Recursive Identification for Hammerstein Systems with Bounded Noise

Mathieu Pouliquen, Eric Pigeon, Olivier Gehan
Control Group, GREYC CNRS UMR 6072
ENSICAEN, 06 Bd du Marechal Juin
14050 Caen Cedex, France
mathieu.pouliquen@unicaen.fr

Abstract—In this paper, we consider the identification of Hammerstein systems in presence of bounded disturbances. The proposed identification algorithm is a recursive algorithm. It ensures that the estimation error is bounded and allows a real-time separation of the nonlinear part and linear part. Furthermore, some stability and convergence results are demonstrated and a simulation example is given to illustrate the performance of the presented method.

I. INTRODUCTION

The Hammerstein systems are the nonlinear systems consisting of a static nonlinearity block followed by a linear dynamic system. The identification of such systems has received much interest for the last decades (see [1], [2], [6], [8], [25], [13], [26], [3], [12], [24], [9], [23], [15], [14]). The reason is that these systems take into account nonlinearities commonly encountered in practice, generated by the technological limitations necessary for the proper functioning of the system (saturation, the limit stops, etc.).

For Hammerstein systems or not, all practical identification algorithms have to deal with measurements corrupted by noise and most of the previous works consider the stochastic noise assumption (except [6] discussed later). In many cases (unknown probability distribution of the disturbances, modeling inaccuracy) such assumption cannot be made and the bounded noise assumption seems more appropriate. To address this identification problem we will use in this paper Set Membership Identification (SMI) algorithms (see [18], [17], [22] and [11]) and particularly the Optimal Bounding Ellipsoid (OBE) algorithms. These algorithms have a low computational complexity and they are appropriated to handle the identification problem in presence of bounded disturbances. Their principle consists in these systems take into account into account nonlinearities commonly encountered in practice, generated by the technological limitations necessary for the proper functioning of the system (saturation, the limit stops, etc.).

For Hammerstein systems or not, all practical identification algorithms have to deal with measurements corrupted by noise and most of the previous works consider the stochastic noise assumption (except [6] discussed later). In many cases (unknown probability distribution of the disturbances, modeling inaccuracy) such assumption cannot be made and the bounded noise assumption seems more appropriate. To address this identification problem we will use in this paper Set Membership Identification (SMI) algorithms (see [18], [17], [22] and [11]) and particularly the Optimal Bounding Ellipsoid (OBE) algorithms. These algorithms have a low computational complexity and they are appropriated to handle the identification problem in presence of bounded disturbances. Their principle consists in these systems take into account nonlinearities commonly encountered in practice, generated by the technological limitations necessary for the proper functioning of the system (saturation, the limit stops, etc.).

For Hammerstein systems or not, all practical identification algorithms have to deal with measurements corrupted by noise and most of the previous works consider the stochastic noise assumption (except [6] discussed later). In many cases (unknown probability distribution of the disturbances, modeling inaccuracy) such assumption cannot be made and the bounded noise assumption seems more appropriate. To address this identification problem we will use in this paper Set Membership Identification (SMI) algorithms (see [18], [17], [22] and [11]) and particularly the Optimal Bounding Ellipsoid (OBE) algorithms. These algorithms have a low computational complexity and they are appropriated to handle the identification problem in presence of bounded disturbances. Their principle consists in these systems take into account nonlinearities commonly encountered in practice, generated by the technological limitations necessary for the proper functioning of the system (saturation, the limit stops, etc.).

For Hammerstein systems or not, all practical identification algorithms have to deal with measurements corrupted by noise and most of the previous works consider the stochastic noise assumption (except [6] discussed later). In many cases (unknown probability distribution of the disturbances, modeling inaccuracy) such assumption cannot be made and the bounded noise assumption seems more appropriate. To address this identification problem we will use in this paper Set Membership Identification (SMI) algorithms (see [18], [17], [22] and [11]) and particularly the Optimal Bounding Ellipsoid (OBE) algorithms. These algorithms have a low computational complexity and they are appropriated to handle the identification problem in presence of bounded disturbances. Their principle consists in these systems take into account nonlinearities commonly encountered in practice, generated by the technological limitations necessary for the proper functioning of the system (saturation, the limit stops, etc.).

