

# Switched Linear System Identification based on Bounded-Switching Clustering\*

Mohammad Gorji Sefidmazgi, Mina Moradi Kordmahalleh, Abdollah Homaifar, and Ali Karimodini\*\*

**Abstract**—This paper aims at identifying switched linear systems, which are described by noisy input/output data. This problem is originally non-convex and ill-posed. The proposed approach utilizes bounded-switching clustering method to convert the problem into a binary integer optimization and least square. This method optimally divides a time series into several clusters whose parameters are piecewise constant in time. Optimal number and order of linear sub-systems as well as the number of switches are selected using Akaike Information Criterion. The performance of the algorithm is evaluated through simulations. Parameters and structures of switched systems are found accurately in the presence of noise.

## I. INTRODUCTION

In recent years, there has been an increasing interest in using switched linear systems to model and analyze control systems [1]. Switched linear systems are an important class of hybrid systems, which consist of sub-systems and a switching mechanism that activates one of the sub-systems. In many practical cases, it is not possible to directly obtain the model of sub-systems and extract the switching mechanism analytically. Instead, the system can be described by time series obtained from experimental data. However, this does not remove the need for an explicit model to analyze the system. Thus, it is desired to extract the sub-systems and the switching mechanism from the given time series, which are often corrupted by measurement noise. Therefore, given inputs/outputs data for a switched linear system, this paper aims at identifying model parameters and switching sequence that can explain the available experimental data.

The switched model identification has been widely studied over the last decade in control system literature. A review of these methods can be found in [2]. Each method analyzes the problem from a different viewpoint. The algebraic approach [3] converts the problem to a linear system of equations, but the solution is sensitive to the noise. Furthermore, the

complexity of these approaches grows exponentially with the number of sub-models and length of data [4]. There are methods based on machine learning tools such as K-means or support vector machines [5]–[9] in which the data is mapped into a feature space where the clusters become separable. For an acceptable performance, these approaches need fair sampling of each cluster and the closeness of data points which belong to the same sub-system [10]. The Bayesian approach [11] and the maximum likelihood [12] analyze the problem in a probabilistic basis. In these methods, a particular probability distribution is assumed for the given data and then, the data are assigned to the clusters with higher probability of membership. In Markov methods [13], data is assumed to be normally distributed, and the switching sequence is modeled as a Markov model with a transition probability matrix. Defining a suitable likelihood function, the parameters and the transition probabilities are estimated by employing an optimization technique. As discussed, these approaches make specific statistical assumptions on distribution of the given dataset, which may not be practical in many cases.

In terms of minimizing the modelling error, some methods have been developed in [4], [14]–[17] to directly solve the essentially non-convex problem of switched linear system modelling, but in these methods there is no guarantee of finding the global solution. The method in [18] that used mixed integer programming will guarantee the convergence to the global optimum. However, due to high complexity, this method is only suitable for relatively few data and thus, it should be used with a sliding window for longer data. In some other methods, the non-convex optimization problem is transformed to a convex optimization problem by introducing relaxation on the optimization [10], [19]–[23]. Hence, the global solution can be found for the convex optimization, and under certain conditions, the result of transformed convex optimization problem is equivalent to the solution of original non-convex problem of identifying the switched linear systems.

Difficulties associated in tuning some hyperparameters of aforementioned approaches need to be addressed. For example it is necessary to tune the length of sliding windows [18], penalization factor [20]–[23], noise level [5], [10], forgetting factor [17] or learning rate [3]. However, these hyperparameters are real-valued numbers with infinite possible choices.

In order to address these problems, in this paper we use Bounded-Switching (BS) approach [24] to cluster the given data of a switched linear system. This method assumes

\*This work is partially supported by the Expeditions in Computing by the National Science Foundation under Award Number CCF-1029731 and DOT under UTC-60040605 and Air Force Research Laboratory and OSD under agreement number FA8750-15-2-0116. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of NSF, DoT, Air Force Research Laboratory, OSD or the U.S. Government.

\*\*M. Gorji Sefidmazgi, M. Moradi Kordmahalleh, A. Homaifar and A. Karimodini are with the Autonomous Control and Information Technology Center, Department of Electrical and Computer Engineering, North Carolina A&T State University, Greensboro, NC 27410 USA.  
{mgorjise, mmoradik}@aggies.ncat.edu,  
{homaifar, akarimod}@ncat.edu  
Corresponding Author: A. Homaifar, (phone: +13362853271)

that the system parameters are piecewise constant in time, and there are limited numbers of switches among the sub-systems. The proposed method uses BS-clustering to convert the optimization of identification problem to the combination of least square and binary integer programming, which can be solved iteratively. The proposed algorithm processes the data in batch mode, and identification of the switching times and sub-system parameters take place at the same time. This method does not require restrictive assumptions on the statistical distribution of data. More importantly, this identification method does not require tuning of any real-valued hyperparameters. By adding statistical assumptions on the residuals (differences between observations and estimated switched model), the information theory methods are used to tune the number, order of sub-systems and the total number of switches among them. Therefore, it is only necessary to tune a small subset of integer-valued hyperparameters from a finite set of admissible structures.

