Proceedings of the 40th IEEE
Conference on Decision and Control
Orlando, Florida USA, December 2001



# Identification of Hybrid Systems via Mixed-Integer Programming

Alberto Bemporad[1], Jacob Roll[2], Lennart Ljung[2]

**Abstract**

This paper addresses the problem of identification of hybrid dynamical systems, by focusing the attention on hinging hyperplanes (HHARX) and Wiener piecewise affine (W-PWARX) autoregressive exogenous models. In particular, we provide algorithms based on mixed-integer linear or quadratic programming which are guaranteed to converge to a global optimum.

## 1 Introduction

Hybrid systems are systems with both continuous and discrete dynamics, the former typically associated with physical principles, the latter with logic devices. Most literature of hybrid systems has dealt with modeling [1, 2], stability analysis [3, 4], control [2, 5, 6], verification [7–9], and fault detection [10, 11]. The different tools rely on a model of the hybrid system. Getting such a model from data is an identification problem, which does not seem to have received enough attention in the hybrid systems community, except for the recent contribution [12]. On the other hand, in other fields there has been extensive research on identification of general nonlinear black-box models [13]. A few of these techniques lead to piecewise affine (PWA) models of nonlinear dynamical systems [14–18], and thanks to the equivalence between PWA systems [1,19,20] and several classes of hybrid systems, they can be used to obtain hybrid models.

As will be pointed out, if the *guardlines* (i.e., the partition of the PWA mapping) are known, the problem of identifying PWA systems can easily be solved using standard techniques. However, when the guardlines are unknown the problem becomes much more difficult. The basic difficulty can be expressed as follows: There are two possibilities: (1) Either a grid defining the cells over which the system is constant is defined a priori or (2) this grid is estimated along with the linear models. The former approach gives an easy estimation process for the linear submodels, but suffers from the curse of dimensionality in the sense that the number of a priori given cells will have to be very large for reasonable flexibility even in the case of moderately many regressors. The second approach allows for efficient use of fewer cells, but leads to potentially (very) many local minima, which may make it difficult to apply local search routines. This dilemma we address in this paper. It should be said right away that we will not offer any practical solution here. Instead we point to reformulations for two subclasses of PWA systems that lead to mixed-integer linear or quadratic programming problems that can be solved for the global optimum. We illuminate this approach with some examples and analysis and also discuss how these insights can be used for new ideas on efficient algorithms.

[1]Dip. Ingegneria dell'Informazione, Università di Siena, Italy, bemporad@dii.unisi.it. ETH Zurich, Switzerland, bemporad@aut.ee.ethz.ch

[2]Division of Automatic Control, Linköping University, Sweden roll,ljung@isy.liu.se

## 2 PWARX Models

To begin with, let us consider systems on the form

$$y_t = g(\phi_t) + e_t \qquad (1)$$

where $\phi_t \in \mathbb{R}^n$ is our regression vector, $e_t$ is white noise, and $g$ is a PWA function of the form

$$g(\phi) = d_j'\phi + c_j \quad \text{if} \quad \bar{H}_j\phi \le \bar{D}_j \qquad (2)$$

where $d^j \in \mathbb{R}^n$, $c^j \in \mathbb{R}$, and the sets $\mathcal{C}_j \triangleq \{\phi : \bar{H}_j\phi \le \bar{D}_j\}$, $j = 1, \ldots, s$ are a polyhedral partition of the $\phi$-space. To allow for a more compact notation, we let $\varphi_t = \left[\begin{smallmatrix} 1 \\ \phi_t \end{smallmatrix}\right]$, $\theta_j = \left[\begin{smallmatrix} c_j \\ d_j \end{smallmatrix}\right]$, and $H_j = [-\bar{D}_j \ \bar{H}_j]$. In this way (2) can be written as

$$g(\varphi) = \varphi'\theta_j \quad \text{if} \quad H_j\varphi \le 0 \qquad (3)$$

$\varphi_t$ could, e.g., consist of old inputs and outputs, i.e., $\varphi_t = [1 \ y_{t-1} \ \ldots \ y_{t-n_a} \ u_{t-1} \ \ldots \ u_{t-n_b}]$. In this case we call the systems PWARX (PieceWise affine AutoRegressive eXogenous) systems. We do not assume that $g$ is necessarily continuous over the boundaries, commonly referred to as *guardlines*. Without this assumption, definition (2) is not well posed in general, as the function can be multiply defined over common boundaries of the sets $\mathcal{C}_j$. Although one can avoid this issue by replacing some of the "$\le$" inequalities into "$<$" in the definition of the regions $\mathcal{C}_j$, this issue is not of practical interest from a numerical point of view.

