

# SET MEMBERSHIP IDENTIFICATION OF PIECEWISE AFFINE MODELS

**Alberto Bemporad, Andrea Garulli, Simone Paoletti,  
Antonio Vicino**

*Dipartimento di Ingegneria dell'Informazione,  
Università di Siena  
Via Roma 56, 53100 Siena, Italy  
{bemporad,garulli,paoletti,vicino}@dii.unisi.it*

**Abstract:** This paper addresses the problem of identification of piecewise affine (PWA) models, which involves the joint estimation of both the parameters of the affine submodels and the partition of the PWA map from data. According to ideas from set-membership identification, the key approach is to characterize the model by its maximum allowed prediction error, which is used as a tuning knob for trading off between prediction accuracy and model complexity. At initialization, the proposed procedure for PWA identification exploits a technique for partitioning an infeasible system of linear inequalities into a (possibly minimum) number of feasible subsystems. This provides both an initial clustering of the datapoints and a guess of the number of required submodels, which therefore is not fixed a priori. A refinement procedure is then applied in order to improve both data classification and parameter estimation. The partition of the PWA map is finally estimated by considering multicategory classification techniques. *Copyright ©2003 IFAC*

**Keywords:** Nonlinear identification, hybrid systems, bounded error, data classification, parameter estimation

## 1. INTRODUCTION

Black-box identification of nonlinear systems has been widely addressed in different contexts. A large number of model classes have been considered and their properties deeply investigated, see the survey papers (Sjöberg *et al.*, 1995; Juditsky *et al.*, 1995) and references therein. In this paper, we deal with the problem of identifying a piecewise affine (PWA) model of a discrete-time nonlinear system from input-output data. Recently, this problem has deserved more and more attention, given the equivalence of PWA systems with several classes of hybrid systems (Bemporad *et al.*, 2000; Heemels *et al.*, 2001). Identification of PWA models is a challenging problem, as it involves the simultaneous estimation of both the parameters of the affine submodels and the partition of the PWA map. It is crucial to point out that the first

issue is closely related to the problem of classifying the data, *i.e.* the problem of correctly assigning each datapoint to one affine submodel. In (Ferrari-Trecate *et al.*, 2003) the classification problem is reduced to an optimal clustering problem, in which the number of clusters is fixed. Once the datapoints have been classified, linear regression is used to compute the final submodels. In (Münz and Krebs, 2002) the authors propose an identification algorithm consisting of analysis of the knowledge available a priori, data clustering, and optimization of the cluster shapes. The user can affect the accuracy of the identification by modifying the number of clusters. In (Bemporad *et al.*, 2001) the attention is focused on two subclasses of PWA models, for which the identification problem can be formulated as a suitable mixed-integer linear, or quadratic, programming problem that can be solved for the global optimum.

The procedure for PWA identification proposed in this paper does not fix the number of affine submodels a priori, rather this number is estimated from data, together with the parameters of the submodels and the partition of the PWA map. The key approach in this work is the selection of a bound on the prediction error, which induces a set of linear inequality constraints on the parameters of the PWA model to be estimated. Unless a single affine model fits all the data within the chosen bound, the whole set of constraints is, in general, infeasible. Hence, in Section 3 a suitable strategy is suggested for picking a number of submodels which is consistent with the available data and the bounded-error condition. In particular, a partition of the above system of linear inequalities into a minimum number of feasible subsystems (Amaldi and Mattavelli, 2002) is sought. Given any solution of this problem, the partition of the inequalities provides the initial classification of the datapoints, whereas a set of feasible parameter vectors for the corresponding affine submodel is associated to each feasible subsystem according to the bounded-error condition. In Section 4 a refinement procedure alternating between datapoint reassignment and parameter update is proposed in order to improve both data classification and parameter estimation, and to possibly reduce the number of submodels. Notice that the final number of submodels and the corresponding parameter vectors will depend on the selected bound on the prediction error, so that this allows one to trade off between the complexity and the accuracy of the model. The estimation of the partition of the PWA map is addressed in Section 5. The final clusters of regression vectors are separated via classification methods such as Linear (Vapnik, 1998) or Multicategory (Bredensteiner and Bennett, 1999) Support Vector Machines. The identified PWA model associates to each submodel a set of feasible parameters, thus allowing for evaluation of the parametric uncertainty associated with it (Milanese and Vicino, 1991).

