Available online at www.sciencedirect.com
SCIENCE DIRECT®

Automatica 40 (2004) 37–50

automatica

www.elsevier.com/locate/automatica

Identification of piecewise affine systems via mixed-integer programming☆

Jacob Roll[a,*], Alberto Bemporad[b], Lennart Ljung[a]

[a] Division of Automatic Control, Linköping University, Linköping SE-581 83, Sweden
[b] Dipartimento di Ingegneria dell'Informazione, University of Siena, Via Roma 56, Siena 53100, Italy

Received 10 June 2002; received in revised form 4 February 2003; accepted 18 August 2003

Abstract

This paper addresses the problem of identification of hybrid dynamical systems, by focusing the attention on hinging hyperplanes and Wiener piecewise affine autoregressive exogenous models, in which the regressor space is partitioned into polyhedra with affine submodels for each polyhedron. In particular, we provide algorithms based on mixed-integer linear or quadratic programming which are guaranteed to converge to a global optimum. For the special case where the estimation data only seldom switches between the different submodels, we also suggest a way of trading off between optimality and complexity by using a change detection approach.

Keywords: System identification; Piecewise affine systems; Mixed-integer programming; Global optimization; Change detection; Wiener models

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 of the literature on hybrid systems has dealt with modeling (Branicky, Borkar, & Mitter, 1998; Heemels, De Schutter, & Bemporad, 2001), stability analysis (Branicky, 1998; Johansson & Rantzer, 1998), control (Bemporad & Morari, 1999; Branicky et al., 1998; Lygeros, Tomlin, & Sastry, 1999), verification (Bemporad, Torrisi, & Morari, 2000c; Chutinan & Krogh, 2003), and fault detection (Bemporad, Mignone, & Morari, 1999; Lunze, 2000). 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 contributions (Bemporad, Garulli, Paoletti, & Vicino, 2003; Ferrari-Trecate, Muselli, Liberati, & Morari, 2003). On the other hand, in other fields there has been extensive research on identification of general nonlinear black-box models (Sjöberg et al., 1995). A few of these techniques lead to piecewise affine (PWA) models of nonlinear dynamical systems (Batruni, 1991; Breiman, 1993; Choi & Choi, 1994; Ernst, 1998; Gad, Atiya, Shaheen, & El-Dessouki, 2000; Heredia & Arce, 1996; Hush & Horne, 1998; Julián, Desages, & Agamennoni, 1999; Julián, Jordán, & Desages, 1998; Kahlert & Chua, 1992; Medeiros, Resende, & Veiga, 1999; Murray-Smith & Johansen, 1997; Pucar and Sjöberg, 1998; Skeppstedt, Ljung, & Millnert, 1992; Strömberg, Gustafsson, & Ljung, 1991). Owing to the equivalence between PWA systems (Bemporad, Ferrari-Trecate, & Morari, 2000a; Heemels et al., 2001; Sontag, 1996) 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 hyperplanes defining the partition of the PWA mapping, see Eq. (2)) are known, the problem of identifying PWA systems can easily be solved using standard techniques for linear systems. However, when the guardlines are unknown the problem becomes much more difficult. There are two alternatives to tackle such a problem:

(1) Define a priori a grid of cells within which the system dynamics is linear;
(2) Estimate the grid along with the linear models.

☆ This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor Brett Ninness under the direction of Editor Torsten Söderström.

* Corresponding author. Tel.: +46-13-281338; fax: +46-13-282622.
E-mail addresses: roll@isy.liu.se (J. Roll), bemporad@unisi.it (A. Bemporad), ljung@isy.liu.se (L. Ljung).

0005-1098/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2003.08.006

The former approach is used, e.g., in Julián et al. (1999), and gives a simple 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. Depending on how the partition is determined, one can distinguish between four different types of approaches:

- All parameters, both the parameters determining the partition (guardlines) and the parameters of the submodels, are identified simultaneously (Batruni, 1991; Gad et al., 2000; Julián et al., 1998; Pucar & Sjöberg, 1998). This category includes, e.g. neural networks with PWA activation functions.
- All parameters are identified simultaneously for a model class with a very simple partition, and new submodels/regions are added when needed (Breiman, 1993; Ernst, 1998; Heredia & Arce, 1996; Hush & Horne, 1998; Julián et al., 1998; Pucar & Sjöberg, 1998).
- The partition and submodels are identified iteratively or in several steps, each step considering either the partition or the submodels (Bemporad et al., 2003; Ferrari-Trecate et al., 2003; Medeiros et al., 1999; Murray-Smith & Johansen, 1997; Skeppstedt et al., 1992).
- The partition is determined using only information about the distribution of the regression vectors (Choi & Choi, 1994; Strömberg et al., 1991).

Most of these approaches (Batruni, 1991; Breiman, 1993; Choi & Choi, 1994; Ernst, 1998; Gad et al., 2000; Heredia & Arce, 1996; Hush & Horne, 1998; Julián et al., 1998; Pucar & Sjöberg, 1998) assume that the system dynamics is continuous, while, e.g., Bemporad et al. (2003) and Ferrari-Trecate et al. (2003) allow for discontinuities. For a more detailed description of the different approaches; see Roll (2003).

A common problem for the contributions mentioned above is that they can only guarantee suboptimal solutions. In contrast to this, in this paper we focus on the approach where both the partition and the submodels are identified simultaneously, and point to reformulations of the identification problem for two subclasses of PWA models that lead to mixed-integer linear or quadratic programming problems that can be solved for the *global* optimum. These classes are the *hinging hyperplane ARX (HHARX) models* and *piecewise affine Wiener models (W-PWARX)*. Although the worst-case complexity is high, these algorithms may be useful in cases where relatively few data are available (e.g., where it is very costly to obtain data), and where it is of importance to get a model which is as good as possible. As we will see, however, for one of the two model classes, namely Wiener models, the worst-case complexity will not be exponential, but polynomial. We will also discuss some ideas on how complexity can be drastically reduced for the case of slowly varying PWA systems.

This paper extends results previously presented in Bemporad (2000), Bemporad, Roll, and Ljung (2000b), Bemporad, Roll, and Ljung (2001) and Roll (2001), and is organized as follows. In Section 2 we introduce PWARX systems and set up the identification problem. In Sections 3 and 4 we present methods to obtain the global solution to the identification problem for the class of HHARX models, and more efficient suboptimal extensions in Section 5. Section 6 deals with identification methods for W-PWARX models, while Section 7 presents simple results to translate the identified models into other existing model classes for analysis/synthesis purposes.

2. PWARX models

To begin with, let us consider systems in the form

$$y_t = g(\phi_t) + e_t, \tag{1}$$

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

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

where $d_j \in \mathbb{R}^n$, $c_j \in \mathbb{R}$, $\bar{H}_j \in \mathbb{R}^{M_j\times n}$, $\bar{D}_j \in \mathbb{R}^{M_j}$, "$\leqslant$" denotes componentwise inequality, and the sets $\mathscr{C}_j \triangleq \{\phi\colon \bar{H}_j\phi \leqslant \bar{D}_j\}$, $j = 1,\ldots,s$ are a polyhedral partition of the ϕ-space. The subscripts in, e.g., $\bar{H}_j$ refer to the different parts of the partition, while superscripts, e.g., $\bar{H}_j^i$ will be used to denote the ith row of $\bar{H}_j$. To allow for a more compact notation, we let

$$\varphi_t = \begin{bmatrix} 1 \\ \phi_t \end{bmatrix}, \quad \theta_j = \begin{bmatrix} c_j \\ d_j \end{bmatrix}, \quad \text{and} \quad H_j = [-\bar{D}_j \ \bar{H}_j].$$

In this way (2) can be written as

$$g(\varphi) = \varphi'\theta_j \quad \text{if } H_j\varphi \leqslant 0. \tag{3}$$

