Ville-Pekka Eronen | Marko M. Mäkelä | Tapio Westerlund

Nonsmooth Extended Cutting Plane Method for Generally Convex MINLP Problems

Turku Centre *for* Computer Science

TUCS Technical Report
No 1055, July 2012

Nonsmooth Extended Cutting Plane Method for Generally Convex MINLP Problems

Ville-Pekka Eronen
University of Turku, Department of Mathematics and Statistics
FI-20014 Turku, Finland
vpoero@utu.fi

Marko M. Mäkelä
University of Turku, Department of Mathematics and Statistics
FI-20014 Turku, Finland
makela@utu.fi

Tapio Westerlund
Åbo Akademi University, Process Design and Systems Engineering Laboratory
Biskopsgatan 8, FI-20500 Åbo, Finland
twesterl@abo.fi

TUCS Technical Report
No 1055, July 2012

Abstract

In this article a generalization of the αECP algorithm to cover nondifferentiable Mixed-Integer NonLinear Programming (MINLP) problems is studied. In the generalization constraint functions are required to be f°-pseudoconvex instead of pseudoconvex functions. This enables the functions to be nonsmooth. The objective function is first assumed to be linear but also f°-pseudoconvex case is considered. Furthermore, the gradients used in the αECP algorithm are replaced by the subgradients of Clarke subdifferential. With some additional assumptions the resulting algorithm shall be proven to converge to a global minimum.

Keywords: Nonsmooth MINLP; Mixed-integer programming; Nonsmooth optimization; Extended cutting plane algorithm; αECP; Subgradient; Pseudoconvex function; Generalized convexity

TUCS Laboratory
TOpGroup

1 Introduction

Practical optimization applications often include both discrete and continuous variables with a nonlinear objective function. The αECP is an algorithm designed for solving this kind of Mixed-Integer NonLinear Programming (MINLP) problems with smooth (i.e. continuously differentiable) objective and constraint functions [16]. The basic idea is to relax nonlinear functions and solve only MILP subproblems. The nonlinearities are taken into account by adding cutting planes created from violated constraints. In each step a cutting plane is created or updated. Ideally the cutting planes cut off the infeasible points while leaving the feasible points in the problem.

The αECP is a generalization of the ECP method in which the cutting planes are not updated. The ECP method finds a global minimum of an MINLP problem if the objective and constraint functions are convex and continuously differentiable [15], whereas the αECP finds a global minimum if the objective and the constraint functions are pseudoconvex [16].

Other algorithms for solving MINLP problems include Branch-and-Bound (see e.g. [6, 12]) and Outer Approximation (see e.g. [2, 5, 17]) type methods. In these methods a NonLinear Programming (NLP) subproblem is solved at each iteration. This is different from αECP where only MILP subproblems are solved. Consequently, the αECP usually performs well compared to others if the problem includes nonlinear functions whose evaluation is time-costly. An example of such situation is a chromatographic separation problem [3].

There are relatively few articles about nonsmooth MINLP problems. In [5] a certain class of nonsmooth functions were addressed with OA method and penalty function formulation. In [4] the ECP method was generalized for nonsmooth convex problems. Furthermore, it was shown in [4] that OA method cannot be generalized to nonsmooth functions by simply substituting the gradients by subgradients.

In this article we shall generalize the αECP method to make it suitable for nonsmooth nonconvex MINLP problems. Similarly to the generalization of the ECP method [4] the main trick is to use a subgradient from the Clarke subdifferential instead of a gradient. This trick is a common way to generalize algorithms to cover nonsmooth problems being exemplified by the subgradient methods [13]. Under some theoretical conditions, the resulting algorithm will first be proven to converge to a global minimum if the object function is linear and the constraint functions are *f°-pseudoconvex*. The f°-pseudoconvexity is a natural generalization of pseudoconvexity to nonsmooth case in the sense that a continuously differentiable f°-pseudoconvex function is pseudoconvex [11]. The algorithm will be further generalized to the problems with f°-pseudoconvex objective functions.

The paper is organized as follows. Section 2 is devoted to the basic results that are needed later. In Section 3 the generalization of the αECP method is derived and shown to converge to a global minimum if the objective function is linear. In Section 4 the algorithm is further generalized to problems with f°-pseudoconvex objective function. In Section 5 the algorithm is applied to three example problems. Section 6 summarizes the results.

2 Preliminaries

First of all we present some preliminaries from nonsmooth analysis.

Definition 2.1. The function $f : \mathbb{R}^n \to \mathbb{R}$ is *upper semicontinuous* at x if for every sequence (x^i) converging to x we have

$$\limsup_{i\to\infty} f(x^i) \leq f(x).$$

Definition 2.2. A function $f : \mathbb{R}^n \to \mathbb{R}$ is *locally Lipschitz continuous* at a point $x \in \mathbb{R}^n$ if there exist scalars $K > 0$ and $\delta > 0$ such that

$$|f(y) - f(z)| \leq K\|y - z\| \quad \text{for all } y, z \in B(x;\delta), \tag{1}$$

where $B(x;\delta) \subset \mathbb{R}^n$ is an open ball with center x and radius δ.

Function f is called locally Lipschitz continuous if it is locally Lipschitz continuous at every $x \in \mathbb{R}^n$. Note that continuously differentiable or convex functions are always locally Lipschitz continuous [1].

Definition 2.3. [1] Let function f be locally Lipschitz continuous at $x \in \mathbb{R}^n$. The *Clarke generalized directional derivative* of f at point x to direction $d \in \mathbb{R}^n$ is

$$f^\circ(x;d) = \limsup_{\substack{y\to x\\ t\to 0^+}} \frac{f(y+td) - f(y)}{t}. \tag{2}$$

With the generalized directional derivative we can generalize the ordinary gradient.

DEFINITION 2.4. [1] The *Clarke subdifferential* of locally Lipschitz continuous function f at point x is the set

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

A vector $\xi \in \partial f(x)$ is called a *subgradient* of function f at point x.

Next, we present some useful results concerning generalized directional derivative and subdifferential.

THEOREM 2.5. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz continuous at $x \in \mathbb{R}^n$ with constant K. Then*

(i) *$\partial f(x)$ is a nonempty, convex and compact set such that $\partial f(x) \subset \operatorname{cl} B(0; K)$.*

(ii) *$f^\circ(x; d)$ is upper semicontinuous as a function of $(x; d)$.*

Proof. See [10] pages 30 and 32. □

From (i) we see that $\|\xi(x)\| \leq K$ for any $\xi(x) \in \partial f(x)$. If f is continuously differentiable then $\partial f(x) = \{\nabla f(x)\}$ and $f'(x; d) = f^\circ(x; d)$, where f' is the ordinary directional derivative. In general, if the equality $f'(x; d) = f^\circ(x; d)$ holds for every $d \in \mathbb{R}$, then the f is said to be *subdifferentially regular* at point x. If f is subdifferentially regular at every $x \in \mathbb{R}^n$ then it is said to be subdifferentially regular. With the aid of this property we may calculate the subdifferential of maximum of a finite number of locally Lipschitz continuous functions quite easily.

THEOREM 2.6. *Let $f_i : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz continuous at x for all $i = 1, \ldots, m$. Then the function*

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

is locally Lipschitz continuous at x and

$$\partial f(x) \subset \operatorname{conv} \{\partial f_i(x) \mid f_i(x) = f(x),\ i = 1, \ldots, m\}. \tag{3}$$

In addition, if f_i is subdifferentially regular at x for all $i = 1, \ldots, m$, then f is also subdifferentially regular at x and equality holds in (3).

Proof. See [1] page 47. □

If in Theorem 2.6 $f_i(x) = f(x)$ then we say that f_i is *active* at point x.

Next, we will consider a generalization of convexity. It is well known that a continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is *pseudoconvex* if for any $x, y \in \mathbb{R}^n$ the inequality $f(y) < f(x)$ implies $\nabla f(x)^T(y - x) < 0$. Next, we will present an analogous concept for possible nonsmooth locally Lipschitz continuous functions.

