2014 American Control Conference (ACC)
June 4-6, 2014. Portland, Oregon, USA

A Finite Element Based Method for Identification of Switched Linear Systems*

Mohammad Gorji Sefidmazgi, Mina Moradi Kordmahalleh, Abdollah Homaifar, and Ali Karimoddini

Abstract— **Non-stationary time series analysis is important in the study of complex systems. Finding mathematical models for such complex systems with transitions between different phases is an ill-posed problem. This paper brings the problem of time series analysis into the context of hybrid modeling. Approximating the hybrid system by a switched linear system, the problem is reduced to identifying the switching times and model parameters. To address this problem, the non-stationary time series clustering technique based on Finite Elements is used for modeling of switched linear systems. The advantage of this method is that it is not necessary to add restrictive statistical assumptions on system variables. Illustrative examples have been provided to verify the proposed algorithm.**

I. INTRODUCTION

Time series analysis plays a crucial role in many areas of science and engineering. Indeed, time series are sets of observations measured from variables of interest. In many complex systems such as climate, economy, and biological systems, time series are characterized as non-stationary. In these cases, the statistical model of time series changes during time. Developing mathematical models for time series that undergo transitions between different segments is a pattern recognition problem. Usually this problem is solved by including statistical assumptions on variables in time series, such as methods based on Hidden Markov or Gaussian Mixture Model [1, 2].

A similar problem exists in the context of hybrid modeling and control theory [3] in which having a rich set of identification input/output data, different operation modes of the system should be identified. Hybrid systems are mixtures of time-driven and event-driven dynamics, which simultaneously and interactively coexist in the system. For example, the trajectory of a bouncing ball resulted from the switching between free fall and elastic contact could be modeled as a hybrid system, where elastic contacts are the events that change the continuous dynamics of the system. The point is that, unlike the bouncing ball, physical modeling is not possible in most practical problems [4-6]. For example, we may assume pressure of a boiler–turbine unit as an output of a hybrid system whose switches takes places when operating point of the turbine changes [7]. In this case, the problem is how to identify the parameters of the complex nonlinear dynamics, which is described by time series and can be modeled by hybrid systems [8-14]. Approximating the system by a *switched linear system*, the problem is reduced to identifying the switching times and model parameters. The time-driven dynamics of the system in each discrete mode is called a *sub-model*. The dynamical models of sub-models can be described in the form of autoregressive exogenous (ARX) or state space equations. A moving window is utilized in [15] to select batch of data for subspace identification and then clustered the data to find sub-models. Recursive methods based on least mean square [16] and gradient descent [17] are also proposed in the literature. In [18], it is assumed that data belong to Gaussian distribution and switching between sub-models follows a Markov process. The L_1-norm is used in [19] for regularization in order to penalize model parameter changes over time.

Time series clustering based on Finite Elements Method (FEM) is an effective method for analysis of complex systems [2]. This method was used for analyzing time-varying linear trend [20, 21] and parameterization of climate time series using multivariate autoregressive factor (VARX) in climate systems [22]. This method integrates L2-norm regularization (Tikhonov) for smoothing in clustering problem to setup a simpler optimization problem and, finally uses the FEM for mapping the problem to a lower dimensional space such that it can be solved numerically. The major difference with methods mentioned above is that FEM-clustering processes the data in batch mode and identifying the switch times and sub-model parameters take place at the same time. Restrictive assumptions on the data to be Gaussian or the switches to be Markovian are not necessary. We propose a method to identify a switched linear system using FEM-clustering approach in order to find the switching times and parameters of piecewise affine dynamics of a hybrid system. In contrast with time series analysis, a rich set of input data should be applied to the system to excite different operating modes for identification of the system dynamics. Here, the identification data can be presented by time series in the form of input/output pairs.

This paper is organized as follows. Section II formulates the problem mathematically. Section III explains the proposed clustering technique based on Finite Element Method and derives necessary mathematics. In section IV, the method is applied to two switched linear systems, one that presented in [19], and the results are discussed. Main conclusions of this work are shown in section V.