When the regression vector φ_t consists of previous inputs and outputs,

$$\varphi_t = [1 \ y_{t-1} \ \cdots \ y_{t-n_a} \ u_{t-1} \ \cdots \ u_{t-n_b}]' \tag{4}$$

we say that system (1) is a PWARX (PieceWise affine AutoRegressive eXogenous) system. We do not assume that g is necessarily continuous over the boundaries defined by the polyhedra, 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 $\mathscr{C}_j$. Although one can avoid this issue by replacing some of the "$\leqslant$" inequalities into "$<$" in the definition of the regions $\mathscr{C}_j$, 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 φ_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

$$\min_{\theta_j,H_j} \frac{1}{2N}\sum_{t=1}^{N}\left(\sum_{j=1}^{s}\|y_t-\varphi_t'\theta_j\|\mathscr{I}_j(\varphi_t)\right) \tag{5a}$$

$$\text{subject to } \mathscr{I}_j(\varphi_t)=\begin{cases}1 & \text{if } H_j\varphi_t\leqslant 0\\ 0 & \text{otherwise}\end{cases} \tag{5b}$$

$$+ \text{ linear bounds over } \theta_j,\ H_j, \tag{5c}$$

where $\theta_j, H_j, j=1,\dots,s$, are the unknowns. In (5a), we will focus on the 1-norm ($|\cdot|$) and the squared Euclidean norm ($\|\cdot\|_2^2$), as they allow to express (5) as a *mixed-integer linear or quadratic program* (MILP/MIQP), respectively, for which efficient solvers exist (Dash Associates, 1999; ILOG, Inc., 1999; Sahinidis, 2000).[1] We distinguish between two main cases:

A. Known guardlines: H_j (i.e. the partition of the φ-space) are known, θ_j have to be estimated. If using 2-norm in (5a), we can see that this is an ordinary least-squares problem which can be solved quite efficiently.

B. Unknown guardlines: Both H_j and θ_j are unknown. This is a much harder problem, since it is in general a highly nonconvex problem with several local minima. However, if bounds on θ_j and H_j are known, the optimization problem (5) 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 (5) is related to the number of samples N and regions s, and the number of parameters H_j, θ_j that are unknown. Note that in general, the guardlines $H_j^i\varphi\leqslant 0$, cannot be determined exactly from a given finite estimation data set, as the pairs y_t, φ_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 (1993). They are defined as a sum of hinge functions $g_i(\varphi)=\pm\max\{\varphi'\theta_i^+,\varphi'\theta_i^-\}$, which each consists of two half-hyperplanes, parameterized by θ_i^+ and θ_i^-, respectively (see Fig. 1). The $\pm$ sign is needed to represent both convex and nonconvex functions. Using an alternative parameterization we obtain the following HHARX (hinging-hyperplane autoregressive exogenous) model

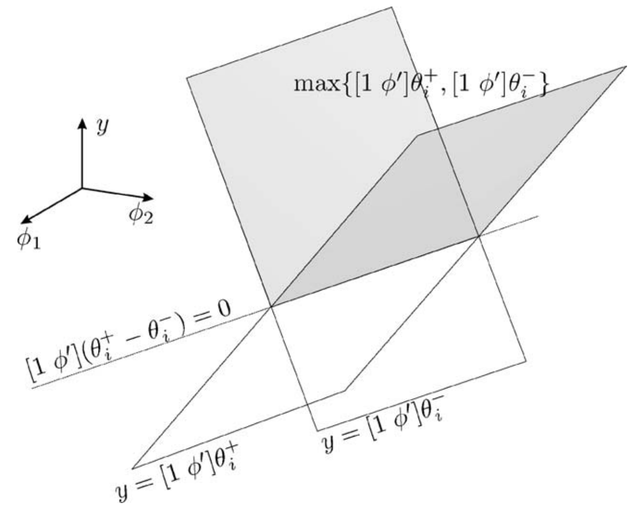

Fig. 1. Hinging hyperplanes and hinge function $y=\pm\max\{\varphi'\theta_i^+,\varphi'\theta_i^-\}$, where $\varphi=[1\ \phi']'$.

$$y_t=\varphi_t'\theta_0+\sum_{i=1}^{M}s_i\max\{\varphi_t'\theta_i,0\}+e_t, \tag{6}$$

where s_i is either $+1$ or -1, depending on the sign of the max function, and for simplicity here it is fixed a priori.[2] Since $-z+\max\{z,0\}=\max\{-z,0\}$, $\forall z\in\mathbb{R}$, there are redundancies in (6) (i.e., the structure is not globally identifiable, so the same system can be described by several different sets of parameter values), which can be partially avoided by introducing the requirement

$$w'\theta_1\geqslant\cdots\geqslant w'\theta_M\geqslant 0,\qquad i\in[1,M], \tag{7}$$

where w is any nonzero vector in $\mathbb{R}^n$, 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 (1993). Later, in Pucar and Sjöberg (1998) it was shown that the original algorithm is a special case of Newton's method, and a modification was provided which guarantees convergence to a *local* minimum. Other algorithms have been proposed based on tree HH models (Ernst, 1998). In this paper, we propose an alternative approach based on mixed-integer programming, which

[1] The problem could also be 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.

[2] The results of this paper can be easily extended to the case where s_i is also a parameter to be optimized, at the price of introducing additional binary and continuous variables in the resulting optimization problem. In the present setting, the identification algorithm presented in the following section provides the global optimum for the problem of finding the best HH model consisting of M_1 convex hinging pairs and M_2 concave hinging pairs, where $M_1+M_2=M$, and convexity/concavity depends on the chosen s_i.

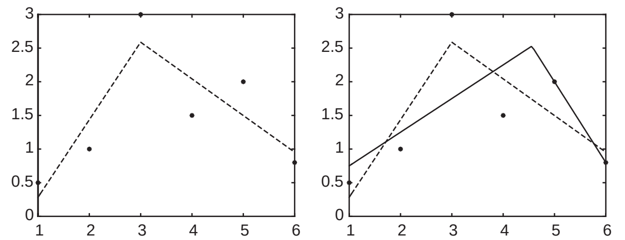

Fig. 2. Identification of a single hinge function. (a) Data samples ($*$) and globally optimal model (dashed). (b) Locally (solid) and globally (dashed) optimal model.

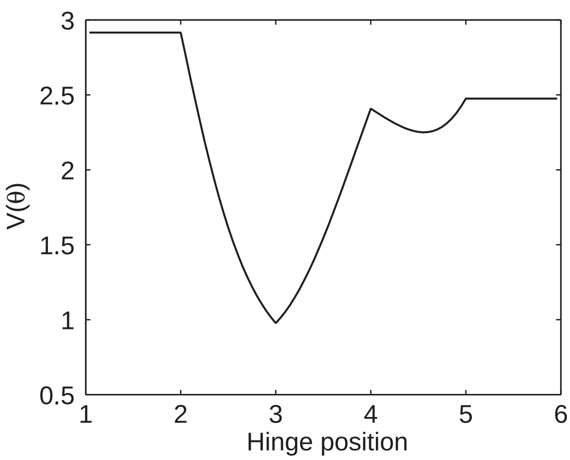

Fig. 3. Cost as a function of hinge position for Example 1.

provides a *global* minimum, at the price of an increased computational effort.

For a noiseless system consisting of one single hinge, the method proposed in Breiman (1993) was shown to converge to the global minimum. However, for noisy systems or systems with multiple hinges, local minima may lead to problems even in very simple cases, as the following example shows.

