Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

An alternating optimization method for switched linear systems identification

Federico Bianchi * **Alessandro Falsone** * **Luigi Piroddi** *
Maria Prandini *

* *Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Via G. Ponzio, 34/5, 20133, Milano, Italy. (e-mail: name.surname@polimi.it)*

Abstract: The identification of switched systems involves solving a mixed-integer optimization problem to determine the parameters of each mode dynamics (continuous part) and assign the data samples to the modes (discrete part), so as to minimize a cost criterion measuring the quality of the model on a set of input/output data collected from the system. Oftentimes, some *a priori* information on the switching mechanism is available, *e.g.*, in the form of a minimum dwell time. This information can be encoded in a suitable constraint and included in the optimization problem, but this introduces a coupling between the discrete and continuous optimization variables that makes the problem harder to solve. In this paper, we propose an iterative approach to the identification of switched systems that alternates a minimization step with respect to the continuous parameters of the modes, and a minimization step with respect to the discrete variables defining the sample-mode mapping. Constraints originating from prior knowledge on the switching mechanism are taken into account after the (unconstrained) discrete optimization step through a post-processing phase. These three phases are repeated until a stopping criterion is met. A comparative numerical analysis of the proposed method shows its improved performance with respect to competitive approaches in the literature.

Keywords: Model identification, Switched systems, Hybrid systems.

1. INTRODUCTION

Data collected from a dynamical system can be characterized by the presence of some heterogeneity that is hard to match with a single model and can be better described by using a set of models (modes) and a switching law among them. In this work, we consider switched systems where the continuous dynamics of the modes are described by a set of linear regression models, defining the relationship between the regression vector $\boldsymbol{x} \in \mathbb{R}^{n_x}$ and the output $y \in \mathbb{R}$. As for the switching mechanism governing the mode activation, it can be characterized as an exogenous signal or alternatively be determined by a polyhedral partition of the regressor domain mapping each regressor vector to a mode. We refer to these cases as Switched Affine (SA) and PieceWise Affine (PWA) systems, respectively. In this paper, we investigate in particular the identification of switched systems expressed in input-output form, namely Switched AutoRegressive with eXogenous inputs (SARX) and PieceWise ARX (PWARX) systems.

The identification of these classes of systems requires to jointly classify data (assigning each sample to a mode) and estimate the model parameters for each mode. This can be recast as an optimization problem with continuous and discrete decision variables, which is hard to solve in practice, see, *e.g.*, (Garulli et al., 2012; Lauer and Bloch, 2019) for a comprehensive review. Solution methods for the resulting mixed integer optimization program can be classified into two categories: methods that carry out the optimization with respect to continuous and discrete variables simultaneously, (Roll et al., 2004; Bemporad et al., 2005; Ma and Vidal, 2005; Bako et al., 2011; Lauer et al., 2011; Piga and Tóth, 2013; Ohlsson and Ljung, 2013; Breschi et al., 2016; Bemporad et al., 2018; Bianchi et al., 2020), and methods that address separately data classification and model parameters estimation, (Ferrari-Trecate et al., 2003; Juloski et al., 2005; Hartmann et al., 2015; Pillonetto, 2016). We next briefly review those methods for switched linear systems identification that are most relevant to our work and, indeed, represent competitors to our approach.

In (Bako et al., 2011), a two-step method is proposed that alternates between discrete state estimation and parameter update via recursive least squares. The clustering criterion used in (Bako et al., 2011), which depends on both the prediction errors of the local models and the regressor-centroid distances, has been further refined in (Breschi et al., 2016), where the authors also proposed to solve the recursive least squares problems via a QR factorization algorithm. Both methods are computationally efficient and accurate, provided that an adequate initialization is used (*e.g.*, in case of batch model identification, a good initialization can be be obtained running the method on a subset of the training data set). However, both approaches classify data samples sequentially, and once a sample is assigned to a mode, then, this assignment is not changed anymore unless the algorithm is restarted.

The RANdom SAmpling Consensus (RANSAC) approach in (Fischler and Bolles, 1981) is a parameter estimation

method designed to cope with the presence of outliers in the observed data. Starting from a random subset of the data set, the parameters are estimated. Then, other data points are tested for compliance with the model and, in case the test is positive, they are added to the subset. Samples that are not assigned are considered as outliers. The procedure is repeated starting from different random subsets, the run resulting in the smallest outlier set is selected and the corresponding model is returned. The k-RANSAC algorithm proposed in (Hartmann et al., 2015) extends this approach to the identification of multiple models in the data set. The k-RANSAC starts by running the RANSAC algorithm so as to identify a first model and the corresponding outlier set. Then, data points assigned to the first model are discarded and the RANSAC algorithm is applied only to the outlier set. This second run will return a second model and a second outlier set, and the procedure is repeated. The rationale is that samples generated by different modes will be considered as outliers when tested for compliance with the wrong mode. Unfortunately, this method is strongly affected by the threshold employed for testing compliance and, similarly to the previous ones, by the incremental nature of the procedure, which never questions the assignment of a sample to a mode again once it has been decided. Moreover, the greedy approach adopted by the k-RANSAC is well known to be sub-optimal.