*Resrach supported by NSF Foundation.
M. Gorji Sefidmazgi, M. Moradi Kordmahalleh, A. Homaifar and A. Karimoddini are with the Autonomous Control and Information Technology center, Electrical Engineering Department, North Carolina A&T State University, Greensboro, NC 27410 USA.
mgorjise@aggies.ncat.edu, mmoradik@aggies.ncat.edu, akarimod@ncat.edu
Corresponding Author: A. Homaifar (homaifar@ncat.edu, phone: +13362853271)

II. PROBLEM DEFINITION

We define piecewise autoregressive exogenous (PWARX) model such that the system has a different ARX model for each of S sub-models. The time driven dynamics of each *sub-model* could be represented in discrete time form (1), where y is the output and u is the input of the system (Fig. 1).

$$\begin{aligned} y_{t+1} &= [a_1^s \quad a_2^s \quad \dots \quad a_n^s][y_t \quad y_{t-1} \quad \dots \quad y_{t-n}]^T \\ &+ [b_1^s \quad b_2^s \quad \dots \quad b_m^s][u_t \quad u_{t-1} \quad \dots \quad u_{t-m}]^T \end{aligned} \tag{1}$$

where $s \in \{1,2,...,S\}$ are the indices of sub-models. Without loss of generality, assume that sampling time is 1. In (1), $\{a_j^s\}_{j=1}^n$ and $\{b_j^s\}_{j=1}^m$ are the parameters of sub-model s. Also we assume $y \in \mathbb{R}$ and $u \in \mathbb{R}^p$. The activation sequence of sub-model appeared in the output and also the coefficients of PWARX models are unknown. Given the discrete pairs of data in the form of $\{u_t, y_t\}_{t=0}^{T-1}$ generated by system (1), our aim is to estimate model parameters and switching times between S sub-models. Each cluster is the output corresponding to the response of a sub-model [16]. In this paper, we assume that all of S sub-models are stable and observable [15]. The orders n and m are finite and not necessarily equal for all of the sub-models and they are assumed to be known as prior information. The inputs and outputs are corrupted with disturbances and measurement noises in the form of white noise.

A challenging problem is to find the number of sub-models. The problem of finding a switched model such as (1) from data, has multiple solutions. For example if n and m 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. If we assign finite and fixed values to n and m, there are still 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 will assume that an upper bound $S \ll T$ on the number of sub-models exists and then we find an appropriate number of sub-models by trial and error [16].

III. FINITE ELEMENTS METHOD FOR DATA CLUSTERING

The basic idea of FEM-clustering is to assume a model for the data in each cluster and then, find the best switch times and model parameters by solving a convex optimization problem. Instead of restrictive assumptions on the statistical distribution of the data, this method considers reasonable assumptions on the persistency of clusters, which makes it possible to solve the problem. Then, using FEM, the problem can be mapped to a finite dimensional space. Finally, the minimization problem can be converted to a linear quadratic programming (LQP) and an unconstrained optimization. These two problems are solved iteratively to determine the parameters of interest [2]. In the following, we show how this method is applicable to solve switched linear system identification. To discuss the details, assume that each of S sub-models can be represented as:

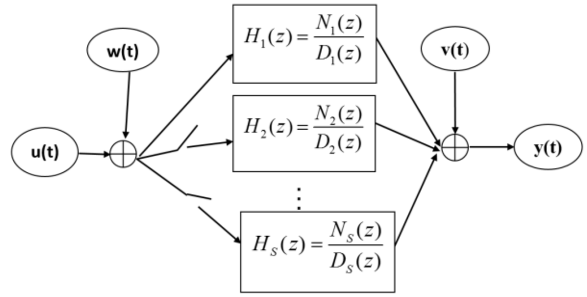

Figure 1. Schematic of a switched linear system

$$\begin{aligned} \varphi(t) = [&y(t) \quad y(t-1) \quad \cdots \quad y(t-n), \\ &u(t) \quad u(t-1) \quad \dots \quad u(t-m)] \end{aligned} \qquad t = 0,...,T-1 \tag{2}$$

$$x(t) = y(t+1) \qquad t = 0,...,T-1 \tag{3}$$

$$\theta_s = [a_1^s \quad a_2^s \quad \dots \quad a_n^s \, b_1^s \quad b_2^s \quad \dots \quad b_m^s] \qquad s = 1,...,S \tag{4}$$

Therefore, the output of each sub-model is in the form of

$$x(t) = \theta_s . \varphi(t)^T + error \qquad \begin{aligned} s &= 1,...,S \\ t &= 0,...,T-1 \end{aligned} \tag{5}$$

where θ_s are unknown, and the error is the result of noise and disturbance. Now, we define *distance function* between measured output and output of sub-model s in the form of Euclidean distance:

$$d(x(t), \theta_s) = \left\| x(t) - \theta_s . \varphi^T(t) \right\|^2 \qquad \begin{aligned} s &= 1,...,S \\ t &= 0,...,T-1 \end{aligned} \tag{6}$$