For Hammerstein systems or not, all practical identification algorithms have to deal with measurements corrupted by noise and most of the previous works consider the stochastic noise assumption (except [6] discussed later). In many cases (unknown probability distribution of the disturbances, modeling inaccuracy) such assumption cannot be made and the bounded noise assumption seems more appropriate. To address this identification problem we will use in this paper Set Membership Identification (SMI) algorithms (see [18], [17], [22] and [11]) and particularly the Optimal Bounding Ellipsoid (OBE) algorithms. These algorithms have a low computational complexity and they are appropriated to handle the identification problem in presence of bounded disturbances. Their principle consists in these systems take into account nonlinearities commonly encountered in practice, generated by the technological limitations necessary for the proper functioning of the system (saturation, the limit stops, etc.).

For Hammerstein systems or not, all practical identification algorithms have to deal with measurements corrupted by noise and most of the previous works consider the stochastic noise assumption (except [6] discussed later). In many cases (unknown probability distribution of the disturbances, modeling inaccuracy) such assumption cannot be made and the bounded noise assumption seems more appropriate. To address this identification problem we will use in this paper Set Membership Identification (SMI) algorithms (see [18], [17], [22] and [11]) and particularly the Optimal Bounding Ellipsoid (OBE) algorithms. These algorithms have a low computational complexity and they are appropriated to handle the identification problem in presence of bounded disturbances. Their principle consists in these systems take into account nonlinearities commonly encountered in practice, generated by the technological limitations necessary for the proper functioning of the system (saturation, the limit stops, etc.).

For Hammerstein systems or not, all practical identification algorithms have to deal with measurements corrupted by noise and most of the previous works consider the stochastic noise assumption (except [6] discussed later). In many cases (unknown probability distribution of the disturbances, modeling inaccuracy) such assumption cannot be made and the bounded noise assumption seems more appropriate. To address this identification problem we will use in this paper Set Membership Identification (SMI) algorithms (see [18], [17], [22] and [11]) and particularly the Optimal Bounding Ellipsoid (OBE) algorithms. These algorithms have a low computational complexity and they are appropriated to handle the identification problem in presence of bounded disturbances. Their principle consists in these systems take into account nonlinearities commonly encountered in practice, generated by the technological limitations necessary for the proper functioning of the system (saturation, the limit stops, etc.).

For Hammerstein systems or not, all practical identification algorithms have to deal with measurements corrupted by noise and most of the previous works consider the stochastic noise assumption (except [6] discussed later). In many cases (unknown probability distribution of the disturbances, modeling inaccuracy) such assumption cannot be made and the bounded noise assumption seems more appropriate. To address this identification problem we will use in this paper Set Membership Identification (SMI) algorithms (see [18], [17], [22] and [11]) and particularly the Optimal Bounding Ellipsoid (OBE) algorithms. These algorithms have a low computational complexity and they are appropriated to handle the identification problem in presence of bounded disturbances. Their principle consists in these systems take into account nonlinearities commonly encountered in practice, generated by the technological limitations necessary for the proper functioning of the system (saturation, the limit stops, etc.).

For Hammerstein systems or not, all practical identification algorithms have to deal with measurements corrupted by noise and most of the previous works consider the stochastic noise assumption (except [6] discussed later). In many cases (unknown probability distribution of the disturbances, modeling inaccuracy) such assumption cannot be made and the bounded noise assumption seems more appropriate. To address this identification problem we will use in this paper Set Membership Identification (SMI) algorithms (see [18], [17], [22] and [11]) and particularly the Optimal Bounding Ellipsoid (OBE) algorithms. These algorithms have a low computational complexity and they are appropriated to handle the identification problem in presence of bounded disturbances. Their principle consists in these systems take into account nonlinearities commonly encountered in practice, generated by the technological limitations necessary for the proper functioning of the system (saturation, the limit stops, etc.).

For Hammerstein systems or not, all practical identification algorithms have to deal with measurements corrupted by noise and most of the previous works consider the stochastic noise assumption (except [6] discussed later). In many cases (unknown probability distribution of the disturbances, modeling inaccuracy) such assumption cannot be made and the bounded noise assumption seems more appropriate. To address this identification problem we will use in this paper Set Membership Identification (SMI) algorithms (see [18], [17], [22] and [11]) and particularly the Optimal Bounding Ellipsoid (OBE) algorithms. These algorithms have a low computational complexity and they are appropriated to handle the identification problem in presence of bounded disturbances. Their principle consists in these systems take into account nonlinearities commonly encountered in practice, generated by the technological limitations necessary for the proper functioning of the system (saturation, the limit stops, etc.).

