Ville-Pekka Eronen | Marko M. Mäkelä | Tapio Wester-
lund

On Convex Nonsmooth MINLP

Turku Centre for Computer Science

TUCS Technical Report
No 988, November 2010

On Convex Nonsmooth MINLP

Ville-Pekka Eronen
University of Turku, Department of Mathematics
FI-20014 Turku, Finland
vpoero@utu.fi
Marko M. Mäkelä
University of Turku, Department of Mathematics
FI-20014 Turku, Finland
makela@utu.fi
Tapio Westerlund
Åbo Akademi University, Process Design and Systems Engineering Laboratory
Biskopsgatan 8, FI-20500 bo, Finland
tapio.westerlund@abo.fi

TUCS Technical Report
No 988, November 2010

Abstract

In this article generalization of some MINLP algorithms to cover convex nonsmooth problems is studied. In the extended cutting plane method, gradients are replaced by the subgradients of the convex function and the resulting algorithm shall be proven to converge to a global optimum. It is shown through a counterexample that this type of generalization is insufficient with certain versions of the outer approximation algorithm. However, with some modifications to the OA method a special type of nonsmooth functions for which the subdifferential at any point is a convex combination of a finite number of subgradients at the point can be considered.

Keywords: Convex nonsmooth MINLP; Convex programming; Extended cutting plane algorithm; MINLP; Nonsmooth optimization; Outer approximation algorithm; Subgradient

TUCS Laboratory
TUCS Laboratory

1 Introduction

Several methods for Mixed-Integer NonLinear Programming (MINLP) problems have been developed during the past few decades and they can be divided into three main classes, namely Cutting Plane (CP) (see e.g. [16, 17]), Branch-and-Bound (BB) (see e.g. [12, 5]) and Outer Approximation (OA) (see e.g. [3, 4, 18]) type of methods .

At the same time, NonSmooth Optimization (NSO) has also been widely studied. Different kind of subgradient (see e.g. [1, 2, 15]) and bundle methods (see e.g. [6, 8, 9, 10, 11, 13, 14]) have been introduced to solve continuous, possible nonsmooth optimization problems. However, the combination of these two optimization areas is amazingly rare, although, both bundle methods for NSO and cutting plane methods for MINLP have their origin in the same classical cutting plane method of Kelley [7]. In addition, [4] the OA method was considered with exact penalty function and some variants that are nonsmooth.

In this paper, generalizations of some MINLP algorithms to cover convex nonsmooth MINLP problems are studied. In the Extended Cutting Plane (ECP) [17] method, gradients are replaced by the subgradients of convex function and the resulting algorithm shall be proven to converge to global optimum. It is shown through a counterexample that this type of generalization is insufficient with certain versions of outer approximation algorithm. However, with some modifications to the OA method, a special type of nonsmooth functions for which subdifferentials at any point is a convex combination of finite number of subgradients at the point can be considered.

Branch-and-bound is not discussed extensively here, since the generalization is quite evident. Branch-and-bound methods solves the nonsmooth MINLP if continuous NSO problems can be solved to global optimum in each node.

The paper is organized as follows. Section 2 introduces the general problem discussed. In section 3 the generalization of the ECP method is shown to converge to a global optimum. In section 4 it is proved that the simple generalization discussed above is insufficient for the OA method. The algorithm is then modified a bit to cover special types of nonsmooth functions. Section 5 illustrates the generalizations through an example and section 6 summarizes the results.

2 The optimization problem

In this article the following MINLP problem is considered

$$\begin{array}{lll} \min & f(x,y) & \\ \text{s.t.} & g(x,y) \leq 0 & (\hat{P}) \\ & x \in X, y \in Y, & \end{array}$$

where functions $f : \mathbb{R}^n \times \mathbb{Z}^m \to \mathbb{R}$ and $g_j : \mathbb{R}^n \times \mathbb{Z}^m \to \mathbb{R}, j = 1, \dots, J$ are convex, possibly nonsmooth functions, X is a compact, convex polyhedral set

$$X = \{x \mid Ax \leq b\},$$

where A is $(p \times n)$ matrix and b is $(p \times 1)$ matrix and Y is a finite set in $\mathbb{Z}^m$. It is assumed that the functions g_j are nonlinear. If there are linear constraints with both continuous and integer variables those would be treated in sequel quite similarly as the constraints in the set X.

The aim of this paper is to generalize the well-known ECP [17] and OA [4] methods for MINLP problems to cover also nonsmooth problems of form $(\hat{P})$.

Nonsmoothness implies that gradients may not exist at every point in the set $X \times Y$. To deal with this problem subgradients for convex funtions are used instead.

Definition 2.1. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. The *subgradient* of the function f at the point x^0 is any vector $\xi(x^0)$ that satisfies the condition

$$f(x^0) + \xi(x^0)^T(x - x^0) \leq f(x), \quad \text{for all } x \in \mathbb{R}^n. \tag{1}$$

The set of all vectors $\xi(x^0)$ satisfying condition (1) is called *subdifferential* at the point x^0 and it is denoted by $\partial f(x^0)$.

