Recursive Bias-Correction Method for Identification of Piecewise Affine Output-Error Models

Manas Mejari, Valentina Breschi and Dario Piga

***Abstract*—Learning *PieceWise Affine Output-Error* (PWA-OE) models from data requires to estimate a finite set of affine output-error sub-models as well as a partition of the regressors space over which the sub-models are defined. For an output-error type noise structure, the algorithms based on ordinary least squares (LS) fail to compute a consistent estimate of the sub-model parameters. On the other hand, the prediction error methods (PEMs) provide a consistent parameter estimate, however, they require to solve a non-convex optimization problem for which the numerical algorithms may get trapped in a local minimum, leading to inaccurate estimates. In this paper, we propose a recursive bias-correction scheme for identifying PWA-OE models, retaining the computational efficiency of the standard LS algorithms while providing a consistent estimate of the sub-model parameters, under suitable assumptions. The proposed approach allows one to recursively update the estimates of the sub-models parameters and to cluster the regressors. Linear multi-category techniques are then employed to estimate a partition of the regressor space based on the estimated clusters. The performance of the proposed algorithm is demonstrated via an academic example.**

***Index Terms*—Identification; Switched systems.**

I. INTRODUCTION

GIVEN their universal approximation properties [5], *PieceWise Affine* (PWA) maps can be used to describe the behavior of many nonlinear dynamical systems. Moreover, due to the equivalence between PWA and *hybrid* linear models, tools developed for analysis and control of hybrid systems can be utilized for systems represented in PWA form [4].

The problem of identifying hybrid systems and, in particular, PWA models, involves the estimation of the parameters defining the affine sub-models as well as the discrete mode sequence representing the active sub-model at each time step. This is known to be an NP-hard problem [12]. The heuristics developed over the years to solve this problem include set-membership approach [3], which imposes that the identification error is bounded by a chosen parameter under bounded noise assumption; sparse optimization algorithms [18], [20] which formulate an over-parametrized least-squares problem with a regularization term penalizing switching between sub-models; mixed-integer programming methods [17], [22] which aim at global optimal solution by solving a mixed-integer program; and clustering-based algorithms [1], [2], [6], [9], that involve simultaneous clustering of data and estimation of model parameters, and for PWA models, an additional stage to compute a polyhedral partition of the input domain using linear classification techniques.

The underlying assumption in these approaches is that the output is generated by *AutoRegressive with eXogenous inputs* (ARX) local models, allowing one to use ordinary *Least Squares* (LS) to retrieve a *consistent* estimate of the sub-models parameters, provided that the correct mode sequence has been reconstructed. This hypothesis might be too restrictive in practice, requiring a more general noise model structure for the affine sub-models. In this paper, we relax the ARX modeling assumption on the affine sub-models and consider a more general PWA *output-error* (PWA-OE) structure. The choice of PWA-OE model makes the identification problem more challenging, as the inputs of the PWA map depend on the unmeasured "noise-free" regressor.

Very few works in the literature have addressed the identification of PWA-OE models. A *Prediction Error Method* (PEM) is proposed in [23] for identifying piecewise linear models under the assumption that the discrete mode sequence is known, while [7] considers more general Box-Jenkins switching models, along with the estimation of the hidden mode sequence. The PEM approach leads to a non-convex optimization problem which thus requires an accurate initial guess so that the numerical algorithm is not trapped in a local minimum. This limitation is overcome in [8], where an *Instrumental Variable* (IV) scheme has been proposed. However, a correct choice of IVs is very critical as they need to be uncorrelated with the system noise and correlated with the inputs. Alternatively, in [11], the PWA-OE identification problem is tackled in an iterative fashion within a Bayesian framework. The devised algorithm, however, is suboptimal.

In this work, we propose a *recursive bias-correction* approach, combined with a clustering based algorithm for the identification of PWA-OE models. The main idea of bias-correction methods is to remove the bias from the LS estimates in order to obtain a consistent estimate of the model parameters. Bias-correction methods have been proposed in the literature for the identification of LTI systems [10], [24], LPV models [16], [19], and nonlinear systems [21]. For the identification of PWA-OE models a *batch* bias-correction approach is recently proposed by authors in [14]. This paper extends the ideas presented in [14] as follows: (i) in [14], the bias-corrected estimates are computed based on the assumption that the variance of the noise is known, which is seldom true in practice. In this work, we relax this assumption and present an algorithm to estimate the noise variance from data; (ii)

This work was partially supported by the European H2020-CS2 project ADMITTED, Grant agreement no. GA832003.

M. Mejari and D. Piga are with IDSIA Dalle Molle Institute for Artificial Intelligence, SUPSI-USI, Via Cantonale 2C, 6928 Manno, Switzerland. {manas.mejari,dario.piga}@supsi.ch

V. Breschi is with Dipartimento di Elettronica e Informazione, Politecnico di Milano, 20133 Milan, Italy. valentina.breschi@polimi.it

a recursive formulation for computing the bias-corrected LS estimates is presented, which makes the algorithm tailored for *on-line* identification and suitable to handle large dataset.

The paper is organized as follows. The identification problem of PWA-OE models is formalized in §II. Bias-corrected least-squares estimates for the affine sub-models parameters are derived in §III and its recursive formulation with noise variance estimation is presented in §IV. §V presents a two-stage recursive algorithm for simultaneous estimation of the submodel parameters and the unknown mode sequence in the first stage and for computing a partition of the regresor space in the second stage. A numerical case study is reported in §VI. Conclusions and directions for future work are drawn in §VII.

