A switched nonlinear system identification method with switching location refinement

Miao Yu, Federico Bianchi, and Luigi Piroddi

***Abstract*— The identification of switched nonlinear systems involves solving a combinatorial problem that simultaneously addresses sample-mode assignment and nonlinear model structure selection. The complexity of this problem is often prohibitive, since mode switchings can take place at arbitrary times. To reduce it to an affordable level, one can constrain the mode switchings to occur only at few specific instants. This approach is effective if combined with a refinement strategy, that corrects the number and locations of the switchings. In this paper, one such strategy is discussed, which employs a local optimization process to correct the position of switchings, and is also capable of detecting redundant modes. An iterative method, applying an identification step and a refinement step at all iterations, is tested on a numerical example to illustrate the effectiveness of the refinement strategy. The method does not require prior assumptions on the number of modes.**

I. INTRODUCTION

Hybrid systems can describe dynamically rich phenomena that combine continuous and discrete dynamics [5]. An often used representation employs different continuous local models (*modes*), and a switching mechanism between them. In switched systems, the latter is determined by an exogenous finite-valued switching signal which identifies at all times which mode is active. On the other hand, piecewise affine models switch according to a polyhedral partition of the state-input domain.

The identification of hybrid systems involves both the estimation of the local dynamics and the switching mechanism, configuring a complex optimization problem. Various approaches have been introduced in the recent literature to address this task (see, *e.g.*, [12], [4], [6]). Comparatively fewer works address the case with nonlinear local models, and typically in a non-parametric setting using kernel functional expansions (see, *e.g.*, [1], [7], [8], [9]). Employing a parametric framework can lead to more compact and interpretable models, but adds another dimension to the problem. Indeed, one can resort *e.g.* to polynomial Nonlinear AutoRegressive with eXogenous input (NARX) models to represent the local dynamics associated to the modes, [10], [11], approximating the system nonlinearities by means of finite-dimensional parameterized polynomial expansions. In that case, however, the identification procedure is effective only if it includes model structure selection or at least regularization to suitably compress the model size.

The authors are with the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, via Ponzio 34/5, 20133 Milano (Italy) {miao.yu, federico.bianchi, luigi.piroddi}@polimi.it

One such method for switched NARX (SNARX) systems is discussed in [2]. The SNARX identification problem is formulated as a combinatorial problem, where the discrete variables are associated to the switching signal *and* the structure of the NARX local models. To manage the combinatorial complexity, mode switching is allowed only at specific instants, which results in a significant reduction of the discrete variables. A randomized method is employed to simultaneously identify the switching signal, thereby assigning a mode to each sampling time (*sample-mode assignment*), and the local nonlinear models associated to the various modes (*local model identification*). The method is shown to be very accurate if the set of candidate switching times includes the true ones, and to provide satisfactory approximations otherwise.

In [3] a refinement stage is added to correct the switching locations, and the identification and refinement step are alternated until convergence. In this way, both the switching signal and the local models can be more accurately estimated. The rationale of the refinement strategy is to augment the set of candidate switching times by sampling more densely in the neighborhood of the detected mode switchings (denoted *active* switchings). Instead, candidate switching times not corresponding to actual detected switchings are removed. This heuristic yields limited improvements as it does not rely on a precise relocation of a detected mode switching, but simply increases the number of candidate switching locations near it. This adds further complexity to the combinatorial problem and generates small time intervals, which complicates the sample-mode assignment at the next run of the identification step. Furthermore, the method rests on the assumption that the number of modes is known.

We here introduce an improved refinement step, designed to correct not only the number and location of the switchings but also the number of modes. The refinement step has been redesigned based on a completely different principle, whereby instead of adding new candidate switching instants near an active switching point, the location of the latter is corrected with a local optimization approach. This prevents the complexity increase mentioned before, and avoids the generation of small intervals. Indeed, the number of switching time instants is rapidly reduced, by removing redundant ones and correcting the location of the active ones. Additionally, the proposed method removes the assumption that the number of modes be known, in that it allows to efficiently detect and eliminate the redundant modes.

The proposed algorithm has been tested on a benchmark example to demonstrate its effectiveness.

II. The SNARX model identification problem

A. The SNARX model class

A SNARX is a switched system where the different modes are described by NARX models, [10], [11]. The switching signal $\boldsymbol{\sigma} \in \{1, ..., N_M\}^N$ associates a mode to each sample, N being the length of the observation window and N_M the number of modes. A NARX model is formulated as an input-output recursion, where the current output value is a nonlinear function of past input and output samples, plus noise. Accordingly, a SNARX model can be written as:

$$y(t) = f^{\sigma(t)}(\boldsymbol{x}(t)) + e(t) \tag{1}$$

where $f^i(\cdot)$ defines the nonlinear dynamics of the ith mode, $i = 1, \dots, N_M$, $\boldsymbol{x}(t) = [y(t-1) \dots y(t-n_y)\, u(t-1) \dots u(t-n_u)]$ is a finite-dimensional vector including the past input and output samples (up to the maximum lags n_u and n_y, respectively), and finally $e(t)$ is a white noise signal. More in detail, the nonlinear dynamics are represented as a linear combination of nonlinear basis functions:

$$f^i(\boldsymbol{x}(t)) = \boldsymbol{\varphi}(\boldsymbol{x}(t))^T \boldsymbol{\vartheta}^i \tag{2}$$

where $\boldsymbol{\varphi}(\boldsymbol{x}(t)) = [\varphi_1(\boldsymbol{x}(t)) \dots \varphi_n(\boldsymbol{x}(t))]^T$ is the vector of basis functions and $\boldsymbol{\vartheta}^i = [\vartheta_1^i \dots \vartheta_n^i]^T$ is the parameter vector associated to the ith mode. Since the model equation takes the form of a linear regression, the $\varphi_j(\boldsymbol{x}(t))$ functions are also referred as regressors. In the context of NARX models, the nonlinearity is often represented as a polynomial functional expansion, the regressors being monomials of $\boldsymbol{x}$. However, since the number of polynomial terms increases rapidly with the number of arguments ($n_y + n_u$), and the degree n_d of the expansion, model structure selection techniques are generally employed to mitigate overparametrization issues and enforce model robustness. The maximum values of n_y, n_u, and n_d are set by the user.

B. SNARX model identification

Let $\{(u(t), y(t))\}_{t=1:N}$ be a set of input-output data pairs obtained from an unknown switched nonlinear system. The identification of a SNARX model from these data consists of two tasks: a) sample-mode assignment, and b) local model identification. The latter task involves both model structure selection and parameter estimation.

As a result of model structure selection, not all regressors $\varphi_j(\boldsymbol{x}(t))$, $j = 1, \dots, n$, are included in a local model. Accordingly, a binary vector $\boldsymbol{s} \in \{0,1\}^n$ can be used to characterize its structure (0 indicating term rejection from the model). Overall, the regressor composition of a SNARX system can be encoded in a binary matrix $\mathcal{S} \in \boldsymbol{\mathcal{S}} = \{0,1\}^{n \times N_M}$, where the kth column defines the structure of the NARX model associated to the kth mode. The switching signal is instead defined by a vector of integers $\boldsymbol{\sigma} = [\sigma(1) \dots \sigma(N)] \in \Sigma = \{1, 2, ..., N_M\}^N$, $\sigma(t)$ indicating the mode associated to sample t. The structure of a SNARX system is then completely characterized by $\lambda = (\boldsymbol{\sigma}, \mathcal{S})$, with $\lambda \in \Lambda = \Sigma \times \boldsymbol{\mathcal{S}}$. Given a candidate SNARX model structure λ, one can univocally determine the local model parameters and compute the model output $\hat{y}(t)$. Finally, one can evaluate the accuracy associated to the SNARX model structure λ, *e.g.* in terms of the mean square prediction error $\mathcal{L}(\lambda)$ achieved by $\hat{y}(t)$. In the following, we will use the performance index

$$\mathcal{J}(\lambda) = e^{-K_\lambda \mathcal{L}(\lambda)}, \tag{3}$$

which is conveniently scaled from 0 to 1, to rate a SNARX model structure.

C. SNARX model identification as a combinatorial problem

Given a SNARX structure $\lambda = (\boldsymbol{\sigma}, \mathcal{S})$, the parameter estimation task for each mode can be univocally addressed by applying ordinary Least Squares on the data portion pertaining to that mode (as of $\boldsymbol{\sigma}$), using the regressors assigned to that mode according to $\mathcal{S}$ (see, *e.g.*, [2]). Therefore, the SNARX identification problem ultimately amounts to finding the value $\lambda^* = (\boldsymbol{\sigma}^*, \mathcal{S}^*) \in \Lambda$ that maximizes the performance index $\mathcal{J}(\lambda)$. An exhaustive approach to the solution of this combinatorial problem is typically unfeasible due to the size of set Λ. It is typically the length N of vector $\boldsymbol{\sigma}$ that blows up the combinatorial complexity. On the other hand, the number of actual samples in which a mode change occurs is normally much smaller than N.

This suggests the following two-step workaround [3]. First, the problem is addressed assuming that switchings can occur only at a few given time instants, specified in the ordered set $\mathcal{T}_s = \{t_k\}_{k=1:N_s}$, with $t_1 = 1$, $t_k < t_{k+1}$ for $k = 1, \dots, N_s - 1$, and $t_{N_s} \leq N$. Set $\mathcal{T}_s$ includes the original guess on the possible switching times, based on the available prior knowledge (*e.g.*, minimum dwell-time). The definition of $\mathcal{T}_s$ leads to a partition of the observation window $\{1, \dots, N\}$ into N_s sub-periods $I_k = [t_k, t_{k+1} - 1]$ for $k = 1, \dots, N_s - 1$, and $I_{N_s} = [t_{N_s}, N]$[1]. Accordingly, $\sigma(t) = \sigma(t_k)$ for $t \in I_k$, *i.e.* mode $\sigma(t_k)$ is active in the entire time interval, and one need only specify the values of $\sigma(t)$ for the time samples in $\mathcal{T}_s$. This greatly reduces the size of the solution space, and allows to effectively tackle the combinatorial problem [2]. However, the result is sub-optimal since the coarse subdivision of the data in sub-periods allows only an approximate estimation of both the switching locations and the local models, which are generally identified on mixed data.

