43rd IEEE Conference on Decision and Control
December 14-17, 2004
Atlantis, Paradise Island, Bahamas

Data Classification and Parameter Estimation for the Identification of Piecewise Affine Models

A. Bemporad, A. Garulli, S. Paoletti and A. Vicino

Abstract— **This paper proposes a three-stage procedure for parametric identification of PieceWise affine AutoRegressive eXogenous (PWARX) models. The first stage simultaneously classifies the data points and estimates the number of submodels and the corresponding parameters by solving the MIN PFS problem (Partition into a MINimum Number of Feasible Subsystems) for a set of linear complementary inequalities derived from input-output data. Then, a refinement procedure reduces misclassifications and improves parameter estimates. The last stage determines a polyhedral partition of the regressor set via two-class or multi-class linear separation techniques. As a main feature, the algorithm imposes that the identification error is bounded by a fixed quantity δ. Such a bound is a useful tuning parameter to trade off between quality of fit and model complexity. Ideas for efficiently addressing the MIN PFS problem, and for improving data classification are also discussed in the paper. The performance of the proposed identification procedure is demonstrated on experimental data from an electronic component placement process in a pick-and-place machine.**

I. INTRODUCTION

PieceWise Affine (PWA) systems are defined by partitioning the state+input space into a finite number of polyhedral regions, and by considering linear/affine subsystems sharing the same continuous state in each region [1]. Such systems represent an attractive model structure for system identification. Thanks to the universal approximation properties of PWA maps, any nonlinear dynamics can be modelled with arbitrary accuracy by a PWA system. In addition, the equivalence between PWA systems and several classes of hybrid systems [2], [3] makes PWA system identification techniques suitable for obtaining hybrid models from data. PWA systems have indeed many applications in different contexts such as neural networks, electrical networks, time-series analysis, function approximation.

Identification of PWA models is a challenging problem that involves the estimation of both the parameters of the affine submodels, and the coefficients of the hyperplanes defining the partition of the state+input space (or the regressor space, for models in regression form). The main difficulty lies in the fact that the identification problem underlies a classification problem, in which each data point must be associated to one region and to the corresponding submodel. The problem is even harder when also the number of submodels must be estimated. Several techniques for identifying PWA models of nonlinear dynamic systems can be found in the literature. A nice overview of different approaches to PWA system identification is presented in [4]. Recently, novel contributions to this topic have been proposed in both the hybrid systems and the nonlinear identification communities. Estimation techniques for PieceWise affine ARX (PWARX) models include the clustering-based procedure [5], the algebraic procedure [6], the adapted weights procedure [7], and the Bayesian procedure [8]. In [9] the identification problem is formulated for two subclasses of PWA models, namely Hinging Hyperplane ARX (HHARX) and Wiener PWARX (W-PWARX) models, and solved via mixed-integer linear or quadratic programming.

A bounded-error approach to the identification of PWARX models, inspired by ideas from set-membership identification (see [10], [11] and references therein), is taken in [12]. The main feature of such an approach is to impose that the identification error is bounded by a given quantity δ for all the samples in the estimation data set. As in [13], at initialization the estimation of the number of submodels, data classification and parameter estimation are performed simultaneously by partitioning a set of linear complementary inequalities derived from data into a minimum number of feasible subsystems (MIN PFS problem). Then, a refinement procedure is proposed in order to reduce misclassifications and to improve parameter estimates. Region estimation is lastly performed via two-class [14], [15] or multi-class [16], [17] linear separation techniques. The bound δ is used as a tuning knob to trade off between quality of fit and model complexity: The larger δ, the smaller the required number of submodels at the price of a worse fit of the data.

In this paper, several improvements of the identification procedure [12] are presented and discussed. The paper is organized as follows. The identification problem is formulated in Section II. In Section III, the greedy algorithm proposed in [13] for solving the MIN PFS problem is modified in order to obtain a number of submodels that is expected to be closer to the minimum. Moreover, new ideas for dealing with undecidable data points in the data classification phase are proposed. In Section IV, the performance of the enhanced identification procedure is tested on experimental data from an electronic component placement process in a pick-and-place machine [18]. Conclusions are given in Section V.

II. PROBLEM FORMULATION

Given a discrete-time nonlinear dynamic system with input $\mathbf{u}_k \in \mathbb{R}^p$, output $y_k \in \mathbb{R}$, and possibly discontinuous

The authors are with Dipartimento di Ingegneria dell'Informazione, Università di Siena, Via Roma 56, 53100 Siena, Italy. E-mail: {bemporad,garulli,paoletti,vicino}@dii.unisi.it.

