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

July 2015

Delft Center for Systems and Control
Delft University of Technology
Mekelweg 2, 2628 CD Delft
The Netherlands
phone: +31-15-278.51.19 (secretary)
fax: +31-15-278.66.79
URL: http://www.dcsc.tudelft.nl

*See http://pub.deschutter.info/abs/15_030.html for up-to-date information on this report.

Optimistic optimization for continuous nonconvex piecewise affine functions ⋆

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

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

Abstract

This paper considers the global optimization problem 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 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 founded on optimistic optimization, which is based on hierarchical partitioning of the feasible set and which focuses on the most promising region when searching for the global optimum. The advantage of optimistic optimization is that one can guarantee bounds on the suboptimality with respect to the global optimum given a finite computational budget (e.g. the number of iterations). In particular, the gap between the best value returned by the proposed algorithm and the real optimum can be made arbitrarily small as the computational budget increases. We derive the analytic expressions for the core parameters required by optimistic optimization for continuous PWA functions. The efficiency of the resulting algorithm is illustrated with numerical examples.

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 [1,19,22]; they also appear as cost functions of numerous optimization problems, see [8,17,21]. Moreover, the model predictive control law for discrete-time linear time-invariant systems with constraints on inputs and states can be explicitly expressed as a continuous PWA function of the initial state [2]. During the last decades, optimization of PWA functions has been investigated by several authors. A traditional technique for the optimization of *convex* PWA function subject to linear constraints consists in transforming the problem into single equivalent linear programming (LP) problem and then applying LP methods. Moreover, some LP methods are extended to directly deal with the optimization of convex PWA function without resorting to LP problems, e.g. the simplex algorithm [11] and the interior point algorithm [6].

The optimization of *nonconvex* PWA functions are often described as mixed integer linear programming (MILP) problems [7,23]. 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.

In this paper, 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 [15,16] 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. 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. A sequence of feasible solutions are generated during the process of iterations and the best solution is returned at the end of the algorithm. 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. In our previous paper [24], we have extended optimistic

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

Email addresses: j.xu-3@tudelft.nl (Jia Xu), a.j.j.vandenboom@tudelft.nl (Ton van den Boom), b.deschutter@tudelft.nl (Bart De Schutter).

optimization to solve the model predictive control problem for max-plus linear systems. Moreover, optimistic optimization has been used to solve the consensus problem in multi-agent systems [4]. The ideas of optimistic optimization have also been applied to planning algorithms, called optimistic planning [5,14].

In order to use optimistic optimization, we need a suited semi-metric that is the foundation of the requirements for optimistic optimization. A partition of the given polytope is also required to perform the search process. The partitioning should generate well-shaped cells that shrink with the depth. In this paper, we will deal with these challenges specifically for continuous PWA functions. We develop a dedicated semi-metric based on the knowledge of the Lipschitz constants of the PWA function. In addition, to establish the partitioning framework, we first employ Delaunay triangulation to divide the polytope into a mesh of simplices and next use edgewise subdivision to subdivide the simplices into smaller simplices that satisfy the requirements for optimistic optimization. We illustrate the effectiveness of the resulting algorithm with numerical examples and show that using optimistic-optimization-based algorithm 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. The second example shows that the proposed approach is also efficient for the optimization of max-min-plus-scaling functions, which are equivalent to continuous PWA functions.

This paper is organized as follows. In Section 2, we give some definitions. In Section 3, we describe the optimization problem of continuous PWA functions. In Section 4, we introduce optimistic optimization and the partitioning framework. We also derive the analytic expressions for the core parameters of optimistic optimization. In Section 5, the proposed approach is assessed with numerical examples.

2 Preliminaries

Let $\mathbb{R}^n$ and $\|\cdot\|_2$ denote the n-dimensional Euclidean space and the Euclidean norm. This section presents some necessary definitions, which are based on [20].

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 $\mathcal{P} = \{x \in \mathbb{R}^n | Ax \leq b\}$ is called a polytope, for some matrix A and some vector b. The polytope $\mathcal{P}$ can also be defined as the convex hull of a finite number of points and can be written as*

$$\mathcal{P} = \left\{ \sum_{i=1}^{V_\mathcal{P}} \lambda_i v_i \middle| \lambda_i \geq 0, i = 1, \ldots, V_\mathcal{P}, \sum_{i=1}^{V_\mathcal{P}} \lambda_i = 1 \right\} ,$$

where v_i denotes the i-th vertex of $\mathcal{P}$ and $V_\mathcal{P}$ is the total number of vertices of $\mathcal{P}$.