And in general for S sub-models, the cost function can be defined as:

$$\sum_{t=0}^{T-1} \sum_{s=1}^{S} \mu_s(t) . d(x(t), \theta_s) \xrightarrow[\mathrm{M}(t),\Theta]{} \min \qquad \begin{aligned} s &= 1,...,S \\ t &= 0,...,T-1 \end{aligned} \tag{7}$$

where $\Theta=[\theta_1, ..., \theta_{S^*}]$ is the set of S^* model parameters for S sub-models. Also $\mu_s(t)$ is the sub-model membership function for sub-model s and $M(t)=[\mu_1(t), ..., \mu_S(t)]$. The following constrains can be defined for $\mu_s(t)$

$$\mu_s(t) \in \{0,1\} \tag{8}$$

$$\sum_{s=1}^{S} \mu_s(t) = 1 \qquad \begin{aligned} s &= 1,...,S \\ t &= 0,...,T-1 \end{aligned} \tag{9}$$

Finding the optimal solution for this problem requires searching for all possible discrete state paths, which is time-consuming. In general, solving the optimization problem in (7) is an ill-posed problem as shown in [16]. Thus, the problem requires additional assumptions to be soluble which is known as *Regularization* [23]. In FEM-clustering method, it is assumed that the sub-model membership functions $\mu_s(t)$ are smooth and their derivatives are bounded. Since $\mu_s(t)$ are defined for all times in $[0,T]$, hence they belong to infinite-dimensional *Hilbert Space*. The smoothness assumption in Hilbert Space is defined as:

$$\left\| \frac{\partial \mu_s(t)}{\partial t} \right\|_{\mathbf{H}(0,T)} = \int_0^T \left(\frac{\partial \mu_s(t)}{\partial t}\right)^2 dt < +\infty \tag{10}$$

where $H(0,T)$ is the Hilbert space defined over the time interval $[0,T]$. Equation (10) is derived based on the definition of norm in Hilbert space [23]. For given observations, this constraint limits the total number of switches between the sub-models and add *persistency* to $\mu_s(t)$[2]. The new optimization problem using *Tikhonov Regularization* can be defined as (11).

$$\sum_{s=1}^{S}\sum_{t=0}^{T-1}\left[\mu_s(t).d(x(t),\theta_s)+\delta\left(\frac{\partial\mu_s(t)}{\partial t}\right)^2\right]\xrightarrow[M(t),\Theta]{}\min \qquad (11)$$

Here $\delta>0$ is the *Regularization Factor* which controls the persistency of the clusters. The optimization problem in (11) is not soluble numerically, since the $\mu_s(t)$ belong to infinite dimensional function space. Therefore, it is necessary to map the problem into a lower dimensional space such that solution can be found numerically. *Galerkin* method can convert a continuous operator problem (such as a differential equation) to a discrete problem. It is widely used in *Finite Element Method (FEM)* literature for solving differential equations [24]. The unknown function (such as the solution of an ODE) in Hilbert Space can be mapped to a lower dimensional space which is spanned by *Finite Element Basis Functions*. One choice of FEM basis function is in the form of N triangular functions, which are called *hat functions* and are common in the FEM literature. We use these continuous basis functions which have local support on $[0,T]$ in our problem, as in (12) and Fig. 2 [2].

$$\alpha_k(t)=\begin{cases}\dfrac{t_2-t}{\Delta} & k=1, t\in[t_1,t_2]\\ \dfrac{t-t_{k-1}}{\Delta} & 2\le k\le N-1, t\in[t_{k-1},t_k]\\ \dfrac{t_{k+1}-t}{\Delta} & 2\le k\le N-1, t\in[t_k,t_{k+1}]\\ \dfrac{t-t_{N-1}}{\Delta}, & k=N, t\in[t_{N-1},t_N]\end{cases} \qquad (12)$$

Applying discretization to $M(t)$ yields (13) and (14)

$$\mu_s(t)=\tilde{\mu}_s(t)+error=\frac{1}{\Delta}\sum_{k=1}^{N}\tilde{\mu}_s^{(k)}.\alpha_k(t)+error \qquad (13)$$

$$\tilde{\mu}_s^{(k)}=\sum_{t=0}^{T}\mu_s(t).\alpha_k(t) \qquad (14)$$

where $\tilde{\mu}_s^{(k)}$ are scalars called *Galerkin coefficients*. Now, the problem of finding $\mu_s(t)$ can be transformed to the estimation of Galerkin coefficients. y substituting (13) and (14) in (11) and using the compactness of finite elements basis function, one can find an optimization problem in the form of linear quadratic programming problem as in (15) after some mathematical simplifications

$$\sum_{s=1}^{S}[\beta(\theta_s)^T\bar{\mu}_s+\delta\bar{\mu}_s^{\,T}H\bar{\mu}_s]\xrightarrow[\bar{\mu},\Theta]{}\min \qquad (15)$$

