Synthesis of Hybrid Automata with Affine Dynamics from Time-Series Data

Miriam García Soto  
IST Austria  
Klosterneuburg, Austria  
miriam.garciasoto@ist.ac.at

Thomas A. Henzinger  
IST Austria  
Klosterneuburg, Austria  
tah@ist.ac.at

Christian Schilling  
University of Konstanz  
Konstanz, Germany  
christian.schilling@uni-konstanz.de

ABSTRACT
Formal design of embedded and cyber-physical systems relies on mathematical modeling. In this paper, we consider the model class of hybrid automata whose dynamics are defined by affine differential equations. Given a set of time-series data, we present an algorithmic approach to synthesize a hybrid automaton exhibiting behavior that is close to the data, up to a specified precision, and changes in synchrony with the data. A fundamental problem in our synthesis algorithm is to check membership of a time series in a hybrid automaton. Our solution integrates reachability and optimization techniques for affine dynamical systems to obtain both a sufficient and a necessary condition for membership, combined in a refinement framework. The algorithm processes one time series at a time and hence can be interrupted, provide an intermediate result, and be resumed. We report experimental results demonstrating the applicability of our synthesis approach.

KEYWORDS
synthesis, hybrid automaton, linear dynamics, membership

1 INTRODUCTION
Formal design and verification of embedded control systems require a mathematical model capturing the dynamics of each component in the system. In general, embedded systems combine analog and digital components. The analog components evolve continuously in real time, while the digital components evolve in discrete time. An appropriate mathematical formalism for modeling systems with mixed continuous and discrete behavior is a hybrid automaton [16].

In this paper we propose an automated approach to synthesizing a hybrid automaton with affine continuous dynamics (abbreviated \(\text{ADHA}\)) from time-series data in an online fashion. The design of models from observed data has been extensively studied in control theory for autoregressive systems [4, 6, 11, 28, 34], which can be seen as discrete dynamical systems, in contrast to the continuous dynamics captured by a hybrid automaton. Most of these approaches process a single time-series or all data at once. In a setting where not all data is available at once, it is desirable to have an online approach that processes time-series data sequentially and iteratively updates a model; only a few approaches support this feature [15, 32, 33, 35].

Our synthesis approach operates in two phases. In the first phase we transform a (discrete) time-series into a piecewise continuous trajectory \(f\), for which we present an optimization procedure that allows to specify the error between the data and the trajectory. The trajectories \(f\) we consider are piecewise-affine (\(\text{PWA}\)) functions where each piece is the solution of an affine dynamical system of the form \(\dot{x} = Ax + b\). \(\text{PWA}\) trajectories can model a large class of physical processes and approximate generic nonlinear systems.

In the second phase, which is independent of how the continuous \(\text{PWA}\) trajectory \(f\) has been obtained, we synthesize an \(\text{ADHA}\) from \(f\). More precisely, we construct an \(\text{ADHA}\) from an existing \(\text{ADHA}\) (initialized with the “empty” \(\text{ADHA}\)) in two stages: 1) (membership) we determine whether the new trajectory is already captured by an execution of the model, up to a predefined precision, and 2) (model update) if the trajectory is not captured, we modify the model such that, after the modification, the new model captures the trajectory (and all trajectories that had been captured before).

We propose a three-step algorithm for the membership problem (“is a \(\text{PWA}\) trajectory captured by an \(\text{ADHA}\?)”). The first step is a reachability analysis inside a tube around the trajectory that we use to provide a negative answer. This problem has been studied in [33] for the class of hybrid automata with piecewise-constant dynamics. The second step is an optimization-based analysis that we use to provide a positive answer. The third step is a refinement procedure to deal with cases when the first two steps were not conclusive.

If we find that the \(\text{PWA}\) trajectory is not captured by the model in the membership query, we apply a model update by adding behavior to the automaton. We first try to relax the continuous constraints of the automaton (called invariants and guards). If this relaxation is not sufficient to capture the trajectory, we also apply structural changes to the automaton (adding transitions and locations).

In summary, we present algorithms to solve the following problems for \(\text{PWA}\) trajectories and \(\text{ADHAs}\) with a given precision:

- transforming time-series data to \(\text{PWA}\) trajectories (Section 4)
- membership of a \(\text{PWA}\) trajectory in an \(\text{ADHA}\) (Section 5)
- synthesizing an \(\text{ADHA}\) from \(\text{PWA}\) trajectories (Section 6)

Together, our algorithms form an end-to-end approach to the synthesis of an \(\text{ADHA}\) from time-series data with a given precision.

Related work. The synthesis of hybrid systems has been explored previously in different fields and is known as identification in the area of control theory (see the surveys [11, 28]) and as process mining and model learning to a broader research community. Most of the techniques focus on input-output models, such as switched autoregressive exogenous (\(\text{SARX}\)) [15, 25] and (\(\text{PWARX}\)) models [6, 9, 10, 17, 23, 31]. \(\text{SARX}\) models constitute a subclass of linear hybrid automata (which, unlike the \(\text{ADHA}\), only has dynamics with constant derivatives) with deterministic switching behavior and \(\text{PWARX}\) models are piecewise ARX models where the regressor space forms a state-space polyhedral partition. The aforementioned methods mainly consider single-input single-output (\(\text{SISO}\)) systems, whereas a few of them consider multiple-input multiple-output (\(\text{MIMO}\)) systems [4, 18, 34]. Other techniques identify piecewise affine systems in state-space form [2, 22, 34]. The identification techniques can also be classified into optimization-based methods [19, 26] clustering-based procedures [9, 23] and
algebraic approaches [4, 24]. Most of these methods are proposed for offline identification, with some exceptions [15, 32, 35]. We propose an online approach that synthesizes hybrid automata with affine dynamics, which are systems in state-space form.

In the field of computer science, we find techniques for learning models from traces, which refers to approaches based on learning finite-state machines [3] or other machine-learning techniques. Most approaches learn a (simpler) linear hybrid automaton. The work in [20] describes an abstract framework, based on heuristics, to learn offline from input-output traces by first learning the discrete structure and later adding continuous dynamics. Bartocci et al. learn shape expressions, which have a similar expressiveness [5]. A recent online approach provides soundness and precision guarantees [33]. However, that approach is restricted to linear hybrid automata, i.e., constant dynamics. We consider affine dynamics and follow a principled search algorithm for the automaton modification.

We are not aware of approaches that transform time-series data to continuous affine dynamical systems. Some approaches consider discrete-time models, such as the work by Willems for LTI systems [36], and other approaches for SABX models based on convex optimization [27] or generalized principal component analysis [4].

2 BASIC DEFINITIONS

Sets. Let $\mathbb{R}$, $\mathbb{R}_{\geq 0}$, and $\mathbb{N}$ denote the set of real numbers, non-negative real numbers, and natural numbers, respectively. Given a set $X$, the power set $\mathcal{P}(X)$ is the set of all subsets of $X$. We write $x$ for points $(x_1, \ldots, x_n) \in \mathbb{R}^n$. Given a point $x \in \mathbb{R}^n$ and $t \in \mathbb{R}_{\geq 0}$, we define the ball of radius $\varepsilon$ around $x$ as $B_t(x) = \{ y \in \mathbb{R}^n : \| x - y \| < \varepsilon \}$, where $\| \cdot \|$ is the infinity norm. Given two sets $P, P' \subseteq \mathbb{R}^n$, we define the distance between $P$ and $P'$ as $d(P, P') := \inf \{ \| x - y \| : x \in P, y \in P' \}$. Let $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$ be constant and $x$ be a variable in $\mathbb{R}^n$, and let $(a, x)$ denote the dot product of $a$ and $x$; then $(a, x) \cdot b$ is a linear constraint where $a \in \mathbb{R}^n$, the set $\{ (a, x) \cdot b = 0 \}$ is a hyperplane, and the set $\{ (a, x) \cdot b \leq 0 \}$ is a half-space. A (convex) polytope is a compact intersection of linear constraints. Equivalently, a polytope is the convex hull of a set of vertices $v_1, \ldots, v_m \in \mathbb{R}^n$, written $\text{chull}(\{v_1, \ldots, v_m\})$. For a polytope $P$ we denote the set of its linear constraints by $\text{constr}(P)$ and the set of its vertices by $\text{vert}(P)$. Let $\text{cpoly}(y)$ be the set of convex polytopes over $\mathbb{R}^n$.