Example 1. Consider the problem of fitting a hinge function to the six data samples given in Fig. 2(a), using a 2-norm criterion. Fig. 2(a) also shows the corresponding globally optimal function, with the optimal cost 0.98. In Fig. 3 the cost is plotted as a function of the position of the hinge, and we can see that there is a local minimum between 4 and 5 with the cost 2.25. The corresponding function is plotted in Fig. 2(b). Furthermore, simple calculations show that Breiman's method will not converge to the optimal solution (regardless of the initial value), but will in most cases converge to the local minimum. The modified method provided in Pucar and Sjöberg (1998) will converge to the global optimum if starting sufficiently close to it, but will converge to the local optimum if the hinge is originally placed between 4 and 5.

Consider the problem of estimating a HH function of the form (6) from the estimation data set $\{y_t, \varphi_t\}_{t=1}^N$. Let us introduce the notation

$$\Theta = (\theta_0 \ \dots \ \theta_M);$$

$$g(\varphi_t, \Theta) = \varphi_t'\theta_0 + \sum_{i=1}^{M} s_i \max\{\varphi_t'\theta_i, 0\}.$$

We choose the optimal parameters Θ^* by solving

$$\Theta^* \triangleq \arg\min V(\Theta) \triangleq \sum_{t=1}^{N} |y_t - g(\varphi_t, \Theta)| \tag{8a}$$

subject to

$$\theta^{j-} \leqslant \theta_j \leqslant \theta^{j+},$$

$$\underline{1}'\theta_i \geqslant 0, \quad i \in [1, M], \tag{8b}$$

where the inequalities in (8b) are componentwise. As we will see, (8) can be reformulated as an MILP. Another possibility is to use the squared Euclidean norm $(y_t - g(\varphi_t, \Theta))^2$, which gives a problem that can be recast as an MIQP.

Note that in (8) we search for parameters θ_j that lie in the given range $\theta^{j-} \leqslant \theta_j \leqslant \theta^{j+}$. Given that θ_j relates linearly y_t to past inputs and outputs (cf. Eq. (6)), it is reasonable to assume that overestimates of θ_j are available from the nature of the system generating the data. In case the prescribed bounds θ^{j-}, θ^{j+} are selected excessively tight, the identification problem setup (8) would simply lead to a suboptimal solution. This situation may be easily detected if any of the above constraints is active at the optimizer, in which case the optimization may be repeated with larger bounds.

4.1. Optimization problem

MILP formulation: To recast (8) as an MILP, we introduce the binary variables δ_{it} (taking values in $\{0, 1\}$):

$$[\delta_{it} = 0] \leftrightarrow [\varphi_t'\theta_i \leqslant 0], \quad i \in [1, M], \ t \in [1, N] \tag{9}$$

and the new continuous variables z_{it}

$$z_{it} = \max\{\varphi_t'\theta_i, 0\} = \varphi_t'\theta_i\delta_{it}. \tag{10}$$

Relations (9) and (10) can be transformed into mixed-integer linear inequalities, by using a slight modification of standard techniques described in Bemporad and Morari (1999) (see also Roll, 2003). By assuming that the bounds over θ_i are all finite, Eqs. (9) and (10) are equivalent[3] to the inequalities

$$z_{it} \geqslant 0,$$

$$z_{it} \leqslant M_{it}^{\theta}\delta_{it},$$

$$\varphi_t'\theta_i \leqslant z_{it},$$

$$(1 - \delta_{it})m_{it}^{\theta} + z_{it} \leqslant \varphi_t'\theta_i, \tag{11}$$

[3] For the equivalence to hold, the last inequality of (11) should be strict; otherwise δ_{it} will not be uniquely determined when $\varphi_t'\theta_i = 0$. However, because of the continuity of the hinge functions, it does not matter in this case if δ_{it} is 0 or 1, and therefore the nonstrict inequality will be used to facilitate implementation.

where M_{it}^{θ} and m_{it}^{θ} are upper and lower bounds on $\varphi_t'\theta_i$, respectively, derived from the bounds on θ_i.

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

Proposition 2. *The optimum of problem* (8) *is equivalent to the optimum of the MILP*

$$\begin{aligned}
\min_{\varepsilon_t,\theta_i,z_{it},\delta_{it}} \quad & \sum_{t=1}^{N} \varepsilon_t \\
\text{s.t.} \quad & \varepsilon_t \geqslant y_t - \varphi_t'\theta_0 - \sum_{i=1}^{M} s_i z_{it}, \\
& \varepsilon_t \geqslant \varphi_t'\theta_0 + \sum_{i=1}^{M} s_i z_{it} - y_t, \\
& \delta_{it} \in \{0,1\},\ 0 \leqslant z_{it} \leqslant M_{it}^{\theta},\ \theta^{i-} \leqslant \theta_i \leqslant \theta^{i+}, \\
& \varepsilon_t \geqslant 0 \text{ and inequalities } (11),\ (7),
\end{aligned} \tag{12}$$

where δ_{it}, z_{it}, θ_i, ε_t are the optimization variables, $t = 1, \ldots, N$, $i = 1, \ldots, M$, and N, M, s_i, y_t, ϕ_t are given.

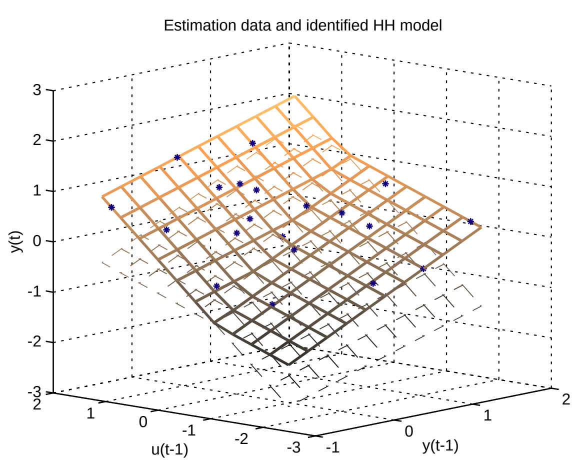

Fig. 4. Identification of model (13)—noiseless case. Identified HH model.

Example 3. 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. 5(a), by solving an MILP with 66 variables (of which 20 integers) and 168 constraints. The problem was solved by using Cplex 6.5 (ILOG, Inc., 1999) (1014 LP solved in 0.68 s on a Sun Ultra 10), and, for comparison, using BARON (Sahinidis, 2000) (73 LP solved in 3.00 s, same machine), which results in a zero output prediction error (Fig. 5(b)). The fitted HH model is shown in Fig. 4. After adding white Gaussian noise e_t with zero mean and variance 0.01 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. 6(a), and produces the validation data reported in Fig. 6(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. 7 (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 reachability analysis and formal verification tools, where a good performance of open-loop prediction is a critical requirement.

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

$$\begin{aligned}
\min_{\theta_i,\delta_{it},z_{it}} \quad & V(\Theta) \triangleq \sum_{t=1}^{N} \Big(y_t - \big(\varphi_t'\theta_0 + \sum_{i=1}^{M} s_i z_{it}\big)\Big)^2 \\
\text{s.t.} \quad & (11),\ (7).
\end{aligned} \tag{16}$$

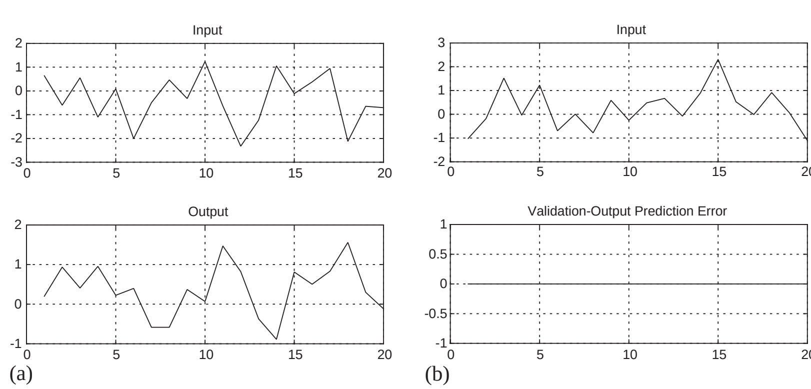

Fig. 5. Identification of model (13)—noiseless case. (a) Estimation data. (b) Validation data.

