

A Bayesian Approach to Identification of Hybrid Systems

A. Lj. Juloski, *Student Member IEEE*, S. Weiland and W. P. M. H. Heemels

Abstract—In this paper we present a novel procedure for the identification of hybrid systems in the piece-wise ARX form. The procedure consists of three steps: 1) parameter estimation, 2) classification of data points and 3) partition estimation. Our approach to parameter estimation is based on the gradual refinement of the a-priori information about the parameter values, using the Bayesian inference rule. Particle filters are used for a numerical implementation of the proposed parameter estimation procedure. Data points are subsequently classified to the mode which is most likely to have generated them. A modified version of the multi-category robust linear programming (MRLP) classification procedure, adjusted to use the information derived in the previous steps of the identification algorithm, is used for estimating the partition of the PWARX map. The proposed procedure is applied for the identification of the component placement process in pick-and-place machines.

I. INTRODUCTION

In this paper we present a novel procedure for the identification of hybrid systems in the Piece-Wise AutoRegressive eXogenous (PWARX) form. PWARX models are a generalization of classical ARX models, obtained when the regressor space is partitioned into a finite number of polyhedral regions, and where an ARX model is defined on each of the regions. PWARX models represent a broad class of systems, as they form a subclass of Piece-Wise Affine (PWA) models, which are in turn equivalent to other hybrid modelling formalisms, like Mixed Logic Dynamics (MLD) and Linear Complementarity (LC) frameworks [1].

The task of the identification procedure is to determine the parameters of the ARX models together with the regions of the regressor space where each of the models is valid. The complexity of the identification problem stems from the fact that it is not a priori known which data is generated by which ARX model, and that the determination of the model parameters and data classification have to be accomplished simultaneously [2].

The identification of PWARX models has been considered before, and to date several approaches exist, like the clustering approach [2], the greedy approach [3] and the algebraic approach [4]. The clustering procedure has been successfully applied for the identification of an experimental setup modelling electronic component placement process in pick-and-place machines [5], [6].

The main conclusions drawn from this experimental work were:

This research is financially supported by STW/PROGRESS grant EES.5173

Aleksandar Juloski and Siep Weiland are with the Department of Electrical Engineering, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands, E-mail: {a.juloski, s.weiland}@tue.nl. Maurice Heemels is with the Embedded Systems Institute, P.O. Box 513, 5600 MB Eindhoven, The Netherlands, E-mail: {maurice.heemels@embeddedsystems.nl}.

- based on the physical operation of the system, a certain amount of a priori knowledge is frequently available.
- by proper choice of excitation signals it is sometimes possible to excite only specific and well chosen modes of the physical system.
- it is often possible to interpret the modes of the identified model in terms of modes of the physical system.
- by a careful analysis of the responses of identified models, with the physical interpretation, it is possible to single out modes which are correctly and incorrectly identified.

These conclusions indicate that in practical situations there exists a strong connection between the structure of the identified hybrid models and the physical system. Ideally, one would like to exploit this connection, for a number of reasons:

- to supply the available a priori knowledge to the identification procedure, in order to improve the identification results.
- to incrementally build hybrid models of increasing complexity: by exciting different modes of the system, the model can be obtained from a series of experiments, rather than from one experiment where all modes of interest have to be excited.
- to identify models that allow an interpretation in physical terms.
- to design targeted identification experiments that would improve upon modes that are not correctly identified, while keeping the correctly identified ones unchanged.

All of the mentioned procedures for hybrid identification approach the problem as a black-box identification problem, and provide no easy way of accomplishing the above stated goals.

The idea of using and refining the a priori knowledge is central in the procedure proposed in this paper. We will treat the unknown parameters as random variables, to reflect the fact that parameter values are known only with a certain degree of confidence. Parameters will be described with probability density functions (pdf). A priori knowledge on parameters can be incorporated by choosing an appropriate pdf. As new information becomes available, knowledge on the parameters can be updated, and a *posteriori* pdf can be obtained using Bayes rule.

Another central idea of the proposed procedure is that the available information on the parameters can be used to associate the data to the mode that most likely generated it. In this way we aim to resolve the complexity issues.

To implement the proposed procedure we use particle filters to represent and compute probability density functions. For the region estimation we use the modification of the Multi-category Robust Linear Programming classification procedure (MRLP) [7].

The class of PWARX models and the identification problem will be formally introduced in section II. In sections IV and V we discuss the parameter estimation algorithm, and a particle filtering approach, as a way to implement it. In section VI we present a modified MRLP procedure. In section VII we give an example that illustrates the presented ideas. The procedure is applied to the experimental setup in section VIII. Conclusions are given in section IX.