For Hammerstein systems or not, all practical identification algorithms have to deal with measurements corrupted by noise and most of the previous works consider the stochastic noise assumption (except [6] discussed later). In many cases (unknown probability distribution of the disturbances, modeling inaccuracy) such assumption cannot be made and the bounded noise assumption seems more appropriate. To address this identification problem we will use in this paper Set Membership Identification (SMI) algorithms (see [18], [17], [22] and [11]) and particularly the Optimal Bounding Ellipsoid (OBE) algorithms. These algorithms have a low computational complexity and they are appropriated to handle the identification problem in presence of bounded disturbances. Their principle consists in these systems take into account nonlinearities commonly encountered in practice, generated by the technological limitations necessary for the proper functioning of the system (saturation, the limit stops, etc.).

II. PROBLEM STATEMENT

The Hammerstein system is described by

\[ A(q)y_t = q^{-d}B(q)x_t + v_t \]  \hspace{1cm} (1)

with \[ B(q) = b_0 + b_1q^{-1} + \cdots + b_nq^{-n_b} \] \hspace{1cm} and \[ A(q) = 1 + a_1q^{-1} + \cdots + a_nq^{-n_a} \]

\[ x_t = h(u_t) \] \hspace{1cm} (2)

where \( u_t \) and \( y_t \) respectively the system input and the system output. \( x_t \) is the internal signal and \( v_t \) is an unknown but bounded disturbing term (noise measurement, state disturbances, modeling inaccuracy, etc.).

\( h(.) \) is a static nonlinear function without memory. It is represented by a sum of the known nonlinear functions \( h_1(\cdot), h_2(\cdot), \ldots, h_{n_h}(\cdot) \):

\[ h(u_t) = \mu_1h_1(u_t) + \mu_2h_2(u_t) + \cdots + \mu_{n_h}h_{n_h}(u_t) \]  \hspace{1cm} (3)

The objective: The objective is to estimate recursively with a one shot algorithm the linear part and the non-linear. This must be done from the input-output measurements of the system described by (1), (2) and (3).

Remark 1: In [6] a similar problem is treated. The differences in [6] are: the linear part is a OE model, the algorithm is not a recursive algorithm, several experiments are needed to identify some points of \( h(.) \) and the separation of the linear and nonlinear part is not realized in the same time.

In this paper, so as to be able to identify the parameters, some assumptions are needed:

A1 The delay \( d \), the orders of the linear part and of the nonlinear part \( (n_a, n_b \text{ and } n_h) \) are supposed to be available;

A2 The coefficient \( b_i \) are such that \( b_0 = 1 \);

A3 Regardless of the disturbance nature, an upper bound is supposed to be available on the magnitude of \( v_t \):

\[ |v_t| \leq \delta_v \] \hspace{1cm} (4)

(The bound \( \delta_v \) could be time-varying as in [4] or [6].)
Remark 2: The assumption [A2] is a common assumption necessary so as to satisfy the identifiability condition. Without this assumption we are not able to distinguish the solution $G(q)h(u_t)$ from the solution $(aG(q))\frac{1}{a}h(u_t)$ with $a \neq 0$. Note that there exist other similar assumptions in literature to get a unique parametrization (see [1], [6] or [3] for example).

III. IDENTIFICATION ALGORITHM AND ANALYSIS

From (1), (2) and (3) the output of the Hammerstein system is expressed as follows:

$$y_t = \sum_{k=0}^{n_b} b_k \sum_{i=1}^{n_h} \mu_i h_i(u_{t-k-1}) + \sum_{k=1}^{n_u} a_k y_{t-k} + v_t$$

By setting

$$\theta^* = \begin{pmatrix} \mu_1 \\ b_{n_h} \mu_1 \\ \vdots \\ \mu_n \\ b_{n_h} \mu_n \\ a_1 \\ \vdots \\ a_{n_u} \end{pmatrix}, \quad \phi = \begin{pmatrix} h_1(u_{t-d}) \\ \vdots \\ h_1(u_{t-d-n_b}) \\ \vdots \\ h_{n_h}(u_{t-d-n_b}) \\ \vdots \\ h_{n_h}(u_{t-d-n_b}) - y_{t-1} \end{pmatrix}$$

the system is rewritten in the usual over-parameterized form

$$y_t = \phi^T \theta^* + v_t$$

$$\theta^* \in \mathbb{R}^n$$ is the unknown parameter vector to be identified with $n = (1 + n_h)n_b + n_u$ the number of parameters.