Definition 2.7. Let $S \subset \mathbb{R}^n$ be a convex set. A locally Lipschitz continuous function $f : S \to \mathbb{R}$ is *f°-pseudoconvex* if for any $x, y \in S$

$$f(y) < f(x) \quad \text{implies} \quad f^\circ(x; y - x) < 0.$$

The concept of f°-pseudoconvexity has been presented in e.g. [8]. If f is continuously differentiable then $f^\circ(x; y - x) = \nabla f(x)^T(y - x)$ and it is easy to see that f is pseudoconvex if it is f°-pseudoconvex. Furthermore, it can be shown that convex function is f°-pseudoconvex function.

Theorem 2.8. *If function $f : \mathbb{R}^n \to \mathbb{R}$ is convex or pseudoconvex then it is also f°-pseudoconvex.*

Proof. See [11] page 6. □

The maximum of a finite number of pseudoconvex functions is not necessarily pseudoconvex. This can be seen easily by noting that the maximum function is not always differentiable everywhere. However, for f°-pseudoconvex functions this is not a problem and it turns out that the maximum of a finite number of f°-pseudoconvex functions is f°-pseudoconvex.

Theorem 2.9. *Let $f_i : \mathbb{R}^n \to \mathbb{R}$ be f°-pseudoconvex for all $i = 1, \ldots, m$. Then the function*

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

is f°-pseudoconvex.

Proof. See [11] page 12. □

Note that the maximum function of pseudoconvex functions f_i, $i = 1, \ldots, m$ is always f°-pseudoconvex due to Theorems 2.8 and 2.9. Since a pseudoconvex function is always subdifferentially regular, a gradient of an active function f_i at x belongs to the subdifferential $\partial f(x)$ by Theorem 2.6. This fact will be used frequently in the numerical examples.

The following lemma will turn out useful.

LEMMA 2.10. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be f°-pseudoconvex function. Let $A, C \subset \mathbb{R}^n$ be nonempty compact sets such that there exists $a \in \mathbb{R}$ with $f(z) \geq a$ for all $z \in A$ and $f(w) < a$ for all $w \in C$. Then, there exists $\delta > 0$ such that*

$$\sup_{\substack{z \in A \\ \xi \in \partial f(z) \\ w \in C}} \xi^T(w - z) = -\delta.$$

Proof. Suppose that $f(z) \geq a > f(w)$ for all $z \in A$ and $w \in C$. The definition of subdifferential and f°-pseudoconvexity of f implies that

$$\xi^T(w - z) \leq f^\circ(z; w - z) < 0 \tag{4}$$

for all $z \in A$, $\xi \in \partial f(z)$ and $w \in C$. Then, we have

$$\sup_{\substack{z \in A \\ \xi \in \partial f(z) \\ w \in C}} \xi^T(w - z) \leq 0.$$

On contrary to the lemma, assume that there exists no $\delta > 0$ defined above. Thus, we have sequences $(z^i) \subset A$, $(w^i) \subset C$ and $(\xi(z^i))$ such that $\xi(z^i) \in \partial f(z^i)$ and $\xi(z^i)^T(w^i - z^i) \to 0$ when $i \to \infty$. Since the sets A and C are compact sequences (z^i) and (w^i) can be chosen to be converging sequences. Denote $\hat{z} = \lim_{i\to\infty} z^i$ and $\hat{w} = \lim_{i\to\infty} w^i$. Then $\hat{z} \in A$ and $\hat{w} \in C$ since the sets A and C are closed. Thus, we have $f(\hat{w}) < a \leq f(\hat{z})$. By the first inequality of (4) we have

$$0 = \lim_{i\to\infty} \xi(z^i)^T(w^i - z^i) \leq \limsup_{i\to\infty} f^\circ(z^i; w^i - z^i).$$

The upper semicontinuity of f° leads to

$$0 \leq \limsup_{i\to\infty} f^\circ(z^i; w^i - z^i) \leq f^\circ(\hat{z}; \hat{w} - \hat{z}).$$

This contradicts with the inequality (4) and, thus, the lemma is proven. □

Note that A or C may be a singleton.

3 The generalization of the αECP-algorithm

We are considering MINLP problems of type

$$\begin{aligned} \min \quad & f(z) \\ \text{s.t.} \quad & g_j(z) \leq 0 \quad j = 1, \ldots, J \quad (GP) \\ & z \in L, \end{aligned}$$

where $z = (x, y)$, $x \in \mathbb{R}^n$ and $y \in \mathbb{Z}^m$. The set $L = X \times Y$, where $X \subset \mathbb{R}^n$ is supposed to be a compact convex polytope and $Y \subset \mathbb{Z}^m$ is finite. The constraint functions g_j are supposed to be f°-pseudoconvex. Denote

$$\hat{g}(z) = \max_{j=1,\ldots,J} \{g_j(z)\} \text{ and } N = \{z \mid \hat{g}(z) \leq 0\}.$$

Then, the feasible set of problem (GP) can be written as $N \cap L$. Since L is compact and N is closed the feasible set $N \cap L$ is compact.

Requirements for the function f depends on section. In this section we assume that the objective function of the problem (GP) is linear. That is, we consider the problem

$$\begin{aligned} \min \quad & c^T z \\ \text{s.t.} \quad & z \in N \cap L. \quad (P) \end{aligned}$$

From Theorem 2.8 we deduce that this formulation includes the problem considered earlier in [14], where the constraints were pseudoconvex functions.

3.1 The generalized αECP algorithm

Our Nonsmooth α Extended Cutting Plane method (NαECP) solves the problem (P) like αECP [14], but now the subgradients are used instead of gradients. As in the original αECP method the nonlinear constraints $g_j(z) \leq 0$, $j = 1, \ldots, J$ are first omitted from the problem. This results in problem

$$\begin{aligned} \min \quad & c^T z \\ \text{s.t.} \quad & z \in L. \quad (P^0) \end{aligned}$$

This is a MILP problem and it can be solved by any decent MILP solver. Let the solution to the problem (P^0) be z^0. If $\hat{g}(z^0) \leq 0$ then z^0 is a global minimum of problem (P) since $z^0 \in N \cap L$ and it is a global minimum in $L \supset N \cap L$. If $\hat{g}(z^0) > 0$ then we create a cutting plane

$$l^0_{j_0}(z) = g_{j_0}(z^0) + \alpha^0_{j_0 0}\xi_{j_0}(z^0)^T(z - z^0) \leq 0, \tag{5}$$

where g_{j_0} is one of the most violated constraints, that is, $\hat{g}(z^0) = g_{j_0}(z^0)$. Furthermore, $\xi_{j_0}(z^0) \in \partial g_{j_0}(z^0)$ and $\alpha^0_{j_0 0}$ is set at first to 1. The superscript of α denotes the iteration at which it was introduced. The first subscript denotes the index of constraint and the second subscript indicates how many times it has been updated. The second subscript is needed in the algorithm but we usually omit it from theoretical considerations. The cutting plane (5) will be added to the problem (P^0) resulting in problem (P^1)

$$\begin{array}{ll} \min & c^T z \\ \text{s.t.} & l^0_{j_0}(z) \leq 0 \quad (P^1) \\ & z \in L, \end{array}$$

which is again a MILP problem. Similarly, at iteration i we will solve the MILP problem

$$\begin{array}{ll} \min & c^T z \\ \text{s.t.} & l^k_{j_k}(z) \leq 0 \quad k = 0, 1, \ldots, i-1 \quad (P^i) \\ & z \in L. \end{array}$$

If the solution z^i does not satisfy the nonlinear constraints, then we will add the linear constraint

$$l^i_{j_i}(z) = g_{j_i}(z^i) + \alpha^i_{j_i 0}\xi_{j_i}(z^i)^T(z - z^i) \leq 0 \tag{6}$$

to the new problem (P^{i+1}), where again, $\alpha^i_{j_i 0}$ is set to 1, $g_{j_i}(z^i) = \hat{g}(z^i)$ and $\xi_{j_i}(z^i) \in \partial g_{j_i}(z^i)$. The cutting plane (6) is called *valid* if it does not cut off any points from the original feasible set $N \cap L$. Next, we shall prove that there always exists a certain value $\alpha^i_{j_i}$ such that the cutting plane (6) is valid. The existence follows from the f°-pseudoconvexity of the constraint functions.