### 2.1 Identification of PWARX Models

Now suppose that we are given $y_t$ and $\varphi_t$, $t = 1, \dots, N$, and want to find the PWARX model that best matches the given data. The identification of model (3) can be carried out by solving the optimization problem

$$\min \frac{1}{2N} \sum_{t=1}^{N} \left( \sum_{j=1}^{s} \|y_t - \varphi_t' \theta_j\| \mathcal{J}_j(\varphi_t) \right) \tag{4a}$$

$$\text{subj. to } \mathcal{J}_j(\varphi_t) = \begin{cases} 1 & \text{if } H_j \varphi_t \leq 0 \\ 0 & \text{otherwise} \end{cases} \tag{4b}$$

$$+ \text{ linear bounds over } \theta_j,\ H_j \tag{4c}$$

where $\theta_j,\ H_j,\ j = 1, \dots, s$ are the unknowns. In (4), we will focus on the 1-norm ( $|\cdot|$ ) and the squared Euclidean norm ( $\|\cdot\|_2^2$ ), as they allow to express (4) as a *mixed-integer linear or quadratic program* (MILP/MIQP), respectively, for which efficient solvers exist [21–24]. The problem can be also recast as an MILP by using infinity norm over time (i.e. $\max_{t=1,\dots,N}$ instead of $\sum_{t=1}^{N}$), although this would be highly sensitive to possible outliers in the estimation data. We distinguish between two main cases:

**A. Known Guardlines** $H_j$ (i.e., the partition of the $\varphi$-space) are known, $\theta_j$ have to be estimated. If using 2-norm in (4), we can see that this is an ordinary least-squares problem which can be solved efficiently.

**B. Unknown Guardlines** Both $H_j$ and $\theta_j$ are unknown. This is a much harder problem, since it is nonconvex and the objective function generally contains several local minima. However, the optimization problem (4) can be recast as an MILP or MIQP. In the following sections, we focus on two subsets of PWA functions, namely the Hinging Hyperplanes (HH) and Wiener processes with PWA static output mapping, and detail the mixed-integer program associated to the identification problem. In general, the complexity of the mixed-integer program needed to solve (4) is related to the number of samples $N$ and regions $s$, and the number of parameters $H_j$, $\theta_j$ that are unknown. Note that in general, the guardlines $H_j^i \varphi \leq 0$, (where $M^i$ denotes the $i$-th row of $M$) cannot be determined exactly from the given estimation data set, as the pairs $y_t, \varphi_t$ are a discrete set of points which can be divided by a continuum of possible guardlines.

## 3 Hinging Hyperplane Models

*Hinging hyperplane* (HH) models were introduced by Breiman [14]. They are based the hinge function $g_i(\varphi) = \pm \max\{\varphi' \theta_i^+, \varphi' \theta_i^-\}$ (see Fig. 1). The $\pm$ sign is needed to represent both convex and nonconvex functions. However, since it will only have a minor effect on the computations in this paper, we will exclude

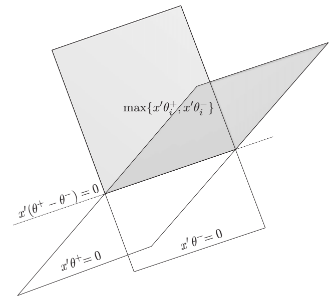



**Figure 1:** Hinging hyperplanes and hinge function

it for notational simplicity. We obtain the following HHARX (Hinging-Hyperplane AutoRegressive eXogenous) model

$$y_t = \varphi_t' \theta_0 + \sum_{i=1}^{M} \max\{\varphi_t' \theta_i, 0\} + e_t \tag{5}$$

Since $-z + \max\{z, 0\} = \max\{-z, 0\}$, $\forall z \in \mathbb{R}$, there are redundancies in (5), which can be partially avoided by introducing the requirement

$$w' \theta_i \geq 0,\ i \in [1, M] \tag{6}$$

where $w$ is any nonzero vector of $\mathbb{R}^n$, e.g., $w = \underline{1} \triangleq [1\ 1\ \dots\ 1]'$ (or any random vector).

## 4 Identification Algorithms for HH Models

The first algorithm for estimating HH models was proposed by Breiman [14]. Later, in [15] it is shown that the original algorithm is a special case of Newton's method, and provide a modification which guarantees convergence to a *local* minimum. Other algorithms have been proposed based on tree HH models [25]. In this paper, we propose an alternative approach based on mixed-integer programming, which provides a *global* minimum, at the price of an increased computational effort.

Consider the problem of estimating a HH function of the form (5) from the estimation data set $\{y_t, \varphi_t\}_{t=1}^{N}$. We choose the optimal parameters $\Theta^*$ by solving

$$\Theta^* \triangleq \arg\min V(\Theta) \triangleq \sum_{t=1}^{N} |y_t - g(\varphi_t, \Theta)| \tag{7a}$$

$$\text{subj. to} \quad \begin{cases} \theta^{j-} \leq \theta_j \leq \theta^{j+} \\ \underline{1}' \theta_i \geq 0,\ i \in [1, M] \end{cases} \tag{7b}$$

where the inequalities in (7b) are componentwise. As we will see, (7) can be reformulated as an MILP. Another possibility is to use the squared Euclidean norm $(y_t - g(\varphi_t, \Theta))^2$, as the problem can be recast as an MIQP.