II. PROBLEM FORMULATION

We consider the following discrete-time, single-input single-output *PieceWise Affine* (PWA) data-generating system $\mathcal{S}_\text{o}$, with *Output-Error* (OE) structure,

$$y_\text{o}(k) = f(x_\text{o}(k)), \tag{1a}$$

$$y(k) = y_\text{o}(k) + e_\text{o}(k), \tag{1b}$$

where $y_\text{o}(k) \in \mathbb{R}$ and $y(k) \in \mathbb{R}$ are the *noise-free* and noise-corrupted output of the system at time $k \in \mathbb{N}$ respectively, $x_\text{o}(k) \in \mathcal{X} \subseteq \mathbb{R}^{n_x}$ denotes the *noise-free regressor*, which depends on the past n_a values of the output y_o, current and past n_b values of the input u, namely,

$$x_\text{o}(k) = [y_\text{o}(k-1) \cdots y_\text{o}(k-n_a)\ u(k)\ u(k-1) \cdots u(k-n_b)]^\top \tag{1c}$$

and $e_\text{o}(k) \in \mathbb{R}$ is a zero-mean additive white Gaussian noise with variance σ_e^2. The PWA map $f : \mathcal{X} \to \mathbb{R}$ is defined as

$$f(x_\text{o}) = \begin{cases} (\theta_1^\text{o})^\top \begin{bmatrix} x_\text{o} \\ 1 \end{bmatrix} & \text{if } x_\text{o} \in \mathcal{X}_1, \\ \vdots \\ (\theta_s^\text{o})^\top \begin{bmatrix} x_\text{o} \\ 1 \end{bmatrix} & \text{if } x_\text{o} \in \mathcal{X}_s, \end{cases} \tag{1d}$$

where $s \in \mathbb{N}$ is the number of *modes* (*i.e.*, the number of affine functions defining f), $\theta_i^\text{o} \in \mathbb{R}^{(n_x+1)\times 1}$ is the parameter vector associated to the i-th affine submodel, and the set $\mathcal{X}_i \subseteq \mathcal{X}$ is a polyhedron defined as:

$$\mathcal{X}_i \doteq \{x_\text{o} \in \mathbb{R}^{n_x} :\ \mathcal{H}_i x_\text{o} \le \mathcal{D}_i\}, \tag{1e}$$

with $\{\mathcal{H}_i\}_{i=1}^s$ and $\{\mathcal{D}_i\}_{i=1}^s$ being real matrices. The polyhedra $\{\mathcal{X}_i\}_{i=1}^s$ form a *complete* polyhedral partition[1] of the regressor space $\mathcal{X}$. Note that, unlike PWA-ARX models, the polyhedral partition $\{\mathcal{X}_i\}_{i=1}^s$ is defined over the space of *noise-free* regressors.

In order to describe the true PWA-OE system $\mathcal{S}_\text{o}$ in (1), the following parameterized model structure $\mathcal{M}_\theta$ is introduced,

$$y(k) = \begin{cases} \theta_1^\top \begin{bmatrix} x(k) \\ 1 \end{bmatrix} + \epsilon(k), & \text{if } x(k) \in \mathcal{X}_1, \\ \vdots \\ \theta_s^\top \begin{bmatrix} x(k) \\ 1 \end{bmatrix} + \epsilon(k), & \text{if } x(k) \in \mathcal{X}_s, \end{cases} \tag{2a}$$

where $\epsilon(k)$ is the residual term (not necessarily a white noise as in PWA-ARX structures) modelling the mismatch between true system and model output, $x(k)$ is the regressor vector constructed from past measured outputs and inputs, *i.e.,*

$$x(k) = [y(k-1) \cdots y(k-n_a)\ u(k)\ u(k-1) \cdots u(k-n_b)]^\top \tag{2b}$$

The identification problem addressed in this paper is formalized as follows,

Problem 1: Given a set of N input-output observations $\{u(k), y(k)\}_{k=1}^N$, generated by the system $\mathcal{S}_\text{o}$ in (1), compute a *consistent* estimate of the true parameters $\{\theta_i^\text{o}\}_{i=1}^s$ characterizing the affine sub-models of the PWA map f, and find a polyhedral partition $\{\mathcal{X}_i\}_{i=1}^s$ of the regressor space $\mathcal{X}$, over which the local affine sub-models are defined.

To this aim, a novel *recursive* identification algorithm based on *bias-corrected* LS is presented in the next sections. The main assumptions used in the rest of the paper are as follows:

1. The parameters n_a, n_b and s are known, which are shared by the true system $\mathcal{S}_\text{o}$ and the model $\mathcal{M}_\theta$.
2. The output y_o and the input u are bounded and statistically independent of the noise e_o.
3. The input u is such that all s modes are sufficiently excited, *i.e.*, sufficient number of data samples N_i are available for each mode and $N \to \infty$ also implies that $N_i \to \infty$ for $i = 1, \dots s$.

III. BIAS-CORRECTED LEAST SQUARES ESTIMATES FOR PWA-OE MODELS

For an output-error model structure, ordinary least squares give an asymptotically biased estimate of the model parameters [13]. The bias is due to the correlation between measured regressors $x(k)$ and $\epsilon(k)$ which is not a white noise. In this section, we quantify the bias in the LS estimates and show how to eliminate it.

Let us define the *active* mode $\sigma(k) \in \{1, \cdots s\}$, as

$$\sigma(k) = i \Leftrightarrow x_\text{o}(k) \in \mathcal{X}_i,$$

which represents the partition to which the regressor $x_\text{o}(k)$ belongs to at time k. Suppose, for now, that the sequence of active modes $\{\sigma(k)\}_{k=1}^N$ is known/estimated [2]. Let N_i be the number of regressor/ouput data points associated to the i-th mode, with $\sum_{i=1}^s N_i = N$. For a given mode sequence, let us define $\mathbb{Y}_i \in \mathbb{R}^{N_i}$ as the output vector constructed by stacking all the outputs associated with the i-th mode, namely,

$$y(k) \text{ is a row of } \mathbb{Y}_i \Leftrightarrow \sigma(k) = i,$$

and, analogously, let $\mathbb{X}_i \in \mathbb{R}^{N_i\times(n_x+1)}$ (with $n_x = n_a + n_b + 1$) be the regressor matrix constructed from the sequence $\{x(k)\}_{k=1}^N$ by stacking the extended regressors associated to the i-th affine submodel, *i.e.*, $\left[\begin{smallmatrix} x^\top(k) & 1 \end{smallmatrix}\right]^\top$ is a row of $\mathbb{X}_i \Leftrightarrow \sigma(k) = i$. Similarly, we define the noise-free output vector $\mathbb{Y}_{i,\text{o}} \in \mathbb{R}^{N_i}$, the noise-free regressor matrix $\mathbb{X}_{i,\text{o}} \in \mathbb{R}^{N_i\times(n_x+1)}$ and the measurement noise vector $\mathbb{E}_{i,\text{o}} \in \mathbb{R}^{N_i}$, by stacking the noise-free outputs, noise-free extended regressors and the samples of the measurement noise e_o associated to the i-th mode, respectively.