It can be proven that the subdifferential of a convex function is a nonempty, convex and compact set. Also, a gradient of a differentiable convex function is always a subgradient [11].

In our attempt to generalize the ECP and OA methods, gradients are replaced by subgradients. In the following, we shall prove that generalization will be successful with the ECP method presented in [17] but not with the OA method presented in [4]. However, after some modifications to the OA method it will also cover problems of form $(\hat{P})$ with some further assumptions on object and constraint functions. Next, some useful definitions and lemmas needed in the convergence proofs are presented.

Definition 2.2. Function $f : \mathbb{R}^n \to \mathbb{R}$ is *locally Lipschitz continuous* at point x^0, if there exist $\epsilon > 0$ and K such that the inequality

$$|f(y) - f(z)| \leq K \|y - z\| \tag{2}$$

holds for all $y, z \in B(x^0, \epsilon)$. Here K is called a local Lipschitz constant of function f at point x^0 and $B(x^0, \epsilon)$ is an open ball with a center x^0 and radius ϵ.

Note that convex functions are locally Lipschitz continuous [11].

Lemma 2.1. Let K be a local Lipschitz constant of a convex function f at point x^0. Then, the inequality

$$\left\|\xi(x^0)\right\| \leq K \tag{3}$$

holds for all $\xi(x^0) \in \partial f(x^0)$.

Proof: See [11] page 14. □

A convex combination of sets $A_1, \ldots, A_I$ is denoted by
conv $\{A_1, \ldots, A_I\} :=$

$$\left\{\lambda_1 a_1 + \cdots + \lambda_I a_I \mid a_i \in A_i,\ \sum_{i=1}^{I} \lambda_i = 1,\ \lambda_i \geq 0,\ i = 1, \ldots, I\right\}.$$

Lemma 2.2. If functions f_i, $i = 1, \ldots, I$ are convex then the function

$$f(x) := \max_{i=1,\ldots,I} \{f_i(x)\}$$

is also convex and

$$\partial f(x) = \text{conv}\left\{\partial f_i(x) \mid i \in \mathcal{I}(x)\right\},$$

where $\mathcal{I}(x) = \{i \mid f_i(x) = f(x)\}$.

Proof: See [11] page 47. □

3 Generalization of ECP method

In the ECP method [17] the objective function of problem $(\hat{P})$ is turned into a constraint by introducing a constraint

$$g_{J+1}(x, y, \mu) := f(x, y) - \mu \leq 0$$

and minimizing variable μ. Function g_{J+1} is convex since it is a sum of two convex functions $f(x, y)$ and $-\mu$. Also, the scalars μ_{min} and μ_{max} such that the inequalities

$$\mu_{min} \leq f(x, y) \leq \mu_{max} \tag{4}$$

hold for all $x \in X$ and $y \in Y$ for which $g_j(x, y) \leq 0$, $j = 1, \ldots, J$, should be known a priori. This is necessary in order to keep the decision space compact and to guarantee the convergence to a global optimum.

Introducing variable $z = (x, y, \mu) \in \mathbb{R}^n \times \mathbb{Z}^m \times \mathbb{R}$ will simplify notations. Now, the constraint functions g_j are treated as functions of z. The usage of the function

$$\tilde{g}(z) := \max_{j=1,\ldots,J+1} \{g_j(z)\}$$

will facilitate the proof. According to lemma 2.2 $\tilde{g}$ is convex and

$$\partial \tilde{g}(z) = \text{conv}\left\{ \partial g_i(z) \mid g_i(z) = \tilde{g}(z) \right\}.$$

If we denote $L := X \times Y \times [\mu_{min}, \mu_{max}]$ and $N := \{z \mid \tilde{g}(z) \leq 0\}$ then the problem $(\hat{P})$ can be stated as follows

$$\begin{array}{ll} \min & z_{m+n+1} \\ \text{s.t.} & z \in N \cap L. \qquad (P) \end{array}$$

Theorem 3.1. Point $(\hat{x}, \hat{y}, \hat{z}_{m+n+1})$ is a global optimum of problem (P) iff the point $(\hat{x}, \hat{y})$ is a global optimum of problem $(\hat{P})$.