The first $(1 + n_b)n_b$ terms of the vector $\theta^*$ can be inserted in a $(1 + n_b) \times n_b$ matrix as follows

$$M_{\theta^*} = \begin{pmatrix} \theta^*(1) & \cdots & \theta^*((1+n_b)(n_b-1)+1) \\ \vdots \\ \theta^*(1+n_b) & \cdots & \theta^*((1+n_b)n_b) \end{pmatrix}$$

which can be rewritten as

$$M_{\theta^*} = \begin{pmatrix} 1 \\ b_1 \\ \vdots \\ b_{n_h} \end{pmatrix} \begin{pmatrix} \mu_1 & \cdots & \mu_n \end{pmatrix}$$

Such structure will be used to separate the parameters $\{b_i\}$ and $\{\mu_i\}$ in subsection III-C. By now, our aim in what follows is to

- propose an estimation algorithm based on (6) and (4);
- estimate a parameter vector structured as (7);
- give a proof of stability and convergence of this algorithm.

The algorithm considered in this paper recursively estimates the parameters $\theta^*$. For this purpose the one-step-ahead output predictor is defined as follows ([16]):

$$\hat{y}_t = y_t - v_t = \phi^T \theta^*$$

Before introducing the parameter adaptation algorithm, let define the a priori and a posteriori predictors as

$$\begin{cases} \hat{y}_{t-1} = \phi^T \hat{\theta}_{t-1} \\ \hat{y}_{t-1} = \phi^T \hat{\theta}_{t} \end{cases}$$

$\hat{\theta}_t$ represents an estimation of the parameter vector at the actual time $t$. Let define also $\hat{\theta}^u_t$ and $\hat{\theta}^u_{t-1}$ as follows:

- $\hat{\theta}^u_t$ is an estimation which satisfies (7);
- $\hat{\theta}^u_{t-1}$ is an estimation which does not satisfy (7).

Subscripts $^u$ and $^u_{t-1}$ respectively stand for "structured" and "un-structured".

A. The adaptation algorithm.

To address the identification problem we will take inspiration from the OBE identification methods presented in [5], [21] and [20]. For each time $t$ the proposed algorithm is divided in two successive stages:

- **stage 1** Update of the estimated parameter vector taking into account the new observation vector $\phi$ and the new output $y_t$: estimation of $\hat{\theta}^u_t$ from $\hat{\theta}^u_{t-1}$;
- **stage 2** Structuring of the estimated solution (separate the linear part and the non-linear part): estimation of $\hat{\theta}^u_t$ from $\hat{\theta}^u_{t-1}$.

The estimated parameter vector must satisfy (7) and has to maintain the prediction error below a bound defined from the upper bound on the disturbance $v_t$. Under some conditions describes below, the following estimation algorithm provides such estimation:

**stage 1**

$$\begin{cases} \hat{\theta}^u_t = \hat{\theta}^u_{t-1} + \Gamma_t \hat{y}_t \hat{\theta}^u_{t-1} \\ \Gamma_t = \frac{P_t^u}{\lambda + \phi^T \phi_t P_t^u} \phi_t \sigma_t \\ \hat{y}_t = y_t - \hat{y}_{t-1} \hat{\theta}^u_{t-1} \\ (P_t^u)^{-1} = \lambda (P_{t-1}^u)^{-1} + \phi_t \sigma_t \phi_t^T \\ P_t^u = \frac{1}{1 + \lambda} P_{t-1}^u + \frac{1}{\lambda} W_t \end{cases}$$

Some useful insights on the algorithm are given below.

**about stage 1:**

- $\hat{\theta}^u_{t-1}$ and $P_{t-1}^u$ are computed at the previous iteration.
- $\hat{y}_t, \hat{\theta}^u_{t-1}$ is the a priori prediction error, $0 < \lambda \leq 1$ is the forgetting factor and $\sigma_t$ is a switching flag given by:

$$\sigma_t = \begin{cases} \frac{\lambda}{\phi_t^T P_t^u \phi_t} \left( \frac{\hat{y}_t \hat{\theta}^u_{t-1}}{\sigma_t} \right) - 1 \\ \text{if } \left( \frac{\hat{y}_t \hat{\theta}^u_{t-1}}{\sigma_t} \right) > 1 \\ \text{and } (\phi_t^T P_{t-1}^u \phi_t > 0) \\ 0 \text{ otherwise} \end{cases}$$

Its role is specified below.