Trees. A tree is a directed acyclic graph $T = (N, E)$ with finite set of nodes $N$, including a root node, and edges $E \subseteq N \times N$. A given node $v \in N$, the child nodes are $\text{children}(v) = \{ v' \in N : (v, v') \in E \}$.

Functions, dynamical systems, and trajectories. Given a function $f$, $\text{dom}(f)$ denote its domain. Let $f \downarrow D$ denote the restriction of $f$ to domain $D \subseteq \text{dom}(f)$. Given two functions $f$ and $g$ with $\text{dom}(f) = \text{dom}(g)$, the distance between $f$ and $g$ is denoted by $d(f, g)$ and defined as $\max_{x \in \text{dom}(f)} \| f(t) - g(t) \|$. We typically have $\text{dom}(f) = [0, T]$, where the initial and final states of $f$ correspond to $f(0)$ and $f(T)$ and are denoted by $f_0$ and $f_{end}$, respectively. A time series is a sampling $s : D \rightarrow \mathbb{R}^n$ over a finite time domain $D$.

A function $f : [0, T] \rightarrow \mathbb{R}^n$ is a piecewise-affine (pwa) trajectory with $k$ pieces if it is continuous and there is a tuple $(\mathcal{A}, \mathcal{B}, \mathcal{S})$ where $\mathcal{I}$ is a finite set of consecutive time intervals $[t_0, t_1], \ldots, [t_{k-1}, t_k]$ with $[0, T] = \bigcup_{i=0}^{k} [t_{i-1}, t_i]$, $\mathcal{A}$ and $\mathcal{B}$ are $k$-tuples of matrices $A_i \in \mathbb{R}^{n \times n}$ and vectors $b_i \in \mathbb{R}^n$, respectively, $i = 0, \ldots, k$, and $f|_{[t_{i-1}, t_i]}$ is a solution of the affine dynamical system $\dot{x} = A_i x + b_i$, where $\dot{x}$ denotes the derivative of $x$ with respect to $t$. We assume that pwa trajectories are given as the above tuple. We call $f|_{[t_{i-1}, t_i]}$ the pieces of $f$, and $t_k(f) := \{ t_0, \ldots, t_k \}$ the switching times of $f$. Each piece of $f$ is called an affine trajectory. A linear trajectory $f$ is a special case of an affine trajectory where $b = 0$.

2.1 Hybrid automaton with affine dynamics

We consider a particular class of hybrid automata [16] with invariants and guards given by linear constraints and with continuous dynamics given by affine differential equations.

Definition 2.1. An $n$-dimensional hybrid automaton with affine dynamics (ADHA) is a tuple $\mathcal{H} = (Q, E, \text{Flow}, \text{Inv}, \text{Grd})$, where 1) $Q$ is a finite set of locations, 2) $E \subseteq Q \times Q$ is a transition relation, 3) $\mathbb{R}^n$ is the continuous state space, 4) $\text{Flow} : Q \rightarrow \mathbb{R}^{n \times n} \times \mathbb{R}^n$ is the injective flow function that returns a matrix $A$ and a vector $b$, and we write $\text{Flow}_A(q) \in \mathbb{R}^{n \times n}$ and $\text{Flow}_b(q) \in \mathbb{R}^n$ to refer to each component, 5) $\text{Inv} : Q \rightarrow \mathcal{cpoly}(\mathbb{R}^n)$ is the invariant function, and 6) $\text{Grd} : E \rightarrow \mathcal{cpoly}(\mathbb{R}^n)$ is the guard function.

A path $\pi$ in $\mathcal{H}$ of length $k$ is a sequence of locations $q_1, \ldots, q_k$ in $Q$ such that $(q_i, q_{i+1}) \in E$ for each $1 \leq i < k$. We write $\pi(q_1) \subseteq \mathcal{H}$ for the set of paths in $\mathcal{H}$. Given a path $\pi = q_1, \ldots, q_k$, we define $\text{len}(\pi) = k$ as the length of $\pi$, and $\text{last}(\pi) = q_k$ as the last location.

Next we define an execution of an ADHA, describing the evolution of the continuous state subject to time passing and discrete switches.

Definition 2.2. An execution $\sigma$ of an ADHA $\mathcal{H}$ is a pwa trajectory $\sigma : I \rightarrow \mathbb{R}^n$ such that there is a path $\pi = q_1, \ldots, q_k$ in $\mathcal{H}$ and a sequence of time points $t_0, t_1, \ldots, t_k$ satisfying 1) $I = [t_0, t_k]$, 2) $\sigma(t) \in \text{Inv}(q_i)$ for every $1 \leq i \leq k$ and $t \in [t_{i-1}, t_i]$, 3) $\sigma(t_i) \in \text{Grd}(q_i, q_{i+1})$ for every $1 \leq i < k$, and 4) $\sigma(t) = \text{Flow}_A(q_i) \cdot \sigma(t) + \text{Flow}_b(q_i)$ for every $1 \leq i \leq k$ and $t \in [t_{i-1}, t_i]$.

Thus switches between dynamics are state-dependent. We call $t_k(\sigma) = \{ t_0, \ldots, t_k \}$ the switching times of $\sigma$. We say that $\sigma$ follows $\pi$, written $\sigma \sim \pi$, and denote the set of executions by $\text{exec}(\mathcal{H})$.

3 PROBLEM STATEMENT

Our overall goal is to synthesize a hybrid automaton from data, given in the form of time series, such that the synthesized automaton captures the dynamical behavior of the data up to a given precision. We split up this problem into two phases. In the first phase, given a time series $s$ and a value $\delta \in \mathbb{R}_{\geq 0}$, we find a pwa trajectory $f$ that is $\delta$-close to all points in $s$.

Definition 3.1. Given a time series $s$ with domain $D \subseteq [0, T]$, a pwa trajectory $f$ with $\text{dom}(f) = [0, T]$, and a value $\delta \in \mathbb{R}_{\geq 0}$, we say that $\delta$-captures $s$ if $\| (s(t) - f(t)) \| \leq \delta$ for each $t \in D$.

In the second phase, given another value $\varepsilon \in \mathbb{R}_{\geq 0}$, we construct a hybrid automaton from this pwa trajectory.

Definition 3.2. Given a pwa trajectory $f$ and a value $\varepsilon \in \mathbb{R}_{\geq 0}$, we say that an ADHA $\mathcal{H}$ $\varepsilon$-captures $f$ if there exists an execution $\sigma \in \text{exec}(\mathcal{H})$ such that $\| f(t) - \sigma(t) \| \leq \varepsilon$.

The definition extends to a set $F$ of piecewise-affine trajectories, i.e., $\mathcal{H}$ $\varepsilon$-captures $F$ if $\mathcal{H}$ $\varepsilon$-captures each $f$ in $F$. A possible problem to consider is: Given a set of pwa trajectories $F$ and $\varepsilon \in \mathbb{R}_{\geq 0}$, construct an adha $\mathcal{H}$ such that $\mathcal{H}$ $\varepsilon$-captures $F$. The construction of a universal
automaton, describing every possible behavior, trivially satisfies the constraint but is not a useful model. Our goal is to construct a model with a reasonable amount of behavior by introducing a minimality criterion that we formally discuss later.

**Problem 1 (Synthesis).** Given a set of PWA trajectories \( F \) and \( \epsilon \in \mathbb{R}_{\geq 0} \), construct an ADHA \( \mathcal{H} \) such that \( \mathcal{H} \) \( \epsilon \)-captures \( F \) and satisfies a minimality criterion.

We propose an approach that processes one trajectory \( f \) in \( F \) at a time and proceeds in two stages. Given a hybrid automaton \( \mathcal{H} \) and a PWA trajectory \( f \), in the first stage we check whether \( \mathcal{H} \) \( \epsilon \)-captures \( f \), which we call a membership query. In the second stage, if \( f \) is not \( \epsilon \)-captured, we modify \( \mathcal{H} \) such that it \( \epsilon \)-captures \( f \). This modification may consist of several changes to the model: increasing the invariants and guards, adding new transitions, and adding new locations. We prioritize the modifications in the order given above to minimize the number of locations.