Proof: Due to box constraints (4) and constraints of problems $(\hat{P})$ and (P) a point (x, y) is feasible in problem $(\hat{P})$ iff $(x, y, f(x, y))$ is feasible in problem (P). If $(\tilde{x}, \tilde{y}, \tilde{z}_{m+n+1})$ is a global optimum of problem (P) then, $\tilde{z}_{m+n+1} = f(\tilde{x}, \tilde{y})$ due to constraint $f - z_{m+n+1} \leq 0$. Now, assume that $(\hat{x}, \hat{y})$ is a global optimum of the problem $(\hat{P})$ and $(\tilde{x}, \tilde{y}, f(\tilde{x}, \tilde{y}))$ is a global optimum of problem (P). Since $(\hat{x}, \hat{y}, f(\hat{x}, \hat{y}))$ is feasible in (P) the inequality $f(\tilde{x}, \tilde{y}) \leq f(\hat{x}, \hat{y})$ holds. Since the point $(\tilde{x}, \tilde{y})$ is feasible in problem $(\hat{P})$ also $f(\hat{x}, \hat{y}) \leq f(\tilde{x}, \tilde{y})$. Hence, $f(\hat{x}, \hat{y}) = f(\tilde{x}, \tilde{y})$ and the theorem is proven. □

In the ECP method, MILP problem where nonlinear constraint $\tilde{g}$ is linearized at previous solution points, is solved in each step. In the beginning the nonlinear constraints are left out and the problem

$$\begin{array}{ll} \min & z_{m+n+1} \\ & z \in L \qquad (P^0) \end{array}$$

is solved. Then, at iteration $i > 0$ we solve the problem

$$\begin{array}{ll} \min & z_{m+n+1} \\ \text{s.t.} & \tilde{g}(z^k) + \tilde{\xi}(z^k)^T (z - z^k) \leq 1,\ k = 0, \ldots, i-1 \quad (P^i) \\ & z \in L, \end{array}$$

where z^k is the solution of problem (P^k). The algorithm goes as follows:

ECP Algorithm 3.1.

Step 0. Set $i = 0$, give some small feasibility tolerance $\epsilon_g > 0$ and create problem (P^0).

Step 1. Solve MILP problem (P^i). Let the solution point be z^i.

Step 2. Check whether z^i satisfies the nonlinear constraints that is $\tilde{g}(z^i) \leq \epsilon_g$. If the nonlinear constraints are satisfied stop, point z^i is a global optimum of the problem (P).

Step 3. Create a new problem (P^{i+1}) by adding the constraint $l^i(z) = \tilde{g}(z^i) + \tilde{\xi}(z^i)^T(z - z^i) \leq 0$, where $\tilde{\xi}(z^i) \in \partial\tilde{g}(z^i)$ is arbitrary.

Step 4. Set $i = i + 1$ and go to step 1.

Note, that in step 1 the optimization problem can be solved by using any suitable MILP solver.

Next, the convergence proof of the algorithm to a global optimum is given. The proof necessitates that $\epsilon_g = 0$. First, we show that cutting planes introduced in the algorithm does not cut off any part of the feasible set of the problem (P). Let z^i be a solution point of the problem (P^i). Due to convexity of the function $\tilde{g}$ the inequality

$$l^i(z) = \tilde{g}(z^i) + \tilde{\xi}(z^i)^T(z - z^i) \leq \tilde{g}(z) \leq 0 \tag{5}$$

holds for all $z \in N \cap L$. Thus, any feasible point of (P) remains feasible in MILP problems. If we denote the feasible set of problem (P^i) by

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

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

$$N \cap L \subseteq \cdots \subseteq \Omega^i \subseteq \cdots \subseteq \Omega^0. \tag{6}$$

If the algorithm ends with finite amount of iterations say at i, the last solution point is feasible in problem (P). This point is also a global optimum of the problem (P), since the point is optimum at the feasible set of problem (P^i) and this set includes the feasible set of problem (P) according to (6).

Next, we assume that the algorithm doesn't stop after finite amount of iterations. Since $l^i(z^i) = \tilde{g}(z^i) > 0$ the constraint (5) makes the point z^i infeasible in subsequent MILP problems. Hence, the algorithm generates a sequence of different points. The solution sequence belongs to the compact set L. Since L is bounded, the classical Bolzano-Weierstrass theorem implies that the sequence has an accumulation point. It remains to prove that the accumulation point is feasible in (P).

Lemma 3.1. An accumulation point generated by the ECP algorithm in problem (P) is a feasible point.

Proof: Let the accumulation point be $\hat{z}$. Since the sequence (z^i) belongs to the closed set L, also $\hat{z} \in L$. To prove the feasibility of $\hat{z}$ it suffices to prove that $\hat{z} \in N$. Let $\epsilon > 0$ be arbitrary and $\delta > 0$ be so small that local Lipschitz condition (2) with Lipschitz constant K holds for $\tilde{g}$ in $B(\hat{z}, \delta)$. Denote $\hat{\epsilon} = \max\left\{\frac{\epsilon}{3K}, \delta\right\}$. Since $\hat{z}$ is the accumulation point, there exist indices $k > j$ such that for MILP solution points z^j and z^k the relation $z^j, z^k \in B(\hat{z}, \hat{\epsilon})$ holds. Since the algorithm does not terminate after a finite amount of iterations, points in the solution sequence (z^i) are in the closed set $\{z \mid \tilde{g}(z) \geq 0\}$. The accumulation point $\hat{z}$ also belongs to this set, because the set is closed. Since $k > j$, the inequality $l^j(z^k) \leq 0$ holds, and thus the inequality