[1] The collection $\{\mathcal{X}_i\}_{i=1}^s$ is a complete partition of $\mathcal{X}$ if $\bigcup_{i=1}^s \mathcal{X}_i = \mathcal{X}$ and $\mathring{\mathcal{X}}_i \cap \mathring{\mathcal{X}}_j = \emptyset$, $\forall i \neq j$, with $\mathring{\mathcal{X}}_i$ denoting the interior of $\mathcal{X}_i$.

[2] The algorithm to estimate the mode sequence is presented in Section V.

A. Computation of the bias in the least squares estimate

The LS estimate of the i-th sub-model parameter is given by,

$$\theta_i^{\mathrm{LS}} = \underbrace{\left(\frac{\mathbb{X}_i^\top \mathbb{X}_i}{N_i}\right)^{-1}}_{\Gamma_{N_i}} \frac{\mathbb{X}_i^\top \mathbb{Y}_i}{N_i}, \tag{3}$$

where Γ_{N_i} is assumed to be invertible. The difference between the LS estimate θ_i^{LS} and the true parameter θ_i^{o} is expressed as follows [14]:

$$\theta_i^{\mathrm{LS}} - \theta_i^{\mathrm{o}} = \underbrace{\left(\frac{\mathbb{X}_i^\top \mathbb{X}_i}{N_i}\right)^{-1} \frac{\mathbb{X}_i^\top \Delta\mathbb{X}_i}{N_i}\theta_i^{\mathrm{o}}}_{B_\Delta(\theta_i^{\mathrm{o}}, \mathbb{X}_i, \Delta\mathbb{X}_i)} + \underbrace{\left(\frac{\mathbb{X}_i^\top \mathbb{X}_i}{N_i}\right)^{-1} \frac{\mathbb{X}_i^\top \mathbb{E}_{i,\mathrm{o}}}{N_i}}_{B_{e_\mathrm{o}}}, \tag{4}$$

where $\Delta\mathbb{X}_i$ is defined as the difference between noise-free and noisy regressor matrices, namely,

$$\Delta\mathbb{X}_i = \mathbb{X}_{i,\mathrm{o}} - \mathbb{X}_i. \tag{5}$$

Since the measurement noise e_o is assumed to be zero-mean white noise statistically independent of u, and since the regressor $x(k)$ depends only on *past* outputs, the noise vector $\mathbb{E}_{i,\mathrm{o}}$ is uncorrelated with the regressors $\mathbb{X}_i$. Thus, the term B_{e_o} in (4) asymptotically (as $N_i \to \infty$) converges to 0 with probability 1 (w.p. 1). However, the term $B_\Delta(\theta_i^{\mathrm{o}}, \mathbb{X}_i, \Delta\mathbb{X}_i)$ is not guaranteed to converge to 0 in general, inducing a non-zero bias in the LS estimate. This proves that the LS estimate θ_i^{LS} in (3) is not consistent, *i.e.*, $\lim_{N_i \to \infty} \theta_i^{\mathrm{LS}} \neq \theta_i^{\mathrm{o}}$, even for the known true mode sequence. Note that, the bias due to a possible mismatch between the mode associated with $x(k)$ and $x_\mathrm{o}(k)$ is not quantified in (4). Nonetheless, the proposed recursive algorithm (presented in Section V) provides a heuristic to reduce the effect of this additional bias.

B. Bias elimination in the least squares estimate

In order to eliminate the bias in LS estimates, we define the corrected LS estimate θ_i^{CLS} as follows:

$$\theta_i^{\mathrm{CLS}} = \theta_i^{\mathrm{LS}} - B_\Delta(\theta_i^{\mathrm{CLS}}, \mathbb{X}_i, \Delta\mathbb{X}_i), \tag{6}$$

where,

$$B_\Delta(\theta_i^{\mathrm{CLS}}, \mathbb{X}_i, \Delta\mathbb{X}_i) = \left(\frac{\mathbb{X}_i^\top \mathbb{X}_i}{N_i}\right)^{-1} \frac{\mathbb{X}_i^\top \Delta\mathbb{X}_i}{N_i}\theta_i^{\mathrm{CLS}} \tag{7}$$

The main idea behind the definition of θ_i^{CLS} in (6) is that, the LS estimate is corrected by eliminating the bias term B_Δ, which is evaluated at the estimate θ_i^{CLS}, rather than at the *unknown* true model parameter θ_i^{o}. By substituting equations (3) and (7) in (6) and with simple algebraic manipulations, we obtain

$$\theta_i^{\mathrm{CLS}} = \left(\frac{\mathbb{X}_i^\top \mathbb{X}_i + \mathbb{X}_i^\top \Delta\mathbb{X}_i}{N_i}\right)^{-1} \frac{\mathbb{X}_i^\top \mathbb{Y}_i}{N_i}. \tag{8}$$

Note that, since $\Delta\mathbb{X}_i$ depends on the unmeasured noise-free regressors $\mathbb{X}_{i,\mathrm{o}}$, the estimates θ_i^{CLS} in (8) can not be computed based on the available input/output measurements. To overcome this problem, the term $\mathbb{X}_i^\top \Delta\mathbb{X}_i$ in (8) is replaced by a *bias-eliminating matrix* Ψ_i, which is constructed in such a way that it can be computed from the available information and it satisfies the following property:

$$\mathbf{C1}: \lim_{N_i \to \infty} \frac{1}{N_i} \mathbb{X}_i^\top \Delta\mathbb{X}_i = \lim_{N_i \to \infty} \frac{1}{N_i} \Psi_i \quad \text{w.p. } 1.$$

A bias-eliminating matrix Ψ_i which satisfies condition $\mathbf{C1}$, can be constructed by evaluating the expected value of the matrix $E\left\{\mathbb{X}_i^\top \Delta\mathbb{X}_i\right\}$, such that $E\left\{\Psi_i\right\} = E\left\{\mathbb{X}_i^\top \Delta\mathbb{X}_i\right\}$ is satisfied by construction.