### 4.1 Optimization Problem

***MILP Formulation.*** To recast (7) as an MILP, we introduce the 0-1 variables $\delta_{it}$:

$$[\delta_{it} = 0] \leftrightarrow [\varphi_t' \theta_i \leq 0], \ i \in [1, M], \ t \in [1, N] \tag{8}$$

and the new continuous variables $z_{it}$

$$z_{it} = \max\{\varphi_t' \theta_i, 0\} = \varphi_t' \theta_i \delta_{it} \tag{9}$$

The relations (8) and (9) can be transformed into mixed-integer linear inequalities, by using standard techniques [6]. By assuming that the bounds over $\theta_i$ are all finite, Eq. (8) is equivalent to the inequalities

$$\begin{aligned} &\varphi_t' \theta_i \leq M_{it}^\theta \delta_{it} \\ &\varphi_t' \theta_i \geq \varepsilon + (m_{it}^\theta - \varepsilon)(1 - \delta_{it}) \end{aligned} \tag{10}$$

where $\varepsilon$ is a small positive scalar (e.g., the machine precision), and $M_{it}^\theta$ and $m_{it}^\theta$ are upper and lower bounds on $\varphi_t' \theta_i$, derived from the bounds on $\theta_i$. Similarly, (9) is equivalently rewritten as

$$\begin{aligned} -M_{it}^\theta \delta_{it} + z_{it} &\leq 0 \\ m_{it}^\theta \delta_{it} - z_{it} &\leq 0 \\ -M_{it}^\theta (1 - \delta_{it}) - z_{it} &\leq -\varphi_t' \theta_i \\ m_{it}^\theta (1 - \delta_{it}) + z_{it} &\leq \varphi_t' \theta_i \end{aligned} \tag{11}$$

Finally, by introducing auxiliary slack variables $\epsilon_t \geq |y_t - g(\varphi_t, \Theta)|$, $t = 1, \ldots, N$, the following holds:

**Proposition 1** *The optimum of problem* (7) *is equivalent to the optimum of the following MILP*

$$\begin{aligned} &\min_{\epsilon_t, \theta_i, z_{it}, \delta_{it}} \sum_{t=1}^N \epsilon_t \\ &\text{subj. to } \epsilon_t \geq y_t - \varphi_t' \theta_0 - \sum_{i=1}^N z_{it} \\ &\qquad\qquad \epsilon_t \geq \varphi_t' \theta_0 + \sum_{i=1}^N z_{it} - y_t \\ &\qquad\qquad (10), (11), (6) \end{aligned} \tag{12}$$

**Example 1**. Consider the following HHARX model

$$\begin{aligned} y_t = &\ 0.8 y_{t-1} + 0.4 u_{t-1} - 0.1 + \\ &+ \max\{-0.3 y_{t-1} + 0.6 u_{t-1} + 0.3, 0\} \end{aligned} \tag{13}$$

The model is identified on the data reported in Fig. 3(a), by solving an MILP with 66 variables (of which 20 integers) and 168 constraints. The problem is solved by using Cplex 6.5 [24] (1014 LP solved in 0.68 s on a Sun Ultra 10 running Matlab 5.3), and, for comparison, using BARON [22] (73 LP solved in 3.00 s, same machine), which results in a zero output prediction error

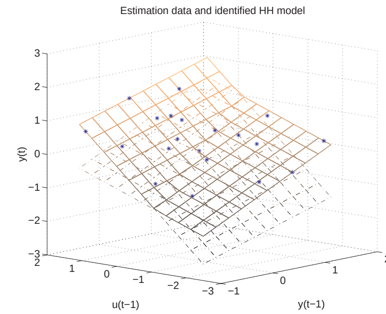



**Figure 2:** Identification of model (13) – noiseless case. Identified HH model.

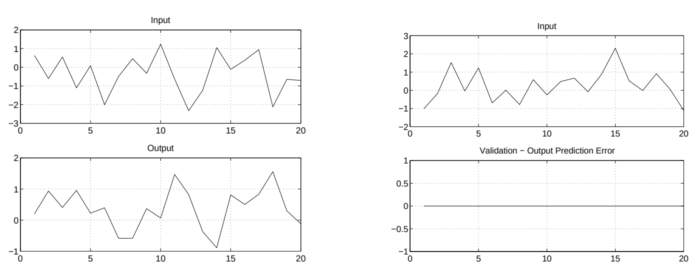
(Fig. 3(b)). The fitted HH model is reported in Fig. 2. After adding noise $e_t \in N(0, 0.1)$ to the output $y_t$, the following model

$$\begin{aligned} y_t = &\ 0.83 y_{t-1} + 0.34 u_{t-1} - 0.20 + \\ &+ \max\{-0.34 y_{t-1} + 0.62 u_{t-1} + 0.40, 0\} \end{aligned} \tag{14}$$

is identified in 1.39 s (3873 LP solved) using Cplex (7.86 s, 284 LP using BARON) on the estimation set reported in Fig. 4(a), and produces the validation data reported in Fig. 4(b). For comparison, we identified the linear ARX model

$$y_t = 0.82 y_{t-1} + 0.72 u_{t-1} \tag{15}$$

on the same estimation data, obtaining the validation data reported in Fig. 5 (higher order ARX models did not produce significant improvements). Clearly, the error generated by driving the ARX model in open-loop with the validation input $u_t$ is much larger, and would not make (15) suitable for instance for formal verification tools, where a good performance of open-loop prediction is a critical requirement.