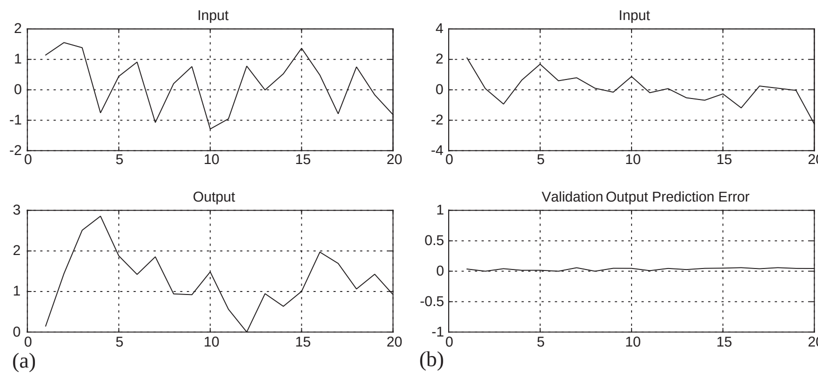

Fig. 6. Identification of model (13)—noisy case. (a) Estimation data. (b) Validation data.

Note that the problem is not strictly positive definite, for instance the cost function does not depend on θ_i, δ_{it} (which only appear in the constraints). For numerical reasons, a term σI, where $\sigma > 0$ is a small number, may be added to the Hessian associated to the MIQP (16).

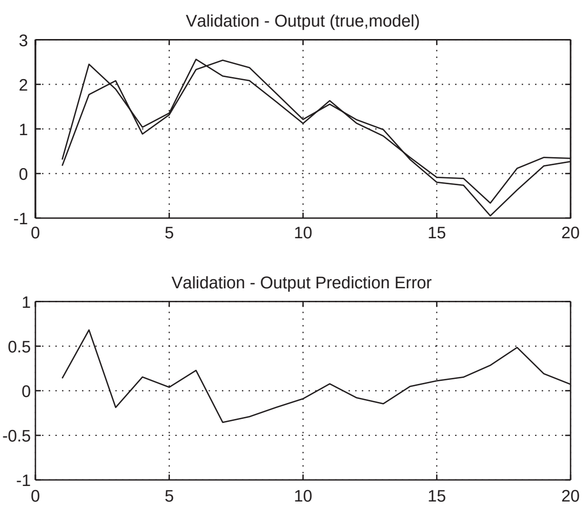

Fig. 7. Identification of a linear ARX model—same estimation and validation data as in Fig. 6.

Example 4. Consider again the PWARX system (13). In Fig. 8 we compare the performance in terms of LP/QPs and total computation time of the linear criterion (12) vs. the quadratic criterion (16). The reported numbers are computed on a Sun Ultra 60 (2×360 MHz) using the solver BARON (Sahinidis, 2000), by averaging the number of LP/QPs and computation times, respectively, for 10 estimation data sets generated by feeding random Gaussian inputs u_t and zero output noise to system (13).

4.2. Complexity

The complexity of the MILP or MIQP problems is well known to be $\mathcal{NP}$-hard, and in particular it is worst-case exponential in the number MN of binary variables, even if there are good solvers available (Dash Associates, 1999; ILOG, Inc., 1999; Sahinidis, 2000). Therefore, the approach is computationally affordable only for problems with few data, or if data are clustered together. An example of the latter approach is given in Section 5, where a piecewise affine function is identified over a sliding window.

4.3. Discontinuous HHARX models

In HHARX models, the output y_t is a continuous function of the regressor ϕ_t. On the other hand, hybrid systems may

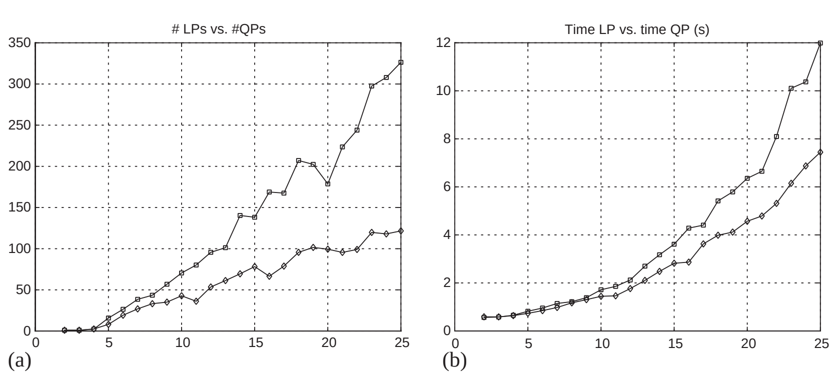

Fig. 8. Identification of model (13)—MILP (diamonds) vs. MIQP (squares). The horizontal axes show the number of estimation data samples. (a) Average number of LPs and QPs. (b) Average computation time.

consist of PWA discontinuous mappings. In order to tackle discontinuities, we can modify the HH model (6) in the form

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

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

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

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

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

where μ_i, $i = 1, \ldots, M$ are additional free vectors of parameters, $\mu_i^- \leqslant \mu_i \leqslant \mu_i^+$, $\underline{1}'\mu_i \geqslant 0$. Similarly to (12), both the identification problems (17) and (18) can be again recast as an MILP. With respect to (12), the MILP has μ_i or a_i as additional optimization variables. Note that the problem may 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 set, for instance a range of possible values a disturbance may take. In the present context of HHARX models, we wish to find an uncertainty description of the form

$$g(\varphi_t, \Theta^-) \leqslant y_t \leqslant g(\varphi_t, \Theta^+), \quad \forall t \geqslant 0 \tag{19}$$

for an inclusion-type of description, or the form

$$y_t = g(\varphi_t, \Theta^*) + n_t, \quad n^- \leqslant n_t \leqslant n^+ \tag{20}$$

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 other data than the estimation data cannot be guaranteed, unless additional hypotheses on the model which generates the data are assumed. Nevertheless, a pair of extreme models Θ^-, Θ^+ can be obtained by solving (12) or (16) with the additional linear constraints

$$y_t \geqslant g(\varphi_t, \Theta), \quad \forall t \in [1, N] \tag{21}$$

for estimating Θ^- and

$$y_t \leqslant g(\varphi_t, \Theta), \quad \forall t \in [1, N] \tag{22}$$

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

(1) First, identify a model Θ^* by solving (12) or (16) and then compute

$$n^+ \triangleq \max_{t=1,\ldots,N} y_t - g(\varphi_t, \Theta^*),$$

$$n^- \triangleq \min_{t=1,\ldots,N} y_t - g(\varphi_t, \Theta^*). \tag{23}$$

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

5. Using change detection to reduce complexity

Some PWA systems of interest may not switch so frequently between the different dynamics of the different submodels. For such systems, it is possible to use a change detection algorithm to roughly find the timepoints when switches occur, and use this information to reduce the complexity of (12) or (16) by forcing several samples, lying in the same interval between two switches, to belong to the same subsystem. Here we propose to use an MILP algorithm over a sliding window as a change detection algorithm. The formulation (12) is used, taking only data from time $t_0, \ldots, t_0 + L - 1$ into account, where L is the length of the window. Furthermore, only one switch is allowed in each window. Hence, the MILP solved takes the form

$$\begin{aligned} \min_{\varepsilon_t, \theta_i, z_{it}, \delta_t} \quad & \sum_{t=t_0}^{t_0+L-1} \varepsilon_t \\ \text{s.t.} \quad & \varepsilon_t \geqslant y_t - \varphi_t'\theta_0 - z_{1t} + z_{2t}, \\ & \varepsilon_t \geqslant \varphi_t'\theta_0 + z_{1t} - z_{2t} - y_t, \\ & \delta_{t_0} \leqslant \ldots \leqslant \delta_{t_0+L-1}, \\ & \text{inequalities (11) with } \delta_{1t} = \delta_{2t} = \delta_t. \end{aligned} \tag{24}$$

Note that we only need two hinges (one positive and one negative) and L discrete variables since only one switch is allowed, compared to the MN discrete variables needed in (12). (If the PWARX structure to be identified just contains positive hinges, we would only need one (positive) hinge in (24).) Furthermore, the inequalities $\delta_{t_0} \leqslant \cdots \leqslant \delta_{t_0+L-1}$ also help to reduce the complexity drastically.