$$\bar{\mu}_s=[\tilde{\mu}_s^{(1)},...\tilde{\mu}_s^{(k)},...,\tilde{\mu}_s^{(N)}] \qquad (16)$$

$$\beta(\theta_s)=[\sum_{t_1}^{t_2}\alpha_1(t).d(x(t),\theta_s)\quad ...\quad \sum_{t_{k-1}}^{t_{k+1}}\alpha_k(t).d(x(t),\theta_s)...\quad ...\sum_{t_{N-1}}^{t_N}\alpha_N(t).d(x(t),\theta_s)] \qquad (17)$$

where H is a tri-diagonal matrix called *stiffness matrix* [24]. The constraints (8) and (9) on sub-model membership functions can be converted to a set of constrains on the Galerkin coefficients as in (19) and (20).

$$\tilde{\mu}_s^{(k)}\in\{0,1\} \qquad \forall k=1,...,N \qquad (19)$$

$$\sum_{s=1}^{S}\tilde{\mu}_s^{(k)}=1 \qquad s=1,...,S \qquad (20)$$

Now the optimization in (15) can be iteratively solved with respect to two sets of unknown parameters ($\bar{\mu}$ and θ) by solving one and substituting in the other one [2].

The optimization problem with respect to $\bar{\mu}_s$ is in the form of summation of S linear binary quadratic programming (LQP) optimizations. We augment $\bar{\mu}_s$ in a vector form as λ and convert the problem into single LQP as follows:

$$\frac{1}{2}\lambda^T G\lambda+B^T\lambda\xrightarrow[\lambda]{}\min \qquad (21)$$

$$\lambda=[\bar{\mu}_1,...,\bar{\mu}_s,...\bar{\mu}_S]^T \qquad (22)$$

$$B=[\beta(\theta_1),...,\beta(\theta_s),...\beta(\theta_S)] \qquad (23)$$

$$G=2\delta\begin{bmatrix}H & 0 & \dots & 0\\ 0 & H & \ddots & \vdots\\ \vdots & \ddots & \ddots & 0\\ 0 & \cdots & 0 & H\end{bmatrix} \qquad (24)$$

As a result, for the new optimization problem of Eq. (21),

$$F\lambda=Q \qquad (25)$$

$$F=[\underbrace{I_{N\times N}\quad I_{N\times N}\quad ...\quad I_{N\times N}}_{S\ \text{times}}] \qquad (26)$$