In the next three sections we present algorithmic approaches to transforming time series to PWA trajectories, solving the membership query, and performing the model update.

### 4 FROM TIME SERIES TO PWA TRAJECTORY

In the first phase of our algorithmic framework we construct a PWA trajectory from a time series \( s \). Recall that \( f \) is supposed to be the solution of a piecewise-affine dynamical system, i.e., of a sequence of contiguous solutions of systems of the form \( \dot{x}(t) = Ax(t) + b \) with \( x(0) = x_0 \). We simplify the problem of finding \( f \) by only considering switching times of \( f \) from the domain of \( s \).

We thus need to solve the following simpler problem. Given a time series \( s \) with domain \( D \) and a value \( \delta \in \mathbb{R}_{\geq 0} \), find an affine dynamical system \( \dot{x}(t) = Ax(t) + b \) and an initial state \( x(0) = x_0 \) such that the solution \( g \) satisfies \( ||s(t) - g(t)|| \leq \delta \) for every \( t \in D \), or determine that no such system exists. We pose the problem of finding \( g \) as a parameter identification problem where the parameters are the coefficients of \( A \), \( b \), and \( x_0 \). This can be written as a query to an optimization tool in combination with an ODE solver (we refer to Section 7 for implementation details). Given concrete parameter values, i.e., instances of \( A \), \( b \), and \( x_0 \), the ODE solver can compute the solution \( g \) corresponding to the affine dynamical system. We can hence evaluate the norm \( ||s(t) - g(t)|| \) at all time points \( t \in D \). The optimization tool thus has to find a solution \( g \) such that this norm at those time points is less than \( \delta \).

We can use the above algorithm for solving the original problem of finding a PWA trajectory. The main idea is to maximize the duration in which we can use the same dynamics. Denote the time points of \( s \) by \( t_0 < \cdots < t_k \). We first find the maximum time point \( t_i \) such that the above-described algorithm finds a solution (e.g., using binary search). Then we iteratively solve the same problem for the time-series suffix from \( t_i \) to \( t_k \), until finally \( t_k = t_k^{*} \). Note that we only need to identify \( x_0 \) for the first piece, as for subsequent pieces the initial state is determined by \( x_0 \) and the previous dynamics.

### 5 MEMBERSHIP QUERY

In this section we formalize and solve the membership query. Given an ADHA \( \mathcal{H} \), a PWA trajectory \( f \), and a value \( \epsilon \in \mathbb{R}_{\geq 0} \), the fundamental problem we need to solve is to determine if \( \mathcal{H} \) \( \epsilon \)-captures \( f \). We reduce this problem to checking whether for a given PWA trajectory \( f \) and a given path \( \pi \) in \( \mathcal{H} \) there exists an execution \( \sigma \) following \( \pi \) such that \( d(f, \sigma) \leq \epsilon \). We apply this check to every path \( \pi \) in \( \mathcal{H} \) of length equal to the number of pieces in \( f \). We provide a solution by restricting \( f \) and \( \sigma \) to switch synchronously, which allows us to evaluate the pieces consecutively.

**Definition 5.1.** An execution \( \sigma \) of an ADHA \( \mathcal{H} \) is synchronized with a PWA trajectory \( f \), denoted by \( \sigma[f] \), if \( \text{dom}(\sigma) = \text{dom}(f) \) and \( t_\delta(\sigma) = t_\delta(f) \).

**Problem 2 (Membership).** Given a path \( \pi \) in an ADHA \( \mathcal{H} \), a PWA trajectory \( f \), and \( \epsilon \in \mathbb{R}_{\geq 0} \), determine if there exists a synchronized execution \( \sigma \) of \( \mathcal{H} \) with \( \sigma \sim \pi \) and \( d(f, \sigma) \leq \epsilon \).

Our membership algorithm uses reachability analysis to approximate the states that the synchronized executions of \( \mathcal{H} \) can reach.

**Definition 5.2.** Given an \( n \)-dimensional PWA trajectory \( f \) and \( \epsilon \in \mathbb{R}_{\geq 0} \), an \( \epsilon \)-tube of \( f \) is the function \( T(f, \epsilon) : \mathbb{R}_{\geq 0} \rightarrow \mathcal{P}(\mathbb{R}^n) \) such that \( T(f, \epsilon)(t) = \overline{B}_{\epsilon}(f(t)) \).

**Definition 5.3.** Given an ADHA \( \mathcal{H} \), a path \( \pi \) in paths(\( \mathcal{H} \)), a PWA trajectory \( f \), and \( \epsilon \in \mathbb{R}_{\geq 0} \), the synchronized reachable set, starting from a set \( P \subseteq \mathbb{R}^n \) and following \( \pi \), is defined as

\[
\text{SReach}(P, \pi, f, \epsilon) = \{ x \in \mathbb{R}^n : \exists \sigma \in \text{exec}(\mathcal{H}), \sigma \sim \pi, \sigma[f] \in P, \sigma(t) \in T(f, \epsilon)(t) \}.
\]

**Proposition 5.4.** Let \( \mathcal{H} \) be an ADHA, \( f \) be a PWA trajectory, \( P \subseteq \mathbb{R}^n \), and \( \epsilon \in \mathbb{R}_{\geq 0} \). If \( \text{SReach}(P, \pi, f, \epsilon) \) is nonempty for some \( \pi \in \text{paths}(\mathcal{H}) \), then \( \mathcal{H} \) \( \epsilon \)-captures \( f \).

We inductively construct the synchronized reachable set for a PWA trajectory \( f \) by computing the synchronized reachable set for each affine piece of \( f \). Concretely, given an initial set \( P \), a path \( \pi = q_1, \ldots, q_k \) in \( \mathcal{H} \), and a PWA trajectory \( f \) with \( t_\delta(f) = t_0, \ldots, t_k \), we define the synchronized reachable sets

\[
P_0 := P, \quad P_i := \text{SReach}(P_{i-1}, q_i, f, t_{i-1}, t_i, \epsilon) \text{ for } 1 \leq i \leq k.
\]

Observe that \( P_k \) is equal to \( \text{SReach}(P, \pi, f, \epsilon) \).

### 5.1 Membership query for single trajectories

We now present a method to approximate the synchronized reachable set for a PWA trajectory \( f \) with just one piece, starting from a polytope \( P \) and following a path \( \pi \) of length one in \( \mathcal{H} \), that is, \( \text{SReach}(P, \pi, f, \epsilon) \). This is a special case of Problem 2 where \( f \) is an affine trajectory and the path \( \pi \) in \( \mathcal{H} \) is a single location \( q \). As observed before, checking emptiness of the synchronized reachable set is equivalent to checking whether there exists an affine trajectory \( \sigma \) in the \( \epsilon \)-tube of \( f \), starting from the given polytope \( P \), with the same time domain as \( f \), and following the dynamics of \( q \).

**Remark 1.** Without loss of generality we restrict ourselves to linear dynamics, which are equivalent to affine dynamics under an appropriate transformation: Add an extra variable \( y \) to an affine system \( \dot{x} = Ax + b \) as \( \dot{x} = Ax + By \) where \( y \) is constant 1 (i.e., \( y = 0 \)). Hence we also consider hybrid automata with linear dynamics (LADH), which means that the flow function has the signature \( \mathcal{F} : Q \rightarrow \mathbb{R}^{n \times n} \).
(a) An affine trajectory \( f \) (black) and the \( \epsilon \)-tube around \( f \) (gray). The light red tube consists of all possible executions \( \sigma \) following some other affine dynamics emerging from \( T(f, \epsilon)(0) \). The execution \( \sigma \) (yellow) always stays inside the gray tube.

(b) Tube partition at a point in time. Green: under-approximation of states whose executions stay inside. Red: over-approximation of states whose executions eventually leave. Yellow: undecided.

Figure 1: Illustration of reachability computations.

Figure 1(a) illustrates that computing the exact synchronized reachable set is not trivial. Hence we settle for an approximate solution by successive polytope refinements into three regions, corresponding to the respective executions emerging from those regions, as illustrated in Figure 1(b): an under-approximation of the states in \( P \) whose executions definitely stay inside the tube, an over-approximation of the states in \( P \) whose executions definitely leave the tube, and the remaining states that are undetermined. In summary, we want to achieve the following goals:

(G1) determine whether \( \text{SReach}(P, q, f, \epsilon) \) is empty,
(G2) (approximately) compute \( \text{SReach}(P, q, f, \epsilon) \), and
(G3) refine the polytope \( P \) to improve the approximation.

