**Delft University of Technology**

**Delft Center for Systems and Control**

Technical report 15-030

# Optimistic optimization for continuous nonconvex piecewise affine functions*

J. Xu, T. van den Boom, and B. De Schutter

*If you want to cite this report, please use the following reference instead:*
J. Xu, T. van den Boom, and B. De Schutter, "Optimistic optimization for continuous nonconvex piecewise affine functions," *Automatica*, Mar. 2021. Article 109476. doi:10.1016/j.automatica.2020.109476

Delft Center for Systems and Control
Delft University of Technology
Mekelweg 2, 2628 CD Delft
The Netherlands
phone: +31-15-278.24.73 (secretary)
URL: https://www.dcsc.tudelft.nl

*This report can also be downloaded via https://pub.deschutter.info/abs/15_030.html

# Optimistic optimization for continuous nonconvex piecewise affine functions ⋆

Jia Xu [a], Ton van den Boom [b], Bart De Schutter [b]

[a] *Department of Control Science and Engineering, Tongji University, Shanghai, China*

[b] *Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands*

**Abstract**

This paper considers global optimization of a continuous nonconvex piecewise affine (PWA) function over a polytope. This type of optimization problem often arises in the context of control of continuous PWA systems. In literature, it has been shown that the given problem can be formulated as a mixed integer linear programming (MILP) problem, the worst-case complexity of which grows exponentially with the number of polyhedral subregions in the domain of the PWA function. In this paper, we propose a solution approach that is more efficient for continuous PWA functions with a large number of polyhedral subregions. The proposed approach is based on optimistic optimization, which uses hierarchical partitioning of the feasible set and which can guarantee bounds on the suboptimality of the returned solution with respect to the global optimum given a prespecified finite number of iterations. Since the function domain is a polytope with arbitrary shape, we introduce a partitioning approach by employing Delaunay triangulation and edgewise subdivision. Moreover, we derive the analytic expressions for the core parameters required by optimistic optimization for continuous PWA functions. The numerical example shows that the resulting algorithm is faster than MILP solvers for PWA functions with a large number of polyhedral subregions.

*Key words:* Piecewise affine function; optimistic optimization; simplicial subdivision.

## 1 Introduction

Piecewise affine (PWA) functions are widely used in various fields for approximating nonlinearities, see [1,5,16,19]; they also appear as cost functions of numerous optimization problems, see [8,18]. The optimization of *nonconvex* PWA functions are often described as mixed integer linear programming (MILP) problems [7,21]. However, the worst-case complexity of MILP solvers grows exponentially with the number of polyhedral subregions of the PWA functions, which usually make the problem solving process less efficient.

We focus on the optimization problem of a continuous and nonconvex PWA function over a given polytope and propose to apply optimistic optimization to seek the global optimal solution. Optimistic optimization [14,15] is a class of algorithms that start from a hierarchical partition of the feasible set and gradually focuses on the most promising area until they eventually perform a local search around the global optimum of the function. The gap between the best value returned by the algorithm and the real global optimum can be expressed as a function of the number of iterations, which can be specified in advance. Optimistic optimization can be applied to the general problem of black-box optimization of a function given evaluations of the functions over general search spaces. Until now, in the literature on optimistic optimization, the feasible set is often assumed to be a hypercube or a hyperbox. In our previous work [22,23], we have extended optimistic optimization to solve the model predictive control problem for max-plus linear and continuous PWA systems. In [23], the PWA-MPC problem is recast as an optimization problem of a continuous PWA objective function. Particularly, the linear constraints on states and inputs are treated as soft constraints and are replaced by adding a penalty function to the objective function. As a result, the feasible set becomes a hyperbox for which the hierarchical partitioning can be efficiently developed.

In this paper, the linear constraints on decision variables are considered as hard constraints and, for the first time in the literature on optimistic optimization, a polytopic feasible set is considered. This extension from a

⋆ This paper was not presented at any IFAC meeting. Corresponding author J. Xu.

*Email addresses:* 615xujia@tongji.edu.cn (Jia Xu), a.j.j.vandenboom@tudelft.nl (Ton van den Boom), b.deschutter@tudelft.nl (Bart De Schutter).

hyperbox feasible set to a polytopic one is not trivial but useful because a polytopic feasible set allows to include general affine constraints on the control variables rather than only single bound constraints. A partition of the given polytope is required to perform the search process. The partitioning should generate well-shaped cells that shrink with the depth. We first employ Delaunay triangulation to divide the polytope into a mesh of simplices and next repeatedly use edgewise subdivision to subdivide the simplices into smaller simplices that satisfy the requirements for optimistic optimization. For this partitioning approach, we develop analytic expressions for the core parameters of optimistic optimization based on the knowledge of the Lipschitz constants of the PWA function. The numerical example shows that using optimistic-optimization-based algorithms for the optimization of a continuous and nonconvex PWA function over a given polytope is more efficient than transforming into an MILP problem if the number of polyhedral subregions of the PWA function is large.

This paper is organized as follows. In Section 2 and 3, we give some definitions and describe the optimization problem of continuous PWA functions. In Section 4, we introduce background of an optimistic optimization algorithm. In Section 5, we propose a partitioning approach for which we develop the analytic expressions for the core parameters of optimistic optimization. In Section 6, the proposed approach is assessed with a numerical example. Finally, Section 7 concludes the paper.

## 2 Preliminaries

For any $x \in \mathbb{R}^n$, define $\|x\|_2 = (|x_1|^2 + \cdots + |x_n|^2)^{\frac{1}{2}}$. This section presents some necessary definitions, which are based on [4,17].

**Definition 1 (Polyhedron)** *A polyhedron is a convex set given as the intersection of a finite number of half-spaces.*

**Definition 2 (Polytope)** *A bounded polyhedron $\mathcal{P} = \{x \in \mathbb{R}^n | Ax \le b\}$ is called a polytope, for some matrix $A$ and some vector $b$.*

**Definition 3 (Simplex)** *An $m$-simplex $\mathcal{S} \subset \mathbb{R}^n$ with $0 \le m \le n$ is the convex hull of $m+1$ affinely independent points $v_0, \ldots, v_m \in \mathbb{R}^n$ which are its vertices. If $m = n$, the set $\mathcal{S}$ is simply called a simplex of $\mathbb{R}^n$. Let $e_i = v_i - v_{i-1}$, $i = 1, \ldots, n$. The $n$-dimensional volume of $\mathcal{S}$ is* $\mathrm{vol}(\mathcal{S}) = \frac{1}{n!}\big|\det(e_1, e_2, \ldots, e_n)\big|$.