## 2. PROBLEM STATEMENT

Consider the discrete-time nonlinear dynamic system

$$y_k = F(u^{k-1}, y^{k-1}) + e_k, \quad (1)$$

where  $F$  is a (possibly non-smooth) nonlinear function,  $u^{k-1}$  and  $y^{k-1}$  are, respectively, past system inputs and outputs up to time  $k-1$ , and  $e_k$  is additive noise. Assuming that a finite collection of input-output samples  $(u_k, y_k)$  is given, the aim is to estimate a PieceWise Affine (PWA) model  $\hat{y}_k = f(x_k)$  of system (1), where  $f$  is the PWA map

$$f(x) = \begin{cases} \theta'_1 \begin{bmatrix} x \\ 1 \end{bmatrix} & \text{if } x \in \mathcal{X}_1 \\ \vdots & \vdots \\ \theta'_s \begin{bmatrix} x \\ 1 \end{bmatrix} & \text{if } x \in \mathcal{X}_s, \end{cases} \quad (2)$$

$x_k \in \mathbb{R}^n$  is a suitable regression vector depending on  $u^{k-1}$  and  $y^{k-1}$ ,  $\theta_i \in \mathbb{R}^{n+1}$ ,  $i = 1, \dots, s$ , are parameter

vectors, and  $\{\mathcal{X}_i\}_{i=1}^s$  is a partition of the regressor set  $\mathcal{X} \subseteq \mathbb{R}^n$ , *i.e.* the region of validity of the PWA model<sup>1</sup>. Each *region*  $\mathcal{X}_i$  is assumed to be a convex polyhedron, represented in the form

$$\mathcal{X}_i = \left\{ x \in \mathbb{R}^n : H_i \begin{bmatrix} x \\ 1 \end{bmatrix} \preceq 0 \right\},$$

where  $H_i \in \mathbb{R}^{q_i \times (n+1)}$ ,  $i = 1, \dots, s$ . In this paper, the focus is on PWARX (PieceWise affine AutoRegressive eXogenous) models, for which  $x_k$  is the standard regression vector, *i.e.*

$$x_k = [y_{k-1} \dots y_{k-n_a} \ u_{k-1} \dots u_{k-n_b}]'.$$

In this case,  $n = n_a + n_b$ , and the parameter vectors  $\theta_i$  contain the coefficients of the *ARX submodels*. For a more compact notation, hereafter the extended regression vector  $\varphi_k = [x_k' \ 1]'$  will be considered. The key approach in this paper consists in selecting a bound on the *prediction error*,

$$|y_k - f(x_k)| \leq \delta, \quad (3)$$

for a fixed  $\delta > 0$ . Accordingly, the considered identification problem can be stated as follows.

*Problem 1.* Given  $N$  datapoints  $(y_k, x_k)$ ,  $k = 1, \dots, N$ , and  $\delta > 0$ , estimate a minimum positive integer  $s$ , a partition  $\{\mathcal{X}_i\}_{i=1}^s$ , and parameter vectors  $\{\theta_i\}_{i=1}^s$ , such that the corresponding PWA model (2) of system (1) satisfies condition (3).

From the above formulation, it is clear that one seeks the “simplest” PWA model that is consistent with the data and condition (3), where, for a given  $\delta$ , “simplicity” is measured in terms of the number of affine submodels. Notice that the bound  $\delta$  is not necessarily given a priori, it is rather a tuning knob of the procedure. A reliable choice can often be made a posteriori by performing a series of trials for different values of  $\delta$ , and then selecting a value that provides a suitable trade-off between model complexity (in terms of number of submodels) and quality of fit (in terms of mean square error).

The proposed procedure for solving Problem 1 consists of the following steps: *Initialization*, exploiting a greedy strategy for partitioning a (possibly infeasible) system of linear inequalities into a minimum number of feasible subsystems; *Refinement of the estimates*, alternating between datapoint reassignment and parameter update; *Estimation of the regions*, exploiting multicategory classification techniques.