0-7803-8682-5/04/$20.00 ©2004 IEEE

dynamics, let $\mathbf{u}^{k-1}$ and $\mathbf{y}^{k-1}$ be, respectively, past inputs and outputs generated by the system up to time $k-1$. A PWARX model establishes a relationship between past observations $(\mathbf{u}^{k-1}, \mathbf{y}^{k-1})$ and future outputs y_k in the form

$$y_k = f(\mathbf{x}_k) + \varepsilon_k \,, \tag{1}$$

where $\varepsilon_k \in \mathbb{R}$ is the error term, $\mathbf{x}_k \in \mathbb{R}^n$ is the regression vector with fixed structure depending only on past n_a outputs and n_b inputs:

$$\mathbf{x}_k = [\, y_{k-1} \;\ldots\; y_{k-n_a} \;\mathbf{u}'_{k-1} \;\ldots\; \mathbf{u}'_{k-n_b} \,]' \tag{2}$$

(hence, $n = n_a + p \cdot n_b$), and $f : \mathscr{X} \to \mathbb{R}$ is the piecewise affine map

$$f(\mathbf{x}) = \begin{cases} \varphi'\theta_1 & \text{if} \quad \mathbf{x} \in \mathscr{X}_1 \\ \vdots & \vdots \\ \varphi'\theta_s & \text{if} \quad \mathbf{x} \in \mathscr{X}_s \,, \end{cases} \tag{3}$$

which is defined over the regressor set $\mathscr{X} = \bigcup_{i=1}^{s} \mathscr{X}_i \subseteq \mathbb{R}^n$ where the PWARX model is valid. In (3), s is the number of submodels (or discrete modes), φ is the extended vector $\varphi = [\, \mathbf{x}' \; 1 \,]'$, and $\theta_i \in \mathbb{R}^{n+1}$, $i = 1, \ldots, s$, are the parameter vectors of each affine ARX submodel. The sets $\mathscr{X}_i$ have disjoint interiors ($\mathscr{X}_i^\circ \bigcap \mathscr{X}_j^\circ = \emptyset, \forall\, i \neq j$, where $\mathscr{X}_i^\circ$ denotes the interior of $\mathscr{X}_i$), and are convex polyhedra,

$$\mathscr{X}_i = \left\{ \mathbf{x} \in \mathbb{R}^n : H_i \varphi \preceq \mathbf{0} \right\}, \tag{4}$$

where $H_i \in \mathbb{R}^{q_i \times (n+1)}$, $i = 1, \ldots, s$ and "$\preceq$" denotes componentwise inequality. With definition (4), f could be multi-valued over common boundaries of the regions $\mathscr{X}_i$. This issue can be easily overcome by making some of the inequalities strict in the definitions of the polyhedra $\mathscr{X}_i$.

The identification problem considered in this paper consists in finding a PWARX model (1)-(3) that fits the available data points $(y_k, \mathbf{x}_k)$, $k = 1, \ldots, N$ according to the following condition:

$$|y_k - f(\mathbf{x}_k)| \leq \delta \,, \quad \forall\, k = 1, \ldots, N \,, \tag{5}$$

for a chosen $\delta > 0$. That is, the identified PWARX model is characterized by imposing a bound $\delta > 0$ on the error term ε_k in (1) for all the samples in the estimation data set. The number s of submodels is not fixed a priori, but is estimated together with the parameters of the model. To obtain a model which is as simple as possible (where "simplicity" is measured in terms of number of submodels), the minimum s for which (5) can be satisfied, is sought. The identification problem can be formally stated as follows:

Problem 1: Given N data points $(y_k, \mathbf{x}_k)$, $k = 1, \ldots, N$, and $\delta > 0$, estimate the minimum positive integer s, a set of parameter vectors $\{\theta_i\}_{i=1}^{s}$, and a polyhedral partition $\{\mathscr{X}_i\}_{i=1}^{s}$ of the regressor set $\mathscr{X}$, such that the corresponding PWARX model (1)-(3) satisfies condition (5).

Note that the bound δ is used as a tuning knob of the identification procedure that can be adjusted in order to find the desired trade off between model complexity and quality of fit: The smaller δ, the larger is typically the number of submodels needed to satisfy (5), while on the other hand, the larger δ, the worse is the quality of fit [13], [19]. Note that the case of different bounds for each data point can be cast into (5) by suitably scaling the data.

The solution of Problem 1 underlies a classification problem, in which each data point is associated to one submodel. According to the bounded-error condition (5), a set of feasible parameters can be also associated to each submodel. This enables the evaluation of the related parametric uncertainty [10].

III. THE IDENTIFICATION PROCEDURE

The procedure introduced in [12] for solving Problem 1 consists of the following three steps:

1) *Initialization*. Data classification and parameter estimation are carried out simultaneously, together with the estimation of the number of submodels, by partitioning a set of linear inequalities derived from data into a minimum number of feasible subsystems (MIN PFS problem).
2) *Refinement*. Misclassifications are reduced and parameter estimates are improved.
3) *Region estimation*. Clusters of regression vectors are linearly separated via two-class or multi-class linear separation techniques.

