A Cutting-Plane Global Optimization Algorithm for a Special Non-Convex Problem

Artur Alves Pessoa[1*], José Miguel Aroztegui Massera[2] and Thiago José Machado[3]

[1]Departamento de Engenharia de Produção, Universidade Federal Fluminense, Niterói, Rio de Janeiro, Brazil.
[2]Departamento de Computação Científica, Universidade Federal da Paraíba, João Pessoa, Paraíba, Brazil.
[3]Departamento de Engenharia Mecânica, Universidade Federal da Paraíba, João Pessoa, Paraíba, Brazil.

*Corresponding author(s). E-mail(s): arturpessoa@id.uff.br;
Contributing authors: jose.miguel@ci.ufpb.br;
thiagao.matematica@gmail.com;

Abstract

This study establishes the convergence of a cutting-plane algorithm tailored for a specific non-convex optimization problem. The presentation begins with the problem definition, accompanied by the necessary hypotheses that substantiate the application of a cutting plane. Following this, we develop an algorithm designed to tackle the problem. Lastly, we provide a demonstration that the sequence generated by the algorithm converges to a solution to the problem. To illustrate the convergence of the algorithm visually, numerical experiments are conducted using Matlab.

Keywords: Cutting Plane Algorithm, Global optimization, Non-convex Optimization

1 Problem definition

Let $g : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ be a given function, non-convex w.r.t. the last input of g. The problem we focus on is defined as follows:

$$(\mathrm{P}): \max_{x\in\mathbb{R}^n,\lambda\in\mathbb{R}} \lambda \tag{1}$$

$$\textit{subject to}: \quad g(x,\lambda) \leqslant 0 \tag{2}$$

Let $\Omega = \{(x,\lambda) : g(x,\lambda) \leqslant 0\}$ be the feasible set of (P) and we write (x^*,λ^*) as the optimal of (P). We are going to fix a particular $\lambda_0 \in \mathbb{R}$ such that $\lambda_0 \geqslant \lambda^*$. The fundamental hypothesis about g are:
(H1) $f_{\lambda_0}(x) := g(x,\lambda_0)$ is convex.
(H2) $\exists M \geqslant 0$: $\forall(x,\lambda)$ such that $\lambda_0 \geqslant \lambda$ and $g(x,\lambda) \leqslant 0$, $f_{\lambda_0}(x) \leqslant M(\lambda_0 - \lambda)$.

In this work, g is the maximum of m scalar functions $g_1,\dots,g_m$. The constraint (2) is equivalent to the m inequalities $g_i(x,\lambda) \leqslant 0$ for all $i \in \{1,\dots,m\}$. If g_i satisfy (H1) and (H2) for a given λ_0 and $M_i \geqslant 0$, g also satisfy (H1) and (H2). In this case, the constant M in (H2) can be $M = \max\{M_1,\dots,M_m\}$. More discussions about the computation of M is given in section 3.

2 A valid cutting-plane

Let λ^* be an optimal objective function value of (P). In this section we define an hyperplane in the variables x and λ that separates an infeasible point (x^0,λ_0) with $\lambda_0 \geqslant \lambda^*$ from all feasible point of (P).

Theorem 1. *If (x^*,λ^*) be an optimal of (P), (x^0,λ_0) is such that $g(x^0,\lambda_0) > 0$, $\lambda_0 \geqslant \lambda^*$ and $s_0 \in \partial f_{\lambda_0}(x^0)$, then:*

$$g(x^0,\lambda_0) + s_0^\top(x - x^0) \leqslant M(\lambda_0 - \lambda) \tag{3}$$

is false for (x^0,λ_0) and is true for all $(x,\lambda) \in \Omega$.