This paper therefore proposes a method to identify a switched linear system using the BS-clustering in order to find the switching times and the parameters of the switched sub-systems. Akaike Information Criterion (AIC) is used to find the optimal number and structure of the sub-systems and number of switches. The algorithm is evaluated through simulations.

The paper is organized as follows. In section II, the identification problem of switched linear systems is defined mathematically. Section III explains the concept of bounded-switching. Section IV describes the Bounded-Switching clustering for identification problem. Section V explains the model selection with AIC. The performance of the algorithm is evaluated through simulations in section VI. Section VII provided conclusions and future work.

## II. PROBLEM FORMULATION

Consider a switched linear system that consists of  $S$  sub-systems, each one is modeled as an autoregressive exogenous (ARX) model, together form a switched autoregressive exogenous (SARX) model. The time driven dynamics of each sub-model could be represented in discrete time as follows:

$$y(t) = [a_1^s \ a_2^s \ \dots \ a_m^s] \cdot [y(t-1) \ y(t-2) \ \dots \ y(t-m)]^T + [b_1^s \ b_2^s \ \dots \ b_n^s] \cdot [u(t-1) \ u(t-2) \ \dots \ u(t-n)]^T + e(t) \quad (1)$$

Where  $y \in \mathbb{R}$  is the output,  $u \in \mathbb{R}$  is the input of the system,  $s \in \{1, 2, \dots, S\}$  are the indices of sub-models,  $e(t)$  is the additive noise, and  $t$  is the time index. Here,  $\{a_j^s\}_{j=1}^m$  and  $\{b_j^s\}_{j=1}^n$  are the parameters of the  $s$ th sub-model. The activation sequence of sub-models and coefficients of the SARX model are unknown. Given the discrete pairs of data in the form of  $\{u_t, y_t\}_{t=1}^T$  generated by the sub-systems of a switched system (1), our aim is to estimate model parameters of the switched system  $\{a_j^s\}_{j=1}^m$  and  $\{b_j^s\}_{j=1}^n$ , the orders  $m$  and  $n$ , the time instants that switching among  $S$  sub-models occur, and the activated sub-models at any given time  $t$ . The

data set  $\{u_t, y_t\}_{t=1}^T$  should be clustered into several sets of input/output pairs. In this paper, we assume that all of  $S$  sub-models are stable and observable. Also we assume that the orders  $m$  and  $n$  are unknown, but they are finite and equal for all sub-models. We also assume that the inputs and outputs are corrupted with disturbances and measurement noises appear in the form of additive white noise. Defining the regression function,  $\varphi(t)$ , and parameter vectors,  $\theta_s$ , for the  $s$ th sub-model:

$$\varphi(t) = \begin{bmatrix} y(t-1) & y(t-2) & \dots & y(t-m) \\ u(t-1) & u(t-2) & \dots & u(t-n) \end{bmatrix}^T \quad t = \{1, \dots, T\} \quad (2)$$

$$\theta_s = [a_1^s \ a_2^s \ \dots \ a_m^s \ b_1^s \ b_2^s \ \dots \ b_n^s]^T \quad s = \{1, \dots, S\} \quad (3)$$

Eq. (1) can be rewritten in a compact form:

$$y(t) = \varphi^T(t) \cdot \theta_s + e(t) \quad t = \{1, \dots, T\} \quad (4)$$

Now, the identification of this switched linear system can be formulated as a minimization problem:

$$\text{Min}_{\mu_s(t), \theta_s} J(\mu_s(t), \theta_s) \quad (5)$$

Here, the aim is to find  $\mu_s(t)$  and  $\theta_s$  that minimize  $J(\mu_s(t), \theta_s)$ , where

$$J(\mu_s(t), \theta_s) = \sum_{t=1}^T \sum_{s=1}^S \mu_s(t) \cdot [y(t) - \varphi^T(t) \cdot \theta_s]^2 \quad (6)$$

