2024 European Control Conference (ECC)
June 25-28, 2024. Stockholm, Sweden

State-Space Piece-Wise Affine System Identification with Online Deterministic Annealing

Christos N. Mavridis*, Aris Kanellopoulos*, John S. Baras†, and Karl Henrik Johansson*

Abstract— **We propose an online identification scheme for discrete-time piece-wise affine state-space models based on a system of adaptive algorithms running in two timescales. A stochastic approximation algorithm implements an online deterministic annealing scheme at a slow timescale, estimating the partition of the augmented state-input space that defines the switching signal. At the same time, an adaptive identification algorithm, running at a higher timescale, updates the parameters of the local models based on the estimate of the switching signal. Identifiability conditions for the switched system are discussed and convergence results are given based on the theory of two-timescale stochastic approximation. In contrast to standard identification algorithms for piece-wise affine systems, the proposed approach progressively estimates the number of modes needed and is appropriate for online system identification using sequential data acquisition. This progressive nature of the algorithm improves computational efficiency and provides real-time control over the performance-complexity trade-off, desired in practical applications. Experimental results validate the efficacy of the proposed methodology.**

I. Introduction

Switched systems constitute a class of universal approximation models with important applications in identification, verification, and control synthesis of hybrid, and complex nonlinear systems [1]–[3]. Piece-Wise Affine (PWA) systems, in particular, are a special class of switched systems, defined as a collection of affine dynamical systems, often called modes, indexed by a discrete-valued switching variable that depends on a partitioning of the state-input domain into a finite number of polyhedral regions [1], [2].

Most existing identification approaches for switched systems can be categorized by the problem formulation as optimization-based [4], algebraic [5], [6], or clustering-based [7]–[9], and by the the method used as offline [7], [10] or online (recursive) [6], [11]. In particular, clustering-based methods are optimization-based methods that make use of unsupervised learning techniques to estimate the partition of the domain that is needed for the switching signal [7]–[9], [11], [12]. It is worth noticing that most such approaches are offline methods that first classify each observation and estimate the local model parameters (either simultaneously or iteratively), and then reconstruct the partition of the switching signal. The local models are reconstructed using classical realization theory results, while the partition reconstruction is typically addressed by resorting to standard linear classification algorithms, such as support vector machines or neural network classifiers. Overall, a research trend can be identified towards efficient optimization tools to tackle the identification problem of switched and PWA systems. Hence, the tradeoff between computational complexity and quality of a suboptimal solution is still a key issue driving the research endeavors in this field.

In this work, we follow a progressive clustering-based method to identify a PWA system in a state space representation, extending our prior work on PWA system identification in the input-output form [11]. The estimation of the partition defining the switching signal is based on a Voronoi tessellation of the observation (state-input) space with respect to a progressively growing set of codevectors that are computed using an online deterministic annealing learning algorithm [11], [13]–[15]. This method progressively estimates the optimal codevectors and simulates an annealing process that induces a series of bifurcation phenomena, according to which, the number of codevectors K is adjusted, thus estimating the number of modes in a PWA system.

Adopting the above adaptive partitioning framework, we propose an online identification scheme for discrete-time state-space PWA models on a system of adaptive algorithms running in two timescales. A stochastic approximation algorithm based on online deterministic annealing runs at a slow timescale estimating the partition of the space that defines the switching signal, as well as the number of modes (Section IV). At the same time, a second stochastic approximation algorithm based on standard recursive system identification methods, running at a higher timescale, updates the parameters of the local models based on the estimate of the switching signal (Section V-A). The identifibility and convergence properties of this system of recursive algorithms are studied in Section III and Section V-B, respectively. In contrast to standard identification algorithms for piece-wise affine systems, the proposed approach is appropriate for real-time system identification using sequential data acquisition, and provides computational efficiency compared to standard algebraic, mixed-integer programming, and clustering-based methods. In addition, the progressive nature of the algorithm provides real-time control over the performance-complexity trade-off, desired in practical applications. Experimental results validate the efficacy of the proposed approach.

II. Switched and Piecewise Affine Systems

A switched affine system consists of multiple affine systems indexed by a discrete-valued switching signal. In discrete-time, a state-space representation of such a system

*Division of Decision and Control Systems, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm. emails:{mavridis,arisk,kallej}@kth.se.

†Department of Electrical and Computer Engineering and the Institute for Systems Research, University of Maryland, College Park, USA. email: baras@umd.edu.

is

$$
\begin{aligned}
x_{t+1} &= A_{\sigma_t}x_t + B_{\sigma_t}u_t + f_{\sigma_t} + w_t \\
y_t &= C_{\sigma_t}x_t + D_{\sigma_t}u_t + g_{\sigma_t} + v_t, \quad t \in \mathbb{Z}_+,
\end{aligned}
\tag{1}
$$

where $x_t \in \mathbb{R}^n$ is the state vector of the system, evolving on the same space for all modes, $u_t \in \mathbb{R}^p$ is the input, $y_t \in \mathbb{R}^q$ is the output, and $w_t \in \mathbb{R}^n$ and $v_t \in \mathbb{R}^q$ are noise terms. The signal $\sigma_t \in \{1, \ldots, s\}$ represents the discrete state of the system and defines the mode (affine dynamics) which is active at time t. The matrices $A_i \in \mathbb{R}^{n\times n}$, $B_i \in \mathbb{R}^{n\times p}$, $C_i \in \mathbb{R}^{q\times n}$, $D_i \in \mathbb{R}^{q\times p}$, $f_i \in \mathbb{R}^n$, and $g_i \in \mathbb{R}^q$ define the affine dynamics for each mode $i \in \{1, \ldots, s\}$.