The general identification framework recently proposed in (Bemporad et al., 2018) encompasses also the SA and PWA model classes addressed in this paper. It adopts a coordinate descent algorithm that alternates between the estimation of the model parameters and the sample assignment, until convergence. Note that the sample assignment to the modes is refined at each iteration, as opposed to the previous methods. The sample assignment phase amounts to a multi-stage discrete optimization problem, whose decision variables represent the switching signal, which is solved via dynamic programming (DP). However, DP is known to be computationally demanding when the number of decision variables is large. Besides, it is not viable – unless by introducing additional decision variables and thereby falling into the previous issue – in the presence of hard constraints involving decision variables associated with different stages, such as dwell time constraints. Less restrictive time constraints, as, *e.g.*, setting a (low) switching probability, can be accounted for in the cost function, through some suitably chosen term penalizing the mode switchings.

This paper presents a novel iterative method for the batch identification of SARX and PWARX models, which, unlike (Bemporad et al., 2018), is able to account for constraints on the switching mechanism. The method operates by assuming an initial sample-mode mapping (the available data set is initially partitioned at random into subsets, each assigned to a different mode), and iteratively correcting it to improve the switched model performance. At each iteration, three tasks are performed, namely a *parameter estimation* on the partitioned data, a *sample assignment* to the modes, and *post-processing* of the sample assignment. In the first phase, the estimation of the parameters of each mode is carried out with a least squares approach on the samples assigned to that mode, according to the current sample-mode mapping. Then, the one-step-ahead prediction error is calculated for each identified mode on the entire data set, and each sample is provisionally assigned to the mode which exhibits the lowest prediction error. Finally, in the last phase, misclassified samples are corrected by taking into account the constraints on the switching mechanism. This iterative process is carried out until a stopping criterion is met.

Numerical results show that the post-processing introduced in the last phase leads to superior performance with respect to competitor algorithms.

The paper is organized as follows. Section 2 introduces the SARX/PWARX identification problem which is then tackled with the method proposed in Section 3. Section 4 presents the results obtained by applying the proposed method on some numerical examples and compares them with alternative approaches. Finally, some concluding remarks and possible work extensions end the paper.

2. PROBLEM STATEMENT

Consider the following Single-Input Single-Output (SISO) switched linear dynamical system

$$y(t) = \boldsymbol{\vartheta}_{\lambda_t}^\top \boldsymbol{x}(t) + e(t), \tag{1}$$

where

$$\boldsymbol{x}(t) = [1, y(t-1), \dots, y(t-n_y), u(t-1), \dots, u(t-n_u)],$$

is the regression vector (containing also the constant 1, to account for different offsets of the modes), $y(t) \in \mathbb{R}$ is the output, $u(t) \in \mathbb{R}$ is the input, $\lambda_t \in \mathcal{M} = \{1, \dots, m\}$ represents the active mode at time $t \in \mathbb{N}$, $\boldsymbol{\vartheta}_i \in \mathbb{R}^{n_x}$ is the parameter vector defining the i-th mode, $i \in \mathcal{M}$, and $e(t)$ is an additive white noise process. Integers n_y and n_u denote the system orders and $n_x = n_y + n_u + 1$. If λ_t is an exogenous signal, then (1) is referred to as a SARX system. Instead, if the value of λ_t encodes $\boldsymbol{x}(t)$ belonging to the i-th set $\mathcal{X}_i$ of a given polyhedral partition $\{\mathcal{X}_i\}_{i \in \mathcal{M}}$ of $\mathbb{R}^{n_x}$, *i.e.*,

$$\lambda_t = i \iff \boldsymbol{x}(t) \in \mathcal{X}_i, \tag{2}$$

then (1) is a PWARX model.

In the rest of the paper we consider the following assumption, which is fairly common in the literature.

Assumption 1. The number of modes m and the orders n_y and n_u are known. □

Under Assumption 1, the identification problem for the switched system in (1) can be formalized as follows.

Given a data set of N time-ordered and consecutive regressor-output samples $\mathcal{D} = \{y(t), \boldsymbol{x}(t)\}_{t=1}^N$, find the values of $\boldsymbol{\vartheta}_i$, $i \in \mathcal{M}$, and $\lambda_t \in \mathcal{M}$, $t = 1, \dots, N$, which solve the following optimization problem

$$\begin{aligned} \min_{\substack{\{\boldsymbol{\vartheta}_i\}_{i\in\mathcal{M}},\\ \{\lambda_t\}_{t=1}^N}} \quad & \sum_{i\in\mathcal{M}} \sum_{t:\lambda_t=i} \Big(y(t) - \boldsymbol{\vartheta}_i^\top \boldsymbol{x}(t)\Big)^2 \\ \text{subject to:} \quad & \boldsymbol{\vartheta}_i \in \mathbb{R}^{n_x},\ i \in \mathcal{M} \\ & \lambda_t \in \mathcal{M},\ t = 1, \dots, N \\ & (\lambda_1, \dots, \lambda_N) \in \Lambda, \end{aligned} \tag{3}$$

where Λ encodes additional constraints on the sample assignment variables $\lambda_1, \dots, \lambda_N$, which are specific for SARX/PWARX models.