Where  $[\theta_1, \dots, \theta_S]$  is the set of all parameters of  $S$  sub-models. Also  $\mu_s(t) \in \{0, 1\}$  is the sub-model activation function, shows which sub-model is active at time  $t$ , and has the following property:

$$\sum_{s=1}^S \mu_s(t) = 1 \quad (7)$$

Alternatively, the minimization in (5) can be rewritten in a more compact form [24]. Lets define Distance Function  $d_s(t)$  as Euclidean distance between the output of the system  $y(t)$  and model of  $s$ th sub-system as below:

$$d_s(t) = [y(t) - \varphi^T(t) \cdot \theta_s]^2 \quad t = \{1, \dots, T\}, s = \{1, \dots, S\} \quad (8)$$

The values of distance function and sub-model activation function are augmented in two vectors for  $s = \{1, \dots, S\}$

$$D_s = [d_s(1) \ d_s(2) \ \dots \ d_s(T)] \quad (9)$$

$$M_s = [\mu_s(1) \ \mu_s(2) \ \dots \ \mu_s(T)] \quad (10)$$

The optimization in (5) can be rewritten as

$$\min_{D_s, M_s} \sum_{s=1}^S D_s \cdot M_s^T \quad (11)$$

Finding the optimal solution for this problem requires searching among all possible discrete state paths, which is very time-consuming. In general, the cost function in (5) is a

non-convex optimization problem [19]. In this work, the non-convex optimization problem is converted to a binary integer programming and a least square problem, in order to make the identification problem solvable.

### III. REGULARIZATION

There are two sets of unknown parameters: sub-model parameters  $\theta_s$  and sub-model activation function  $\mu_s(t)$  as mentioned in section II. When  $\mu_s(t)$  is known, finding  $\theta_s$  using least square is straightforward. The difficulty of identification problem arises in estimating the  $\mu_s(t)$ . Suppose that the values of sub-model parameters  $\theta_s$  are known and equal to  $\theta_s^*$  for  $s = 1, \dots, S$ . In this case, the optimization in (5) should be solved to find  $\mu_s(t)$  by assigning each data to one of the  $S$  clusters. One method for finding  $\mu_s(t)$  is to assign data to the sub-model that its output has shortest distance to the output of the system  $y(t)$  [4]:

$$\mu_s(t) = \begin{cases} 1 & s = \arg \min_{j=1, \dots, S} [y(t) - \varphi^T(t) \cdot \theta_j^*]^2 \\ 0 & \text{otherwise} \end{cases} \quad t = \{1, \dots, T\} \quad (12)$$

However the solution in (12) may lead to finding  $\mu_s(t)$  with high number of switches in the activation sequence of sub-models. This means that  $\mu_s(t)$  might rapidly oscillate between  $S$  clusters, even if the real number of switches among sub-models is limited. In this case, small changes in the data might lead to a large changes in the values of  $\mu_s(t)$ , and thus, the solution does not continuously depends on the data [24]. This property of solution in (12) violates the Hadamards definition of well-posed problem [25] and thus the optimization in (5) is ill-posed. To overcome this problem, the regularization techniques is necessary, which incorporate additional assumptions to make the problem well-posed.

One type of possible regularization is to limit the total number of switches between the sub-models. Some methods add minimum number of switches by making the vector of variations in  $\theta_s$  sparse, i.e. most of the elements of this vector becomes zero [10], [21]–[23]. In these methods, the number of sub-models cannot be fixed. Furthermore, in [10], [22], [23] the sub-models are non-returning, i.e. if a sub-system has been activated previously, and again re-activated, the algorithm considers it as a new sub-system rather than a returning sub-system. A regularization method is proposed in [20] by adding L2-norm penalization on variations in  $\mu_s(t)$  between two consecutive time steps. Method in [10] requires prior knowledge on noise level. The methods in [20]–[23] need tuning of the real-valued penalization factor. Furthermore, the direct mathematical relation between the value of the penalization factor and the total number of switches cannot be found in these approaches. To address these problems, here we adopt Bounded-Switching (BS) technique, which we can either set or find the total number of sub-models and also detect returning sub-models without the need for tuning penalization factor or threshold.