$$\tilde{g}(\hat{z}) \leq \left|\tilde{g}(\hat{z}) - l^j(z^k)\right| = \left|\tilde{g}(\hat{z}) - \tilde{g}(z^j) - \tilde{\xi}(z^j)^T(z^k - z^j)\right|$$

is also true. From the triangle and Cauchy-Schwarz inequalities it follows that

$$\begin{aligned} & \left|\tilde{g}(\hat{z}) - \tilde{g}(z^j) - \tilde{\xi}(z^j)^T(z^k - z^j)\right| \\ \leq \quad & \left|\tilde{g}(\hat{z}) - \tilde{g}(z^j)\right| + \left\|\tilde{\xi}(z^j)\right\| \left\|z^k - z^j\right\|. \end{aligned}$$

Then the local Lipschitz continuity and lemma 2.1 implies

$$\begin{aligned} & \left|\tilde{g}(\hat{z}) - \tilde{g}(z^j)\right| + \left\|\tilde{\xi}(z^j)\right\| \left\|z^k - z^j - \hat{z} + \hat{z}\right\| \\ \leq \quad & K\left\|\hat{z} - z^j\right\| + K\left\|z^k - \hat{z}\right\| + K\left\|\hat{z} - z^j\right\| \\ < \quad & K\frac{\epsilon}{3K} + K\frac{\epsilon}{3K} + K\frac{\epsilon}{3K} = \epsilon. \end{aligned}$$

Thus, $\tilde{g}(\hat{z}) = 0$ proving the lemma. □

We sum up the convergence result in the following theorem.

Theorem 3.2. ECP algorithm 3.1 with $\epsilon_g = 0$ converges to a global optimum when solving problem (P).

Proof: As stated earlier, the added cutting planes does not cut off any part of the feasible set of problem (P). Thus, the feasible set of problem (P) is a subset of the feasible set of any occuring MILP problem. If the solution of the MILP problem is feasible in (P) it is also a global optimum. If the algorithm stops after finite amount of iterations a feasible point will be obtained. If the algorithm generates an infinite sequence it will have an accumulation point that satisfies the constraints according to lemma 3.1. Hence, the algorithm converges to a global optimum. □

It is good to note that there can be many accumulation points, but all the accumulation points are, possibly different, global optima.

As a consequence of the above results, the ECP method presented in [17] can be applied also to nonsmooth problems without any modifications.

4 About the generalization of outer approximation

Next, the classical outer approximation method presented in [4] is considered. It is assumed here that the objective function and the constraint functions of the problem $(\hat{P})$ are continuously differentiable. A short presentation of the algorithm follows.

The OA method proceeds by alternating between a continuous nonlinear programming problem and a MILP problem. The nonlinear problem corresponds to problem $(\hat{P})$ with fixed integer value $y = y^i$, that is, the problem

$$\begin{array}{lll} \min & f(x, y^i) & \\ \text{s.t.} & g_j(x, y^i) \leq 0,\ j = 1, \dots, J & \qquad (NLP_{y^i}) \\ & x \in X & \end{array}$$

is solved at iteration i. If (NLP_{y^i}) is feasible, an upper bound UB^i is obtained for the objective function. The best upper bound is stored for the MILP problem as

$$UBD^i = \min \left\{ UB^j \mid j \leq i \text{ and } (NLP_{y^j}) \text{ is feasible} \right\}.$$

If (NLP_{y^i}) is infeasible a feasibility problem (F_{y^i}) is solved instead. A special case of feasibility problems presented in [4] is

$$\begin{array}{lll} \min & \mu & \\ \text{s.t.} & g_j(x, y^i) \leq \mu,\ j = 1, \dots, J & \qquad (F_{y^i}) \\ & x \in X. & \end{array}$$

Indices for which (NLP_{y^i}) is feasible is stored in set T^i:

$$T^i = \left\{ j \mid j \leq i,\ (NLP_{y^j}) \text{ was feasible} \right\}.$$

Similarly, S^i stores indices where (NLP_{y^i}) was infeasible:

$$S^i = \left\{ j \mid j \leq i,\ (NLP_{y^j}) \text{ was infeasible} \right\}.$$

Let x^i be the solution of (NPL_{y^i}) or (F_{y^i}). The MILP problem is then formulated as follows:

$$\begin{array}{lll} \min & \mu & \\ \text{s.t.} & \mu < UBD^i & \\ & f(x^i, y^i) + \nabla f(x^i, y^i)^T \begin{pmatrix} x - x^i \\ y - y^i \end{pmatrix} \leq \mu\ \forall i \in T^i & \qquad (M^i) \\ & g_j(x^i, y^i) + \nabla g_j(x^i, y^i)^T \begin{pmatrix} x - x^i \\ y - y^i \end{pmatrix} \leq 0\ \forall i \in T^i \cup S^i,\ j = 1, \dots, J & \\ & x \in X, y \in Y. & \end{array}$$