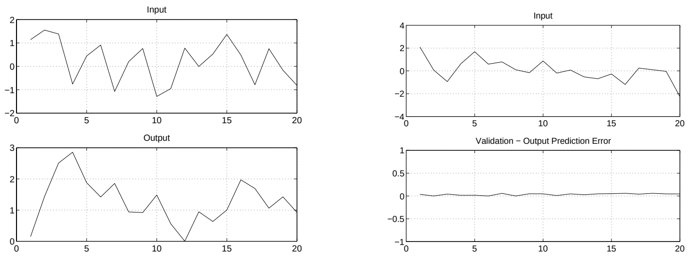



(a) Estimation data (b) Validation data

**Figure 3:** Identification of model (13) – noiseless case

***MIQP Formulation.*** When the squared 2-norm is used in the objective function, the optimization problem can be recast as the MIQP

$$\begin{aligned} &\min_{\theta_i, \delta_{it}, z_{it}} V(\Theta) \triangleq \sum_{t=1}^N (y_t - (\varphi_t' \theta_0 + \sum_{i=1}^M z_{it}))^2 \\ &\text{subj. to } (10), (11), (6) \end{aligned} \tag{16}$$




(a) Estimation data (b) Validation data

**Figure 4:** Identification of model (13) – noisy case

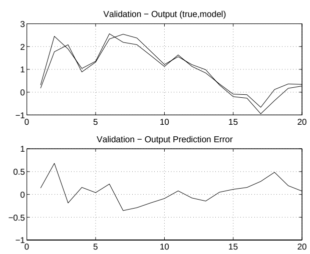



**Figure 5:** Identification of a linear ARX model – same estimation and validation data as in Fig. 4

Note that the problem is not strictly positive definite, for instance the cost function does not depend on $\theta_i$, $\delta_{it}$ (which only appear in the constraints). For numerical reasons, a term $\sigma I$, where $\sigma$ is a small number, may be added to the Hessian associated to the MIQP (16).

### 4.2 Complexity

Despite the good solvers available [21,22,24], the complexity of the MILP or MIQP problems is well known to be $\mathcal{NP}$-hard, and in particular it is exponential in the number $MN$ of binary variables. Therefore, the approach is computationally affordable only for model with a few data, or if data are clustered together (e.g., 100 data are averaged into 10 data).

### 4.3 Discontinuous HHARX Models

In HHARX models, the output $y_t$ is a continuous function of the regressor $\phi_t$. On the other hand, hybrid systems often consist of PWA discontinuous mappings. In order to tackle discontinuities, we can modify the HH model (5) in the form

$$g(\varphi_t, \Theta) = \varphi_t'\theta_0 + \sum_{i=1}^{M}(\varphi_t'\theta_i + a_i)\delta_{it}(\varphi_t) \quad (17a)$$

$$[\delta_{it}(\varphi_t) = 0] \leftrightarrow [\varphi_t'\theta_i \leq 0],\ i \in [1, M],\ t \in [1, N] \quad (17b)$$

where $a_i$, $i = 1, \ldots, M$ are additional free parameters, $a_i^- \leq a_i \leq a_i^+$; or, more in general, in the form

$$g(\varphi_t, \Theta) = \varphi_t'\theta_0 + \sum_{i=1}^{M}(\varphi_t'\theta_i)\delta_{it}(\varphi_t) \quad (18a)$$

$$[\delta_{it}(\varphi_t) = 0] \leftrightarrow [\varphi_t'\mu_i \leq 0],\ i \in [1, M], t \in [1, N] \quad (18b)$$

where $\mu_i$, $i = 1, \ldots, M$ are additional free vectors of parameters, $\mu_i^- \leq \mu_i \leq \mu_i^+$, $\underline{1}'\mu_i \geq 0$. Similarly to (12), both the identification problems (17) and (18) can be again recast as an MILP. With respect to (12), the MILP has $\mu_i$ or $a_i$ as additional optimization variables. Note that the problem in general does not have a unique solution, just as for general PWARX systems.