### IV. BOUNDED-SWITCHING CLUSTERING

The analysis of time series with switched models based on bounded-switching method is introduced in [24] and is used for climate data analysis [26] and is extended to dynamic systems modeling [27]. The basic idea is to assume an upper bound on the total number of switches in the system to make the ill-posed clustering problem a solvable optimization. In this method, the optimization is solved by *coordinate descent algorithm* [28], i.e. iteratively with respect to two sets of unknown parameters, sub-model parameters and activation functions. As mentioned earlier, if  $\mu_s(t)$  is known, finding the sub-model parameters,  $\theta_s$ , is straightforward. In this case,  $\theta_s$  can be directly found from (5) in closed form by the least square method [20]:

$$\theta_s = \left[ \sum_{t=1}^T \mu_s(t) \cdot \varphi(t) \cdot \varphi^T(t) \right]^{-1} \cdot \left[ \sum_{t=1}^T \mu_s(t) \cdot \varphi(t) \cdot y(t) \right] \quad s \in \{1, \dots, S\} \quad (13)$$

In the case that the first term in (13) is singular, one can apply recursive least square to find  $\theta_s$ . Next, suppose that  $\theta_s$  is known for all of the sub-models. Here, we show how  $\mu_s(t)$  can be found using a clustering algorithm by defining the functions  $q_s(t) \in \{0, 1\}$  as an upper bound for variations of  $\mu_s(t)$  over time as follows:

$$|\mu_s(t+1) - \mu_s(t)| \leq q_s(t) \quad s \in \{1, \dots, S\}, t \in \{1, \dots, T-1\} \quad (14)$$

Eq. (14) can be extended as [24]:

$$\begin{cases} \mu_s(t+1) - \mu_s(t) - q_s(t) \leq 0 \\ -\mu_s(t+1) + \mu_s(t) - q_s(t) \leq 0 \end{cases} \quad s \in \{1, \dots, S\}, t \in \{1, \dots, T-1\} \quad (15)$$

At switch times between two sub-systems e.g.  $s_1$  to  $s_2$ , the value of  $\mu_{s_1}(t)$  changes from 1 to 0 and the value of  $\mu_{s_2}(t)$  changes from 0 to 1. Thus, in order to limit the total number of switches to be less than  $K$ , one can add a constraint on the  $q_s(t)$  as follows:

$$\sum_{s=1}^S \sum_{t=1}^{T-1} q_s(t) = 2K \quad (16)$$

where  $K$  is an integer number. The value of  $K$  might be a small number (near zero) for systems with small number of switches or it might be a large number (near  $T$ ) in a fast switching system. Next, we should include  $q_s(t)$  in the set of parameters to be estimated. We define the vector  $Q_s$  containing all values of  $q_s(t)$  over time [24]

$$Q_s = [q_s(1) \quad q_s(2) \quad \dots \quad q_s(T-1)] \quad (17)$$

Defining vector  $\bar{M}$  containing all of  $\mu_s(t)$  and  $q_s(t)$

$$\bar{M} = \left[ \underbrace{M_1 \quad M_2 \quad \dots \quad M_S}_{T \times S} \quad \underbrace{Q_1 \quad Q_2 \quad \dots \quad Q_S}_{(T-1) \times S} \right] \quad (18)$$

And the vector  $\bar{D}$  containing all the values of  $d_s(t)$

$$\bar{D} = \left[ \underbrace{D_1 \ D_2 \ \dots \ D_S}_{T \times S} \ \underbrace{0 \ 0 \ \dots \ 0}_{(T-1) \times S} \right] \quad (19)$$

Now, the optimization in (11) can be rewritten by  $\bar{M}$  and  $\bar{D}$  as [24]

$$\min_{\bar{M}} \bar{D} \cdot \bar{M}^T \quad (20)$$

The vector  $\bar{M}$ , which contains all of  $\mu_s(t)$  and  $q_s(t)$ , can be found by solving this minimization. Since all the values of  $\mu_s(t)$  and  $q_s(t)$  are binary, the vector  $\bar{M}$  contains only binary numbers and thus (20) is a binary integer programming which is solvable by standard approaches of optimization. The order of this optimization problem is  $(2T - 1)S$ . The equations (7), (15) and (16) are the set of linear constraints that include  $2S(T - 1)$  inequality and  $T + 1$  equality constraints. In this work, we use the toolbox developed in [29] which utilizes branch-and-bound algorithm for solving binary integer programming with complexity of  $O(2^n)$ . Note that our method based on binary integer programming is completely different from the method in [18] which applied the mixed integer programming for the dataset in a sliding window. In fact, our proposed method in this paper applies binary integer programming to all dataset and thus it doesn't require tuning the length of window. The proposed identification algorithm is summarized in Algorithm 1. First an initial value for  $\mu_s(t)$  is chosen and then two iterative steps of eq. (13) and eq. (20) are followed until the algorithm converges. It is shown that BS-clustering converges to at least a local solution. For finding the global solution, the algorithm should be started from initially generated  $\mu_s(t)$ .