II. PROBLEM STATEMENT

We consider piece-wise AutoRegressive eXogenous (PWARX) models of the form:

$$y(k) = f(x(k)) + e(k) \quad (1)$$

where $k \geq 0$, and:

$$f(x) = \begin{cases} \theta_1^\top \begin{bmatrix} x \\ 1 \end{bmatrix} & \text{if } x \in \mathcal{X}_1 \\ \vdots & \\ \theta_s^\top \begin{bmatrix} x \\ 1 \end{bmatrix} & \text{if } x \in \mathcal{X}_s \end{cases} \quad (2)$$

is a piece-wise affine map, and $x(k)$ is a vector of regressors, defined as

$$x(k) = \begin{bmatrix} y(k-1) & \dots & y(k-n_a) \\ u(k-1) & \dots & u(k-n_b) \end{bmatrix}^\top, \quad (3)$$

where $y(k)$ and $u(k)$ are the output and the input of the system at time k , respectively.

The parameters n_a and n_b in (3) and the number of modes s are assumed to be known. Therefore, $\theta_i \in \Theta_i \subseteq \mathbb{R}^{n+1}$, where $n = n_a + n_b$. The sets \mathcal{X}_i are assumed to be bounded polyhedrons, described by:

$$\mathcal{X}_i = \{x \mid H_i x \leq h_i\} \quad (4)$$

where H_i, h_i are of appropriate dimensions, and the inequality holds element-wise. The set $\mathcal{X} = \bigcup_{i=1}^s \mathcal{X}_i$ is a bounded polyhedron, and we assume that $\{\mathcal{X}_i\}_{i=1}^s$ is a partition of \mathcal{X} . We also assume that parameter vectors θ_i and θ_j are different over different \mathcal{X}_i :

Assumption II.1 $\theta_i \neq \theta_j$, whenever $i \neq j$.

The realization of the additive noise e in (1) is assumed to be a sequence of independent, identically distributed random values, with a known probability density function p_e .

The identification problem consists of estimating the values of the unknown parameter vectors θ_i , for $i = 1, \dots, s$, and the regions $\{\mathcal{X}_i\}_{i=1}^s$, described by (4), given the data pairs $(x(k), y(k))$, for $k = 1, \dots, T$.

III. PRELIMINARIES

When the partition $\{\mathcal{X}_i\}_{i=1}^s$ is known we can define the mode $\mu(k)$ of the data pair $(x(k), y(k))$, $k = 1, \dots, T$ uniquely as:

$$\mu(k) := i \text{ if } x(k) \in \mathcal{X}_i. \quad (5)$$

When the partition of the regressor space is not known, we will treat the mode as a random variable with integer values, that can be described with the following probability density function (pdf):

$$p(\mu(k) = i) = p_\mu^{i;k} \quad (6)$$

where $p_\mu^{i;k} \geq 0$ and

$$\sum_{i=1}^s p_\mu^{i;k} = 1, \quad \forall k = 1, \dots, T. \quad (7)$$

The mode can now be estimated as:

$$\hat{\mu}(k) = \arg \max_i p_\mu^{i;k} \quad (8)$$

Note that this estimate is not necessarily unique, as it may happen that $p_\mu^{i;k} = p_\mu^{j;k}$ for $i \neq j$. In that case either of the values i, j can be taken as $\mu(k)$.

Before we start the identification procedure we need to formulate our assumptions on the mode sequence, i.e. we need the *a priori* probability density of $\mu(k)$. If some *a priori* information of the mode of the data pair $(x(k), y(k))$ is available it can be incorporated in the coefficients $p_\mu^{i;k}$. For example, if it is known that $\mu(k) = i$ then

$$p_\mu^{i;k} = 1, \quad p_\mu^{j;k} = 0, \quad j \neq i. \quad (9)$$

When no *a priori* information about the mode sequence distribution is known, we will assume that all modes are equally probable, i.e. we set

$$p_\mu^{i;k} = \frac{1}{s}, \quad i = 1, \dots, s, \quad k = 1, \dots, T. \quad (10)$$

The unknown parameters θ_i will be treated as random variables and we will describe them with their probability density functions (pdf) p_{θ_i} . Ideally, if the pdf p_{θ_i} takes the form

$$p_{\theta_i}(\theta) = \delta(\theta - \theta_i^0), \quad (11)$$

where δ is the Dirac delta distribution, then $\theta_i = \theta_i^0$, with probability one, and the parameter value is known exactly.

From the probability density functions of the parameters, different estimates of the parameters can be easily obtained. For instance, the *expectation* of θ_i is given as

$$\hat{\theta}_i^E = E[\theta_i] = \int \theta p_{\theta_i}(\theta) d\theta \quad (12)$$

and the *maximum a posteriori probability* (MAP) estimate is given as:

$$\hat{\theta}_i^{MAP} = \arg \max_{\theta \in \Theta_i} p_{\theta_i}(\theta). \quad (13)$$

The covariance matrix V_i , which is a measure of the quality of the estimate $\hat{\theta}_i^E$, is defined as:

$$V_i^E = \int_{\Theta_i} (\theta - \hat{\theta}_i^E)(\theta - \hat{\theta}_i^E)^\top p_{\theta_i}(\theta) d\theta. \quad (14)$$

V_i^{MAP} can be defined in the analogous way for $\hat{\theta}_i^{MAP}$. The *quality* of the estimate of θ_i is defined as the spectral radius of the covariance matrix of the estimate:

$$\rho_i^E = \rho(V_i^E) = \lambda_{max}(V_i^E) \quad (15)$$

where $\lambda_{max}(\cdot)$ denotes the maximal eigenvalue. ρ_i^{MAP} is defined in an analogous way for V_i^{MAP} . This measure is useful for comparing two different estimates of the same parameter vector, where the smaller value of ρ indicates the better estimate.