### 4.4 Robust HHARX Models

In formal verification methods, model uncertainty needs to be handled in order to provide safety guarantees. Typically, the model is associated with a bounded uncertainty. In the present context of HHARX models, we wish to find an uncertainty description of the form

$$g(\varphi_t, \Theta^-) \leq y_t \leq g(\varphi_t, \Theta^+),\ \forall t \geq 0 \quad (19)$$

for an inclusion-type of description, or the form

$$y_t = g(\varphi_t, \Theta^*) + n_t,\ n^- \leq n_t \leq n^+ \quad (20)$$

for an additive-disturbance-type of description. Clearly, since the model is identified from a finite estimation data set, fulfillment of (19) or (20) for all $t$ and for all initial conditions cannot be guaranteed, unless additional hypotheses on the model which generates the data are assumed. Nevertheless, a pair of extreme models $\Theta^-$, $\Theta^+$ can be obtained by solving (12) or (16) with the additional linear constraints

$$y_t \geq g(\varphi_t, \Theta),\ \forall t \in [1, N] \quad (21)$$

for estimating $\Theta^-$, and

$$y_t \leq g(\varphi_t, \Theta),\ \forall t \in [1, N] \quad (22)$$

for estimating $\Theta^+$. An additive-disturbance description can be instead computed in two alternative ways:

1. First, identify a model $\Theta^*$ by solving (12) or (16) and then compute

$$\begin{aligned} n^+ &\triangleq \max_{t=1,\ldots,N} y_t - g(\varphi_t, \Theta^*) \\ n^- &\triangleq \min_{t=1,\ldots,N} y_t - g(\varphi_t, \Theta^*) \end{aligned} \quad (23)$$

2. Modify the MILP (12) by setting replacing $\epsilon_t$ with one variable $\epsilon$ only, and minimize $\epsilon$. The corresponding optimum $\epsilon^*$ provides a nominal model such that the bound on the norm of the additive disturbance $n_t$ is minimized.

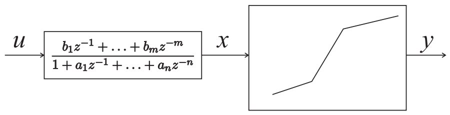



**Figure 6:** Wiener process with PWA static output mapping

## 5 Piecewise Affine Wiener Models

Let us now turn to the class of models shown in Fig. 6, described by the relations

$$\begin{aligned} A(z)x_t &= B(z)u_t \\ y_t &= f(x_t) \end{aligned} \tag{24a}$$

where $A(z) = 1 + \sum_{l=1}^{n_a} a_l z^{-l}$, $B(z) = \sum_{l=1}^{n_b} b_l z^{-l}$, and $z^{-1}$ is the delay operator, $z^{-1}x_t = x_{t-1}$. We assume that $f(x)$ is a piecewise affine, invertible function (without restrictions we can assume that $f$ is strictly increasing), and parameterize its inverse as

$$x_t = y_t - \alpha_0 + \sum_{i=1}^{M} \pm \max\{\beta_i y_t - \alpha_i, 0\} \tag{24b}$$

Both signs $\pm$ are allowed in order to be able to represent nonconvex functions. We assume that the number $M^+$ of positive signs is known (without restrictions we can let these be the first terms of the sum). As $\max\{-z, 0\} = -z + \max\{z, 0\}$ for all $z \in \mathbb{R}$, without loss of generality we can also assume $\beta_i \geq 0$.

## 6 Identification of W-PWARX Models

As seen from Fig. 6, a Wiener model consists of a linear dynamic system followed by an output nonlinearity. In some cases, the two can be identified separately: first the inverse nonlinearity is estimated by supplying a quasi-static input, and then a linear dynamic model is identified by using standard linear techniques [26]. On the other hand, in some other cases the input signal cannot be designed arbitrarily, as input/output estimation data are simply supplied by other sources. Then one algorithm which estimates the whole Wiener process is desirable. Here, we describe an algorithm based on mixed-integer programming, which identifies W-PWARX models of the form (24). Such PWA form is particularly useful when the identified system models an unknown part of a larger hybrid model. We assume that we are given an estimation data set $\{y_t, u_t\}_{t=1}^N$.

Define $a_h = a_h^+ - a_h^-$, $a_h^+, a_h^- \geq \gamma$, where $\gamma > 0$ is any positive scalar. Then

$$\begin{aligned} & a_h \max\{\beta_i y_{t-h} - \alpha_i, 0\} = \\ & = \max\{a_h^+ \beta_i y_{t-h} - a_h^+ \alpha_i, 0\} - \\ & \quad - \max\{a_h^- \beta_i y_{t-h} - a_h^- \alpha_i, 0\} \\ & = \max\{c_{ih}^+ y_{t-h} - d_{ih}^+, 0\} - \max\{c_{ih}^- y_{t-h} - d_{ih}^-, 0\} \end{aligned}$$

where

$$\begin{aligned} c_{ih}^{\pm} &\triangleq a_h^{\pm}\beta_i, \\ d_{ih}^{\pm} &\triangleq a_h^{\pm}\alpha_i, \ i \in [1, M], \ h \in [1, n_a] \end{aligned}$$