The discrete state σ_t can be either an exogenous input, e.g. triggered by some event, or a function of the system state and input. In particular, when σ_t is defined according to a polyhedral partition of the state and input space, i.e., when

$$
\sigma_t = i \iff \begin{bmatrix} x_t \\ u_t \end{bmatrix} \in R_i \subset R, \tag{2}
$$

where $\{R_i\}_{i=1}^s$ are convex polyhedra defining a complete partition of the state-input domain $R \subseteq \mathbb{R}^{n+p}$, the switched system is called Piece-Wise Affine (PWA).

A. Uniqueness of Realization

The problem of identifying a state-space representation of a switched affine system can be quite challenging. Traditionally, it has been handled linked to applying results from classical realization theory to each linear subsystem [16]. However, identifiability issues arise regarding the characterization of minimality of discrete-time switched linear systems, as will be discussed in Section III-A. The first issue relates to the known fact that realizations of a switched affine system are not unique [17]. The lack of uniqueness is related to the (i) the minimal realizations of the local linear systems from input-output observations are non-unique, and (ii) a realization of a switched affine system can be constructed for any arbitrary number of modes $s' \geq s$ [17]. The effect of the number of modes to the realization of system (1) will be discussed in Section V. To ensure uniqueness of the realizations, given that all subsystems $i \in \{1, \ldots, s\}$ share the same state space, we make the following assumptions:

Assumption 1: We assume that $C_i = C$, $\forall i \in \{1, \ldots, s\}$ holds for system (1).

Assumption 2: Moreover, we assume no affine dynamics f_{σ_t}, g_{σ_t}, no feed-forward terms D_{σ_t}, $C = I$ is the identity matrix, i.e., that the states are fully observable, and that the error terms w_t and v_t share the same statistics for every mode of the system.

Assumption 1 enforces that the set of observations is acquired using the same observation mechanism, which leads to the realization of (1) being unique. Assumptions 2 are made to simplify the presentation of the proposed methodology without loss of generality.

In view of Assumptions 1, 2, and by defining an augmented vector $r_t \in \mathbb{R}^{n+p}$ as

$$
r_t = \begin{bmatrix} x_t \\ u_t \end{bmatrix} \in \mathbb{R}^{n+p}, \tag{3}
$$

the PWA system of the form (1) becomes:

$$
\begin{cases}
x_{t+1} &= A_1 x_t + B_1 u_t + w_t, \quad \text{if } r_t \in R_1 \\
\quad\vdots & \qquad \vdots \\
x_{t+1} &= A_s x_t + B_s u_t + w_t, \quad \text{if } r_t \in R_s,
\end{cases}
\tag{4}
$$

where $R_i \subset R$ are polyhedra in $\mathbb{R}^{n+p}$ for all $i = 1, \ldots, s$, such that $R_i \cap R_j = \emptyset$ for $i \neq j$, and $\bigcup_i R_i = R$. In the rest of the paper, we will focus our attention to system (4).

III. Identification of Switched Systems

In addition to the realizations of the local systems being non-unique, minimality and identifiability of the switched system does not necessarily imply that of the local subsystems [18]. In particular, without additional assumptions, estimating the parameters of a linear switched system of the form (4) by first estimating the parameters of the corresponding linear subsystems independently may fail.

A. Identifiability, Minimality, and Persistence of Excitation

In this section, we describe the conditions, under which, the local linear models of (4) can be identified, even when a subset of them is not controllable (minimal) in isolation. These conditions will take the form a persistence of excitation criterion, according to Theorem 1 below.

Theorem 1: Let a bounded-input bounded-output linear discrete-time system of the form:

$$
\begin{aligned}
x_{t+1} &= Ax_t + Bu_t, \quad t \in \mathbb{Z}_+ \\
y_t &= x_t,
\end{aligned}
\tag{5}
$$

where $x_t \in \mathbb{R}^n$, $u_t \in \mathbb{R}^p$, $A \in \mathbb{R}^{n\times n}$, and $B \in \mathbb{R}^{n\times p}$. Denote $r_t = [x_t^{\mathrm{T}} u_t^{\mathrm{T}}]^{\mathrm{T}}$. Then, if there exist some $\alpha, \beta, T > 0$ such that

$$
\alpha I_{n+p} \preceq \sum_{\tau=t}^{t+T} r_\tau r_\tau^{\mathrm{T}} \preceq \beta I_{n+p}, \quad \forall t \geq 0, \tag{6}
$$

the augmented parameter matrix $\hat{\Theta}_t = [\hat{A}_t|\hat{B}_t]$ updated by the recursion

$$
\hat{\Theta}_{t+1} = \hat{\Theta}_t - \gamma\left(\hat{\Theta}_t r_t - x_{t+1}\right) r_t^{\mathrm{T}}, \quad t \geq 0, \tag{7}
$$

for some $\gamma > 0$, asymptotically converges to $\Theta = [A|B]$.