In this paper, new ideas for efficiently addressing the MIN PFS problem, and for improving data classification are discussed. Note that condition (5) naturally leads to a set-membership or bounded-error approach to the identification problem. Hence, in the following, pointwise parameter estimates will be computed by using the ℓ_∞ projection estimator [10]. Given a set $\mathscr{D}$ of data points $(y_k, \mathbf{x}_k)$, the ℓ_∞ projection estimate is defined as

$$\Phi_p(\mathscr{D}) = \arg\min_{\theta} \max_{(y_k, \mathbf{x}_k) \in \mathscr{D}} |y_k - \varphi_k'\theta| \,, \tag{6}$$

where $\varphi_k = [\, \mathbf{x}'_k \; 1 \,]'$. Problem (6) can be easily solved via linear programming. Note that (6) could be replaced by any other projection estimate, such as least squares.

Notation: $\#\mathscr{A}$ denotes the cardinality of a finite set $\mathscr{A}$, and $\mathscr{A} \backslash \mathscr{B}$ denotes the difference of two sets $\mathscr{A}$ and $\mathscr{B}$.

A. Initialization

The estimation of the number of submodels, data classification and parameter estimation are simultaneously addressed by solving the following MIN PFS problem with linear complementary inequalities.

Problem 2: Given $\delta > 0$ and the (possibly infeasible) system of N linear complementary inequalities:

$$|y_k - \varphi_k'\theta| \leq \delta \,, \quad k = 1, \ldots, N, \tag{7}$$

find a partition of (7) into the minimum number s of feasible subsystems.

Problem 2 is NP-hard. A greedy randomized approach for tackling it has been proposed in [13]. This approach divides the partition problem into a sequence of subproblems, each

TABLE I

MODIFIED GREEDY ALGORITHM FOR THE MIN PFS PROBLEM WITH COMPLEMENTARY INEQUALITIES

Algorithm 1

Let $\mathscr{T}_1 = \{1,\dots,N\}$ and $\ell = 0$
REPEAT
  Let $\ell = \ell + 1$ and $\Sigma_\ell = \{|y_k - \varphi_k'\theta| \le \delta : k \in \mathscr{T}_\ell\}$
  Find a solution θ_ℓ to the MAX FS problem for Σ_ℓ (*Algorithm 2*)
  Let $i = 1$
  WHILE $i < \ell$
    Let $\mathscr{I}_{i\ell} = \{k \in \mathscr{T}_i : |y_k - \varphi_k'\theta_\ell| \le \delta\}$
    IF $\#\mathscr{I}_{i\ell} > \#\mathscr{I}_i$ THEN let $\theta_i = \theta_\ell$ and $\ell = i$
    Let $i = i + 1$
  END WHILE
  Let $\mathscr{I}_\ell = \{k \in \mathscr{T}_\ell : |y_k - \varphi_k'\theta_\ell| \le \delta\}$ and $\mathscr{T}_{\ell+1} = \mathscr{T}_\ell \backslash \mathscr{I}_\ell$
UNTIL $\mathscr{T}_{\ell+1} = \emptyset$
RETURN $s = \ell$ and $\mathscr{I}_i$, $i = 1,\dots,s$

one consisting in finding a parameter vector θ that satisfies the maximum number of linear complementary inequalities (MAX FS problem). Starting from (7), feasible subsystems of maximum cardinality are iteratively extracted (and the corresponding inequalities removed), until the remaining subsystem is feasible. Since the MAX FS problem is again NP-hard, a suboptimal approach to its solution, based on a randomized and thermal relaxation method for solving systems of linear inequalities, has been also proposed. The greedy randomized algorithm [13] has a limited computational burden, but is not guaranteed to find the minimum number of subsystems. In particular, it was observed in extensive trials that both the variance of the results can be quite large (*i.e.*, the number of extracted subsystems may differ considerably from trial to trial), and the average number of extracted subsystems can be rather far from the minimum [19].

1) A greedy algorithm for the MIN PFS problem: Based on the above discussion, the greedy algorithm in [13] has been modified as shown in Table I. It differs from the original version for the addition of the WHILE loop. Let Σ_ℓ be the system consisting of the remaining inequalities after having extracted $\ell - 1$ feasible subsystems from (7), and let θ_ℓ be a (suboptimal) solution of the MAX FS problem for system Σ_ℓ (see later Algorithm 2 in Table II). The solution θ_ℓ is applied to the systems Σ_i with $i < \ell$ (WHILE loop). Note that Σ_ℓ is a subsystem of Σ_i for all $i < \ell$, so that θ_ℓ satisfies at least as many complementary inequalities in Σ_i as in Σ_ℓ. Let i^* be the first index i, if any, such that θ_ℓ satisfies a larger number of complementary inequalities in Σ_i than those satisfied by θ_i. Then, the best current solution θ_{i^*} for system Σ_{i^*} is set equal to θ_ℓ, and ℓ is reset to i^*. Since the number of data points is finite, this algorithm terminates in a finite number of steps. Improvements obtained by the proposed modification to the original algorithm are twofold. First, the cardinality of successively extracted subsystems decreases, as one would expect if each MAX FS problem were solved exactly. Second, it favors the construction of subsystems with larger cardinality (*e.g.*, when subsystems of complementary inequalities that could be satisfied by the same parameter vector are extracted at different MAX FS instances due to the suboptimality of Algorithm 2).