Let also

$$\begin{aligned} c_{i0} &= c_{i0}^+ = c_{i0}^- \triangleq \beta_i \\ d_{i0} &= d_{i0}^+ = d_{i0}^- \triangleq \alpha_i \\ d_{0h} &\triangleq a_h \alpha_0 \\ d_{00} &\triangleq \alpha_0 \\ \bar{d}_0 &\triangleq \sum_{h=0}^{n_a} d_{0h} = \left(1 + \sum_{h=1}^{n_a} a_h\right)\alpha_0 \end{aligned}$$

For each $-\max$ and $+\max$ function in (24b), and for each $t$, we introduce the integer variables $\delta_{it} \in \{0, 1\}$

$$[\delta_{it} = 1] \leftrightarrow [\beta_i y_t - \alpha_i \geq 0], \ i \in [1, M], \ t \in [1, N] \tag{25}$$

Without loss of generality, we can assume that the $M^+$ first breakpoints in the PWA output nonlinearity are ordered, and similarly for the $M - M^+$ last breakpoints. Clearly, the logic constraint

$$[\delta_{it} = 1] \quad \rightarrow \quad [\delta_{jt} = 1] \tag{26}$$

should hold for all $i, j \leq M^+$ such that $j < i$, and for all $i, j > M^+$ such that $j < i$. Each constraint (26) is translated into

$$\delta_{it} - \delta_{jt} \leq 0, \tag{27}$$

and a minimal set of inequalities is obtained by collecting (27) only for pairs of consecutive indices $i, j$. Moreover, since the output data $y_t$ can be ordered, we can also get additional relations on $\delta_{it}$ by using (25). In fact, if $\delta_{it_0} = 1$ and $y_{t_1} > y_{t_0}$, it must follow that $\delta_{it_1} = 1$. We can translate these relations into

$$\delta_{it_0} - \delta_{it_1} \leq 0, \ \forall t_1 \neq t_0 : y_{t_1} \geq y_{t_0} \tag{28}$$

As $a_h^+, a_h^- > 0$, from (25) it also follows

$$[\delta_{it} = 1] \leftrightarrow \left[c_{ih}^{\pm} y_t - d_{ih}^{\pm} \geq 0\right] \tag{29}$$

Let us also introduce the auxiliary continuous variables

$$\begin{aligned} z_{it0} &\triangleq (c_{i0}y_t - d_{i0})\delta_{it} \\ z_{ith} &\triangleq [(c_{ih}^+ - c_{ih}^-)y_{t-h} - (d_{ih}^+ - d_{ih}^-)]\delta_{i(t-h)}, \ h \in [1, n_a] \end{aligned} \tag{30}$$

Using the same techniques as in (10) and (11), we can translate (29) and (30) to linear inequalities.

Now,

$$\begin{aligned} x_t &= y_t - d_{00} + \sum_{i=1}^{M} \pm z_{it0} \\ a_h x_{t-h} &= a_h y_{t-h} - d_{0h} + \sum_{i=1}^{M} \pm z_{ith} \end{aligned} \tag{31}$$

By (24) and (31),

$$x_t = y_t - d_{00} + \sum_{i=1}^{M} \pm z_{it0} = \sum_{k=1}^{n_b} b_k u_{t-k} - \\ - \sum_{h=1}^{n_a} \left( a_h y_{t-h} - d_{0h} + \sum_{i=1}^{M} \pm z_{ith} \right)$$

which provides the relation

$$y_t = - \sum_{h=1}^{n_a} a_h y_{t-h} + \sum_{k=1}^{n_b} b_k u_{t-k} + \bar{d}_0 - \sum_{i=1}^{M} \sum_{h=0}^{n_a} \pm z_{ith} \tag{32}$$

In order to fit the estimation data to model (32), we solve the mixed-integer quadratic program (MIQP)

$$\min \frac{1}{N} \sum_{t=1+\max\{n_a,n_b\}}^{N} \left| y_t + \sum_{h=1}^{n_a} a_h y_{t-h} - \right. \\ \left. - \sum_{k=1}^{n_b} b_k u_{t-k} - \bar{d}_0 + \sum_{i=1}^{M} \sum_{h=0}^{n_a} \pm z_{ith} \right|^2 \tag{33}$$

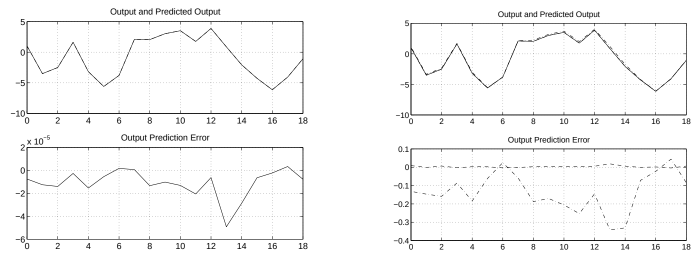
subj. to linear constr.from (27), (28), (29), and (30)

with respect to the variables $a_h$, $b_k$, $c_{i0}$, $d_{i0}$, $\bar{d}_0$, $c_{ih}^{\pm}$, $d_{ih}^{\pm}$, $z_{ith}$, and the binary variables $\delta_{it}$. The solution to (33) provides the optimal parameters $a_h^*$, $b_h^*$, and $\alpha_0^* \triangleq \frac{\bar{d}_0^*}{1+\sum_{h=1}^{n_a} a_h^*}$, $\alpha_i^* \triangleq d_{i0}^*$, $\beta_i^* \triangleq c_{i0}^*$. Finally, we can obtain the estimation $f^*(x)$ by inverting (24b) (see [27] for details).