## 3. INITIALIZATION

In the first part of the proposed identification procedure, the problem of estimating the hyperplanes defining the polyhedral partition of the regressor set is provisionally not addressed. The focus is only on classifying the datapoints according to the fact that

<sup>1</sup>  $\bigcup_{i=1}^s \mathcal{X}_i = \mathcal{X}$  and  $\mathcal{X}_i \cap \mathcal{X}_j = \emptyset$  if  $i \neq j$

they are fitted by the same affine submodel. In this phase, it is reasonable to look for the minimum number of submodels (namely  $s$ ) fitting all (or most of, due to possible outliers) the datapoints. By requiring condition (3), the initial classification problem can be stated as follows.

*Problem 2.* Given  $\delta > 0$  and the system of  $N$  complementary inequalities

$$|y_k - \varphi'_k \theta| \leq \delta, \quad k = 1, \dots, N, \quad (4)$$

find a partition of this system into a minimum number  $s$  of feasible subsystems.

The above formulation enables to address simultaneously the two fundamental issues of data classification and parameter estimation. Given any solution of Problem 2, the partition of the complementary inequalities provides the classification of the datapoints, whereas each feasible subsystem defines the set of feasible parameter vectors for the corresponding affine submodel, according to the bounded-error condition. This naturally leads to a set-membership or bounded-error approach to the identification problem, see, e.g., (Milanese and Vicino, 1991; Milanese *et al.*, 1996). Since each complementary inequality in system (4) corresponds to the pair of linear inequalities

$$\begin{cases} \varphi'_k \theta \leq y_k + \delta \\ \varphi'_k \theta \geq y_k - \delta, \end{cases} \quad (5)$$

Problem 2 turns out to be an extension of the combinatorial problem of finding a *Partition* of an infeasible system of linear inequalities into a *MINimum* number of *Feasible Subsystems* (MIN PFS problem), with the additional constraint that two paired linear inequalities (5) must be included in the same subsystem (*i.e.* they must be simultaneously satisfied by the same parameter vector  $\theta$ ). The MIN PFS problem is NP-hard. On the other hand, in order to initialize the identification procedure, one is interested in finding approximate solutions of Problem 2 rapidly. To this aim, the greedy approach proposed in (Amaldi and Mattavelli, 2002), which efficiently provides good approximate solutions, is used in this paper. This approach divides the overall partition problem into a sequence of subproblems, each subproblem consisting in finding one parameter vector  $\theta$  that satisfies the maximum number of complementary inequalities. Starting from system (4), maximum feasible subsystems are iteratively extracted (and the corresponding inequalities removed), until the remaining subsystem is feasible. The problem of finding one  $\theta$  that satisfies as many pairs of complementary inequalities (4) as possible, extends the combinatorial problem of finding a *MAXimum Feasible Subsystem* of an infeasible system of linear inequalities (MAX FS problem). Based on the consideration that also MAX FS is NP-hard, the approach proposed in (Amaldi and Mattavelli, 2002) tackles the above extension of MAX FS using a randomized and thermal variant of the classical Agmon-

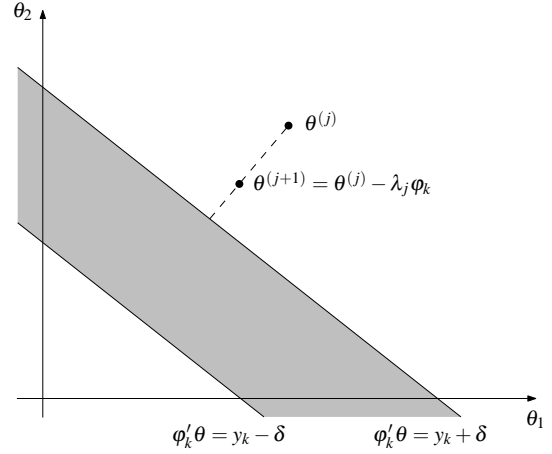


Fig. 1. Geometric interpretation in the parameter space of a single iteration of the relaxation method for the extended MAX FS (case  $\theta \in \mathbb{R}^2$ )