The second step consists in refining the set $\mathcal{T}_s$ based on the result of the identification step, removing redundant sample times and correcting the location of the remaining ones. Then the identification step is repeated with the new set $\mathcal{T}_s$ and the method is iterated until convergence.

In the following sections, we briefly review the method introduced in [2] for dealing with the identification step, and propose a novel scheme for the refinement step. Besides refining the switching time locations, the proposed scheme is designed to detect and remove redundant modes as well, so that the method can be employed to estimate the number of modes as well (and does not need prior information on this parameter).

[1] Notice that if the length of the I_k sub-periods matches the minimum dwell-time, at most one switching can occur in each sub-period.

III. The identification step

This section briefly reviews the randomized approach for SNARX identification introduced in [2], based on the assumption that mode switchings can occur only at the time instants in $\mathcal{T}_s$. This algorithm generally provides excellent results if $\mathcal{T}_s^\circ \subseteq \mathcal{T}_s$, $\mathcal{T}_s^\circ$ being the set of the real switching times, and performs reasonably otherwise, albeit with some unavoidable approximation errors.

The method reformulates the identification problem as the tuning of a probability distribution over Λ. Briefly, the probability distribution is tuned based on a sample-and-evaluate strategy, whereby various SNARX model structures are extracted from the distribution and collectively evaluated to infer how to correct the distribution for better performance. Eventually, the distribution converges to a limit distribution corresponding to a specific NARX structure.

Let $\mathbb{P}_\Phi(\lambda) = \mathbb{P}_{\boldsymbol{\xi}}(\boldsymbol{\sigma}) \cdot \mathbb{P}_{\boldsymbol{\varsigma}}(\mathcal{S})$ be a probability distribution defined over Λ, where $\Phi = (\boldsymbol{\xi}, \boldsymbol{\varsigma})$ is the associated random variable. Here, $\boldsymbol{\xi} \in \Sigma$ and $\boldsymbol{\varsigma} \in \boldsymbol{\mathcal{S}}$ are associated to the switching signal and the SNARX regressor composition, respectively. The adopted factorization of $\mathbb{P}_\Phi(\lambda)$ rests on the independence assumption between the mode sequence $\boldsymbol{\sigma}$ and the SNARX regressor composition $\mathcal{S}$. The average performance of Φ can be calculated as $\mathbb{E}[\mathcal{J}(\Phi)] = \sum_{\lambda\in\Lambda} \mathbb{P}_\Phi(\lambda)\mathcal{J}(\lambda)$. The best probability distribution $\mathbb{P}_\Phi^*$ maximizes $\mathbb{E}[\mathcal{J}(\Phi)]$. Accordingly, λ^* can be characterized as

$$\lambda^* = \arg\max_{\lambda\in\Lambda} \ \mathbb{P}_\Phi^*(\lambda) \tag{4}$$

The randomized approach consists in extracting samples of Φ and evaluating them to approximate $\mathbb{E}[\mathcal{J}(\Phi)]$, using this information to correct the distribution, until convergence to a limit distribution (*i.e.*, to a specific SNARX structure λ^*).

In order to tackle the optimization of $\mathbb{E}[\mathcal{J}(\Phi)]$, a suitable parametrization of $\mathbb{P}_\Phi(\lambda)$ is needed.

A. Parametrization of $\mathbb{P}_\Phi(\lambda)$

Recalling that mode switchings can occur only at time instants belonging to $\mathcal{T}_s$, we associate a Categorical random variable $\xi(k) \sim \texttt{Cat}(\boldsymbol{\eta}_k)$ to each $\sigma(t_k)$, $t_k \in \mathcal{T}_s$, where $\boldsymbol{\eta}_k = [\eta_k^1, \eta_k^2, ..., \eta_k^{N_M}]$, and η_k^i denotes the probability that $\sigma(t_k)$ (actually the entire interval I_k) is assigned to mode i. Clearly, $\sum_{i=1}^{N_M} \eta_k^i = 1$. By assuming independence between the elements $\sigma(t_k)$, $\mathbb{P}_{\boldsymbol{\xi}}(\boldsymbol{\sigma})$ takes the form

$$\mathbb{P}_{\boldsymbol{\xi}}(\boldsymbol{\sigma}) = \prod_{t_k\in\mathcal{T}_s} \eta_k^{\sigma(t_k)}. \tag{5}$$

Similarly, a Bernoulli random variable $\rho_j^i \sim \texttt{Be}(\mu_j^i)$ is associated to each element in $\mathcal{S}$, where the success probability μ_j^k represents the belief that the jth regressor φ_j is included in the ith local model. By assuming independence between each element in $\mathcal{S}$, $\mathbb{P}_{\boldsymbol{\varsigma}}(\mathcal{S})$ takes the form

$$\mathbb{P}_{\boldsymbol{\varsigma}}(\mathcal{S}) = \prod_{i=1}^{N_M} \prod_{j:\mathcal{S}_{ji}=1} \mu_j^i \prod_{j:\mathcal{S}_{ji}=0} (1-\mu_j^i). \tag{6}$$

The probability distribution $\mathbb{P}_\Phi(\lambda)$ is thus parameterized by the set of scalar parameters η_k^i and μ_j^i, denoted Mode Extraction Probabilities (MEPs) and Regressor Inclusion Probabilities (RIPs), respectively.