**Definition 4 (Polyhedral partition)** *Given a polyhedron $\mathcal{P} \subseteq \mathbb{R}^n$, then a polyhedral partition of $\mathcal{P}$ is a finite collection $\{\mathcal{P}_i\}_{i=1}^N$ of nonempty polyhedra satisfying (i) $\bigcup_{i=1}^N \mathcal{P}_i = \mathcal{P}$; (ii) $(\mathcal{P}_i \setminus \partial\mathcal{P}_i) \bigcap (\mathcal{P}_j \setminus \partial\mathcal{P}_j) = \emptyset$ for all $i \neq j$ where $\partial$ denotes the boundary.*

**Definition 5 (PWA function)** *A scalar-valued function $f : \mathcal{P} \to \mathbb{R}$, where $\mathcal{P} \subseteq \mathbb{R}^n$ is a polyhedron, is PWA if there exists a polyhedral partition $\{\mathcal{P}_i\}_{i=1}^N$ of $\mathcal{P}$ such that $f$ is affine on each $\mathcal{P}_i$, i.e., $f(x) = \alpha_{(i)}^T x + \beta_{(i)}$, for all $x \in \mathcal{P}_i$, with $\alpha_{(i)} \in \mathbb{R}^n$, $\beta_{(i)} \in \mathbb{R}$, for $i = 1, \ldots, N$. If a PWA function $f$ is continuous on the boundary of any two neighboring regions, then $f$ is said to be continuous PWA. A vector-valued function is continuous PWA if each of its components is continuous PWA.*

## 3 Problem statement

Consider the following optimization problem

$$\min_{x} \quad f(x) \tag{1}$$

$$\text{subject to} \quad Ax \le b \ , \tag{2}$$

where $A \in \mathbb{R}^{m \times n_x}$ and $b \in \mathbb{R}^m$ are the constraint matrix and vector, and $f$ is a scalar-valued continuous PWA function given by $f(x) = \alpha_{(i)}^T x + \beta_{(i)}$, $\forall x \in \mathcal{P}_i$, with $\alpha_{(i)} \in \mathbb{R}^{n_x}$, $\beta_{(i)} \in \mathbb{R}$, $i = 1, \ldots, N$. We assume that the feasible set $\mathcal{X} = \{x \in \mathbb{R}^{n_x} | Ax \le b\} \subset \mathcal{P}$ is nonempty and bounded. From Definition 2, $\mathcal{X}$ is a polytope.

In this paper, we consider the case that $f$ is continuous and nonconvex and that the number $N$ of polyhedral subregions is much larger than $n_x$. For this case, one possible solution approach consists in transforming the problem (1)-(2) into an MILP problem. The number of auxiliary variables and linear constraints in the resulting MILP description is proportional to $N$. So the complexity of the resulting MILP problem grows in the worst case exponentially in $N$. In the next section, we will introduce an optimistic optimization algorithm for the problem (1)-(2). The knowledge of a Lipschitz constant of $f$ is important for designing the two key parameters $\nu$ and $\rho$ of optimistic optimization. For any $x, y \in \mathcal{P}_i$, we have

$$\big|f(x) - f(y)\big| \le \|\alpha_{(i)}\|_2 \|x - y\|_2 \ . \tag{3}$$

It is easy to verify that $\max\limits_{i=1,\ldots,N} \|\alpha_{(i)}\|_2$ is the smallest Lipschitz constant of $f$ [10].

## 4 Deterministic optimistic optimization

In this section, we introduce the background of the deterministic optimistic optimization (DOO) algorithm [14]. The notations $f$ and $\mathcal{X}$ remain generic in this section.

DOO algorithm is based on a given partitioning of $\mathcal{X}$. For any integer $h \in \{0, 1, \ldots\}$, the space $\mathcal{X}$ is recursively split into $K^h$ cells where $K$ is a finite positive integer denoting the maximum number of child cells of a parent cell. The partitioning may be represented by a tree structure. The whole set $\mathcal{X}$ is denoted as $X^{0,0}$ and corresponds to the root node $(0, 0)$ of the tree. Each cell at any depth $h$ is denoted as $X^{h,d}$ with $d \in \{0, \ldots, K^h - 1\}$ and corresponds to a node $(h, d)$ in the tree. A cell $X^{h,d}$ at depth $h$ is split into $K$ child cells $\{X^{h+1,i}\}_{i=1}^K$. Each cell $X^{h,d}$ is characterized by a representative point $x^{h,d} \in X^{h,d}$ in which $f$ may be evaluated.

Four necessary assumptions are stated in [14] regarding the function $f$ and the partitioning. Those assumptions are expressed in terms of a semi-metric. However, as discussed in [12], this semi-metric actually just seems to

**Algorithm 1** Deterministic Optimistic Optimization (DOO)

**Given**: partitioning of $\mathcal{X}$, number $n$ of iterations
Initialize the tree $\mathcal{T} \leftarrow \{(0,0)\}$ (root node)
**for** $t = 1$ to $n$ **do**
  $(h^\dagger, d^\dagger) \leftarrow \arg\min_{(h,d)\in\mathcal{L}} f(x^{h,d}) - \nu\rho^h$
  Expand $(h^\dagger, d^\dagger)$ by adding its $K$ children to $\mathcal{T}$
**end for**
**Return** $x(n) = \arg\min_{(h,d)\in\mathcal{T}} f(x^{h,d})$

link the function and the partitioning and it is not used in the implementation of the algorithm. So in [12] the assumptions for the DOO algorithm are merged into a single one by discarding the semi-metric. In this paper, we use the setting in [12].