Property 1: The bias-eliminating matrix Ψ_i is given by,

$$\Psi_i = -\sigma_e^2 N_i \underbrace{\left[\begin{array}{c|c} I_{n_a \times n_a} & \mathbf{0}_{n_a \times (n_b+2)}, \\ \hline \mathbf{0}_{(n_b+2) \times n_a} & \mathbf{0}_{(n_b+2) \times (n_b+2)}, \end{array}\right]}_{J} \tag{9}$$

where σ_e^2 is the variance of the measurement noise e_o and N_i is the number of training data points associated with the i-th mode. The matrix Ψ_i in (9) satisfies condition $\mathbf{C1}$.
See [15, Property 1] for the proof.

C. Bias-corrected LS estimate

The bias-corrected LS estimate θ_i^{BC} is thus given by,

$$\theta_i^{\mathrm{BC}} = \theta_i^{\mathrm{LS}} - B_\Delta(\theta_i^{\mathrm{BC}}, \Psi_i(N_i)), \tag{10}$$

where

$$B_\Delta(\theta_i^{\mathrm{BC}}, \Psi_i(N_i)) = \left(\frac{\mathbb{X}_i^\top \mathbb{X}_i}{N_i}\right)^{-1} \frac{\Psi_i}{N_i}\theta_i^{\mathrm{BC}}, \tag{11}$$

i.e., the bias-corrected LS estimate is obtained by replacing $\mathbb{X}_i^\top \Delta\mathbb{X}_i$ in (8) with the bias-eliminating matrix Ψ_i (eq. (9)),

$$\theta_i^{\mathrm{BC}} = \left(\frac{\mathbb{X}_i^\top \mathbb{X}_i + \Psi_i}{N_i}\right)^{-1} \frac{\mathbb{X}_i^\top \mathbb{Y}_i}{N_i}. \tag{12}$$

Property 2: Under the assumption that the limit $\lim_{N_i \to \infty} \left(\frac{\mathbb{X}_i^\top \mathbb{X}_i + \Psi_i}{N_i}\right)^{-1}$ exists, the bias-corrected estimate θ_i^{BC} in (12) is a *consistent* estimate of the true model parameter θ_i^{o}, *i.e.*, $\lim_{N_i \to \infty} \theta_i^{\mathrm{BC}} = \theta_i^{\mathrm{o}}$, w.p. 1.
See [15, Property 2] for the proof.

IV. RECURSIVE BIAS-CORRECTED LEAST SQUARES ESTIMATES

Based on the results summarized in the previous section, we derive a recursive formulation for the bias-corrected LS estimate (12) and the noise variance estimate $\hat{\sigma}_e^2$.

Let the active mode at time instance k be i, *i.e.*, $\sigma(k) = i$. Note that, the bias-corrected estimate for the i-th mode θ_i^{BC} is a fixed-point of (10). Thus, at time k, $\theta_i^{\mathrm{BC}}(k)$ can be computed by solving (10) recursively as follows:

$$\theta_i^{\mathrm{BC}}(k) = \theta_i^{\mathrm{LS}}(k) - B_\Delta(\theta_i^{\mathrm{BC}}(\tau), \Psi_i(\kappa_i)), \tag{13}$$

where $\theta_i^{\mathrm{LS}}(k)$ denotes the LS estimate of the i-th sub-model at time k, $\theta_i^{\mathrm{BC}}(\tau)$ denotes the bias-corrected LS estimate computed at the previous time instance τ at which the active mode was i, *i.e.*, $\sigma(\tau) = i$. Note that, τ denotes the *latest* time instance, previous to $k > \tau$ at which the active mode was i, *i.e.*, $\{\nexists\, t : \sigma(t) = i, \tau < t < k\}$, and κ_i denotes the total number of data points associated to i-th mode upto time

k, *i.e.*, $\kappa_i = \sum_{t=1}^{k} \mathbf{1}_{\{\sigma(t)=i\}}$, with $\mathbf{1}_{\{\cdot\}}$ denoting the indicator function.

By substituting the bias-correcting matrix $\Psi_i(\kappa_i) = -\sigma_e^2 \kappa_i J$ (see (9)) and the bias term $B_\Delta(\theta_i^{\mathrm{BC}}(\tau), \Psi_i(\kappa_i)) = \left(\sum_{t=1}^{k} \left(\left[\begin{smallmatrix} x(t) \\ 1 \end{smallmatrix}\right] \left[\begin{smallmatrix} x^\top(t) & 1 \end{smallmatrix}\right] \right) \mathbf{1}_{\{\sigma(t)=i\}} \right)^{-1} \Psi_i(\kappa_i)\theta_i^{\mathrm{BC}}(\tau)$ (see (11)) in (13), we obtain the following recursive update for $\theta_i^{\mathrm{BC}}(k)$ at time k:

$$\theta_i^{\mathrm{BC}}(k) = \theta_i^{\mathrm{LS}}(k) + \sigma_e^2 \kappa_i P_i(k) J \theta_i^{\mathrm{BC}}(\tau), \tag{14}$$

where $P_i(k) = \left(\sum_{t=1}^{k} \left(\left[\begin{smallmatrix} x(t) \\ 1 \end{smallmatrix}\right] \left[\begin{smallmatrix} x^\top(t) & 1 \end{smallmatrix}\right] \right) \mathbf{1}_{\{\sigma(t)=i\}} \right)^{-1}$ denotes the matrix computed at time k, using the regressors associated with the i-th sub-model.