Proof: We construct the system

$$
\hat{x}_{t+1} = \hat{A}x_t + \hat{B}u_t, \quad t \in \mathbb{Z}_+, \tag{8}
$$

where $\hat{A} \in \mathbb{R}^{n\times n}$, and $\hat{B} \in \mathbb{R}^{n\times p}$. Subtracting (5) from (8), we get:

$$
e_{t+1} = \bar{\Theta} r_t, \quad t \in \mathbb{Z}_+, \tag{9}
$$

where $e_t = \hat{x}_t - x_t \in \mathbb{R}^n$ is the observation error, $r_t = [x_t^{\mathrm{T}}|u_t^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{n+p}$ is the augmented state-input vector as defined in (3), and $\bar{\Theta} = [(\hat{A} - A)|(\hat{B} - B)]$ is an augmented matrix of the system parameters of size $n \times (n + p)$. Then (7) is equivalent to:

$$
\bar{\Theta}_{t+1} = \bar{\Theta}_t - \gamma e_{t+1} r_t^{\mathrm{T}}, \quad t \geq 0. \tag{10}
$$

Notice that (10) can be written in the form of a linear time-varying dynamical system:

$$\bar{\Theta}_{t+1} = \bar{\Theta}_t (I_{n+p} - \gamma r_t r_t^{\mathrm{T}}),\ t \geq 0. \tag{11}$$

By vectorizing $\bar{\Theta}_t$ such that $\bar{\theta}_t = vec(\bar{\Theta}_t)$, (11) becomes:

$$\bar{\theta}_{t+1} = (I_{n(n+p)} - \gamma \psi_t \psi_t^{\mathrm{T}}) \bar{\theta}_t = \Xi_t \bar{\theta}_t,\ t \geq 0, \tag{12}$$

where $\otimes$ denotes the Kronecker product, and $\psi_t = [r_t^{\mathrm{T}} \otimes I_n]^{\mathrm{T}}$ is a $n(n+p) \times n$ matrix. We will show that (12) is exponentially stable in the large as long as (6) is satisfied. Let the Lyapunov function $V(t, \bar{\theta}) = \bar{\theta}_t^{\mathrm{T}} \bar{\theta}_t$. It is obvious that there exist $k_1, k_2 > 0$ such that $k_1 \|\bar{\theta}\|^2 \leq V(t, \bar{\theta}) \leq k_2 \|\bar{\theta}\|^2$. Notice that $V(t+1, \bar{\theta}_{t+1}) - V(t, \bar{\theta}_t) = \bar{\theta}_t^{\mathrm{T}} \Xi_t^{\mathrm{T}} \Xi^t \bar{\theta}_t$. As a result, by summing the differences for T timesteps, we get:

$$\begin{aligned} V(t{+}T+1, &\bar{\theta}_{t+T+1}) - V(t, \bar{\theta}_t) = \\ &= \sum_{\tau=t}^{t+T} V(\tau+1, \bar{\theta}_{\tau+1}) - V(\tau, \bar{\theta}_\tau) \\ &= \sum_{\tau=t}^{t+T} \bar{\theta}_\tau^{\mathrm{T}} \left(\Xi_\tau^{\mathrm{T}} \Xi_\tau - I_{n(n+p)} \right) \bar{\theta}_\tau \\ &= \bar{\theta}_t^{\mathrm{T}} \left[\sum_{\tau=t}^{t+T} \Phi(\tau; t)^{\mathrm{T}} \left(\Xi_\tau^{\mathrm{T}} \Xi_\tau - I_{n(n+p)} \right) \Phi(\tau; t) \right] \bar{\theta}_\tau \\ &\leq -\alpha_1 \bar{\theta}_t^{\mathrm{T}} I_{n(n+p)} \bar{\theta}_t = -\alpha_1 V(t, \bar{\theta}_t), \end{aligned} \tag{13}$$

for some $0 < \alpha_1 < 1$. Here $\Phi(\tau; t) = \Xi_t \Xi_{t+1} \dots \Xi_{\tau-1}$ is the transition matrix of (12), and the inequality follows from condition (6). Notice that the first inequality in (6) is equivalent to $\alpha I_{n+p} \preceq \sum_{\tau=t}^{t+T} r_\tau^{\mathrm{T}} r_\tau$ and directly implies that $\alpha_2 I_{n(n+p)} \preceq \sum_{\tau=t}^{t+T} \psi_\tau^{\mathrm{T}} \psi_\tau$, for some $\alpha_2 > 0$, as well. As a result $\sum_{\tau=t}^{t+T} \Xi_\tau^{\mathrm{T}} \Xi_\tau \preceq \alpha_3 T I_{n(n+p)}$ for some $0 < \alpha_3 < 1$, and, therefore, $\sum_{\tau=t}^{t+T} \left(\Xi_\tau^{\mathrm{T}} \Xi_\tau - I_{n(n+p)} \right) \preceq -\alpha_4 T I_{n(n+p)}$ for some $0 < \alpha_4 < 1$. Finally this implies that $\left[\sum_{\tau=t}^{t+T} \Phi(\tau; t)^{\mathrm{T}} \left(\Xi_\tau^{\mathrm{T}} \Xi_\tau - I_{n(n+p)} \right) \Phi(\tau; t) \right] \leq -\alpha_1 I_{n(n+p)}$ for some $0 < \alpha_1 < 1$ [19]. Notice that the second inequality of (6) is necessary to ensure non-singularity of the transition matrix $\Phi(\tau; t)$ [20]. Finally, as an immediate result of (13), $V(t+T+1, \bar{\theta}_{t+T}+1) \leq (1 - \alpha_1) V(t, \bar{\theta}_t)$, $\forall t \geq 0$, which implies uniform asymptotic stability in the large, and, due to linearity, exponential stability in the large. ∎