For a fixed mode i, the inner sum of the cost function in (3) measures the fitness (in terms of the overall Squared Prediction Error (SPE)) of $\boldsymbol{\vartheta}_i$ for all the samples assigned to mode i by λ_t. The outer summation then adds the SPE of all modes to get an overall identification error of the switched model. Problem (3) thus aims at finding the partition of the data set $\mathcal{D}$ among modes, that minimizes the sum of the SPEs over all the modes, compatibly with the assignment constraints in Λ.

Concerning the latter, for SARX models Λ may encode dwell time constraints like, *e.g.*,

$$\Lambda = \Big\{(\lambda_1, \dots, \lambda_N) \in \mathcal{M}^N : \lambda_{t+1} \neq \lambda_t \implies \lambda_{t+1} = \cdots = \lambda_{\min\{t+D,N\}},\ t = 1, \dots, N\Big\},$$

for a minimum dwell time of D time instants. In the case of a PWARX, Λ should be chosen so that the sets of data samples $\{(y(t), \boldsymbol{x}(t)) : \lambda_t = i\}$, $i \in \mathcal{M}$, are pairwise linearly separable, since the switching mechanism originates from a polyhedral partition $\{\mathcal{X}_i\}_{i\in\mathcal{M}}$ of the regression vector space. Given that the optimal solution to (3) returns a mapping of the samples into the modes, finding the polyhedral partition $\{\mathcal{X}_i\}_{i\in\mathcal{M}}$ becomes a linear supervised classification problem with m classes. The interested reader is referred to, *e.g.*, (Paoletti, 2004), for a comprehensive review on two-class and multi-class classification methods.

Note that problem (3) is a mixed-integer optimization program as it contains both continuous ($\boldsymbol{\vartheta}_i \in \mathbb{R}^{n_x}$, $i \in \mathcal{M}$) and discrete ($\lambda_t \in \mathcal{M}$, $t = 1, \dots, N$) variables, and is therefore hard to solve in practice. In the next section we propose an iterative algorithm to approximately solve it.

3. THE PROPOSED METHOD

We start by observing that problem (3) has two sets of decision variables: the model parameters $\boldsymbol{\vartheta}_i$, $i \in \mathcal{M}$, and the assignment variables λ_t. We then propose to alternatively minimize the cost function with respect to $\boldsymbol{\vartheta}_i$ with λ_t fixed (*parameter estimation phase*), then with respect to λ_t while keeping $\boldsymbol{\vartheta}_i$ fixed (*sample assignment phase*), and finally we seek to enforce the Λ constraints (*post-processing phase*). The overall procedure is summarized in Algorithm 1 and discussed next.

The algorithm requires an initialization of the sample-mode mapping. In the absence of *a priori* knowledge on the switching mechanism, a randomly chosen mode in $\mathcal{M}$ can be assigned to each sample $(y(t), \boldsymbol{x}(t))$, for all $t = 1, \dots, N$, as done in Step 1 of Algorithm 1. Alternatively, unsupervised classification algorithms can be employed to look for patterns in $\mathcal{D}$. Then, the generic k-th iteration of the proposed procedure unfolds as follows.

In the parameter estimation phase, given the current sample assignment $\lambda_t(k)$, we minimize the cost function of (3) with respect to $\boldsymbol{\vartheta}_i$, $i \in \mathcal{M}$. Since λ_t is fixed and equal to $\lambda_t(k)$, problem (3) becomes

$$\begin{aligned} \min_{\{\boldsymbol{\vartheta}_i\}_{i\in\mathcal{M}}} \quad & \sum_{i\in\mathcal{M}} \sum_{t:\lambda_t(k)=i} \Big(y(t) - \boldsymbol{\vartheta}_i^\top \boldsymbol{x}(t)\Big)^2 \\ \text{subject to:} \quad & \boldsymbol{\vartheta}_i \in \mathbb{R}^{n_x},\ i \in \mathcal{M}, \end{aligned} \tag{4}$$