---

**Algorithm 1** Identification of SARX based on BS-clustering

---

- 1: Choose the number of sub-models  $S$ , and the number of switches  $K$
  - 2: Set iteration number  $L = 1$
  - 3: Generate initial  $\mu_s^{[1]}(t)$  satisfying (7), find the sub-model parameters  $\theta_s^{[1]}$  by (13)
  - 4: Having  $\theta_s^{[L]}$ , find the sub-model activation function  $\mu_s^{[L+1]}(t)$  by solving (20) using constraints in (7), (15) and (16).
  - 5: Having  $\mu_s^{[L+1]}(t)$ , find the sub-model parameters  $\theta_s^{[L+1]}$  by (13)
  - 6: Set  $L = L + 1$  and then repeat steps 5 and 6 until the algorithm converges, i.e.  $\left| J(\mu_s^{[L+1]}, \theta_s^{[L+1]}) - J(\mu_s^{[L]}, \theta_s^{[L]}) \right| \geq \epsilon$ , where  $\epsilon$  is a very small number.
- 

## V. MODEL SELECTION

In the proposed algorithm, we have assumed that the structure of the model, i.e. number  $S$  and order of sub-models  $m$ ,  $n$  and number of switches  $K$  are given. But, if these parameters are not known, then selecting an appropriate model structure in switched model identification is a challenging problem as there exist multiple solutions. For example, if sub-model orders ( $m$  and  $n$ ) are not fixed, we can find a trivial switched model consisting of one single sub-model

with sufficiently large orders that fits all the finite set of measurements. When finite and fixed values are assigned to  $m$  and  $n$ , still there exist infinitely many switched models that explain the data. For example, a trivial solution is a switched model with  $T$  sub-models that can model a system with  $T$  pairs of data, which is the largest possible number of sub-models. Therefore, we should assume an upper bound on the number of sub-models and find a model without underfitting or overfitting.

The AIC is a well-known approach to perform a trade-off between the goodness of fit and the complexity of models. It is widely used in system identification literature to find the order of dynamical systems. Also, it can find the total number of switches in time series with a switched model (i.e. non-stationary time series) [30]. Here, we extend these applications to find number and order of sub-models and also number of switches and sub-models in SARX system. For this purpose, let the index of active sub-model at  $t$  be:

$$s^*(t) = \arg \max_{s=1, \dots, S} \mu_s(t) \quad (21)$$

The output of identified model is derived from:

$$\hat{y}(t) = \varphi^T(t) \cdot \theta_{s^*(t)} \quad (22)$$

Thus the loss function is defined as:

$$V = \frac{1}{T} \sum_{t=1}^T (y(t) - \hat{y}(t))^2 \quad (23)$$

If the model errors is assumed to be normally distributed, the AIC can be calculated as [31]:

$$AIC = \ln(V) + \frac{2 \times \text{number of parameters}}{T} \quad (24)$$

In (24), we need to know the number of estimated parameters for each identified structures. Let  $\omega = m + n$  be the number of parameters for each sub-model. For  $S$  sub-models, the total number of parameters is  $S \times \omega$ . In addition, there are  $K$  switching times which should be estimated too. Thus:

$$\text{number of parameters} = S \times \omega + K \quad (25)$$

The best model is the one with lowest AIC. Hence, if the model structure is not given, a proper solution is to calculate the values of AIC for different admissible structures and choose the structure which has the minimum AIC.

## VI. SIMULATION

In this section, two examples are shown to verify the performance of the proposed algorithm. The first example is a switched-linear systems borrowed from [3] with two first order sub-systems. Example 2 identifies a switched system with four sub-models which have higher orders than the first example.

*Example 1:* We applied the identification method on a first order SARX system with two sub-models borrowed from [3]. The length of data is 1000 and the sequence of two sub-models  $C_1$  and  $C_2$  is periodic with a period of 30  $s$  and the input is  $u(t) = N(0, 1)$ . The noise is Gaussian with  $w(t) = N(0, 0.3^2)$ . The model in each cluster is in following form:

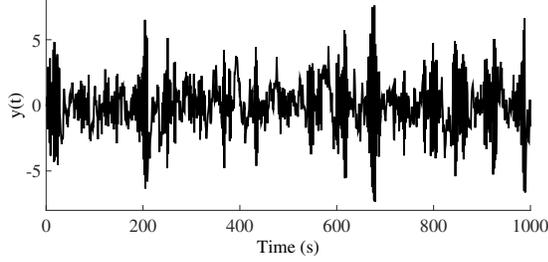


Fig. 1. Output of the switched system with two sub-models in Example 1

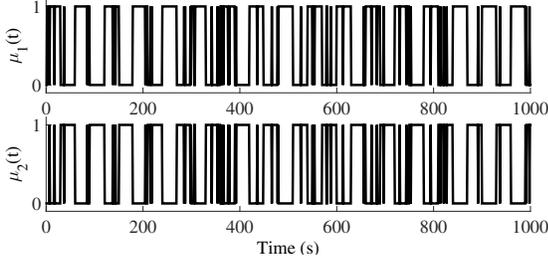


Fig. 2. Detected sub-model activation functions without using any regularization in Example 1

$$y(t) = \begin{cases} -0.9y(t-1) + u(t-1) + w(t), & t \in \mathcal{C}_1 \\ 0.7y(t-1) - u(t-1) + w(t), & t \in \mathcal{C}_2 \end{cases} \quad (26)$$

The output of the system is shown in Figure 1. Assume that the true parameters of sub-models are given as (26). In this case, if we follow (12) to find sub-model activation function  $\mu_s(t)$ , the detected clusters are shown in Figure 2. It can be seen that there are too many switches between the clusters and the result is not correct. Furthermore, if another realization of noise  $w(t)$  with the same variance is added to the output, it will cause changes in  $\mu_s(t)$ . This shows that the clustering problem is ill-posed and regularization is necessary. Next, the BS-clustering is applied to the system with different numbers of sub-models  $S = \{2, 3, 4, 5\}$ , number of switches  $K = \{30, 31, 32, 33, 34, 35\}$ , orders  $m = \{1, 2\}$ , and  $n = \{1, 2\}$ . As noted, totally there are 96 different cases and for each case, the value of AIC is obtained. The results are presented in Figure 3. In this figure, each line shows the values of AIC with respect to  $K$  for a specific  $S$ ,  $m$  and  $n$ . The smallest AIC is -2.2042 which is found for  $S = 2$ ,  $K = 33$ ,  $m = 1$  and  $n = 1$  which is shown by an arrow in Figure 3. Figure 4 shows

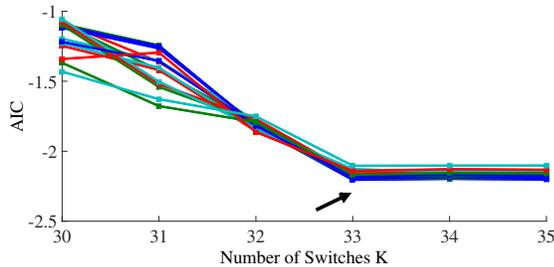


Fig. 3. The values of AIC for different structures of switched model in Example 1

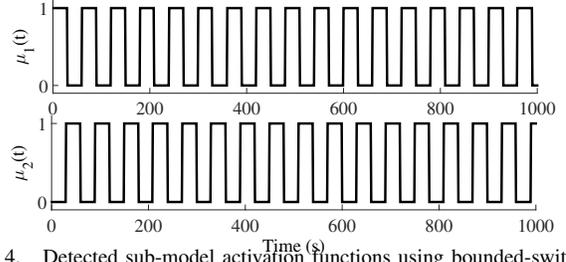


Fig. 4. Detected sub-model activation functions using bounded-switching regularization in Example 1

$\mu_s(t)$  for two sub-models and 33 switches with smallest AIC. The BS-clustering is converged after five iterations and the parameters of the two sub-models are found as  $[-0.8948, 1.0256; 0.7046, -1.0092]$  which has good accuracy with respect to true values in (26). The identification results of our proposed method is comparable with the results in [3], but does not require to tune the real-valued learning rate as in [3]. In addition, we have estimated the degree and number of sub-models in our method.

*Example 2:* This example has four sub-models with different time-lags. The input  $u$  is  $\pm 1$  PRBS (Pseudo random binary signal). The additive noise is a signal with Gaussian distribution and zero-mean. The signal-to-noise ratio for the output signal is 25 dB.

$$y(t) = \begin{cases} 0.1y(t-3) + 0.3u(t-1) & t \in [1, 20) \cup [80, 100) \cup [170, 200) \\ -0.4y(t-1) + 0.6u(t-2) & t \in [20, 50) \cup [100, 120) \cup [200, 220) \\ -0.1y(t-1) + 0.8u(t-1) & t \in [50, 60) \cup [120, 150) \cup [220, 240) \\ 0.2y(t-2) + 0.6u(t-1) & t \in [60, 80) \cup [150, 170) \cup [240, 250) \end{cases} \quad (27)$$