To initialize the identification procedure we will assume that an *a priori* pdf of the parameters is available. A priori pdfs of the parameters can be obtained from physical insights, or by using some of the black-box procedures for hybrid identification [2–4].

IV. PARAMETER ESTIMATION

We assume that the initial (a priori) pdf of $\mu(k)$ is given by (6) for each k , and that a priori pdfs p_{θ_i} are available. Our parameter estimation algorithm will have T iterations, and in each iteration we will refine the pdf of one of the parameters. We will denote the pdf of the parameter θ_i at step k by $p_{\theta_i}(\cdot; k)$, with $p_{\theta_i}(\cdot; 0)$ denoting the a priori pdf of θ_i .

In the k -th iteration of the algorithm we will compute the a posteriori pdf of $\mu(k)$, using available a priori pdf (6) and available pdfs of the parameter vectors, assign the data pair $(x(k), y(k))$ to the mode i that most likely generated it, and compute the a posteriori pdf of θ_i , using as a fact that the pair $(x(k), y(k))$ was generated by mode i .

The probability that the mode of data pair $(x(k), y(k))$ is i , $p(\mu(k) = i | (x(k), y(k)))$ can be computed from the a priori probability that $\mu(k) = i$, using Bayes rule:

$$p(\mu(k) = i | (x(k), y(k))) = \frac{p((x(k), y(k)) | \mu(k) = i) p_{\mu}^{i;k}}{\sum_{j=1}^s p_{\mu}^{j;k} p((x(k), y(k)) | \mu(k) = j)}. \quad (16)$$

To compute the probability $p((x(k), y(k)) | \mu(k) = i)$ we use the available information on the pdf of the parameter θ_i from the step $k - 1$, i.e. $p_{\theta_i}(\theta; k - 1)$. Noting that

$$p((x(k), y(k)) | \theta) = p_e(y(k) - \theta x(k)), \quad (17)$$

and using the formula of the total probability we compute:

$$p((x(k), y(k)) | \mu(k) = i) = \int_{\Theta_i} p_e(y(k) - \theta x(k)) p_{\theta_i}(\theta; k - 1) d\theta \quad (18)$$

Note that (16) is just a normalized version of (18), but for the expression (16) to hold it is necessary that the

denominator is not equal to zero. The interpretation of this condition is that the likelihood that the data pair $(x(k), y(k))$ is generated by at least one of the available parameters should be bigger than zero.

Finally, if the data pair $(x(k), y(k))$ is assigned to mode i , this assignment is used to update the pdf of the parameter vector θ_i . Again, using Bayes rule we compute:

$$p_{\theta_i}(\theta | (x(k), y(k)); k) = \frac{p((x(k), y(k)) | \theta; k - 1) p_{\theta_i}(\theta; k - 1)}{\int_{\Theta_i} p((x(k), y(k)) | \theta; k - 1) p_{\theta_i}(\theta; k - 1) d\theta}, \quad (19)$$

which is the *a posteriori* pdf p_{θ_i} at step k .

Now, we are ready to formally state the algorithm for parameter estimation.

Algorithm IV.1 (parameter estimation)

- **step 1:** obtain the *a priori* probability density functions $p_{\theta_i}(\cdot; 0)$ and $p_{\mu}(\cdot; 0)$ for $i = 1, \dots, s$; set $k = 1$.
- **step 2:** for the data pair $(x(k), y(k))$ compute the likelihood $p(\mu(k) = i | (x(k), y(k)))$, using (16). Assign the data pair to the mode with the highest likelihood, i.e.

$$\hat{\mu}(k) = \arg \max_{1 \leq i \leq s} p(\mu(k) = i | (x(k), y(k))) \quad (20)$$

- **step 3:** compute the *a posteriori* pdf of the parameter $\theta_{\hat{\mu}(k)}$, $p_{\theta_{\hat{\mu}(k)}}(\cdot; k)$ using (19); for all $j \neq \hat{\mu}(k)$, set $p_{\theta_j}(\cdot; k) = p_{\theta_j}(\cdot; k - 1)$.
- **step 4:** $k = k + 1$; goto step 2 until $k > T$ \diamond

The schematic representation of the algorithm IV.1, for the case $s = 2$ is given in figure 1.