Let the solution of (M^i) be (x^{i+1}, y^{i+1}). Next, the problem $(NLP_{y^{i+1}})$ or, if infeasible, the problem $(F_{y^{i+1}})$ is solved. The OA algorithm proceeds as follows:

Algorithm 4.1. (Fletcher & Leyffer)
Step 0. Let $y^0 \in \mathbb{Z}^m$ be given. Set $i = 0$, $T^{-1} = \emptyset$, $S^{-1} = \emptyset$ and $UBD^{-1} = \infty$.

Step 1. Solve the problem (NLP_{y^i}), or the feasibility problem (F_{y^i}), if (NLP_{y^i}) is infeasible, and let the solution be x^i.

Step 2. Linearize the objective and constraint functions about (x^i, y^i). Set $T^i = T^{i-1} \cup \{i\}$ or $S^i = S^{i-1} \cup \{i\}$ as appropriate.

Step 3. If (NLP_{y^i}) is feasible and $f(x^i, y^i) < UBD^{i-1}$ then update current best point by setting $x^* = x^i$, $y^* = y^i$ and $UBD^i = f(x^i, y^i)$.

Step 4. Solve the problem (M^i), giving a new integer assignment y^{i+1} to be tested in the algorithm. If (M^i) is infeasible stop: the solution is (x^*, y^*), else set $i = i + 1$ and go to step 1.

In addition to the assumptions of problem $(\hat{P})$, the constraints are supposed to satisfy the Slater constraint qualification condition in problems (NLP_{y^i}) and (F_{y^i}) for every y^i. This means that for every y^i there should exist a point $x \in X$ such that $g_j(x, y^i) < 0$ for all $j = 1, \ldots, J$. This requirement enables us to use KKT-conditions.

The proof that OA method converges is based on the following three steps.

Lemma 4.1. If (NLP_{y^i}) is infeasible and (x^i, y^i) is the solution to (F_{y^i}) problem, then $y = y^i$ is infeasible in the constraints

$$g_j(x^i, y^i) + \nabla g_j(x^i, y^i)^T \begin{pmatrix} x - x^i \\ y - y^i \end{pmatrix} \leq 0,\ j = 1, \ldots, J$$

for all $x \in X$.

Proof: See [4] lemma 1. □

Lemma 4.2. If (NLP_{y^i}) is feasible and (x^i, y^i) is the solution point, then $y = y^i$ is infeasible in the constraints

$$\begin{aligned}
&\mu < UBD^i \\
&f(x^i, y^i) + \nabla f(x^i, y^i)^T \begin{pmatrix} x - x^i \\ y - y^i \end{pmatrix} \leq \mu \\
&g_j(x^i, y^i) + \nabla g_j(x^i, y^i)^T \begin{pmatrix} x - x^i \\ y - y^i \end{pmatrix} \leq 0,\ j = 1, \ldots, J
\end{aligned}$$

for all $x \in X$.

Proof: This proof can be found in [4] as a part of the proof of theorem 2. □

Two previous lemmas show that the integer part of the solution of the feasible (M^i) problem is different from the previous solutions. Since Y is finite the algorithm ends up with an infeasible (M^i) problem. Now, it remains to prove that when (M^i) is infeasible the upper bound UBD^i is a global optimum. This follows from the fact that linearizations in (M^i) do not cut off any feasible point of problem (P) if the objective function f is smaller than UBD^i at the feasible point. This in turn follows from the convexity of the objective and constraint functions, specially, from inequality (1) for gradients. The proof can be found in [4] (theorem 2).

4.1 Counterexample

Next, we shall show that substitution of the gradient by an arbitrary subgradient in outer approximation is insufficient. Let

$$g(x,y) = \max\left\{-\frac{3}{2} - x + y, -\frac{7}{2} + y + x\right\}$$

be the convex, nonsmooth constraint function and consider the following problem

$$\begin{aligned} \min \quad & 2x - y \\ \text{s.t.} \quad & g(x,y) \le 0 \\ & y - 4x - 1 \le 0 \qquad\qquad (E) \\ & 0 \le x \le 2, y \in Y, \end{aligned}$$

where $Y = \{0,1,2,3,4,5\}$. Let the initial point be $y^0 = 3$.