Definition 3 (Simplex) *An m-simplex $\mathcal{S} \subset \mathbb{R}^n$ with $0 \leq m \leq n$ is the convex hull of $m+1$ affinely independent points $v_0, \ldots, v_m \in \mathbb{R}^n$ which are its vertices. It can be written as*

$$\mathcal{S} = \left\{ \sum_{i=0}^{m} \lambda_i v_i \middle| \lambda_i \geq 0, i = 0, \ldots, m, \sum_{i=0}^{m} \lambda_i = 1 \right\} .$$

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 volume of $\mathcal{S}$ is

$$\mathrm{vol}(\mathcal{S}) = \frac{1}{n!} \left| \det(e_1, e_2, \ldots, e_n) \right| . \tag{1}$$

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) $\mathcal{P}_i \bigcap \mathcal{P}_j = \emptyset$ for all $i \neq j$.*

Definition 5 (PWA function) *A 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 $i = 1, \ldots, N$.*

3 Optimization of PWA functions

Consider the following optimization problem

$$\min_x \quad f(x) \tag{2}$$

$$\text{subject to} \quad Ax \leq b , \tag{3}$$

where $A \in \mathbb{R}^{m \times n}$ 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 $\alpha_{(i)} \in \mathbb{R}^n$, $\beta_{(i)} \in \mathbb{R}$, $i = 1, \ldots, N$. We assume that the feasible set $\mathcal{X} = \{x \in \mathbb{R}^n | Ax \leq b\} \subset \mathcal{P}$ is nonempty and bounded. From Definition 2, $\mathcal{X}$ is a polytope. If f is convex, then the problem (2)-(3) is equivalent to LP problems, which can be solved very efficiently.

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. For this case, one possible solution approach consists in transforming the problem (2)-(3) 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 optimistic optimization for the problem (2)-(3). The knowledge of a Lipschitz constant of f is important for designing the semi-metric ℓ, that is a key ingredient of optimistic optimization. For any $x, y \in \mathcal{P}_i$, we have

$$\left| f(x) - f(y) \right| \leq \|\alpha_{(i)}\|_2 \|x - y\|_2 . \tag{4}$$

It is easy to verify that $\max_{i=1,\ldots,N} \|\alpha_{(i)}\|_2$ is the smallest Lipschitz constant of f [12].

4 Optimistic optimization of PWA functions

We first introduce the background of optimistic optimization and next derive the analytic expressions for the core parameters satisfying the requirements.

4.1 Optimistic optimization

Now we introduce optimistic optimization [15] for the minimization of any function f over a given set $\mathcal{X}$. The notations f and $\mathcal{X}$ remain generic in this subsection. The implementation of optimistic optimization is based on a hierarchical partitioning of $\mathcal{X}$. For any integer $h = 0, 1, \ldots$, the set $\mathcal{X}$ is split into K^h cells with K a finite positive integer. This partition may be represented by a tree structure, thus K is the number of branches at each fork. Each cell is denoted as $X^{h,d}$, $d = 0, \ldots, K^h - 1$, and corresponds to a node (h, d) in the tree (with h the depth and d the index). Each node (h, d) contains K children nodes $(h+1, d_i)$, $i = 1, \ldots, K$, and the cells $\{X^{h+1,d_i} | i = 1, \ldots, K\}$ form a partition of the parent cell $X^{h,d}$. The root node of the tree corresponds to the whole region $\mathcal{X}$, denoted as $X^{0,0}$. Each cell $X^{h,d}$ is labeled by a point $x^{h,d}$ where f may be evaluated.

Definition 6 (Semi-metric) *A semi-metric on a set $\mathcal{X}$ is a function $\ell : \mathcal{X} \times \mathcal{X} \to \mathbb{R}^+$ satisfying the following conditions for any $x, y \in \mathcal{X}$: 1) $\ell(x, y) = \ell(y, x) \geq 0$; 2) $\ell(x, y) = 0$ if and only if $x = y$.*

Requirements for optimistic optimization. The following conditions, expressed in terms of a semi-metric ℓ, need to be satisfied:

1. There exists a semi-metric ℓ defined on $\mathcal{X}$ such that for all $x \in \mathcal{X}$, $f(x) - f(x^*) \leq \ell(x, x^*)$, where $f(x^*) = \min_{x \in \mathcal{X}} f(x)$.