Motzkin-Schoenberg relaxation method for solving systems of linear inequalities. The proposed method consists in a simple iterative procedure that generates a sequence of estimates. Starting with an arbitrary initial guess  $\theta^{(1)} \in \mathbb{R}^{n+1}$  (*e.g.*, randomly selected, or computed by least squares), at each iteration one complementary inequality is selected from the system at hand according to a prescribed rule (*e.g.*, cyclicly, or uniformly at random without replacement), while all the others are relaxed. Assume that at iteration  $j$  the complementary inequality  $|y_k - \varphi'_k \theta| \leq \delta$  is considered. Then the current estimate  $\theta^{(j)}$  is updated as follows:

$$\begin{aligned} &\text{if } \varphi'_k \theta^{(j)} > y_k + \delta \\ &\quad \text{then } \theta^{(j+1)} = \theta^{(j)} - \lambda_j \varphi_k \\ &\text{else if } \varphi'_k \theta^{(j)} < y_k - \delta \\ &\quad \text{then } \theta^{(j+1)} = \theta^{(j)} + \lambda_j \varphi_k \\ &\text{else } \theta^{(j+1)} = \theta^{(j)} \end{aligned}$$

with  $\lambda_j > 0$ . Geometrically, the complementary inequality  $|y_k - \varphi'_k \theta| \leq \delta$  defines a hyperstrip in the parameter space (see Figure 1). If the current estimate  $\theta^{(j)}$  belongs to the hyperstrip, then it is unchanged. Otherwise,  $\theta^{(j+1)}$  is obtained by making a step along the line drawn from  $\theta^{(j)}$  in the direction orthogonal to the hyperstrip. The size of the step  $\lambda_j$  decreases exponentially with the violation of the considered complementary inequality, which is computed as follows:

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

The basic idea is indeed to favor updates of the current estimate  $\theta^{(j)}$ , which aim at correcting unsatisfied inequalities with a relatively small violation. Decreasing attention to unsatisfied inequalities with large violations (whose correction is likely to corrupt other inequalities that the current estimate satisfies) is obtained by introducing a decreasing temperature parameter  $T$ , which the violations are compared to.

The algorithm is stopped after a predefined maximum number of cycles  $C$  through all the inequalities at hand. The solution returned is the estimate that, during the process, has satisfied the largest number of complementary inequalities. Nevertheless, this is not guaranteed to be optimal, due to the randomness of the method. For the choice of the maximum number of cycles  $C$  and the initial temperature parameter  $T_0$ , as well as for more details concerning the implementation of the algorithm, the reader is deferred to (Amaldi and Mattavelli, 2002).

### 3.1 Comments about the initialization

Denote by  $\hat{s}$  the number of feasible subsystems of system (4) provided by the application of the greedy algorithm for the extended MIN PFS described in Section 3. Due to the suboptimality of the greedy strategy for MIN PFS, and the randomness of the method used to tackle the MAX FS, the estimate of the number of affine submodels needed to fit the data and the classification of the datapoints thus obtained, suffer two drawbacks. First, this strategy is not guaranteed to yield minimum partitions, *i.e.* the number of submodels  $\hat{s}$  could be larger than the minimum number  $s$ . Indeed, due to the randomness of the method used to tackle the MAX FS, two subsets of complementary inequalities that could be satisfied by one and the same parameter vector, may be extracted at two different steps. Second, since some datapoints might be consistent with more than one submodel, the cardinality and the composition of the clusters could depend on the order in which the feasible subsystems are extracted. In order to cope with these drawbacks, a procedure for the refinement of the estimates will be proposed in the next section. As it will be shown, such a procedure improves both the data classification and the quality of fit by properly reassigning the datapoints, and selecting pointwise estimates of the parameter vectors that characterize each submodel. Notice that one could decide to stop the algorithm when the cardinalities of the extracted clusters become too small. This might be useful in order to penalize submodels that account for just a few datapoints (most likely outliers).