As a result of Theorem 1, we make the following assumption to ensure identifiability of (4):

Assumption 3: All linear subsystems $i \in \{1, \dots, s\}$ of (4) are asymptotically bounded, and a bounded control input u_t is designed such that for every mode $i \in \{1, \dots, s\}$ of (4), there exist some $\alpha_i, \beta_i, T_i > 0$ for which the following persistence of excitation condition holds:

$$\alpha_i I_{n+p} \preceq \sum_{\tau=t}^{t+T_i} \begin{bmatrix} x_\tau x_\tau^{\mathrm{T}} & x_\tau u_\tau^{\mathrm{T}} \\ u_\tau x_\tau^{\mathrm{T}} & u_\tau u_\tau^{\mathrm{T}} \end{bmatrix} \preceq \beta_i I_{n+p},\ \forall t \geq 0. \tag{14}$$

Remark 1: The condition (14) implies that not every subsystem in (4) should be controllable (minimal), as long as the boundaries of each mode (region R_i in the state-input system) are visited often enough and from a rich-enough set of different states.

Remark 2: The assumption of assymptotic boundedness and controllability (thus, minimality) for all subsystems of (4) would simplify the condition (14) to a persistence of excitation criterion for the input u_t for each subsystem separately. Although this assumption is usually adopted, it is a limiting assumption in a practical sense. The assumption that all the local systems share the same state space of order n is a modeling assumption that facilitates the identification of the switched signal as a partition of the state-input space. However, it allows for situations when the minimal realization of some of the local models is of order $n' < n$, as long as the switched system as a whole is identifiable.

B. Identification as an Optimization Problem

Consider the state-input observations of system (4) written in the form

$$\begin{aligned} y_t &= \Theta_i r_t + w_t, \\ &= [r_t^{\mathrm{T}} \otimes I_n] \theta_i + w_t,\ \text{if } r_t \in R_i, t \geq 0, \end{aligned} \tag{15}$$

for all $i = 1 \dots, s$, where $y_t := x_{t+1}$, $\Theta_i = [A_i | B_i]$, $\theta_i = vec(\Theta_i) \in \mathbb{R}^{n(n+p)}$, r_t is as defined in (3), and $\otimes$ denotes the Kronecker product. Under the identifiability conditions discussed in Section III-A (Assumption 3), the general identification problem for a PWA system in the state-space as given in (4) can be formulated as a stochastic optimization problem over the parameters $\left\{ n, s, \left\{ \hat{\theta}_i \right\}_{i=1}^s, \{R_i\}_{i=1}^s \right\}$, as follows:

$$\min_{n, s, \{\theta_i\}, \{R_i\}} \mathbb{E} \left[\sum_{i=1}^s \mathbb{1}_{[r \in R_i]} d(y, [r_t^{\mathrm{T}} \otimes I_n] \theta_i) \right], \tag{16}$$

where $y \in \mathbb{R}^n$ and $r \in \mathbb{R}^{n+p}$ represent random variables, realizations of which constitute the system observations, the nonnegative measure d is an appropriately defined dissimilarity measure, and the expectation is taken with respect to the joint distribution of $(y, r) \in \mathbb{R}^{2n+p}$ that depends on the system dynamics, the control input, and the noise term.

It is clear that the optimization problem (16) is intractable as is. In particular, notice that the both the model order n, and the number of modes s, completely alter the cardinality and the domain of θ_i, $i \in \{1, \dots, s\}$ that represent the dynamics of the system. In addition, a parametric representation for the polyhedral regions $R_i, i \in \{1, \dots, s\}$, that form a partition of $R \subseteq \mathbb{R}^{n+p}$ satisfying $R_i \subset R$, $R_i \cap R_j = \emptyset$ for $i \neq j$, and $\bigcup_i R_i = R$, should be defined. Finally, the expectation operation cannot be analytically computed in general. On the other hand, under Assumption 2, because of full state observability ($C = I$), knowledge of the dimension n of the state space can be assumed a priori.

IV. Mode Identification with Online Deterministic Annealing

In this section we will adopt the online deterministic annealing method [11], [14], [15] to solve the problem of finding s and $\{R_i\}_{i=1}^s$ given that $\{\theta_i\}_{i=1}^s$ are known. We introduce a set of variables $\{\rho_i\}_{i=1}^K$, $\rho_i \in R$ each one representing a region

$$\Sigma_i = \left\{ r \in R : i = \arg\min_j d(r, \rho_j) \right\}, \tag{17}$$

for a given dissimilarity measure d. The measure d can be designed such that the Voronoi regions Σ_i are polyhedral, e.g., Euclidean distance or any Bregman divergence, as will be explained in Section IV-A. In this sense, each R_i can be mapped to a region Σ_j (for $K = s$) or a union of adjacent sets $\{\Sigma_j\}$ (for $K > s$), as will be explained in Section V-B.

Problem (16) then becomes a clustering problem:

$$\min_{\{\hat{r}_i\}} \mathbb{E}\left[\sum_{i=1}^K \mathbb{1}_{[r\in\Sigma_i]} d(X, \mu_i)\right], \tag{18}$$

on the augmented space of the random variable $X = [\theta^{\mathrm{T}} r^{\mathrm{T}}]^{\mathrm{T}} \in S \subseteq \mathbb{R}^{(n+1)(n+p)}$ defined in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where μ_i is the augmented codevector:

$$\mu_i := \begin{bmatrix} \hat{\theta}_i \\ \hat{r}_i \end{bmatrix} \in S, \; i = 1, \ldots, K, \tag{19}$$

with $\hat{\theta}_i$ being an estimate of θ_i (so far we assume $\hat{\theta}_i = \theta_i$). Here the measure $d : S \times S \to [0, \infty)$ is a dissimilarity measure defined on S. Problem (18) is a hard clustering problem with respect to the parameters $\{\hat{r}_i\}_{i=1}^K$. The lowest possible number K should also be computed.

