Identification of Piecewise Affine State-Space Models via Expectation Maximization*

Rafael Rui[1], Tohid Ardeshiri[2] and Alexandre Bazanella[1]

Abstract— **This paper deals with the identification of piecewise affine state-space models. These models are obtained by partitioning the state or input domain into a finite number of regions and by considering affine submodels in each region. The proposed framework uses the Expectation Maximization (EM) algorithm to identify the parameters of the model. In most of the current literature, a discrete random variable with a discrete transition density is introduced to describe the transition between each submodel, leading to a further approximation of the dynamical system by a jump Markov model. On the contrary, we use the cumulative distribution function (CDF) to compute the probability of each submodel given the measurement at that time step. Then, given the submodel at each time step the latent state is estimated using the Kalman smoother. Subsequently, the parameters are estimated by maximizing a surrogate function for the likelihood. The performance of the proposed method is illustrated using the simulated model of the JAS 39 Gripen aircraft.**

Index Terms— **Piecewise affine, expectation maximization, state-space models**

I. INTRODUCTION

Our goal in this paper is the identification of piecewise affine state-space (PWASS) models. The PWASS models are the next natural step in the approximation of nonlinear state-space models (SSM); Instead of linearizing a nonlinear dynamical system by a single linear SSM, one can divide it into several affine submodels turning the identification of a nonlinear system into the identification of several affine submodels. PWASS models are a particular case of piecewise affine (PWA) models. Such models are used to approximate nonlinear dynamical systems and have been considered in several fields, such as automatic control [1], signal processing [2] and computer vision [3]. They are obtained by partitioning the state or the input domain into a finite number of polyhedral regions; and by considering an affine submodel in each region.

The identification problem of PWA models is a challenging problem that involves the estimation of both the parameters of the affine submodels and the coefficients that define the boundaries of each region of the state domain. The tutorial paper [4] discusses the main issues and difficulties connected to hybrid systems identification.

[1]Rafael Rui and Alexandre Bazanella are with Department of Electrical Engineering, Universidade Federal do Rio Grande do Sul, Porto Alegre 90040-060, Brazil `rafael.rui, bazanella@ufrgs.br`

[2]Tohid Ardeshiri is with the Department of Electrical Engineering, Linköping University, 58183 Linköping, Sweden, `tohid@isy.liu.se`

The majority of the identification methods proposed in the literature focus on the particular case of piecewise affine autoregressive with exogenous inputs (PWARX) models [5], [6], [7]. A PWASS model can be written as a PWARX model, and the equivalence between them is well-known [8]. The authors in [8] have shown that any observable PWASS model admits a representation as a PWARX model. It has been observed that PWARX systems are strictly contained in the class of PWASS models and that the number of submodels (and thus the number of parameters) might grow considerably when a PWASS model is converted into a minimum-order equivalent PWARX representation. On the other hand, given a PWARX model with N_r submodels, it is possible to find an equivalent PWASS model with N_r submodels, but the equivalent PWASS model will not necessarily be a minimal realization [8].

Another reason that makes the PWASS representation be more appealing compared to PWARX representation is that the majority of existing hybrid and piecewise analysis and control methods such as [9], [10] are based on SSM; and also SSM are more suitable to deal with multiple input and multiple output (MIMO) systems [11].

The identification of PWASS models undergoes some identifiability problems [12], [10]. If one assumes that either the state is not fully measurable or that one does not fix some parameters, then the model will suffer from a realization problem, i.e., one can only determine the model up to a linear state transformation. For that reason, the few works that deal with the identification of PWASS models have to make some assumption such as: minimum dwell-time assumption [13], observability and controllability assumptions for each submodel [14], known switching times assumption [15], and full state knowledge assumption [11].

In most of the current literature, a discrete random variable with a discrete transition density is introduced to describe the transition among the different regions, leading to a further approximation of the dynamical system by a jump Markov model [16]. That is, the PWASS model is seen as an extension of hidden Markov models (HMMs) in which each HMM state is associated with a linear dynamical process. Several nonlinear and linear HMMs structures have been considered and their properties investigated in the literature, see, e.g., [17], [18], [19], and references therein.

In this paper, we identify the parameters of PWASS models. We assume that the boundaries of each region that form the piecewise function are known and use an approximate Bayesian smoother within the Expectation-Maximization (EM) algorithm to identify the model parameters. In our

approach, the submodels are a function of the state, that is, given the state the active submodel is totally defined. The EM algorithm is used to identify the parameters of the submodels that form the PWASS model. In the E-step of the EM algorithm, we have used the cumulative distribution function (CDF) to compute the probability of each submodel given all available measurements. In the M-step, the parameters are estimated by maximizing a surrogate function for the likelihood function.

This paper is organized as follows; in section II is presented the model structure that we will be working with. In section III the EM algorithm for PWASS is presented and discussed. In section IV the EM algorithm is used to identify a PWASS model. Finally, the concluding remarks are given in section V.

II. PROBLEM DEFINITION

Consider the following SSM

$$\mathbf{x}_{t+1} = \mathcal{F}(\mathbf{x}_t) + B\mathbf{u}_t + \boldsymbol{\omega}_t, \tag{1a}$$