## 4. REFINEMENT OF THE ESTIMATES

The initialization of the identification procedure, described in Section 3, provides the clusters  $\mathcal{D}_i^{(0)}$ , which consist of all the datapoints  $(y_k, x_k)$  corresponding to the  $i$ -th extracted feasible subsystem of system (4),  $i = 1, \dots, \hat{s}$ . Moreover, each feasible subsystem defines the set of feasible parameter vectors for the corresponding affine submodel. As discussed in Section 3.1, a procedure for the refinement of the estimates is necessary in order to improve both data classification and quality of fit, as well as to possibly reduce the number of submodels. The proposed basic procedure consists of two steps to be iterated. In the first step, using the current

estimated parameter vectors, datapoints are grouped together in the same cluster only if they can be fitted by the same affine submodel. In the second step, new pointwise parameter estimates are computed for each submodel. The *projection estimate* is used, defined as:

$$\Phi_p(\mathcal{D}) = \arg \min_{\theta} \max_{(y_k, x_k) \in \mathcal{D}} |y_k - \phi'_k \theta|, \quad (6)$$

where  $\mathcal{D}$  is a cluster of datapoints  $(y_k, x_k)$ . Notice that the computation of the projection estimate can be formulated as a suitable *linear programming* (LP) problem. The basic refinement procedure can be formalized as follows.

#### (0) Initialization

Set  $t = 1$  and select a termination threshold  $\gamma \geq 0$

For  $i = 1, \dots, \hat{s}$ , set  $\hat{\theta}_i^{(1)} = \Phi_p(\mathcal{D}_i^{(0)})$

#### (1) Datapoint reassignment

For each datapoint  $(y_k, x_k)$ ,  $k = 1, \dots, N$ :

- If  $|y_k - \phi'_k \hat{\theta}_i^{(t)}| \leq \delta$  for only one  $i = 1, \dots, \hat{s}$ , then assign  $(y_k, x_k)$  to  $\mathcal{D}_i^{(t)}$
- If  $|y_k - \phi'_k \hat{\theta}_i^{(t)}| > \delta$  for all  $i = 1, \dots, \hat{s}$ , then mark  $(y_k, x_k)$  as *infeasible*
- Otherwise, mark  $(y_k, x_k)$  as *undecidable*

#### (2) Parameter update

For  $i = 1, \dots, \hat{s}$ , compute  $\hat{\theta}_i^{(t+1)} = \Phi_p(\mathcal{D}_i^{(t)})$

#### (3) Termination

If  $\|\hat{\theta}_i^{(t+1)} - \hat{\theta}_i^{(t)}\| / \|\hat{\theta}_i^{(t)}\| \leq \gamma$  for all  $i = 1, \dots, \hat{s}$ , then exit; Otherwise, set  $t = t + 1$  and go to step 1

In order to avoid that the procedure does not terminate, a maximum number  $t_{\max}$  of refinements can be predefined. The underlying idea is that, as the new parameter estimates  $\hat{\theta}_i^{(t+1)}$  are computed based on the clusters  $\mathcal{D}_i^{(t)}$ , some infeasible, as well as undecidable, datapoint may become *feasible*, *i.e.* it may be assigned to one cluster  $\mathcal{D}_i^{(t+1)}$ , thus improving the quality of the classification. Notice that the use of the projection estimate in step 2 guarantees that no feasible datapoint at refinement  $t$  becomes infeasible at refinement  $t + 1$ , since

$$\max_{(y_k, x_k) \in \mathcal{D}_i^{(t)}} |y_k - \phi'_k \hat{\theta}_i^{(t+1)}| \leq \max_{(y_k, x_k) \in \mathcal{D}_i^{(t)}} |y_k - \phi'_k \hat{\theta}_i^{(t)}|,$$

and the right-hand side of the above inequality is less than or equal to  $\delta$ . Motivations for the distinction among infeasible, undecidable, and feasible datapoints are twofold. Infeasible datapoints are not consistent with any submodel, and may be outliers (especially if the corresponding violation is large). Hence, it is reasonable to expect that neglecting them in the parameter update helps to improve the quality of fit. The undecidable datapoints are instead consistent with more than one submodel. Neglecting them helps to reduce the number of misclassifications. As it will be clarified in the next section, this will favor a better estimation of the PWA partition.

