming. Mixed Integer Nonlinear Prog. IMA Vol. Math. Appl. **154**, 41–59 (2012)
5. Duran, M., Grossmann, I.E.: An Outer approximation algorithm for a class of mixed-integer nonlinear programs. Math. Program. **36**, 307–339 (1986)
6. Eronen, V.-P., Mäkelä, M.M., Westerlund, T.: On the generalization of ECP and OA methods to nonsmooth convex MINLP problems. Optimization **63**, 1057–1073 (2014)
7. Eronen, V.-P., Mäkelä, M.M., Westerlund, T.: Extended cutting plane method for a class of nonsmooth nonconvex MINLP problems. Optimization (2013). doi:10.1080/02331934.2013.796473
8. Fletcher, R., Leyffer, S.: Solving mixed-integer nonlinear programs by outer approximation. Math. Prog. **66**, 327–349 (1994)
9. Decomposition in general mathematical programming: Flippo, O.E., Rinnooy Kan, A.H.G. Math. Prog. **60**, 361–382 (1993)
10. Floudas, C.A.: Nonlinear and Mixed Integer Optimization: Fundamentals and Applications. Oxford University Press, New York (1995)
11. Geoffrion, A.M.: Generalized Benders decomposition. J. Optim. Theory. Appl. **10**(4), 237–260 (1972)
12. Grossmann, I.E.: Review of nonlinear mixed-integer and disjunctive programming techniques. Optim. Eng. **3**, 227–252 (2002)
13. Grossmann, I. E., Sahinidis, N. V. (eds): Special Issue on Mixed-Integer Programming and Its Application to Engineering, Part I, Optim. Eng., vol. 3(4), Kluwer Academic Publishers, Netherlands (2002)
14. Grossmann, I. E., Sahinidis, N. V. (eds): Special Issue on Mixed-Integer Programming and Its Application to Engineering, Part II, Optim. Eng., vol. 4(1), Kluwer Academic Publishers, Netherlands (2002)
15. Leyffer, S.: Integrating SQP and branch-and-bound for mixed integer nonlinear programming. Comput. Optim. Appl. **18**, 295–309 (2001)
16. Liberti, L., Pantelides, C.: An exact reformulation algorithm for large nonconvex NLPs involving bilinear terms. J. Global. Optim. **36**, 161–189 (2006)
17. Linderoth, J.T., Ralphs, T.K.: Noncommercial software for mixed-integer linear programming. In: Karlof, J. (ed.) Integer Programming: Theory and Practice, Operations Research Series, pp. 253–303. CRC Press, Boca Raton (2005)
18. Michelon, P., Maculan, N.: Lagrangean decomposition for integer nonlinear programming with linear constraints. Math. Program. **52**, 303–313 (1991)
19. Nowak, I., Vigerske, S.: LaGO: a (heuristic) branch and cut algorithm for nonconvex MINLPs. Cent. Eur. J. Oper. Res **16**(2), 127–138 (2008)
20. Phelps, R.R.: Convex functions, Monotone Operators and Differentiability. Lecture Notes in Math, vol. 1364. Springer, New York (1989)
21. Tawarmalani, M., Sahinidis, N.V.: Global optimization of mixed-integer nonlinear programs: A theoretical and computational study. Math. Program. **99**, 563–591 (2004)
22. Tawarmalani, M., Sahinidis, N.V.: Convexification and Global Optimization in Continuous and Mixed-Integer Nonlinear Programming: Theory, Algorithms, Software, and Applications. Kluwer Academic Publishers, Netherlands (2002)
23. Wei, Z., Ali, M.M.: Outer approximation algorithm for one class of convex mixed-integer nonlinear programming problems with partial differentiability. J. Optim. Theory. Appl. doi:10.1007/s10957-015-0715-y
24. Westerlund, T., Pettersson, F.: An extended cutting plane method for solving convex MINLP problems. Computer. Chem. Eng. **19**, 131–136 (1995)
25. Westerlund, T., Pörn, R.: Solving pseudo-convex mixed integer optimization problems by cutting plane techniques. Optim. Eng. **3**, 253–280 (2002)
26. Zălinescu, C.: Convex Analysis in General Vector Spaces. World Scientific, Singapore (2002)