**Requirement 1** Given the partitioning of $\mathcal{X}$, let $d_h^*$ be the index of the cell at depth $h$ containing a global optimizer $x^*$, i.e., $x^* \in X^{h,d_h^*}$, and let $x^{h,d_h^*}$ be the representative point of the cell $X^{h,d_h^*}$. Then there exists $\nu > 0$ and $\rho \in (0,1)$ such that for any $h \in \{0,1,\dots\}$, we have $f(x^{h,d_h^*}) - f(x^*) \le \nu\rho^h$.

DOO is summarized in Algorithm 1 [14]. Given a finite number $n$ of iterations, DOO generates a sequence of feasible solutions during the iterations and returns the best solution $x(n)$ at the end. Starting with the root node $\{(0,0)\}$, DOO incrementally updates the tree $\mathcal{T}$ for $t = 1,\dots,n$. At each iteration $t$, DOO selects a leaf [1] of the current tree with the minimum value of $f(x^{h,d}) - \nu\rho^h$ to expand by adding its $K$ children to the current tree. Expanding a leaf $(h,d)$ corresponds to splitting the cell $X^{h,d}$ into $K$ subcells and evaluating the function $f$ at the representative points of the children cells.

Requirement 1 implies that any cell containing $x^*$ satisfies $f(x^{h,d_h^*}) - \nu\rho^h \le f(x^*)$. Consequently, a cell $X^{h',d'}$ such that $f(x^{h',d'}) - \nu\rho^{h'} > f(x^*)$ will never be selected to split because there always exists a cell containing $x^*$ such that $f(x^{h,d_h^*}) - \nu\rho^h < f(x^{h',d'}) - \nu\rho^{h'}$. More specifically, DOO only expands nodes of the set $I \triangleq \bigcup_{h\ge 0} I_h$ where $I_h = \{(h,d) | f(x^{h,d}) - f(x^*) \le \nu\rho^h\}$. The elements of $I_h$ can be considered as $\nu\rho^h$-near-optimal solutions. The more near-optimal solutions, the slower the convergence speed of the algorithm. In general, the number of near-optimal solutions will increase if the number of optimal solutions increases. Therefore, the algorithm is in general more efficient for the problem with a unique optimal solution than the case where the optimal solution is not unique. A measure (called near-optimality dimension) is defined in [14] to characterize the number of near-optimal solutions and to derive bounds on the difference between the optimal solution and the solution returned by the algorithm. In this paper, we adapt the definition of near-optimality dimension in [12] to make it equivalent to the definition in [14]. Proposition 7 provides a guarantee on the performance of DOO.

**Definition 6** *The near-optimality dimension of $f$ is the smallest $\eta \ge 0$ such that there exists a positive constant $C$ such that the maximum number of cells $X^{h,d}$ at any depth $h$ for which $f(x^{h,d}) - f(x^*) \le \nu\rho^h$ is less than $C(\nu\rho^h)^{-\eta}$.*

**Proposition 7** *For a given finite number $n$ of iterations, let $x^*$ be a global minimizer and let $x(n)$ be the solution returned by the algorithm after $n$ iterations.*
*(i) Let $(h_{\max}, d_{\max})$ be the deepest node that has been expanded by the algorithm up to $n$ iterations. Then we have $f(x(n)) - f(x^*) \le \nu\rho^{h_{\max}}$.*
*(ii) If $\eta > 0$, then $f(x(n)) - f(x^*) \le \left(\frac{C}{1-\rho^\eta}\right)^{1/\eta} n^{-1/\eta}$.*
*(iii) If $\eta = 0$, then $f(x(n)) - f(x^*) \le \nu\rho^{n/C-1}$.*

**PROOF.** (i) Since DOO only expands the nodes of the set $I$, we have $f(x^{h_{\max},d_{\max}}) - f(x^*) \le \nu\rho^{h_{\max}}$. Note that $x(n)$ is the returned solution with minimum function value of $f$ among the expanded nodes, so $f(x^*) \le f(x(n))$ and $f(x(n)) \le f(x^{h_{\max},d_{\max}})$. Hence, $f(x(n)) - f(x^*) \le \nu\rho^{h_{\max}}$. Furthermore, $f(x(n)) - \nu\rho^{h_{\max}}$ and $f(x(n))$ are respectively a lower and an upper bound of $f(x^*)$. In addition, the distance between the two bounds is bounded by $\nu\rho^{h_{\max}}$.
(ii) From Definition 6, we have $|I_h| \le C(\nu\rho^h)^{-\eta}$. Define an indicator function $\mathbf{1}_{I_h}(h,d)$ as: if $(h,d)$ has been expanded, $\mathbf{1}_{I_h}(h,d) = 1$, else $\mathbf{1}_{I_h}(h,d) = 0$. When $\eta > 0$, the number of node expansions $n$ satisfies $n = \sum_{h=0}^{h_{\max}} \sum_{d=0}^{K^h-1} \mathbf{1}_{I_h}(h,d) \le \sum_{h=0}^{h_{\max}} |I_h| \le C\nu^{-\eta} \sum_{h=0}^{h_{\max}} (\rho^{-\eta})^h \le C\nu^{-\eta} \frac{\rho^{-\eta h_{\max}}}{1-\rho^\eta}$. Thus, we have $(\nu\rho^{h_{\max}})^\eta \le \frac{C}{n(1-\rho^\eta)}$. Combined with (i), this yields $f(x(n)) - f(x^*) \le \left(\frac{C}{1-\rho^\eta}\right)^{1/\eta} n^{-1/\eta}$.
(iii) When $\eta = 0$, we have $n \le C(h_{\max} + 1)$. Thus, $h_{\max} \ge \frac{n}{C} - 1$. Since $\rho \in (0,1)$, we obtain $f(x(n)) - f(x^*) \le \nu\rho^{h_{\max}} \le \nu\rho^{n/C-1}$.

## 5 Optimistic optimization of PWA functions

In this section, we first develop a partitioning approach for the polytopic feasible set $\mathcal{X} = \{x \in \mathbb{R}^{n_x} | Ax \le b\}$.