$$\mathbf{y}_t = C\mathbf{x}_t + \boldsymbol{\nu}_t, \tag{1b}$$

where the state vector $\mathbf{x}_t \in \mathbb{R}^{n_x}$ is partitioned by two scalar variables η_t and ζ_t, and a vector $\boldsymbol{\chi}_t \in \mathbb{R}^{(n_x-2)}$ such that $\mathbf{x}_t \triangleq \left[\eta_t, \zeta_t, \boldsymbol{\chi}_t^{\mathrm{T}}\right]^{\mathrm{T}}$; $\mathbf{y}_t \in \mathbb{R}^{n_y}$ is the measurement; $C \in \mathbb{R}^{n_y \times n_x}$ is the known measurement matrix; $B \in \mathbb{R}^{n_x \times n_u}$ is the input matrix, $\mathbf{u}_t \in \mathbb{R}^{n_u}$ is the input. The initial state has a prior distribution $\mathbf{x}_1 \sim \mathcal{N}(\hat{\mathbf{x}}_{1|0}, P_{1|0})$, where the subscript "$t_1|t_2$" is read "at time t_1 using measurements up to time t_2", and $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ means a Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance Σ. The process noise $\{\boldsymbol{\omega}_t \in \mathbb{R}^{n_x} | 1 \le t \le T\}$ and the measurement noise $\{\boldsymbol{\nu}_t \in \mathbb{R}^{n_y} | 1 \le t \le T\}$ are mutually independent Gaussian noise sequences. The nonlinear function $\mathcal{F}(\cdot)$ is the state transition matrix given by

$$\mathcal{F}(\mathbf{x}_t) \triangleq \begin{bmatrix} \Phi^{\mathrm{T}}\mathbf{x}_t \\ \boldsymbol{\phi}^{\mathrm{T}}\mathbf{z}_t + f(\eta_t) \\ F\mathbf{x}_t \end{bmatrix}, \tag{2}$$

with $\Phi \in \mathbb{R}^{n_x}$, $F \in \mathbb{R}^{(n_x-2) \times n_x}$, $\mathbf{z}_t \triangleq \left[\zeta_t, \boldsymbol{\chi}_t^{\mathrm{T}}\right]^{\mathrm{T}}$, $\boldsymbol{\phi} \in \mathbb{R}^{(n_x-1)}$, and $f(\eta_t)$ is a piecewise affine function such that

$$f(\eta_t) = \begin{cases} f_1 = a_1\eta_t + b_1 & \text{if } l_1 < \eta_t \le l_2, \\ \vdots \\ f_{N_r} = a_{N_r}\eta_t + b_{N_r} & \text{if } l_{N_r} < \eta_t < l_{N_r+1}, \end{cases} \tag{3}$$

where l_i, $i = 1, \ldots, N_r+1$ are the boundaries of each region that form the piecewise function. Rewriting (2) for a region $\mathcal{R}_i \triangleq \{\eta_t : l_i < \eta_t \le l_{i+1}\}$ using (3) we obtain

$$\mathcal{F}(\mathbf{x}_t) = \begin{bmatrix} \Phi^{\mathrm{T}}\mathbf{x}_t \\ \boldsymbol{\phi}^{\mathrm{T}}\mathbf{z}_t + a_i\eta_t + b_i \\ F\mathbf{x}_t \end{bmatrix} \tag{4}$$

$$= \overbrace{\begin{bmatrix} \Phi^{\mathrm{T}} \\ [a_i \ \boldsymbol{\phi}^{\mathrm{T}}] \\ F \end{bmatrix}}^{A_i} \mathbf{x}_t + \overbrace{\begin{bmatrix} \mathbf{0} \\ b_i \\ \mathbf{0} \end{bmatrix}}^{\mathbf{b}_i} \tag{5}$$

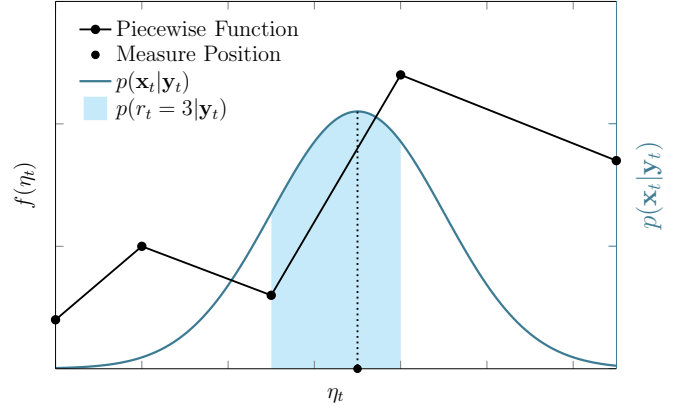$$= A_i\mathbf{x}_t + \mathbf{b}_i. \tag{6}$$

Fig. 1: The shaded area represents the probability that the system be in determinate region delimited by the piecewise function.

Therefore, for a given region $\mathcal{R}_i$, the model (1) is a conditionally affine SSM

$$\mathbf{x}_{t+1} = A_i\mathbf{x}_t + B\mathbf{u}_t + \mathbf{b}_i + \boldsymbol{\omega}_t, \tag{7a}$$

$$\mathbf{y}_t = C\mathbf{x}_t + \boldsymbol{\nu}_t. \tag{7b}$$

The index $i \in \{1, \cdots, N_r\}$ determines which piecewise affine dynamic is activated at time t. We will assume that the matrix C is such that a direct measurement of the variable η_t (up to a scalar product) is available in the measurement vector $\mathbf{y}_t$ and it is denoted by y_t.