Proof. Setting $x = x^0$ and $\lambda = \lambda_0$ in (3) we get $g(x^0,\lambda_0) \leqslant 0$ which is false since $g(x^0,\lambda_0) > 0$. Since (H1), $f_{\lambda_0}(x) = g(x,\lambda_0)$ is convex, then

$$p(x;x^0) := g(x^0,\lambda_0) + s_0^\top(x - x^0) = y \tag{4}$$

is an extreme support to the graph of $f_{\lambda_0}(x)$. In the special case of g smooth, $s_0 = \nabla f_{\lambda_0}(x^0)$. In such situation $y = p(x,x^0)$ is tangent hyperplane to the the graph of $f_{\lambda_0}(x)$ at $x = x^0$. Additionally, since f_{λ_0} is convex,

$$p(x;x^0) \leqslant f_{\lambda_0}(x), \quad \forall x. \tag{5}$$

Now, since $(x,\lambda) \in \Omega$, $g(x,\lambda) \leqslant 0$ and consequently, $\lambda^* \geqslant \lambda$. By hypothesis, $\lambda_0 \geqslant \lambda^*$, then $\lambda_0 \geqslant \lambda$. So, using (4), (5) and (H2), we arrive to (3). □

3 Computation of M

If M satisfy (H2) and $M \leqslant M'$, then M' also satisfy (H2). Hence, all M satisfying (H2) must verify $M^* \leqslant M$ where M^* is the solution of:

$$(P_M) : \min_{x,\lambda,M} M \tag{6}$$

$$s.t.\ g(x,\lambda) \leqslant 0 \tag{7}$$

$$g(x,\lambda_0) \leqslant M(\lambda_0 - \lambda) \tag{8}$$

where $\lambda_0 \geqslant \lambda \geqslant \lambda^*$. Problem (P_M) akin to (P), share the constraint $g(x,\lambda) \leqslant 0$. Consequently, it is observed that (P_M) do not exhibit a simpler structure than (P). Therefore, we abstain from pursuing this approach for computing M^*.

The subsequent theorem provides a valid value for M.

Theorem 2. *If g_i is a smooth scalar functions for $i \in \{1,\ldots,m\}$, $g = \max\{g_1,\ldots,g_m\}$,*

$$M := \max\left\{\left|\frac{\partial g_i}{\partial \lambda}(x,\lambda)\right| : i \in \{1,\ldots,m\},\ (x,\lambda) \in \Omega\right\} \tag{9}$$

then, $(H2)$.

Proof. Consider $\lambda' \leqslant \lambda_0$, $g(x,\lambda') \leqslant 0$, then using the fundamental theorem of calculus, for all $i = 1,\ldots,m$:

$$g_i(x,\lambda_0) = g_i(x,\lambda') + \int_{\lambda'}^{\lambda_0} \frac{\partial g_i}{\partial \lambda}(x,\lambda)d\lambda \tag{10}$$

$$\leqslant g(x,\lambda') + \int_{\lambda'}^{\lambda_0} \left|\frac{\partial g_i}{\partial \lambda}(x,\lambda)\right| d\lambda \tag{11}$$

$$\leqslant 0 + \int_{\lambda'}^{\lambda_0} M d\lambda = M(\lambda_0 - \lambda'). \tag{12}$$

Consequently, $g(x,\lambda_0) \leqslant M(\lambda_0 - \lambda')$ since $g = \max\{g_1,\ldots,g_m\}$. $\square$

4 A cutting-plane algorithm

Let S_0 be a compact polyhedron in $\mathbb{R}^n \times \mathbb{R}$ containing (x^*,λ^*) and Ω. Assume $\varepsilon > 0$ and (x^0,λ_0) be an infeasible point of (P): $g(x^0,\lambda_0) > \varepsilon$.

We are going to consider the following class of algorithms to solve (P):

To have a particular algorithm of the class, we have to choose one subgradient $s_k \in \partial_x\, g\left(x^k,\lambda_k\right)$. In the case of g is smooth, s_k is unique and equal to $\nabla_x\, g(x_k,\lambda^k)$.