In each position t_0 of the window, the fit of the local HHARX model (i.e. the optimal value of the cost function in (24)) is compared to the fit of a linear model over the same window. The value of the relative improvement of the cost function

$$k_{t_0} = 1 - \frac{V^*_{\text{HHARX}}}{V^*_{\text{ARX}}} \tag{25}$$

is assigned to the time point of the change, and as the window is moving, these values are summed up (for each time point). If the sum of the relative improvements for a certain time point exceeds a prescribed threshold K_0, this time point will be considered as a possible switching time.

The advantage of using (24) instead of a standard change detection algorithm, e.g. Brandt's GLR method (see, e.g. Gustafsson, 2000), is that the latter does not require linear

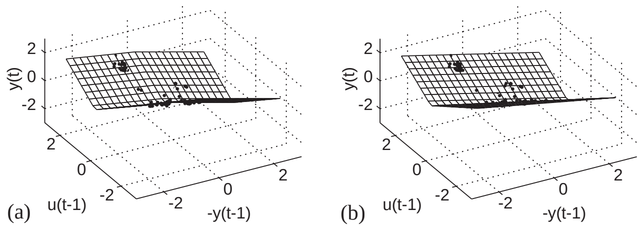

Fig. 9. Identification of (26). (a) System function (26) and estimation data. (b) Identified model.

Table 1
Identification of (26)—values of the objective function (8a)

	True system	Identified model
Estimation data	7.8213	7.4833
Validation data	7.8668	8.6777

separability between the classes; nor does it take the continuity of the PWA function into account.

After having obtained the estimated possible time points of the switches as described above, we solve (12) or (16), but using the same δ variable for all samples lying in the same time interval between two consecutive possible switches. This will force the samples to belong to the same submodel, and will reduce the complexity considerably. To summarize, the algorithm consists of two phases:

(1) Use a sliding window with a local MILP algorithm to detect possible switches and divide the time series into segments.
(2) Use an MILP to simultaneously assign the different segments to different submodels and estimate the parameters of the submodels.

Once again, note that in the first step, the MILP solved just uses two hinges, independently of how many hinge functions the final global model contains.

Example 5. The system

$$\begin{aligned} y_t &= -0.3 + 1.2y_{t-1} - u_{t-1} \\ &\quad + \max\{-1.2 + 2u_{t-1}, 0\} \\ &\quad - \max\{-0.2y_{t-1}, 0\} + e_t, \end{aligned} \tag{26}$$

where e_t is white Gaussian noise with variance 0.01, is identified using 100 data samples. The true system function and the data samples are shown in Fig. 9(a). The proposed sliding window algorithm was used with $L = 15$ and $K_0 = 1$. This resulted in the system shown in Fig. 9(b). Table 1 shows the values of the objective function (8a) for the true system and the identified model, for the estimation data and a set of validation data. As can be seen, the identified model shows a good performance. The computation time running CPLEX on a 333 MHz Pentium II laptop was 144 s (42 s for the sliding windows and 102 s for the final large MILP). This should be compared to solving the MILP (12) directly, which did not return a solution within a maximum allotted time of 3 h on the same computer.

5.1. Complexity

The advantage of the described sliding window algorithm, compared to solving (12) or (16) directly, lies in the reduction of the computational complexity. In the sliding window phase, the complexity is linear in the number of data N when using a window of fixed size, as opposed to the exponential complexity of (12) and (16).

For the second phase, the complexity is closely related to the number of possible switches. Here, the thresholding procedure makes it possible to explicitly trade off between complexity of the algorithm and optimality: The higher the threshold value, the fewer possible switch times will be considered. If it is high enough, no switches will be allowed, which means that all samples will be forced to belong to the same submodel, and we will end up with a linear model. If, on the other hand, the threshold value is chosen to be zero, every time point will be considered as a possible switch time, and we will again get the globally optimal solution.

As previously mentioned, the described algorithm requires the system to switch only seldom, staying in each submodel for a period at least in the order of the window length, L. The general issue of designing input signals having the desired properties of sufficiently exciting the modes of the system and letting the system switch seldom is a subject for future research.

5.2. Approximating general nonlinear systems

To give another example of the described sliding window algorithm, the problem of approximating a simple nonlinear system is considered. The capability of approximating arbitrary nonlinear systems is an interesting issue. Since HH functions have the universal approximation property (see, e.g., Lin & Unbehauen, 1992), they can (under mild conditions) approximate any function arbitrarily well, given a large enough number of hinges. As a very simple illustration, a quadratic NARX (nonlinear ARX) system is approximated by a HHARX model in the following example.

Example 6. Consider the system

$$y_t = -0.5y_{t-1}^2 + 0.7u_{t-1} + e_t, \tag{27}$$

where e_t is white Gaussian noise with variance 0.01, is identified using 100 data samples. The input is designed to make the output change sign only seldom (about every 25 samples). The true system function and the data samples are shown in Fig. 10(a). Using the sliding window algorithm with only one hinge in the final step, $L = 10$, and a threshold $K_0 = 1$, resulted in the system shown in

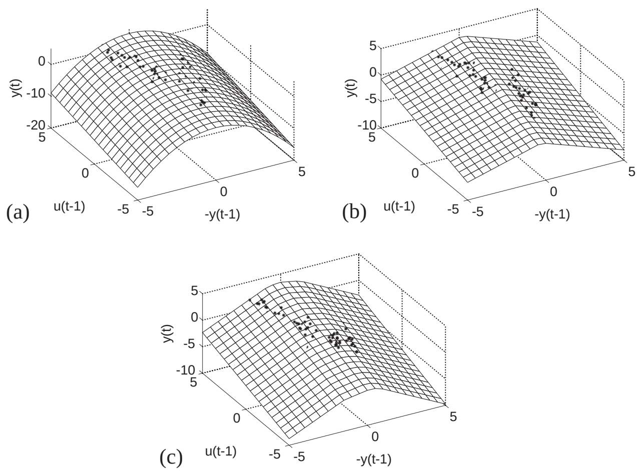

Fig. 10. Identification of (27). (a) System function (27) and estimation data. (b) Identified model using one hinge. (c) Identified model using three hinges.

Fig. 10(b). We can see that the parabola is approximated by the hinge in a natural way. The computation time running CPLEX on a 333 MHz Pentium II laptop (128 MB RAM) was 19.7 s (17.7 s for the sliding windows and 2 s for the final large MILP). Solving the MILP (12) directly required about 1300 s of computations.

If we instead use three hinges to approximate the true system function, we get the result shown in Fig. 10(c). The computation time was 152 s (20 s for the sliding windows and 132 s for the final large MILP).

6. Piecewise affine Wiener models

Let us now turn to the class of Wiener models. These models form a common class of nonlinear models that consist of a linear dynamical system followed by a static nonlinearity (see Fig. 11), that we assume here invertible. Identification of Wiener models have been discussed quite extensively in the literature (see, e.g., Hagenblad & Ljung, 2000; Kalafatis, Wang, & Cluett, 1997; Lovera, Gustafsson, & Verhaegen, 2000; Wigren, 1993).