2. There exists a decreasing sequence $\{\delta(h)\}_{h=0}^{\infty}$ with $\delta(h) > 0$, such that for any depth $h = 0, 1, \ldots$, for any cell $X^{h,d}$ at depth h, we have $\sup_{x \in X^{h,d}} \ell(x, x^{h,d}) \leq \delta(h)$, where $\delta(h)$ is called the maximum diameter of the cells at depth h.

3. There exists a scalar $\nu > 0$ such that any cell $X^{h,d}$ at any depth h contains an ℓ-ball of radius $\nu\delta(h)$ centered in $x^{h,d}$.

The requirements guarantee bounds on the suboptimality with respect to the global optimum and on the computational budget (e.g. number of iterations). In particular, Requirement 1 regard the local property of f near the optimum with respect to the semi-metric ℓ. Requirements 2-3 guarantee that the partitioning of the feasible set generates well-shaped cells that shrink with further partitioning. Loosely speaking, this means that the value of $\delta(h + 1)/\delta(h)$ should be less than a given constant that is strictly smaller than 1. The scalar ν corresponds to the ratio of the radius of the inscribed ball of any cell and the maximum distance between any two points in that cell.

Algorithm 1 Determine Optimistic Optimization

Given: number of iterations n, partitioning of $\mathcal{X}$

Initialize the tree $\mathcal{T} = \{(0, 0)\}$ (root node)

for $t = 1$ to n **do**

  Select the leaf (h, d) with minimum $b^{h,d}$ value

  Expand this leaf by adding its K children to $\mathcal{T}$

end for

Return $x(n) = \arg\max_{(h,d) \in \mathcal{T}} f(x^{h,d})$

The optimistic optimization algorithm is summarized in Algorithm 1 [15]. For each cell $X^{h,d}$, define $b^{h,d} = f(x^{h,d}) - \delta(h)$. From the Requirement 1-2, for any cell $X^{h,d}$ containing the optimal solution x^*, we have $b^{h,d} \leq f(x^{h,d}) - \ell(x, x^{h,d}) \leq f(x^*)$, $\forall x \in X^{h,d}$. Hence, the minimum $b^{h,d}$ is actually a lower bound of the minimum value of f. Expanding a leaf (h, d) corresponds to splitting the cell $X^{h,d}$ into K sub-cells.

4.2 Hierarchical partition of the feasible set

As mentioned in Section 3, the feasible set $\mathcal{X} = \{x \in \mathbb{R}^n | Ax \leq b\}$ of the problem (2)-(3) is a polytope. To implement the algorithm, we need to partition $\mathcal{X}$ into a collection of simplices that satisfy the requirements for optimistic optimization. The partitioning consists of two parts: initialization of the partition (i.e. dividing the polytope $\mathcal{X}$ into simplices) and refinement of the partition (i.e. subdividing each simplex into smaller simplices). In this paper, we propose to use Delaunay triangulation to divide $\mathcal{X}$ into a set of simplices. Next for the subdivision of each simplex, we use the edgewise subdivision method.

Edgewise subdivision [10] divides a simplex $\mathcal{S}$ of $\mathbb{R}^n$ into k^n n-simplices, where each edge of $\mathcal{S}$ is cut into k equal pieces. A ready-to-implement algorithm for edgewise subdivision is presented in [13].

Properties of edgewise subdivision. For every integer $k \geq 1$, the edgewise subdivision of $\mathcal{S}$ has the following properties [10]:

(1) all generated simplices have the same n-dimensional volume; (2) all generated simplices fall into at most $n!/2$ congruence classes; (3) the faces of $\mathcal{S}$ are subdivided the same way; (4) repeated subdivision has the same effect as increasing k.

To satisfy the Requirements 2-3 for optimistic optimization, the diameter of simplices should decrease with further refinement and the interior angles of simplices should not tend to zero. By selecting an appropriate k, the properties of edgewise subdivision will allow the satisfaction of the requirements. The selection of k will be discussed in the next subsection.

4.3 PWA optimistic optimization

In this subsection, we prove that optimistic optimization can be used to solve the problem (2)-(3). We first derive a suited semi-metric ℓ for the PWA function over $\mathcal{X}$ and next give the procedure to determine $\delta(h)$ and ν. Some of the symbols and acronyms that occur frequently in this section are listed in Appendix B.