Algorithm 1 Cutting-Plane Algorithm (CPA)

Require: g, S_0, M, ε

```
k ← 0
(x^k, λ_k) ← arg max {λ | (x, λ) ∈ S_k}
while g(x^k, λ_k) > ε do
    Choose s_k ∈ ∂_x g(x^k, λ_k).
    S_{k+1} ← S_k ∩ {x | g(x^k, λ_k) + s_k^⊤ (x − x^k) ⩽ M(λ_k − λ)}
    (x^{k+1}, λ_{k+1}) ← arg max {λ | (x, λ) ∈ S_{k+1}}
    k ← k + 1
end while
```

5 The convergence of CPA

To prove the convergence of the sequence generated by the Algorithm 1, we require the following additional hypothesis:

(H3) There exists a compact set $S \subset \mathbb{R}^n \times \mathbb{R}$ such that $\Omega \subset S$.

(H4) $\forall (x,\lambda) \in S$, $\forall s \in \partial_x\, g(x,\lambda)$, $\exists K > 0$ such that $\left\| \begin{bmatrix} s \\ M \end{bmatrix} \right\| \leqslant K$.

The proof of the next theorem follows the steps in the proof presented in [1].

Theorem 3. *If $(x^k, \lambda_k) \in S_k$ is such that*

$$\lambda_k = \max\{\lambda \,|\, (x,\lambda) \in S_k\}$$

where

$$S_k = S_{k-1} \cap \{(x,\lambda) \,|\, g(x^{k-1}, \lambda_{k-1}) + s_{k-1}^\top (x - x^{k-1}) \leqslant M(\lambda_{k-1} - \lambda)\}$$

then the sequence $\{(x_k, \lambda_k)\}$ contains a subsequence that converges to a point $(\overline{x}, \overline{\lambda})$ in Ω with $\overline{\lambda} \geqslant \lambda$ for all $(x,\lambda) \in \Omega$.

Proof. Let λ^* be the optimal value of (P). Since (x^k, λ_k) maximizes λ over all $(x,\lambda) \in S_k$, and $\Omega \subseteq S_k \subset S_{k-1}$, we have that $\lambda^* \leqslant \lambda_k \leqslant \lambda_{k-1}$. Thus, if the sequence $\{(x^k, \lambda_k)\}$ contains a subsequence that converges to a point $(\overline{x}, \overline{\lambda})$, in Ω, then it follows from the method of computation that $\overline{\lambda} = \lambda^*$.

To prove that $\{(x^k, \lambda_k)\}$ contains the desired subsequence we proceed as follows. Note first that since $(x^k, \lambda_k) \in S_k$, it must satisfy

$$g(x^i, \lambda_i) + s_i^\top (x^k - x^i) \leqslant M(\lambda_i - \lambda_k), \qquad i \in \{0, 1, \ldots, k-1\},$$

or equivalently

$$g(x^i, \lambda_i) \leqslant \begin{bmatrix} s_i^\top \; M \end{bmatrix} \left(\begin{bmatrix} x^i \\ \lambda_i \end{bmatrix} - \begin{bmatrix} x^k \\ \lambda_k \end{bmatrix} \right), \qquad i \in \{0, 1, \ldots, k-1\}. \tag{13}$$

If the desired convergence does not occur, then there exists $\varepsilon > 0$ independent of k such that $\varepsilon \leqslant g(x^i, \lambda_i)$, then using Cauchy-Schwarz inequality and (H4):

$$\varepsilon \leqslant g(x^i, \lambda_i) \leqslant K \left\| \begin{bmatrix} x^i \\ \lambda_i \end{bmatrix} - \begin{bmatrix} x^k \\ \lambda_k \end{bmatrix} \right\|, \quad i \in \{0, 1, \ldots, k-1\}. \tag{14}$$

Consequently, every subsequence of indices $\{k_p\}$ also satisfy:

$$\frac{\varepsilon}{K} \leqslant \left\| \begin{bmatrix} x^{k_q} \\ \lambda_{k_q} \end{bmatrix} - \begin{bmatrix} x^{k_p} \\ \lambda_{k_p} \end{bmatrix} \right\|, \quad q < p, \tag{15}$$

so that $\{(x^k, \lambda_k)\}$ does not contain a Cauchy subsequence. But this is impossible since the sequence $\{(x^k, \lambda_k)\}$ is in the compact S, as assumed in (H3). Then, $\{(x^k, \lambda_k)\}$ contains a subsequence which converges to $(\overline{x}, \overline{\lambda}) \in S$ and $\{g(x^k, \lambda_k)\}$ converges to zero, therefor $(\overline{x}, \overline{\lambda}) \in \Omega$. □