From (8) the unstructured a posteriori prediction error $\tilde{y}_t, \tilde{\theta}^u_{t-1}$ can be written as

$$\tilde{y}_t, \tilde{\theta}^u_{t-1} = (1 - \phi_t^T \Gamma_t) \hat{y}_t, \hat{\theta}^u_{t-1}$$

4920
Using the value of $\sigma_t$ for $\sigma_t \neq 0$ (11) yields

\[
|\tilde{y}_t, \hat{\theta}_s| = 1. \quad \text{This clearly shows that, provided that} \ \phi_t^T P_{t-1} \phi_t > 0, \ \text{the introduction of} \ \sigma_t \ \text{in the first part of the algorithm ensures the following property:}
\]

\[
\left\{ \begin{array}{l}
\frac{\hat{\sigma}_t}{\sigma_t} > 1 \ \text{then} \ \sigma_t 
\neq 0 \ \text{and} \ \frac{\hat{\sigma}_t}{\sigma_t} = 1 \\
\frac{\hat{\sigma}_t}{\sigma_t} \leq 1 \ \text{then} \ \sigma_t = 0 \ \text{and} \ \frac{\hat{\sigma}_t}{\sigma_t} \leq 1
\end{array} \right.
\]

$\delta$ is then a bound on the magnitude of the unstructured a posteriori prediction error. This bound has to be specified taking into account the bound $\delta_s$.

about stage 2:

The second part of the algorithm as to realize the structuring of $\hat{\theta}_s^{\mu*}$. This is done by a function $S(\cdot)$. $\hat{\theta}_s = S(\hat{\theta}_s^{\mu*})$. This function will be detailed in subsection III-C (see table I page 4).

The interest of matrix $W_t$ is to maintain the stability of the algorithm. In fact, the use of that structuring step introduce a correction term $w_t$ in the following manner

\[
W_t = \hat{\theta}_s - \hat{\theta}_s^{\mu*}
\]

and that correction term disturbs the algorithm behavior. So as to take into account this disturbance $W_t$ is designed in the following manner:

\[
W_t: \left\{ \begin{array}{l}
W_t = 0 \ \text{if} \ w_t = 0 \\
W_t \ \text{such that} \ w_t^T W_t^{-1} w_t \leq 1 \ \text{otherwise}
\end{array} \right.
\]

(12)

$0 < \zeta < 1$ is a design parameter whose role will be discussed later.

Let notice that without the second stage, the above algorithm is similar to OBE algorithms presented in [5], [21] and [4]. The main interest of the proposed algorithm is the introduction of the correction term $w_t$ in the second stage. In this paper this correction term corresponds to a perturbation introduces during the real-time separation of the linear part and the non-linear part. In the next subsection it is shown that stability and convergence properties can be demonstrated even in presence of that correction term.

B. Stability and convergence properties.

The analysis of the estimation algorithm is performed via the analysis of the estimation error $\hat{\theta}_s - \theta^* - \hat{\theta}_s^{\mu*}$ in the first part of the analysis some geometrical interpretations of the adaptation algorithm are provided. To this end let us define for each time $t$ the observation set $\mathcal{S}_t$:

\[
\mathcal{S}_t = \left\{ \theta \in \mathbb{R}^n : \frac{\hat{\theta}_s - \theta - \hat{\theta}_s^{\mu*}}{\sigma_t} \leq 1 \right\}
\]

Given $y_t$ and $\phi_t$, $\mathcal{S}_t$ is the set of all possible $\theta$ which are consistent with the chosen threshold $\delta$. An important property of that observation set is given in the following propertie.

Propertie 1: Assume that $\delta \geq \delta_s$, then

\[
\theta^* \in \mathcal{S}_t
\]

$\mathcal{S}_t$ is then a strip in parameter space bounded by two parallel hyperplanes and which contains $\theta^*$. The principle of OBE type algorithms is to find a parameter vector included in $\mathcal{S}_t$. It is not necessary to know exactly the value of $\delta$, however more $\delta$ is close to $\delta_s$, more the identification is accurate.

Let define the un-structured ellipsoid $\mathcal{E}_t^{\mu*}$:

\[
\mathcal{E}_t = \left\{ \theta \in \mathbb{R}^n : \sigma_t P_t^{-1} (\theta - \hat{\theta}_s) \leq 1 \right\}
\]

and the structured ellipsoid $\mathcal{E}_t^*$:

\[
\mathcal{E}_t^* = \left\{ \theta \in \mathbb{R}^n : \sigma_t P_t^{-1} (\theta - \hat{\theta}_s) \leq 1 \right\}
\]

(15)

The following theorem shows that the proposed identification algorithm provides rules for computing $\hat{\theta}_s$ and $P_t^*$ so that $\mathcal{E}_t^*$ contains the intersection ($\mathcal{S}_t \cap \mathcal{E}_t^*_{t-1}$).