Figure 5 shows the AIC with respect to  $K$  for different values of  $S = \{2, 3, 4, 5\}$ ,  $K = \{9, 10, 11, 12, 13\}$ ,  $m = \{1, 2, 3\}$ , and  $n = \{1, 2, 3\}$ . The minimum value of AIC is -6.3064 marked with an arrow and was found for  $S = 4$ ,  $K = 11$ ,  $n = 2$  and  $m = 3$ . Figure 6 shows the sub-model activation function for each of the 4 sub-models. The output of the switched system  $y(t)$ , and the output of switched model  $\hat{y}(t)$  are shown in Figure 7. The values of these four sub-model parameters  $\{a_1^s, a_2^s, a_3^s, b_1^s, b_2^s\}$  are found and given as below which are close to their actual values. Clearly, by increasing the admissible structures ( $S$ ,  $K$ ,  $m$  and  $n$ ), the number of calculated AIC values and thus the algorithm complexity increase.

$$\begin{pmatrix} -0.0541 & -0.0027 & 0.0937 & 0.3068 & 0.0043 \\ -0.3905 & 0.0101 & 0.0056 & 0.0024 & 0.6015 \\ -0.0840 & 0.0041 & -0.0042 & 0.8005 & -0.0063 \\ -0.0064 & 0.2007 & 0.0006 & 0.6075 & 0.0026 \end{pmatrix}$$

## VII. CONCLUSIONS

In this paper, an algorithm was introduced for identification of switched linear systems based on the bounded-switching clustering technique. Compared to conventional

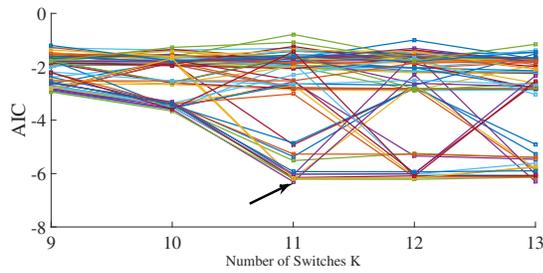


Fig. 5. The Values of AIC for different structures of switched model in Example 2

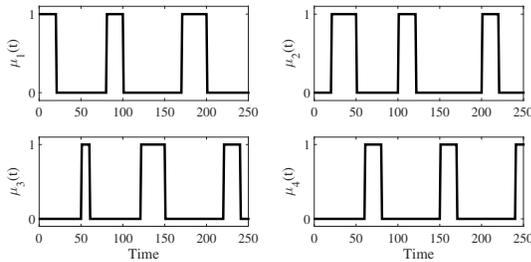


Fig. 6. Detected sub-model activation functions using bounded-switching regularization in Example 2

approaches, the proposed method does not require any statistical assumption on the data and tuning of any real-valued hyperparameter. Instead, it utilizes Akaike Information Criterion to find the optimal number and order of sub-models as well as number of switches. The proposed identification method was tested through MATLAB simulation where the sub-model parameters and their orders are found to be very accurate.

## REFERENCES

- [1] D. Liberzon, *Switching in Systems and Control*. Birkhuser, 2012.
- [2] A. Garulli, S. Paoletti, and A. Vicino, "A survey on switched and piecewise affine system identification," in *16th IFAC Symposium on System Identification*, 2012, pp. 344 – 355.
- [3] R. Vidal, "Recursive identification of switched ARX systems," *Automatica*, vol. 44, no. 9, pp. 2274 – 2287, 2008.
- [4] F. Lauer, "Estimating the probability of success of a simple algorithm for switched linear regression," *Nonlinear Analysis: Hybrid Systems*, vol. 8, no. 0, pp. 31 – 47, 2013.
- [5] V. L. Le, F. Lauer, and G. Bloch, "Identification of linear hybrid systems: A geometric approach," in *American Control Conference (ACC)*, 2013, pp. 830–835.
- [6] N. Canty, T. O'Mahony, and M. T. Cychowski, "An output error algorithm for piecewise affine system identification," *Control Engineering Practice*, vol. 20, no. 4, pp. 444 – 452, 2012.
- [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] Z. Lassoued and K. Abderrahim, "A new clustering technique for the identification of pwarx hybrid models," in *9th Asian Control Conference (ASCC)*, 2013, pp. 1–6.
- [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] N. Ozay, M. Sznaiier, C. M. Lagoa, and O. I. Camps, "A sparsification approach to set membership identification of switched affine systems," *Automatic Control, IEEE Transactions on*, vol. 57, no. 3, pp. 634–648, 2012.
- [11] A. L. Juloski, S. Weiland, and W. Heemels, "A bayesian approach to identification of hybrid systems," *Automatic Control, IEEE Transactions on*, vol. 50, no. 10, pp. 1520–1533, 2005.