There is no feasible points in problem (NLP_{y^0}) and thus problem (F_{y^0})

$$\begin{aligned} \min \quad & \mu \\ \text{s.t.} \quad & \max\left\{\frac{3}{2} - x, -\frac{1}{2} + x\right\} \le \mu \\ & 2 - 4x \le 0 \\ & 0 \le x \le 2 \end{aligned}$$

will be solved. The solution is $x^0 = 1$ with $\mu = \frac{1}{2}$. Then g will be linearized at point $(x^0, y^0) = (1,3)$ for the next MILP problem. Both of the functions $-\frac{3}{2} - x + y$ and $-\frac{7}{2} + y + x$ has the same value $\frac{1}{2}$ at the point (x^0, y^0) and thus according to lemma 2.2 the subdifferential is

$$\partial g(1,3) = \left\{(\alpha, 1)^T \mid \alpha \in [-1,1]\right\}. \tag{7}$$

Since we may select an arbitrary subgradient we can choose for example $\xi(z^0) = (1,1)^T$. Thus, the new linear constraint is

$$\frac{1}{2} + (1,1)(x-1, y-3)^T \le 0 \Rightarrow x + y - \frac{7}{2} \le 0.$$

The resulting problem (M^0) is

$$\begin{aligned} \min \quad & 2x - y \\ \text{s.t.} \quad & x + y - \frac{7}{2} \leq 0 \\ & y - 4x - 1 \leq 0 \\ & 0 \leq x \leq 2, y \in Y. \end{aligned}$$

The solution point of problem (M^0) is $(\frac{1}{2}, 3)$. Here $y^1 = y^0$ and $\mathrm{F}_{y^1} \equiv \mathrm{F}_{y^0}$, thus outer approximation may generate an infinite loop between points $(1, 3)$ and $(\frac{1}{2}, 3)$. Both of them are infeasible, but the problem (E) has a feasible point $(0, 1)$ for example. Thus, outer approximation with subgradients could not find an optimum.

4.2 Modified OA method

If the integer variables of the solutions obtained in problems (M^i) are explicitly cut off with integer cuts from the following problems (M^i) then the generalization would be sufficient. This follows from the inequality (1) which ensures that feasible solutions for which f is smaller than UBD^i are not cut off in problem (M^i). However, as pointed out in [4], integer cuts are practical only with binary variables.

When tracking down why properties proven for continuously differentiable functions in [4] are not valid here, the reason was noticed to be the KKT-conditions.

We know that for differentiable functions the KKT-conditions for (F_{y^i}) problem are the equations

$$\nabla \mu + \sum_{j=1}^{J} \lambda_j \nabla (g_j(x, y^i) - \mu) + \eta = 0 \tag{8}$$

$$\lambda_j (g_j(x, y^i) - \mu) = 0, \; j = 1, \ldots, J \tag{9}$$

$$\lambda_j \geq 0, \; j = 1, \ldots, J,$$

where $\eta \in N_X(x)$ and $N_X(x)$ is the normal cone of the set X at the point x. Since X is a convex set the inequality

$$\eta^T (x^0 - x) \leq 0 \tag{10}$$

holds for all $x^0 \in X$ and $\eta \in N_X(x)$. For nonsmooth functions the relation corresponding to (8) reads

$$0 \in \partial \mu + \sum_{j=1}^{J} \lambda_j \partial (g_j(x, y^i) - \mu) + N_X(x). \tag{11}$$

Specially, for arbitrary subgradients, equality like (8) does not hold. If we could always find such subgradients that the equality holds for (11) also the proof in [4] would hold and outer approximation would converge also in the nonsmooth case.

A special class of nonsmooth functions can be dealt with in the following way. Consider a nonsmooth function f for which at any point the subdifferential can be generated by a convex combination of finite amount of subgradients. That is, for a point z there exist subgradients $\xi^1(z), \ldots, \xi^K(z)$ such that

$$\partial f(z) = \left\{ \alpha_k \xi^k(z) \mid \sum_{k=1}^{K} \alpha_k = 1,\ \alpha_k \geq 0,\ k = 1, \ldots, K \right\}. \tag{12}$$

An example of such a function is the maximum of a finite number of convex differentiable functions [11]. It should be noted that this type of functions can be dealt with also with the exact penalty function formulation used in [4]. In the generalization of the OA method to work with this kind of functions possibly many linearizations are made instead of one in step 2 of algorithm 4.1. For example, consider the function f that should be linearized at point (x^i, y^i) for the next MILP problem (M^i) and its subdifferential, which is of the form (12). Then, new linearizations for (M^i) would be

$$f(x^i, y^i) + \xi^1(x^i, y^i)^T \begin{pmatrix} x - x^i \\ y - y^i \end{pmatrix} \leq 0$$

$$\vdots$$

$$f(x^i, y^i) + \xi^K(x^i, y^i)^T \begin{pmatrix} x - x^i \\ y - y^i \end{pmatrix} \leq 0.$$

For a certain linearized function subgradients in different linearizations must be such that the subdifferential can be constructed with a convex combination of those subgradients.

Next, we shall prove that with this generalization lemma 4.1 holds.