Theorem 1: Consider the class of systems defined in section II and the identification algorithm given by (8), (9), (10) and (12). If $\delta \geq \delta_s$ hold and suppose that

\[
\theta^* \in \mathcal{E}_t^*_{t-1}
\]

then for all initial conditions

- a- An outer bounding ellipsoid of ($\mathcal{S}_t \cap \mathcal{E}_t^*_{t-1}$) is given by $\mathcal{E}_t^*$;
- b- $\theta^* \in \mathcal{E}_t^*$

(17)

If furthermore, $\{\phi_t\}$ is a persistently exciting sequence of order $o_e \geq n$, i.e there exist $\alpha > 0$ and $\beta > 0$ such that for all $t \geq o_e$

\[
a_{h} \leq \sum_{i=0}^{\beta-1} \phi_{t+i} \sigma_{t+i} \phi_{t+i} \leq \beta a_{h}
\]

(18)

and if there exists $\beta_w > 0$ such that for all $t$

\[
0 \leq w_t \leq \beta_w a_{h}
\]

(19)

Then the following property holds:

- c- The volume of the ellipsoid $\mathcal{E}_t^*$ is bounded.

A sketch of proof is given in appendix A page 6. A consequence of this result is the fact that if in the beginning (i.e. at the time $t = 0$) $\theta^*$ is included in $\mathcal{E}_0^*$ (this is true with for example $\hat{\theta}_s^0 = 0$ and $P_0 = \mu I_n$ with $\mu$ large enough) then $\theta^*$ belongs to $\mathcal{E}_t^*$ for all $t$.

The second part of the analysis is performed via the analysis of the estimation error $\hat{\theta}_s^*$ which is a non linear function of $\hat{\theta}_s^{t-1}$, $w_t$ and $v_t$. Theorem 2 focuses on the behavior of $\hat{\theta}_s^{t-1}$.

It is shown that under some conditions the center $\hat{\theta}_s^{t-1}$ of the ellipsoid $\mathcal{E}_t^*$ converges next to the solution $\theta^*$.

Theorem 2: Consider the class of systems defined in section II and the identification algorithm given by (8), (9), (10) and (12). If $\delta \geq \delta_s$ hold and if conditions (18) and (19) in Theorem 1 hold, then for all initial conditions

- a- for all $t \geq o_e$

\[
|\hat{\theta}_s^t|^2 \leq \gamma_1 (1 - \zeta_t^2) |\hat{\theta}_s^{t-1}|^2 + \gamma_2 \frac{1 - (1 - \zeta_t^2) \beta}{1 - \zeta_t^2} \beta
\]

(20)
with $\gamma_1$ and $\gamma_2$ two finite scalars.

If, furthermore, there exists a time $t_s$ such that for all $t \geq t_s$

$$w_t = 0$$

(21)

the following properties hold

- b- for all $t \geq t_s + \alpha_e$

$$|\hat{\theta}_t|^2 \leq \gamma_1 (1 - \zeta)^{t-t_s} \lambda^{t-t_s}$$

(22)

with $\gamma_1$ a finite scalar.

- c- one has

$$\lim_{t \to \infty} \left| \frac{\tilde{y}_t, \hat{\theta}_t}{\sigma} \right| \leq 1$$

(23)

A sketch of proof is given in appendix B page 6.

Remark 3: Note that if $\delta = \delta_s$ and $w_i = 0$ then from (22) we have

lim$_{t \to \infty} \hat{\theta}_t = \hat{\theta}$ where $\hat{\theta}$ parametrizes a model such that:

$$\hat{\theta}(q)y_t = q^{-d} \hat{B}(q) \hat{x}_t + \tilde{z}_t$$

with

$$\hat{\theta}_0 = \hat{B}(u_t)$$

(24)

which is coherent with the objective stated in section II.

C. Design of function $S(.)$.

In our algorithm, at each time $t$ the linear part and the non-linear part are separated. This is the aim of the function $S(.)$. This function must satisfy three conditions:

- $\hat{\theta}_t = S(\hat{\theta}_t)$ must be structured as (7);

- $\hat{\theta}_t$ must satisfy

$$\left| \frac{\tilde{z}_t, \hat{\theta}_t}{\sigma} \right| \leq 1;$$

- The correction term $w_t = \hat{\theta}_t - \hat{\theta}_t$ must have a finite upper bound (condition (19) on $W_t$ in Theorem 1).