Algorithm 1 SARX/PWARX identification

1: $\lambda_t(0) \leftarrow$ random element from $\mathcal{M}$, $t = 1, \dots, N$
2: $k = 0$
3: **repeat**
4: % Parameter estimation phase
5: **for** $i \in \mathcal{M}$ **do**
6: $\boldsymbol{\vartheta}_i(k) = \arg\min_{\boldsymbol{\vartheta}_i \in \mathbb{R}^{n_x}} \sum_{t:\lambda_t(k)=i} \Big(y(t) - \boldsymbol{\vartheta}_i^\top \boldsymbol{x}(t)\Big)^2$
7: **end for**
8: % Sample assignment phase
9: **for** $t = 1$ **to** N **do**
10: $\hat{\lambda}_t(k+1) = \arg\min_{\lambda_t \in \mathcal{M}} \Big(y(t) - \boldsymbol{\vartheta}_{\lambda_t}^\top(k) \boldsymbol{x}(t)\Big)^2$
11: **end for**
12: % Post-processing phase
13: $\{\lambda_t(k+1)\}_{t=1}^N \leftarrow \mathcal{F}(\{\hat{\lambda}_t(k+1)\}_{t=1}^N)$
14: $k \leftarrow k + 1$
15: **until** stopping criterion is met

which is separable over i and thus reducible to m Least Squares (LS) problems, one per mode (cf. Step 6 in Algorithm 1).

In the sample assignment phase, given the current model parameters $\boldsymbol{\vartheta}_i(k)$ for each mode $i \in \mathcal{M}$, we minimize the cost function of (3) with respect to λ_t, $t = 1, \dots, N$. Setting $\boldsymbol{\vartheta}_i = \boldsymbol{\vartheta}_i(k)$ and swapping the two summation in the cost function, (3) becomes

$$\begin{aligned} \min_{\{\lambda_t\}_{t=1}^N} \quad & \sum_{t=1}^N \sum_{i:\lambda_t=i} \Big(y(t) - \boldsymbol{\vartheta}_i^\top(k) \boldsymbol{x}(t)\Big)^2 \\ \text{subject to:} \quad & \lambda_t \in \mathcal{M},\ t = 1, \dots, N \\ & (\lambda_1, \dots, \lambda_N) \in \Lambda, \end{aligned}$$

which is equivalent to

$$\begin{aligned} \min_{\{\lambda_t\}_{t=1}^N} \quad & \sum_{t=1}^N \Big(y(t) - \boldsymbol{\vartheta}_{\lambda_t}^\top(k) \boldsymbol{x}(t)\Big)^2 \\ \text{subject to:} \quad & \lambda_t \in \mathcal{M},\ t = 1, \dots, N \\ & (\lambda_1, \dots, \lambda_N) \in \Lambda. \end{aligned} \tag{5}$$

Problem (5) is, however, hard to solve due to the presence of the constraints encoded by Λ, which couple the λ_t's decision variables. We then propose to neglect the Λ constraint, thus making problem (5) separable into N sub-problems, one per sample (cf. Step 10 in Algorithm 1), with the following interpretation: each data point $y(t)$ is provisionally assigned to the mode that best explains it, or, equivalently, to the model with the smallest prediction error at t.

Finally, we seek to enforce the assignment variable constraint Λ via an *ad hoc* post-processing phase (cf. Step 13 in Algorithm 1), which depends on the assumed system class and is discussed in the next subsections. The heuristic procedure $\mathcal{F}(\cdot)$ is directly applied to the assignment returned by the previous phase and implements the post-processing phase.

The algorithm iteratively performs the three described tasks and terminates whenever the data partition does not change across two subsequent iterations, the change in the cost function across iterations is below a user-defined threshold, or a maximum number of iterations is reached.

3.1 Switched affine systems with dwell time

Suppose that (1) is a SARX system and let D be the largest odd integer smaller than or equal to the system dwell time. Then, the heuristic procedure (coded by $\mathcal{F}(\cdot)$ in Algorithm 1) attempts to correct wrong assignments $\hat{\lambda}_t(k)$ by reassigning the mode of isolated samples. This indirectly favors the obtainment of the dwell time constraint. Let $w = D - 1$ and consider the following vector of assignment variables

$$L_t(k) = \left(\hat{\lambda}_{t-w}(k), \ldots, \hat{\lambda}_{t+w}(k)\right) \in \mathcal{M}^{2D-1}$$

of size $2D-1$ and centered at t, for iteration k. Then, for each $t = w+1, \ldots, N-w$, the map $\mathcal{F}(\{\hat{\lambda}_t(k)\}_{t=1}^N)$ assigns to $\lambda_t(k)$ the most frequent mode value within $L_t(k)$. In case of a tie, we set $\lambda_t(k)$ equal to the mode with the best fitting at t, among the most frequent modes in $L_t(k)$. For $t < w$ and $t > N - w$, we set $\lambda_t(k) = \hat{\lambda}_t(k)$.