Assuming that the piecewise function is continuous, that means,

$$a_i l_{i+1} + b_i = a_{i+1}l_{i+1} + b_{i+1}, \quad i = 1, \ldots, N_r - 1, \tag{8}$$

we can write each one of the f_i's in (3) as

$$f_1 = a_1\eta_t + b_1, \tag{9}$$

$$f_i = a_i\eta_t + b_i, \quad i = 2, \ldots, N_r, \tag{10}$$

where

$$b_i = -a_i l_i + b_1 + a_1 l_1 + \sum_{j=1}^{i-1} a_j(l_{j+1} - l_j), \tag{11}$$

The identification problem consists of finding the parameters

$$\theta \triangleq \left\{\{a_i\}_{i=1}^{N_r}, \{b_i\}_{i=1}^{N_r}, F, \Phi, \boldsymbol{\phi}\right\} \tag{12}$$

based on $\mathbf{y}_{1:T}$ (and $\mathbf{u}_{1:T}$), where $\mathbf{y}_{1:T}$ is a collection of T observations. For the continuous piecewise function, using (11) the identification problem will be to find

$$\theta' = \left\{\{a_i\}_{i=1}^{N_r}, b_1, F, \Phi, \boldsymbol{\phi}\right\}. \tag{13}$$

III. SOLUTION USING EM

The EM algorithm is an iterative method which is useful for approaching maximum likelihood estimate of unknown parameter θ in probabilistic models involving latent variables defined by

$$\hat{\theta}_{ML} = \arg\max_{\theta \in \Theta} p_\theta(\mathbf{y}_{1:T}|\theta). \tag{14}$$

TABLE I: E-step of EM algorithm

1: **Inputs:** $A_i, \mathbf{b}_i, l_i, \quad i = 1 \ldots, N_r \quad, \quad B, \quad C, \quad Q, \quad R, \quad \mathbf{u}_{1:T}, \mathbf{y}_{1:T}, \widehat{\mathbf{x}}_{1|0}, P_{1|0}$
2: **for** $j = 1$ to M **do**
3:     $\widehat{\mathbf{x}}_{1|1} \leftarrow \widehat{\mathbf{x}}_{1|0}$
4:     $P_{1|1} \leftarrow P_{1|0}$
5:     **for** $t = 1$ to $T - 1$ **do**
      Sample Submodel Trajectory $r_{1:T}$
6:       $r_t^{(j)} \sim \int_{l_i}^{l_{i+1}} \mathcal{N}(\eta_t; y_t, [R]_{(1,1)}) \, \mathrm{d}\eta_t, \quad i = 1, \ldots, N_r$
      Kalman filter prediction step
7:       $\widehat{\mathbf{x}}_{t+1|t} \leftarrow A_{r_t^{(j)}} \widehat{\mathbf{x}}_{t|t} + B\mathbf{u}_t + \mathbf{b}_{r_t^{(j)}}$
8:       $P_{t+1|t} \leftarrow A_{r_t^{(j)}} P_{t|t} A_{r_t^{(j)}}^{\mathrm{T}} + Q$
      Kalman filter update step
9:       $\Sigma_{t+1} \leftarrow CP_{t+1|t}C^T + R$
10:       $K_{t+1} \leftarrow P_{t+1|t}C^T \Sigma_{t+1}^{-1}$
11:       $\widehat{\mathbf{x}}_{t+1|t+1} \leftarrow \widehat{\mathbf{x}}_{t+1|t} + K_{t+1}(\mathbf{y}_{t+1} - C\widehat{\mathbf{x}}_{t+1|t})$
12:       $P_{t+1|t+1} \leftarrow P_{t+1|t} - P_{t+1|t}C^T \Sigma_{t+1}^{-1} C P_{t+1|t}$
13:     **end for**
      RTS smoother
14:     $P_{T|1:T} \leftarrow P_{T|T}$
15:     $\widehat{\mathbf{x}}_{T|1:T} \leftarrow \widehat{\mathbf{x}}_{T|T}$
16:     **for** $t = T - 1$ to 1 **do**
17:       $G_t \leftarrow P_{t|t}A_{r_t}^T [P_{t+1|t}]^{-1}$
18:       $\hat{P}_{t,t-1|T}^{(j)} \leftarrow \hat{P}_{t+1|T}G_t^T$
19:       $\widehat{\mathbf{x}}_{t|1:T}^{(j)} \leftarrow \widehat{\mathbf{x}}_{t|t} + G_t(\hat{\mathbf{x}}_{t+1|1:T} - \widehat{\mathbf{x}}_{t+1|t})$
20:       $\hat{P}_{t|1:T}^{(j)} \leftarrow P_{t|t} + G_t(\hat{P}_{t+1|1:T} - P_{t+1|t})G_t^T$
21:     **end for**
22:     $r_{1:T}^{(j)} \leftarrow \left\{ r_t^{(j)} \right\}, t = 1, \ldots, T$
23: **end for**
24: **Outputs:** $r_{1:T}^{(j)}, \hat{P}_{t|1:T}^{(j)}, \widehat{\mathbf{x}}_{t|1:T}^{(j)}, \hat{P}_{t,t-1|T}^{(j)}, j = 1, \ldots, M$

The main idea of the EM algorithm [20] is to compute an auxiliary function $\mathbf{Q}(\theta, \hat{\theta}_k)$ as a surrogate for the likelihood $p_\theta(\mathbf{y}_{1:T}|\theta)$. In the expectation step (E-step) of the EM approach, first the following expectation should be computed

$$\mathbf{Q}(\theta, \hat{\theta}_k) = \mathrm{E}_{\hat{\theta}_k} \left[\log p(\mathbf{x}_{1:T}, \mathbf{y}_{1:T}|\theta)|\mathbf{y}_{1:T} \right], \quad (15)$$

where $\mathrm{E}_{\hat{\theta}_k}[\cdot]$ denotes the expectation with respect to the latent variable whose posterior is computed using the previous estimate of θ denoted by $\hat{\theta}_k$. However, the joint distribution $p(\mathbf{x}_{1:T}, \mathbf{y}_{1:T}|\theta)$ can not be computed analytically due to the nonlinearity of the state transition function. As a remedy, we use an approximation which uses a latent variable representing the submodel at each time step denoted by r_t and with the probability density function (PDF)

$$p(r_t) = \iint_{\eta_t \in \mathcal{R}_i} p(\mathbf{x}_t) \, \mathrm{d}\eta_t \, \mathrm{d}\mathbf{z}_t. \quad (16)$$