THEOREM 3.1. *Let $z^i \in L$ be a solution to the problem (P^i) and j_i be such that $g_{j_i}(z^i) = \hat{g}(z^i) > 0$. Then, there exists $M > 0$ such that the cutting plane*

$$g_{j_i}(z^i) + \alpha^i_{j_i}\xi_{j_i}(z^i)^T(z - z^i) \leq 0 \tag{7}$$

is valid if $\alpha^i_{j_i} > M$.

Proof. Since $g_{j_i}(z^i) > 0$, the inequality $g_{j_i}(z) < g_{j_i}(z^i)$ holds for every feasible z. From the f°-pseudoconvexity of function g_{j_i} and Definition 2.4, we deduce that

$$\xi_{j_i}(z^i)^T(z - z^i) \leq g^\circ_{j_i}(z^i; z - z^i) < 0$$

for any $\xi_{j_i}(z^i) \in \partial g_{j_i}(z^i)$. Since $N \cap L$ and $\partial g_{j_i}(z^i)$ are compact (Theorem 2.5 (i)) the continuous function $G(z, \xi) = \xi^T(z - z^i)$ attains its maximum value on the compact set $(N \cap L) \times \partial g_{j_i}(z^i)$. Denote this negative scalar $-\varepsilon$. If we choose $\alpha^i_{j_i} \geq \frac{g_{j_i}(z^i)}{\varepsilon}$, then for any $z \in N \cap L$ and $\xi_{j_i}(z^i) \in \partial g_{j_i}(z^i)$ we have

$$\begin{aligned} & g_{j_i}(z^i) + \alpha^i_{j_i}\xi_{j_i}(z^i)^T(z - z^i) \\ \leq \quad & g_{j_i}(z^i) + \alpha^i_{j_i} \max_{\substack{z \in N \cap L \\ \xi \in \partial g_{j_i}(z^i)}} \xi^T(z - z^i) \\ \leq \quad & g_{j_i}(z^i) + \frac{g_{j_i}(z^i)}{\varepsilon}(-\varepsilon) = 0. \end{aligned}$$

Thus, we see that any point $z \in N \cap L$ satisfies the inequality (7). Consequently, we may choose $M = \frac{g_{j_i}(z^i)}{\varepsilon}$ and the theorem is proved. □

From the proof of Theorem 3.1 we see that the cutting plane (7) is valid if

$$\alpha^i_{j_i} \geq \max_{\substack{z \in N \cap L \\ \xi \in \partial g_{j_i}(z^i)}} -\frac{g_{j_i}(z^i)}{\xi^T(z - z^i)}. \tag{8}$$

It is difficult to provide sufficiently large $\alpha^i_{j_i}$ from the inequality (8). In practise, a small value $\varepsilon_z > 0$ is given and a cutting plane from point z^i is considered valid if it's shortest distance to z^i is less or equal to ε_z and $\alpha^i > 0$. More precisely, denote $T_{ij} = \left\{z \mid \xi_{j_i}(z^i)^T(z - z^i) = 0\right\}$, where $\xi_{j_i}(z^i) \in \partial g_{j_i}(z^i)$ is chosen to the cutting plane. The cutting plane (with $\|\xi_{j_i}(z^i)\| \neq 0$) can be written as

$$\begin{aligned} & g_{j_i}(z^i) + \alpha^i_{j_i}\xi_{j_i}(z^i)^T(z - z^i) = \\ & g_{j_i}(z^i) + \left\|\xi_{j_i}(z^i)\right\| \alpha^i_{j_i} \frac{\xi_{j_i}(z^i)^T}{\|\xi_{j_i}(z^i)\|}(z - z^i) \leq 0. \end{aligned}$$

Note that if $\|\xi_{j_i}(z^i)\| = 0$ then $0 \in \partial g_{j_i}(z^i)$. Then it follows that z^i is a global minimum of g_{j_i} due to f°-pseudoconvexity of g_{j_i} (see [11] page 7). This would imply that $N = \emptyset$ as $g_{j_i}(z^i) > 0$. Thus, the assumption $\|\xi_{j_i}(z^i)\| \neq 0$

is sensible. If any point $\hat{z}$ with $\xi_{j_i}(z^i)^T(\hat{z}-z^i) < 0$ that has distance to T_{ij} greater than or equal to ε_z, in other words

$$\left| \frac{\xi_{j_i}(z^i)^T}{\|\xi_{j_i}(z^i)\|}(\hat{z}-z^i) \right| \geq \varepsilon_z,$$

should not be cutted off, then

$$g_{j_i}(z^i) + \left\|\xi_{j_i}(z^i)\right\| \alpha^i_{j_i} \frac{\xi_{j_i}(z^i)^T}{\|\xi_{j_i}(z^i)\|}(\hat{z}-z^i) \leq g_{j_i}(z^i) + \left\|\xi_{j_i}(z^i)\right\| \alpha^i_{j_i}(-\varepsilon_z) \leq 0.$$

This leads to inequality

$$\alpha^i_{j_i} \geq \frac{g_{j_i}(z^i)}{\|\xi_{j_i}(z^i)\| \, \varepsilon_z}, \tag{9}$$

which was presented in [16] for the smooth case. To make the algorithm faster, $\alpha^i_{j_i}$ is set to 1 at first and it is updated in later iterations until it satisfies the inequality (9). More accurately, a coefficient $\beta > 1$ is given and $\alpha^i_{j_i}$ coefficients are updated according to equation

$$\alpha^i_{j_i(k+1)} = \beta \alpha^i_{j_i k} \tag{10}$$

when necessary. Now we are ready to present the NαECP algorithm.

Algorithm 3.1.
0. Give tolerance parameter $\varepsilon_g > 0$, update parameter $\beta > 1$ and set $i = 0$. Create the problem (P^0) by omitting the nonlinear constraints from (P).
1. Solve the MILP problem (P^i).
2. If the problem has a solution denote it z^i and go to the next step. If not and some α coefficients does not satisfy inequality (9), update them according to (10), set $i = i+1$ and return to step 1. Otherwise, the problem has no solution.
3. If $\hat{g}(z^i) > \varepsilon_g$, then create problem P^{i+1} by adding a cutting plane (6) from the point z^i and function g_{j_i} with $g_{j_i}(z^i) = \hat{g}(z^i)$. Set $\alpha^i_{j_i} = 1$, $i = i+1$ and go to step 1. If $\hat{g}(z^i) \leq \varepsilon_g$ and all α coefficients satisfy inequality (9) then z^i is the final solution and the algorithm stops. Otherwise, update all α coefficients for which the inequality (9) is not satisfied, set $i = i+1$ and got to step 1.

3.2 The convergence proof

Next, we will prove that the generalized algorithm NαECP converges to a global minimum if $\varepsilon_g = 0$. In order to be able to proof the convergence rigorously, we assume that sufficiently large coefficients α^i are known beforehand and they are used immediatedly when the cutting planes are created. Thus, the cutting planes are always valid and we do not need to update α values. If all constraint functions are convex, then $\alpha = 1$ is always sufficiently large [4].

The difficulty with choosing $\varepsilon_g = 0$ is that the algorithm may not stop after a finite number of iterations. On the other hand, if $\varepsilon_g > 0$ the algorithm will always stop after a finite number of iterations but the solution may not be feasible.

The assumption that sufficiently large α values are known beforehand is unrealistic (unless functions are convex), but the convergence proof shows that the algorithm used in practice has reasonable origins.

If we denote the feasible set of problem (P^i) by

$$\Omega^i = \left\{ z \mid z \in \Omega^{i-1}, l^{i-1}_{j_{i-1}}(z) \leq 0 \right\}.$$

and $\Omega^0 = L$ we get a sequence

$$N \cap L \subset \cdots \subset \Omega^i \subset \cdots \subset \Omega^0. \tag{11}$$

The property (11) follows from Theorem 3.1 which states that the feasible points are not cut off by the cutting planes as α coefficients are assumed to be sufficiently large. The next theorem follows quite easily from the relation (11).