2) A relaxation algorithm for the MAX FS problem: The second improvement discussed above is also pursued by suitably modifying the algorithm for the MAX FS problem in [13]. The modified version, shown in Table II, differs from the original version for the addition of the final IF statement. The algorithm requires to define a maximum number of cycles $C > 0$, an initial temperature parameter $T_0 > 0$, an initial estimate $\theta^{(0)} \in \mathbb{R}^{n+1}$, and $\rho \in (0,1)$. During each of the C cycles, all the complementary inequalities of the current subsystem Σ_ℓ of (7) are selected in the order defined by a prescribed rule (*e.g.*, cyclicly, or uniformly at random without replacement). A sequence of parameter estimates $\theta^{(j)}$, where j is the iteration counter, is generated. At iteration j, the violation of the k-th complementary inequality is defined as

$$v_j^k = \begin{cases} \varphi_k'\theta^{(j-1)} - y_k - \delta & \text{if } \varphi_k'\theta^{(j-1)} > y_k + \delta \\ \varphi_k'\theta^{(j-1)} - y_k + \delta & \text{if } \varphi_k'\theta^{(j-1)} < y_k - \delta \\ 0 & \text{otherwise}. \end{cases} \quad (8)$$

If the cycle counter c is greater than ρC, the current best solution θ_{best} (*i.e.*, the one that has satisfied the largest number of complementary inequalities so far) is replaced by the projection estimate (6), that is likely to satisfy a larger number of complementary inequalities of Σ_ℓ. It was found experimentally that suitable values for ρ lie between 0.7 and 0.8. Indeed, the current solution $\theta^{(j)}$ (and hence the number of satisfied complementary inequalities of Σ_ℓ) would not change significantly as c approaches C, because the temperature parameter T, to which violations are compared, becomes smaller and smaller. By resetting $\theta^{(j)}$ to the current best solution θ_{best} at the exit of a cycle when c approaches C, one focuses the future search in a neighborhood of θ_{best}, where it is more likely to satisfy a larger number of complementary inequalities. For the choice of T_0, the reader is referred to [13]. In general, the larger the value of C, the better the solution, at the price of a longer computation time. Typical good choices for C are $C = 10 \div 20$ [19].

B. Refinement

Algorithm 1 returns the number s of submodels, and the sets of indices $\mathscr{I}_i$, $i = 1,\dots,s$, characterizing the s feasible subsystems extracted from (7). These provide the initial classification of the N data points $(y_k, \mathbf{x}_k)$ into the s clusters $\mathscr{D}_i^{(0)} = \{(y_k, \mathbf{x}_k) : k \in \mathscr{I}_i\}$, $i = 1,\dots,s$. As it was discussed in [12], there may exist data points $(y_k, \mathbf{x}_k)$ that satisfy the complementary inequality $|y_k - \varphi_k'\theta_i| \le \delta$ for more than one $i = 1,\dots,s$. These data points will be termed *undecidable*. The ambiguity in classifying undecidable data points cannot be overcome by exploiting the partition of the regressor set, because this is not available at this stage of

TABLE II

MODIFIED RANDOMIZED AND THERMAL RELAXATION ALGORITHM FOR THE MAX FS PROBLEM WITH COMPLEMENTARY INEQUALITIES

Algorithm 2

GIVEN: $C > 0$, $T_0 > 0$, $\theta^{(0)} \in \mathbb{R}^{n+1}$, $\rho \in (0,1)$
Let $j = 0$, $\theta_{best} = \theta^{(0)}$ and $\mathscr{I}_{best} = \{k \in \mathscr{T}_\ell : |y_k - \varphi_k' \theta_{best}| \leq \delta\}$
FOR $c = 0$ TO $C-1$ DO
  Let $\mathscr{T} = \mathscr{T}_\ell$ and $T = (1 - c/C)T_0$
  REPEAT
    Let $j = j+1$
    Pick an index k from $\mathscr{T}$ according to the prescribed rule
    Compute v_j^k and $\lambda_j = \text{sign}(v_j^k)(T/T_0)\exp(-|v_j^k|/T)$
    Let $\theta^{(j)} = \theta^{(j-1)} - \lambda_j \varphi_k$ and $\mathscr{I}^{(j)} = \{k \in \mathscr{T}_\ell : |y_k - \varphi_k' \theta^{(j)}| \leq \delta\}$
    IF $\#\mathscr{I}^{(j)} > \#\mathscr{I}_{best}$ THEN let $\theta_{best} = \theta^{(j)}$ and $\mathscr{I}_{best} = \mathscr{I}^{(j)}$
    Let $\mathscr{T} = \mathscr{T} \backslash \{k\}$
  UNTIL $\mathscr{T} = \emptyset$
  IF $c > \rho C$ THEN
    Let $\mathscr{D} = \{(y_k, \mathbf{x}_k) : k \in \mathscr{I}_{best}\}$
    Let $\theta_{best} = \Phi_p(\mathscr{D})$ and $\theta^{(j)} = \theta_{best}$
    Let $\mathscr{I}_{best} = \{k \in \mathscr{T}_\ell : |y_k - \varphi_k' \theta_{best}| \leq \delta\}$
  END IF
END FOR
RETURN θ_{best}