We next discuss in detail how to achieve these goals.

5.2 Emptiness of SReach

We now work toward an algorithm for achieving goal (G1). As argued before, solving the emptiness problem exactly is not trivial. A sufficient condition is to compute an over-approximation and show emptiness for that set. \( \text{SReach}(P, q, f, \epsilon) \) is empty if and only if there exists a time point \( t \in [0, T] \) such that \( \text{SReach}(P, q, f|_{[0,t]}, \epsilon) \) is empty. We can generalize this observation to sets of points \( P' \subseteq P \). Observe that \( \text{SReach}(P, q, f, \epsilon) = \text{SReach}(P', q, f, \epsilon) \cup \text{SReach}(P \setminus P', q, f, \epsilon) \), so if for each point \( x \in P \) there exists a time point \( t \) such that the execution emerging from \( x \) leaves the tube, we can remove the set \( P' \) from \( P \). We recall a classic result.

**Definition 5.5.** The reachable region from \( P \subseteq \mathbb{R}^n \) following the linear dynamics described by \( A \in \mathbb{R}^{n \times n} \) at time \( t \) is defined as \( \text{Reach}(P, A, t) := \{ e^{At} \cdot x : x \in P \} \).

With \( A = \text{Flow}(q) \in \mathbb{R}^{n \times n} \), we know that \( \text{Reach}(P, A, t) \) includes the points of all executions \( \sigma \) at time \( t \) such that \( \sigma \sim q \) starting from \( x_0 \in P \). Moreover, \( \sigma(t) \) belongs to the \( \epsilon \)-tube around \( f \) at time \( t \). Therefore, \( \text{Reach}(P, A, t) \cap T(f, \epsilon)(t) \) is an over-approximation of \( \text{SReach}(P, q, f|_{[0,t]}, \epsilon) \), providing a sufficient emptiness check.

Algorithm 1 Over-approximation of \( \text{SReach} \)

**Require:** Matrices \( A, B \in \mathbb{R}^{n \times n} \), a point \( x_0 \in \mathbb{R}^n \), a polytope \( P_0 \), values \( \epsilon, T \in \mathbb{R}_{\geq 0} \), and a number \( m \in \mathbb{N}^+ \).
1. \( \delta := T / m \) [uniform time step for sampling]
2. for \( j \in [1, \ldots, m] \) do
3. \( x := \text{ExpMatrix}(B j \delta) \cdot x_0 \)
4. \( B_j := \text{Ball}(x, \epsilon) \) [tube at time point \( j \delta \)]
5. \( P_j := \text{SReach}(P_{j-1}, A, \delta) \cap B_j \)
6. if \( \text{isEmpty}(P_j) \) then
7. \( \text{return } P_j \)
8. \( \text{return } P_m \)

**Proposition 5.6.** Emptiness of \( \text{SReach}(P, A, t) \cap T(f, \epsilon)(t) \) implies emptiness of \( \text{SReach}(P, q, f, \epsilon) \), which implies emptiness of \( \text{SReach}(P, q, f, \epsilon) \).

Proposition 5.6 suggests an algorithm for showing emptiness of \( \text{SReach}(P, A, t) \cap T(f, \epsilon)(t) \) at sampled time points \( t \in [0, T] \). Observe that a finer time sampling provides a more accurate approximation, and a better chance to show emptiness if \( \text{SReach}(P, q, f, \epsilon) = \emptyset \). For a uniform sequence of \( m \) time points of delay \( \delta \), Algorithm 1 performs the above sufficient check numerically. Recall that \( A \) is the flow of location \( q \), \( \text{Flow}(q) \), and the linear trajectory \( f \) is given as the tuple \( ([0, T]), \{B\} \). Algorithm 1 takes as input two matrices \( A, B \in \mathbb{R}^{n \times n} \), a point \( x_0 \in \mathbb{R}^n \), a polytope \( P_0 \subseteq \mathbb{R}^n \), two values \( \epsilon, T \in \mathbb{R}_{\geq 0} \), and a natural number \( m > 0 \). For the \( j \)-th time step \( \delta = T / m \), the algorithm computes \( e^{B j \delta} \cdot x_0 \), where \( e^{B j \delta} \) is obtained with the function \( \text{ExpMatrix}(B j \delta) \). Then \( \text{Ball}(x, \epsilon) \) constructs the ball \( B_j(x) \), and \( \text{SReach}(P_{j-1}, A, \delta) \) computes the set \( \text{SReach}(P_{j-1}, A, \delta) \), which is intersected with the ball for constructing \( P_j \).

**Proposition 5.7 (Soundness).** Algorithm 1 returns an empty set only if \( \text{SReach}(P, q, f, \epsilon) \) is empty.

**Proof.** Assume that the algorithm returns an empty set but \( \text{SReach}(P, q, f, \epsilon) \) is nonempty. Then there is a point \( x \in P \) with \( e^{A t} \cdot x \in T(f, \epsilon)(t) \) for every \( t = j \delta, 1 \leq j \leq m \). Hence \( e^{A j \delta} \cdot x \in P_j \), which contradicts the condition in line 6.

**Proposition 5.8 (Robust completeness).** Let \( P \) be a polytope in \( \mathbb{R}^n \), \( A \in \mathbb{R}^{n \times n} \), \( f \) a linear trajectory with \( \text{dom}(f) = [0, T] \), and \( \epsilon_0 > 0 \) such that for every \( x \in P \) there exists \( t \in [0, T] \) with \( d(\text{Reach}([x], A, t), T(f, \epsilon)(t)) > \epsilon_0 \). Then there exists a finite number \( m \) such that Algorithm 1 returns an empty set.

**Proof.** Fix \( x \in P \) and \( t \in [0, T] \) such that \( d(\text{Reach}([x], A, t), T(f, \epsilon)(t)) > \epsilon_0 \). Then, by continuity of the distance function, there exists a time \( t^{\text{exit}} \in [0, t] \) such that \( d(\text{Reach}([x], A, t^{\text{exit}}), T(f, \epsilon)(t^{\text{exit}})) = 0 \) and for every \( t' \in [t^{\text{exit}}, t] \), \( d(\text{Reach}([x], A, t'), T(f, \epsilon)(t')) > 0 \). Let us denote \( t - t^{\text{exit}} = \delta_x \). Compute the infimum of \( \delta_x \) for every \( x \in P \), denoted as \( \delta^* \). Then, choose \( m > T / \delta^* \).

**Remark 2.** The assumption on Proposition 5.8 about \( \epsilon_0 \) is necessary in general because \( P \) and the \( \epsilon \)-tube image are compact. Since \( P \) and \( P' := \{ x \in P : \text{Reach}([x], A, t) \subseteq T(f, \epsilon)(t) \forall t \} \) are topologically closed, \( P' \setminus P \) is not topologically closed.
Algorithm 2 Synchronization check for linear trajectories

Require: Two matrices $A, B \in \mathbb{R}^{n \times n}$, states $x_0, y_0 \in \mathbb{R}^n$, and values $\epsilon, T \in \mathbb{R}_{\geq 0}$.

1. $\sigma(t) := \text{ExpMatrix}(A) \cdot y_0$
2. $f(t) := \text{ExpMatrix}(B) \cdot x_0$
3. $h(t) := f(t) - \sigma(t)$
4. if $\|x_0 - y_0\| \leq \epsilon$ and $\|f(t) - \sigma(t)\| \leq \epsilon$ then
5. $v := 0$
6. for $1 \leq i \leq n$ do
7. $v_{\text{max}} := \max(\text{Abs}(\text{Proj}(h(t), i)), [0, T])$
8. $v := \text{max}(v_{\text{max}}, 0)$
9. if $0 \leq v$ then
10. return True
11. return False

Algorithm 1 is a sufficient check: the result is only if SReach($f, q, f, \epsilon$) is empty. Next we consider membership of $\sigma$ in the $\epsilon$-tube of $f$ where $\sigma$ starts from a fixed point $x$ in $P$.