B. Tuning of $\mathbb{P}_\Phi(\lambda)$

The MEPs and RIPs are iteratively tuned by extracting and evaluating sample values $\lambda = (\boldsymbol{\sigma}, \mathcal{S})$ of $\Phi = (\boldsymbol{\xi}, \boldsymbol{\varsigma})$. Specifically, they are updated based on the following equations:

$$\eta_k^i \leftarrow \eta_k^i + \chi\delta_k^i, \qquad \mu_j^i \leftarrow \mu_j^i + \chi\ell_j^i \tag{7}$$

where $\chi > 0$ and

$$\delta_k^i = \mathbb{E}_{\mathbb{P}_\Phi}\left[\mathcal{J}(\Phi) \mid \xi(k) = i\right] - \mathbb{E}_{\mathbb{P}_\Phi}\left[\mathcal{J}(\Phi) \mid \xi(k) \neq i\right]$$
$$\ell_j^i = \mathbb{E}_{\mathbb{P}_\Phi}\left[\mathcal{J}(\Phi) \mid \rho_j^i = 1\right] - \mathbb{E}_{\mathbb{P}_\Phi}\left[\mathcal{J}(\Phi) \mid \rho_j^i = 0\right]$$

Note that δ_k^i is positive only if on average it is more convenient to assign sub-period I_k to mode i, rather than not. Similarly, a positive ℓ_j^i indicates that the SNARX model performs better on average when it includes the ith regressor in the jth mode, as opposed to not including it. In practice, the update terms δ_k^i and ℓ_j^i can only be approximated with averages taken on a limited number of extracted samples. This is the reason why a conservative update policy is adopted, limiting the effect of the correction terms with parameter χ. A theoretical proof of the local convergence properties of this update policy is discussed in [3].

Remark 1: The independence assumptions between the introduced random variables are only meant to provide a convenient sampling mechanism of the model structures λ. They do not limit *per se* the nature of the underlying system.

C. Algorithm termination

The algorithm terminates when the update step does not modify the MEPs and RIPs anymore. Then, one can determine the switching signal by setting $\hat{\sigma}(t_k) = \arg\max_i(\eta_k^i)$, and $\hat{\sigma}(t) = \hat{\sigma}(t_k)$ for $t < t_{k+1}$ (or $t \leq N$ if $k = N_s$). The structures of the local models are defined by the estimated RIP values. For further details we address the reader to [3].

IV. The refinement step

The refinement steps first detects and removes redundant modes. Then, a local optimization approach is used to correct the locations of the active switchings, while redundant switchings are removed.

A. Removing redundant modes

Redundant modes are detected according to the following three rules. First, irrespective of their structure and parameters, modes that are not assigned to any sub-period (*i.e.*, any mode j such that $j \notin \cup_{t_k\in\mathcal{T}_s}\hat{\Sigma}(t_k)$, where $\hat{\Sigma}(t_k) = \{i \mid \eta_k^i > 0\}$) can be immediately removed.

Second, if for any two modes i and j it holds that $\mathcal{T}_s^i = \mathcal{T}_s^j$, where $\mathcal{T}_s^l = \{t_k \in \mathcal{T}_s \mid l \in \hat{\Sigma}(t_k)\}$, then one mode is redundant. In that case, only the mode that best fits the common assigned data is retained, and the other is removed.

Another more frequent –and less trivial to detect– condition is that, due to redundancy and noise in the data, the data pertaining to a single mode are fragmented between multiple similar local models. To assess model similarity we

employ a Student's t-test to pairwise compare the identified local models. For fairness, we calculate the residuals of both models on the same data-set, consisting of all sub-periods assigned to either of them. In case the test indicates that modes i and j are statistically indistinguishable, then j is replaced by i in $\hat{\Sigma}(t_k)$ for all k, and the j-th sub-model is removed.

B. Refinement of the switching time locations

After the identification step, one of the following cases applies to each $t_k \in \mathcal{T}_s \setminus \{t_1\}$:

1) Let $|\hat{\Sigma}(t_k)| = |\hat{\Sigma}(t_{k-1})| = 1$. Then, if $\hat{\Sigma}(t_k) = \hat{\Sigma}(t_{k-1})$, no actual mode change occurs at t_k (*inactive* switching), and t_k can be removed: $\mathcal{T}_s = \mathcal{T}_s \setminus \{t_k\}$.
2) If $\hat{\Sigma}(t_k) \cap \hat{\Sigma}(t_{k-1}) = \emptyset$, then a mode change occurs at t_k (*active* switching), and t_k is retained in $\mathcal{T}_s$. However, due to the coarse solution of the mode assignment problem, the switching location must be regarded as approximate and requires a local refinement.
3) If neither condition applies, it is not yet possible to establish if the switching is active, and t_k is retained.