Proposition 7 *For any $x, y \in \mathcal{X}$, define $\ell(x,y) = \alpha\|x-y\|_2$, where $\alpha = \max_{i=1,\ldots,N} \|\alpha_{(i)}\|_2$ and $\alpha_{(i)}$ are defined as in Definition 5. Then the function ℓ is a semi-metric defined on $\mathcal{X}$. Let $x^* \in \mathcal{X}$ be a global optimizer for the problem (2)-(3). Then we have $f(x) - f(x^*) \leq \alpha\|x - x^*\|_2$.*

PROOF. It is easy to verify that the function ℓ satisfies the conditions of Definition 6 of a semi-metric and $f(x) - f(x^*) \leq \alpha\|x-x^*\|_2$ is implied by (4). □

The design of the semi-metric ℓ 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, although in general this results in the loss of accuracy.

Proposition 7 shows that Requirement 1 for optimistic optimization is satisfied. Next, we verify Requirements 2-3, which concern the shape of the partition of $\mathcal{X}$. As mentioned in Section 4.2, we use Delaunay triangulation to initialize the partitioning and then use edgewise subdivision to refine the partitioning. We will prove that the partitioning following this strategy satisfy the requirements for optimistic optimization.

Using Delaunay triangulation, the feasible set $\mathcal{X}$ 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 (1)-(4) of edgewise subdivision given in Section 4.2 are essential for the remaining proof. Edgewise subdivision divides $\mathcal{X}_s$ into k^n n-simplices, so the branching factor K of optimistic optimization equals k^n. Note that h is the depth of the subdivision (indicator of the recursion of edgewise subdivision) and d is the index of simplex 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 $N_s \leq n!/2$ be the number of congruence classes that all simplices generated by repeated edgewise subdivision of $\mathcal{X}_s$ fall into (see Property (2)). Let $C_{s,i}$, $i = 1,\ldots,N_s$, be the representative simplices, scaled such that their maximal edge length equals 1, of each congruence classes. Define the ratio between the maximum and minimum volumes among representative simplices as

$$\gamma_s = \max_{i,j=1,\ldots,N_s} \frac{\mathrm{vol}(C_{s,i})}{\mathrm{vol}(C_{s,j})} \ . \tag{5}$$

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

$$\rho_s = \min_{i=1,\ldots,N_s} \rho_{s,i} \ . \tag{6}$$

Let $v_{s,0},\ldots,v_{s,n}$ be the vertices of $\mathcal{X}_s$. Let $v_{s,0}^{h,d},\ldots,v_{s,n}^{h,d}$ be the vertices of $X_s^{h,d}$. Define $e_{s,i} = v_{s,i} - v_{s,i-1}$ and $e_{s,i}^{h,d} = v_{s,i}^{h,d} - v_{s,i-1}^{h,d}$, $i = 1,\ldots,n$. Then taking into account the proof of the independence lemma in [10] as well as the fact that repeated subdivision has the same effect as increasing k (see Property (4)), there exists a permutation $\pi_s^{h,d}$ of $\{1,\ldots,n\}$ such that $e_{s,i}^{h,d} = \frac{1}{k^h} e_{s,\pi_s^{h,d}(i)}$. Note that we have

$$v_{s,i}^{h,d} - v_{s,0}^{h,d} = e_{s,i}^{h,d} + e_{s,i-1}^{h,d} + \cdots + e_{s,1}^{h,d} \ . \tag{7}$$

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 (7), we have $|v_{s,j}^{h,d} - v_{s,i}^{h,d}| = |e_{s,j}^{h,d} + e_{s,j-1}^{h,d} + \cdots + e_{s,i+1}^{h,d}| = \frac{1}{k^h}|e_{s,\pi_s^{h,d}(j)} + e_{s,\pi_s^{h,d}(j-1)} + \cdots + e_{s,\pi_s^{h,d}(i+1)}|$. Define

$$\theta_{s,\min} = \min_{i=1,\ldots,n} |e_{s,i}| \ , \qquad \theta_{s,\max} = \sum_{i=1}^{n} |e_{s,i}| \ . \tag{8}$$

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

$$\frac{1}{k^h}\theta_{s,\min} \leq |v_{s,j}^{h,d} - v_{s,i}^{h,d}| \leq \frac{1}{k^h}\theta_{s,\max} \ . \tag{9}$$