Theorem 3.2. *Suppose that the Algorithm 3.1 ends with a finite number of iterations or an accumulation point $\hat{z}$ of solution sequence (z^i) is feasible. Then the last iteration point or the accumulation point is a global minimum.*

Proof. If the algorithm ends up with a finite number of iterations say at i, the last solution point is feasible in problem (P). This point is also a global minimum of the problem (P), since the point is a global minimum at the feasible set of problem (P^i) and this set includes the feasible set of problem (P) according to (11).

Assume now that the accumulation point $\hat{z}$ is feasible. Denote $c^T z^i = Z^i$ and $c^T \hat{z} = \hat{Z}$. By relation (11) the sequence (Z^i) is nondecreasing and for any i the Z^i is a lower bound for the optimal objective function value. By the continuity of the objective function we have $Z^i \to \hat{Z}$ as $i \to \infty$. Hence, there exists a lower bound that is arbitrarily close to $\hat{Z}$ and, thus, $\hat{z}$ is a global minimum. □

Next, we will study the case where the Algorithm 3.1 does not converge to a global minimum after a finite number of iterations. It turns out that then we will have a sequence of solutions (z^i) with an accumulation point $\hat{z}$ and this accumulation point is feasible in problem (P).

First, we prove with the following lemma and theorem that there exists an accumulation point.

LEMMA 3.3. *If $z^i \notin N \cap L$ is the solution of problem (P^i), then it is infeasible in the subsequent MILP problems.*

Proof. Since $z^i \notin N \cap L$ the cutting plane

$$l^i_{j_i}(z) = g_{j_i}(z^i) + \alpha^i_{j_i} \xi_{j_i}(z^i)^T (z - z^i) \leq 0 \tag{12}$$

with $g_{j_i}(z^i) > 0$ will be introduced to the subsequent MILP problems. Since

$$l^i_{j_i}(z^i) = g_{j_i}(z^i) + \alpha^i_{j_i} \xi_{j_i}(z^i)^T (z^i - z^i) = g_{j_i}(z^i) > 0$$

we see that z^i does not satisfy inequality (12). Thus, z^i is infeasible in all subsequent MILP problems. □

THEOREM 3.4. *If NαECP-algorithm does not stop after a finite number of iterations, then it generates a solution sequence (z^i) with an accumulation point.*

Proof. By Lemma 3.3 the sequence (z^i) contains infinite number of different points. Since $(z^i) \subset L$ and L is compact, the classical Bolzano-Weierstrass Theorem states that the sequence (z^i) has an accumulation point. □

Now we know that an accumulation point exists. Next, we will prove that any accumulation point of the sequence (z^i) is feasible in problem (P). We will need the following lemma for this.

LEMMA 3.5. *Let $\hat{z} \in L \setminus (N \cap L)$ and j be such that $g_j(\hat{z}) > 0$. Let $\varepsilon > 0$ be such that $g_j(z) > 0$ if $z \in \operatorname{cl} B(\hat{z};\varepsilon)$. Then there exists $D \in \mathbb{R}$ such that*

$$\frac{g_j(z)}{|\xi_j(z)^T(w-z)|} \leq D, \tag{13}$$

for all $z \in \operatorname{cl} B(\hat{z};\varepsilon) \cap L$, $w \in N \cap L$ and $\xi_j(z) \in \partial g_j(z)$.

Proof. By local Lipschitz continuity of function g_j we have

$$g_j(z) \leq g_j(\hat{z}) + K\varepsilon. \tag{14}$$

Denote $A = \operatorname{cl} B(\hat{z};\varepsilon) \cap L$ and $C = N \cap L$. The set A is compact and continuous function g_j attains a minimum value on it. Denote this value g_0. Since $A \subset \operatorname{cl} B(\hat{z};\varepsilon)$ we have $g_0 > 0$. Thus, $g_j(z) \geq g_0$ for all $z \in A$ and $g_j(w) \leq 0 < g_0$ for all $w \in C$. Since g_j is f°-pseudoconvex, according to Lemma 2.10 there exists $\delta > 0$ such that

$$\sup_{\substack{z \in A \\ \xi \in \partial g_j(z) \\ w \in C}} \xi^T(w-z) = -\delta.$$

Hence, for all $z \in B(\hat{z};\varepsilon)$ and $w \in N \cap L$ we have by f°-pseudoconvexity of g_j that

$$\begin{aligned} \left|\xi_j(z)^T(w-z)\right| &= -\xi_j(z)^T(w-z) \\ &\geq -\sup_{\substack{z \in A \\ \xi \in \partial g_j(z) \\ w \in C}} \xi^T(w-z) = \delta. \end{aligned}$$

By (14) we may choose $D = \frac{g_j(\hat{z})+K\varepsilon}{\delta}$. □

Essentially the Lemma 3.5 states that for every $z \in L \setminus (N \cap L)$ there exists $\varepsilon > 0$ such that a finite α is sufficiently large for cutting planes created at $\operatorname{cl} B(z;\varepsilon) \cap L$. This can be seen by rewriting the inequality (8) as follows

$$\alpha^i_{j_i} \geq \max_{\substack{z \in N \cap L \\ \xi \in \partial g_{j_i}(z^i)}} \frac{-g_{j_i}(z^i)}{\xi^T(z-z^i)} = \max_{\substack{z \in N \cap L \\ \xi \in \partial g_{j_i}(z^i)}} \frac{g_{j_i}(z^i)}{|\xi^T(z-z^i)|},$$

where the equality is a consequence of the negativity of $\xi^T(z-z^i)$. Now, we are ready to prove that an accumulation point is feasible.

THEOREM 3.6. *If the point $\hat{z}$ is an accumulation point of the sequence of MILP solutions generated by NαECP-algorithm, then it is a feasible point of the problem (P).*

Proof. On the contrary, assume that the accumulation point $\hat{z} \notin N \cap L$, that is, $\hat{g}(\hat{z}) > 0$. Due to the finite number of constraint functions there exists j such that g_j is linearized infinitely many times and $g_j(\hat{z}) > 0$. Since $\mathbb{R}^n \setminus \{z \mid g_j(z) \leq 0\}$ is an open set, there exists $\varepsilon > 0$ such that $g_j(z) > 0$ for all $z \in \operatorname{cl} B(\hat{z}; \varepsilon)$ and Lipschitz condition (1) holds for g_j and any $y, z \in \operatorname{cl} B(\hat{z}; \varepsilon)$ with constant $K > 0$.

By Lemma 3.5 there is an upper bound for the quotient

$$\frac{g_j(z)}{|\xi_j(z)^T(w - z)|},$$

when $z \in \operatorname{cl} B(\hat{z}; \varepsilon) \cap L$ and $w \in N \cap L$. Thus, there exists a finite $\hat{\alpha}$ for which cutting planes generated at $B(\hat{z}; \varepsilon) \cap L$ on function g_j are valid. Assume that for those cutting planes the chosen α values are smaller than $\tilde{\alpha}$, that is, $\alpha \in [\hat{\alpha}, \tilde{\alpha}]$. Denote

$$\delta = \min\left\{\varepsilon, \frac{g_j(\hat{z})}{2\tilde{\alpha}K + K}\right\}.$$

Let $k > i$ and $z^k, z^i \in B(\hat{z}; \delta)$, where z^k is the solution point to the problem (P^k) and g_j is linearized at i. Then by inequality (1) we have

$$g_j(z^i) \geq g_j(\hat{z}) - K \left\|\hat{z} - z^i\right\| > g_j(\hat{z}) - K\delta. \tag{15}$$

Also,

$$\alpha_j^i \xi_j(z^i)^T(z^k - z^i) \geq -\tilde{\alpha} \left\|\xi_j(z^i)\right\| \left\|z^k - \hat{z} + \hat{z} - z^i\right\|,$$

since $\alpha_j^i \leq \tilde{\alpha}$. By the triangle inequality and Theorem 2.5 (i) we have

$$-\tilde{\alpha} \left\|\xi_j(z^i)\right\| \left\|z^k - \hat{z} + \hat{z} - z^i\right\| \geq -\tilde{\alpha}K(\left\|z^k - \hat{z}\right\| + \left\|\hat{z} - z^i\right\|) > -2\tilde{\alpha}K\delta.$$

Together with (15) we have

$$\begin{aligned} g_j(z^i) + \alpha_j^i \xi_j(z^i)^T(z^k - z^i) &> g_j(\hat{z}) - K\delta - 2\tilde{\alpha}K\delta \\ &\geq g_j(\hat{z}) - (K + 2\tilde{\alpha}K)\frac{g_j(\hat{z})}{2\tilde{\alpha}K + K} = 0. \end{aligned}$$

Thus, z^k would not be a feasible point in the problem (P^k), which contradicts with the assumption. Hence, any accumulation point is feasible. □

We summarize the convergence results in the following theorem.