$$Q=\begin{matrix}1 & 1 & \dots & 1\end{matrix}^T_{N\times 1} \qquad (27)$$

$$\lambda_j\ge 0 \qquad j=1,...,N\times S \qquad (28)$$

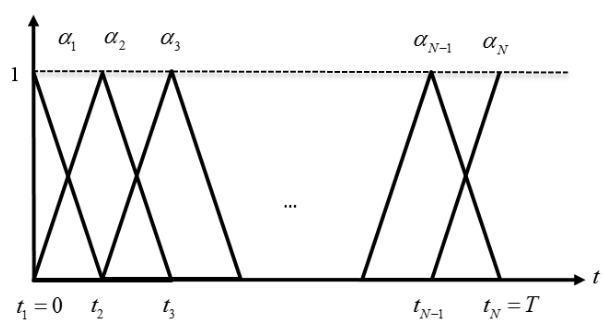

Figure 2.Finite Element Basis Functions

$$H=\begin{bmatrix} \int_{t_1}^{t_2}(\frac{\partial\alpha_1}{\partial t})^2\,dt & \int_{t_1}^{t_2}(\frac{\partial\alpha_1}{\partial t})(\frac{\partial\alpha_2}{\partial t})dt & 0 & \cdots & 0 & 0 \\ \int_{t_1}^{t_2}(\frac{\partial\alpha_1}{\partial t})(\frac{\partial\alpha_2}{\partial t})dt & \int_{t_1}^{t_3}(\frac{\partial\alpha_2}{\partial t})^2\,dt & \ddots & 0 & 0 & 0 \\ 0 & \ddots & \ddots & \int_{t_{k-1}}^{t_k}(\frac{\partial\alpha_{k-1}}{\partial t})(\frac{\partial\alpha_k}{\partial t})dt & 0 & 0 \\ 0 & 0 & \int_{t_{k-1}}^{t_k}(\frac{\partial\alpha_{k-1}}{\partial t})(\frac{\partial\alpha_k}{\partial t})dt & \int_{t_{k-1}}^{t_{k+1}}(\frac{\partial\alpha_k}{\partial t})^2\,dt & \ddots & 0 \\ \vdots & 0 & 0 & \ddots & \ddots & \int_{t_{N-1}}^{t_N}(\frac{\partial\alpha_{N-1}}{\partial t})(\frac{\partial\alpha_N}{\partial t})dt \\ 0 & 0 & 0 & 0 & \int_{t_{N-1}}^{t_N}(\frac{\partial\alpha_{N-1}}{\partial t})(\frac{\partial\alpha_N}{\partial t})dt & \int_{t_{N-1}}^{t_N}(\frac{\partial\alpha_N}{\partial t})^2\,dt \end{bmatrix}=\begin{bmatrix} \frac{1}{\Delta} & \frac{-1}{\Delta} & 0 & \cdots & 0 & 0 \\ \frac{-1}{\Delta} & \frac{2}{\Delta} & \ddots & 0 & 0 & 0 \\ 0 & \ddots & \ddots & \frac{-1}{\Delta} & 0 & 0 \\ 0 & 0 & \frac{-1}{\Delta} & \frac{2}{\Delta} & \ddots & 0 \\ \vdots & 0 & 0 & \ddots & \ddots & \frac{-1}{\Delta} \\ 0 & 0 & 0 & 0 & \frac{-1}{\Delta} & \frac{1}{\Delta} \end{bmatrix} \quad (18)$$

The resulting optimization problem is a large sparse integer quadratic programming, which include constraints in the form of equalities and inequalities. In the past, different techniques such as interior point method and active set method have been applied to solve the LQP [25]. The large sparse problem in (21) with the constrains of (25)-(28) is solved by the toolbox developed in [26]. This toolbox is specially designed to solve large sparse optimization problems and has options to choose among several solvers. After finding Galerkin coefficient, we can build $\mu_s(t)$ using FEM basis function and (13).

Since $d(x,\theta)$ is continuous and differentiable, solution of (11) with respect to Θ can be found analytically. Using least mean square, we can derive (29). The proof is provided in appendix I.

$$\theta_s=\left[\sum_{t=0}^{T-1}\mu_s(t).\varphi^T(t).\varphi(t)\right]^{-1}\left[\sum_{t=0}^{T-1}\mu_s(t).x(t).\varphi^T(t)\right] \quad (29)$$

Table I shows the general procedure for solving the optimization problem. It has been proven that the optimization problem in (15) is monotone and converges to a local solution. By running the algorithm multiples times with different initial conditions, the algorithm can converge to global solution [2].

Decreasing N reduces the order of LQP and consequently the complexity of calculations. On the other hand, it decreases the accuracy of the identification process and it may result in losing some very short clusters; and consequently finding another set of sub-model parameters which are not optimal. A challenging problem arises when choosing S. As mentioned in section II, there are two trivial solutions when the cost function equals to zero. Since the number of sub-models is unknown in advance, trial and error along with human judgment is used to select S subjectively. If the sub-model membership function at time t for sub-model s is 1, it means the datum at t belongs completely to that sub-model Similarly, membership function equal to 0 means that data does not belong to sub-model s at all. Thus, we can start with an upper bound on S and run the algorithm. A criterion for choosing S is based on the sub-model membership functions having equal to one at least in some instances. It means that if a sub-model membership function for one of the sub-models is found equal to 0 over $[0,T]$, the number of sub-models should be

Table 1. GENERAL PROCEDURE OF OPTIMIZATION

1	Set S the number of sub-models, δ regularization factor, Δ width of hat functions
2	Iteration number, L=1
3	Choose random initial $\bar{\mu}_s^{initial}$ for s=1,…,S satisfying related constraints.
4	Solve the minimization problem with respect to Θ for fixed $\bar{\mu}_s^{initial}$ to find $\Theta_s^{initial}$
5	Solve the minimization problem with respect to $\bar{\mu}_s$ for a fixed Θ_s to find $\bar{\mu}_s^{[L+1]}$
6	Solve the minimization problem with respect to Θ_s for a fixed $\bar{\mu}_s$ to find $\Theta_s^{[L+1]}$
7	Go to step 5 and continue until pre-defined number of iterations

decreased by one Increasing δ leads to an increase in persistency of clusters, this means that the system tends to stay in current sub-model instead of jumping to another. If we set δ a small number near zero, then there will be many switches between sub-models in the results. However, when δ is a large number then the number of switches decreases. In this work, we assume δ is a small number such that Tikhonov regularization slightly biases the solution. Therefore, the regularization term in (11) has a small bias on the solution, but make the problem soluble. Usually the value around 0.1 is enough for our application.