A. Online Deterministic Annealing

To construct a recursive stochastic optimization algorithm to solve problem (18) and progressively estimate the number K of the augmented codevectors $\{\mu_i\}_{i=1}^K$, we adopt the online deterministic annealing approach [14], [15]. Recall that the observed data are represented by the random variable $X : \Omega \to S \subseteq \mathbb{R}^{(n+1)(n+p)}$ defined in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and the augmented codevectors $\{\mu_i\}_{i=1}^K$ are treated as constant parameters to be estimated. According to the online deterministic annealing principles [14], [15], we extend this approach and define a probability space over an infinite number of codevectors, while constraining their distribution using a maximum-entropy principle at different levels. First we define a quantizer $Q : S \to \mathrm{ri}(S)$ as a discrete random variable in the same probability space with countably infinite domain $\mu := \{\mu_i\}$. Then we formulate the multi-objective optimization

$$\min_\mu F_\lambda(\mu) := (1-\lambda) D(\mu) - \lambda H(\mu), \; \lambda \in [0, 1), \tag{20}$$

where $D(\mu) := \mathbb{E}\left[d\left(X, Q\right)\right] = \int p(x) \sum_i p(\mu_i|x) d(x, \mu_i) \mathrm{d}x$ takes the place of the objective in (18), and $H(\mu) := \mathbb{E}\left[-\log P(X, Q)\right]$ is the Shannon entropy which can be written as $H(\mu) = H(X) \int p(x) \sum_i p(\mu_i|x) \log p(\mu_i|x) \mathrm{d}x$. This is now a problem of finding the locations $\{\mu_i\}$ and the corresponding probabilities $\{p(\mu_i|x)\} := \{p(Q = \mu_i | X = x)\}$.

The Lagrange multiplier $\lambda \in [0, 1)$ controls the trade-off between D and H. As λ is varied, we essentially transition from one Pareto solution of the multi-objective optimization to another. The entropy term, however, introduces several properties to the approach that can be useful in many applications [14], [15], [21]–[23]. First, it introduces robustness with respect to initial conditions [14], [24]. Second, as we will discuss in Section IV-B, reducing the values of λ defines a direction that resembles an annealing process [14], [25] and induces a bifurcation phenomenon such that the number K of the codevectors is finite and depends on the value of λ.

To solve (20) for a given value of λ, we successively minimize F_λ first with respect to the association probabilities $\{p(\mu_i|x)\}$, and then with respect to the codevector locations μ. In particular, as shown in [14], [15], the sequence $\mu_i(n)$ constructed by the stochastic approximation updates

$$\begin{cases} \rho_i(t+1) & = \rho_i(t) + \beta(t) \left[\hat{p}(\mu_i|x_t) - \rho_i(t)\right] \\ \sigma_i(t+1) & = \sigma_i(t) + \beta(t) \left[x_t \hat{p}(\mu_i|x_t) - \sigma_i(t)\right], \end{cases} \tag{21}$$

where $x_t \sim X$, $\sum_t \beta(t) = \infty$, $\sum_t \beta^2(t) < \infty$, and the quantities $\hat{p}(\mu_i|x_t)$ and $\mu_i(t)$ are recursively updated as follows:

$$\mu_i(t) = \frac{\sigma_i(t)}{\rho_i(t)}, \quad \hat{p}(\mu_i|x_t) = \frac{\rho_i(t) e^{-\frac{1-\lambda}{\lambda} d(x_t, \mu_i(t))}}{\sum_i \rho_i(t) e^{-\frac{1-\lambda}{\lambda} d(x_t, \mu_i(t))}}, \tag{22}$$

converges almost surely to a solution of (20) provided that the dissimilarity measure d belongs to the family of Bregman divergences—information-theoretic dissimilarity measures play an important role in learning applications and include the widely used Euclidean distance and Kullback-Leibler divergence [14]. Therefore, throughout this paper, we will assume that the dissimilarity measure d in (17) is a Bregman divergence. In addition, the following remark holds:

Remark 3 ([11]): The partition $\{\Sigma_i\}$ induced by (17) and a dissimilarity measure d that belongs to the family of Bregman divergences, is separated by hyperplanes, such that each Σ_i is a polyhedral region for a bounded domain R.

B. Bifurcation and The Number of Modes

According to the online deterministic annealing approach, we solve a sequence of optimization problems (20) with decreasing values of λ. This process grants λ the name of a 'temperature' parameter, and induces a bifurcation phenomenon, where, as λ is lowered below some critical values, the unique values of the set $\{\mu_i\}$ that solves (20) (referred to as "effective codevectors"), form a finite set $K(\lambda)$ of increasing cardinality, which defines the estimated number of modes s [11], [14], [15], [26]. In addition, from Remark 3, it follows that the partition $\{\Sigma_i\}$ of R defined in (17) is polyhedral. Therefore, each region R_i in (4) can be mapped to a region Σ_j, if the number of effective codevectors is $K = s$, or a union of adjacent sets $\{\Sigma_j\}$ (for $K > s$). In this case, the inverse process of increasing the temperature parameter λ to merge adjacent sets Σ_i, Σ_j can be followed. For more details, the readers are referred to [11].