Hence, $r_t \in \{1, 2, \cdots, N_r\}$ is a categorical random variable. We define $r_{1:t} \triangleq \{r_1, \ldots, r_t\}$ as the submodel trajectory up to time t. Also, we can compute the posterior distribution of r_t given the measurement $\mathbf{y}_t$ via the following marginalization

$$p(r_t|\mathbf{y}_t) \propto \iint_{\eta_t \in \mathcal{R}_i} p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t) \, \mathrm{d}\eta_t \, \mathrm{d}\mathbf{z}_t. \quad (17)$$

Assuming an uninformative prior on $\mathbf{x}_t$ and the fact that η_t is directly measured, where its corresponding measurement is denoted by y_t, it gives the following categorical posterior

distribution for r_t

$$p(r_t|\mathbf{y}_t) = p(r_t = i|\mathbf{y}_t), \quad i = 1, \cdots, N_r$$
$$= \int_{\eta_t \in \mathcal{R}_i} \mathcal{N}(\eta_t; y_t, [R]_{(1,1)}) \, \mathrm{d}\eta_t, \quad i = 1, \cdots, N_r \quad (18)$$

where $[R]_{(1,1)}$ is the element of the first row and first column of R. Fig. 1 shows the PDF $p(\mathbf{x}_t|\mathbf{y}_t)$ and a piecewise function with four regions. The shaded area represents the probability that the system be in region three given the measurement $\mathbf{y}_t$, i.e., $p(r_t = 3|\mathbf{y}_t)$. We will use the joint distribution $p_\theta(\mathbf{x}_{1:T}, \mathbf{y}_{1:T}, r_{1:T}|\theta)$ instead of $p_\theta(\mathbf{x}_{1:T}, \mathbf{y}_{1:T}|\theta)$ in (15). This joint density is given by

$$p(\mathbf{x}_{1:T}, \mathbf{y}_{1:T}, r_{1:T}|\theta)$$
$$= p(\mathbf{x}_{1:T}|r_{1:T}, \theta)p(\mathbf{y}_{1:T}|\mathbf{x}_{1:T}, \theta)p(r_{1:T}|\mathbf{y}_{1:T}, \mathbf{x}_{1:T}) \quad (19)$$

However, we notice that $r_t|\mathbf{x}_t$ is deterministic i.e., given $\mathbf{x}_t$, r_t is totally determined. As a remedy we use $p(r_{1:T}|\mathbf{y}_{1:T})$ instead of $p(r_{1:T}|\mathbf{y}_{1:T}, \mathbf{x}_{1:T})$ in (19). Since the transition density $p(\mathbf{x}_t|\mathbf{x}_{t-1}, r_{t-1}, \theta)$ is known to be $\mathcal{N}(\mathbf{x}_t|A_{r_{t-1}}\mathbf{x}_{t-1} + B\mathbf{u}_{t-1} + \mathbf{b}_{r_{t-1}}, Q)$, the joint distribution $p(\mathbf{x}_{1:T}, \mathbf{y}_{1:T}, r_{1:T}|\theta)$ can be approximated as in

$$p(\mathbf{x}_{1:T}, \mathbf{y}_{1:T}, r_{1:T}|\theta)$$
$$\approx p(\mathbf{x}_{1:T}|r_{1:T}, \theta)p(\mathbf{y}_{1:T}|\mathbf{x}_{1:T}, \theta)p(r_{1:T}|\mathbf{y}_{1:T}) \quad (20)$$
$$= p(\mathbf{x}_1)p(\mathbf{y}_1|\mathbf{x}_1)p(r_1|\mathbf{y}_1)$$
$$\times \prod_{t=2}^{T} p(\mathbf{x}_t|\mathbf{x}_{t-1}, r_{t-1}, \mathbf{y}_{t-1}, \theta)p(r_t|\mathbf{y}_t)p(\mathbf{y}_t|\mathbf{x}_t), \quad (21)$$

where $p(\mathbf{y}_t|\mathbf{x}_t) = \mathcal{N}(\mathbf{y}_t; C\mathbf{x}_t, R)$. Hence, in the E-step of the EM approach we calculate the expected value of the log-likelihood $p(r_{1:T}, \mathbf{x}_{1:T}, \mathbf{y}_{1:T}|\theta)$ with respect to the observed data and the previous estimate $\hat{\theta}_k$:

$$\mathbf{Q}(\theta, \hat{\theta}_k) = \mathrm{E}_{\hat{\theta}_k} \left[\log p(\mathbf{x}_{1:T}, \mathbf{y}_{1:T}, r_{1:T}|\theta)|\mathbf{y}_{1:T} \right] \quad (22)$$
$$= \iint \log p(\mathbf{x}_{1:T}, \mathbf{y}_{1:T}, r_{1:T}|\theta)$$
$$\times p(\mathbf{x}_{1:T}, r_{1:T}|\mathbf{y}_{1:T}, \hat{\theta}_k) \, \mathrm{d}\mathbf{x}_{1:T} \, \mathrm{d}r_{1:T}, \quad (23)$$

then in the M-step the parameters are estimated by

$$\hat{\theta}_{EM} = \arg\max_{\theta \in \Theta} \mathbf{Q}(\theta, \hat{\theta}_k). \quad (24)$$