The LS estimate $\theta_i^{\mathrm{LS}}(k)$ at time k can be updated using standard recursive least squares as follows:

$$\begin{aligned} \theta_i^{\mathrm{LS}}(k) &= \theta_i^{\mathrm{LS}}(\tau) + P_i(k) \left[\begin{smallmatrix} x(k) \\ 1 \end{smallmatrix}\right] \left(y(k) - \left[\begin{smallmatrix} x^\top(k) & 1 \end{smallmatrix}\right] \theta_i^{\mathrm{LS}}(\tau) \right), \\ P_i(k) &= P_i(\tau) - \frac{P_i(\tau) \left[\begin{smallmatrix} x(k) \\ 1 \end{smallmatrix}\right] \left[\begin{smallmatrix} x^\top(k) & 1 \end{smallmatrix}\right] P(\tau)}{1 + \left[\begin{smallmatrix} x^\top(k) & 1 \end{smallmatrix}\right] P_i(\tau) \left[\begin{smallmatrix} x(k) \\ 1 \end{smallmatrix}\right]}. \end{aligned} \tag{15}$$

A. Estimation of the noise variance

The bias corrected estimate in (14) depends on the noise variance σ_e^2, which is often unknown in practice and thus, it has to be estimated from data. At each time instance k, the estimate of the noise variance $\hat{\sigma}_e^2(k)$ can be computed as:

$$\begin{aligned} \hat{\sigma}_e^2(k) &= \frac{1}{k} \sum_{t=1}^{k} \left(y(t) - \hat{y}_{\mathrm{o}}(t) \right)^2 \\ &= \frac{1}{k} \left(y(k) - \hat{y}_{\mathrm{o}}(k) \right)^2 + \frac{k-1}{k} \hat{\sigma}_e^2(k-1), \end{aligned} \tag{16}$$

where, $\hat{y}_{\mathrm{o}}(k) = (\theta_{\sigma(k)}^{\mathrm{BC}})^\top \left[\begin{smallmatrix} \hat{x}_{\mathrm{o}}(k) \\ 1 \end{smallmatrix}\right]$ denotes the noise-free simulated one-step ahead output of the PWA-OE model and $\hat{\sigma}_e^2(k-1)$ is the estimate of the noise variance computed at the previous step $k-1$. Note that, the simulated output $\hat{y}_{\mathrm{o}}(k)$ at time k depends on the *active* mode $\sigma(k)$, the reconstructed noise-free regressor $\hat{x}_{\mathrm{o}}(k)$ and the corresponding active sub-model parameter $\theta_{\sigma(k)}^{\mathrm{BC}}$, which are estimated by the recursive algorithm presented in the next section. Thus, based on the estimate of the noise-variance (16), the bias-corrected LS estimate at time k is computed as follows:

$$\theta_i^{\mathrm{BC}}(k) = \theta_i^{\mathrm{LS}}(k) + \hat{\sigma}_e^2(k-1) \kappa_i P_i(k) J \theta_i^{\mathrm{BC}}(\tau). \tag{17}$$

V. PWA-OE IDENTIFICATION ALGORITHM

The recursive bias-corrected estimates and the noise variance estimate derived in the previous section are embedded into a recursive clustering-based algorithm for estimating the underlying discrete mode sequence in order to identify the overall PWA-OE model.

A. Recursive clustering and parameter estimation

Algorithm 1 summarizes the main ideas of the proposed approach which involves the computation of the bias-corrected estimates of the affine sub-model parameters $\{\theta_i^{\mathrm{BC}}\}_{i=1}^{s}$, the noise variance estimate $\hat{\sigma}_e^2$ as well as the unknown mode sequence $\{\sigma(k)\}_{k=1}^{N}$ and the clusters $\{\mathcal{C}_i\}_{i=1}^{s}$ characterizing the regressor space partition. The cluster $\mathcal{C}_i$ is constructed by stacking all (estimated) noise-free regressors $\hat{x}_{\mathrm{o}}(k)$ associated to i-th mode and its centroid c_i is defined as $c_i = \frac{1}{N_i} \sum_{\hat{x}_{\mathrm{o}}(k) \in \mathcal{C}_i} \hat{x}_{\mathrm{o}}(k)$.

For each time index $k = \max(n_a, n_b) + 1, \cdots, N$, the simulated noise-free regressor $\hat{x}_{\mathrm{o}}(k)$ is computed based on the simulated noise-free outputs $\hat{y}_{\mathrm{o}}(k-1), \cdots, \hat{y}_{\mathrm{o}}(k-n_a)$ (Step 4.1), with $\hat{y}_{\mathrm{o}}(k) = y(k)$ for $k = 1, \cdots, \max(n_a, n_b)$. At Step 4.2, the prediction-error $e_i(k)$ is computed for each mode $i = 1, \cdots, s$. Note that, $e_i(k)$ is the output-error computed based on the bias-corrected estimate θ_i^{BC} and $\hat{x}_{\mathrm{o}}(k)$. Step 4.3 selects the *best* active mode $\sigma(k)$ to which the regressor $\hat{x}_{\mathrm{o}}(x)$ is associated with. In particular, the clustering criterion for choosing the *best* mode at Step 4.3, minimizes the prediction-error $e_i(k)$ and the distance between $\hat{x}_{\mathrm{o}}(k)$ and the centroid c_i of the cluster $\mathcal{C}_i$. A positive hyper-parameter λ is used to weigh these two terms. Note that, by using the estimated noise-free $\hat{x}_{\mathrm{o}}(k)$, we aim at reducing the effect of a possible mismatch between the mode associated with $x(k)$ and $x_{\mathrm{o}}(k)$. Based on the active mode $\sigma(k)$ selected at Step 4.3, the regressors $\hat{x}_{\mathrm{o}}(k)$ is assigned to cluster $\mathcal{C}_{\sigma(k)}$ at Step 4.4 whose cardinality $N_{\sigma(k)}$ is increased by 1 (Step 4.5). The matrix $P_{\sigma(k)}$ and LS estimate $\theta_{\sigma(k)}^{\mathrm{LS}}$ are then updated according to (15) at Steps 4.6 and 4.7 respectively. The bias-corrected LS estimate $\theta_{\sigma(k)}^{\mathrm{BC}}$ is updated at Step 4.8 (see 17) and the output $\hat{y}_{\mathrm{o}}(k)$ (used to construct the regressor $\hat{x}_{\mathrm{o}}$) is computed at Step 4.9. Based on the output $\hat{y}_{\mathrm{o}}(k)$, the estimate of the noise variance $\hat{\sigma}_e^2(k)$ is updated at Step 4.10. The centroid $c_{\sigma(k)}$ is finally updated at Step 4.11.