Lemma 4.3. If (NLP_{y^i}) is infeasible and (x^i, y^i) is the solution to problem (F_{y^i}), then $y = y^i$ is infeasible in the constraints

$$g_j(x^i, y^i) + \xi_j^k(x^i, y^i)^T \begin{pmatrix} x - x^i \\ y - y^i \end{pmatrix} \leq 0,\ j = 1, \ldots, J,\ k = 1, \ldots, K^{ij}, \tag{13}$$

where $\xi_j^1, \ldots \xi_j^{K^{ij}}$ are such subgradients that any vector from the subdifferential $\partial g_j(x^i, y^i)$ is a convex combination of those subgradients.

Proof: Suppose as a contrary, that y^i is feasible in constraints (13). Since x^i is optimal in (F_{y^i}), the KKT-conditions hold and

$$0 \in \partial\mu + \sum_{j=1}^{J} \lambda_j \partial(g_j(x^i, y^i) - \mu) + N_X(x^i).$$

From the component corresponding to μ we see that

$$\sum_{j=1}^{J} \lambda_j = 1. \tag{14}$$

From the other components we see that there exist appropriate subgradients $\xi_j(x^i, y^i)$ and a normal vector η such that the equation

$$0 = \sum_{j=1}^{J} \lambda_j \xi_j(x^i, y^i) + \eta \tag{15}$$

holds. For all j we can choose constants α_{jk} such that

$$\begin{aligned} &\xi_j(x^i, y^i) = \sum_{k=1}^{K^{ij}} \alpha_{jk} \xi_j^k(x^i, y^i) \\ &\sum_{k=1}^{K^{ij}} \alpha_{jk} = 1,\ j = 1, \ldots, J \\ &\alpha_{jk} \geq 0,\ j = 1, \ldots, J,\ k = 1, \ldots, K^{ij}. \end{aligned}$$

Multiplying all the inequalities in (13) with corresponding constants α_{jk}, inserting $y = y^i$ and summing over k we obtain

$$\begin{aligned} &\sum_{k=1}^{K^{ij}} \alpha_{jk} g_j(x^i, y^i) + \sum_{k=1}^{K^{ij}} \alpha_{jk} \xi_j^k(x^i, y^i)^T \begin{pmatrix} x - x^i \\ y^i - y^i \end{pmatrix} \leq 0,\ j = 1, \ldots, J \\ \Rightarrow\quad & g_j(x^i, y^i) + \xi_j(x^i, y^i)^T \begin{pmatrix} x - x^i \\ 0 \end{pmatrix} \leq 0,\ j = 1, \ldots, J. \end{aligned}$$

Multiplying these equations with the KKT-multipliers λ_j and summing them up results in equation

$$\sum_{j=1}^{J} \lambda_j g_j(x^i, y^i) + \sum_{j=1}^{J} \lambda_j \xi_j(x^i, y^i)^T \begin{pmatrix} x - x^i \\ 0 \end{pmatrix} \leq 0.$$

From the KKT-condition (9) and equation (14) we see that $\sum_j \lambda_j g_j(x^i, y^i) = \mu$. Also, since $\eta \in N_X(x^i)$ equation $\eta^T(x - x^i) \leq 0$ holds. Summing this equation and making substitution $\sum_j \lambda_j g_j(x^i, y^i) = \mu$ results in equation

$$\mu + \sum_j \lambda_j \xi_j(x^i, y^i)^T \begin{pmatrix} x - x^i \\ 0 \end{pmatrix} + \eta^T(x - x^i) \leq 0. \tag{16}$$

From equation (15) we know that

$$(\sum_j \lambda_j \xi_j(x^i, y^i) + \eta)^T(x - x^i) = 0,$$

thus inequality (16) results in $\mu \leq 0$ implying there is a feasible solution for problem (NLP_{y^i}). This is impossible since problem (F_{y^i}) was solved. Hence, the lemma is proved. □

Similarly, corresponding lemma 4.2 holds. The proof of this lemma is very similar to the previous one. Also, if a point is feasible and the objective function f is smaller than UBD^i at the point, then the point is feasible in problem (M^i). This follows from the definition of a subgradient and convexity of the objective and constraint functions. Thus, the modified OA method converges. Also, if functions are continuously differentiable this procedure reduces to the OA method in [4].

5 An example problem

Next, the example problem (E) is solved with the modified outer approximation and extended cutting plane methods. First, we continue the solving process of outer approximation from point $(1, 3)$ obtained in previous section. Now, for constraint g_1 two cutting planes are generated, one with gradient $(1, 1)^T$ and the other with $(-1, 1)^T$, resulting in the constraints

$$\begin{aligned} &x + y - \frac{7}{2} \leq 0 \quad \text{and} \\ &-x + y - \frac{3}{2} \leq 0. \end{aligned}$$

The resulted MILP problem is

$$\begin{aligned} \min \quad & 2x - y \\ \text{s.t.} \quad & x + y - \frac{7}{2} \leq 0 \\ & -x + y - \frac{3}{2} \leq 0 \\ & y - 4x - 1 \leq 0 \\ & 0 \leq x \leq 2, y \in Y, \end{aligned}$$

and the solution is $(0, 1)^T$. Problem (NLP_{y^1}) has the same optimum and (M^1) is infeasible. Thus, the algorithm ends up with the solution $(0, 1)^T$ which is a global optimum.