5.3 Approximation of SReach

We can achieve goal (G2) (and hence goal (G1)) for a singleton set $P = \{x\}$. In other words, for a fixed starting point $\sigma(0) = x$ we can decide if $\sigma(t) \in \mathbb{T}(f, \epsilon)(t)$ for every $t \in [0, T]$. We consider the case where $x \in \mathbb{T}(f, \epsilon)(0)$. We can easily determine if $\sigma(t) \in \mathbb{T}(f, \epsilon)(T)$ (e.g., by executing Algorithm 1 with $m = 1$). In the nontrivial case that $\sigma(T) \in \mathbb{T}(f, \epsilon)(T)$, the goal is to compute the maximum of $d(\sigma(t), \mathbb{T}(f, \epsilon)(t))$ over time interval $[0, T]$. Our approach to that problem involves solving 2n optimization problems.

Proposition 5.9 (Theorem 4 in [14]). Let $x \in \mathbb{R}^n$ be a point and $A \in \mathbb{R}^{n \times n}$ a matrix with rational coefficients. Then $\text{Reach}(\{x\}, A, t)$ is computable for every time $t$.

We summarize the procedure in Algorithm 2. The inputs are two matrices $A, B \in \mathbb{R}^{n \times n}$, states $x_0, y_0 \in \mathbb{R}^n$, and values $\epsilon, T \in \mathbb{R}_{\geq 0}$. Recall that $A$ is the flow of location $q$, Flow($q$), and the linear trajectory $f$ is given as the tuple $([0, T]), (B)$. Initially, the algorithm defines the linear trajectories $\sigma(t)$ and $f(t)$ and their difference $h(t)$. If the norm of this difference is less than $\epsilon$ for $t = 0$ and $t = T$, the algorithm computes the maximum (Max function) over $[0, T]$ of the absolute values (Abs function) for each coordinate $i$ of $h(t)$, that is, $\text{Proj}(h(t), i)$. The algorithm returns True if the maximum distance between $\sigma$ and $f$ is less than $\epsilon$, and False otherwise. Thus the algorithm determines emptiness of SReach($x$, $q, f, \epsilon$) for $x \in P$.

Proposition 5.10. Algorithm 2 returns False if and only if SReach($x$, $q, f, \epsilon$) is empty.

We assume a numerically sound optimization tool in practice. Algorithm 2 gives us a way to obtain an under-approximation of SReach($P, q, f, \epsilon$): apply Algorithm 2 to every vertex of $P$ and construct the convex hull of the vertices for which Algorithm 2 returns True. Next we prove that this set is contained in SReach($P, q, f, \epsilon$).

Proposition 5.11. Let $P$ be a convex polytope, $A = \text{Flow}(q)$, $f$ be a linear trajectory with domain $[0, T]$, and $\epsilon$ be a value in $\mathbb{R}_{\geq 0}$. Then, SReach($P, q, f, \epsilon$) = Reach($P, A, T)$ if SReach($v$, $q, f, \epsilon$) is not empty for every $v \in \text{vert}(P)$.

Proof. The inclusion SReach($P, q, f, \epsilon$) \subseteq Reach($P, A, T$) is obvious. Let SReach($v$, $q, f, \epsilon$) \neq \emptyset for every $v \in \text{vert}(P)$. We want to show that for every point $x \in P$, Reach($x$, $A, t$) belongs to $\mathbb{T}(f, \epsilon)(t)$ for all $t \in [0, T]$. Assume there exist $x \in P$ and $t \in [0, T]$ with Reach($x$, $A, t$) \subseteq $\mathbb{T}(f, \epsilon)(t)$, so SReach($x$, $q, f, \epsilon$) = \emptyset. We know that Reach($x$, $A, t$) \subseteq Reach($P, A, t$). Since Reach($P, A, t$) \subseteq $\mathbb{T}(f, \epsilon)(t)$. Moreover, $\mathbb{T}(f, \epsilon)(t)$ is convex for each $t \in [0, T]$. For any polytope $P$ and convex set $C$ it holds that $P \subseteq C$ if and only if $\text{vert}(P) \subseteq C$. Therefore, Reach($P, A, t$) \subseteq $\mathbb{T}(f, \epsilon)(t)$ if and only if Reach($\text{vert}(P), A, t$) \subseteq $\mathbb{T}(f, \epsilon)(t)$, i.e., Reach($v$, $A, t$) \subseteq $\mathbb{T}(f, \epsilon)(t)$ for each $v \in \text{vert}(P)$ and $t \in [0, T]$. By assumption, SReach($v$, $q, f, \epsilon$) \neq \emptyset for each $v \in \text{vert}(P)$. Using Proposition 5.10, Reach($v$, $A, t$) \subseteq $\mathbb{T}(f, \epsilon)(t)$ for each $v \in \text{vert}(P)$. Hence Reach($P, A, t$) \subseteq $\mathbb{T}(f, \epsilon)(t)$ for each $x \in P$: a contradiction.

Corollary 5.12. If Algorithm 2 returns True for all vertices $v \in \text{vert}(P)$, then SReach($P, q, f, \epsilon$) = Reach($P, A, T$).

5.4 Polytope refinement

Recall that $P$ is a polytope, $f$ is a linear trajectory $f(t) = e^{Bt} \cdot x_0$ with time domain $[0, T]$, $q$ is a location in some LDHA $\mathcal{H}$ with Flow($q$) = $A$, $\epsilon$ is a value in $\mathbb{R}_{\geq 0}$, and $m > 0$ is a natural number. We can use Algorithm 1 from Section 5.2 to obtain an over-approximation $P$ of the synchronized reachable set if $P$ is nonempty, we can use Algorithm 2 from Section 5.3 for every vertex of $P$, and if the algorithm returns True for some vertex, we have a nonempty under-approximation and can conclude membership of $f$ in $\mathcal{H}$. If Algorithm 2 returns False for all vertices, we cannot conclude.

Next we propose a new procedure, which together with Proposition 5.11 suggests an algorithm for computing a more precise under-approximation of SReach($P, f, \epsilon$). Intuitively, recalling Figure 1(b), this refinement narrows the discrepancy between the over-approximation (yellow) and the under-approximation (green).

First we observe that the over-approximation $P$ is a convex polytope. The idea is to contract this polytope to a new polytope. Given a value $\delta \in \mathbb{R}_{\geq 0}$, we define the $\delta$-contraction of $P$ as follows.

Definition 5.13. Let $P$ be a polytope, $\epsilon$ be a value in $\mathbb{R}_{\geq 0}$, and $\text{const}(f) = \{a_1 \cdot x \sim b_1, \ldots, a_m \cdot x \sim b_m\}$. The $\delta$-contraction of $P$ is the polytope $P_\delta := \{x \in \mathbb{R}^n : a_1 \cdot x \sim c_1, \ldots, a_m \cdot x \sim c_m\} \delta$ where $c_j = b_j \text{ if } ~\sim = \text{'}$ and $c_j = b_j - \frac{\delta}{\|a_j\|_2}$ if $\sim = \text{'$}', \text{ for each } 1 \leq j \leq m$.

We can hence take the over-approximation $P$, compute the $\delta$-contraction $P'$, and then apply Algorithm 2 to all vertices of $P$ and $P'$. Ultimately we may have to repeat this contraction several times (at most $\lceil d / \delta \rceil$ times, where $d$ is the diameter of $P$). In the end, since we know that the true synchronized reachable set is convex, we can take the convex hull of all those vertices for which Algorithm 2 returned True (i.e., these vertices belong to the synchronized reachable set). We summarize the refinement in Algorithm 3, where the procedure CONTRACT applies a $\delta$-contraction.

In principle, now that we have two polytopes $P$ and $P'$ over-approximating and under-approximating the synchronized reachable set, respectively, a natural additional refinement procedure
Algorithm 3 Polytope refinement

Require: A polytope \( P \), two matrices \( A, B \in \mathbb{R}^{n \times n} \), a state \( x_0 \in \mathbb{R}^n \), and three values \( \epsilon, T, \delta \in \mathbb{R}_{\geq 0} \).