### 5.1 *Hierarchical partition of a polytope*

The partitioning consists of two stages: (i) dividing the polytope $\mathcal{X}$ into a collection of simplices; (ii) subdividing each simplex into smaller simplices. In this paper, we propose to use Delaunay triangulation for the first stage and next to use edgewise subdivision repeatedly for the second stage. Delaunay triangulation [6] divides a polytope into a mesh of simplices where simplices with very short edges are created as less as possible. This property conforms to the requirement of optimistic optimization that at any depth there is always a ball with some depth-dependent radius that fits within the subcell. Edgewise subdivision [9] divides a simplex $\mathcal{S}$ of $\mathbb{R}^{n_x}$ into $k^{n_x}$ $n_x$-simplices, where each edge of $\mathcal{S}$ is cut into $k$

[1] A leaf of a tree is a node with no children. The set $\mathcal{L}$ contains the leaves of $\mathcal{T}$.

equal pieces. A ready-to-implement algorithm for edgewise subdivision is presented in [11]. Below we introduce the definition and some properties of edgewise subdivision. Those properties are needed in the next section for the development of expressions for $\nu$ and $\rho$ of DOO.

**Definition 8** *[3] Two non-degenerate (volume is not zero) simplices $\mathcal{S}, \mathcal{S}'$ are called congruent to each other if there exists a translation vector $v \in \mathbb{R}^{n_x}$, a scaling factor $c > 0$, and an orthogonal matrix $Q \in \mathbb{R}^{n_x \times n_x}$ such that[2] $\mathcal{S}' = v + cQ\mathcal{S}$. In this case $\mathcal{S}$ and $\mathcal{S}'$ are elements of the same congruence class.*

**Properties of edgewise subdivision**. For every integer $k \geq 1$, the edgewise subdivision of $\mathcal{S}$ has the following properties [9]: (i) all generated simplices have the same $n$-dimensional volume; (ii) all generated simplices fall into at most $n_x!/2$ congruence classes; (iii) the faces of $\mathcal{S}$ are subdivided the same way; (iv) repeated subdivision has the same effect as increasing $k$.

The property (iv) that repeated subdivision has the same effect as increasing $k$, means that instead of dividing an $n_x$-simplex $\mathcal{S}$ into $k^{n_x}$ $n_x$-simplices and subsequently subdividing each sub-simplex into $l^{n_x}$ $n_x$-simplices, we can subdivide $\mathcal{S}$ into $(kl)^{n_x}$ $n_x$-simplices and reach the same result.

### *5.2 PWA optimistic optimization*

In this section, Proposition 10 gives analytic expressions for $\nu$ and $\rho$ required by DOO. Lemma 11 guarantees that the sub-simplices generated by the developed partitioning approach do not become too slim with very short edges.

By performing Delaunay triangulation, $\mathcal{X}$ is divided into a mesh of simplices $\{\mathcal{X}_s | s = 1, \ldots, N_\mathrm{t}\}$. Every simplex $\mathcal{X}_s$ in the simplicial mesh is taken as the original simplex on which repeated edgewise subdivision is performed. Properties (i)-(iv) of edgewise subdivision given in Section 5.1 are essential for the remaining proof. For any integer $k \geq 1$, edgewise subdivision divides $\mathcal{X}_s$ into $k^{n_x}$ $n_x$-simplices; so the maximum number $K$ of child cells of a parent cell equals $k^{n_x}$. Note that $h \in \{0, 1, \ldots\}$ is the depth of the subdivision (indicator of the recursion of edgewise subdivision) and $d \in \{0, \ldots, K^h - 1\}$ is the index of simplices at a given depth $h$. Let $X_s^{h,d}$ be a simplex at depth $h$ generated by repeated edgewise subdivision of $\mathcal{X}_s$. Let $L_s^{h,d}, r_s^{h,d}, x_s^{h,d}$ be the maximum edge length, inradius (radius of the inscribed hyper-ball) and incenter (center of the inscribed hyper-ball) of $X_s^{h,d}$. Let $N_s \leq n_x!/2$ be the number of congruence classes that all simplices generated by repeated edgewise subdivision of $\mathcal{X}_s$ fall into (see Property (ii)). Note that the simplices in each congruence class are the same up to translation, scaling, and rotation. Let $C_{s,i}$, $i = 1, \ldots, N_s$, be a representative simplex[3] of each congruence class. Define the ratio between the maximum and minimum volumes among representative simplices as

$$\gamma_s = \max_{i,j=1,\ldots,N_s} \frac{\mathrm{vol}(C_{s,i})}{\mathrm{vol}(C_{s,j})} \ . \tag{4}$$

Let $\tau_{s,i}$ be the inradius of $C_{s,i}$ and denote

$$\tau_s = \min_{i=1,\ldots,N_s} \tau_{s,i} \ . \tag{5}$$

Let $v_{s,0}, \ldots, v_{s,n_x}$ be the vertices of $\mathcal{X}_s$. Let $v_{s,0}^{h,d}, \ldots, v_{s,n_x}^{h,d}$ be the vertices of $X_s^{h,d}$. Define $e_{s,i} = v_{s,i} - v_{s,i-1}$ and $e_{s,i}^{h,d} = v_{s,i}^{h,d} - v_{s,i-1}^{h,d}$, $i = 1, \ldots, n_x$. Then taking into account the proof of the independence lemma in [9] as well as the fact that repeated subdivision has the same effect as increasing $k$ (see Property (iv)), there exists a permutation $\pi_s^{h,d}$ of $\{1, \ldots, n_x\}$ such that $e_{s,i}^{h,d} = \frac{1}{k^h} e_{s,\pi_s^{h,d}(i)}$. Note that we have

$$v_{s,i}^{h,d} - v_{s,0}^{h,d} = e_{s,i}^{h,d} + e_{s,i-1}^{h,d} + \cdots + e_{s,1}^{h,d} \ . \tag{6}$$

Now select an arbitrary edge of $X_s^{h,d}$ and let $v_{s,i}^{h,d}$ and $v_{s,j}^{h,d}$ with $j > i$ be the corresponding vertices. By (6), we have $\left|v_{s,j}^{h,d} - v_{s,i}^{h,d}\right| = \left|e_{s,j}^{h,d} + e_{s,j-1}^{h,d} + \cdots + e_{s,i+1}^{h,d}\right| = \frac{1}{k^h}\left|e_{s,\pi_s^{h,d}(j)} + e_{s,\pi_s^{h,d}(j-1)} + \cdots + e_{s,\pi_s^{h,d}(i+1)}\right|$. Define

$$\theta_{s,\min} = \min_{i=1,\ldots,n_x} |e_{s,i}| \ , \qquad \theta_{s,\max} = \sum_{i=1}^{n_x} |e_{s,i}| \ . \tag{7}$$