**Example 2**. A Wiener model constituted by a first-order linear system and a nonlinearity with two breakpoints is identified, using $N = 20$ estimation data points. The system is first identified using noiseless data, and then using noisy measurements $\tilde{y}_t = y_t + e_t$, where $e_t$ are independent and uniformly distributed on a symmetric interval around 0. The MIQP problem (33) is solved by running BARON [22] on a Sun Ultra 10. The resulting estimates are shown in Table 1. The estimated parameters are overall very close to the true values, the closer the lower the intensity of the output noise, as it should be expected. The estimated model was also tested on a set of validation data, and we report in Fig. 7 the resulting one-step-ahead predicted

| Par. | True | $e_t = 0$ | $\|e_t\| < 0.01$ | $\|e_t\| < 0.1$ |
|---|---|---|---|---|
| $a_1$ | -0.5 | -0.5000 | -0.4990 | -0.5360 |
| $b_1$ | 2 | 2.0000 | 2.0024 | 2.0003 |
| $\alpha_0$ | -2 | -2.0000 | -2.0001 | -1.7748 |
| $\alpha_1$ | 0.5 | 0.5000 | 0.5095 | 0.5509 |
| $\alpha_2$ | -1.5 | -1.5000 | -1.4924 | -1.4999 |
| $\beta_1$ | 0.5 | 0.5000 | 0.5016 | 0.5028 |
| $\beta_2$ | 0.5 | 0.5000 | 0.4988 | 0.4876 |
| CPU | - | 45.44 s | 51.33 s | 90.34 s |

**Table 1:** Estimation results




(a) System estimated with noiseless data.

(b) System estimated with output noise $|e_t| \leq 0.01$ (dashed), and $|e_t| \leq 0.1$ (dot-dashed)

**Figure 7:** Validation results

output and output error. Note that such a good performance cannot be achieved by using standard linear identification techniques.

### 6.1 Complexity Analysis

By imposing the constraints expressed by (27) and (28), the degrees of freedom for the integer variables, and hence the complexity, are reduced considerably. In fact, instead of having to test $2^{MN}$ different cases in the *worst case*, only $\binom{M+N}{M} \cdot \binom{M}{M^+}$ combinations would be tested. For example, for $N = 20$ and $M = 2$ this means that the number of possible combinations of integer variables decreases from approximately $10^{12}$ to 462. In general, for a fixed $M$ the worst-case complexity grows as $N^M$. Note that this simplification is possible since the nonlinearity is one-dimensional, which allows an ordering of the breakpoints and of the output data.

## 7 State-Space Realizations

Similarly to the linear ARX case, state-space realizations of HHARX and W-PWARX models can be obtained using different hybrid state-space discrete-time paradigms introduced recently in the hybrid systems literature [1]. In particular, realization into Mixed-Logical Dynamical (MLD), Piecewise Affine (PWA), Min-Max-Plus Scaling (MMPS), Linear Complementarity (LC), and Extended Linear Complementarity

(ELC) forms are discussed in [28, 29].

## 8 Conclusions

In this paper we have addressed the problem of identification of hybrid dynamical systems, by focusing our attention on piecewise affine (PWARX), hinging hyperplanes (HHARX), and Wiener piecewise affine (W-PWARX) autoregressive exogenous models. In particular, for the two latter classes we have provided globally convergent algorithms based on mixed-integer linear or quadratic programming. Several problems remain open, such as the choice of persistently exciting input signals $u$ for identification (i.e., that allow for the identification of all the affine dynamics), and criteria like Akaike's criterion for choosing the best order and number of hinging pairs in HHARX models. Future research will also be devoted to faster suboptimal algorithms, by combining MIP solvers and clustering techniques.

### References


[1] W.P.M.H. Heemels, B. De Schutter, and A. Bemporad. Equivalence of hybrid dynamical models. *Automatica*, 37(7):1085–1091, July 2001.

[2] M.S. Branicky, V.S. Borkar, and S.K. Mitter. A unified framework for hybrid control: model and optimal control theory. *IEEE Trans. Automatic Control*, 43(1):31–45, 1998.

[3] M. Johansson and A. Rantzer. Computation of piece-wise quadratic Lyapunov functions for hybrid systems. *IEEE Trans. Automatic Control*, 43(4):555–559, 1998.

[4] M.S. Branicky. Multiple Lyapunov functions and other analysis tools for switched and hybrid systems. *IEEE Trans. Automatic Control*, 43(4):475–482, April 1998.