When the greedy algorithm provides an overestimation of the number of submodels needed to fit the data (see the discussion in Section 3.1), further steps are

required in order to possibly reduce the complexity of the model. To this aim, the similarity of the parameter vectors and the cardinality of the clusters can be exploited. Indeed, if two subsets of complementary inequalities can be satisfied by one and the same parameter vector, it is likely that the corresponding parameter estimates are very similar, so that they can be merged into one subset. Notice that, in this case, a large number of undecidable datapoints should show up. On the other hand, if during the refinement of the estimates the cardinality of one cluster becomes too small with respect to  $N$ , the corresponding submodel can be discarded, since it accounts only for few datapoints (most likely outliers). Additional steps to the basic refinement procedure are thus the following ( $\alpha$  and  $\beta$  are fixed nonnegative thresholds).

- **Similarity of the parameter vectors**

If  $\alpha_{i^*,j^*} \triangleq \min_{1 \leq i < j \leq \hat{s}} \mu(\hat{\theta}_i^{(t)}, \hat{\theta}_j^{(t)}) \leq \alpha$ , then merge the submodels  $i^*$  and  $j^*$ , and set  $\hat{s} = \hat{s} - 1$

- **Cardinality of the clusters**

If  $\beta_{i^*} = \min_{1 \leq i \leq \hat{s}} \text{card}(\mathcal{D}_i^{(t)})/N \leq \beta$ , then discard the  $i^*$ -th submodel, set  $\hat{s} = \hat{s} - 1$ , and reassign only the undecidable datapoints as in step 1

The similarity of the parameter vectors is tested before step 1. Here  $\mu(\hat{\theta}_i^{(t)}, \hat{\theta}_j^{(t)})$  is a suitable measure of the distance between  $\hat{\theta}_i^{(t)}$  and  $\hat{\theta}_j^{(t)}$ , e.g.,

$$\mu(\hat{\theta}_i^{(t)}, \hat{\theta}_j^{(t)}) \triangleq \|\hat{\theta}_i^{(t)} - \hat{\theta}_j^{(t)}\| / \min\{\|\hat{\theta}_i^{(t)}\|, \|\hat{\theta}_j^{(t)}\|\}.$$

Two submodels  $i^*$  and  $j^*$  can be merged by computing the new parameter vector as  $\Phi_p(\mathcal{D}_{i^*}^{(t-1)} \cup \mathcal{D}_{j^*}^{(t-1)})$ . The cardinality of the clusters is instead tested after step 1. The thresholds  $\alpha$  and  $\beta$  should be suitably chosen in order to possibly reduce the number of submodels still preserving a good fit of the data. Indeed, it is clear that, if such thresholds are too large, the number of submodels might decrease under  $s$ . In this case, the number of infeasible datapoints considerably increases, since some significant dynamics is no more in the model. One could use this information in order to adjust  $\alpha$  and  $\beta$ , and then repeat the refinement. Current research is aimed at deriving rules for the automatic selection and update of  $\alpha$  and  $\beta$ , in order to completely automatize the procedure.

## 5. ESTIMATION OF THE REGIONS

The final step of the identification procedure consists in estimating the partition of the regressor set. This step can be performed by considering pairwise the clusters  $\mathcal{F}_i = \{x_k | (y_k, x_k) \in \mathcal{D}_i\}$  (where  $\mathcal{D}_i$ ,  $i = 1, \dots, \hat{s}$ , is the final classification of the feasible datapoints provided by the refinement procedure), and by computing a separating hyperplane for each of such pairs. If two clusters  $\mathcal{F}_i$  and  $\mathcal{F}_j$  are not linearly separable, it is reasonable to look for a *generalized* separating hyperplane of the two, i.e. a hyperplane