We notice that the only part that depends on θ in the joint posterior distribution (21) is the term $p(\mathbf{x}_t|\mathbf{x}_{t-1}, r_{t-1}, \mathbf{y}_{t-1}, \theta)$. Thus the auxiliary quantity of the EM algorithm can be written as (omitting terms independent of θ)

$$\mathbf{Q}(\theta, \hat{\theta}_k) = \mathrm{E}_{\hat{\theta}_k} \left[\log p(\mathbf{x}_{1:T}, \mathbf{y}_{1:T}, r_{1:T},|\theta)|\mathbf{y}_{1:T} \right] \quad (25)$$
$$= \iint \sum_{t=2}^{T} \left(\log \mathcal{N}(\mathbf{x}_t|A_{r_{t-1}}\mathbf{x}_{t-1} + B\mathbf{u}_{t-1} + \mathbf{b}_{r_{t-1}}, Q) \right)$$
$$\times p(\mathbf{x}_{1:T}, r_{1:T}|\mathbf{y}_{1:T}, \hat{\theta}_k) \, \mathrm{d}\mathbf{x}_{1:T} \, \mathrm{d}r_{1:T}, \quad (26)$$

where the first term of (26) can be written as (omitting constant terms denoted by $\stackrel{\pm}{=}$)

$$\log \mathcal{N}(\mathbf{x}_t | A_{r_{t-1}}\mathbf{x}_{t-1} + B\mathbf{u}_{t-1} + \mathbf{b}_{r_{t-1}}, Q)$$
$$\stackrel{\pm}{=} \langle \Psi_{r_{t-1}}(\theta), s(\mathbf{x}_t, \mathbf{x}_{t-1}) \rangle + \Xi_{r_{t-1}}(\theta), \quad (27)$$

where $\langle a, b \rangle \triangleq \mathrm{tr}(a^T b) = a^T \cdot b$ denotes inner product and

$$\Psi_{r_t}(\theta) \triangleq \begin{bmatrix} A_{r_t}^T Q^{-1} \\ \mathbf{b}_{r_t}^T Q^{-1} \\ -\frac{1}{2} A_{r_t}^T Q^{-1} A_{r_t} \\ -(B\mathbf{u}_{t-1} + \mathbf{b}_{r_t})^T Q^{-1} A_{r_t} \end{bmatrix}, \quad (28)$$

$$s(\mathbf{x}_t, \mathbf{x}_{t-1}) \triangleq \begin{bmatrix} \mathbf{x}_{t-1}\mathbf{x}_t^T \\ \mathbf{x}_t^T \\ \mathbf{x}_{t-1}\mathbf{x}_{t-1}^T \\ \mathbf{x}_{t-1}^T \end{bmatrix}. \quad (29)$$

Also, $\Psi_{r_t}(\theta)$ and $s(\mathbf{x}_t, \mathbf{x}_{t-1})$ are the natural parameter and the sufficient statistic, respectively. Further,

$$\Xi_{r_t}(\theta) \triangleq -\mathbf{u}_{t-1}^T B^T Q^{-1} \mathbf{b}_{r_t} - \frac{1}{2}\mathbf{b}_{r_t}^T Q^{-1}\mathbf{b}_{r_t} \quad (30)$$

denotes the log-partition function. The auxiliary quantity of the EM algorithm can thus be written as (omitting constant terms)

$$\mathbf{Q}(\theta, \hat{\theta}_k) = \mathrm{E}_{\hat{\theta}_k}[\log p(\mathbf{x}_{1:T}, \mathbf{y}_{1:T}, r_{1:T}|\theta)|\mathbf{y}_{1:T}] \quad (31)$$

$$\stackrel{\pm}{=} \sum_{t=2}^{T} \left(\langle \Psi_{r_t}(\theta), \mathrm{E}_{\hat{\theta}_k}[s(\mathbf{x}_t, \mathbf{x}_{t-1})|\mathbf{y}_{1:T}] \rangle + \mathrm{E}_{\hat{\theta}_k}[\Xi_{r_t}(\theta)|\mathbf{y}_{1:T}] \right)$$
$$(32)$$

$$\stackrel{\pm}{=} \sum_{t=2}^{T} \langle \Psi_{r_t}(\theta), \mathrm{E}_{\hat{\theta}_k}[s(\mathbf{x}_t, \mathbf{x}_{t-1})|\mathbf{y}_{1:T}] \rangle + \sum_{t=2}^{T}\Xi_{r_t}(\theta). \quad (33)$$

In order to complete the calculation of $\mathbf{Q}(\theta, \hat{\theta}_k)$, we need to evaluate the expected value in (33) with respect to the observed data. This amounts to calculating the integral in (26). As in (21), the second term of (26) can be factorized as

$$p(r_{1:T}, \mathbf{x}_{1:T}|\mathbf{y}_{1:T}, \hat{\theta}_k) = p(\mathbf{x}_{1:T}|\mathbf{y}_{1:T}, r_{1:T}, \hat{\theta}_k) \prod_{t=1}^{T} p(r_t|\mathbf{y}_t),$$
$$(34)$$

where it was assumed that

$$p(r_{1:T}|\mathbf{y}_{1:T}) \approx \prod_{t=1}^{T} p(r_t|\mathbf{y}_t). \quad (35)$$

Since $r_1, \cdots, r_t \in \{1, 2, \cdots, N_r\}$, the number of possible trajectories up to time t is N_r^t, *i.e.*, it grows exponentially with time. The integral (26) can be computed using various integration methods. We use Monte-Carlo integration, where the samples are drawn from

$$r_t^{(j)} \sim p(r_t|\mathbf{y}_t), \quad j = 1, \cdots, M. \quad (36)$$

When the samples are drawn, the remaining integral can be computed analytically with the approximation of the posterior of each trajectory to be Gaussian. With such assumption, we can use the Kalman Smoother also called Rauch-Tung-Striebel (RTS) smoother for linear and Gaussian