the identification procedure. When solving Problem 2, undecidable data points are classified depending on the order in which the feasible subsystems are extracted from (7). As an alternative, each undecidable data point $(y_k, \mathbf{x}_k)$ could be associated a posteriori to the submodel i^* such that the identification error is minimized, namely:

$$i^* = \arg \min_{i=1,\ldots,s} |y_k - \varphi_k' \theta_i| \,. \tag{9}$$

Both criteria may lead to misclassifications when the partition of the regressor set is estimated (see Fig. 1). Thus, in [12], undecidable data points were discarded during the classification process. Although this approach works well in many cases, a non-negligible amount of information is lost when a large number of undecidable data points shows up. In view of the estimation of the regions, the idea in this paper is to exploit the spatial localization of the regression vectors in order to associate undecidable data points to submodels. To this aim, a new step is added to the refinement procedure proposed in [12].

1) Dealing with undecidable data: The complete refinement procedure is detailed in Table III. Its aim is to iteratively improve both data classification and quality of fit by properly reassigning the data points (steps 2 and 4), and updating the parameter estimates (step 5). It also enables to reduce the number of submodels by exploiting parameter similarities (step 1, where μ is a suitable distance in $\mathbb{R}^{n+1}$) and cluster cardinalities (step 3). While a thorough description of Algorithm 3 and of the parameters involved can be found in [19], and partially in [12], the focus here is only on the additional step 4, which tries to overcome the ambiguity related to the undecidable data points by exploiting spatial localization in the regressor set. At iteration $t = 1, 2, \ldots$, a cluster $\mathscr{F}_i^{(t)}$ of feasible regression vectors can be associated

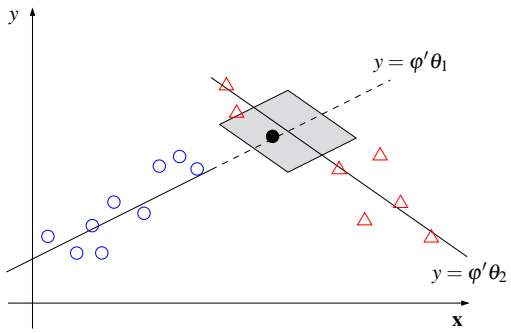

Fig. 1. PWA model with two discrete modes, $\mathbf{x} \in \mathbb{R}$. The gray set represents the region of all possible undecidable data points for a fixed δ. By applying both Algorithm 1 and the criterion (9), the only undecidable data point in the data set (the black circle) is associated to the first submodel. This yields two non-linearly separable clusters of points (circles and triangles).

to each cluster $\mathscr{D}_i^{(t)}$ of feasible data points as follows:

$$\mathscr{F}_i^{(t)} = \{\mathbf{x}_k : (y_k, \mathbf{x}_k) \in \mathscr{D}_i^{(t)}\}, \quad i = 1, \ldots, s. \tag{10}$$

In step 4, for each undecidable data point $(y_k, \mathbf{x}_k)$ one computes the set $\mathscr{C}(\mathbf{x}_k)$ of the c feasible regression vectors that are closest to $\mathbf{x}_k$. Here, c is a fixed positive integer, and the Euclidean distance is used. The feasible points around $\mathbf{x}_k$ are expected to provide useful information for correctly classifying the undecidable data point $(y_k, \mathbf{x}_k)$. A set $\mathscr{C}(\mathbf{x}_k)$ may in principle contain regression vectors from different clusters $\mathscr{F}_i^{(t)}$. Hence, the sets $\mathscr{C}_i(\mathbf{x}_k) = \mathscr{C}(\mathbf{x}_k) \bigcap \mathscr{F}_i^{(t)}$ are computed, $i = 1, \ldots, s$, and the index i^* maximizing $\#\mathscr{C}_i(\mathbf{x}_k)$ is considered. If the undecidable data point $(y_k, \mathbf{x}_k)$ is consistent with the i^*-th submodel, then it is assigned to $\mathscr{D}_{i^*}^{(t)}$, otherwise it is left undecidable. Good choices for the parameter c in step 4 depend on the density of the data set. A small c may originate sets $\mathscr{C}(\mathbf{x}_k)$ not containing enough points for correct classification. On the other hand, for large values of c, a set $\mathscr{C}(\mathbf{x}_k)$ might contain points distant from $\mathbf{x}_k$. In this case, the data point $(y_k, \mathbf{x}_k)$ could be badly assigned to a "far" cluster, or left undecidable. Indeed, if many data points are still classified as undecidable at the exit of the refinement procedure, one can reduce c, and repeat Algorithm 3.