B. Partitioning the regressor space

Given the clusters $\{\mathcal{C}_i\}_{i=1}^{s}$ obtained from Algorithm 1, the partition $\{\mathcal{X}_i\}_{i=1}^{s}$ of the regressor space $\mathcal{X}$ can be computed using the linear multicategory discrimination algorithm proposed in [6]. To this end, we search for a piecewise-affine separator function $\phi : \mathbb{R}^{n_x} \to \mathbb{R}$ defined as

$$\phi(\hat{x}) = \max_{i=1,\ldots,s} \left(\left[\begin{smallmatrix} \hat{x}_{\mathrm{o}}^\top & -1 \end{smallmatrix}\right] \left[\begin{smallmatrix} \omega^i \\ \gamma^i \end{smallmatrix}\right] \right), \tag{18}$$

where $\omega^i \in \mathbb{R}^{n_x}$ and $\gamma^i \in \mathbb{R}$ are unknown parameters to be computed. Let N_i denote the cardinality of the i-th cluster $\mathcal{C}_i$ and $M_i \in \mathbb{R}^{N_i \times n_x}$ be the matrix obtained by stacking the regressors $\hat{x}_{\mathrm{o}}^\top(k)$ belonging to $\mathcal{C}_i$ in its rows.

The parameters $\{\omega^i, \gamma^i\}_{i=1}^{s}$, are computed via the solution of the following convex optimization problem [6]:

$$\begin{aligned} \min_{\xi} \ & \frac{\kappa}{2} \sum_{i=1}^{s} \left(\|\omega^i\|_2^2 + (\gamma^i)^2 \right) + \\ & \sum_{i=1}^{s} \sum_{\substack{j=1 \\ j \neq i}}^{s} \frac{1}{N_i} \left\| \left(\left[\begin{smallmatrix} M_i & -\mathbf{1}_{N_i} \end{smallmatrix}\right] \left[\begin{smallmatrix} \omega^j - \omega^i \\ \gamma^j - \gamma^i \end{smallmatrix}\right] + \mathbf{1}_{N_i} \right)_+ \right\|_2^2, \end{aligned} \tag{19}$$

where ξ is the set of optimization variables, *i.e.*, $\xi = \left[\begin{smallmatrix} (\omega^1)^\top & \ldots & (\omega^s)^\top & \gamma^1 & \ldots & \gamma^s \end{smallmatrix}\right]^\top$, and for a given $x \in \mathbb{R}^n$, $(x)_+$ denotes a vector whose i-th element is $\max\{x_i, 0\}$. This problem can be solved recursively via [6, Algorithm 3] while Algorithm 1 is run.

VI. NUMERICAL EXAMPLE

The effectiveness of the proposed identification algorithm is demonstrated via a numerical example. All computations are carried out on an i7 1.9-GHz Intel core processor with 32 GB of RAM running MATLAB R2019a.

The following data-generating system is considered [14],

$$y_{\rm o}(k)=\begin{cases}\left[\,-0.4\ \ 1\ \ 1.5\,\right] x_{\rm o}(k), \text{ if } 4y_{\rm o}(k-1)-u(k-1)+10<0,\\ \left[\,0.5\ \ -1\ \ -1.5\,\right] x_{\rm o}(k), \text{ if } 4y_{\rm o}(k-1)-u(k-1)+10<0,\\ \qquad\qquad \&5y_{\rm o}(k-1)+u(k-1)-6\leq 0,\\ \left[\,-0.3\ \ 0.5\ \ -1.7\,\right] x_{\rm o}(k), \text{if } 5y_{\rm o}(k-1)+u(k-1)-6>0,\end{cases}$$

$$y(k)=y_{\rm o}(k)+e_{\rm o}(k), \tag{20}$$

characterized by $s=3$ discrete modes. The training dataset consist of $N=5000$ input/output samples gathered from the system (20), with the input signal u generated from a uniform random distribution taking values in the interval $[-4,4]$. The noise $e_{\rm o}$ corrupting the output signal is generated by a zero-mean white Gaussian process with variance $\sigma_e^2=0.64$, which corresponds to the *Signal-to-Noise Ratio* $\text{SNR}=10\log\frac{\sum_{k=1}^N y_{\rm o}^2(k)}{\sum_{k=1}^N e_{\rm o}^2(k)}=11.7$ dB.

Algorithm 1 Recursive bias-corrected parameter estimation and clustering.

Input: Observations $\{x(k),y(k)\}_{k=1}^N$; number of modes s; model orders n_a,n_b; tuning parameter λ; initial guess of the model parameters $\{\theta_i^{\rm BC},\theta_i^{\rm LS}\}_{i=1}^s$; initial noise variance estimate $\hat{\sigma}_e^2$; initial cluster centroids $\{c_i\}_{i=1}^s$.

1. **set** $\mathcal{C}_i=\emptyset$, $N_i=0$, $\forall i=1,\dots,s$;
2. **set** $P_i=I_{n_a+n_b+2}$, $\forall i=1,\dots,s$;
3. **set** $\hat{y}_{\rm o}(k)=y(k)$, $k=1,\cdots,\max(n_a+n_b)$;
4. **for** $k=\max(n_a+n_b)+1,\dots,N$ **do**

4.1. **let**
$$\hat{x}_{\rm o}(k)=[\hat{y}_{\rm o}(k-1)\cdots\hat{y}_{\rm o}(k-n_a)\ u(k)\cdots u(k-n_b)]^\top;$$

4.2. **let** $e_i(k)\leftarrow y(k)-(\theta_i^{\rm BC})^\top\begin{bmatrix}\hat{x}_{\rm o}(k)\\1\end{bmatrix};\forall i=1,\cdots s.$

4.3. **let** $\sigma(k)\leftarrow\arg\min_{i=1,\dots,s}\ \lambda e_i^2(k)+\|\hat{x}_{\rm o}(k)-c_i\|_2^2$;

4.4. **update** cluster $\mathcal{C}_{\sigma(k)}\leftarrow\mathcal{C}_{\sigma(k)}\cup\{\hat{x}_{\rm o}(k)\}$;