1: \( V^+ := \emptyset \)
2: while True do
3: \( V^- := \emptyset \)
4: for \( v \in \text{ver}(P) \) do
5: if Algorithm 2\((A, B, x_0, v, \epsilon, T)\) then
6: \( V^+ := V^+ \cup \{v\} \)
7: else
8: \( V^- := V^- \cup \{v\} \)
9: if \( V^- = \emptyset \) then
10: return chull\((V^+)\)
11: end
12: end
13: end
14: end

Algorithm 4 Membership query for a single piece

Require: A polytope \( P_0 \), two matrices \( A, B \in \mathbb{R}^{n \times n} \), a state \( x_0 \in \mathbb{R}^n \), three values \( \epsilon, T, m \in \mathbb{R}_{\geq 0} \), and a value \( m \in \mathbb{N}^+ \).

1: \( F_1 := \text{Algorithm 1}\((A, B, x_0, P_0, \epsilon, T, m)\) \)
2: if isempt\((F_1)\) then
3: return \( \emptyset \)
4: end
5: \( x_1 := \text{ExpMatrix}\((AT)\) \cdot x_0 \)
6: \( F_1 := \text{Algorithm 3}\((F_1, -A, -B, x_1, \epsilon, T, \delta)\) \)
7: return \( P_1 := \text{chull}\(F_1\) \)

can be conceived where one iteratively tries to enlarge the under-approximation or shrink the over-approximation. We did not investigate this direction because the above scheme is already very precise in practice. (In fact, we rather observed that the approximations become too precise; see the further discussion in Section 7.1.)

5.5 Summary

Algorithm 4 summarizes the overall procedure for computing both an under-approximation and an over-approximation of the synchronized reachable set for a linear trajectory \( f \). We first use Algorithm 1 to compute the over-approximation \( F_1 \). If the over-approximation is empty, we can conclude that \( f \) is not \( \epsilon \)-captured. Otherwise, taking the end state \( x_1 \) of \( f \) and inverting the dynamics \( \left( f_{\text{inv}}(x) = -f(x) \right) \), we use Algorithm 3 to compute the under-approximation \( P_1 \).

We illustrate the algorithm and the generalization to multiple pieces with the following parametric linear trajectories:

\[
\begin{align*}
\dot{x} &= \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} x, & x(0) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad (2) \\
\dot{y} &= \begin{pmatrix} 0 & 1 - \alpha \\ -1 & 0 \end{pmatrix} y, & y(0) = y_0 \quad (3)
\end{align*}
\]

System (2) is fixed and takes the role of the linear trajectory \( f \) while system (3) models the location of an LDHA. In the following, we fix the parameter value \( \alpha \) and ask whether there exists an initial state \( y_0 \) such that the corresponding execution of system (3) is synchronized with \( f \). In Figure 2 we plot the executions for \( \alpha = 0.01 \) and the same initial state \( y_0 = x(0) \). It can be seen that for a time horizon of 4\( \pi \) we need to choose \( \epsilon \) larger than \( \sim 0.08 \).

In Figure 3 we plot the results for \( \epsilon = 0.1 \) and \( \alpha = 0 \) or \( \alpha = 0.01 \), respectively. In Figure 3(a) we see the under-approximation computed by the algorithm in light green. The dark green set is a simplified under-approximation that we use to handle complexity, further described in Section 7.1. Similarly, the dark yellow set is the over-approximation computed by the algorithm, while the light yellow set is a simplified over-approximation. It can be seen that the gap between the over-approximation (dark yellow) and the under-approximation (light green) is very narrow, indicating that the refinement procedure (Algorithm 3) is precise. Also note that the true synchronized reachable set in this case is a Euclidean ball, while the executions all follow the same dynamics as \( f \), those executions starting from a state outside this ball rotate around \( f \) and eventually leave the tube (since the tube does not rotate).

In Figure 3(c) we plot the intermediate sets for the same executions but modeled as PWAs trajectories with 23 pieces, starting with the set at time 0. In theory, the settings with a single piece and 23 pieces of the same dynamics are equivalent; however, due to the simplifications of the approximations for each piece, the approximations lose precision in the latter case. Still, the approximations in the last piece are sufficiently precise to prove that the under-approximation (green set) is nonempty and hence we can conclude with a positive answer to the membership query. In the last subplot we depict a random sampling from the over-approximation where we apply Algorithm 2 to check whether the state indeed corresponds to a synchronized execution (green dot) or not (red dot). Figure 3(b) shows the setting for \( \alpha = 0.01 \) with similar results.

6 MODEL UPDATE

In this section we describe a procedure to solve Problem 1 and propose a minimality criterion. The procedure tackles the problem by evaluating the given PWAs trajectories in an online fashion.

6.1 Lexicographic order to rank model updates

For a given ADHA \( \mathcal{H} \), a pwa trajectory \( f \), and a value \( \epsilon \in \mathbb{R}_{\geq 0} \), our procedure searches for a path \( \pi \) in \( \mathcal{H} \) such that the membership query of \( f \) in \( \mathcal{H} \) is positive. If there is no such path in \( \mathcal{H} \), the procedure modifies \( \mathcal{H} \) such that the modified ADHA includes such a path. The path selection and the corresponding modifications of the ADHA are chosen in the following order: 1) increasing invariants
Observe that \( \text{exec}(\mathcal{H}) \subseteq \text{exec}(\text{Mod}^{P,f}(\mathcal{H})) \). Next we define a tree capturing the ADHA updates for every affine piece in \( f \).

**Definition 6.2.** Given an \( n \)-dimensional ADHA \( \mathcal{H}_0 \) and a pwa trajectory \( f \) with \( k \) pieces, an exploration tree for \( \mathcal{H}_0 \) and \( f \) is \( \mathcal{T} = (N,E) \) with \( k \) layers (not counting the root node as a layer). Each node \( v \in N \) is represented as a tuple \((\pi, \mathcal{H}, \text{mod}, s)\) where \( \pi \) is a path in an ADHA \( \mathcal{H} \), \text{mod} is a triple of integers \((n_1, n_2, n_3)\), and \( s \) is a four-valued variable called status (with meanings 0: ‘unexplored’, 1: ‘activated’, 2: ‘explored’, 3: ‘deactivated’).

Observe that exploration trees for \( \mathcal{H}_0 \) and \( f \) can only differ in the status. The set of all exploration trees for \( \mathcal{H}_0 \) and \( f \) is denoted by \( \mathcal{S}(\mathcal{H}_0, f) \), and we call all trees belonging to \( \mathcal{S}(\mathcal{H}_0, f) \) similar. We may add a subscript to the elements in the node \( v = (\pi, \mathcal{H}, \text{mod}, s) \) (i.e., write \( \pi_v \) etc.) for clarity. We define, for an initial polyhedron \( P \) and a value \( \epsilon \in \mathbb{R}_{\geq 0} \), an exploration tree \( T_0^{P,f} \in \mathcal{S}(\mathcal{H}_0, f) \) such that the root node is \((\emptyset, \mathcal{H}_0, (0, 0, 0), 0)\), where \( \emptyset \) is the empty path. Each node \((\pi, \mathcal{H}, (n_1, n_2, n_3), 0)\) in layer \( i - 1 \), for \( 0 < i \leq k \), where \( \mathcal{Q}_i = \{q_1, \ldots, q_m\} \), has \( m + 1 \) child nodes. The first \( m \) nodes are:

\[
(\pi \cdot q_1, \text{Mod}_{q_1}^{P,f}(\mathcal{H}), (n_1 + a_1, n_2 + b_1, n_3), 0), \ldots, \\
(\pi \cdot q_m, \text{Mod}_{q_m}^{P,f}(\mathcal{H}), (n_1 + a_m, n_2 + b_m, n_3), 0),
\]

where \( a_j = 0 \) if \((\mathcal{I}(\pi, q_j)) \in \mathcal{E}_Q \) and \( q_j = 1 \) otherwise, and \( b_j \) is the number of constraint modifications for invariants and guards with respect to \( \mathcal{H} \), for every \( 1 \leq j \leq m \). The last child node is:

\[
(\pi \cdot q, \text{Mod}_{q}^{P,f}(\mathcal{H}), (n_1 + 1, n_2 + 1, n_3), 0),
\]

where \( q \) is a new location with \( \text{Flow}(q) = (A_1, b_1) \).