THEOREM 3.7. *If $\varepsilon_g = 0$ and all the cutting planes are valid the NαECP-algorithm converges to a global minimum.*

Proof. If the algorithm stops after a finite number of iterations the result follows from Theorem 3.2. If the algorithm does not stop after a finite number of iterations, it creates a sequence with a feasible accumulation point by Theorems 3.4 and 3.6. A feasible accumulation point is then a global minimum according to Theorem 3.2. □

There may occur many accumulation points. However, since Theorems 3.4 and 3.6 considers any of them, all of the accumulations points will be global minima.

Note, that there may not exist a global M for α values for which any cutting plane would be valid. In fact, it is an open question whether there exists a problem that generates sequences (z^i) and values (M^i) with $\lim_{i\to\infty} M^i = \infty$. Nevertheless, Theorems 3.4 and 3.6 guarantee that the sequence (z^i) has an accumulation point and it is feasible. Hence, a global minimum will be found in this case as well.

For the next section we need the following result. Again, we assume that sufficiently large α coefficients are known beforehand.

THEOREM 3.8. *If $\varepsilon_g > 0$ then the NαECP algorithm terminates after a finite number of iterations.*

Proof. On the contrary, assume that the algorithm does not terminate after a finite number of iterations. Then the solution sequence (z^i) has an accumulation point by Theorem 3.4. Denote this accumulation point $\hat{z}$. Theorem 3.6 implies that $\hat{g}(\hat{z}) \leq 0$. By the continuity of function $\hat{g}$ there exists $\delta > 0$ such that

$$|\hat{g}(z) - \hat{g}(\hat{z})| < \varepsilon_g$$

if $z \in B(\hat{z}; \delta)$. By the definition of accumulation point there exist some points (actually infinitely many) in the sequence (z^i) that belongs to the set $B(\hat{z}; \delta)$. Thus, there exists a point $z^j \in (z^i)$ for which we have

$$\hat{g}(z^j) \leq \hat{g}(z^j) - \hat{g}(\hat{z}) < \varepsilon_g, \tag{16}$$

where $\hat{g}(z^j) > 0$ since $(z^i) \cap N \cap L = \emptyset$. The inequality (16) implies that the algorithm terminates after the jth iteration. Thus, there can be at most finite number of iterations. $\square$

4 f°-pseudoconvex objective function

Next we develop an algorithm that can deal also problems with an f°-pseudoconvex objective function. As in the previous section, the constraint functions are assumed to be f°-pseudoconvex. This algorithm can be seen as a generalization to the algorithm presented in [16], where the object and constraint functions were pseudoconvex.

Our algorithm solves problem (GP) as a series of MINLP subproblems with linear objective function and f°-pseudoconvex constraints. This kind of problem was already studied in Section 3. Throughout this section we assume that in the MINLP problems $\varepsilon_g > 0$ and sufficiently large α coefficients are known beforehand. It is convenient to introduce the notation $\tilde{N} = \{z \mid \hat{g}(z) \leq \varepsilon_g\}$. We shall call a point $z \in \tilde{N} \cap L$ an ε_g-*feasible* point. From Theorem 3.8 we deduce that MINLP problems will be solved after a finite number of iterations. On the other hand, the solution point may not be feasible in the original problem (GP) but it will be ε_g-feasible. The first problem is

$$\begin{array}{ll} \min & \mu \\ \text{s.t.} & \mu \geq \mu_0 \quad (P_0^{00}) \\ & z \in N \cap L, \end{array}$$

where the constant μ_0 guarantees that the problem has a bounded solution. Note that the purpose of the problem (P_0^{00}) is merely to find an ε-feasible solution. Denote the solution point (z^0, μ^0). Now, we have an (ε_g-feasible) upper bound $f(z^0)$ which we denote f_1. Also, we denote $z_1^1 = z^0$. Next, we will solve the problem

$$
\begin{aligned}
\min \quad & \mu \\
\text{s.t.} \quad & \mu \le f_1 \\
& f(z) \le f_1 \qquad\qquad\qquad\qquad (P_1^{11}) \\
& f_1 + \xi(z_1^1)^T(z - z_1^1) \le \mu \qquad\qquad (17) \\
& z \in N \cap L,
\end{aligned}
$$

where $\xi(z_1^1) \in \partial f(z_1^1)$. Note that constraint (17) guarantees that (P_1^{11}) has a bounded solution. Therefore, we were allowed to discard the constraint $\mu \ge \mu_0$. Let the next solution point be (z^1, μ^1). Note that $f(z^1) \le f_1 + \varepsilon_g$. If $f(z^1) < f_1$, we have found a new upper bound $f(z^1) = f_2$. Then, f_1 will be replaced by f_2 and we denote $z_2^1 = z^1$. Also, the constraint (17) will be replaced by the constraint

$$f_2 + \xi(z_2^1)^T(z - z_2^1) \le \mu.$$

If $f_1 \le f(z^1)$, then denote $z_1^2 = z^1$ and add a new constraint

$$f_1 + \xi(z_1^2)^T(z - z_1^2) \le \mu$$

to the problem. At iteration i we will solve the problem

$$
\begin{aligned}
\min \quad & \mu \\
\text{s.t.} \quad & \mu \le f_r \\
& f(z) \le f_r \qquad\qquad\qquad\qquad (P_r^{qi}) \\
& f_r + \xi(z_r^j)^T(z - z_r^j) \le \mu, \quad \text{for all } j \in J_r \qquad (18) \\
& z \in N \cap L,
\end{aligned}
$$

where i is the total number of iterations, r is the number of upper bounds obtained so far and $J_r = \{1, \dots, q\}$, where $q - 1$ indicates how many times we have solved MINLP problem with the upper bound f_r. Note that the only constraints that limits the optimal μ from below are (18). Then, we may deduce that if (z^i, μ^i) is the solution to the problem (P_r^{qi}), we have

$$\sup_{j \in J_r} \left\{ f_r + \xi(z_r^j)^T(z^i - z_r^j) \right\} = \mu^i. \qquad (19)$$

The algorithm proceeds as follows.

ALGORITHM 4.1.
0. Give positive accuracy tolerances ε_g, ε_f and set $i = q = r = 0$ and $f_0 = \infty$. Create the problem (P_0^{00}) by replacing the objective function f by μ and add the constraint $\mu \geq \mu_0$.
1. Solve the MINLP problem (P_r^{qi}) with the algorithm 3.1. Denote the solution (z^i, μ^i) and $f = f(z^i)$. If $i = 0$ remove constraint $\mu \geq \mu_0$.
2. If $i > 0$ and $|f - \mu^i| \leq \varepsilon_f$ then z^i is the final solution and algorithm stops.
3. If $f < f_r$ set $f_{r+1} = f$, $q = 1$ and $z_{r+1}^1 = z^i$. Update the constraints by replacing old f_r by f_{r+1}. Discard the constraints (18) and add the constraint $f_{r+1} + \xi(z_{r+1}^1)^T(z - z_{r+1}^1) \leq \mu$. Set $r = r + 1$.
4. If $f \geq f_r$ set $q = q+1$, $z_r^q = z^i$ and add the constraint $f_r + \xi(z_r^q)^T(z - z_r^q) \leq \mu$ to the problem.
5. Set $i = i + 1$ and go to step 1.