For the ECP method the first MILP problem is

$$\begin{aligned} \min \quad & 2x - y \\ \text{s.t.} \quad & y - 4x - 1 \leq 0 \\ & 0 \leq x \leq 2, y \in Y. \end{aligned}$$

The solution point $(1, 5)^T$ is not feasible in (E) since $g(1, 5) = \frac{5}{2} \geq 0$. Both the linear functions of g have the same value and the subdifferential is the same as in (7). According to algorithm 3.1 we may choose any subgradient from the subdifferential. We will see how algorithm proceeds if we would choose subgradients $(1, 1)$, $(0, 1)$ or $(-1, 1)$ respectively. Results are summarized in table 1.

ξ^i	i	x^i	y^i	$g(x^i, y^i)$	$2x - y$
	0	1	5	$\frac{5}{2}$	-3
(1,1)	1	$\frac{1}{2}$	3	1	$-\frac{3}{2}$
	2	0	1	$-\frac{1}{2}$	-1
(0,1)	1	$\frac{1}{4}$	2	1	$-\frac{7}{4}$
	2	0	1	$-\frac{1}{2}$	-1
(-1,1)	1	0	1	$-\frac{1}{2}$	-1

Table 1: Quantities in the ECP algorithm with different choices of subgradient

As can be seen from table 1, the ECP method converged to optimal point in this example with all chosen subgradients. However, different choices of subgradients resulted in different amount of iterations. The global optimum was found fastest with the subgradient $(-1, 1)^T$.

6 Conclusions

The ECP and OA methods with nonsmooth MINLP problems were studied. The ECP method could solve the problem if subgradients are used in linearizations when the gradient does not exist. An example showed that different choices of subgradients may affect how fast the global optimum is found. The OA method presented in [4] could not solve the problem generally, but with slight modifications to the algorithm a special class of nonsmooth functions could be dealt with.

7 Acknowledgements

This research was supported by the grant # 127992 of the Academy of Finland.

References

[1] BECK, A., AND TEBOULLE, M. Mirror descent and nonlinear projected subgradient methods for convex optimization. *Operations Research Letters 31*, 3 (2003), 167–175.

[2] BEN-TAL, A., AND NEMIROVSKI, A. Non-Euclidean restricted memory level method for large-scale convex optimization. *Mathematical Programming 102*, 3 (2005), 407–456.

[3] 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.

[4] FLETCHER, R., AND LEYFFER, S. Solving mixed integer nonlinear programs by outer approximation. *Mathematical programming 66*, (1994), 327–349.

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

[6] GAUDIOSO, M., AND MONACO, M. F. Variants to the cutting plane approach for convex nondifferentiable optimization. *Optimization 25*, (1992), 65–75.

[7] KELLEY, J. E. The cutting-plane method for solving convex programs. *Journal of SIAM 8*, 4 (1960), 703–712.

[8] KIWIEL, K. C. Proximity control in bundle methods for convex nondifferentiable minimization. *Mathematical Programming 46*, (1990), 105–122.

[9] LUKŠAN, L., AND VLČEK, J. Globally convergent variable metric method for convex nonsmooth unconstrained minimization. *Journal of Optimization Theory and Applications 102*, 3 (1999), 593–613.

[10] MÄKELÄ, M. M. Survey of bundle methods for nonsmooth optimization. *Optimization Methods and Software 17*, 1 (2002), 1–29.

[11] MÄKELÄ, M. M., AND NEITTAANMÄKI P. *Nonsmooth Optimization: Analysis and Algorithm with Applications to Optimal Control*. World Scientific Publications Publishing Co., Singapore, 1992.

[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] SAGASTIZÁBAL, C., AND SOLODOV, M. An infeasible bundle method for nonsmooth convex constrained optimization without a penalty function or a filter. *SIAM Journal on Optimization 16*, 1 (2005), 146–169.

[14] SCHRAMM, H., AND ZOWE, J. A version of the bundle idea for minimizing a nonsmooth function: Conceptual idea, convergence analysis, numerical results. *SIAM Journal on Optimization 2*, 1 (1992), 121–152.

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

[16] 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).

[17] WESTERLUND, T., AND PETTERSON, F. An extended cutting plane method for solving convex MINLP problems. *Computers & Chemical Engineering 19*, 11 (1995), 131–136.

[18] YUAN, X., PIBOULEAU, L., AND DOMENECH, S. Experiments in process synthesis via mixed-integer programming. *Chemical Engineering and processing 25*, 2 (1989), 99–116.

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 978-952-12-2495-9
ISSN 1239-1891