Next, the location of the switchings in $\mathcal{T}_s$ is refined obtaining set $\mathcal{T}_s^{new}$. More in detail, for a given $t_k \in \mathcal{T}_s$ let $\hat{t}_k \in B(t_k)$, where $B(t_k)$ is a conveniently small neighborhood of t_k (*e.g.*, one can set $B(t_k) = [\max(1, t_k - \lfloor w/2 \rfloor), \min(t_k + \lfloor w/2 \rfloor, N)]$, where w determines the length of the time window centered over t_k, save for border issues). Now, for each pair $(i_{k-1}, i_k) \in \hat{\Sigma}(t_{k-1}) \times \hat{\Sigma}(t_k)$, let

$$L^{(i_{k-1},i_k)} = \sum_{t=t_{k-1}}^{t_{k+1}-1} \varepsilon_{i_{k-1}}(t)^2$$

if $i_{k-1} = i_k$, where $\varepsilon_q(t) = y(t) - \boldsymbol{\varphi}(\boldsymbol{x}(t))^T \hat{\boldsymbol{\vartheta}}^q$ is the residual associated to mode q, and

$$L^{(i_{k-1},i_k)} = \min_{\hat{t}_k \in B(t_k)} \bar{L}^{(i_{k-1},i_k)}(\hat{t}_k)$$

otherwise, where

$$\bar{L}^{(i_{k-1},i_k)}(\hat{t}_k) = \sum_{t=t_{k-1}}^{\hat{t}_k - 1} \varepsilon_{i_{k-1}}(t)^2 + \sum_{t=\hat{t}_k}^{t_{k+1}-1} \varepsilon_{i_k}(t)^2. \quad (8)$$

$L^{(i_{k-1},i_k)}$ evaluates the model accuracy (according to the currently available estimated local models) in the interval from t_{k-1} to t_{k+1}, for an optimal positioning of the switching (in case of an active switching), assuming that the left sub-period is assigned to mode i_{k-1} and the right one to i_k. Accordingly, the optimal mode assignment for the two sub-periods is given by

$$(i_{k-1}^\star, i_k^\star) = \arg \min_{(i_{k-1},i_k) \in \hat{\Sigma}(t_{k-1}) \times \hat{\Sigma}(t_k)} L^{(i_{k-1},i_k)}.$$

Then, the mode assignments are reset to $\hat{\Sigma}(t_{k-1}) = \{i_{k-1}^\star\}$ and $\hat{\Sigma}(t_k) = \{i_k^\star\}$. Additionally, if $i_{k-1}^\star = i_k^\star$, t_k is evaluated an inactive switching, and therefore removed. Conversely, if $i_{k-1}^\star \neq i_k^\star$, the switching time t_k is corrected to

$$t_k^\star = \arg \min_{\hat{t}_k \in B(t_k)} \bar{L}^{(i_{k-1}^\star, i_k^\star)}(\hat{t}_k)$$

Algorithm 1 Switching refinement

Require: $\mathcal{T}_s, \boldsymbol{\eta}_k, k = 1, .., N_s, \boldsymbol{\vartheta}^i, i = 1, .., N_M, w$
Ensure: $\mathcal{T}_s^{new}, \hat{\boldsymbol{\sigma}}$

for all $t_k \in \mathcal{T}_s$ **do** ▷ Mode assignment
  $\hat{\Sigma}(t_k) \leftarrow \{i \in \{1, \dots, N_M\} \mid \eta_k^i > 0\}$;
end for
$\mathcal{T}_s^{new} \leftarrow \{t_1\}$;
for $k = 2$ **to** $|\mathcal{T}_s|$ **do** ▷ Switching refinement
  $L^\circ \leftarrow \infty$;
  $B \leftarrow [\max(1, t_k - \lfloor w/2 \rfloor), \min(t_k + \lfloor w/2 \rfloor, N)]$;
  for all $(i_{k-1}, i_k) \in \hat{\Sigma}(t_{k-1}) \times \hat{\Sigma}(t_k)$ **do**
    if $i_{k-1} = i_k$ **then** ▷ Inactive switching
      $L \leftarrow \sum_{t=t_{k-1}}^{t_{k+1}-1} \varepsilon_{i_{k-1}}(t)^2$;
    else ▷ Active switching
      $L \leftarrow \min_{\hat{t} \in B} \sum_{t=t_{k-1}}^{\hat{t}-1} \varepsilon_{i_{k-1}}(t)^2 + \sum_{t=\hat{t}}^{t_{k+1}-1} \varepsilon_{i_k}(t)^2$;
    end if
    if $L < L^\circ$ **then**
      $L^\circ \leftarrow L$; $i_{k-1}^\star \leftarrow i_{k-1}$; $i_k^\star \leftarrow i_k$;
    end if
  end for
  if $i_{k-1}^\star \neq i_k^\star$ **then**
    $t_k^\star \leftarrow \arg\min_{\hat{t} \in B} \sum_{t=t_{k-1}}^{\hat{t}-1} \varepsilon_{i_{k-1}^\star}(t)^2 + \sum_{t=\hat{t}}^{t_{k+1}-1} \varepsilon_{i_k^\star}(t)^2$;
    $\mathcal{T}_s^{new} \leftarrow \mathcal{T}_s^{new} \cup \{t_k^\star\}$;
    $\hat{\sigma}_{t_k} \leftarrow i_k^\star$;
  end if
  if $k = 2$ **then** $\hat{\sigma}_{t_1} \leftarrow i_{k-1}^\star$; **end if**
end for