The paths from root to leaves in an exploration tree represent all possible paths in updated ADHAs, given the initial ADHA \( \mathcal{H}_0 \), for exploring membership of \( f \). An upper bound on the number of paths is \((m + k)^3\), where \( m = |\mathcal{Q}_0| \) is the number of locations in \( \mathcal{H}_0 \) and \( k \) is the number of pieces in \( f \). The complexity for the membership query is in \( O(p(n)) \) for some polynomial \( p \) in the dimension \( n \). Hence the complexity for a membership check in each node of the exploration tree is upper-bounded by \( O((m + k)^3 k p(n)) \).

We introduce a strategy for partial exploration that minimizes automaton modifications (according to \( \text{mod} \)). Given \( \mathcal{H}_0 \) and \( f \), a decision strategy is a function \( D : \mathcal{S}(\mathcal{H}_0, f) \times N \rightarrow N \) that determines the next node to be analyzed in an exploration tree. A decision strategy is combined with a tree update in order to activate and explore nodes or discard useless nodes. We say that a node is unexplored when its status is 0. We can explore a node when it is activated (status 1). After exploration, if the membership query is positive, we set the status to 2 (explored) and otherwise to 3 (deactivated). Child nodes of deactivated nodes need not be explored further.

**Definition 6.3.** Given an ADHA \( \mathcal{H}_0 \), a pwa trajectory \( f \), a polyhedron \( P \), and a value \( \epsilon \in \mathbb{R}_{\geq 0} \), an \( \epsilon \)-tree update function, \( \text{update}_\epsilon : \mathcal{S}(\mathcal{H}_0, f) \times N \rightarrow \mathcal{S}(\mathcal{H}_0, f) \), maps a tree \( T \) and a node \( v \) in the \( i \)-th layer to a similar tree such that \( s_v = 3 \) if \( \text{SReach}(P, \pi_v, f|_{\{q_{\ell}^{(i)}\}}, \epsilon) \neq \emptyset \), and \( s_v = 2 \) and \( s_{v'} = 1 \) for every node \( v' \in \text{child}(v) \) otherwise, and leaves the status of all other nodes unchanged.

Given a set of nodes \( W \), we denote by \( \text{Act}(W) \) the set of nodes with activated status, i.e., \( \{v \in W : s_v = 1\} \). Our decision strategy
minimizing $\text{mod}$ is $D(T, v) = \arg \min_{v' \in \text{Act}(\mathcal{N})} \text{mod} v'$, assuming that arg min returns one node if several nodes minimize the $\text{mod}$ value.

Example 6.4. Figure 4 shows an example of an intermediate state of an exploration tree for a given ADHA and a PWA trajectory with two pieces. The root node has been described before. For the remaining tree nodes we represent the automata $\mathcal{H}_i$ only symbolically. The first piece of $f$ follows the dynamics $\dot{x} = -x$ for 0.5 time units. The available choices for the first automaton mode are $q_1$, $q_2$, or a new mode $q_3$; hence the root node expands to three new nodes. The node with path $q_1$ requires a modification of the invariant of $q_1$ because dwelling in that mode for 0.5 time units is not possible otherwise. However, since the final reachable states do not intersect with the $\epsilon$-tube (negative membership query), this node status is set to 3 (deactivated) and none of the child nodes are explored further. The node with the new location $q_3$ has a “location entry” in the modification tuple. The node with path $q_2$ does not require any modifications (i.e., $\mathcal{H}_2 = \mathcal{H}$) and is hence chosen as the next node for exploration. Now consider the second piece with dynamics $\dot{x} = 2x$ for 0.7 time units. Again we have the choice between the existing locations and a new location $q_3$. The exploration works like before, only that this time we need to add a transition from $q_2$ to the next location in all three cases (since $q_2$ does not have any outgoing transitions in $\mathcal{H}_2 = \mathcal{H}$ yet). The 2-leaf with the path $q_2q_1$ has the highest priority and we perform a membership query for it. In this case, the query returns a positive answer and the algorithm outputs the automaton $\mathcal{H}_4$, which looks like $\mathcal{H}$ but with an additional transition and an extended invariant in location $q_1$.

Algorithm 5 shows the procedure for a model update given an initial ADHA $\mathcal{H}$, a PWA trajectory $f$, and a value $\epsilon \in \mathbb{R}_{\geq 0}$. The function $\text{InitTree}$ constructs the exploration tree $T_0^{P, f}$ for the polyhedron $P = T(f, \epsilon)(0)$. Then the algorithm starts exploring from the root node (line 2) and subsequently explores nodes driven by the decision strategy (line 5), which chooses activated nodes with minimum $\text{mod}$ component and iteratively activates every child nodes and deactivates the current node or sets it to explored (line 6). The algorithm returns the updated model $\mathcal{H}'$. Finally, Problem 1 is solved by iteratively running Algorithm 5 over every trajectory $f \in F$ and modifying the ADHA, as shown in Algorithm 6.

Algorithm 5 Hybrid model update

Require: An ADHA $\mathcal{H}$, a PWA trajectory $f$ and $\epsilon \in \mathbb{R}_{\geq 0}$.
Ensure: An ADHA $\mathcal{H}'$ such that it $\epsilon$-captures $f$.

1: $T := \text{InitTree}(\mathcal{H}, f, \epsilon)$
2: $v := \text{RootNode}(T)$
3: $T := \text{upd}_\epsilon(T, v)$
4: while $v$ not in bottom layer with $s_v = 2$ do
5:  $v := D(T, v)$
6:  $T := \text{upd}_\epsilon(T, v)$
7:  $\mathcal{H}' := \mathcal{H}_v$
8: return $\mathcal{H}'$

Algorithm 6 Synthesis of ADHA from PWA trajectories

Require: A finite set of PWA trajectories $F$ and $\epsilon \in \mathbb{R}_{\geq 0}$.

1: $\mathcal{H} := \emptyset$ (empty automaton with no location)
2: for $f$ in $F$ do
3:  $\mathcal{H} := \text{Algorithm 5}(\mathcal{H}, f, \epsilon)$
4: return $\mathcal{H}$

Proof. Given an ADHA $\mathcal{H}$, a PWA trajectory $f$, and a value $\epsilon \in \mathbb{R}_{\geq 0}$, Algorithm 5 provides an updated ADHA $\epsilon$-capturing $f$ and minimizing the number of modifications.

Proposition 6.5. Given an ADHA $\mathcal{H}$, a PWA trajectory $f$, and a value $\epsilon \in \mathbb{R}_{\geq 0}$, Algorithm 5 provides an updated ADHA $\epsilon$-capturing $f$ and minimizing the number of modifications.

Theorem 6.6. Given an ADHA $\mathcal{H}$, a set $F$ of PWA trajectories, and a value $\epsilon \in \mathbb{R}_{\geq 0}$, Algorithm 6 solves Problem 1.
7 IMPLEMENTATION AND CASE STUDY

In this section we describe our implementation and evaluate it: in the first two examples we obtain the PWA trajectories from random executions with perturbed dynamics from a given ADHA model; in a third example we construct the PWA trajectories from time series.

7.1 Implementation

We implemented our approach in HySynth [1] where we wrote the high-level synthesis algorithm in Python and the low-level algorithms in Julia. For the ODE optimization (both in Section 4 and Algorithm 2) we use the libraries Optim.jl [21] (which uses Brent’s method [8] to find a root in a bracketing interval and guarantees convergence for functions computable within the interval) and DifferentialEquations.jl [30] as follows (assuming linear dynamics without loss of generality). Given two n-dimensional executions $x = A x, x(0) = x_0$ and $y = By, y(0) = y_0$, we construct a 3n-dimensional execution $z = C z, z(0) = z_0$ where

$$C = \begin{pmatrix} A & 0 & 0 \\ 0 & B & 0 \\ A & -B & 0 \end{pmatrix},$$

$$z_0 = \begin{pmatrix} x_0 \\ y_0 \\ x_0 - y_0 \end{pmatrix}.$$ 

We are interested in the projection of $z(t)$ onto the last n dimensions. Calling this projection $w(t)$, the norm of $w(t)$ describes the distance between $x(t)$ and $y(t)$, i.e., $\|w(t)\| = d(x(t), y(t))$. We query the solver for each dimension of $w(t)$ to find the maximum distance.