Note that problems (P_r^{qi}) can be solved to nearly global feasible solution in a finite number of iterations if $\varepsilon_g > 0$.

In the following, we will prove that our algorithm converges. We assume that the sequence of MINLP solutions (z^i) contains only one accumulation point and $\varepsilon_f = 0$. Then the algorithm will converge to a point $z \in \tilde{N} \cap L$ which is ε_g-feasible. Also, there are no feasible points with objective function value lower than $f(z)$. We shall call such point z an *ε_g-feasible global minimum* of problem (GP).

First, we note that if the point z is feasible in problem (GP) and $f(z) < f_r$, then it is feasible in problems (P_r^{qi}) for all q and i. The constraints (18) are satisfied with $\mu \leq f_r$ since

$$\xi(z_r^j)^T(z - z_r^j) \leq f^\circ(z_r^j; z - z_r^j) < 0$$

for all $j \in J_r$. This is due to f°-pseudoconvexity of f and inequalities $f(z) < f_r \leq f(z_r^j)$, $j \in J_r$. The other constraints are satisfied since z is feasible in (GP) and $f(z) < f_r$.

We denote $\mu_{\tilde{z}}$ the optimal value of μ in problem (P_r^{qi}) with additional constraint $z = \tilde{z}$, where $\tilde{z}$ is feasible in (P_r^{qi}). Consequently $\mu_{\tilde{z}} \geq \mu^i$.

Next, we will justify the stopping criterion $|f_r - \mu^i| \leq \varepsilon_f = 0$.

THEOREM 4.1. *If $\mu = f_r$, then f_r is an ε_g-feasible global minimum.*

Proof. On the contrary, assume that there exists a feasible point $\tilde{z}$ with $f(\tilde{z}) < f_r$. Then, $\tilde{z}$ is feasible in current MINLP problem. Denote $A = \tilde{N} \cap L \cap \{z \mid f(z) \geq f_r\}$. By Lemma 2.10 there exists $\delta > 0$ such that

$$-\delta = \sup_{\substack{z \in A \\ \xi \in \partial f(z)}} \xi^T(\tilde{z} - z) \geq \sup_{j \in J_r} \xi(z_r^j)^T(\tilde{z} - z_r^j) = \mu_{\tilde{z}} - f_r.$$

Thus, at the feasible point $\tilde{z}$ we would have $\mu_{\tilde{z}} \leq f_r - \delta < \mu$ which is impossible since μ was assumed to be a minimum. Hence, there exists no feasible point $\tilde{z}$ with $f(\tilde{z}) < f_r$. □

If the Algorithm 4.1 does not stop after a finite number of iterations there will be an infinite sequence of solutions $(z_r^1)_{r=1}^\infty \subset (z^i)$ or for some r there will be an infinite sequence of solutions $(z_r^j)_{j=1}^\infty$. In the first case the sequence contains only different points as $f(z_{r_1}^1) > f(z_{r_2}^1)$ whenever $r_1 < r_2$. In the latter case, if there exists $j > k$ with $z_r^k = z_r^j$ then from the constraint (18) we see that

$$f_r + \xi(z_r^k)^T(z_r^j - z_r^k) \leq \mu,$$

implying $f_r \leq \mu$. With the constraint $\mu \leq f_r$ this results to $\mu^i = f_r$ and the algorithm stops. Hence, if the sequence (z_r^j) is infinite for some r, then it contains only different points.

In the following we shall show that if the sequence of solutions has one accumulation point then the algorithm will end up with an ε_g-feasible global minimum.

Theorem 4.2. *If in the Algorithm 4.1 there will be an infinite number of iterations for some r, then f_r is an ε_g-feasible global minimum.*

Proof. The infinite sequence (z_r^j) belongs to the compact set L. Every point of the sequence is different and, thus, there exists an accumulation point $\hat{z}$. Let $\varepsilon > 0$ be such that the Lipschitz condition holds in $B(\hat{z}; \varepsilon)$ with constant K for function f. Let $\delta = \min\left\{\varepsilon, \frac{\varepsilon}{2K}\right\}$ and $k > j$ be indexes such that $z_r^j, z_r^k \in B(\hat{z}; \delta)$. Then, at problem $(P_r^{(k-1)i})$ the constraint

$$f_r + \xi(z_r^j)^T(z - z_r^j) \leq \mu$$

is satisfied with $z = z_r^k$ and $\mu = \mu^i$. Thus,

$$f_r + \xi(z_r^j)^T(z_r^k - z_r^j) \leq \mu^i \quad \text{implying}$$
$$\xi(z_r^j)^T(z_r^k - z_r^j) \leq \mu^i - f_r.$$

Both sides of the inequality are non-positive implying that

$$\left|\xi(z_r^j)^T(z_r^k - \hat{z} + \hat{z} - z_r^j)\right| \geq \left|\mu^i - f_r\right|.$$

The leftside is

$$\begin{aligned}\left|\xi(z_r^j)^T(z_r^k - \hat{z} + \hat{z} - z_r^j)\right| &\leq \left\|\xi(z_r^j)\right\|\left(\left\|z_r^k - \hat{z}\right\| + \left\|\hat{z} - z_r^j\right\|\right) \\ &< K2\delta \leq 2K\frac{\varepsilon}{2K} = \varepsilon.\end{aligned}$$

Hence, $|\mu^i - f_r| < \varepsilon$.

Now, assume that there exists $\tilde{z} \in N \cap L$ with $f(\tilde{z}) < f_r$ and denote $A = \tilde{N} \cap L \cap \{z \mid f(z) \geq f_r\}$. By Lemma 2.10 there exists $\delta > 0$ such that

$$-\delta = \sup_{\substack{z\in A \\ \xi\in\partial f(z)}} \xi^T(\tilde{z} - z) \geq \sup_{j\in J_r} \xi(z_r^j)^T(\tilde{z} - z_r^j) = \mu_{\tilde{z}} - f_r,$$

where $J_r = \mathbb{N}\setminus\{0\}$. Let l be such that $\left|\mu^l - f_r\right| < \delta$. Then, $\mu^l > f_r - \delta \geq \mu_{\tilde{z}}$ which is impossible since μ^l is the minimum at iteration l. Hence, the point $\tilde{z}$ does not exist and the theorem is proved. □

From Theorem 4.2 we deduce that if the sequence (z_r^j) is infinite for some r then the algorithm will converge despite of the number of accumulation points. This is not the case when the upper bound is improved infinitely many times.

Note that if the upper bound will be improved infinitely many times there exists $\hat{f} = \lim_{r\to\infty} f_r$. This follows from the facts that the continuous function f is bounded below on a compact set L and the sequence (f_r) is decreasing.

Theorem 4.3. *If $r \to \infty$, and there is only one accumulation point, then the Algorithm 4.1 converges to an ε_g-feasible global minimum.*

Proof. All the points in the subsequence (z_r^1) are different. As the points are in the compact set L there exists an accumulation point $\hat{z}$. Let $\varepsilon > 0$ be such that the Lipschitz condition holds in $B(\hat{z};\varepsilon)$ with constant K for function f. Let $\delta = \min\left\{\varepsilon, \frac{\varepsilon}{2K}\right\}$ and $\tilde{r}$ be an index such that $z_{\tilde{r}}^1, z_{\tilde{r}+1}^1 \in B(\hat{z};\delta)$. Since there is exactly one accumulation point, such $\tilde{r}$ is guaranteed to exist. Thus we have

$$\begin{aligned}&\xi(z_{\tilde{r}}^1)^T(z_{\tilde{r}+1}^1 - z_{\tilde{r}}^1) \leq \mu^i - f_{\tilde{r}} < 0 \quad \text{implying} \\ &\left|\xi(z_{\tilde{r}}^1)^T(z_{\tilde{r}+1}^1 - z_{\tilde{r}}^1)\right| \geq \left|\mu^i - f_{\tilde{r}}\right|\end{aligned}$$

The leftside is

$$\begin{aligned}\left|\xi(z^1_{\tilde r})^T(z^1_{\tilde r+1}-\hat z+\hat z-z^1_{\tilde r})\right| &\le \left\|\xi(z^1_{\tilde r})\right\|(\left\|z^1_{\tilde r+1}-\hat z\right\|+\left\|\hat z-z^1_{\tilde r}\right\|)\\ &< K2\delta \le 2K\frac{\varepsilon}{2K}=\varepsilon.\end{aligned}$$

Thus, $|\mu^i - f_{\tilde r}| < \varepsilon$.