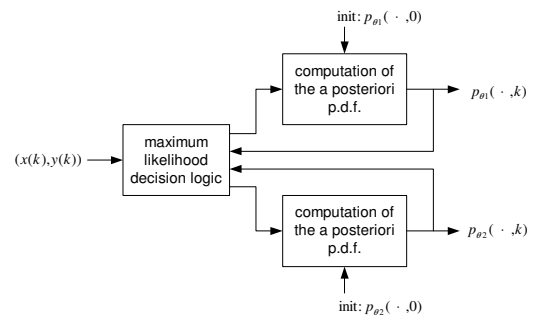


Fig. 1. Schematic representation of Algorithm IV.1 for two modes

V. PARTICLE FILTERING APPROXIMATION

Analytical solutions to (16),(18),(19) are in general intractable. We opt for the particle filtering approach as an efficient way to numerically compute probability density functions [8]. The basic idea of particle filtering methods is to approximate the pdf $p_{\theta_i}(\cdot; k)$ defined over the dense set Θ_i with a pdf supported in a finite number of points of Θ_i

(these points are called particles). We therefore approximate p_{θ_i} as:

$$p_{\theta_i}(\theta; k) \approx \hat{p}_{\theta_i}(\theta; k) := \sum_{l=1}^N w_i^{l,k} \delta(\theta - \theta_i^{l,k}). \quad (21)$$

where $\{\theta_i^{l,k}\} \subset \Theta_i$, $l = 1, \dots, N$ is a set of particles at step k , and $w_i^{l,k} > 0$ is a weight associated with the particle $\theta_i^{l,k}$.

Algorithms that sample particles $\theta_i^{l,k}$ according to any given probability density function can be found in the literature (i.e. Metropolis-Hastings algorithm, Gibbs sampler, etc. [9]).

Estimates of (12) and (14) can be obtained from (21) in a straightforward way. Combining (21) with (18) we obtain the following approximation to (18):

$$p((x(k), y(k)) | \mu(k) = i) \approx \sum_{l=1}^N w_i^{l,k-1} p_e(y(k) - \theta_i^{l,k-1} x(k)) \quad (22)$$

To compute the recursion (19) we use a modification of the Sample Importance Resampling (SIR) particle filtering algorithm [8].

Algorithm V.1 (SIR particle filtering)

- FOR $l = 1 : N$
 - diversify particles: $\theta_i^{l,k} = \theta_i^{l,k-1} + \varepsilon^l$, where $\varepsilon_l \sim \mathcal{N}(0, \Sigma_\varepsilon)$
 - compute $w_i^{l,k} = p((x(k), y(k)) | \theta_i^{l,k})$ using (17)
- END FOR
- compute total weight $q = \sum_{l=1}^N w_i^{l,k}$
- normalize: $w_i^{l,k} := q^{-1} w_i^{l,k}$, for $l = 1, \dots, N$
- resample the distribution (21) to obtain the new set of particles $\theta_i^{l,k}$, where $w_i^{l,k} = N^{-1} \diamond$

Algorithms for sampling distributions of the type (21) are standard (see for instance [8, algorithm 2]). Since we are using the SIR algorithm for estimating the constant parameter it is necessary to diversify the particles [10]. For this purpose we add the normally distributed random term ε^l to each particle in the first step of the Algorithm V.1. The variance matrix Σ_ε is the tuning parameter in the algorithm. This method of particle diversification is simple, but increases the variance of the estimates. Other particle filtering algorithms with better statistical properties but higher computational load, can be found in the literature (see for instance [10]).

VI. PARTITION ESTIMATION

Once the pdfs of the parameters $p_{\theta_i}(\cdot; T)$ are available data points can be attributed to the mode with the highest likelihood, using (20). After this classification standard techniques from pattern recognition can be applied to determine the regions $\{\mathcal{X}_i\}_{i=1}^s$ (see e.g. [7]).

However, the method of highest likelihood classification does not necessarily classify the data points to the correct mode. This problem is especially important when the

hyperplanes defined by two parameter vectors θ_i and θ_j intersect over the region \mathcal{X}_j . Then, data points near this intersection may be wrongly attributed to the mode i . This issue will be illustrated in the example in section VII. Wrongly attributed data points may in turn lead to errors in determining separating hyperplanes. In this section we propose a modified version of the MRLP algorithm from [7] that aims to alleviate this problem.

For $i = 1, \dots, s$ define the set \mathcal{D}_i as:

$$\mathcal{D}_i = \{x(k) \mid \hat{\mu}(k) = i\} \quad (23)$$

where $\hat{\mu}$ is computed as in (20), with $p_{\theta_i}(\cdot; T)$.

Definition VI.1 [7] The sets $\{\mathcal{D}_i\}_{i=1}^s$ are *piecewise-linearly separable* if there exist $w_i \in \mathbb{R}^n$, $\gamma_i \in \mathbb{R}$ for $i = 1, \dots, s$ such that $\forall i, j = 1, \dots, s$, $i \neq j$ and $\forall x \in \mathcal{D}_i$ the following holds:

$$\left\langle \begin{bmatrix} x \\ 1 \end{bmatrix}, \begin{bmatrix} w_i \\ \gamma_i \end{bmatrix} \right\rangle > \left\langle \begin{bmatrix} x \\ 1 \end{bmatrix}, \begin{bmatrix} w_j \\ \gamma_j \end{bmatrix} \right\rangle \quad (24)$$

where $\langle \cdot, \cdot \rangle$ denotes the standard inner product in \mathbb{R}^{n+1} .