This is this structured parameter vector $\hat{\theta}_t$ that will be updated at next time.

The proposed function $S(.)$ corresponds to the 5 steps procedure given in table I.

The interest of this procedure is twofold:

- It provides structuring of $\hat{\theta}_t$.

- $\hat{\theta}_t$ is such that the magnitude of the a posteriori prediction error $\tilde{y}_t, \hat{\theta}_t$ is bounded by $\delta$.

The condition on the existence of a finite scalar $\beta_W$ is difficult to guarantee in such a recursive algorithm. The procedure of table I is ill-conditioned if the parameters $\{\mu_i\}$ extracted in step 1 are close to zero. This is the reason why we advice the use of the above procedure and the use of the second stage of the algorithm after a few times. Typically the implementation of the procedure must be realized only after $n_a + n_b + n_a$ samples (this corresponds to the number of parameters to estimate). Numerical simulations have shown the efficiency of that procedure. This is illustrated in the next section.

Remark 4: It is possible to make use of a Singular Values Decomposition (SVD) in step 2. The algorithm will have a higher computational complexity, however it will be numerically more robust when dealing with ill-conditioned data.

Function $S(.)$ - Estimation of a structured parameter vector

step 1: Extract the parameters $\{a_i\}$ and $\{\mu_i\}$ from $\hat{\theta}_t$.

step 2: If $\sum_{i=1}^{n_2} \mu_i^2 \neq 0$, compute a first estimation of $P = \left( \begin{array}{c} b_1 \\ \vdots \\ b_{n_b} \end{array} \right)$ (making use of $M_{\hat{\theta}_t}$). This estimation is denoted $P^{(1)}$.

If $\sum_{i=1}^{n_2} \mu_i^2 = 0$, take $P^{(1)} = 0$.

step 3: Compute the predictor $\hat{y}_t = \phi_t^T P^{(1)} + z_t$.

step 4: Realize a second estimation $P^{(2)}$ as follows:

$$P^{(2)} = P^{(1)} + \frac{\phi_t, \sigma'}{1 + \phi_t^T \phi_t, \sigma'} (y_t - \hat{y}_t)$$

with

$$\sigma'_i = \begin{cases} \frac{1}{\phi_t^T \phi_t} \left( \frac{y_t - \hat{y}_t}{\sigma} - 1 \right) & \text{if } \frac{y_t - \hat{y}_t}{\sigma} > 1 \text{ and } (\phi_t^T \phi_t > 0) \\ 0 & \text{otherwise} \end{cases}$$

step 5: Finally $\hat{\theta}_t$ is the parameterized vector built from the estimates for $\{a_i\}$ (step 1), $\{b_i\}$ (step 4) and $\{\mu_i\}$ (step 1).

TABLE I

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>0.5</td>
</tr>
<tr>
<td>$a_2$</td>
<td>0.2</td>
</tr>
<tr>
<td>$b_1$</td>
<td>0.1</td>
</tr>
</tbody>
</table>

IV. NUMERICAL EXAMPLE

Numerical data have been generated according to (1), (2), (3) and (4). The following system has been considered in [3]:

$$q^{-d} B(q) = q^{-1} (1 + 0.5 q^{-1} + 0.5 q^{-2})$$

$$A(q) = 1 - 0.3 q^{-1} - 0.2 q^{-2} + 0.3 q^{-3}$$

$$h(u_t) = h_1(u_t) + 0.2983 h_2(u_t)$$

with $h_1(u_t) = u_t$ and $h_2(u_t) = 2u_t^2$.

The input $u_t$ is white noise uniformly distributed in $[-1;1]$ of length $N = 4000$. $v_t$ is a noise generated in the following manner:

$$v_t = \delta \frac{1}{2} (e_t + r_t)$$

where $r_t$ is a binary square signal with a period of 20 samples, $e_t$ is a white noise uniformly distributed in $[-1;1]$ and $\delta$ is close to 0.7 in order to have a signal to noise ratio close to SNR = 5dB.

The parameters $\{a_i\}$, $\{b_i\}$ and $\{\mu_i\}$ are estimated with three recursive algorithm:

- Procedure 1: this first procedure corresponds to the identification of an extended ARX structure using (6) and the ordinary least square algorithm. In order to separate the parameters $\{b_i\}$ and $\{\mu_i\}$ a SVD is applied at each iterations.