Now, assume that there exists $\tilde z \in N \cap L$ such that $f(\tilde z) < \hat f = \lim_{r\to\infty} f_r$. Denote $A = \tilde N \cap L \cap \left\{ z \mid f(z) \ge \hat f \right\}$. Then by Lemma 2.10 there exists $\delta > 0$ such that

$$-\delta = \sup_{\substack{z\in A\\ \xi\in\partial f(z)}} \xi(z)^T(\tilde z - z).$$

Let l and r be such that at iteration l we have $\left|\mu^l - f_r\right| < \delta$. Now we have

$$\mu_{\tilde z} = f_r + \max_{j\in J_r} \xi(z^j_r)^T(\tilde z - z^j_r) \le f_r + \sup_{\substack{z\in A\\ \xi\in\partial f(z)}} \xi(z)^T(\tilde z - z) = f_r - \delta.$$

Thus, $\mu_{\tilde z} < \mu^l$ which is impossible since μ^l was the minimum. Hence, $\tilde z$ does not exist and $\hat f$ is an ε_g-feasible global minimum. □

Again, we summarize the convergence proof.

Theorem 4.4. *If $\varepsilon_f = 0$ and the solution sequence contains at most one accumulation point the Algorithm 4.1 converges to an ε_g-feasible global minimum.*

Proof. If the algorithm stops after a finite number of iterations, then $\mu = f_r$ and the minimum is found by Theorem 4.1. If the algorithm does not stop after a finite number of iterations and there is exactly one accumulation point then algorithm will converge to the minimum according to the Theorems 4.2 and 4.3. □

The algorithm will converge after a finite number of iterations if $\varepsilon_f > 0$ which can be proven similarly to the case $\varepsilon_g > 0$ in Theorem 3.8.

5 Numerical examples

In this section we will apply the nonsmooth αECP algorithm to three example problems. The first one is two dimensional problem

$$\begin{aligned} \min \quad & f(x_1,x_2) = \max\left\{\sqrt{1+|x_1|}, \sqrt{1+|x_2|}\right\} \\ \text{s.t.} \quad & -5 \le x_1 \le 5,\ -5 \le x_2 \le 5,\ x_2 \in \mathbb{Z} \qquad (P1)\end{aligned}$$

The objective function f can be considered as a maximum of four pseudo-convex functions

$$g_i^+(x_i) = \begin{cases} \sqrt{1+x_i}, & x_i \geq 0 \\ 1 + \frac{1}{2}x_i, & x_i < 0 \end{cases} \quad \text{and} \quad g_i^-(x_i) = \begin{cases} 1 - \frac{1}{2}x_i, & x_i \geq 0 \\ \sqrt{1-x_i}, & x_i < 0 \end{cases},$$

where $i = 1, 2$. Since the functions g_i^+ and g_i^- are pseudoconvex, f is f°-pseudoconvex by Theorems 2.8 and 2.9. The function f is nonsmooth at the points satisfying the equation $|x_1| = |x_2|$. The gradient of an active function is a subgradient of f at corresponding point by Theorem 2.6. By calculating subgradients analytically this way the $N\alpha$ECP algorithm found the global optimum $(0, 0)$ with the optimal value 1.

We also tried to approximate a subgradient with the standard (forward) finite difference method developed to calculate gradients numerically. It is well known that the finite difference method is not the best way to calculate a subgradient [9]. Indeed, the algorithm stopped at $(-5, -5)$ yielding a non-optimal objective function value 2.45.

The second problem is a modification of a problem presented in [16]. Replacing the term $(x-3)^2$ by $|x-3|$ we get a nonsmooth problem

$$\begin{array}{ll} \min & \dfrac{|x-3| - 10x}{3x+y+1} \\ \text{s.t.} & (x-7)^2 - 5y \leq 0 \\ & x - 1.8y \leq 0 \qquad\qquad (P2) \\ & 1 \leq x, \quad y \leq 8 \\ & x \geq 0, \quad y \in \mathbb{Z}^+. \end{array}$$

The objective function can be viewed as a maximum of pseudoconvex functions and thus, the subgradients can be calculated similarly to the first example problem. We calculated again the subgradients both exactly and with the finite difference method. The global optimum was obtained with both ways. It turns out that in problem 2 the algorithm does not visit any nonsmooth point and hence the finite difference method yields a correct approximation to the gradients.

Finally we consider a real life problem, namely the cyclic scheduling problem taken from [7, 16]. In this problem there are 7 feeds and 4 furnaces and optimal scheduling should be made in order to maximize the profit. The problem has 140 binary variables, 233 continuous variables and 138 constraints. An outline of the problem is in the appendix. Again, we modified the differentiable objective function. Instead of maximizing the profit, we maximize the profit of the least profitable furnace. When we change the maximization to minimization by multiplying the objective function with -1, the objective function transforms to a maximum function of four pseudoconvex functions. Again, we will calculate the subgradients as in previous examples. Both the exact subgradient and the finite difference method results in objective function value -39071. The numerical results of all three problems are summarized in table 1. The problems were solved on HP pavilion dv9500 notebook PC with 1.5 GHz Intel processor. In each problem we

Table 1: Numerical results

problem	optimal value	subgradient	μ^*	iterations	CPU-time(s)
P1	1	exact	1.00	33	0.45
		numerical	2.45	2	0.04
P2	$-\frac{258}{101} \approx -2.55$	exact	-2.55	10	0.17
		numerical	-2.55	10	0.17
P3	best known $= -39071$	exact	-39071	255	85
		numerical	-39071	210	60

had $\varepsilon_z = 0.1$ and $\beta = 1.3$. In problems 1 and 2 parameters $\varepsilon_f = 0.001$, $\varepsilon_g = 0.001$ and the difference 0.001 in the finite difference were used. In problem 3 we had $\varepsilon_g = \varepsilon_f = 10$, and the difference were 0.00001. Also, in problem 3 a line search procedure introduced in [16] was used. This procedure does not affect the theoretical convergence properties. If the procedure is not used the algorithm will take much more time (> 6 hours).

6 Conclusions

The αECP method for nonsmooth MINLP problems was studied. The considered MINLP problem was required to have f°-pseudoconvex constraint functions and linear or f°-pseudoconvex objective function. The algorithms were readily generalized for nonsmooth functions from those presented in [14] and [16]. In theory, the only difference in algorithms was that gradients were replaced by subgradients.

In the case where the objective function is linear the algorithm was proven to converge to global minimum if sufficiently large α coefficients are known beforehand. If the objective function is f°-pseudoconvex the generalized algorithm was proven to converge to an ε_g-feasible global minimum assumed that sufficiently large α coefficients are known beforehand.

A few numerical problems were considered. The results were in accordance with the theory. The algorithm converged to the global optimal solution or at least to the best known solution for all examples when applying subgradients. When the gradients were approximated by finite differences, it was further illustrated that the global optimal solution could not be achieved in all examples.

References

[1] Clarke, F. H. Optimization and Nonsmooth Analysis. Wiley-Interscience, New York, 1983.

[2] Duran, M. A., and Grossmann, I. E. An outer-approximation algorithm for a class of mixed-integer nonlinear programs. *Mathematical Programming 36*, 3 (1986), 307–339.

[3] Emet S., and Westerlund, T. Comparisons of solving a chromatographic separation problem using MINLP methods. *Computers & Chemical Engineering 28*, (2004), 673–682.

[4] Eronen, V-P., Mäkelä, M.M., and Westerlund, T. On the generalization of ECP and OA methods to nonsmooth convex MINLP problems. *Optimization* (2012), doi:10.1080/02331934.2012.712118.