and $t_k^\star$ is added to the refined set $\mathcal{T}_s^{new}$.

The proposed refinement algorithm is summarized in the pseudo-code Algorithm 1.

C. Detection and refinement of close switchings

A last check is carried out to detect and remove duplicate switchings that may result from the previous operation. Observe that the refinement performed by Algorithm 1 is approximate, due to the limited length of the optimization window and the imprecision of the estimated local models. Thus, it may sometimes happen that two subsequent switchings are refined towards the same position, generating an unacceptably small sub-period (*e.g.*, regarding to a minimum dwell time constraint). Close switching times of this type are merged into one, removing the intermediate sub-period.

More precisely, let $t_k, t_{k+1} \in \mathcal{T}_s^{new}$ be such that $t_{k+1} - t_k < w/2$. Then, if $\hat{\sigma}_{t_{k-1}} = \hat{\sigma}_{t_{k+1}}$, both switching locations are simply removed from $\mathcal{T}_s^{new}$. Otherwise, t_k is eliminated and the location of t_{k+1} is refined with the same method explained in the previous sub-section.

V. SIMULATION RESULTS

The SNARX system considered in [3] is employed here, which includes two modes.

$$\mathcal{M}_1^\circ: \quad y(t) = -0.905y(t-1) + 0.9u(t-1) + e(t)$$
$$\mathcal{M}_2^\circ: \quad y(t) = -0.4y(t-1)^2 + 0.5u(t-1) + e(t)$$

where y is the output, u is the input (uniformly distributed in [0,1]), and $e(t)$ is a white Gaussian noise with zero mean and variance 0.012. The data-set consists of $N = 2000$ samples, and includes 4 mode switchings starting from $\mathcal{M}_1^\circ$, at samples 370, 1420, 1520, and 1750.

Initially, the set of candidate switching times is defined as a uniform partition of the time horizon, *i.e.* $\mathcal{T}_s = \{t_k = 100k\}$, $k = 1, 2, ..., 19$, assuming a minimum dwell-time of 100 samples. The local models are quadratic polynomial NARX models with orders $n_y = n_u = 2$, which amounts to $n = 15$ possible regressors for each mode. Initially, the MEPs are set to $1/N_M$, N_M being the assumed number of modes, and the RIPs are all initialized by 0.2.

A. Single run results

We first illustrate how the proposed SNARX identification procedure works, with focus on the refinement of the switching times and the detection of redundant models. We assume that $N_M = 3$. Figure 1 graphically illustrates the evolution of the switching signal estimation during the algorithm iterations, both regarding the identification step and the refinement step. The first application of the identification step (see upper colored bar in Figure 1) shows that $\mathcal{M}_1$ and $\mathcal{M}_2$ are often associated to the same sub-periods, indicating a possible redundancy, which cannot yet be resolved, since $\mathcal{T}_s^1 \subset \mathcal{T}_s^2$ and the two models are not sufficiently similar to satisfy the t-test (they have different structure, as per Table I).

Next, the active switchings and the latent switching times in the mixed sub-periods are refined by Algorithm 1. Specifically, all possible mode switching patterns from I_{k-1} to I_k are tested and the best assignment is selected. Consider for example the first two sub-periods, I_1 (from 1 to 100) and I_2 (from 101 to 200): the value of the local fit index (8) as a function of the switching location in the region around 101 is shown in Figure 2 (left) for all possible mode patterns (both sub-periods can be assigned either to $\mathcal{M}_1$ or $\mathcal{M}_2$). The comparison clearly indicates that the assignment of both sub-periods to $\mathcal{M}_1$ yields the best fitting results. Consequently, the switching can be considered inactive and be removed. On the other hand, when we focus on t_{15}, all possible switching patterns indicate a switching around 1420 from $\mathcal{M}_3$ to either $\mathcal{M}_1$ or $\mathcal{M}_2$. Actually, the two alternatives yield almost equivalent results due to the similarity of the first two modes. The outcome of the refinement procedure is depicted in the second colored bar in Figure 1. Only 6 switching times remain (at 266, 370, 1420, 1570, 1751 and 1948), dividing the observation horizon in 7 sub-periods.