3.2 Piecewise affine systems

Suppose now that (1) belongs to the PWARX class of models. In this case, data points should be clustered in linearly separable regions, each associated to a different mode. To take this constraint on the assignment variables into account, if a sample is surrounded by data belonging to the same mode, then we assume that it belongs to the same mode. This is implemented in the following heuristic procedure (coded by $\mathcal{F}(\cdot)$ in Algorithm 1). Let p be a positive integer and $\mathcal{N}_{p,t}$ the set of the p nearest neighbors of $(y(t), \boldsymbol{x}(t))$ in the $\mathbb{R} \times \mathbb{R}^{n_x}$ space, where the distance between two data points $(y(t_1), \boldsymbol{x}(t_1))$ and $(y(t_2), \boldsymbol{x}(t_2))$ space is given by $\|(y(t_1), \boldsymbol{x}(t_1)) - (y(t_2), \boldsymbol{x}(t_2))\|$. Let

$$L_t(k) = \left(\hat{\lambda}_{t_1}(k), \ldots, \hat{\lambda}_{t_p}(k)\right) \in \mathcal{M}^p$$

be the vector of mode indices associated to the data points $(y(t_1), \boldsymbol{x}(t_1)), \ldots, (y(t_p), \boldsymbol{x}(t_p))$ in $\mathcal{N}_{p,t}$. Then, for each $t = 1, \ldots, N$, the map $\mathcal{F}(\{\hat{\lambda}_t(k)\}_{t=1}^N)$ assigns to $\lambda_t(k)$ the most frequent mode index in $L_t(k)$. As for the SARX model class, in case of a tie, we set $\lambda_t(k)$ equal to the mode with the best fitting at t, among the most frequent modes in $L_t(k)$.

Notice that the distance measure employed above is defined in the joint regressor-output space as opposed to the plain regressor space. Indeed, this often facilitates the classification of samples near the boundaries of the polyhedral regions, especially in case of discontinuities among modes, where data points that are close in the regressor space actually belong to different modes. It is important to stress that here we are addressing the sample assignment, as opposed to the reconstruction of the polyhedral partition $\{\mathcal{X}\}_{i \in \mathcal{M}}$, which can be carried out on the regressor space only once the identification procedure ends.

4. NUMERICAL EXAMPLES

In this section the effectiveness of the proposed method is shown with reference to some numerical examples involving both SARX and PWARX model identification. Furthermore, the method is compared with the recursive techniques presented in (Bako et al., 2011; Breschi et al., 2016), the k-RANSAC (Hartmann et al., 2015), and the approach in (Bemporad et al., 2018). To allow a fair comparison of all the methods in the offline setting discussed in this work, online recursive methods are executed as follows: starting from a user-defined initialization, the identification is first performed on half the data set and then the algorithm is run again 10 times on the whole data set by initializing it each time according to the output of the previous run. As suggested in Bemporad et al. (2018), the method is run 5 times from different random initial sample assignments and the best result is selected. Table 1 lists the parameter settings used for the different methods. The methods are evaluated in terms of the classification error rate, *i.e.* the percentage of misclassified data points, and the fitting error, which is measured as in (Bako et al., 2011) according to

$$FIT = \left(1 - \frac{\|\hat{y} - y\|_2}{\|y - \bar{y}\mathbb{1}_N\|_2}\right), \tag{6}$$

where $\hat{y}$ denotes the estimated model output sequence, $\bar{y}$ is the average of the true output sequence y, and $\mathbb{1}_N$ denotes an N-dimensional vector of ones. For each example, a Monte Carlo analysis is carried out with 1000 different data realizations of length 2000 each and the evaluation indices are averaged over all runs. All tests have been performed in a MATLAB 2017a environment, on an i7 2.20-GHz Intel Core processor with 8 GB of RAM.

4.1 SARX identification

In this study, the following system has been considered (Bako et al., 2011):

$$\begin{aligned}
\text{mode 1: } y(t) = &-0.0322y(t-1) + 0.8017y(t-2) \\
&- 1.2878u(t-1) - 1.1252u(t-2) + e(t) \\
\text{mode 2: } y(t) = &-0.1921y(t-1) + 0.5917y(t-2) \\
&+ 1.1050u(t-1) + 0.0316u(t-2) + e(t) \\
\text{mode 3: } y(t) = &+1.4746y(t-1) - 0.5286y(t-2) \\
&- 0.4055u(t-1) + 0.2547u(t-2) + e(t)
\end{aligned}$$

Table 1. Parameter settings for each method (please, refer to the cited papers for the parameters definition and notations).