Note that $\theta_{s,\min} > 0$. Then we have

$$\frac{1}{k^h}\theta_{s,\min} \leq \left|v_{s,j}^{h,d} - v_{s,i}^{h,d}\right| \leq \frac{1}{k^h}\theta_{s,\max} \ . \tag{8}$$

**Lemma 9** *Denote $L_{s,h} = \max_{d \in D_h} L_s^{h,d}$ and $r_{s,h} = \min_{d \in D_h} r_s^{h,d}$ where $D_h = \{0, \ldots, K^h - 1\}$ is the index set of simplices at depth $h$. Then we have*

$$\frac{L_{s,h+1}}{L_{s,h}} \leq \frac{1}{k}\gamma_s^{1/n_x} \ , \qquad \frac{r_{s,h}}{L_{s,h}} \geq \frac{\theta_{s,\min}\tau_s}{\theta_{s,\max}} \tag{9}$$

*where $\gamma_s$, $\tau_s$, $\theta_{s,\min}$ and $\theta_{s,\max}$ are as defined in (4), (5), (7) and $1/k$ is the factor of edgewise subdivision.*

**PROOF.** Let $X_s^{h,d'}$ be the simplex that has the maximum edge length $L_{s,h}$ among all simplices at depth $h$ and assume that $X_s^{h,d'}$ belongs to congruence class $i$ with a representative simplex $C_{s,i}$. The maximum edge length of $C_{s,i}$ equals 1. From Property (iv), repeated subdivision is equivalent to increasing $k$; so a division at depth $h$ actually corresponds to selecting $k^h$ instead of $k$. Moreover, from Property (i), we have $\mathrm{vol}(X_s^{h,d'}) = \mathrm{vol}(\mathcal{X}_s)/k^{hn_x}$. Scaling $X_s^{h,d'}$ with a factor $1/L_{s,h}$ scales every column in the matrix of which the determinant is taking in the volume formula given in Definition 3, resulting in a multiplication with $(1/L_{s,h})^{n_x}$ compared to the original expression. Hence, we have

$$\mathrm{vol}(C_{s,i}) = \left(\frac{1}{L_{s,h}}\right)^{n_x} \mathrm{vol}(X_s^{h,d'}) = \left(\frac{1}{L_{s,h}}\right)^{n_x} \frac{\mathrm{vol}(\mathcal{X}_s)}{k^{hn_x}} \ . \tag{10}$$

Likewise let $X_s^{h+1,d''}$ be the simplex that has the maximum edge length $L_{s,h+1}$ among all simplices at depth

[2] $v + cQ\mathcal{S} = \{v + cQx | x \in \mathcal{S}\}$.

[3] A representative simplex of a congruence class is then defined as the simplex resulting from scaling any simplex in the class such that its maximum edge length equals 1.

$h+1$ and assume that $X_s^{h+1,d''}$ belongs to congruence class $j$ with a representative simplex $C_{s,j}$. So

$$\mathrm{vol}(C_{s,j}) = \Big(\frac{1}{L_{s,h+1}}\Big)^{n_x} \frac{\mathrm{vol}(\mathcal{X}_s)}{k^{(h+1)n_x}} \quad . \tag{11}$$

Thus (10) and (11) result in

$$\left(\frac{L_{s,h+1}}{L_{s,h}}\right)^{n_x} = \frac{1}{k^{n_x}} \frac{\mathrm{vol}(C_{s,i})}{\mathrm{vol}(C_{s,j})} \tag{12}$$

and then $\frac{L_{s,h+1}}{L_{s,h}} = \frac{1}{k}\left(\frac{\mathrm{vol}(C_{s,i})}{\mathrm{vol}(C_{s,j})}\right)^{1/n_x} \leq \frac{1}{k}\gamma_s^{1/n_x}$. This completes the proof of the first inequality in (9).

Let $X_s^{h,d^\sharp}$ be the simplex that has the shortest inradius $r_{s,h}$ among all simplices at depth $h$ and assume that $X_s^{h,d^\sharp}$ belongs to congruence class $l$ with a representative simplex $C_{s,l}$. The maximum edge length of $C_{s,l}$ equals 1 and the inradius of $C_{s,l}$ is $\tau_{s,l}$. Thus, we have $r_{s,h} = L_s^{h,d^\sharp}\tau_{s,l}$. Due to (5), we also have $r_{s,h} \geq L_s^{h,d^\sharp}\tau_s$. Note that (8) implies that $\frac{1}{k^h}\theta_{s,\min} \leq L_s^{h,d} \leq \frac{1}{k^h}\theta_{s,\max}$ , $\forall d \in D_h$. Hence, $r_{s,h} \geq L_s^{h,d^\sharp}\tau_s \geq \frac{1}{k^h}\theta_{s,\min}\tau_s$ and thus $\frac{r_{s,h}}{L_{s,h}} \geq \frac{\frac{1}{k^h}\theta_{s,\min}\tau_s}{L_{s,h}} \geq \frac{\frac{1}{k^h}\theta_{s,\min}\tau_s}{\frac{1}{k^h}\theta_{s,\max}} \geq \frac{\theta_{s,\min}\tau_s}{\theta_{s,\max}}$. This completes the proof. □

**Proposition 10** *Denote $\alpha = \max_{i=1,\dots N}\|\alpha_{(i)}\|_2$ and $\nu_s = \alpha L_{s,0}$, $\rho_s = \frac{L_{s,h+1}}{L_{s,h}}$ where $L_{s,h}$ is as defined in Lemma 9. Let $\nu = \max\limits_{s=1,\dots,N_t}\nu_s$, $\rho = \max\limits_{s=1,\dots,N_t}\rho_s$. If $k$ is selected as an integer that is strictly larger than $\max\limits_{s=1,\dots N_t}\gamma_s^{1/n_x}$, then for any cell $X^{h,d_h^*}$ that contains a global optimizer $x^*$ with the incenter selected as the representative point $x^{h,d_h^*}$ of the cell $X^{h,d_h^*}$, we have $\nu > 0$, $\rho \in (0,1)$, and $f(x^{h,d_h^*}) - f(x^*) \leq \nu\rho^h$.*