I. SIMULATION

In this section, we present two examples to verify our proposed algorithm. The switching times and model parameters are chosen randomly. Example I has three switched linear system and in example II the memory of system changes.

Example I: we applied the FEM algorithm to a system including three sub-models. The system model and switch times between sub-models are:

$$y(t+1)=\begin{cases} 0.5y(t)-0.3u(t) & 0\le t<35 \text{ or } 90\le t<100 \\ -0.4y(t)+0.6u(t) & 35\le t<70 \\ -0.1y(t)+0.8u(t) & 70\le t<90 \end{cases} \quad (30)$$

An input in the form of 100 pseudo-random binary data (PRBS) with zero mean and unit variance is applied to the system. The input has an additive disturbance with $N(0, 0.05^2)$ and the output has a noise as $N(0, 0.05^2)$. In the simulation, we assume $\Delta = 1$, and $\delta=0.2$. The resulting LQP has the order equals to 303. After 6 iterations, algorithm converges to its final solution. We can run the algorithm several times with different initial conditions for $\mu_s(t)$ to reach minimum cost function. Fig. 3 and 4 show input and output used for system identification. The resulting sub-model parameters for three sub-models are found as

$$\theta_s = [a_1^s \quad b_1^s] =$$
$$[0.4795 \quad -0.2794 \ \mid \ -0.4108 \quad 0.5871 \ \mid \ -0.0965 \quad 0.7910]$$

Fig. 5 shows $\mu_s(t)$ for three sub-models and indicates that the switch times are estimated correctly. On other hand, if we set S>3 in the algorithm, we find similar $\mu_s(t)$, but the value of $\mu_s(t)$ for remaining *S-3* sub-models are zero over all *[0,T]*. Therefore, we conclude that there exist only three sub-models that generated the dataset. For testing the validity of identification, another dataset *u=sin(0.2t) is* applied to the same system. Fig. 6 shows the true and estimated output generated by identified model. The Variance-accounted-for (VAF) is defined as below where $y(t)$ and $\hat{y}(t)$ are true and simulated (by identified model) output of the system. In this case, we found VAF≈97%.

$$VAF = \max\left\{1 - \frac{\text{var}(y(t) - \hat{y}(t))}{\text{var}(y(\text{t}))}, 0\right\} \times 100 \qquad (31)$$

Example II: we applied our algorithm to model as presented in [19]. In this example the time lag of sub-models are different. The input u is ±1 PRBS signal and the additive noise has variance 0.1.

$$y(t+1) = \begin{cases} -0.9y(t) + u(t-1) & 0 \le t < 20 \\ -0.9y(t) + u(t) & 20 \le t < 40 \end{cases} \qquad (32)$$

We set Δ=1 and ε=0.1. The estimated parameters are as

$$\theta_i = [a_1^s \quad b_1^s \quad b_2^s] =$$
$$[-0.8927 \quad 0.0109 \quad 09809 \ \mid \ -0.9299 \quad 1.0019 \quad 0.0847]$$

Which are near the true values [-0.9 , 0 ,1 | -0.9 , 1, 0]. Fig. 8 shows the cluster affiliation functions. Comparing with results in [19], using the same size of data, both methods correctly estimated the switch times. In addition, the coefficients estimated by the FEM-clustering are closer to the actual values. For example, the first coefficients of each sub model are -0.8927 and -0.9299 in FEM, which are close to -0.9, but in [19], the results are further away from -0.9.

Appendix I

Proof of (29). We want to solve (11) analytically with respect to θ. The second term in (11) is not a function of θ and hence can be eliminated; and the problem will be simplified as follows:

$$L = \sum_{s=1}^{S}\sum_{t=0}^{T-1} \mu_s(t).d(x(t), \theta_s) \quad ; s = 1, \ldots, S \qquad (33)$$

Expanding the above equation with respect to s and convert the integral to summation

$$\sum_{t=0}^{T-1} \mu_1(t)[x(t) - \theta_1.\varphi^T(t)]^2 + \ldots + \mu_S(t)[x(t) - \theta_S.\varphi^T(t)]^2 \qquad (34)$$

For each cluster, a separate optimization can be written

$$\Rightarrow \frac{dL}{d\theta_s} = -2\sum_{t=0}^{T-1} \mu_s(t)[x(t) - \theta_s.\varphi^T(t)].\varphi^T(t) = 0$$
$$\Rightarrow \sum_{t=0}^{T-1}[\mu_s(t).x(t).\varphi^T(t) - \mu_s(t).(\theta_s.\varphi^T(t)).\varphi^T(t)] = 0 \qquad (35)$$