SSM [21]. The algorithm in Table I presents the E-step of the proposed EM algorithm. The backward recursion equations for the RTS smoother are given in the lines 14-20 of Table I. Using the RTS smoother we can write the expectation $\mathrm{E}_{\hat{\theta}_k}\left[s(\mathbf{x}_t, \mathbf{x}_{t-1})|r_{1:T}^{(j)}, \mathbf{y}_{1:T}\right]$ as

$$\mathrm{E}_{\hat{\theta}_k}\left[s(\mathbf{x}_t, \mathbf{x}_{t-1})|r_{1:T}^{(j)}, \mathbf{y}_{1:T}\right] =$$

$$\overbrace{\begin{bmatrix} \hat{\mathbf{x}}_{t-1|1:T}^{(j)}(\hat{\mathbf{x}}_{t|1:T}^{(j)})^T + \hat{P}_{t,t-1|1:T}^{(j)} \\ \hat{\mathbf{x}}_{t|1:T}^{(j)} \\ \hat{\mathbf{x}}_{t-1|1:T}^{(j)}(\hat{\mathbf{x}}_{t-1|1:T}^{(j)})^T + \hat{P}_{t-1|1:T}^{(j)} \\ \hat{\mathbf{x}}_{t-1|1:T}^{(j)} \end{bmatrix}}^{S_t^{(j)}}, \quad (37)$$

Here the superscript $^{(j)}$ means that the quantity is conditioned to the j-th sampled trajectory $r_{1:T}$. Finally, inserting (37) in (33) we find

$$\mathbf{Q}(\theta, \hat{\theta}_k) = \frac{1}{M}\sum_{t=2}^{T}\sum_{j=1}^{M}\left(\langle \Psi_{r_t^{(j)}}(\theta), S_t^{(j)} \rangle + \Xi_{r_t^{(j)}}(\theta) \right).$$
$$(38)$$

In the M-step, it is possible to calculate exactly the gradient and the Hessian of (38) with respect to θ and use them within the Newton method to find the maximum of $\mathbf{Q}(\theta, \hat{\theta}_k)$. The quantities used in (38) are the outputs of the algorithm in Table I. It is worth pointing out that the computational complexity to calculate (38) is $\mathcal{O}(TM)$.

IV. NUMERICAL SIMULATION

We will evaluate the methodology presented above through an example of the identification of JAS 39 Gripen's flight dynamic in the longitudinal direction. The discretized SSM of the system is given by [22]

$$\begin{bmatrix} \eta_{t+1} \\ \zeta_{t+1} \end{bmatrix} = \begin{bmatrix} \eta_t + Z_\eta\eta_t + \zeta_t + Z_\zeta\zeta_t \\ f(\eta_t) + M_\zeta\zeta_t \end{bmatrix}$$
$$+ \begin{bmatrix} Z_{\delta_e} & Z_{\delta_c} \\ M_{\delta_e} & M_{\delta_c} \end{bmatrix}\begin{bmatrix} \delta_{et} \\ \delta_{ct} \end{bmatrix} + \mathbf{w}_t, \quad (39a)$$

$$\mathbf{y}_t = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \eta_t \\ \zeta_t \end{bmatrix} + \boldsymbol{\nu}_t, \quad (39b)$$

where the process noise $\boldsymbol{\omega}_t$ and the measurement noise $\boldsymbol{\nu}_t$ are Gaussian with known mean and covariance, η_t is the angle of attack, ζ_t is the pitch rate of the aircraft, δ_{et} and δ_{ct} are the elevator and canard control action. The nonlinear function $f(\eta_t)$ is constructed as a continuous piecewise function with N_r known regions as in (10). The goal is to identify the parameters $Z_\eta = -0.9759$, $Z_\zeta = 1.174$, $M_\zeta = -1.2616$ and the piecewise function values

$$\left[\{f(l_i)\}_{i=1}^{N_r+1}\right] = [-0.3240\ 0.0300\ 0.1260\ 0.9660\ 1.3800],$$

where the boundaries for each region $l = [l_1, \ldots, l_5]$ are
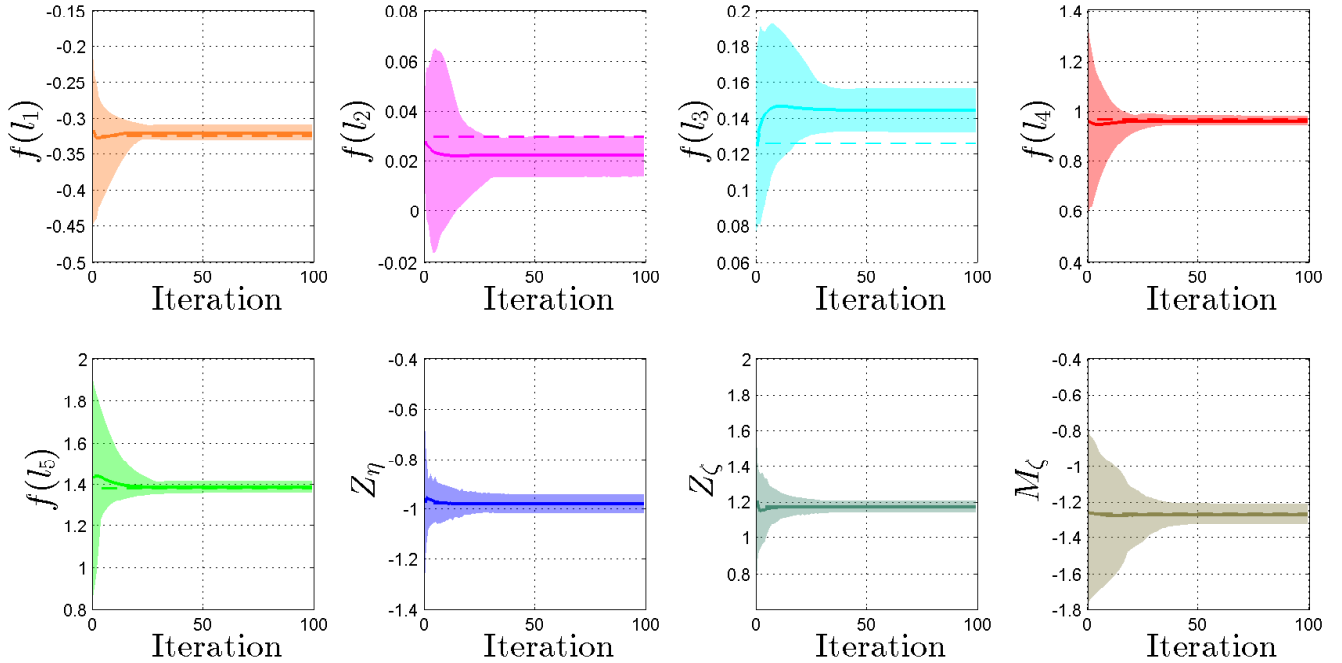
$$l = [-1°, 4°, 7°, 12°, 16°].$$