Given w_i, γ_i the mode of the data point x can be estimated as:

$$\tilde{\mu}(x) = \arg \max_i \left\langle \begin{bmatrix} x \\ 1 \end{bmatrix}, \begin{bmatrix} w_i \\ \gamma_i \end{bmatrix} \right\rangle. \quad (25)$$

and the hyperplane that separates regions \mathcal{X}_i and \mathcal{X}_j is given by:

$$\{x \in \mathbb{R}^n \mid (w_i - w_j)x = \gamma_i - \gamma_j\}. \quad (26)$$

Matrices H_i, h_i describing the region \mathcal{X}_i defined in (4) can be formed as:

$$H_i = \text{col}_j\{(w_i - w_j)^\top\}, \quad h_i = \text{col}_j\{\gamma_i - \gamma_j\}, \quad (27)$$

where $j = 1, \dots, s$, $j \neq i$, and the operator $\text{col}\{\cdot\}$ stacks its arguments into a column vector. Note that only regions with up to $s - 1$ vertices can be described in this way.

If the sets \mathcal{D}_i are not separable some data points are going to violate (24). If the regressor $x \in \mathcal{D}_i$ is classified to the region \mathcal{X}_j (i.e. if $\tilde{\mu}(x) = j$) the *violation* $\zeta_{ij}(x) : \mathcal{D}_i \rightarrow \mathbb{R}$ is given as

$$\zeta_{ij}(x) = (-x(w_i - w_j) + (\gamma_i - \gamma_j) + 1)_+ \quad (28)$$

where $q_+ = \max\{q, 0\}$. Standard MRLP algorithm finds w_i, γ_i by minimizing the sum of averaged violations (28), through a single linear program [7].

In our case we will weight the violations (28) according to the following principle: if the probability that the regressor $x \in \mathcal{D}_i$ belongs to the mode i is approximately equal to the probability that it belongs to mode j , then the corresponding violation $\zeta_{ij}(x)$, if positive, should not be penalized highly. We define the *weighting function* $\xi_{ij} : \mathcal{D}_i \rightarrow \mathbb{R}$ as

$$\xi_{ij}(x(k)) = \log \frac{p((x(k), y(k)) | \mu(k) = i)}{p((x(k), y(k)) | \mu(k) = j)}. \quad (29)$$

Since for any $j \neq i$

$$p((x(k), y(k)) | \mu(k) = i) > p((x(k), y(k)) | \mu(k) = j)$$

the weight (29) is always nonnegative, and is equal to zero when the two likelihoods are exactly equal.

The weighting function (29) takes into account only the relative size of the mode likelihoods. If outliers are present in the data set, mode likelihoods may be negligible, but their ratio, formed as in (29), may still be significant. Another possible choice of the weighting function ξ_{ij} , which also takes the absolute sizes of mode likelihoods into account is:

$$\xi_{ij}(x(k)) = \frac{p((x(k), y(k)) | \mu(k) = i) - p((x(k), y(k)) | \mu(k) = j)}{p((x(k), y(k)) | \mu(k) = i) + p((x(k), y(k)) | \mu(k) = j)}. \quad (30)$$

The optimization problem can be stated as:

$$\min_{w_i, \gamma_i} \sum_{i=1}^s \sum_{\substack{j=1 \\ j \neq i}}^s \sum_{x \in \mathcal{D}_i} \xi_{ij}(x) \zeta_{ij}(x). \quad (31)$$

Problem (31) can be further cast as a linear program, in the same way as in [7].

VII. EXAMPLE

Let data $(x(k), y(k))$ be generated by the system of type (1) where:

$$f(x) = \begin{cases} \begin{bmatrix} 0.5 & 0.5 \end{bmatrix} \begin{bmatrix} x \\ 1 \end{bmatrix}, & \text{if } x \in [-2.5, 0) \\ \begin{bmatrix} -1 & 2 \end{bmatrix} \begin{bmatrix} x \\ 1 \end{bmatrix}, & \text{if } x \in [0, 2.5] \end{cases} \quad (32)$$

and $e(k)$ is a sequence of normally distributed random numbers, with zero mean and variance $\sigma_e^2 = 0.025$. The data set of $T = 100$ data points together with the true model is shown in figure 2.