If the input u and the number of collected data can be chosen arbitrarily, a technique may simply consist of first estimating the nonlinearity by running quasi-static experiments, inverting the nonlinearity, and then identifying the dynamic part of the model from u to x by using standard linear methods. This way of proceeding, however, besides being impractical in several situations where the identification experiment cannot be decided (data are simply provided "as they are"), has the disadvantage of weighting nonlinearly the output prediction errors.

Indeed, a common approach is to first ignore the nonlinearity to come up with a first linear approximation. From the output of this and the measured output a first estimate of the static nonlinearity can be formed. This can then be followed by further iterations. Such approaches may easily suffer from ending up in nonoptimal solutions.

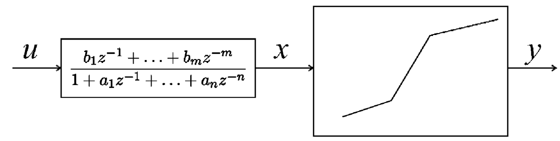

Fig. 11. Wiener process with PWA static output mapping.

In this section we shall consider Wiener models where the static nonlinearity is piecewise affine (W-PWARX models). This gives an overall piecewise affine model with a certain structure, which makes it possible to design an optimal identification algorithm whose worst-case complexity is polynomial in the number of data.

The models considered will be in the form shown in Fig. 11, described by the relations

$$\begin{aligned} &A(z)x_t = B(z)u_t, \\ &y_t = f(x_t), \end{aligned} \tag{28a}$$

where $A(z)=1+\sum_{h=1}^{n_a} a_h z^{-h}$, $B(z)=\sum_{k=1}^{n_b} b_k z^{-k}$ 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

$$x_t = y_t - \alpha_0 + \sum_{i=1}^{M} \pm \max\{\beta_i y_t - \alpha_i, 0\}. \tag{28b}$$

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 $\beta_i \geqslant 0$.

6.1. Identification of W-PWARX models

The algorithm described here is based on mixed-integer programming, which identifies W-PWARX models of the form (28). 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 $\{y_t,u_t\}_{t=1}^N$.

Like in the HHARX case, the first thing to do is to get rid of the max functions. This is done by introducing the discrete variables $\delta_{it}\in\{0,1\}$

$$[\delta_{it}=1] \leftrightarrow [\beta_i y_t - \alpha_i \geqslant 0], \quad i\in[1,M], \quad t\in[1,N]. \tag{29}$$

Before continuing with the usual reformulation into an MIQP, let us consider some additional structure that can be used to reduce the complexity of the problem. 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. Then the logic constraint

$$[\delta_{it} = 1] \rightarrow [\delta_{jt} = 1] \tag{30}$$

should hold for all $i, j \leqslant M^+$ such that $j < i$, and for all $i, j > M^+$ such that $j < i$. Each constraint (30) is translated into

$$\delta_{it} - \delta_{jt} \leqslant 0 \tag{31}$$

and a minimal set of inequalities is obtained by collecting (31) 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 δ_{it} by using (29). 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

$$\delta_{it_0} - \delta_{it_1} \leqslant 0, \quad \forall t_1 \neq t_0 : y_{t_1} \geqslant y_{t_0}. \tag{32}$$

Both (31) and (32) will help to reduce the search space considerably in the optimization.

One specific problem for this model structure is that we will get products between the coefficients a_h of the $A(z)$ polynomial and the coefficients inside the max functions, β_i and α_i. Furthermore, since a_h may be negative, the inequalities in definition (29) of $\delta_i(t)$ may change directions if we multiply by a_h. To avoid these problems, first define $a_h = a_h^+ - a_h^-$, where $a_h^+, a_h^- \geqslant \gamma$, and $\gamma \geqslant 0$ 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

$$c_{ih}^{\pm} \triangleq a_h^{\pm} \beta_i,$$

$$d_{ih}^{\pm} \triangleq a_h^{\pm} \alpha_i, \quad i \in [1, M], \ \ h \in [1, n_a].$$

Let also

$$c_{i0} = c_{i0}^+ = c_{i0}^- \triangleq \beta_i,$$

$$d_{i0} = d_{i0}^+ = d_{i0}^- \triangleq \alpha_i,$$

$$d_{0h} \triangleq a_h \alpha_0,$$

$$d_{00} \triangleq \alpha_0,$$

$$\bar{d}_0 \triangleq \sum_{h=0}^{n_a} d_{0h} = \left(1 + \sum_{h=1}^{n_a} a_h\right) \alpha_0.$$

As $a_h^+, a_h^- > 0$, from (29) it now follows

$$[\delta_{it} = 1] \leftrightarrow [c_{ih}^{\pm} y_t - d_{ih}^{\pm} \geqslant 0]. \tag{33}$$

Let us also introduce the auxiliary continuous variables

$$z_{it0} \triangleq (c_{i0} y_t - d_{i0}) \delta_{it},$$

$$z_{ith} \triangleq [(c_{ih}^+ - c_{ih}^-) y_{t-h} - (d_{ih}^+ - d_{ih}^-)] \delta_{i(t-h)},$$

$$h \in [1, n_a]. \tag{34}$$

Using the same techniques as in Bemporad and Morari (1999), we can translate (33) and (34) to linear inequalities.

Now,

$$x_t = y_t - d_{00} + \sum_{i=1}^{M} \pm z_{it0},$$

$$a_h x_{t-h} = a_h y_{t-h} - d_{0h} + \sum_{i=1}^{M} \pm z_{ith}. \tag{35}$$

By (28) and (35),

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

which provides the relation

$$y_t = -\sum_{h=1}^{n_a} a_h y_{t-h} + \sum_{k=1}^{n_b} b_k u_{t-k} + \bar{d}_0 - \sum_{i=1}^{M} \sum_{h=0}^{n_a} \pm z_{ith}. \tag{36}$$

In order to fit the estimation data to model (36), we solve the mixed-integer quadratic program (MIQP)

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

$$\text{subject to linear constr. from (31)–(34)} \tag{37}$$

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 δ_{it}. The solution to (37) provides the optimal parameters a_h^*, b_k^*, and $\alpha_0^* \triangleq \bar{d}_0^* / (1 + \sum_{h=1}^{n_a} a_h^*)$, $\alpha_i^* \triangleq d_{i0}^*$, $\beta_i^* \triangleq c_{i0}^*$. Finally, we can obtain the estimation $f^*(x)$ by inverting (28b) (see Roll, 2003, for details).

Example 7. 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 (37) is solved by running BARON (Sahinidis, 2000) on a Sun Ultra 10. The resulting estimates are shown in Table 2. The estimated parameters are overall very close to the true values, the closer the lower the intensity of the output noise, as should be expected. The

Table 2
Estimation results

Parameter	True value	$e_t = 0$	$\|e_t\| < 0.01$	$\|e_t\| < 0.1$
a_1	−0.5	−0.5000	−0.4990	−0.5360
b_1	2	2.0000	2.0024	2.0003
α_0	−2	−2.0000	−2.0001	−1.7748
α_1	0.5	0.5000	0.5095	0.5509
α_2	−1.5	−1.5000	−1.4924	−1.4999
β_1	0.5	0.5000	0.5016	0.5028
β_2	0.5	0.5000	0.4988	0.4876
CPU	—	45.44 s	51.33 s	90.34 s

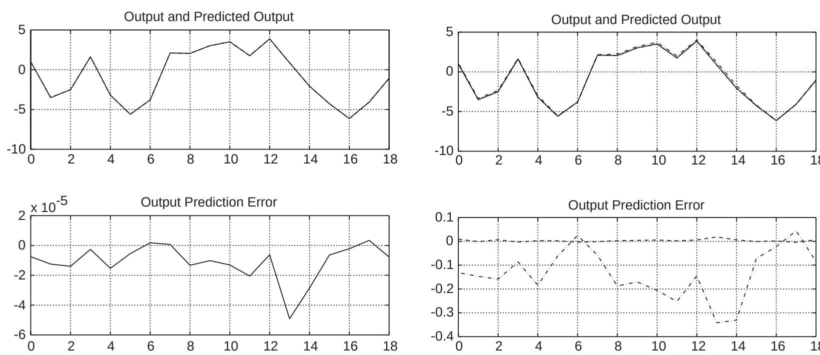

Fig. 12. Example 7—Results for a validation data set. (a) System estimated with noiseless data. (b) System estimated with output noise $|e_t| \leqslant 0.01$ (dashed), and $|e_t| \leqslant 0.1$ (dot-dashed).

estimated model was also tested on a set of validation data, and we report in Fig. 12 the resulting one-step-ahead predicted output and output error. Note that such a good performance cannot be achieved by using standard linear identification techniques.