Remark 3.1: The last step of the identification procedure aims at finding a complete partition of the regressor set into s polyhedral regions (4) such that $\mathbf{x}_k \in \mathscr{X}_i$ if $(y_k, \mathbf{x}_k) \in \mathscr{D}_i$. If this is not possible, the number of misclassifications or some misclassification cost is minimized. This problem can be viewed as that of separating s sets of points by means of linear classifiers (hyperplanes), which is widely addressed in the literature. Robust Linear Programming (RLP) [14] and Support Vector Machines (SVM) [15] methods (and their extensions to the multi-class case [16], [17]) can be employed. The minimization of the number of misclassifications is equivalent to solving a MAX FS problem. The interested reader is referred to [5], [12], [19] for a detailed discussion. □

TABLE III

ALGORITHM FOR THE REFINEMENT OF THE ESTIMATES

Algorithm 3

GIVEN: $\alpha \geq 0$, $\beta \geq 0$, $\gamma > 0$, $c > 0$

Let $t = 1$ and $\theta_i^{(1)} = \Phi_p(\mathscr{D}_i^{(0)})$, $i = 1,\dots,s$

1) Compute $\alpha_{i^*,j^*} = \min_{1\leq i<j\leq s} \mu(\theta_i^{(t)}, \theta_j^{(t)})$
   IF $\alpha_{i^*,j^*} \leq \alpha$
   THEN merge submodels i^* and j^*, and let $s = s-1$
2) For each data point $(y_k, \mathbf{x}_k)$, $k = 1,\dots,N$:
   - IF $\left|y_k - \varphi_k' \theta_i^{(t)}\right| \leq \delta$ for only one $i = 1,\dots,s$ THEN mark $(y_k, \mathbf{x}_k)$ as *feasible* and assign it to $\mathscr{D}_i^{(t)}$
   - IF $\left|y_k - \varphi_k' \theta_i^{(t)}\right| \leq \delta$ for more than one $i = 1,\dots,s$ THEN mark $(y_k, \mathbf{x}_k)$ as *undecidable*
   - OTHERWISE mark $(y_k, \mathbf{x}_k)$ as *infeasible*
3) Compute $\beta_{i^*} = \min_{i=1,\dots,s} \#\mathscr{D}_i^{(t)}/N$
   IF $\beta_{i^*} \leq \beta$
   THEN discard submodel i^*, let $s = s-1$ and go to step 2
4) For each undecidable data point $(y_k, \mathbf{x}_k)$:
   Compute $\mathscr{C}_i(\mathbf{x}_k)$, $i = 1,\dots,s$, and $i^* = \arg\max_{i=1,\dots,s} \#\mathscr{C}_i(\mathbf{x}_k)$
   IF $\left|y_k - \varphi_k' \theta_{i^*}^{(t)}\right| \leq \delta$
   THEN mark $(y_k, \mathbf{x}_k)$ as *feasible* and assign it to $\mathscr{D}_{i^*}^{(t)}$
5) Compute $\theta_i^{(t+1)} = \Phi_p(\mathscr{D}_i^{(t)})$, $i = 1,\dots,s$
6) IF $\left\|\theta_i^{(t+1)} - \theta_i^{(t)}\right\| \leq \gamma \left\|\theta_i^{(t)}\right\|$ for all $i = 1,\dots,s$
   THEN RETURN s, $\theta_i = \theta_i^{(t+1)}$ and $\mathscr{D}_i = \mathscr{D}_i^{(t)}$, $i = 1,\dots,s$
   ELSE let $t = t+1$ and go to step 1

IV. APPLICATION

The proposed identification procedure is applied to the identification of an electronic component placement process in a pick-and-place machine [18]. Pick-and-place machines are used to automatically place electronic components on printed circuit boards. The process consists of a mounting head carrying the electronic component, which is pushed down until it comes in contact with the circuit board, and then is released. Input-output data are collected from a real experimental setup consisting of a mounting head and an impacting surface simulating the printed circuit board. The input to the system is the voltage applied to the motor driving the mounting head. The output of the system is the position of the mounting head. Four operating conditions of the system can be distinguished. In the *free* mode, the mounting head moves unconstrained, *i.e.*, without being in contact with the impacting surface. In the *impact* mode, the mounting head moves in contact with the impacting surface. The *upper* and *lower saturation* modes correspond to situations in which the mounting head cannot move upwards or downwards, respectively, due to physical constraints. The reader is referred to [18] for a complete description of the experimental setup.