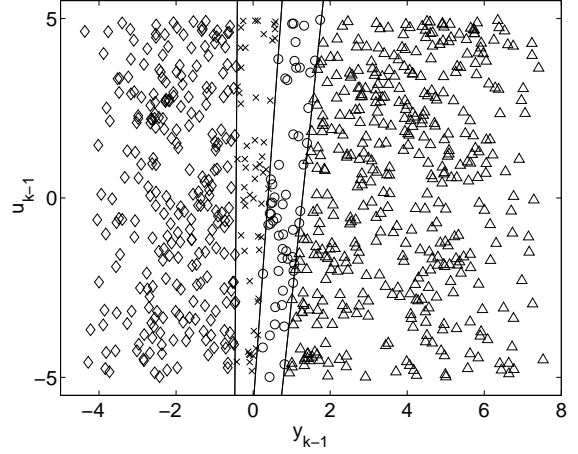


Fig. 2. Final classification of the regression vectors, and partition of the PWA map

that maximizes the number of well-separated points. This amounts to find a solution  $(a, b)$ , with  $a \in \mathbb{R}^n$  and  $b \in \mathbb{R}$ , of the MAX FS problem of system

$$\begin{cases} x'_k a + b \leq -1 & \forall x_k \in \mathcal{F}_i \\ x'_k a + b \geq 1 & \forall x_k \in \mathcal{F}_j. \end{cases}$$

Support Vector Machines (Vapnik, 1998) with a linear kernel are a suitable tool to accomplish this task. Notice that, when the number of misclassified points is large, it likely means that at least one of the two clusters corresponds either to a nonconvex region (which then needs to be split into convex polyhedra), or to nonconnected regions where the submodel is the same (recall that the classification procedure groups together all the datapoints fitted by the same affine submodel). Efficient techniques for detecting and splitting the clusters corresponding to such situations, are currently under investigation. Each region  $\mathcal{F}_i$ ,  $i = 1, \dots, \hat{s}$ , is finally defined by all the hyperplanes separating  $\mathcal{F}_i$  from  $\mathcal{F}_j$ , with  $j \neq i$ . Although computationally appealing, this method has the major drawback of not guaranteeing that the estimated regions form a complete partition of the regressor set. In order to avoid “holes” in the partition, all clusters can be simultaneously involved in a computationally more demanding *multicategory* classification problem, for whose solution both linear and quadratic programming based methods have been proposed (Bredensteiner and Bennett, 1999). Once the partition of the regressor set has been estimated, the undecidable datapoints can be finally classified, and final parameter estimates for each submodel can be computed using (6).

## 6. NUMERICAL EXAMPLE

The PWA identification algorithm has been applied to fit the data generated by the nonlinear system

$$y_k = \sqrt{|y_{k-1}|} - u_{k-1} + e_k. \quad (7)$$

The input signal  $u_k$  was drawn from a uniform distribution on  $[-5, 5]$ , and the noise signal  $e_k$  was drawn

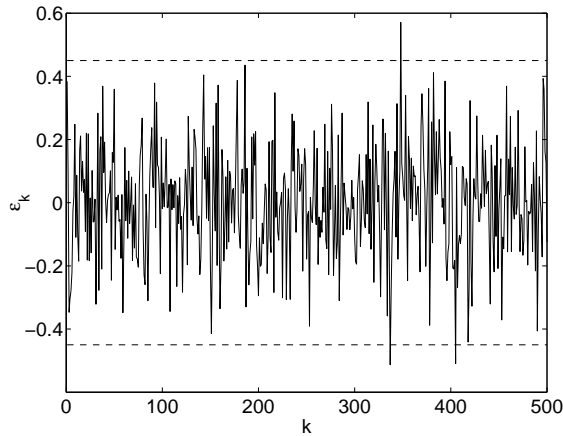


Fig. 3. Plot of the prediction error

from a zero-mean normal distribution with variance  $\sigma^2 = 0.025$ .  $N = 1000$  and  $N_v = 500$  datapoints were used for estimation and validation, respectively. The noise-to-signal ratio (computed as the ratio between the standard deviations of the noise and the output) was about 5%. The algorithm was run with  $\delta$  equal to 0.2 (approximately,  $1.26\sigma$ ), and provided  $s = 4$  submodels. The final classification of the regression vectors  $x_k = [y_{k-1} \ u_{k-1}]'$  used for estimation, and the partition of the PWA map are shown in Figure 2. The model was validated by computing the prediction error  $\varepsilon_k$ , *i.e.* the difference between the measured and the predicted output, whose plot is depicted in Figure 3. It is noticeable that only 3 times out of 500 it falls outside the interval  $[-3\sigma, 3\sigma]$ . The Mean Square Error