6.2. Complexity analysis

By imposing the constraints expressed by (31) and (32), 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

In the recent literature on hybrid systems, several formalisms for describing different system classes have emerged that are tailored to the analysis (stability, reachability) of hybrid systems or the synthesis of control and monitoring schemes. Many of such analysis and synthesis tools are based on a piecewise affine (PWA) (Sontag, 1981) or a mixed logical dynamical (MLD) (Bemporad & Morari, 1999) representation of the hybrid model. In order to bridge the identification process with the following analysis/synthesis task, we provide here simple results that immediately allow mapping the HHARX or W-PWARX models, obtained from data, into MLD and PWA models, more suitable for analysis/synthesis purposes.

7.1. MLD realization

Mixed logical dynamical (MLD) systems (Bemporad & Morari, 1999) are a discrete-time formalism for systems containing both continuous and boolean/discrete variables. The key idea is to transform the Boolean variables into 0-1 integers, and to express the relations as mixed-integer linear inequalities, similarly to what was done in (11) and (31). The MLD model has the form

$$\xi_{t+1} = \Phi\xi_t + \mathscr{G}_1 u_t + \mathscr{G}_2\delta_t + \mathscr{G}_3 z_t, \tag{38a}$$

$$y_t = \mathscr{H}\xi_t + \mathscr{D}_1 u_t + \mathscr{D}_2\delta_t + \mathscr{D}_3 z_t, \tag{38b}$$

$$\mathscr{E}_2\delta_t + \mathscr{E}_3 z_t \leqslant \mathscr{E}_1 u_t + \mathscr{E}_4\xi_t + \mathscr{E}_5, \tag{38c}$$

where $\xi \in \mathbb{R}^{n_c} \times \{0,1\}^{n_\ell}$ is a vector of continuous and binary states, $u \in \mathbb{R}^{m_c} \times \{0,1\}^{m_\ell}$ are the inputs, $y \in \mathbb{R}^{p_c} \times \{0,1\}^{p_\ell}$ the outputs, $\delta \in \{0,1\}^{r_\ell}$, and $z \in \mathbb{R}^{r_c}$ are auxiliary variables.

By defining $\xi_t = [y_{t-1} \; \dots \; y_{t-n_a} \; u_{t-1} \; \dots \; u_{t-n_b}]'$, where $n_a + n_b + 1 = n$ is the dimension of the regressor vector (cf. Eq. (4)), and auxiliary variables δ_{it}, z_{it} similarly to what was done in (9)–(11), it is immediate to prove the following proposition.

Proposition 8. *HHARX models (6) admit an MLD state–space realization with $n_a + n_b$ states.*

The following proposition links W-PWARX systems to MLD systems.

Proposition 9. *W-PWARX models (28) admit an MLD state–space realization with n_a states.*

Proof. Let $\xi_t = \Phi\xi_{t-1} + \mathscr{G}_1 u_t$, $x_t = C\xi_t$, $y_t = C\xi_t - \bar{\alpha}_0 + \sum_{i=1}^{M} \pm\max\{\bar{\beta}_i C\xi_t - \bar{\alpha}_i, 0\}$ be a minimal state–space realization of (28a). Define M auxiliary binary variables $[\delta_{it}{=}1] \leftrightarrow [\bar{\beta}_i C\xi_t - \bar{\alpha}_i \geqslant 0]$ and M continuous variables $z_{it} = (\bar{\beta}_i C\xi_t - \bar{\alpha}_i)\delta_{it}$. With translations into mixed-integer inequalities as in (11) or Bemporad and Morari (1999), the MLD form can be immediately obtained. □

7.2. PWA realization

Analogously to what was defined in (2), a PWA state–space system is defined as

$$\begin{aligned} \xi_{t+1} &= A_j\xi_t + B_j u_t + f_j \\ y_t &= C_j\xi_t + D_j u_t + g_j \end{aligned} \quad \text{for} \begin{bmatrix} \xi_t \\ u_t \end{bmatrix} \in \mathscr{C}_j, \tag{39}$$

where $\xi \in \mathbb{R}^n$, and $\{\mathscr{C}_j\}_{j=0}^{s-1}$ is a polyhedral partition of the combined state-input-space. The following propositions can be obtained as corollaries of the equivalence between MLD and PWA systems (Bemporad et al., 2000a), and allows to construct a PWA state–space realization of (28) via the above MLD realization.

Proposition 10. *HHARX models (6) admit a PWA state–space realization (39) with $n_a + n_b$ states and at most 2^M regions.*

Proposition 11. *W-PWARX models (28) admit a PWA state–space realization (39) with n_a states and at most 2^M regions.*

Linear complementary (LC), Extended linear complementary (ELC), and Min–max-plus scaling (MMPS) state–space realizations can also be obtained by exploiting the equivalences described in Heemels et al. (2001).

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 algorithms that always converge to the global optimum, based on mixed-integer linear or quadratic programming. As a possible step in the direction towards faster suboptimal algorithms based on the mixed-integer approach, we have also proposed a suboptimal sliding window algorithm for HHARX models, for the case when the estimation data not so frequently switches between the different submodels.

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. There may also be possibilities for improving the computational efficiency by exploiting the structures of the specific MILP/MIQP problems.

Acknowledgements

This research was supported by the Swedish Research Council, by the European Community through the IST project "Computation and Control" and by the Italian Research Ministry through the PRIN project "Optimization and robust techniques for control of uncertain systems".

References

Batruni, R. (1991). A multilayer neural network with piecewise-linear structure and back-propagation learning. *IEEE Transactions on Neural Networks*, *2*(3), 395–403.

Bemporad, A. (2000). *Identification of hybrid systems: Global convergence via mixed-integer programming*. Technical report AUT00-28, ETH, Zürich.

Bemporad, A., Ferrari-Trecate, G., & Morari, M. (2000a). Observability and controllability of piecewise affine and hybrid systems. *IEEE Transactions on Automatic Control*, *45*(10), 1864–1876.

Bemporad, A., Garulli, A., Paoletti, S., & Vicino, A. (2003). A greedy approach to identification of piecewise affine models. In O. Maler & A. Pnueli (Eds.), *Hybrid systems: Computation and control*. Lecture Notes in Computer Science, Vol. 2623 (pp. 97–112). Berlin: Springer.

Bemporad, A., Mignone, D., & Morari, M. (1999). Moving horizon estimation for hybrid systems and fault detection. In *Proceedings of American control conference*. Chicago, IL (pp. 2471–2475).

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

Bemporad, A., Roll, J., & Ljung, L. (2000b). *Identification of hybrid systems via mixed-integer programming*. Technical report AUT00-29, ETH, Zürich.

Bemporad, A., Roll, J., & Ljung, L. (2001). Identification of hybrid systems via mixed-integer programming. In *Proceedings of the 40th IEEE conference on decision and control*, Orlando, FL, USA (pp. 786–792).

Bemporad, A., Torrisi, F. D., & Morari, M. (2000c). Optimization-based verification and stability characterization of piecewise affine and hybrid systems. In *Hybrid systems: Computation and control*. Lecture Notes in Computer Science, Vol. 1790 (pp. 45–58). Berlin: Springer.