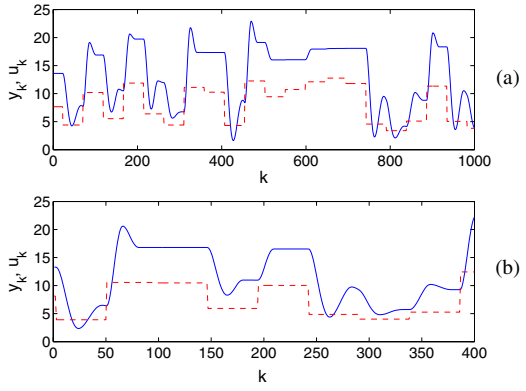

Fig. 2. Data sets used for (a) identification and (b) validation (solid line - system output, dashed line - scaled input).

Only two operating conditions are excited in the considered data set, namely the free mode and the impact mode. Input-output data used for identification and validation are plotted in Fig. 2. Nonlinear phenomena due to dry friction damping are evident in both data sets, *e.g.*, in Fig. 2(a) on the interval $(500, 750)$. A PWARX model structure with orders $n_a = 2$ and $n_b = 2$ is considered. By choosing $\delta = 0.06$, $\delta = 0.05$, and $\delta = 0.04$, models with $s = 2$, $s = 3$, and $s = 4$ discrete modes, respectively, are identified from $N = 1000$ estimation data. For $s = 3$ and $s = 4$, multicategory RLP [16] is used to estimate the partition of the regressor set. Validation is then carried out by evaluating the fit between the measured and the simulated responses using $N_v = 400$ validation data. By letting $\mathbf{y} = (y_1, \dots, y_{N_v})$ be the vector of system outputs, $\bar{y}$ the mean value of $\mathbf{y}$, and $\hat{\mathbf{y}} = (\hat{y}_1, \dots, \hat{y}_{N_v})$ the vector of simulated outputs, the values of the figure of fit [20]

$$FIT = 100 \cdot \left(1 - \frac{\|\hat{\mathbf{y}} - \mathbf{y}\|}{\|\mathbf{y} - \bar{y}\|}\right) \tag{11}$$

for the three identified models are shown in Table IV, where it is apparent that the fit improves as the number of submodels increases, *i.e*, as δ decreases. In Fig. 3, the plots of the simulated responses are graphically compared to the measured response. Fig. 3(a) clearly shows that two affine submodels are not sufficient for accurately reproducing the system dynamics. Very good accordance between the measured and the simulated responses is instead obtained with $s = 3$ and $s = 4$ submodels.

Difficulties of the identified models in reproducing the nonlinear phenomena on the interval $(210, 240)$ are likely to be due to incomplete information provided by the estimation data. Indeed, in the estimation data set shown in Fig. 2(a), all significant transitions of the output from low to high values exhibit an overshoot. Consequently, an overshoot shows up in the simulated responses on the intervals $(60, 140)$ and $(210, 240)$, that are caused by large positive variations of

TABLE IV

FIT BETWEEN THE MEASURED AND THE SIMULATED RESPONSES

	FIT
$s = 2$	81.33%
$s = 3$	90.18%
$s = 4$	93.48%

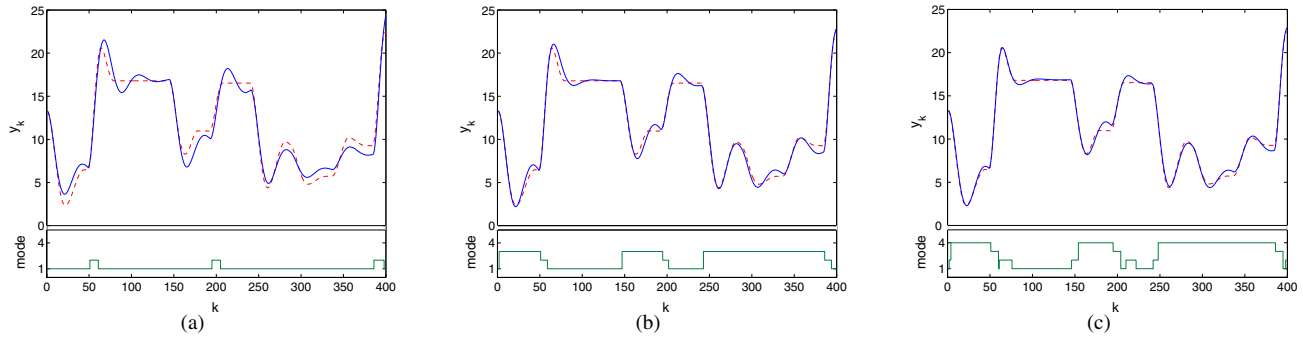

Fig. 3. Simulation results of the identified PWARX models with (a) $s = 2$, (b) $s = 3$, and (c) $s = 4$ submodels (solid line - simulated output, dashed line - system output). The plot at the bottom of each figure shows the evolution of the discrete mode.

the input signal. It is interesting to note that the identified model with $s = 4$ discrete modes is able to reproduce the peak in the interval $(60, 140)$ very accurately. The discrete mode evolution at the bottom of Fig. 3(a) clearly shows that one of the two submodels is active in situations of high incoming velocity of the mounting head (*i.e.*, rapid transitions from low to high values of the mounting head position). One submodel modelling the same situation is also present in the identified models with $s = 3$ and $s = 4$ discrete modes.