6 Numerical examples

6.1 Example 1, n=1.

Let $\omega_1 = 1$, $\omega_2 = 1.3$, $\phi_1 = 0$, $\phi_2 = -0.5\pi$ and

$$g(x, \lambda) = \max\{g_1(x, \lambda),\, g_2(x, \lambda)\}, \tag{16}$$

$$g_1(x, \lambda) = -\frac{1}{2} a \sin(\omega_1 \lambda + \phi_1) - x, \tag{17}$$

$$g_2(x, \lambda) = x - \frac{1}{3} a \sin(\omega_2 \lambda + \phi_2). \tag{18}$$

Given λ_0, $f_{\lambda_0}(x) = \max\{-c_1 - x, x - c_2\}$ where $c_1 = -0.5\, a\, sin(\omega_1 \lambda_0 + \phi_1)$ and $c_2 = -0.\overline{3}\, a\, sin(\omega_2 \lambda_0 + \phi_2)$, then f_{λ_0} is convex.

To obtain $s \in \partial f_{\lambda_0}(x)$, first we compute $i = \arg\max\{g_1(x, \lambda_0), g_2(x, \lambda_0)\}$, then $s = 1$ if $i = 1$ and $s = -1$ if $i = 2$.

Figure 1 shows $\Omega \cap S_0$ where $S_0 = [-a, a] \times [0, 2a]$ and $a = 1.8\pi$. The feasible region of (P), restricted to S_0, is the union of two disconnected regions.

Taking $M = 4.6$ and $\varepsilon = 10^{-6}$, the algorithm makes 45 iterations to finish, reaching $(x^*, \lambda^*) = (-1.1973, 8.9875)$. Figure 2 shows the cuts generated by the algorithm.

Now, we take $a = 3\pi$ and maintain other parameters as before. Figure 3 shows $\Omega \cap S_0$. In this case, the feasible region of (P), restricted to S_0, is the union of three disconnected parts. Using the same M and ε as before, the algorithm performs 251 iterations and obtains $(x^*, \lambda^*) = (-0.0139, 15.7046)$. Figure 4 shows the cuts generated by the algorithm.

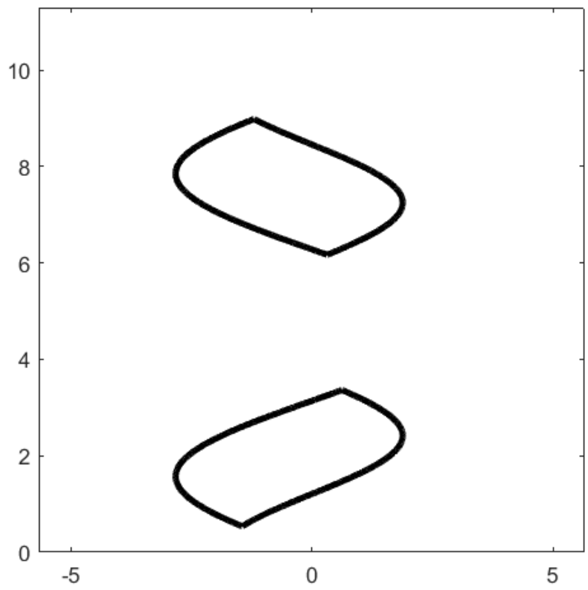

Fig. 1 Example 1. Plot of $\Omega \cap S_0$, $S_0 = [-a, a] \times [0, 2a]$ and $a = 1.8\pi$.