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

Branicky, M. S., Borkar, V. S., & Mitter, S. K. (1998). A unified framework for hybrid control: Model and optimal control theory. *IEEE Transactions on Automatic Control*, *43*(1), 31–45.

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

Choi, C.-H., & Choi, J. Y. (1994). Constructive neural networks with piecewise interpolation capabilities for function approximations. *IEEE Transactions on Neural Networks*, *5*(6), 936–944.

Chutinan, A., & Krogh, B. H. (2003). Computational techniques for hybrid system verification. *IEEE Transactions on Automatic Control*, *48*(1), 64–75.

Dash Associates (1999). *XPRESS-MP user guide*. URL http://www.dashopt.com.

Ernst, S. (1998). Hinging hyperplane trees for approximation and identification. In *Proceedings of the 37th IEEE conference on decision and control*, Tampa, FL, USA Vol. 2 (pp. 1266–1271).

Ferrari-Trecate, G., Muselli, M., Liberati, D., & Morari, M. (2003). A clustering technique for the identification of piecewise affine systems. *Automatica*, *39*(2), 205–217.

Gad, E. F., Atiya, A. F., Shaheen, S., & El-Dessouki, A. (2000). A new algorithm for learning in piecewise-linear neural networks. *Neural Networks*, *13*(4–5), 485–505.

Gustafsson, F. (2000). *Adaptive filtering and change detection*. New York: Wiley.

Hagenblad, A., & Ljung, L. (2000). Maximum likelihood estimation of Wiener models. In *Proceedings of the 39th IEEE conference on decision and control*, Sydney, Australia (pp. 2417–2418).

Heemels, W. P. M. H., De Schutter, B., & Bemporad, A. (2001). Equivalence of hybrid dynamical models. *Automatica*, *37*(7), 1085–1091.

Heredia, E. A., & Arce, G. R. (1996). Piecewise linear system modeling based on a continuous threshold decomposition. *IEEE Transactions on Signal Processing*, *44*(6), 1440–1453.

Hush, D. R., & Horne, B. (1998). Efficient algorithms for function approximation with piecewise linear sigmoidal networks. *IEEE Transactions on Neural Networks*, *9*(6), 1129–1141.

ILOG, Inc. (1999). *CPLEX 6.5 user manual*. ILOG S.A., Gentilly Cedex, France.

Johansson, M., & Rantzer, A. (1998). Computation of piecewise quadratic Lyapunov functions for hybrid systems. *IEEE Transactions on Automatic Control*, *43*(4), 555–559.

Julián, P., Desages, A., & Agamennoni, O. (1999). High level canonical piecewise linear representation using a simplicial partition. *IEEE Transactions on Circuits and Systems I*, *46*, 463–480.

Julián, P., Jordán, M., & Desages, A. (1998). Canonical piecewise-linear approximation of smooth functions. *IEEE Transactions on Circuits and Systems—I: Fundamental Theory and Applications*, *45*(5), 567–571.

Kahlert, C., & Chua, L. O. (1992). The complete canonical piecewise-linear representation—part i: The geometry of the domain space. *IEEE Transactions on Circuits and Systems I*, *39*, 222–236.

Kalafatis, A. D., Wang, L., & Cluett, W. R. (1997). Identification of Wiener-type nonlinear systems in a noisy environment. *International Journal on Control*, *66*(6), 927–941.

Lin, J.-N., & Unbehauen, R. (1992). Canonical piecewise-linear approximations. *IEEE Transactions on Circuits and Systems I*, *39*(8), 697–699.

Lovera, M., Gustafsson, T., & Verhaegen, M. (2000). Recursive subspace identification of linear and nonlinear Wiener state–space models. *Automatica*, *36*(11), 1639–1650.

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

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

Medeiros, M. C., Resende, M. G. C., & Veiga, A. (1999). *Piecewise linear time series estimation with GRASP*. Technical report, AT&T Labs Research, Florham Park, NJ 07932 USA.

Murray-Smith, R., Johansen, T. A. (Eds.) (1997). *Multiple model approaches to modelling and control*. London: Taylor & Francis.

Pucar, P., & Sjöberg, J. (1998). On the hinge-finding algorithm for hinging hyperplanes. *IEEE Transactions on Information Theory*, *44*(3), 1310–1319.

Roll, J. (2001). *Robust verification and identification of piecewise affine systems*. Licentiate thesis, Department of Electrical Engineering, Linköping University. URL http://www.control.isy.liu.se/.

Roll, J. (2003). *Local and piecewise affine approaches to system identification*. Ph.D. thesis, Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden.

Sahinidis, N. V. (2000). *BARON—Branch and reduce optimization navigator*. Technical report, University of Illinois at Urbana-Champaign, Department of Chemical Engineering, Urbana, IL, USA. URL http://archimedes.scs.uiuc.edu/baron/manuse.pdf.

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

Skeppstedt, A., Ljung, L., & Millnert, M. (1992). Construction of composite models from observed data. *International Journal on Control*, *55*(1), 141–152.

Sontag, E. D. (1981). Nonlinear regulation: The piecewise linear approach. *IEEE Transactions on Automatic Control*, *26*(2), 346–358.

Sontag, E. D. (1996). Interconnected automata and linear systems: A theoretical framework in discrete-time. In R. Alur, T. A. Henzinger, & E. D. Sontag (Eds.), *Hybrid systems III—verification and control*, Vol. 1066, Lecture Notes in Computer Science (pp. 436–448). Berlin: Springer.

Strömberg, J.-E., Gustafsson, F., & Ljung, L. (1991). Trees as black-box model structures for dynamical systems. In *Proceedings of European control conference* (pp. 1175–1180). Grenoble, France.

Wigren, T. (1993). Recursive prediction error identification using the nonlinear Wiener model. *Automatica*, *29*(4), 1011–1025.

Jacob Roll was born in Stockholm, Sweden, in 1974. He obtained the M.Sc. degree in Applied Physics and Electrical Engineering in 1999, and the Ph.D. degree in Automatic Control in 2003, both from Linköping University, Sweden, where he currently works as a researcher. His main research interests are in the fields of nonlinear system identification and hybrid systems.

Alberto Bemporad received the master degree in Electrical Engineering in 1993 and the Ph.D. in Control Engineering in 1997 from the University of Florence, Italy. He spent the academic year 1996/1997 at the Center for Robotics and Automation, Department, Systems Science & Mathematics, Washington University, St. Louis, as a visiting researcher. In 1997–1999, he held a postdoctoral position at the Automatic Control

Lab, ETH, Zurich, Switzerland, where he collaborated as a senior researcher in 2000–2002. Since 1999, he is assistant professor at the University of Siena, Italy. He has published papers in the area of hybrid systems, model predictive control, computational geometry, and robotics. He is coauthoring the new version of the Model Predictive Control Toolbox (The Mathworks, Inc.). He is an Associate Editor of the IEEE Transactions on Automatic Control, and Chair of the IEEE Control Systems Society Technical Committee on Hybrid Systems.

Lennart Ljung received his Ph.D. in Automatic Control from Lund Institute of Technology in 1974. Since 1976 he is Professor of the chair of Automatic Control In Linköping, Sweden, and is currently Director of the Competence Center "Information Systems for Industrial Control and Supervision" (ISIS). He has held visiting positions at Stanford and MIT and has written several books on System Identification and Estimation. He is an IEEE Fellow and an IFAC Advisor as well as a member of the Royal Swedish Academy of Sciences (KVA), a member of the Royal Swedish Academy of Engineering Sciences (IVA), and an Honorary Member of the Hungarian Academy of Engineering. He has received honorary doctorates from the Baltic State Technical University in St Petersburg, and from Uppsala University. In 2002 he received the Quazza Medal from IFAC.