Lemma 8 *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 is the index set of simplices at depth h. Then we have*

$$\frac{L_s^{h+1}}{L_s^h} \leq \frac{1}{k}\gamma_s^{1/n} \ , \qquad \frac{r_s^h}{L_s^h} \geq \frac{\theta_{s,\min}\rho_s}{\theta_{s,\max}} \tag{10}$$

where γ_s, ρ_s, $\theta_{s,\min}$ and $\theta_{s,\max}$ are as defined in (5), (6), (8) 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 representative simplex $C_{s,i}$. The maximum edge length of $C_{s,i}$ equals 1. From Property (4), 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 (1), we have $\mathrm{vol}(X_s^{h,d'}) = \mathrm{vol}(\mathcal{X}_s)/k^{hn}$. 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 (1), resulting in a multiplication with $(1/L_s^h)^n$ compared to the original expression. Hence, we have

$$\mathrm{vol}(C_{s,i}) = \Big(\frac{1}{L_s^h}\Big)^n \mathrm{vol}(X_s^{h,d'}) = \Big(\frac{1}{L_s^h}\Big)^n \frac{\mathrm{vol}(\mathcal{X}_s)}{k^{hn}} \ . \tag{11}$$

Likewise let $X_s^{h+1,d''}$ be the simplex that has the maximum edge length L_s^{h+1} among all simlices at depth $h+1$ and assume that $X_s^{h+1,d''}$ belongs to congruence class j with representative simplex $C_{s,j}$. So

$$\mathrm{vol}(C_{s,j}) = \Big(\frac{1}{L_s^{h+1}}\Big)^n \frac{\mathrm{vol}(\mathcal{X}_s)}{k^{(h+1)n}} \ . \tag{12}$$

Thus (11) and (12) result in $\big(\frac{L_s^{h+1}}{L_s^h}\big)^n = \frac{1}{k^n}\frac{\mathrm{vol}(C_{s,i})}{\mathrm{vol}(C_{s,j})}$ and then $\frac{L_s^{h+1}}{L_s^h} = \frac{1}{k}\big(\frac{\mathrm{vol}(C_{s,i})}{\mathrm{vol}(C_{s,j})}\big)^{1/n} \leq \frac{1}{k}\gamma_s^{1/n}$.

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 representative simplex $C_{s,l}$. The maximum edge length of $C_{s,l}$ equals 1 and the inradius of $C_{s,l}$ is $\rho_{s,l}$. Thus, we have $r_s^h = L_s^{h,d^\sharp}\rho_{s,l}$. Due to (6), we also have $r_s^h \geq L_s^{h,d^\sharp}\rho_s$. Note that (9) 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, $r_s^h \geq L_s^{h,d^\sharp}\rho_s \geq \frac{1}{k^h}\theta_{s,\min}\rho_s$ and thus $\frac{r_s^h}{L_s^h} \geq \frac{\frac{1}{k^h}\theta_{s,\min}\rho_s}{L_s^h} \geq \frac{\frac{1}{k^h}\theta_{s,\min}\rho_s}{\frac{1}{k^h}\theta_{s,\max}} \geq \frac{\theta_{s,\min}\rho_s}{\theta_{s,\max}}$. □

Proposition 9 *Define* $\delta_s(h) = \alpha L_s^h$*. If* k *is selected as an integer that is strictly larger than* $\gamma_s^{1/n}$*, then* $\{\delta_s(h)\}_{h=0}^{\infty}$ *is a decreasing positive sequence such that* $\sup_{x\in X_s^{h,d}} \ell(x, x_s^{h,d}) \leq \delta_s(h)$*.*

PROOF. If k is selected as an integer that is strictly larger than $\gamma_s^{1/n}$, from (10), we have $\frac{\delta_s(h+1)}{\delta_s(h)} = \frac{L_s^{h+1}}{L_s^h} \leq \frac{1}{k}\gamma_s^{1/n} < 1$. Hence, $\{\delta_s(h)\}_{h=0}^{\infty}$ is a decreasing positive sequence. Furthermore, for any $x \in X_s^{h,d}$, $\ell(x, x_s^{h,d}) = \alpha\|x - x_s^{h,d}\|_2 \leq \alpha L_s^{h,d} \leq \delta_s(h)$. □

Proposition 10 *Select* ν_s *such that* $0 < \nu_s \leq \frac{\theta_{s,\min}\rho_s}{\theta_{s,\max}}$*. Then* $X_s^{h,d}$ *contains an* ℓ*-ball of radius* $\nu_s\delta_s(h)$ *centered in* $x_s^{h,d}$*.*