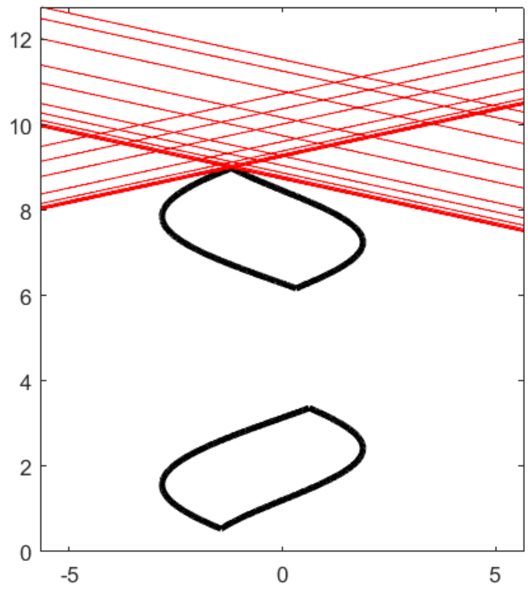

Fig. 2 Example 1. Cuts generated by the algorithm.

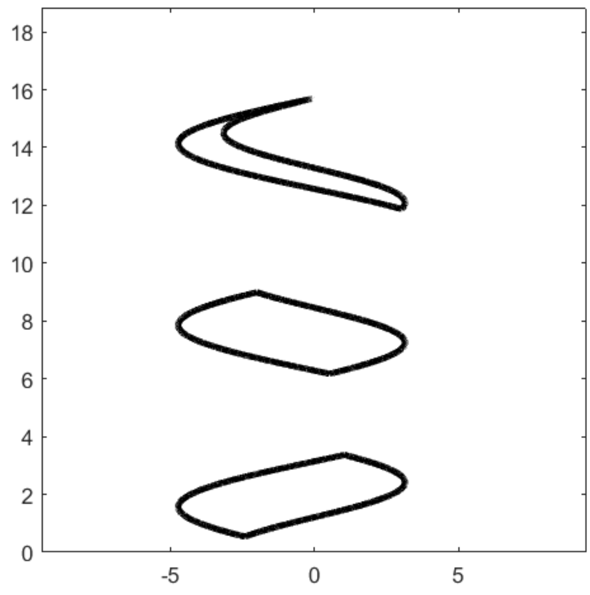

Fig. 3 Example 1. Plot of $\Omega \cap S_0$, $S_0 = [-a, a] \times [0, 2a]$ and $a = 3\pi$.

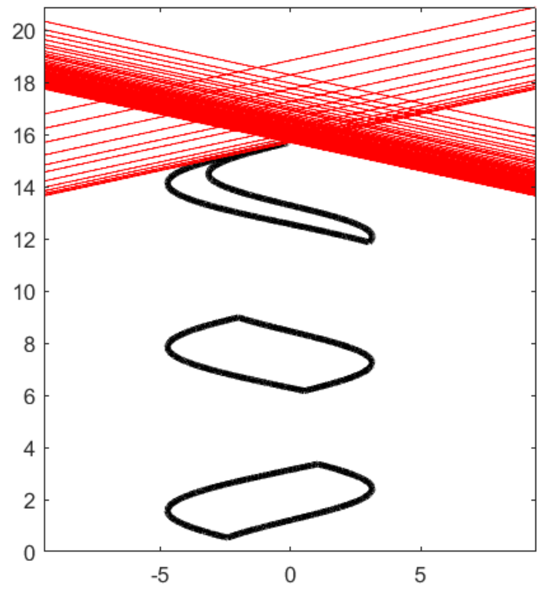

Fig. 4 Example 1. Cuts generated by the algorithm.

7 Conclusions

This research introduces a cutting-plane algorithm designed to address the specific non-convex optimization problem (P). Assuming hypotheses (H1) and (H2), we establish a valid cutting plane that depends on an infeasible point (x^0, λ_0) and a constant M satisfying (H2). A practical formula is provided for computing M for a specific function g. The proposed cutting-plane algorithm is outlined, and its convergence to an optimal solution of (P) is proven. An initial numerical example is presented, showcasing the sequence of cuts and demonstrating how the optimal solution is obtained, even in scenarios involving disconnected feasible sets.

References

[1] Kelley, J.J.E.: The cutting-plane method for solving convex programs. Journal of the Society for Industrial and Applied Mathematics **8**, 703–712 (1960)