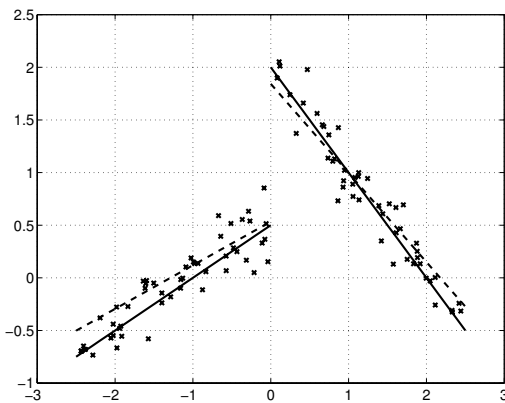


Fig. 2. True model (solid), identified model (dashed) and the data set used for identification

A priori pdfs are chosen as the uniform distributions $p_{\theta_1}(\theta_1) = p_{\theta_2}(\theta_2) = \mathcal{U}([-2.5, 2.5] \times [-2.5, 2.5])$. A particle approximation to this pdf, with $N = 200$ particles for each

pdf, is given in figure 3, left. The particle filtering algorithm V.1 is applied, with $\Sigma_\varepsilon^2 = \text{diag}\{0.001, 0.001\}$ and the final particle distribution at the step $k = 100$ is shown in figure 3, right. The estimates of the parameter vectors are:

$$\theta_1^E = \begin{bmatrix} 0.4143 \\ 0.5340 \end{bmatrix}, \quad \theta_2^E = \begin{bmatrix} -0.8467 \\ 1.8432 \end{bmatrix} \quad (33)$$

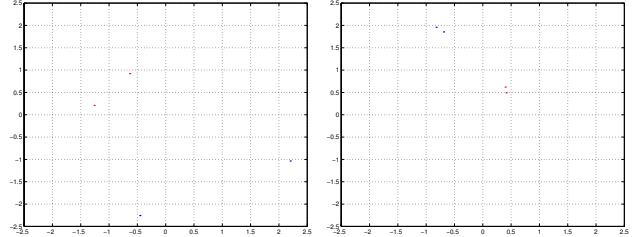


Fig. 3. **left:** Particle approximation to the initial pdfs of the parameters θ_1, θ_2 (red dots: particles of p_{θ_1} , blue dots: particles of p_{θ_2}) **right:** Final pdf of the parameters θ_1 (red dots), θ_2 (blue dots)

Data points are classified using (20), and the results are depicted in figure 4a. Several data points that belong to mode 1 are attributed to mode 2. These points are near the virtual intersection of the two lines defined by the parameter vectors. Figure 4b shows the weighting function (29) for misclassification of points is shown. The weight for misclassification of wrongly attributed points is small in comparison to the weight for misclassification of the correctly attributed points. The region for mode 1 is estimated as $x \geq 0.0228$ while the region corresponding to mode 2 is estimated as $x < 0.0228$. The identified model, together with the true model and the data set is depicted in figure 2.

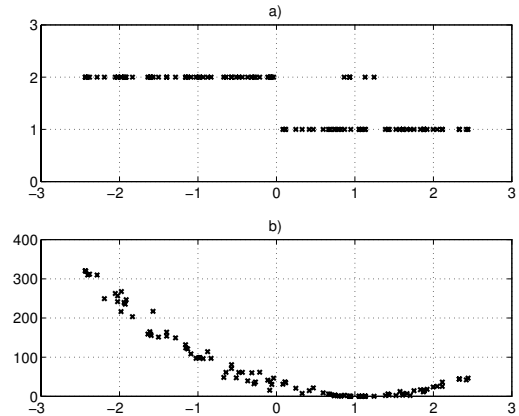


Fig. 4. a) Data points attributed to modes b) Price function for the wrong classification

VIII. EXPERIMENTAL EXAMPLE

In order to demonstrate the proposed identification procedure we applied it to the data collected from the experimental setup made of the mounting head from a pick-and-place machine. The same experimental setup was previously

successfully identified using the clustering procedure [2]. The experimental setup and the identification results with the clustering procedure are described in more detail in [5], [6].

A photo and the schematic representation of the experimental setup are given in figure 5. The setup consists of the mounting head, from an actual pick-and-place machine, which is fixed above the impacting surface. The upper part of the figure 5b (mass M , spring c_1 , friction blocks d_1 and f_1) schematically represents the mechanical assembly of the mounting head, while the lower part (spring c_2 and friction blocks d_2 and f_2) represents the impacting surface.

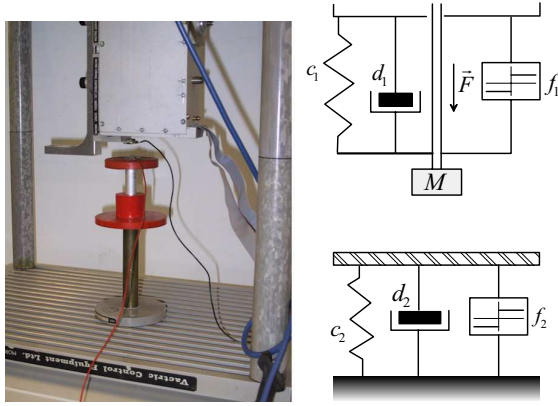


Fig. 5. **left:** a) Photo of the experimental setup **right:** b) Schematic representation of the experimental setup

The dynamics of the experimental setup exhibits, in a first approximation, four different modes of operation:

- **upper saturation:** the head is in the upper retracted position (i.e. can not move upwards, due to the physical constraints)
- **free mode:** the head is not in contact with the impacting surface, but is not in the upper saturation
- **impact mode:** the head is in contact with the impacting surface, but is not in lower saturation
- **lower saturation:** the head is in the lower extended position, (i.e. can not move downwards due to the physical constraints)

We stress that the switch between the impact and free modes does not occur at a constant head position, because of the movement of the impacting surface. For the upper and lower saturations, although they occur at a fixed position, they introduce dynamic behaviors due to the bouncing when hitting the constraints.