Fig. 2: Parameter identification over 100 EM iterations. The lines show the averages (full line) and the true values of the parameters (dashed line), the transparent shaded areas show the upper and lower bounds over 150 independent runs for different parameter initialization.

The system is simulated for $T = 1\,800$ time steps and the sample time is $\Delta t = 1/60s$. The input signal is such that all submodels are activated about the same amount of times. The parameters of the input matrix are: $Z_{\delta_e} = 0.3043$, $Z_{\delta_c} = 0.0289$, $M_{\delta_e} = -31.0898$ and $M_{\delta_c} = 8.2557$. It is worth mentioning that the system described in (39) is unstable, so an LQ regulator is used [22]. This LQ regulator adds a correlation between the control signal and the measurement noise as well as state noise. However, this is ignored here.

We have used (36) to sample $M = 300$ possible trajectories $r_{1:T}$ and evaluate the integral on the right-hand side of (26). The MATLAB function fminunc is used to find the solution of (24). Fig. 2 presents the identification of the model (39) for 150 different realizations and 100 iterations of the EM algorithm. The shaded area in each figure represents the upper and lower bounds over the 150 different realizations. The initialization of θ is chosen randomly and uniformly, but such that each entry laid in an interval equal to 40% of the corresponding entry in the true parameter vector. We have set $\boldsymbol{\omega}_t \sim \mathcal{N}(0, \mathrm{diag}[0.06°, 0.06°])$, $\boldsymbol{\nu}_t \sim \mathcal{N}(0, \mathrm{diag}[0.6°, 0.6°])$ and $\mathbf{x}_1 \sim \mathcal{N}(0, \mathrm{diag}[0.06°])$. Fig. 3 and 4 present one of those 150 realizations presented in the Fig. 2. The Fig. 4a shows the true piecewise function, the initial guess, and the last estimate. In Fig. 4b is presented the estimated value of the piecewise function calculated at each boundary of the function, *i.e.*, $f(l_i)$, versus the EM algorithm iteration. In Fig. 3, the estimated values of the parameters Z_η, Z_ζ, and M_ζ versus the EM algorithm iteration are given.
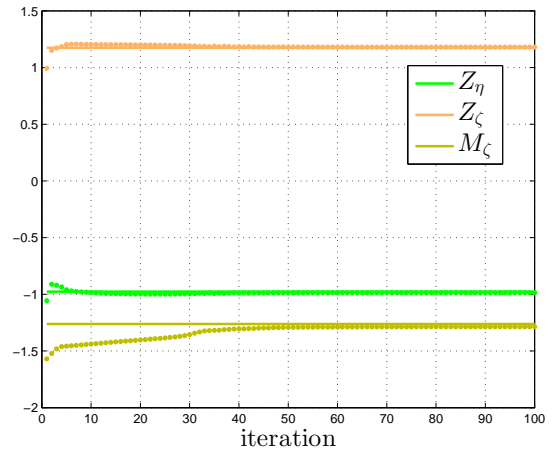
We notice that there is a bias in the final value of the

Fig. 3: True and estimate parameters Z_η, Z_ζ and M_ζ versus the EM iterations.

parameters $f(l_2)$ and $f(l_3)$. However, it is possible to verify in Fig. 4a that the estimated piecewise function can describe the true piecewise function. In that case the estimated values were $\hat{f}(l_2) = 0.02001$ and $\hat{f}(l_3) = 0.1503$, equivalent to an error between the true value and the estimated value of, respectively, 0.33 and 0.19.

V. CONCLUSION

We have proposed a method based on the EM algorithm for identification of PWASS models. We use the direct
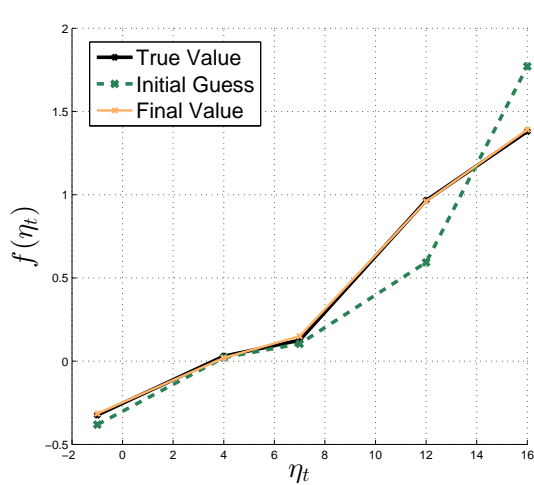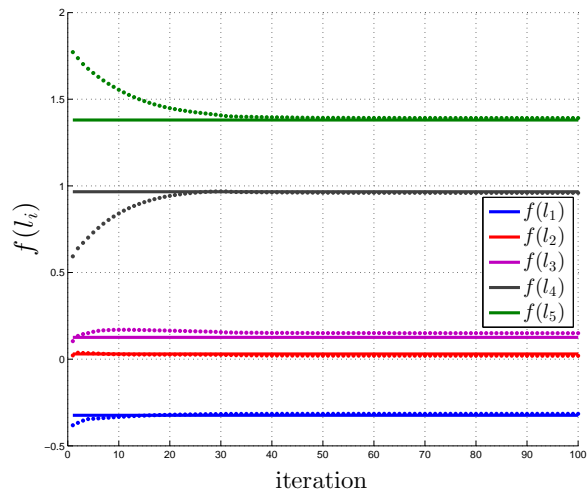
Fig. 4: Fig. 4a : True, initial guess and final estimate of the piecewise function. Fig. 4b: True and estimated values of $f(l_i)$ versus the EM iterations.

but noisy measure of the nonlinear state to calculate the probability for a given region for each time. The proposed EM algorithm was applied to the identification of the JAS 39 Gripen's flight dynamic in the longitudinal direction. In this example, a piecewise affine function with four regions was successfully identified as well as the remaining parameters of the state matrix. The results have shown that the proposed method can be used to identify PWASS models.