[5] Fletcher, R., and Leyffer, S. Solving mixed integer nonlinear programs by outer approximation. *Mathematical Programming 66*, (1994), 327–349.

[6] Fletcher, R., and Leyffer, S. *Numerical Experience with Lower Bounds for MIQP Branch-and-Bound.* University of Dundee, Numerical analysis report, NA/151, (1995).

[7] Jain, V., and Grossmann, I. Cyclic scheduling of continuous parallel-process units with decaying performance. *AIChE Journal 44*, (1999), 1623–1636.

[8] Komlosi, S. Generalized monotonicity and generalized convexity. *Journal of Optimization Theory and Applications 84*, (1995), 361–376.

[9] Lemaréchal, C. *Nondifferentiable Optimization*, in Optimization, Nemhauser, G. L. and Rinnooy Kan, A. H. G. and Todd, M. J (eds.), Elsevier North-Holland, Inc., New York, 1989.

[10] Mäkelä, M. M., and Neittaanmäki P. *Nonsmooth Optimization: Analysis and Algorithm with Applications to Optimal Control.* World Scientific Publications Publishing Co., Singapore, 1992.

[11] Mäkelä, M. M., Karmitsa, N. and Eronen, V-P. *On generalized pseudo- and quasiconvexities for nonsmooth functions.* TUCS report 989, Turku Centre for Computer Science (2011).

[12] Quesada, I., and Grossmann, I. E. An LP/NLP based branch-and-bound algorithm for convex MINLP optimization problems. *Computers & Chemical Engineering 16*, (1992), 937–947.

[13] Shor, N. Z. *Minimization Methods for Non-Differentiable Functions.* Springer-Verlag, Berlin, 1985.

[14] Still, C., and Westerlund, T. *Extended cutting plane algorithm*, in Encyclopedia of Optimization, C. A. Floudas and P.M. Pardalos (eds.), Kluwer Academic Publishers, Dordrecht 2001.

[15] Westerlund, T., and Petterson, F. An extended cutting plane method for solving convex MINLP problems. *Computers & Chemical Engineering 19*, 11 (1995), 131–136.

[16] Westerlund T., and Pörn R. Solving pseudo-convex mixed integer optimization problems by cutting plane techniques. *Optimization and Engineering 3*, 3 (2002), 253–280.

[17] Yuan, X., Pibouleau, L., and Domenech, S. Experiments in process synthesis via mixed-integer programming. *Chemical Engineering and processing 25*, 2 (1989), 99–116.

A Example problem 3

Next we will briefly present the problem in the example 3. For more detailed description we refer to [7]. The decision variables of the problem are

$$
\begin{aligned}
t_{il} &= \text{Total processing time of feed } i \text{ in furnace } l. \\
n_{il} &= \text{Number of subcycles of feed } i \text{ in furnace } l. \\
\Delta t_{il} &= \text{The total time devoted to feed } i \text{ in furnace } l. \\
T_{cycle} &= \text{The common cycle time for all the furnaces.}
\end{aligned}
$$

Variables S_i and y_{ilk} are auxiliary variables. The problem is

$$
\begin{aligned}
\min \quad & \max_l \frac{1}{T_{cycle}} \sum_{i=1}^{7} (Cs_{il} n_{il} - P_{il} D_{il} c_{il} t_{il} + \frac{1}{b_{il}} P_{il} D_{il} a_{il} n_{il} (e^{-b_i \frac{t_{il}}{n_{il}}} - 1)) \\
\text{s.t.} \quad & Flo_i T_{cycle} + S_i = \sum_{l=1}^{4} D_{il} t_{il} \quad \forall i = 1, 2, \ldots, 7 \\
& S_i \leq (Fup_i - Flo_i) T_{cycle} \quad \forall i = 1, 2, \ldots, 7 \\
& n_{il} = \sum_{k=\varepsilon,1,2,\ldots,K} k y_{ilk} \quad \forall i = 1, 2, \ldots, 7,\ l = 1, 2, 3, 4 \\
& \sum_{k=\varepsilon,1,2,\ldots,K} y_{ilk} = 1 \quad \forall i = 1, 2, \ldots, 7,\ l = 1, 2, 3, 4 \\
& \Delta t_{il} = n_{il} \tau_{il} + t_{il} \quad \forall i = 1, 2, \ldots, 7,\ l = 1, 2, 3, 4 \\
& \sum_{i=1}^{7} \Delta t_{il} \leq T_{cycle} \quad \forall l = 1, 2, 3, 4 \\
& t_{il} \leq 40(1 - y_{il\varepsilon}) \quad \forall i = 1, 2, \ldots, 7,\ l = 1, 2, 3, 4 \\
& \sum_{l=1}^{4} n_{il} \geq 1 \quad \forall i = 1, 2, \ldots, 7 \\
& \varepsilon \leq n_{il} \leq K,\ \Delta t_{il} \geq 0,\ t_{il} \geq 0 \quad \forall i = 1, 2, \ldots, 7,\ l = 1, 2, 3, 4 \\
& 35 \leq T_{cycle} \leq 40,\ S_i \geq 0, \quad \forall i = 1, 2, \ldots, 7, \\
& y_{ilk} \in \{0, 1\} \quad \forall i = 1, 2, \ldots, 7,\ l = 1, 2, 3, 4,\ k = \varepsilon, 1, 2, \ldots, K.
\end{aligned}
$$

Note that the only nonlinear part of the problem is the f°-pseudoconvex objective function. In the problem we used the values $K = 4$ and $\varepsilon = 0.01$. The values for parameters $\tau_{il}, D_{il}, a_{il}, b_{il}, c_{il}, P_{il}, Cs_{il}, Flo_i$ and Fup_i can be found in table 2 on the next page.

Table 2: Parameter values of example problem 3

	Furnace	Feed i						
Parameter	l	A	B	C	D	E	F	G
τ_{il}(d)	1	2	3	3	3	1	2	3
	2	3	1	2	2	2	1	1
	3	1	3	1	1	2	1	2
	4	2	1	3	2	2	1	1
D_{il}(ton/d)	1	1300	1200	1100	800	1300	300	700
	2	1100	1050	1000	1000	1200	400	600
	3	900	800	800	1200	1000	300	850
	4	1200	1000	800	700	1200	400	600
a_{il}	1	0.30	0.40	0.35	0.32	0.29	0.35	0.31
	2	0.32	0.38	0.33	0.31	0.28	0.40	0.34
	3	0.31	0.35	0.36	0.36	0.29	0.37	0.31
	4	0.31	0.36	0.35	0.36	0.28	0.39	0.32
b_{il}(1/d)	1	0.10	0.20	0.10	0.20	0.23	0.34	0.20
	2	0.20	0.10	0.20	0.25	0.29	0.27	0.30
	3	0.30	0.20	0.30	0.27	0.28	0.29	0.25
	4	0.20	0.20	0.15	0.25	0.29	0.22	0.28
c_{il}	1	0.20	0.18	0.21	0.20	0.30	0.26	0.16
	2	0.21	0.19	0.23	0.25	0.31	0.27	0.17
	3	0.19	0.18	0.21	0.23	0.30	0.25	0.18
	4	0.20	0.19	0.21	0.24	0.31	0.26	0.17
P_{il}(\$/ton)	1	123	105	110	123	105	110	120
	2	114	132	129	114	132	129	113
	3	110	122	120	110	122	120	117
	4	120	125	129	115	115	128	115
Cs_{il}(\$)	1	100	90	80	75	90	93	78
	2	80	85	75	90	94	78	70
	3	90	90	90	85	93	92	75
	4	80	90	85	80	92	85	72
Flo_i(ton/d)	–	300	400	300	500	500	100	600
Fup_i(ton/d)	–	600	700	600	800	800	400	900

Turku Centre *for* Computer Science

Lemminkäisenkatu 14 A, 20520 Turku, Finland | www.tucs.fi

University of Turku

- Department of Information Technology
- Department of Mathematics

Åbo Akademi University

- Department of Computer Science
- Institute for Advanced Management Systems Research

Turku School of Economics and Business Administration

- Institute of Information Systems Sciences

ISBN 978-952-12-2769-1
ISSN 1239-1891