We use JuliaReach [7] for the set computations and reachability analysis. As mentioned in Section 5.4, the polytopes over- and under-approximating the synchronized reachable sets constructed during the membership query grow in complexity, especially for input trajectories with many pieces. We simplify the sets after each piece, i.e., we under-approximate an over-approximation (for which JuliaReach computes a polytope from support vectors in template directions) and over-approximate an over-approximation (for which we implemented an algorithm from [13] to compute a zonotope in template directions) with octagonal directions (i.e., axis-parallel or diagonal constraints in two dimensions).

7.2 Evaluation

We consider an ADHA that models a heater with two locations “ON” and “OFF”, as depicted in Figure 5(a) with parameter value $a = 0.1$.

Next we describe how we sampled executions from the model. The inputs to the simulation procedure are 1) an ADHA (here: the heater model), 2) a desired path length (here: 6), 3) a maximum dwell time per location (here: 7), 4) a time step (here: 0.05), and 5) a maximum perturbation (here: 0.001). We first sample an initial location $q_0$ and an initial (continuous) state $x_0$ from $Inv(q_0)$. Then we repeat the following loop. Given a location $q$ and a state $x$, we first compute a matrix $A$ by perturbing the dynamics matrix $Flow(q)$ (technically, we only perturb non-zero entries). Then we compute the discrete-time successor of $x$ with the fixed time step $t$ (via $x' := Reach(x, A, t)$), and we check which of the outgoing transitions of $q$ are enabled for this new state. We continue computing successor states and collecting enabled transitions until either the state leaves $Inv(q)$ or we exceed the maximum dwell time. Then we choose a random transition together with a random time point of those that were enabled. The above loop terminates if either there is no transition enabled or we exceed the desired path length.

We applied the above procedure to obtain 100 random executions from the heater model. Then we first learned a model from the first ten executions and then continued modifying the resulting ADHA with the remaining 90 executions, where we used a precision value $\varepsilon = 0.1$. (Note that our algorithmic framework behaves exactly the same way as if we had learned an ADHA from the 100 executions at once. The split into two stages is only for illustrative purposes.) We show the intermediate and the final result obtained with our implementation in Figure 5(b) and Figure 5(c) respectively, and random simulations in Figure 6(a).

The first observation is that the discrete structure of the resulting ADHA matches exactly the structure of the original model. The reason why the dynamics matrices of the locations is the same as in the original model is because we did not perturb the dynamics of the very first execution in order to obtain a legible flow representation. Still, even though the algorithm is confronted with slightly
We have presented an automatic synthesis algorithm for computing a hybrid automaton with affine differential dynamics $\mathcal{H}$ from a set of time series $S$ respectively from a set of piecewise-affine trajectories $F$. Given precision parameters $\delta$ and $\epsilon$, the main feature of our algorithm is that every time series $s$ in $S$ is $\delta$-captured by some trajectory $f$ in $F$ and that $\mathcal{H}$ is guaranteed to $\epsilon$-capture every function $f$ in $F$, that is, $\mathcal{H}$ contains an execution that has distance at most $\epsilon$ from $f$. Another feature of our algorithm is that it works online, meaning that the functions $f$ are processed sequentially and we only modify the intermediate automaton models.

For future work, hardness of the membership problem for the class of automata that we considered is open. We currently do not know if that problem is decidable, and if so, what complexity is required to solve it exactly. Another interesting but challenging
extension of our work is to allow for transition switches not at a single time point but in a whole time interval.

**ACKNOWLEDGMENTS**

This research was supported in part by the Austrian Science Fund (FWF) under grant Z211-N23 (Wittgenstein Award) and the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 754411.

**REFERENCES**

[1] 2021. HySynth. https://github.com/HySynth/HySynth.

[2] Rajeev Alur and Nimit Singhania. 2014. Precise Piecewise Affine Models from Input-Output Data. In EMAF (New Delhi, India). Association for Computing Machinery. Article 3.

[3] Dana Angluin. 1987. Learning Regular Sets from Queries and Counterexamples. Inf. Comput. 75, 2 (1987).

[4] Laurent Bako and René Vidal. 2008. Algebraic Identification of MIMO SÄRM Models. In HSVC, Vol. 4981. Springer.

[5] Ezio Bartocci, Jostyemee Deshmukh, Felix Gigler, Cristinel Mateis, Dejan Nickovic, and Xin Qin. 2020. Mining Shape Expressions From Positive Examples. IEEE Trans. Comput. Aided Des. Integr. Circuits Syst. 39, 11 (2020), 3809–3820.

[6] Alberto Bemporad, Andrea Garulli, Simone Paoletti, and Antonio Vicino. 2005. A bounded-error approach to piecewise affine system identification. IEEE Trans. Automat. Contr. 50, 10 (2005).

[7] Sergiy Bogomolov, Marcelo Forets, Goran Frehse, Kostiantyn Potomkin, and Christian Schilling. 2019. JuliaReach: a toolbox for set-based reachability. In HSVC. ACM.

[8] Richard P. Brent. 1971. An Algorithm with Guaranteed Convergence for Finding a Zero of a Function. Comput. J. 14, 4 (1971).

[9] Giancarlo Ferrari-Trecate and Marco Muselli. 2003. Single-Linkage Clustering for Optimal Classification in Piecewise Affine Regression. IFAC Proceedings Volumes 36, 6 (2003).

[10] Giancarlo Ferrari-Trecate, Marco Muselli, Diego Liberati, and Manfred Morari. 2001. A Clustering Technique for the Identification of Piecewise Affine systems. In HSVC. Springer.

[11] Andrea Garulli, Simone Paoletti, and Antonio Vicino. 2012. A survey on switched and piecewise affine system identification. IFAC Proceedings Volumes 45, 16 (2012).

[12] Ary L. Goldberger, Luis A. N. Amaral, Leon Glass, Jeffrey M. Hausdorff, Plamen Ch. Ivanov, Roger G. Mark, Joseph E. Mietus, George B. Moody, Chung-Kang Peng, and H. Eugene Stanley. 2000. PhysioBank, PhysioToolkit, and PhysioNet. Circulation 101, 23 (2000), e215–e220.

[13] Leonidas J. Guibas, An Thanh Nguyen, and Li Zhang. 2003. Zonotopes as bounding volumes. In SODA. ACM/SIAM.

[14] Emmanuel Hamry. 2008. Reachability in Linear Dynamical Systems. In Logic and Theory of Algorithms. Springer Berlin Heidelberg.

[15] Yaminn Hashambhoy and René Vidal. 2005. Recursive identification of switched ARX models with unknown number of models and unknown orders. In CDC.

[16] Thomas A. Henzinger. 2000. The Theory of Hybrid Automata. Springer.

[17] A. L. Judovick, S. Weiland, and W. P. M. H. Heemels. 2005. Identification of Switched ARX Models with Noise. Asian J. Control 18, 5 (2016).

[18] Kun Huang, A. Wagner, and Yi Ma. 2004. Identification of hybrid linear time-invariant systems via subspace embedding and segmentation (SES). In CDC, Vol. 3.

[19] Fabien Lauer, Gérard Bloch, and René Vidal. 2011. A continuous optimization framework for hybrid system identification. Automatica 47, 3 (2011).

[20] Ramy Medhat, S. Ramesh, Borzoo Bonakdarpour, and Sebastian Fischmeister. 2015. A framework for mining hybrid automata from input/output traces. In EMAF. IEEE.

[21] Patrick Kofod Mogensen and Asbjørn Nilsen Riseth. 2018. Optim: A mathematical optimization package for Julia. Journal of Open Source Software 3, 24 (2018).

[22] Eberhard Münz and Volker Krebs. 2005. Continuous Optimization Approaches to the Identification of Piecewise Affine Systems. IFAC Proceedings Volumes 38, 1 (2005).

[23] Hayato Nakada, Kiyoitsugu Takaba, and Tohru Katayama. 2005. Identification of piecewise affine systems based on statistical clustering technique. Automatica 41, 5 (2005).

[24] Sohail Nazari, Bahador Rashidi, Qing Zhao, and Biao Huang. 2016. An Iterative Algebraic Geometric Approach for Identification of Switched ARX Models with Noise. Asian J. Control 18, 5 (2016).

[25] Necmiye Ozay. 2016. An exact and efficient algorithm for segmentation of ARX models. In ACC. IEEE.

[26] N. Ozay, C. Lagoa, and M. Sznaier. 2009. Robust identification of switched affine systems via moments-based convex optimization. In CDC.