Method	**SARX**	**PWARX**
Proposed method	$w = 9$	$p = 15$
Bako et al. (2011)	$\gamma = 0.9$ $L_j(0) = 100I_{n_x}$ fitting error: prior $\alpha = 0$ $c(0)$: not used	$\gamma = 0.9$ $L_j(0) = 100I_{n_x}$ fitting error: prior $\alpha = 0.3$ $c(0)$: random
Breschi et al. (2016)	$T^{i,j}(0) = 1000I_{n_x}$ $\Lambda_e = var(y)$ $\kappa = 1$ A_i set equal to best linear model $c(0)$: not used $R_i(0)$: not used	$T^{i,j}(0) = 100I_{n_x}$ $\Lambda_e = var(y)$ $\kappa = 1$ A_i set equal to best linear model $c(0)$: random $R_i(0) = I_{n_x}$
k-RANSAC	$\varepsilon = 0.1$ $\lvert\mathcal{S}_k\rvert = 10$	$\varepsilon = 0.1$ $\lvert\mathcal{S}_k\rvert = 10$
Bemporad et al. (2018)	$\ell(\boldsymbol{x}(t), y(t), \boldsymbol{\vartheta}_{s_t}) = \|y(t) - \boldsymbol{\vartheta}_{s_t}^\top \boldsymbol{x}(t)\|^2$ $r(\boldsymbol{\vartheta}_k) = 0$ $\mathcal{L}(S)$ see Section 4.4 in Bemporad et al. (2018)	$\ell(\boldsymbol{x}(t), y(t), \boldsymbol{\vartheta}_{s_t}) = \|y(t) - \boldsymbol{\vartheta}_{y,s_t}^\top \boldsymbol{x}(t)\|^2 + 0.01\|\boldsymbol{x}(t) - \boldsymbol{\vartheta}_{x,s_t}\|^2$ $r(\boldsymbol{\vartheta}_k) = 0$ $\mathcal{L}(S) = 0$

where $e(t)$ is a zero mean white Gaussian noise chosen so as to have 25 dB of Signal to Noise Ratio, and $u(t)$ is a zero mean white Gaussian noise with variance equal to 1.

The system switches randomly 20 times in the interval $(1, 2000)$, with a minimum dwell time D of 10 time instants. Figure 1 shows the estimated discrete state sequence (bottom) obtained by applying the proposed method on the generated data set (top). The triangular markers denote the switching time instants. The method solved efficiently the sample-mode mapping on the considered data realization, as it is also evident from the computed classification error rate, 0.3003%, and FIT criterion, 94.3227%. The identified local model parameters are:

$$\begin{aligned}\boldsymbol{\vartheta}_1 &= [-0.0377, \quad 0.8058, \ -1.2924, \ -1.1378]\\ \boldsymbol{\vartheta}_2 &= [-0.1934, \quad 0.5889, \quad 1.1046, \quad 0.0312]\\ \boldsymbol{\vartheta}_3 &= [\ 1.4873, \ -0.5434, \ -0.4047, \quad 0.2606]\end{aligned}$$

Table 2 reports the aggregated results of the comparative analysis carried out on the methods reported in Table 1. The proposed method significantly outperformed all the compared methods on the considered example when the assignment constraints are taken into account in the post-processing phase, while similar performance with that of (Bako et al., 2011) and (Breschi et al., 2016) were experienced without applying the heuristic procedure $\mathcal{F}(\cdot)$. Furthermore, the comparatively low standard deviation values of the accuracy indicators imply that the proposed approach is more robust than the competitors.

Table 2. SARX Identification. Comparative analysis: aggregated mean ± standard deviation values.

Method	**Cl. error** [%]	**FIT** [%]	**CPU time** [s]
Bako et al. (2011)	8.04 ± 2.79	91.11 ± 5.72	0.33 ± 0.02
Breschi et al. (2016)	7.33 ± 1.88	93.92 ± 2.18	0.31 ± 0.01
k-RANSAC	18.14 ± 7.77	76.77 ± 9.91	0.92 ± 0.03
Bemporad et al. (2018)	1.20 ± 3.38	92.57 ± 3.08	0.88 ± 0.12
Proposed method (no post-processing)	6.66 ± 0.70	94.82 ± 0.50	0.04 ± 0.02
Proposed method	0.58 ± 0.18	91.86 ± 1.72	0.81 ± 0.35

4.2 PWARX identification

In this example, the following system has been considered (Bemporad et al., 2005):

$$y(t) = \begin{cases} -0.4y(t-1) + u(t-1) + 1.5 + e(t) \\ \qquad \text{if } 4y(t-1) - u(t-1) + 10 < 0 \\ 0.5y(t-1) - u(t-1) - 0.5 + e(t) \\ \qquad \text{if } 4y(t-1) - u(t-1) + 10 \geq 0 \\ \qquad \text{and } 5y(t-1) + u(t-1) - 6 \leq 0 \\ -0.3y(t-1) + 0.5u(t-1) - 1.7 + e(t) \\ \qquad \text{if } 5y(t-1) + u(t-1) - 6 > 0 \end{cases}$$

where $u(t)$ and $e(t)$ are sequences of independent random variables uniformly distributed in the ranges $[-4, 4]$ and $[-0.2, 0.2]$, respectively. Figure 2 shows a data realization of length 2000 used for estimation. Figure 3 displays the sample-mode mapping obtained by applying the proposed approach. It is easy to note that the region boundaries have been quite accurately estimated. This reflects on the final classification error rate and FIT values which amount to 0.1501% and 96.2870%, respectively. The identified affine subsystems are:

$$\begin{aligned}\boldsymbol{\vartheta}_1 &= [-0.4047, \quad 1.0016, \quad 1.4867]\\ \boldsymbol{\vartheta}_2 &= [\ 0.5001, \ -0.9979, \ -0.4936]\\ \boldsymbol{\vartheta}_3 &= [-0.3016, \quad 0.5018, \ -1.6894]\end{aligned}$$