**PROOF.** From (12), we can conclude that $\rho_s = \frac{L_{s,h+1}}{L_{s,h}}$ does not depend on $h$. Note that with the given definitions of $\nu$ and $\rho$, they are naturally positive constants. Moreover, if $k$ is selected as an integer that is strictly larger than $\max_{s=1,\dots N_t}\gamma_s^{1/n_x}$, then, from Lemma 9, for any $s$, we have $\rho_s \leq \frac{1}{k}\gamma_s^{1/n_x} < 1$. So $\nu > 0$ and $\rho \in (0,1)$. Assume that $x^*$ is contained in a cell $X_s^{h,d_h^*}$ and the incenter of $X_s^{h,d_h^*}$ is selected as the representative point $x_s^{h,d_h^*}$. Then we have $f(x_s^{h,d_h^*}) - f(x^*) \overset{(3)}{\leq} \alpha\|x_s^{h,d_h^*} - x^*\|_2 \leq \alpha L_s^{h,d_h^*} \leq \alpha L_{s,h}$. From $\rho_s = \frac{L_{s,h+1}}{L_{s,h}}$, we have

$$L_{s,h} = (\rho_s)^h L_{s,0}. \tag{13}$$

Thus, $f(x_s^{h,d_h^*}) - f(x^*) \leq \alpha(\rho_s)^h L_{s,0}$. From $\nu_s = \alpha L_{s,0}$, we have $f(x_s^{h,d_h^*}) - f(x^*) \leq \nu_s(\rho_s)^h$. Let $\nu = \max\limits_{s=1,\dots,N_t}\nu_s$ and $\rho = \max\limits_{s=1,\dots,N_t}\rho_s$. Therefore, for any cell $X^{h,d_h^*}$ that contains $x^*$, we have $f(x^{h,d_h^*}) - f(x^*) \leq \nu\rho^h$. This completes the proof. □

**Lemma 11** *Let $\sigma = \mu_1\mu_2 \min\limits_{s=1,\dots,N_t}\sigma_s$ where $\mu_1 = \min\limits_{s',s''=1,\dots,N_t}\frac{L_{s'',0}}{L_{s',0}}$, $\mu_2 = \min\limits_{s,s''=1,\dots,N_t}\frac{L_{s,h}}{L_{s'',h}}$, and $\sigma_s$ is a positive constant such that $0 < \sigma_s \leq \frac{\tau_s\theta_{s,\min}}{\alpha\theta_{s,\max}}$. Then any cell $X^{h,d}$ at any depth $h$ contains a ball of radius $\sigma\nu\rho^h$ centered in $x^{h,d}$, denoted as $\mathfrak{B}(x^{h,d},\sigma\nu\rho^h) = \{x \in \mathcal{X} \mid \|x - x^{h,d}\|_2 \leq \sigma\nu\rho^h\} \subset X^{h,d}$, where $\nu$ and $\rho$ are defined as in Proposition 10.*

**PROOF.** First we prove that $\mu_2$ is independent of $h$. Similar to the proof of Lemma 9, we get $\mathrm{vol}(X_s^{h,d'}) = (L_{s,h})^{n_x}\mathrm{vol}(C_{s,i})$, $i \in \{1,\dots,N_s\}$ and $\mathrm{vol}(X_{s''}^{h,d''}) = (L_{s'',h})^{n_x}\mathrm{vol}(C_{s'',j})$, $j \in \{1,\dots,N_{s''}\}$. Hence, we have $\frac{L_{s,h}}{L_{s'',h}} = \left(\frac{\mathrm{vol}(X_s^{h,d'})\,\mathrm{vol}(C_{s'',j})}{\mathrm{vol}(X_{s''}^{h,d''})\,\mathrm{vol}(C_{s,i})}\right)^{1/n_x}$ which is independent of $h$. So $\mu_2$ is independent of $h$.

Now, we prove that $\sigma\nu\rho^h \leq \sigma_s\nu_s(\rho_s)^h$ for any $s = 1,\dots,N_t$, where $\nu_s$ and $\rho_s$ are defined as in Proposition 10. The inequality to be proved is rewritten as $\sigma \leq \sigma_s\frac{\nu_s(\rho_s)^h}{\nu\rho^h}$. Let $s'$ and $s''$ denote the indices such that $\nu_{s'} = \max\limits_{1,\dots,N_t}\nu_s$ and $\rho_{s''} = \max\limits_{1,\dots,N_t}\rho_s$. Thus we have $\nu = \nu_{s'}$, $\rho = \rho_{s''}$, and $\frac{\nu_s(\rho_s)^h}{\nu\rho^h} = \frac{\nu_s(\rho_s)^h}{\nu_{s'}(\rho_{s''})^h} = \frac{L_{s,h}}{L_{s',0}(\rho_{s''})^h} = \frac{L_{s'',0}L_{s,h}}{L_{s',0}L_{s'',h}} \geq \mu_1\mu_2$. If $\sigma = \mu_1\mu_2\min\limits_{s=1,\dots,N_t}\sigma_s$, then we have $\sigma\nu\rho^h \leq \sigma_s\nu_s(\rho_s)^h$ for any $s = 1,\dots,N_t$.

Finally, we prove that $\mathfrak{B}(x^{h,d},\sigma\nu\rho^h) \subset X^{h,d}$. For any $x \in \mathfrak{B}(x^{h,d},\sigma\nu\rho^h)$, we have $\|x - x^{h,d}\| \leq \sigma\nu\rho^h \leq \sigma_s\nu_s(\rho_s)^h \overset{(13)}{\leq} \frac{\tau_s\theta_{s,\min}}{\theta_{s,\max}}L_{s,h} \overset{(9)}{\leq} r_{s,h}$ where $r_{s,h}$ defined in Lemma 9 is the minimum among the inradii of simplices at depth $h$. Therefore, $x \in \mathfrak{B}(x^{h,d},\sigma\nu\rho^h)$ implies that $x \in X^{h,d}$. This completes the proof.