VI. ACKNOWLEDGMENT

REFERENCES

[1] C. E. Seah and I. Hwang, "State Estimation for Stochastic Linear Hybrid Systems with Continuous-State-Dependent Transitions: An IMM Approach," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 45, no. 1, pp. 376–392, 2009.

[2] A. Doucet, N. J. Gordon, and V. Kroshnamurthy, "Particle filters for state estimation of jump Markov linear systems," *IEEE Transactions on Signal Processing*, vol. 49, no. 3, pp. 613–624, mar 2001.

[3] R. Vidal and Y. Ma, "A Unified Algebraic Approach to 2-D and 3-D Motion Segmentation and Estimation," *Journal of Mathematical Imaging and Vision*, vol. 25, no. 3, pp. 403–421, oct 2006.

[4] S. Paoletti, A. L. Juloski, G. Ferrari-Trecate, and R. Vidal, "Identification of Hybrid Systems A Tutorial," *European Journal of Control*, vol. 13, no. 2-3, pp. 242–260, jan 2007.

[5] H. Ohlsson and L. Ljung, "Identification of switched linear regression models using sum-of-norms regularization," *Automatica*, vol. 49, no. 4, pp. 1045–1050, apr 2013.

[6] A. Hartmann, J. M. Lemos, R. S. Costa, J. Xavier, and S. Vinga, "Identification of switched ARX models via convex optimization and expectation maximization," *Journal of Process Control*, vol. 28, pp. 9–16, apr 2015.

[7] P. Mihaly, "Minimality and Identifiability of SARX Systems," K. Michel, Ed., jul 2012, pp. 541–546.

[8] S. Weiland, a. L. Juloski, and B. Vet, "On the equivalence of switched affine models and switched ARX models," in *Proceedings of the 45th IEEE Conference on Decision and Control*, no. 1. IEEE, 2006, pp. 2614–2618.

[9] E. Sontag, "Nonlinear regulation: The piecewise linear approach," *IEEE Transactions on Automatic Control*, vol. 26, no. 2, pp. 346–358, apr 1981.

[10] M. Petreczky, L. Bako, and J. H. van Schuppen, "Identifiability of discrete-time linear switched systems," in *Proceedings of the 13th ACM international conference on Hybrid systems: computation and control - HSCC '10*. New York, New York, USA: ACM Press, 2010, p. 141.

[11] L. Bako, F. Lauer, and G. Bloch, "Identification of MIMO switched state-space models," in *2013 American Control Conference*, no. 1. IEEE, jun 2013, pp. 71–76.

[12] R. Vidal, A. Chiuso, and S. Soatto, "Observability and identifiability of jump linear systems," in *Proceedings of the 41st IEEE Conference on Decision and Control, 2002.*, vol. 4, no. December. IEEE, 2002, pp. 3614–3619.

[13] J. Borges, V. Verdult, M. Verhaegen, and M. Botto, "A switching detection method based on projected subspace classification," in *Proceedings of the 44th IEEE Conference on Decision and Control*, vol. 6, no. 1. IEEE, 2005, pp. 344–349.

[14] L. Bako, G. Mercère, R. Vidal, and S. Lecoeuche, "Identification of Switched Linear State Space Models without Dwell Time," in *15th IFAC Symposium on System Identification*, W. Eric, Ed., Saint-Malo, France, jul 2009, pp. 569–574.

[15] V. Verdult and M. Verhaegen, "Subspace identification of piecewise linear systems," in *43rd IEEE Conference on Decision and Control (CDC)*. Paradise Island, Bahamas: IEEE, 2004, pp. 3838–3843.

[16] E. Fox, E. B. Sudderth, M. I. Jordan, and A. S. Willsky, "Bayesian Nonparametric Inference of Switching Dynamic Linear Models," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1569–1585, apr 2011.

[17] M. Petreczky and R. Vidal, "Realization theory of stochastic jump-Markov linear systems," in *2007 46th IEEE Conference on Decision and Control*. IEEE, 2007, pp. 4668–4674.

[18] L. Blackmore, S. Gil, S. Chung, and B. Williams, "Model learning for switching linear systems with autonomous mode transitions," in *2007 46th IEEE Conference on Decision and Control*. IEEE, 2007, pp. 4648–4655.

[19] E. Özkan, V. Šmídl, S. Saha, C. Lundquist, and F. Gustafsson, "Marginalized adaptive particle filtering for nonlinear models with unknown time-varying noise parameters," *Automatica*, vol. 49, no. 6, pp. 1566 – 1575, 2013.

[20] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. . . .* , vol. 39, no. 1, pp. 1–38, 1977.

[21] H. E. Rauch, C. T. Striebel, and F. Tung, "Maximum likelihood estimates of linear dynamic systems," *AIAA Journal*, vol. 3, no. 8, pp. 1445–1450, aug 1965.

[22] R. Larsson, "System Identification of Flight Mechanical Characteristics," Ph.D. dissertation, Linköping University, 2013.