Table 3 reports the aggregated results of the comparative analysis. The algorithm proposed in the paper shows comparable performance with the approach presented in Bemporad et al. (2018), while it outperforms all the other competitors particularly in terms of sample classification. Also note that significant differences in the computed classification error rates do not correspond to equally significant differences in the fitting performance, *i.e.*, FIT values. This is not surprising since the majority of misclassified samples lie on the mode boundaries and hence can be consistent with more than one affine submodel. Nonetheless, reducing the amount of misclassified data points is fundamental if the estimation of the polyhedral partition $\{\mathcal{X}_i\}_{i\in\mathcal{M}}$ is addressed based on the results of the identification procedure.

Table 3. PWARX Identification. Comparative analysis: aggregated mean ± standard deviation values.

Method	**Cl. error** [%]	**FIT** [%]	**CPU time** [s]
Bako et al. (2011)	0.59 ± 0.87	95.85 ± 0.84	0.43 ± 0.01
Breschi et al. (2016)	1.91 ± 0.70	96.04 ± 0.83	0.77 ± 0.03
k-RANSAC	5.63 ± 3.55	82.80 ± 9.94	0.87 ± 0.02
Bemporad et al. (2018)	0.16 ± 0.10	96.25 ± 0.08	2.36 ± 0.52
Proposed method (no post-processing)	2.11 ± 1.04	96.27 ± 0.42	0.04 ± 0.02
Proposed method	0.15 ± 0.14	96.24 ± 0.08	1.12 ± 0.53

5. CONCLUSION

An iterative batch method has been presented for switched system identification. It tackles the typical heterogeneous optimization problem which characterizes the identification of SARX and PWARX models with an alternating optimization strategy over the model parameters first, and then over the sample-mode assignment variables. A distinguished feature of the presented approach is the capability of correcting wrong sample-mode assignments by means of a simple and computationally affordable procedure which can take into account for assignment constraints originating from some prior knowledge on the switching mechanism. Numerical results show that the proposed method is quite accurate and has superior performance with respect to alternative methods in the literature.

Future work will focus on the convergence analysis of the proposed method, as well as the analysis of robustness against a possibly mode dependent noise. Furthermore, the extension to the case when the number of modes is not *a priori* known will be pursued.

REFERENCES

Bako, L., Boukharouba, K., Duviella, E., and Lecoeuche, S. (2011). A recursive identification algorithm for switched linear/affine models. *Nonlinear Analysis: Hybrid Systems*, 5(2), 242–253.

Bemporad, A., Breschi, V., Piga, D., and Boyd, S.P. (2018). Fitting jump models. *Automatica*, 96, 11 – 21.

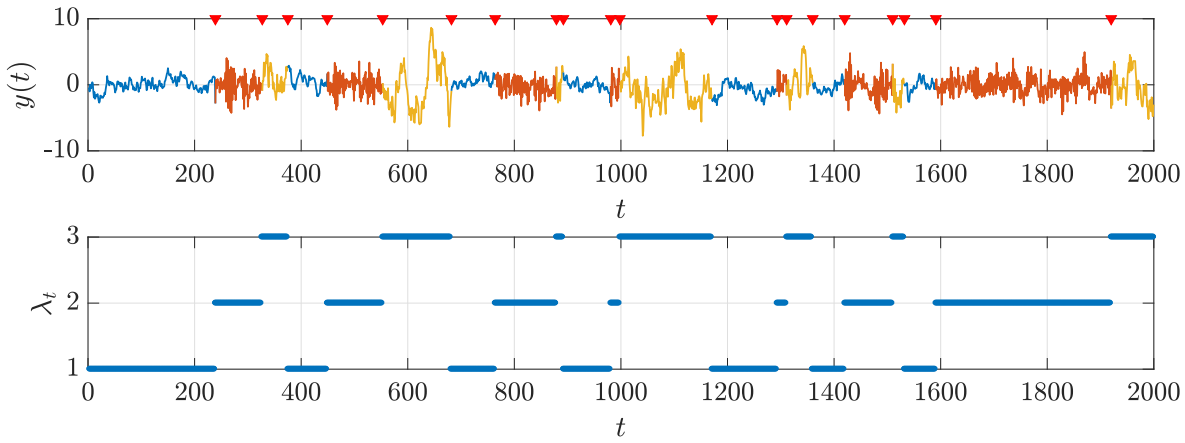

Fig. 1. SARX Identification, single run execution: available data points (top), where different colors denote different modes, and estimated discrete state (bottom). The triangular markers represent the true switching times.