V. Piecewise Affine System Identification

A. Identification of Local Models

Recall that, given knowledge of the partition $\{R_i\}_{i=1}^s$, each local linear model of the PWA system in (4) is completely defined by the parameters $\{\theta_i\}$ of (15). In the following, we develop a stochastic approximation recursion to estimate $\left\{\hat{\theta}_i\right\}$. First we define the error $\epsilon(t) = [r_t^{\mathrm{T}} \otimes I_n]\theta_i(t) - y_t$. A stochastic gradient descent approach aims to minimize

$$\min_{\hat{\theta}_i} \frac{1}{2}\mathbb{E}\left[\|\epsilon(t)\|^2\right], \tag{23}$$

using the recursive updates:

$$\begin{aligned}\hat{\theta}_i(t+1) &= \hat{\theta}_i(t) - \alpha(t)\left(\nabla_{\theta_i}\epsilon(t)\right)\epsilon(t)\\ &= \hat{\theta}_i(t) - \alpha(t)[r_t^{\mathrm{T}} \otimes I_n]^{\mathrm{T}}\epsilon(t),\end{aligned} \tag{24}$$

where $\sum_n \alpha(n) = \infty$, $\sum_n \alpha^2(n) < \infty$. Here the expectation is taken with respect to the joint distribution of (y, r) as explained in III-B. This is a stochastic approximation sequence and converges almost surely to the equillibrium of the differential equation $\dot{\hat{\theta}}_i = h(\hat{\theta}_i)$, $t \geq 0$, which can be shown to be a solution of (23) with standard Lyapunov arguments. For more details the reader is referred to [15], [27]. Moreover, notice that (24) is a vectorized representation of (7), for $\gamma = \alpha(t) > 0$. Therefore, under the PE condition (14), and under the zero-mean noise assumption, it follows that $\hat{\theta}_i$ converge asymptotically to θ_i for all $i = 1, \ldots, s$, i.e., the global minimum of (23) is achieved.

B. Combined Partitioning and Local Model Identification

Notice that the estimation updates of the number of modes s and the partition $\{\Sigma_i\}_{i=1}^s$ in (21) is a stochastic approximation algorithm with a stepsize schedule $\beta(t)$. At the same time, the recursive system identification technique to estimate $\{\theta_i\}_{i=1}^s$ given $\{\Sigma_i\}_{i=1}^s$ in (24) is a stochastic approximation sequence with a stepsize schedule $\alpha(t)$. The two recursive systems can be combined using the theory of two-timescale stochastic approximation if ${}^{\beta(t)}/_{\alpha(t)} \to 0$, i.e., the estimation of the partition $\{\Sigma_i\}_{i=1}^s$ is updated at a slower rate than the updates of the parameters $\{\theta_i\}_{i=1}^s$. This follows directly from Theorem 2 in [15]. In practice, the condition ${}^{\beta(t)}/_{\alpha(t)} \to 0$ is satisfied by stepsizes of the form $(\alpha(t), \beta(t)) = ({}^1/_t, {}^1/_{1+t\log t})$, or $(\alpha(t), \beta(t)) = ({}^1/_{t^{2/3}}, {}^1/_t)$.

VI. Experimental Results

We illustrate the properties and evaluate the performance of the proposed algorithm in the following linearized PWA system:

$$\begin{cases} x_{t+1} &= (I_2 + \mathrm{d}t \begin{bmatrix} 0 & 1\\ 0 & 0\end{bmatrix})x_t + \mathrm{d}t \begin{bmatrix} 0\\ 1\end{bmatrix} u_t, \text{ if } |u_t| > 1\\ x_{t+1} &= (I_2 + \mathrm{d}t \begin{bmatrix} 0 & 1\\ 0 & -1\end{bmatrix})x_t + \mathrm{d}t \begin{bmatrix} 0\\ 0\end{bmatrix} u_t, \text{ if } |u_t| \leq 1,\end{cases} \tag{25}$$

where $x_t \in \mathbb{R}^2$ and $u_t \in \mathbb{R}$. System (25) has three modes ($s = 3$) and the and the switching signal is defined by the regions $R_1 = \left\{[x^{\mathrm{T}}|u^T]^{\mathrm{T}} \in \mathbb{R}^3 : u < -1\right\}$, $R_2 = \left\{[x^{\mathrm{T}}|u^T]^{\mathrm{T}} \in \mathbb{R}^3 : -1 \leq u < 1\right\}$, and $R_3 = \left\{[x^{\mathrm{T}}|u^T]^{\mathrm{T}} \in \mathbb{R}^3 : 1 < u\right\}$. Notice that the linear system of the second mode ($s = 2$) is not minimal. To preserve the PE conditions of Assumption 3, the input signal is chosen as $u_t = 2\cos(2\pi t * \mathrm{d}t)$, $t \in \mathbb{Z}_+$, and the noise term w_t is a zero-mean Gaussian random variable with $\sigma^2 = 0.1$. The evolution of (25) over time, as well as the mode switching behavior, are shown in Fig. 1.