After some mathematical simplifications, we derive:

$$\Rightarrow \theta_s = \left[\sum_{t=0}^{T-1} \mu_s(t).\varphi^T(t).\varphi(t)\right]^{-1} \left[\sum_{t=0}^{T-1} \mu_s(t).x(t).\varphi^T(t)\right] \qquad (36)$$

If there is only one sub-model, the values of sub-model membership function is equal to 1 and above equation is converted to (37) which is widely used in system identification literature [27].

$$\Rightarrow \theta = \left[\sum_{t=0}^{T-1} \varphi^T(t).\varphi(t)\right]^{-1} \left[\sum_{t=0}^{T-1} x(t).\varphi^T(t)\right] \qquad (37)$$

II. Conclusion

In this paper, we applied the FEM-clustering algorithm to identify the system parameters of a hybrid switched linear system. Unlike other methods, no statistical assumption on the system model is required and switching times and model parameters are determined at the same time iteratively. The FEM-clustering converted the ill-posed and non-convex optimization problem into two sets of optimizations. One is in the form of binary linear quadratic programming; and for the other one, we derived an analytical solution using least mean square. The identification of the sub model parameters is accurate and the performance of the algorithm is robust by testing different sets of data. In the future, we will integrate the method with estimating sub-model orders and propose criteria for selecting number of sub-models; and expand our approach for handling general class of hybrid systems.

Acknowledgment

This work is partially supported by the Expeditions in Computing by the National Science Foundation under Award Number: CCF-1029731 and No. DBI-0939454

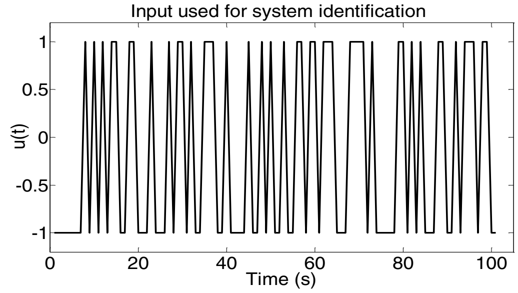

Figure 3: inputs applied to the system in example I

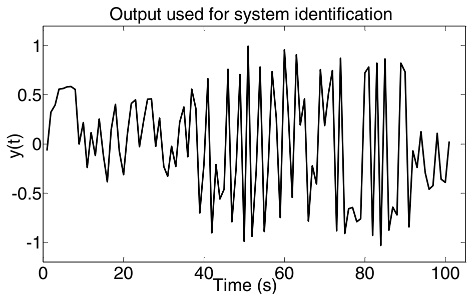

Figure 4: output used for identification in example I

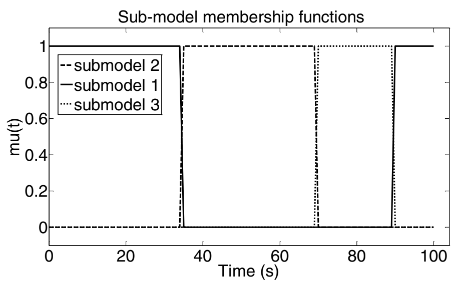

Figure 5: sub-model membership function for example I

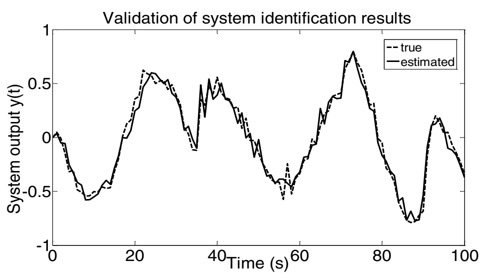

Figure 6. True and estimated output in example I

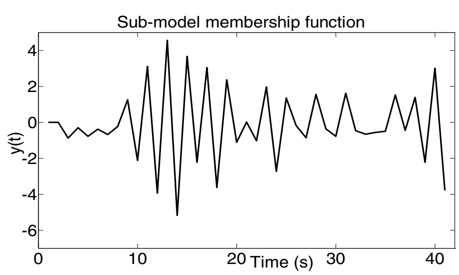

Figure 7: Output used for identification in example II

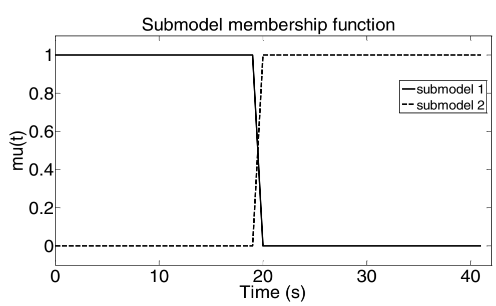

Figure 8. Estimated sub-model membership function in example II

REFERENCES

[1] C. Chatfield, The analysis of time series: an introduction: CRC press, 2003.