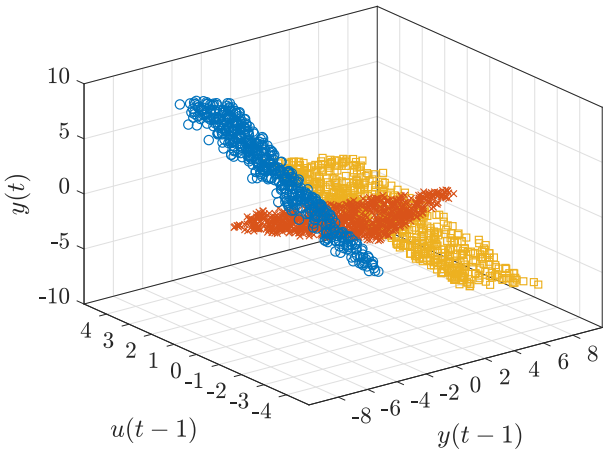

Fig. 2. PWARX Identification, single run execution: available data points. Different colors denote different modes.

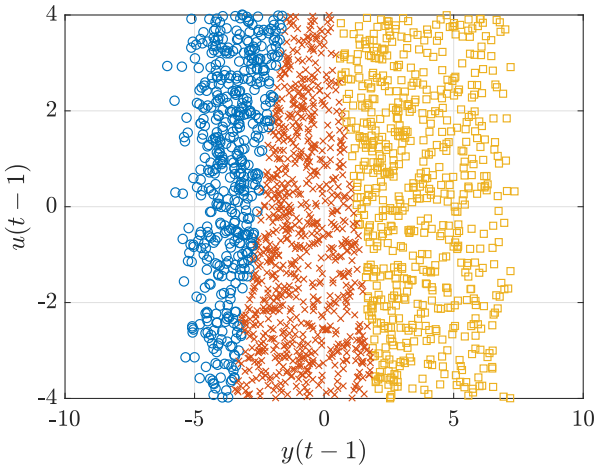

Fig. 3. PWARX Identification: data classification obtained with the proposed approach. Different colors denote different modes.

Bemporad, A., Garulli, A., Paoletti, S., and Vicino, A. (2005). A bounded-error approach to piecewise affine system identification. *IEEE Transactions on Automatic Control*, 50(10), 1567–1580.

Bianchi, F., Prandini, M., and Piroddi, L. (2020). A randomized two-stage iterative method for switched nonlinear systems identification. *Nonlinear Analysis: Hybrid Systems*, 35, 100818.

Breschi, V., Piga, D., and Bemporad, A. (2016). Piecewise affine regression via recursive multiple least squares and multicategory discrimination. *Automatica*, 73, 155–162.

Ferrari-Trecate, G., Muselli, M., Liberati, D., and Morari, M. (2003). A clustering technique for the identification of piecewise affine systems. *Automatica*, 39(2), 205–217.

Fischler, M.A. and Bolles, R.C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), 381–395.

Garulli, A., Paoletti, S., and Vicino, A. (2012). A survey on switched and piecewise affine system identification. In 16^{th} *IFAC Symposium on System Identification*, 344–355. Brussels, Belgium.

Hartmann, A., Lemos, J.M., Costa, R.S., Xavier, J., and Vinga, S. (2015). Identification of switched ARX models via convex optimization and expectation maximization. *Journal of Process Control*, 28, 9–16.

Juloski, A.L., Weiland, S., and Heemels, W. (2005). A Bayesian approach to identification of hybrid systems. *IEEE Transactions on Automatic Control*, 50(10), 1520–1533.

Lauer, F. and Bloch, G. (2019). *Hybrid System Identification*. Springer International Publishing.

Lauer, F., Bloch, G., and Vidal, R. (2011). A continuous optimization framework for hybrid system identification. *Automatica*, 47(3), 608–613.

Ma, Y. and Vidal, R. (2005). Identification of deterministic switched ARX systems via identification of algebraic varieties. In *International Workshop on Hybrid Systems: Computation and Control*, 449–465.

Ohlsson, H. and Ljung, L. (2013). Identification of switched linear regression models using sum-of-norms regularization. *Automatica*, 49(4), 1045–1050.

Paoletti, S. (2004). *Identification of piecewise affine models*. Ph.D. thesis, Dipartimento di Ingegneria dell'Informazione, Univ. Siena, Siena, Italy.

Piga, D. and Tóth, R. (2013). An SDP approach for ℓ_0-minimization: Application to ARX model segmentation. *Automatica*, 49(12), 3646–3653.

Pillonetto, G. (2016). A new kernel-based approach to hybrid system identification. *Automatica*, 70, 21–31.

Roll, J., Bemporad, A., and Ljung, L. (2004). Identification of piecewise affine systems via mixed-integer programming. *Automatica*, 40(1), 37–50.