**Remark 12** *In Proposition 10, the parameter $\alpha$ requires the knowledge of the Lipschitz constants of the PWA function $f$. Actually, it may not always be possible to find the smallest Lipschitz constant of a general PWA objective function. In this case, an upper bound on the Lipschitz constants is also acceptable, but note that a larger $\alpha$ results in a larger $\nu$ and consequently results in a larger number of cells such that $f(x^{h,d}) - f(x^*) \leq \nu\rho^h$. As a result, the algorithm may waste time on exploring too many unnecessary cells which will lower the degree of optimality of the resulting solution for the predefined computational budget.*

In Section 5.2 of [20], it is shown that functions defined over a finite-dimensional and bounded space $\mathcal{X}$ have a near-optimality dimension equal to 0 if the functions have an upper and lower envelope around one global maximizer $x^*$ of the same order, i.e., there exist constants $c \in (0,1)$ and $\delta > 0$, such that for all $x \in \mathcal{X}$:

$$\min(\delta, c\ell(x,x^*)) \leq f(x^*) - f(x) \leq \ell(x,x^*) \tag{14}$$

where $\ell$ is a semi-metric. Clearly, the condition (14) is satisfied by the continuous nonconvex PWA functions considered in this paper. So the near-optimality dimension in our problem is equal to 0.

## 6 Numerical example

In this section, we evaluate the proposed optimistic-optimization-based approach for continuous PWA functions and compare it with other methods. The instances considered include 60 randomly generated continuous PWA functions $f : \mathbb{R}^2 \to \mathbb{R}$ in which the vector pairs $\alpha_{(i)} \in \mathbb{R}^2$, $\beta_{(i)} \in \mathbb{R}$ contain pseudorandom values drawn from the standard normal distribution $\mathcal{N}(0,1)$ with $i = 1, \ldots, N$ where $N$ is also random. Below we compare the efficiency of the DOO algorithm, the MILP method, and the DIRECT algorithm [13]. DIRECT is a direct search algorithm not requiring the knowledge of the Lipschitz constant. It uses an optimistic splitting technique similar to the optimistic optimization algorithm. The corresponding MILP problem is derived based on the techniques in [2] and solved with the `intlinprog` function in Matlab and the `cplex` function in Tomlab. DOO is implemented as the `pwadoo` function in Matlab. Note that `pwadoo` and `intlinprog` are both Matlab functions and `cplex` is implemented in object code, which implies that it will in general run much faster than a equivalent program written in Matlab. DIRECT is performed using the `glbDirect` solver in Tomlab and is implemented in object code. Fig. 1 shows the semi logarithmic plot of CPU time (average over 10 runs). The function values of $f$ returned from different solvers are denoted as $f_{\text{int}}$, $f_{\text{cpl}}$, $f_{\text{oo}}$, and $f_{\text{dir}}$, where $f_{\text{int}}$ and $f_{\text{cpl}}$ of every instance are equal. The iteration in `pwadoo` (`glbDirect`) is stopped if the gap between $f_{\text{cpl}}$ and $f_{\text{oo}}$ ($f_{\text{dir}}$) is less than 5% (the gap is calculated as $100|(f_{\text{oo}} - f_{\text{cpl}})/f_{\text{cpl}}|$). We can see that `pwadoo` is faster than `intlinprog` and even `cplex` for 80% of the instances. Fig. 2 shows the relative error of `pwadoo` and `glbDirect` given different number of iterations for all 60 PWA function instances. We can see that the rate of convergence of `pwadoo` is slower than `glbDirect`. This is because the Lipschitz constant is used in the DOO algorithm. The experiments show that DOO finds an approximation solution close to the optimal solution requiring computation time less than that of the MILP solvers taking to find the optimal solution. Hence, we propose to use DOO instead of the MILP method to solve the optimization problem of the PWA function for the case that $N$ is much larger than the dimension of the feasible set.

## 7 Conclusions

In this paper, we have considered the optimization of a continuous nonconvex PWA function over a polytope. We have proposed an optimistic-optimization-based approach to solve the given problem. In particular, by employing Delaunay triangulation and edgewise subdivision, we have constructed a partition of the feasible set

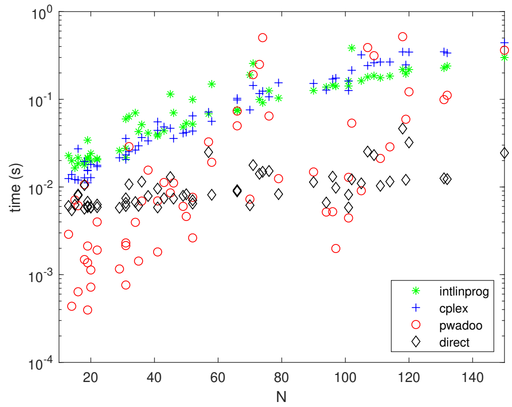
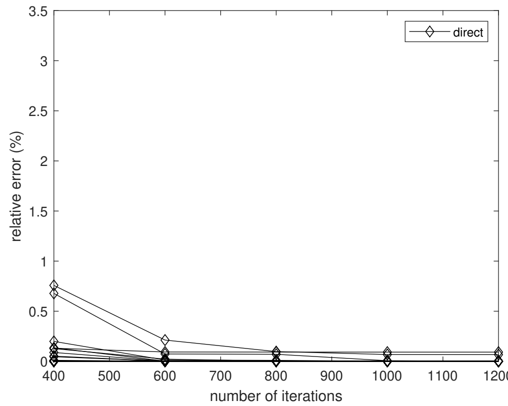
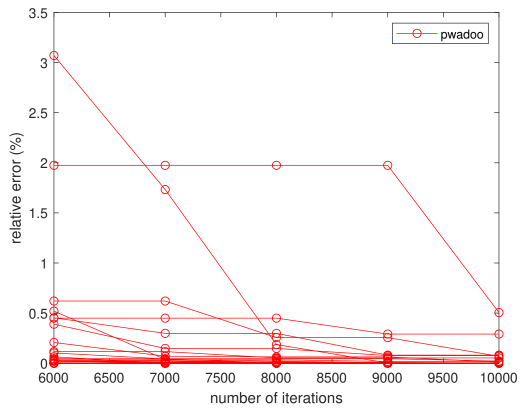