PROOF. An ℓ-ball of radius $\nu_s\delta_s(h)$ centered in $x_s^{h,d}$ can be written as $\mathfrak{B} = \{x \in \mathcal{X}_s | \ell(x, x_s^{h,d}) = \alpha\|x - x_s^{h,d}\|_2 \leq \nu_s\delta_s(h)\}$. If we select $\nu_s \leq \frac{\alpha r_s^{h,d}}{\delta_s(h)}$, then we have $\|x - x_s^{h,d}\|_2 \leq \frac{\nu_s\delta_s(h)}{\alpha} \leq r_s^{h,d}$ for all $x \in \mathfrak{B}$. Hence, $\mathfrak{B} \subset X_s^{h,d}$. Note that $\delta_s(h) = \alpha L_s^h$. From 10, we have $\frac{\alpha r_s^{h,d}}{\delta_s(h)} = \frac{r_s^{h,d}}{L_s^h} \geq \frac{r_s^h}{L_s^h} \overset{(10)}{\geq} \frac{\theta_{s,\min}\rho_s}{\theta_{s,\max}}$. Thus if we choose $0 < \nu_s \leq \frac{\theta_{s,\min}\rho_s}{\theta_{s,\max}}$, then $\nu_s \leq \frac{\alpha r_s^{h,d}}{\delta_s(h)}$ and then $X_s^{h,d}$ contains an ℓ-ball of radius $\nu_s\delta_s(h)$ centered in $x_s^{h,d}$. □

Based on Proposition 9-10, denote $\delta(h) = \max_{s=1,\ldots,N_t} \delta_s(h)$ and $\nu = \min_{s=1,\ldots,N_t} \nu_s$. We can see that Requirements 2-3 are now satisfied. Up to now, we have addressed all the challenges for using optimistic optimization to solve the optimization problem of continuous nonconvex PWA functions. The effectiveness of the proposed approach will be illustrated in the next section.

5 Numerical example

In this section, we evaluate the optimistic-optimization-based approach and compare with other methods. All experiments are implemented in MATLAB 2014b on an 3.1 GHz processor with 3.7 GB RAM.

Example 5.1 The instances considered include a set of randomly generated continuous PWA functions

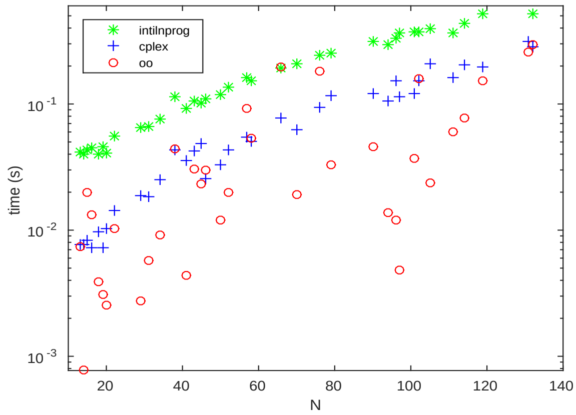

Fig. 1. CPU time of intlinprog, cplex and optimistic optimization (oo) for the optimization of PWA functions

$f : \mathbb{R}^2 \to \mathbb{R}$ in which the vector pairs $\alpha_{(i)} \in \mathbb{R}^2, \beta_{(i)} \in \mathbb{R}$, $i = 1, \ldots, N$, contain pseudorandom values drawn from the standard normal distribution. We totally generate 50 random continuous PWA functions. Because of randomness, it may happen that several PWA functions have the same N (number of polyhedral subregions). We randomly select one from the instances having the same N. Eventually, 36 PWA functions are used. We consider the optimization of these PWA functions over a rectangle region $\mathcal{X} = [x_{\min}, x_{\max}] \times [y_{\min}, y_{\max}]$. Below we compare the efficiency of the optimistic-optimization-based approach and the MILP based approach. The corresponding MILP problem is derived based on the techniques in [3] (see Appendix A for details) and solved with the `intlinprog` function in Matlab Optimization Toolbox and the `cplex` function in TOMLAB. The optimistic-optimization-based approach is implemented as the `pwaoo` functin in MATLAB. Note that the `pwaoo` and `intlinprog` functions are both Matlab functions and the `cplex` function is implemented in C. Fig. 1 shows the semi logarithmic plot of CPU time (average over 10 runs) of the three solvers. The function values of f returned from the three solvers are denoted as f_{oo}, $f_{\text{intlinprog}}$ and f_{cplex}, where $f_{\text{intlinprog}}$ and f_{cplex} of every instance are equal. The iteration in the `pwaoo` function is stopped if the gap between f_{cplex} and f_{oo} is less than 5% (the gap is calculated as $100|(f_{\text{oo}} - f_{\text{cplex}})/f_{\text{cplex}}|$). The gap can be made arbitrarily small by increasing the number of iterations. We can see that the `pwaoo` function is faster than `intlinprog` except some special cases and is faster than `cplex` for 70% of the instances.