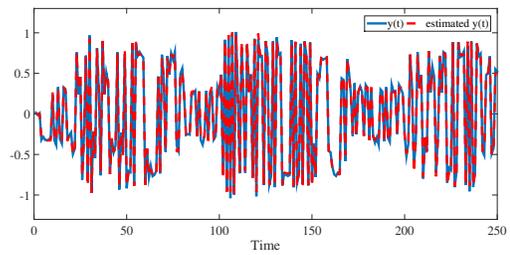


Fig. 7. True and estimated output of the switched system in Example 2

- [12] X. Jin and B. Huang, "Robust identification of piecewise/switching autoregressive exogenous process," *AIChE Journal*, vol. 56, no. 7, pp. 1829–1844, 2010.
- [13] —, "Identification of switched markov autoregressive exogenous systems with hidden switching state," *Automatica*, vol. 48, no. 2, pp. 436–441, 2012.
- [14] A. Bemporad, A. Garulli, S. Paoletti, and A. Vicino, "A bounded-error approach to piecewise affine system identification," *Automatic Control, IEEE Transactions on*, vol. 50, no. 10, pp. 1567–1580, 2005.
- [15] F. Lauer, G. Bloch, and R. Vidal, "A continuous optimization framework for hybrid system identification," *Automatica*, vol. 47, no. 3, pp. 608 – 613, 2011.
- [16] T. P. Dinh, H. M. Le, H. A. L. Thi, and F. Lauer, "A difference of convex functions algorithm for switched linear regression," *Automatic Control, IEEE Transactions on*, vol. 59, no. 8, pp. 2277–2282, 2014.
- [17] L. Bako, K. Boukharouba, E. Duviella, and S. Lecoeuche, "A recursive identification algorithm for switched linear/affine models," *Nonlinear Analysis: Hybrid Systems*, vol. 5, no. 2, pp. 242 – 253, 2011.
- [18] 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.
- [19] L. Bako, "Identification of switched linear systems via sparse optimization," *Automatica*, vol. 47, no. 4, pp. 668 – 677, 2011.
- [20] M. Gorji Sefidmazgi, M. Moradi Kordmahalleh, A. Homaifar, and A. Karimodini, "A finite element based method for identification of switched linear systems," in *American Control Conference (ACC)*, 2014, pp. 2644–2649.
- [21] H. Ohlsson and L. Ljung, "Identification of switched linear regression models using sum-of-norms regularization," *Automatica*, vol. 49, no. 4, pp. 1045 – 1050, 2013.
- [22] D. Piga and R. Tóth, "An SDP approach for  $\ell_0$ -minimization: Application to ARX model segmentation," *Automatica*, vol. 49, no. 12, pp. 3646 – 3653, 2013.
- [23] H. Ohlsson, L. Ljung, and S. Boyd, "Segmentation of ARX-models using sum-of-norms regularization," *Automatica*, vol. 46, no. 6, pp. 1107 – 1111, 2010.
- [24] P. Metzner, L. Putzig, and I. Horenko, "Analysis of persistent non-stationary time series and applications," *Communications in Applied Mathematics and Computational Science*, vol. 7, no. 2, pp. 175–229, 2012.
- [25] B. Rynne and M. Youngson, *Linear Functional Analysis*. Springer, 2007.
- [26] M. Gorji Sefidmazgi, M. Sayemuzzaman, and A. Homaifar, "Non-stationary time series clustering with application to climate systems," in *Advance Trends in Soft Computing*, ser. Studies in Fuzziness and Soft Computing, M. Jamshidi, V. Kreinovich, and J. Kacprzyk, Eds. Springer International Publishing, 2014, vol. 312, pp. 55–63.
- [27] I. Horenko, "On the identification of nonstationary factor models and their application to atmospheric data analysis," *Journal of the Atmospheric Sciences*, vol. 67, no. 5, pp. 1559–1574, 2010.
- [28] D. P. Bertsekas and D. P. Bertsekas, *Nonlinear Programming*. Athena Scientific, 1999.
- [29] I. Gurobi Optimization, "Gurobi optimizer reference manual," 2015. [Online]. Available: <http://www.gurobi.com>
- [30] V. Jandhyala, S. Fotopoulos, I. MacNeill, and P. Liu, "Inference for single and multiple change-points in time series," *Journal of Time Series Analysis*, 2013.
- [31] K. Burnham and D. Anderson, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer, 2002.