$$MSE = \frac{1}{N_v} \sum_{k=1}^{N_v} \varepsilon_k^2$$

was equal to 0.030, which is very close to the variance of the noise. Recall that the prediction error is influenced by both the noise and the model error. The overall computation of the PWA model took about 9 seconds on a 1GHz Pentium III running Matlab 6.1.

## 7. CONCLUSIONS

This paper has addressed the problem of identifying a PWA model of a discrete-time nonlinear system from input-output data. The proposed two-stage procedure first classifies the data into clusters and estimates the parameters of the affine submodels, and then estimates the coefficients of the hyperplanes defining the partition of the PWA map. The key approach is the selection of a bound on the prediction error. This makes it possible to formulate the initial classification problem as an extension of the MIN PFS problem for infeasible systems of linear inequalities. The major capability of this formulation is that it also provides an estimate of the minimum number of submodels needed to fit the data. Other approaches could be used to initialize the identification procedure, *e.g.*, the  $k$ -plane clustering algorithm proposed in (Bradley and Mangasarian, 2000). A procedure for improving both data

classification and parameter estimation was also proposed. It alternates between datapoint reassignment and parameter update. Moreover, the number of submodels is allowed to vary from iteration to iteration. The partition of the PWA map is finally estimated via multicategory classification techniques. Future studies will concern the convergence properties of the refinement procedure and the evaluation of the quality of the identified PWA models.

## REFERENCES

- Amaldi, E. and M. Mattavelli (2002). The MIN PFS problem and piecewise linear model estimation. *Discrete Applied Mathematics* **118**, 115–143.
- Bemporad, A., G. Ferrari-Trecate and M. Morari (2000). Observability and controllability of piecewise affine and hybrid systems. *IEEE Trans. Automatic Control* **45**(10), 1864–1876.
- Bemporad, A., J. Roll and L. Ljung (2001). Identification of hybrid systems via mixed-integer programming. In: *Proc. 40th IEEE Conf. on Decision and Control*. pp. 786–792.
- Bradley, P. S. and O. L. Mangasarian (2000).  $k$ -plane clustering. *Journal of Global Optimization* **16**, 23–32.
- Bredensteiner, E. J. and K. P. Bennett (1999). Multicategory classification by support vector machines. *Computational Optimization and Applications* **12**, 53–79.
- Ferrari-Trecate, G., M. Muselli, D. Liberati and M. Morari (2003). A clustering technique for the identification of piecewise affine systems. *Automatica* **39**(2), 205–217.
- Heemels, W.P.M.H., B. De Schutter and A. Bemporad (2001). Equivalence of hybrid dynamical models. *Automatica* **37**(7), 1085–1091.
- Juditsky, A., H. Hjalmarsson, A. Benveniste, B. Delyon, L. Ljung, J. Sjöberg and Q. Zhang (1995). Nonlinear black-box models in system identification: mathematical foundations. *Automatica* **31**(12), 1725–1750.
- Milanese, M. and A. Vicino (1991). Optimal estimation theory for dynamic systems with set membership uncertainty: an overview. *Automatica* **27**(6), 997–1009.
- Milanese, M., J. P. Norton, H. Piet-Lahanier and E. Walter (eds.) (1996). *Bounding Approaches to System Identification*. Plenum Press. New York.
- Münz, E. and V. Krebs (2002). Identification of hybrid systems using apriori knowledge. In: *Proc. 15th IFAC World Congress*.
- Sjöberg, J., Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P. Glorennec, H. Hjalmarsson and A. Juditsky (1995). Nonlinear black-box modeling in system identification: a unified overview. *Automatica* **31**(12), 1691–1724.
- Vapnik, V. (1998). *Statistical Learning Theory*. John Wiley.