Example 5.2 Any continuous PWA function can be represented as a min-max or max-min composition of its affine components [18], which is similar to the canonical form of max-min-plus-scaling (MMPS) functions. As presented in [9], the optimization of MMPS functions can be written as a finite sequence of LP problems while the worst-case complexity is largely determined by the number of affine terms in equivalent canonical form of

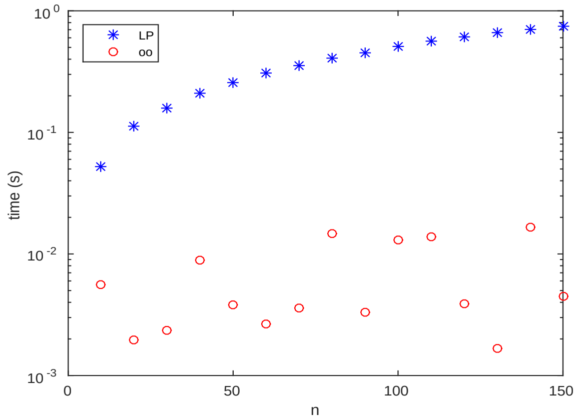

Fig. 2. CPU time of linprog and optimistic optimization (oo) for the optimization of MMPS functions

the MMPS expression. We consider an MMPS function written as $g(x) = \min_{i=1\ldots n} \max_{j=1\ldots n} \{\alpha_{ij}^T x + \beta_{ij}\}, \quad \forall x \in \mathbb{R}^2$, where $\alpha_{ij} \in \mathbb{R}^2, \beta_{ij} \in \mathbb{R}$ contains pseudorandom values drawn from the standard normal distribution. We use the `linprog` function of TOMLAB to solve the set of LPs resulting from the minimization problem of g. The optimistic-optimization-based approach is implemented as the `mmpsoo` function in MATLAB. Fig. 2 shows the semi logarithmic plot of CPU time (average over 10 runs) of the LP based approach and the optimistic-optimization-based approach for increasing n. The gap between the function value g_{lp} returned by `linprog` and g_{oo} returned by `mmpsoo` is guaranteed within 5% (the gap is calculated as $100|(g_{\mathrm{lp}} - g_{\mathrm{oo}})/g_{\mathrm{lp}}|$). We can see that using the `mmpsoo` function is more efficient than solving a sequence of LPs.

6 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, we have developed a dedicated semi-metric needed for optimistic optimization for PWA functions. By employing Delaunay triangulation and edgewise subdivision, we have constructed a partition of the feasible set 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.

In the future, we will investigate applying the optimistic-optimization-based approach to improve the efficiency of current model predictive control approaches for continuous PWA systems. Moreover, the proposed approach is in the deterministic setting. We will also investigate the stochastic setting.

Acknowledgements

Research supported by the Chinese Scholarship Council.

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, F. Borrelli, and M. Morari. Model predictive control based on linear programming - the explicit solution. *IEEE Transactions on Automatic Control*, 47(12):1974–1985, 2002.

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

[4] L. Buşoniu and I.-C. Morărescu. Consensus for black-box nonlinear agents using optimistic optimization. *Automatica*, 50(4):1201–1208, 2014.

[5] L. Buşoniu, R. Munos, and R. Babuska. A review of optimistic planning in Markov decision processes. In *Reinforcement Learning and Adaptive Dynamic Programming for feedback control*, pages 494–516. Wiley, 2012.

[6] M.C. Cavichia and M.N. Arenales. Piecewise linear programming via interior points. *Computers & Operations Research*, 27(13):1303–1324, 2000.

[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] B. De Schutter and T. van den Boom. MPC for continuous piecewise-affine systems. *Systems & Control Letters*, 52(3–4):179–192, July 2004.