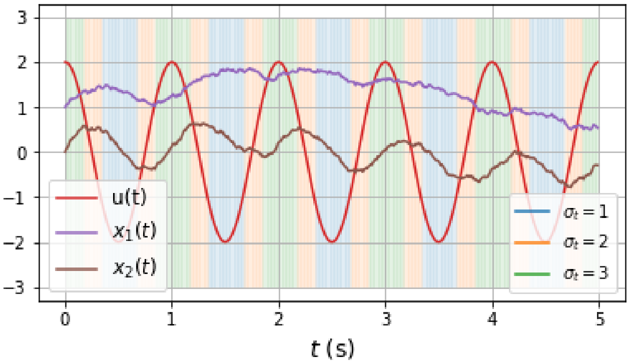

Fig. 1: Time evolution of system (25) for $T = 5$ seconds. The mode-switching behavior is depicted.

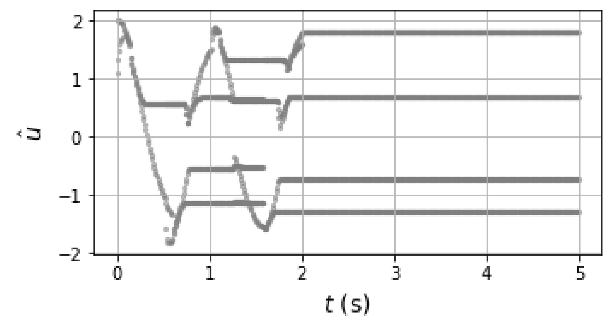

Fig. 2: Mode estimation illustrating the bifurcation phenomenon described in Section IV-B. The evolution of the $\hat{u}_i$ coordinate of the codevectors $\{\mu_i\}$ is depicted.

The system is allowed to run for $T = 5s$ (seconds), with $\mathrm{d}t = 0.01$, i.e., a total of $N = 500$ observations are acquired online. The temperature parameters used for the online deterministic annealing algorithm are $(\lambda_{\max}, \lambda_{\min}, \gamma) = (0.99, 0.1, 0.8)$, and the stepsizes $(\alpha(t), \beta(t)) = ({}^1/_{1+0.01t}, {}^1/_{1+0.9t\log t})$. The estimated parameter $\hat{\theta}_1$ gets updated by the iterations (24). We have assumed $\hat{\theta}_1(0) = [1, 1, 1, 1, 1, 1]^{\mathrm{T}}$. The bifurcation phenomenon is illustrated in Fig. 2 where the third coordinate of the codevectors $\{\hat{r}_i\}$, which gives an estimate of the control input representation of the mode, is depicted.

The estimation error and the estimated modes are shown in Fig. 3. The algorithm identifies a total of $K = 4$ modes, with the modes for which $\hat{\sigma}_t = 1$, and $\hat{\sigma}_t = 2$ representing the same dynamics of the original mode $\sigma_t = 2$. These can be combined with the inverse process explained in Section IV-B, if necessary. In Figure 4, the convergence of the parameters $\left\{\hat{\theta}_i\right\}_{i=1}^4$ are shown.

VII. Conclusion

We proposed a real-time identification scheme for discrete-time piece-wise affine state-space models. In contrast to most standard identification algorithms for piece-wise affine

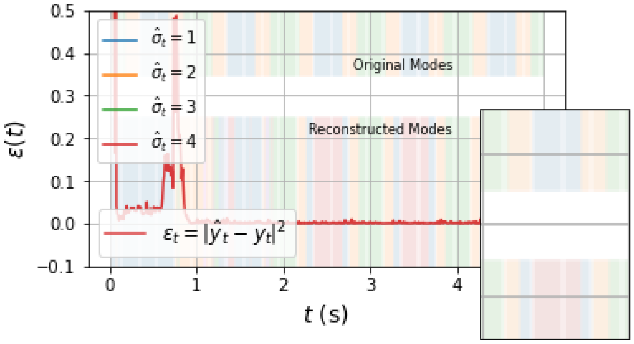

Fig. 3: Estimation error over time. The estimated modes are also compared against the original modes.

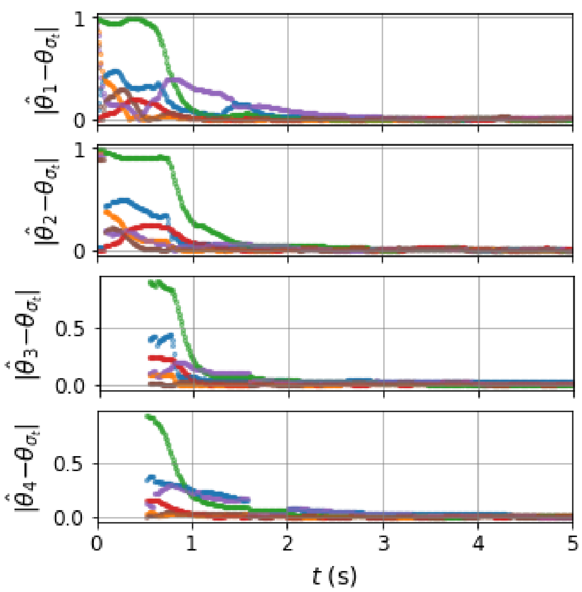

Fig. 4: Convergence of the parameters $\left\{\hat{\theta}_i\right\}_{i=1}^{4}$.

systems, the proposed approach is appropriate for online identification of both the modes and the subsystems of the switched system. The progressive nature of the algorithm also provides real-time control over the performance-complexity trade-off. Future directions include extensions of the proposed approach for identification of partially observable general switched models.

References

[1] D. Liberzon, *Switching in systems and control*. Springer, 2003, vol. 190.