In the second iteration, the 3^{rd} and 5^{th} period are assigned to $\mathcal{M}_3$, whereas the remaining 5 sub-periods return a mixed result. However, since this time $\mathcal{T}_s^1 = \mathcal{T}_s^2$, one among $\mathcal{M}_1$ and $\mathcal{M}_2$ can be eliminated. Applying the refinement rule

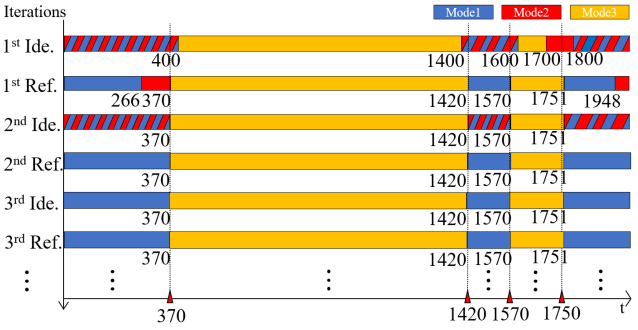

Fig. 1. Evolution of the mode assignment and the switching times along the iterations. Odd and even rows indicate the results of the identification and refinement steps, respectively. The mode assignment is graphically represented using a different color for each mode, as of the colormap on top. Striped regions indicate sub-periods associated to multiple modes.

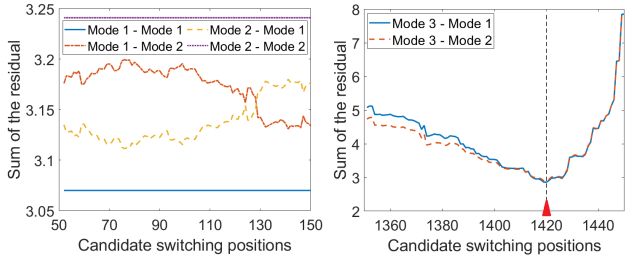

Fig. 2. Refinement of switching times t_2 (left) and t_{15} (right), according to Algorithm 1.

with the two remaining models yields the final switching times, namely 370, 1420, 1570, and 1751, and the algorithm has reached convergence.

TABLE I
MODEL STRUCTURE SELECTION OVER ITERATIONS

Iter.	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$
1	$\boldsymbol{y(t-1)}$, $\boldsymbol{u(t-1)}$	$\boldsymbol{y(t-1)}$, $\boldsymbol{u(t-1)}$, $u(t-2)$	$\boldsymbol{y(t-1)^2}$, $\boldsymbol{u(t-1)}$
2	$\boldsymbol{y(t-1)}$, $\boldsymbol{u(t-1)}$	$\boldsymbol{y(t-1)}$, $\boldsymbol{u(t-1)}$	$\boldsymbol{y(t-1)^2}$, $\boldsymbol{u(t-1)}$
3	$\boldsymbol{y(t-1)}$, $\boldsymbol{u(t-1)}$	-	$\boldsymbol{y(t-1)^2}$, $\boldsymbol{u(t-1)}$

B. Monte Carlo analysis

In this sub-section, a Monte Carlo (MC) analysis is carried out to analyze the performance of the proposed method. The analysis is repeated for different initial assumptions on the upper bound of the number of modes, N_M, to evaluate its robustness regarding the ability to estimate correctly this crucial parameter. Table II reports the aggregate results over 50 runs for each case. The model quality is assessed in terms of the classification error rate (percentage of misclassified samples) and the quality of fit, according to the normalized index

$$FIT = 100\left(1 - \frac{\sum_{t=1}^{N}\|y(t) - \hat{y}(t)\|_2}{\sum_{t=1}^{N}\|y(t) - \bar{y}\|_2}\right),$$

TABLE II

RESULTS OF THE MC ANALYSIS.

# of initial modes	3	4	5
Average # of iterations	2.56	2.74	2.82
FIT [mean (std.)]	74.3 (1.6)	74.3 (1.6)	74.3 (1.6)
Class. error [mean (std.)]	0.26 (0.18)	0.26 (0.18)	0.26 (0.18)
Corr. ident. of N_M	100%	100%	100%
Corr. sel. of $\boldsymbol{s}^{(1)}$	100%	100%	100%
Corr. sel. of $\boldsymbol{s}^{(2)}$	96%	96%	96%

where $\hat{y}(t)$ is the output predicted at time t by the estimated SNARX model and $\bar{y} = \frac{1}{N}\sum_{t=1}^{N} y(t)$ is the output average.

The proposed method performed well both regarding the detection of redundant modes and the estimation of the switching times. Indeed, the correct number of local models and switchings was estimated in all runs. Largely over-estimating the number of modes ultimately does not affect the quality of the results, though more iterations are required on average to recognize the mode redundancy. The switching locations are very accurately estimated, save from an occasional displacement from the ground truth positions by a few samples. Indeed, the refinement rule is based on a fitting condition, which is unavoidably subject to small bias due to the effect of noise. This, however, does not prevent the algorithm from achieving excellent classification accuracy overall. Regarding the estimation of the model structures, sometimes the nonlinearity in $\mathcal{M}_2^\circ$ is not captured in the identified sub-model, depending on the particular realization (see also the analysis in [2] on this aspect). Fortunately, this happens only 2 times out of 50. Compared to [3], the performance is equivalent both in terms of the FIT index and the classification error, although the efficiency is much higher, as discussed next.