Figure 1. CPU time of `intlinprog`, `cplex`, `pwadoo` and `glbDirect` for the optimization of PWA functions ($N$ is the number of polyhedral subregions of PWA functions)




| number of iterations | 400 | 600 | 800 | 1000 | 1200 |
|---|---|---|---|---|---|
| CPU time (s) | 0.0633 | 0.0922 | 0.1259 | 0.1610 | 0.1965 |




| number of iterations | 6000 | 7000 | 8000 | 9000 | 10000 |
|---|---|---|---|---|---|
| CPU time (s) | 0.5365 | 0.6909 | 0.8685 | 1.0663 | 1.2954 |

Figure 2. Relative error and average CPU times of `glbDirect` (top) and `pwadoo` (bottom) for all 60 PWA function instances where `glbDirect` is written in object code and `pwadoo` is written in Matlab

satisfying the requirements for optimistic optimization. We have also derived the analytic expressions for the core parameters. Numerical examples have been implemented to test the proposed approach. Compared with the MILP based methods, the optimistic-optimization-based approach is more efficient especially for large problems. The partitioning scheme developed in this paper is the only way we have found currently satisfying all the requirements of optimistic optimization. In the future, we will search for other suitable partitioning schemes. Furthermore, we will investigate the performance of optimistic optimization algorithms that do not require the knowledge of the Lipschitz constant, such as the simultaneous optimistic optimization (SOO) algorithm. We will apply both DOO and SOO to the optimization problem of continuous PWA functions and compare their performance.

## Acknowledgements

Research supported by the Chinese Scholarship Council grant and the National Science Foundation of China (Grant No. 62003245).

## References

[1] S. Azuma, J. Imura, and T. Sugie. Lebesgue piecewise affine approximation of nonlinear systems. *Nonlinear Analysis-Hybrid Systems*, 4(1):92–102, 2010.

[2] A. Bemporad and M. Morari. Control of systems integrating logic, dynamics, and constraints. *Automatica*, 35(3):407–427, 1999.

[3] J. Bey. Simplicial grid refinement: On Freudenthal's algorithm and the optimal number of congruence classes. *Numerische Mathematik*, 85(1):1–29, 2000.

[4] F. Borrelli. *Constrained Optimal Control of Linear and Hybrid Systems.* Springer, Berlin, 2003.

[5] V. Breschi, D. Piga, and Bemporad A. Piecewise affine regression via recursive multiple least squares and multicategory discrimination. *Automatica*, 73:155–162, 2016.

[6] S.-W. Cheng, T.K. Dey, and J.R. Shewchuk. *Delaunay Mesh Generation.* CRC Press, 2012.

[7] K.L. Croxton, B. Gendron, and T.L. Magnanti. A comparison of mixed-integer programming models for nonconvex piecewise linear cost minimization problems. *Management Science*, 49(9):1268–1273, 2003.

[8] K.L. Croxton, B. Gendron, and T.L. Magnanti. Models and methods for merge-in-transit operations. *Transportation Science*, 37(1):1–22, 2003.

[9] H. Edelsbrunner and D.R. Grayson. Edgewise subdivision of a simplex. *Discrete & Computational Geometry*, 24(4):707–719, 2000.

[10] T. Fujisawa and E.S. Kuh. Piecewise-linear theory of nonlinear networks. *SIAM Journal on Applied Mathematics*, 22(2):307–328, 1972.

[11] E.N. Gonçalves, R.M. Palhares, R.H.C. Takahashi, and R.C. Mesquita. New approach to robust $\mathcal{D}$-stability analysis of linear time-invariant systems with polytope-bounded uncertainty. *IEEE Transactions on Automatic Control*, 51(10):1709–1714, 2006.

[12] J.-B. Grill, M. Valko, and R. Munos. Black-box optimization of noisy functions with unknown smoothness. In *Advances in Neural Information Processing Systems 28*, pages 667–675. 2015.

[13] D.R. Jones, C.D. Perttunen, and B.E. Stuckman. Lipschitzian optimization without the Lipschitz constant. *Journal of Optimization Theory and Application*, 79(1):157–181, 1993.

[14] R. Munos. Optimistic optimization of a deterministic function without the knowledge of its smoothness. In *25th Annual Conference on Neural Information Processing Systems*, pages 783–791, Granada, Spain, December 2011.

[15] R. Munos. From bandits to Monte-Carlo tree search: The optimistic principle applied to optimization and planning. *Foundations and Trends in Machine Learning*, 7(1):1–130, 2014.

[16] D. Piga, A. Bemporad, and A. Benavoli. Rao-Blackwellized sampling for batch and recursive Bayesian inference of Piecewise Affine models. *Automatica*, 117:109002, 2020.

[17] F.P. Preparata and M.I. Shamos. *Computational Geometry: An Introduction.* Springer-Verlag, New York, 1985.

[18] N. Rizk, A. Martel, and A. Ramudhin. A Lagrangean relaxation algorithm for multi-item lot-sizing problems with joint piecewise linear resource costs. *International Journal of Production Economics*, 102(2):344–357, 2006.

[19] S. Sridhar, J. Linderoth, and J. Luedtke. Models and solution techniques for production planning problems with increasing byproducts. *Journal of Global Optimization*, 59(2–3):597–631, 2014.

[20] M. Valko, A. Carpentier, and R. Munos. Stochastic simultaneous optimistic optimization. In *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, volume 28, pages 19–27, Atlanta, Georgia, USA, 2013.

[21] J.P. Vielma, S. Ahmed, and G. Nemhauser. Mixed-integer models for nonseparable piecewise-linear optimization: Unifying framework and extensions. *Operations Research*, 58(2):303–315, 2010.

[22] J. Xu, T. van den Boom, and B. De Schutter. Optimistic optimization for model predictive control of max-plus linear systems. *Automatica*, 74:16–22, December 2016.

[23] J. Xu, T.J.J. van den Boom, L. Buşoniu, and B. De Schutter. Model predictive control for continuous piecewise affine systems using optimistic optimization. In *2016 American Control Conference*, pages 4482–4487, Boston, Massachusetts, July 2016.