[2] S. Paoletti, A. L. Juloski, G. Ferrari-Trecate, and R. Vidal, "Identification of hybrid systems a tutorial," *European journal of control*, vol. 13, no. 2-3, pp. 242–260, 2007.

[3] A. Garulli, S. Paoletti, and A. Vicino, "A survey on switched and piecewise affine system identification," *IFAC Proceedings Volumes*, vol. 45, no. 16, pp. 344–355, 2012.

[4] J. Roll, A. Bemporad, and L. Ljung, "Identification of piecewise affine systems via mixed-integer programming," *Automatica*, vol. 40, no. 1, pp. 37–50, 2004.

[5] R. Vidal, S. Soatto, Y. Ma, and S. Sastry, "An algebraic geometric approach to the identification of a class of linear hybrid systems," in *42nd IEEE International Conference on Decision and Control (IEEE Cat. No. 03CH37475)*, vol. 1. IEEE, 2003, pp. 167–172.

[6] R. Vidal, "Recursive identification of switched arx systems," *Automatica*, vol. 44, no. 9, pp. 2274–2287, 2008.

[7] G. Ferrari-Trecate, M. Muselli, D. Liberati, and M. Morari, "A clustering technique for the identification of piecewise affine systems," *Automatica*, vol. 39, no. 2, pp. 205–217, 2003.

[8] M. Gegundez, J. Aroba, and J. M. Bravo, "Identification of piecewise affine systems by means of fuzzy clustering and competitive learning," *Engineering Applications of Artificial Intelligence*, vol. 21, no. 8, pp. 1321–1329, 2008.

[9] H. Nakada, K. Takaba, and T. Katayama, "Identification of piecewise affine systems based on statistical clustering technique," *Automatica*, vol. 41, no. 5, pp. 905–913, 2005.

[10] V. Verdult and M. Verhaegen, "Subspace identification of piecewise linear systems," in *2004 43rd IEEE Conference on Decision and Control (CDC)(IEEE Cat. No. 04CH37601)*, vol. 4. IEEE, 2004, pp. 3838–3843.

[11] C. N. Mavridis and J. S. Baras, "Identification of piecewise affine systems with online deterministic annealing," in *2023 62nd IEEE Conference on Decision and Control (CDC)*. IEEE, 2023, pp. 4885–4890.

[12] A. M. Ivanescu, T. Albin, D. Abel, and T. Seidl, "Employing correlation clustering for the identification of piecewise affine models," in *Proceedings of the 2011 workshop on Knowledge discovery, modeling and simulation*, 2011, pp. 7–14.

[13] C. N. Mavridis and J. S. Baras, "Convergence of stochastic vector quantization and learning vector quantization with bregman divergences," *IFAC-PapersOnLine*, vol. 53, no. 2, 2020.

[14] ——, "Online deterministic annealing for classification and clustering," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 10, pp. 7125–7134, 2023.

[15] C. Mavridis and J. S. Baras, "Annealing optimization for progressive learning with stochastic approximation," *IEEE Transactions on Automatic Control*, vol. 68, no. 5, pp. 2862–2874, 2023.

[16] M. Petreczky, L. Bako, and J. H. Van Schuppen, "Realization theory of discrete-time linear switched systems," *Automatica*, vol. 49, no. 11, pp. 3337–3344, 2013.

[17] R. Vidal, A. Chiuso, and S. Soatto, "Observability and identifiability of jump linear systems," in *Proceedings of the 41st IEEE Conference on Decision and Control, 2002.*, vol. 4. IEEE, 2002, pp. 3614–3619.

[18] M. Petreczky, L. Bako, and J. H. van Schuppen, "Identifiability of discrete-time linear switched systems," in *Proceedings of the 13th ACM international conference on hybrid systems: Computation and control*, 2010, pp. 141–150.

[19] B. Anderson and J. Moore, "New results in linear system stability," *SIAM Journal on Control*, vol. 7, no. 3, pp. 398–414, 1969.

[20] B. M. Jenkins, A. M. Annaswamy, E. Lavretsky, and T. E. Gibson, "Convergence properties of adaptive systems and the definition of exponential stability," *SIAM journal on control and optimization*, vol. 56, no. 4, pp. 2463–2484, 2018.

[21] C. N. Mavridis, N. Suriyarachchi, and J. S. Baras, "Maximum-entropy progressive state aggregation for reinforcement learning," in *IEEE Conference on Decision and Control*, 2021, pp. 5144–5149.

[22] C. N. Mavridis and J. S. Baras, "Progressive graph partitioning based on information diffusion," in *IEEE Conference on Decision and Control*, 2021, pp. 37–42.

[23] C. N. Mavridis, N. Suriyarachchi, and J. S. Baras, "Detection of dynamically changing leaders in complex swarms from observed dynamic data," in *International Conference on Decision and Game Theory for Security*. Springer, 2020, pp. 223–240.

[24] C. Mavridis, E. Noorani, and J. S. Baras, "Risk sensitivity and entropy regularization in prototype-based learning," in *2022 30th Mediterranean Conference on Control and Automation (MED)*. IEEE, 2022, pp. 194–199.

[25] K. Rose, "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2210–2239, 1998.

[26] C. Mavridis and J. Baras, "Multi-resolution online deterministic annealing: A hierarchical and progressive learning architecture," *arXiv preprint arXiv:2212.08189*, 2022.

[27] V. S. Borkar, *Stochastic approximation: a dynamical systems viewpoint*. Springer, 2009, vol. 48.