[5] J. Lygeros, C. Tomlin, and S. Sastry. Controllers for reachability specifications for hybrid systems. *Automatica*, 35(3):349–370, 1999.

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

[7] A. Chutinan. *Hybrid System Verification Using Discrete Model Approximations*. PhD thesis, Dept. of Electrical and Computer Engineering, Carnegie Mellon Univ. May 1999.

[8] S. Kowalewski, O. Stursberg, M. Fritz, H. Graf, I. Hoffmann, J. Preußig, M. Remelhe, S. Simon, and H. Treseler. A case study in tool-aided analysis of discretely controlled continuos systems: the two tanks problem. In *Hybrid Systems V*, volume 1567 of *Lecture Notes in Computer Science*, pages 163–185. Springer-Verlag, 1999.

[9] A. Bemporad, F.D. Torrisi, and M. Morari. Optimization-based verification and stability characterization of piecewise affine and hybrid systems. In *Hybrid Systems: Computation and Control*, volume 1790 of *Lecture Notes in Computer Science*, pages 45–58. Springer Verlag, 2000.

[10] J. Lunze. Diagnosis of quantized systems based on a timed discrete-event model. *IEEE Trans. on Systems, Man & Cybernetics, Part A*, 30(3):322–335, May 2000.

[11] A. Bemporad, D. Mignone, and M. Morari. Moving horizon estimation for hybrid systems and fault detection. In *Proc. American Contr. Conf.*, pages 2471–2475, Chicago, IL, 1999.

[12] G. Ferrari-Trecate, M. Muselli, D. Liberati, and M. Morari. A clustering technique for the identification of piecewise affine systems. In A. Sangiovanni-Vincentelli and M.D. Di Benedetto, editors, *Hybrid Systems: Computation and Control*, Lecture Notes in Computer Science. Springer Verlag, 2001.

[13] J. Sjöberg, Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P.Y. Glorennec, H. Hjalmarsson, and A. Juditsky. Nonlinear black-box modeling in system identification: a unified overview. *Automatica*, 31(12):1691–1724, 1995.

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

[15] P. Pucar and J. Sjöberg. On the hinge-finding algorithm for hinging hyperplanes. *IEEE Trans. Inform. Theory*, 44(3):1310–1319, May 1998.

[16] C. Kahlert and L.O. Chua. The complete canonical piecewise-linear representation — part i: The geometry of the domain space. *IEEE Trans. Circuits Syst. I*, 39:222–236, 1992.

[17] P. Julián, A. Desages, and O. Agamennoni. High level canonical piecewise linear representation using a simplicial partition. *IEEE Trans. Circuits Syst. I*, 46:463–480, 1999.

[18] P. Julian, M. Jordan, and A. Desages. Canonical piecewise-linear approximation of smooth functions. *IEEE Trans. Circuits and Systems — I: Fundamental Theory and Applications*, 45(5):567–571, May 1998.

[19] E.D. Sontag. Interconnected automata and linear systems: A theoretical framework in discrete-time. In R. Alur, T.A. Henzinger, and E.D. Sontag, editors, *Hybrid Systems III - Verification and Control*, number 1066 in Lecture Notes in Computer Science, pages 436–448. Springer-Verlag, 1996.

[20] A. Bemporad, G. Ferrari-Trecate, and M. Morari. Observability and controllability of piecewise affine and hybrid systems. *IEEE Trans. Automatic Control*, 45(10):1864–1876, 2000.

[21] Dash Associates. *XPRESS-MP User Guide*, 1999. `http://www.dashopt.com`.

[22] N. V. Sahinidis. BARON — Branch And Reduce Optimization Navigator. Technical report, Univ. of Illinois at Urbana-Champaign, Dept. of Chemical Eng., Urbana, IL, USA, 2000.

[23] A. Bemporad and D. Mignone. *MIQP.M: A Matlab function for solving mixed integer quadratic programs*. ETH Zurich, 2000. code available at `http://control.ethz.ch/~hybrid/miqp`.

[24] ILOG, Inc. *CPLEX 6.5 User Manual*. Gentilly Cedex, France, 1999.

[25] S. Ernst. Hinging hyperplane trees for approximation and identification. In *Proc. 37th IEEE Conf. on Decision and Control*, volume 2, pages 1266–1271, 1998.

[26] L. Ljung. *System Identification : Theory for the User*. Prentice Hall, $2^{nd}$ edition, 1999.

[27] J. Roll. Robust verification and identification of piecewise affine systems. Licentiate Thesis, Dept. of Electrical Engineering, Linköping Univ. SE-581 83 Linköping, Sweden, 2001.

[28] A. Bemporad. Identification of hybrid systems: Global convergence via mixed-integer programming. Technical Report AUT00-28, ETH, Zurich, September 2000.

[29] A. Bemporad, J. Roll, and L. Ljung. Identification of hybrid systems via mixed-integer programming. Technical Report AUT00-29, ETH, Zurich, October 2000.