4.5. **let** $N_{\sigma(k)}\leftarrow N_{\sigma(k)}+1$;

4.6. **update** $P_{\sigma(k)}\leftarrow P_{\sigma(k)}-\frac{P_{\sigma(k)}\left[\begin{smallmatrix}x(k)\\1\end{smallmatrix}\right]\left[\begin{smallmatrix}x^\top(k) & 1\end{smallmatrix}\right]P_{\sigma(k)}}{1+\left[\begin{smallmatrix}x^\top(k) & 1\end{smallmatrix}\right]P_{\sigma(k)}\left[\begin{smallmatrix}x(k)\\1\end{smallmatrix}\right]}$;

4.7. **update** LS estimate
$$\theta_{\sigma(k)}^{\rm LS}\leftarrow\theta_{\sigma(k)}^{\rm LS}+P_{\sigma(k)}\left[\begin{smallmatrix}x(k)\\1\end{smallmatrix}\right]\left(y(k)-\left[\begin{smallmatrix}x^\top(k) & 1\end{smallmatrix}\right]\theta_{\sigma(k)}^{\rm LS}\right);$$

4.8. **update** model parameters
$$\theta_{\sigma(k)}^{\rm BC}\leftarrow\theta_{\sigma(k)}^{\rm LS}+\hat{\sigma}_e^2(k-1)N_{\sigma(k)}P_{\sigma(k)}J\theta_{\sigma(k)}^{\rm BC};$$

4.9. **let** $\hat{y}_{\rm o}(k)=(\theta_{\sigma(k)}^{\rm BC})^\top\begin{bmatrix}\hat{x}_{\rm o}(k)\\1\end{bmatrix}$;

4.10. **update** variance
$$\hat{\sigma}_e^2(k)=\frac{k-1}{k}\hat{\sigma}_e^2(k-1)+\frac{1}{k}\left(y(k)-\hat{y}_{\rm o}(k)\right)^2;$$

4.11. **update** centroid $c_{\sigma(k)}\leftarrow c_{\sigma(k)}+\frac{1}{N_{\sigma(k)}}(\hat{x}_{\rm o}(k)-c_{\sigma(k)})$;

4.12. **end for**;

Output: Estimated parameters $\{\theta_i^{\rm BC}\}_{i=1}^s$; clusters $\{\mathcal{C}_1\}_{i=1}^s$; mode sequence $\{\sigma(k)\}_{k=1}^N$; estimated noise variance $\hat{\sigma}_e^2(N)$.

For the identification, we consider the PWA-OE model with $s=3$ modes, and regressor $x(k)=\left[\,y(k-1)\ \ u(k-1)\,\right]^\top$. The model parameters and the mode sequence are estimated by running Algorithm 1 with different initial guesses of the noise variance estimate $\hat{\sigma}_e^2(0)$. The hyper-parameter λ is set to 1.8. The evolution of $\hat{\sigma}_e^2(k)$ for 16 different initial values $\hat{\sigma}_e^2(0)=\{0,0.1,\cdots,1.5\}$ is plotted in Fig. 1. The estimated noise variance converges to $\hat{\sigma}_e^2(N)=0.63$ for every initial condition, thus very close to the true value $\sigma_e^2=0.64$.

For comparison, we run the batch algorithm presented in [14], with $\lambda=1.8$ and the estimated noise variance $\hat{\sigma}_e^2=0.63$, that has to be carried out for 20 iterations in order to achieve a comparable accuracy with the same number of training samples. The estimated sub-model parameters obtained with the recursive bias-correction Algorithm 1 and batch approach [14] are reported in Table I, along with the least squares estimates. It is evident that the LS estimates are biased, while the bias-corrected estimates obtained by both recursive and batch algorithm match closely with the true system parameters. The difference between the recursive approach and batch algorithm in [14] can be appreciated in terms of computational efficiency. The average computational time to run Algorithm 1 is 0.16 sec, which is about 16× faster than the batch algorithm in [14].

TABLE I: True ($\theta_i^{\rm o}$) and estimated model parameters: recursive BC ($\theta_i^{\rm BC}$) *vs* LS ($\theta_i^{\rm LS}$) *vs* batch BC ($\theta_i^{\rm batch}$).

Mode	$\theta^{\rm o}$	$\theta^{\rm BC}$	$\theta^{\rm LS}$	$\theta^{\rm batch}$[14]
$s=1$	−0.4000	−0.3916	−0.2099	−0.4141
	1.0000	0.9973	0.9736	1.0108
	1.5000	1.5525	2.2250	1.4367
$s=2$	0.5000	0.5173	0.3387	0.5128
	−1.0000	−1.0043	−0.9995	−0.9739
	−0.5000	−0.5133	−0.6516	−0.4707
$s=3$	−0.3000	−0.3195	−0.2507	−0.2830
	0.5000	0.5048	0.5126	0.4888
	−1.7000	−1.5856	−1.8489	−1.7898

Based on the clusters estimated through Algorithm 1, the polyhedral partition of the regressor space is computed by solving the linear multicategory discrimination problem (19), with regularization parameter $\kappa=10^{-5}$. The true and the estimated polyhedral partitions of the regressor space are shown in Fig. 2. The estimated PWA-OE model is validated on a new dataset of length $N_{\rm val}=1000$. Based on the estimated submodel parameters and the polyhedral partitions, the output is simulated in open-loop as shown in Fig. 3 with bias-corrected and LS estimates. Only a subset of validation data is plotted for better visualization. The accuracy of the estimated mode sequence is expressed by the *Mode-Fit* (MF)

TABLE II: Mode fit index over validation data

Recursive BC	LS	Batch BC [14]
0.81	0.73	0.83

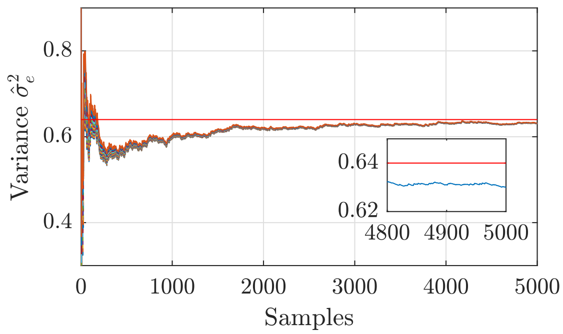

Fig. 1: Estimated variance $\hat{\sigma}_e^2$ for different initial conditions (red solid line: true value).