C. Analysis of the computational load

The previous SNARX system is considered again here with a linearly increasing number of switching times, $N_s^\circ = 5, \ldots, 30$, namely $\mathcal{T}_s^\circ = \{1\} \cup \{150k + 1\}_{k=1,\ldots,N_s^\circ-1}$, accordingly extending the length of the observation window to $N = 150N_s^\circ$. The candidate switching times are randomly initialized from Gaussian distributions $t_k \sim \mathcal{N}(100k, 10)$ where $k = 1, 2, ..., \lfloor N/100 \rfloor - 1$. Batches of MC simulations were carried out with the identification approach in [3] and the refinement method proposed in the paper, in both cases assuming $N_M = 2$. As depicted in Fig. 3, the proposed refinement stage allows the proposed method to complete the task in a shorter time compared to [3], the difference increasing with the complexity of the problem (*i.e.*, as the number of switchings $|\mathcal{T}_s^\circ|$ is increased). This superiority derives from the new optimization approach, in which the switching times are locally refined at the price of a small amount of calculation. In addition, the new method does not require the introduction of further candidate switching times over the iterations, which obviously reduces the computational complexity.

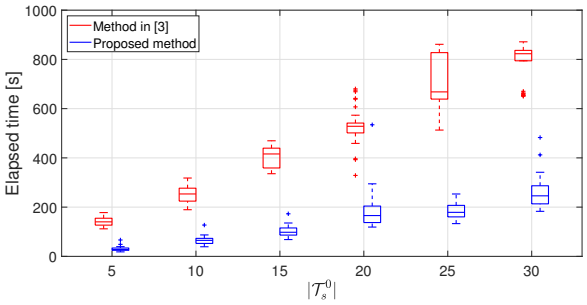

Fig. 3. Computational load analysis: total elapsed time of the method of [3] (red) and the proposed one (blue) for increasing number of switchings.

VI. CONCLUSION

The identification of SNARX models was addressed, without assuming prior knowledge on the number of modes, which is a crucial hypothesis in many approaches. The resulting algorithm alternates between an identification step, operating with fixed switching positions, and a refinement step, designed to revise the switching locations and detect redundant modes. A detailed analysis on a numerical example drawn from the literature indicates that the proposed algorithm provides accurate estimation of the number of modes, the switching locations, as well as the structure and parameters of the sub-models.

REFERENCES

[1] L. Bako, K. Boukharouba, and S. Lecoeuche. An l_0–l_1 norm based optimization procedure for the identification of switched nonlinear systems. In 49^{th} *IEEE Conference on Decision and Control*, pages 4467–4472, 2010.

[2] F. Bianchi, M. Prandini, and L. Piroddi. A randomized approach to switched nonlinear systems identification. In 18^{th} *IFAC Symposium on System Identification (SYSID 2018)*, volume 51 of *IFAC-PapersOnLine*, pages 281–286, Stockholm, Sweden, July 9-11 2018.

[3] F. Bianchi, M. Prandini, and L. Piroddi. A randomized two-stage iterative method for switched nonlinear systems identification. *Nonlinear Analysis: Hybrid Systems*, 35:1–23, February 2020.

[4] A. Garulli, S. Paoletti, and A. Vicino. A survey on switched and piecewise affine system identification. In 16^{th} *IFAC Symposium on System Identification*, pages 344–355, Brussels, Belgium, July 11-13 2012.

[5] R. Goebel, R. G. Sanfelice, and A. R. Teel. Hybrid dynamical systems. *IEEE control systems magazine*, 29(2):28–93, 2009.

[6] F. Lauer and G. Bloch. *Hybrid System Identification*. 2019.

[7] F. Lauer, G. Bloch, and R. Vidal. Nonlinear hybrid system identification with kernel models. In 49^{th} *IEEE Conference on Decision and Control*, pages 696–701, 2010.

[8] V. L. Le, G. Bloch, and F. Lauer. Reduced-size kernel models for nonlinear hybrid system identification. *IEEE Transactions on Neural Networks*, 22(12):2398–2405, 2011.

[9] V. L. Le, F. Lauer, L. Bako, and G. Bloch. Learning nonlinear hybrid systems: from sparse optimization to support vector regression. In *Proceedings of the* 16^{th} *International Conference on Hybrid systems: computation and control*, pages 33–42, 2013.

[10] I. Leontaritis and S. Billings. Input-output parametric models for non-linear systems part I: deterministic non-linear systems. *International Journal of Control*, 41(2):303–328, 1985.

[11] I. Leontaritis and S. Billings. Input-output parametric models for non-linear systems part II: stochastic non-linear systems. *International Journal of Control*, 41(2):329–344, 1985.

[12] S. Paoletti, A. L. Juloski, G. Ferrari-Trecate, and R. Vidal. Identification of hybrid systems a tutorial. *European Journal of Control*, 13(2–3):242–260, 2007.