The control input is the voltage applied to the motor, which is converted up to a negligible time constant to the force \vec{F} . The input signal for the identification experiment should be chosen in a way that modes of interest are sufficiently excited. To obtain the data for identification, the input signal $u(t)$ is chosen as:

$$u(t) = a_k \quad \text{when} \quad t \in [kT, (k+1)T) \quad (34)$$

where $T > 0$ is fixed, and the amplitude a_k is a random variable, with uniform distribution in the interval $[a, b]$. By

properly choosing the boundaries of the interval $[a, b]$ only certain modes of the system are excited. For instance only free and impact modes can be excited, without reaching upper and lower saturations.

Physical insight into the operation of the setup facilitates the initialization of the procedure. For instance, although the mode switch does not occur at a fixed height of the head, with a degree of certainty data points below certain height may be attributed to the free mode, and, analogously data points above certain height may be attributed to the impact mode. Data points that belong to saturations can also be distinguished. This a priori information may be exploited in the following way.

If $m > n + 1$ data pairs, $(x(k_1), y(k_1)), \dots, (x(k_m), y(k_m))$ are attributed to the mode i , the least squares estimate of the value of the parameter vector θ_i^{LS} may be obtained as:

$$\begin{aligned} \theta_i^{LS} &= (\Phi_i^T \Phi_i)^{-1} \Phi_i^T y_i, \quad (35) \\ \Phi_i &= \begin{bmatrix} x(k_1) & x(k_2) & \dots & x(k_m), \\ 1 & 1 & \dots & 1 \end{bmatrix}^T \\ y_i &= [y(k_1) \quad y(k_2) \quad \dots \quad y(k_m)]^T. \end{aligned}$$

The empirical covariance matrix of θ_i^{LS} can be computed as [11]:

$$\tilde{V}_i = \frac{y_i^T (I - \Phi_i (\Phi_i^T \Phi_i)^{-1} \Phi_i^T) y_i (\Phi_i^T \Phi_i)^{-1}}{m - (n + 1)}. \quad (36)$$

This information is sufficient to define the parameter θ_i as a normally distributed random variable:

$$p_{\theta_i}(\cdot; 0) = \mathcal{N}(\theta_i^{LS}, \tilde{V}_i) \quad (37)$$

Samples from the normal distribution can be easily obtained with some of the mentioned algorithms for sampling from general multidimensional distributions (or using built-in MATLAB functions).

We present an identification experiment in which the free, impact and lower saturation modes are excited. Collected data sets consist of 750 points which are divided into two overlapping sets of 500 points: one is used for identification, while the second is used for validation of the identified models. Weighting function (30) is used. As a probability density function of the noise we used $p_e \sim \mathcal{N}(0, 1)$.

The data set used for identification is depicted in figure 6. Portions of the dataset that are used for initialization of free, impact and saturation mode are marked with \times , \circ and $+$, respectively. Models with $s = 3, n_a = 2, n_b = 2$ are identified.

Final classification of data points is depicted in figure 7a. In figure 7b spectral radii of variance matrices $\rho_{1,2,3}^F$ at each step of the classification are depicted. Simulation of the identified model, together with the modes active during the simulation is depicted in figure 8.

From figure 7a we see that data points are classified into three groups, corresponding to the impact, free and saturation modes. From 7b we see that the estimates of the parameters are improving during the iterations of the

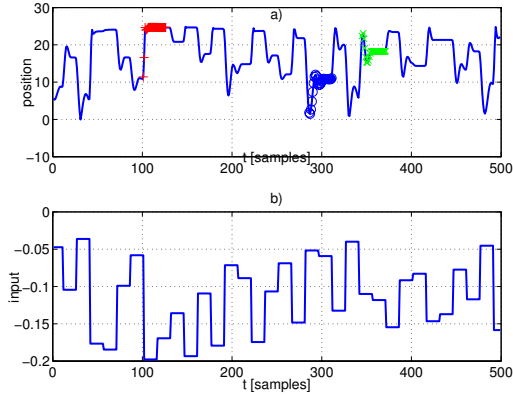


Fig. 6. Identification with saturations. Data set used for identification a) position (portion marked with +: data points used for the initialization of the lower saturation mode; portion marked with o: data points used for initialization of impact mode; portion marked with x: data points used for the initialization of free mode b) input signal