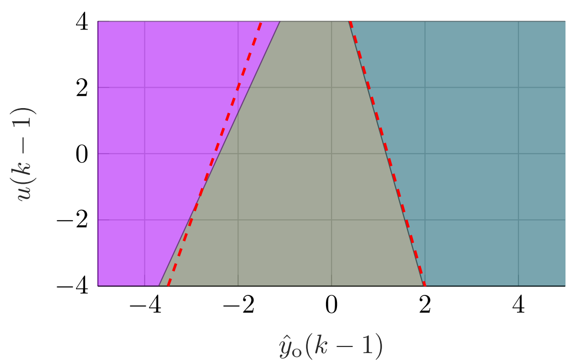

Fig. 2: True (dashed red lines) *vs* estimated partition.

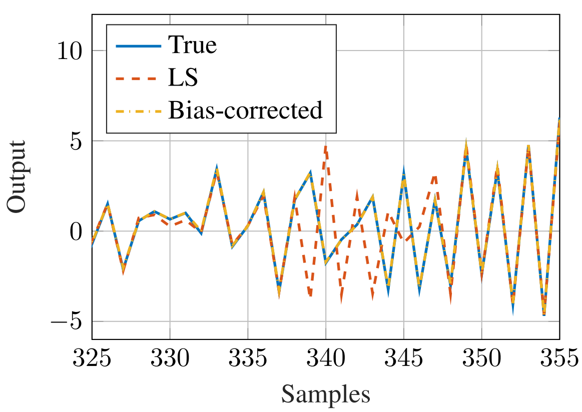

Fig. 3: Validation: true *vs* simulated outputs.

index $\text{MF} = \left(\frac{1}{N_{\text{val}}}\sum_{k=1}^{N_{\text{val}}} \mathbf{1}_{\{\sigma(k)=\sigma^\star(k)\}}\right)$, $\sigma(k)$ and $\sigma^\star(k)$ are the estimated and the true modes at time k, respectively. The MF index achieved over the validation dataset with bias-corrected (recursive and batch) and LS estimates are reported in Table II. Note that, despite old data are never reprocessed in the presented recursive algorithm, it achieves a comparable performance to the batch algorithm proposed in [14].

VII. CONCLUSION

We have presented a recursive clustering-based bias-correction algorithm for the identification of PWA output-error models. The results of the numerical case study show that the presented algorithm allows us to simultaneously estimate the sub-model parameters, unknown noise variance and mode sequence with high accuracy. The recursive nature of the algorithm allows an *on-line* model parameter update as new data becomes available. Future research activities will include quantification of bias due to misclassification of the regressors.

REFERENCES

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

[2] A. Bemporad, V. Breschi, D. Piga, and S. Boyd. Fitting jump models. *Automatica*, 96:11 –21, 2018.

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

[4] A. Bemporad and M. Morari. Control of systems integrating logic, dynamics, and constraints. *Automatica*, 35(3):407–427, 1999.

[5] L. Breiman. Hinging hyperplanes for regression, classification, and function approximation. *IEEE Transactions on Information Theory*, 39(3):999–1013, 1993.

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

[7] V. Breschi, D. Piga, and A. Bemporad. Maximum-a-posteriori estimation of jump Box-Jenkins models. Proc. 58th Conf. on Decision and Control, 2019.

[8] N. Canty, T. O'Mahony, and M. T. Cychowski. An output error algorithm for piecewise affine system identification. *Control Engineering Practice*, 20(4):444–452, 2012.

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

[10] M. Hong, T. Söderström, and W. X. Zheng. Accuracy analysis of bias-eliminating least squares estimates for errors-in-variables systems. *Automatica*, 43:1590–1596, 2007.

[11] A. Lj. Juloski and S. Weiland. A Bayesian approach to the identification of piecewise linear output error models. Proc. 14th IFAC Symposium on System Identification, pages 374–379, Newcastle, Australia, 2006.

[12] F. Lauer. On the complexity of piecewise affine system identification. *Automatica*, 62:148–153, 2015.

[13] L. Ljung. *System identification: theory for the user*. Prentice-Hall Englewood Cliffs, NJ, 1999.

[14] M. Mejari, V. Breschi, V.V. Naik, and D. Piga. A bias-correction approach for the identification of piecewise affine output-error models. To appear, IFAC World Congress, Berlin, Germany, 2020. https://ipg.idsia.ch/preprints/mejari2020a.pdf.

[15] M. Mejari, V. Breschi, and D. Piga. Recursive bias-correction method for identification of piecewise affine output-error models. Technical report, IDSIA Dalle Molle Institute for Artificial Intelligence, 2020. www.dariopiga.com/TR/TR_MeBrPi20.pdf.

[16] M. Mejari, D. Piga, and A. Bemporad. A bias-correction method for closed-loop identification of Linear Parameter-Varying systems. *Automatica*, 87:128–141, 2018.

[17] V. V. Naik, M. Mejari, D. Piga, and A. Bemporad. Regularized moving-horizon piecewise affine regression using mixed-integer quadratic programming. Proc. 25th Mediterranean Conference on Control and Automation, pages 1349–1354, Valletta, Malta, 2017.

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

[19] D. Piga, P. Cox, R. Tóth, and V. Laurain. LPV system identification under noise corrupted scheduling and output signal observations. *Automatica*, 53:329–338, 2015.

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

[21] D. Piga and R. Tóth. A bias-corrected estimator for nonlinear systems with output-error type model structures. *Automatica*, 50(9):2373–2380, 2014.

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

[23] F. Rosenqvist and A. Karlström. Realisation and estimation of piecewise-linear output-error models. *Automatica*, 41(3):545 – 551, 2005.

[24] W. X. Zheng. A bias correction method for identification of linear dynamic errors-in-variables models. *IEEE Transactions on Automatic Control*, 47(7):1142–1147, 2002.