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

[11] R. Fourer. A simplex algorithm for piecewise-linear programming I – Derivation and proof. *Mathematical Programming*, 33(2):204–233, 1985.

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

[13] 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.

[14] K. Máthé, L. Buşoniu, R. Munos, and B. De Schutter. Optimistic planning with a limited number of action switches for near-optimal nonlinear control. In *53rd IEEE Conference on Decision and Control*, pages 3518–3523, Los Angeles, California, USA, December 2014.

[15] 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.

[16] 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.

[17] F.J. Nourie and F. Guder. A restricted-entry method for a transportation problem with piecewise-linear concave costs. *Computers & Operations Research*, 21(7):723–733, 1994.

[18] S. Ovchinnikov. Max-min representation of piecewise linear functions. *Beiträge zur Algebra und Geometrie/ Contributions to Algebra and Geometry*, 43(1):297–302, 2002.

[19] P. Pejovic and D. Maksimovic. A new algorithm for simulation of power electronic systems using piecewise-linear device models. *IEEE Transactions on Power Electronics*, 10(3):340–348, 1995.

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

[21] 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.

[22] 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.

[23] 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.

[24] J. Xu, B. De Schutter, and T. van den Boom. Model predictive control for max-plus-linear systems via optimistic optimization. In *12th International Workshop on Discrete Event Systems*, pages 111–116, Cachan, France, May 2014.

A The mixed integer linear programming (MILP) problem

This appendix presents the MILP form of the problem (2)-(3) based on the techniques in [3]. Define a binary variable $\omega_i \in \{0,1\}$ such that $[\omega_i = 1] \leftrightarrow [x \in \mathcal{P}_i]$. The polyhedral subregion $\mathcal{P}_i$ is of the form $\mathcal{P}_i = \{x \in \mathbb{R}^n | H_i x \leq g_i\}$ where $H_i \in \mathbb{R}^{m_i \times n}$ and $g_i \in \mathbb{R}^m_i$ are the constraint matrix and vector. Define a new variable $z_i = \omega_i x$ and then the corresponding MILP problem of problem (2)-(3) can be written as

$$
\begin{aligned}
&\min_{x,z,\omega} \qquad \sum_{i=1}^N [\alpha_{(i)}^T z_i + \beta_{(i)} \omega_i] \\
&\text{subject to} \\
&Ax \leq b, \\
&\omega_i x_{\min} \leq z_i \leq \omega_i x_{\max}, \\
&x - x_{\max}(1-\omega_i) \leq z_i \leq x - x_{\min}(1-\omega_i), \\
&\textstyle\sum_{i=1}^N \omega_i = 1, \\
&H_i x - g_i \leq U_i(1-\omega_i), \\
&\text{where } \omega_i \in \{0,1\},\ U_i = \max_{x\in\mathcal{X}} H_i x - g_i, i = 1,\ldots,N \\
&x_{\min,j} = \min_{x\in\mathcal{X}} x_j, x_{\max,j} = \max_{x\in\mathcal{X}} x_j, j = 1,\ldots,n.
\end{aligned}
\tag{A.1}
$$

B List of Symbols

$\mathcal{X}$	polytopic feasible set
$\{\mathcal{X}_s \| s = 1,\ldots,N_\mathrm{t}\}$	simplicial mesh of $\mathcal{X}$
$1/k$	factor of edgewise subdivision
n	dimension of $\mathcal{X}$
K	k^n, branching factor of optimistic optimization
h	subdivision depth
d	index of simplices at depth h
$X_s^{h,d}$	simplex at depth h of the edgewise subdivision of $\mathcal{X}_s$
$L_s^{h,d}$	maximum edge length of $X_s^{h,d}$
$r_s^{h,d}$	inradius of $X_s^{h,d}$
$x_s^{h,d}$	incenter of $X_s^{h,d}$
N_s	number of congruence classes of edgewise subdivision of $\mathcal{X}_s$
$C_{s,i}$, $i = 1,\ldots,N_s$	representative simplices of congruence classes
γ_s	ratio between the maximum and minimum volumes among the representative simplices
$\rho_{s,i}$	inradius of $C_{s,i}$
ρ_s	minimum of $\rho_{s,i}$
$v_{s,0},\ldots,v_{s,n}$	vertices of $\mathcal{X}_s$
$v_{s,0}^{h,d},\ldots,v_{s,n}^{h,d}$	vertices of $X_s^{h,d}$