V. CONCLUSIONS

In this paper, a procedure for the identification of PWARX models was presented and discussed. The greedy algorithm for the MIN PFS problem was modified in order to improve the estimation of the number of submodels. Problems arising in the classification of the undecidable data points were highlighted, and a modification of the classification procedure was proposed, which associates undecidable data points to submodels by exploiting spatial localization of the regression vectors. Interesting results were obtained by applying the identification procedure on experimental data. These results demonstrated the effectiveness of the procedure and the use of the bound δ as a tuning knob to trade off between model complexity and quality of fit.

Future research will concern the possibility to include prior knowledge on the system to be identified (*e.g.*, saturations in the case study of Section IV) in the identification procedure, to identify submodels of different orders for each mode, as well as to handle non-convex regions in the partition of the regressor set.

ACKNOWLEDGMENTS

Authors would like to thank Dr. Aleksandar Juloski for providing the data used for identification in Section IV. Data are by courtesy of Assembleon (`www.assembleon.com`).

REFERENCES

[1] E. D. Sontag, "Nonlinear regulation: The piecewise linear approach," *IEEE Trans. Automat. Contr.*, vol. 26, no. 2, pp. 346–358, 1981.

[2] A. Bemporad, G. Ferrari-Trecate, and M. Morari, "Observability and controllability of piecewise affine and hybrid systems," *IEEE Trans. Automat. Contr.*, vol. 45, no. 10, pp. 1864–1876, 2000.

[3] W. P. M. H. Heemels, B. De Schutter, and A. Bemporad, "Equivalence of hybrid dynamical models," *Automatica*, vol. 37, no. 7, pp. 1085–1091, 2001.

[4] J. Roll, "Local and piecewise affine approaches to system identification," Ph.D. dissertation, Department of Electrical Engineering, Linköping University, Sweden, 2003.

[5] 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.

[6] R. Vidal, S. Soatto, Y. Ma, and S. Sastry, "An algebraic geometric approach to the identification of a class of linear hybrid systems," in *Proc. 42nd IEEE Conf. Decision Contr.*, 2003, pp. 167–172.

[7] J. Ragot, G. Mourot, and D. Maquin, "Parameter estimation of switching piecewise linear systems," in *Proc. 42nd IEEE Conf. Decision Contr.*, 2003, pp. 5783–5788.

[8] A. Lj. Juloski, S. Wieland, and W. P. M. H. Heemels, "A Bayesian approach to identification of hybrid systems," in *Proc. 43rd IEEE Conf. Decision Contr.*, 2004.

[9] J. Roll, A. Bemporad, and L. Ljung, "Identification of piecewise affine systems via mixed-integer programming," *Automatica*, vol. 40, pp. 37–50, 2004.

[10] M. Milanese and A. Vicino, "Optimal estimation theory for dynamic systems with set membership uncertainty: an overview," *Automatica*, vol. 27, no. 6, pp. 997–1009, 1991.

[11] M. Milanese, J. P. Norton, H. Piet-Lahanier, and E. W. (eds.), *Bounding Approaches to System Identification*. New York: Plenum Press, 1996.

[12] A. Bemporad, A. Garulli, S. Paoletti, and A. Vicino, "A greedy approach to identification of piecewise affine models," in *Hybrid Systems: Computation and Control*, ser. Lect. Notes Comput. Sci., O. Maler and A. Pnueli, Eds. Springer Verlag, 2003, pp. 97–112.

[13] E. Amaldi and M. Mattavelli, "The MIN PFS problem and piecewise linear model estimation," *Discr. Appl. Math.*, vol. 118, pp. 115–143, 2002.

[14] K. P. Bennett and O. L. Mangasarian, "Robust linear programming discrimination of two linearly inseparable sets," *Optim. Methods Softw.*, vol. 1, pp. 23–34, 1992.

[15] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995.

[16] K. P. Bennett and O. L. Mangasarian, "Multicategory discrimination via linear programming," *Optim. Methods Softw.*, vol. 3, pp. 27–39, 1994.

[17] E. J. Bredensteiner and K. P. Bennett, "Multicategory classification by support vector machines," *Comput. Optim. Appl.*, vol. 12, pp. 53–79, 1999.

[18] A. Lj. Juloski, W. P. M. H. Heemels, and G. Ferrari-Trecate, "Data-based hybrid modelling of the component placement process in pick-and-place machines," *Contr. Eng. Pract.*, vol. 12, no. 10, pp. 1241–1252, 2004.

[19] S. Paoletti, "Identification of piecewise affine models," Ph.D. dissertation, Department of Information Engineering, University of Siena, Italy, 2004.

[20] L. Ljung, *System Identification Toolbox User's Guide*, 6th ed. The MathWorks, Inc., 2003.