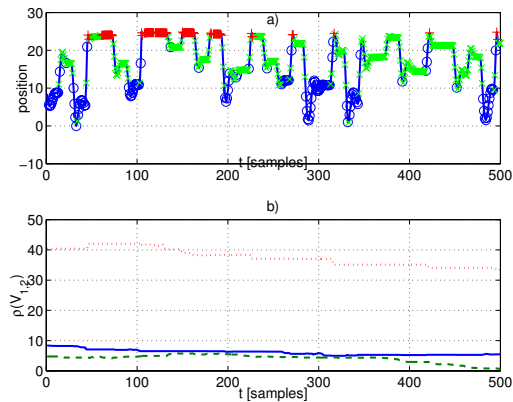


Fig. 7. Identification with saturations. a) Classified data points (o: free mode, x: impact mode; + lower saturation) b) $\rho(V_{1,2,3}^E$ (solid line: free mode; dashed line: impact mode; dotted line: lower saturation)

algorithm. From figure 8 we see that the simulated response is satisfactory, and that the modes active during the simulation correspond well to intuitive classification of data. Response in the free mode does not match the measured response precisely, while the responses in impact and saturation modes are predicted remarkably well.

IX. CONCLUSIONS

In this paper we presented a novel method for the identification of hybrid systems in PWARX form. The presented method facilitates the use of the available a priori information on the system to be identified, but can also be used as the black-box method. Unknown parameters are treated as random variables described with probability density functions (pdf), and the Bayes rule is used to refine a priori pdfs, as new information is considered. Modified MRLP procedure is used for estimation of the regions. The applicability and effectiveness of the developed method is illustrated on an experimental example.

The approach taken in this paper is to refine the parameter

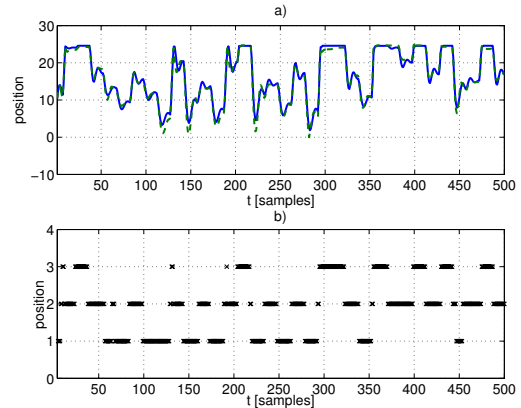


Fig. 8. Identification with saturations. a) Simulation of the identified model (solid line: simulated response, dashed line: measured response b) modes active during the simulation

pdfs by taking one data point at each step of the algorithm. Another possible approach would be to first classify all available data on the basis of a priori knowledge, and compute the a posteriori parameter pdfs in one step. This approach will be considered in the further research. Further research will also focus on the investigation of convergence properties of the presented method, and on generalization to other classes of hybrid systems.

REFERENCES

- [1] W. Heemels, B. De Schutter, and A. Bemporad, "Equivalence of hybrid dynamical models," *Automatica*, vol. 37, pp. 1085–1091, 2001.
- [2] G. Ferrari-Trecate, M. Muselli, D. Liberati, and M. Morari, "A clustering technique for the identification of piecewise affine systems," *Automatica*, vol. 39, no. 2, pp. 205–217, 2003.
- [3] A. Bemporad, A. Garulli, S. Paoletti, and A. Vicino, "A greedy approach to identification of piecewise affine models," in *Hybrid Systems: Computation and Control 2003*, ser. Lecture Notes in Computer Science, O. Maler and A. Pnueli, Eds., vol. 2623, Prague, Czech Republic, 2003, pp. 97–112.
- [4] R. Vidal, S. Soatto, and S. Sastry, "An algebraic geometric approach for identification of linear hybrid systems," in *Proceedings of 42nd IEEE Conference on Decision and Control*, 2003.
- [5] A. Juloski, W. Heemels, and G. Ferrari-Trecate, "Identification of an experimental hybrid system," in *Proc. of IFAC Symposium on Analysis and Design of Hybrid Systems*, Saint Malo, France, 2003.
- [6] —, "Data-based hybrid modelling of the component placement process in pick-and-place machines," *Control Engineering Practice*, to appear, 2004.
- [7] K. Bennet and O. Mangasarian, "Multicategory discrimination via linear programming," *Optimization Methods and Software*, vol. 4, pp. 27–39, 1994.
- [8] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, pp. 174–188, 2002.
- [9] G. Fishman, *Monte Carlo: Concepts, Algorithms and Applications*. Springer, 1996.
- [10] C. Berzuini and W. Gilks, "RESAMPLE-MOVE filtering with cross-model jumps," in *Sequential Monte-Carlo Methods in Practice*, A. Doucet, N. de Freitas, and N. Gordon, Eds. Springer, 2001, ch. 6, pp. 117–138.
- [11] L. Ljung, *System Identification - Theory for the User*. Prentice Hall, 1999.