[2] I. Horenko, "Finite element approach to clustering of multidimensional time series," SIAM Journal on Scientific Computing, vol. 32, pp. 62-83, 2010.

[3] P. J. Antsaklis, "A brief introduction to the theory and applications of hybrid systems," in Proc IEEE, Special Issue on Hybrid Systems: Theory and Applications, 2000.

[4] D. Liberzon, Switching in systems and control: Springer, 2003.

[5] Z. Sun and S. S. Ge, Switched linear systems: Control and design: Springer, 2005.

[6] A. Van der Schaft and H. Schumacher, An introduction to hybrid dynamical systems: Springer, 1999.

[7] M. Keshavarz, M. Barkhordari Yazdi, and M. Jahed-Motlagh, "Piecewise affine modeling and control of a boiler–turbine unit," Applied Thermal Engineering, vol. 30, pp. 781-791, 2010.

[8] M. Morari, M. Baotic, and F. Borrelli, "Hybrid systems modeling and control," European Journal of Control, vol. 9, pp. 177-189, 2003.

[9] J. Roll, A. Bemporad, and L. Ljung, "Identification of piecewise affine systems via mixed-integer programming," Automatica, vol. 40, pp. 37-50, 2004.

[10] G. Ferrari-Trecate, M. Muselli, D. Liberati, and M. Morari, "A clustering technique for the identification of piecewise affine systems," Automatica, vol. 39, pp. 205-217, 2003.

[11] H. Nakada, K. Takaba, and T. Katayama, "Identification of piecewise affine systems based on statistical clustering technique," Automatica, vol. 41, pp. 905-913, 2005.

[12] A. L. Juloski, S. Weiland, and W. Heemels, "A Bayesian approach to identification of hybrid systems," Automatic Control, IEEE Transactions on, vol. 50, pp. 1520-1533, 2005.

[13] A. Bemporad, A. Garulli, S. Paoletti, and A. Vicino, "A bounded-error approach to piecewise affine system identification," Automatic Control, IEEE Transaction., vol. 50, pp. 1567-1580, 2005.

[14] C. Y. Lai, C. Xiang, and T. H. Lee, "Data-based identification and control of nonlinear systems via piecewise affine approximation," Neural Networks, IEEE Transactions on, vol. 22, pp. 2189-2200, 2011.

[15] R. V. Lopes, G. A. Borges, and J. Y. Ishihara, "New algorithm for identification of discrete-time switched linear systems," in American Control Conference (ACC), 2013, 2013, pp. 6219-6224.

[16] L. Bako, K. Boukharouba, E. Duviella, "A recursive identification algorithm for switched linear/affine models," Nonlinear Analysis: Hybrid Systems, vol. 5, pp. 242-253, 2011.

[17] R. Vidal, "Recursive identification of switched ARX systems," Automatica, vol. 44, pp. 2274-2287, 2008.

[18] X. Jin and B. Huang, "Identification of switched Markov autoregressive eXogenous systems with hidden switching state," Automatica, vol. 48, pp. 436-441, 2012.

[19] H. Ohlsson, L. Ljung, and S. Boyd, "Segmentation of ARX-models using sum-of-norms regularization," Automatica, vol. 46, pp. 1107-1111, 2010.

[20] M. Gorji, A. Homaifar, and M. Sayemuzzaman, "Non-stationary time series clustering with application to climatic systems", Third annual world conference on soft computing, San Antonio, 2013.

[21] I. Horenko, "On clustering of non-stationary meteorological time series," Dynamics of Atmospheres and Oceans, vol. 49, pp. 164-187, 2010.

[22] I. Horenko, "On the Identification of Nonstationary Factor Models and Their Application to Atmospheric Data Analysis," Journal of the Atmospheric Sciences, vol. 67, 2010.

[23] A. S. Poznâk, Advanced Mathematical Tools for Automatic Control Engineers: deterministic techniques. Vol. 1: Elsevier Science Limited, 2008.

[24] O. C. Zienkiewicz, R. L. Taylor, and J. Z. Zhu, The Finite Element Method: Its Basis and Fundamentals: Its Basis and Fundamentals: Elsevier Science, 2013.

[25] S. P. Boyd and L. Vandenberghe, Convex Optimization: Cambridge University Press, 2004.

[26] Gurobi, "Gurobi optimizer reference manual," 2014.

[27] M. Verhaegen and V. Verdult, Filtering and System Identification: A Least Squares Approach: Cambridge University Press.