- Procedure 2: this second procedure corresponds to our identification algorithm without stage 2. However the parameters $\{b_i\}$ and $\{\mu_i\}$ are separated using the procedure presented in subsection III-C. In fact this amounts to not use the structured parameter vector estimated at
time $t$ for time $t+1$. The threshold $\delta$ is taken equal to $\delta_v$.

- Procedure 3: this third identification procedure is our identification algorithm using stages 1 and 2. Here again $\delta = \delta_v$.

For each procedure the value of the forgetting factor is $\lambda = 0.99$.

In Fig. 1, 2 and 3 the thick solid lines denote the true parameters and the other lines the parameters estimates. These figures present the estimations for $a_1$, $b_1$ and $\mu_1$. The other components have similar behavior. In Fig. 4, 5 and 6 the solid lines denote thresholds $\pm \delta_v$, dotted lines correspond to the a posteriori prediction error for each identification procedure.

From these figures we conclude the following:

- Procedure 1 is the only procedure for which the a posteriori prediction error is not bound by $\delta$. Moreover procedure 1 clearly delivers biased estimates (see Fig. 1 and 2). This is due to the fact that the disturbing term $v_t$ is not a white noise, the ordinary least square algorithm is not adapted to the considered identification problem.

- The estimations computed with procedures 2 and 3 converge towards a neighborhood of the true parameters. One point that should be emphasized is that the fact that the convergence towards the true parameters is faster with procedure 3. This is a consequence of taking into account the specific structure of the parameters vector i.e. the use of stage 2: the difference between procedures 2 and 3 lies in the introduction of the correction term $w_t$ in procedure 3. This correction term affects the behavior of the algorithm without loss of the stability property as demonstrated in subsection III-B.

This example shows the ability of the proposed algorithm to estimate, through a real-time procedure, a model satisfying 24.

Remark 5: An overestimation of $\delta_v$ (i.e. $\delta > \delta_v$) would generate a larger ellipsoid $\delta_v^2$ and consequently a convergence in a larger neighborhood of the true parameters. For lack of space we do not illustrate this in this example.

V. CONCLUSION AND FUTURE WORKS

A recursive algorithm for Hammerstein systems identification has been proposed. This algorithm is adapted for the bounded noise case. It consists in two successive stages:
the first stage realizes an update of the estimate from new measurements, the second stage allows an online separation of the nonlinear part and the linear part. Stability and convergence properties have been investigated. The second stage of the algorithm consists in the introduction of a "disturbance term" on the parameters computed in the first stage. This "disturbance term" is used here in order to separate the nonlinear part and the linear part. In this paper the linear part is an ARX process, future researches will concentrate on a less restrictive structure: an OE process.

REFERENCES


A- PROOF OF THEOREM 1

For lack of space we do not detail all the proof, we will simply state the essential points.

- First, from (13) and (15) it can easily be shown that $\mathcal{F}_1 \cap \mathcal{E}^{-1}_1 \subset \mathcal{E}^{-1}_2$.
- Second, let consider the following lemma from [19]:

**Lemma 1:** [19] Consider two ellipsoids $\delta_1 = \left\{ \theta : (\theta - \theta_0) \neq P^{-1}_1 (\theta - \theta_0) \leq 1 \right\}$ and their sum $\delta_1 \oplus \delta_2 = \left\{ \theta : \theta_1 + \theta_2 \in \delta_1 \text{ and } \theta_2 \in \delta_2 \right\}$. Then $\forall \epsilon > 0$ such that $0 < \epsilon < 1$, we have $\delta_1 \subseteq \delta_1 \oplus \delta_2$ and $\delta_1 \oplus \delta_2 = \left\{ \theta : (\theta - \theta_0) \neq P^{-1}_1 (\theta - \theta_0) \leq 1 \right\}$, $\theta_1 = \theta_1^T + \theta_2$ and $\lambda = \frac{1}{\lambda_1^2} + \frac{1}{\lambda_2^2}$.

It can be shown that $\mathcal{F}_1 \cap \mathcal{E}^{-1}_1 \subset \mathcal{E}^{-1}_2$ if we take $P_1 = P_1^\parallel$, $P_2 = W_1$, $P_3 = P_3^\parallel$, and $\delta_1 = \delta_1^T + \delta_2 = \delta_1^T + W_1$. From the two previous results it is easy to conclude $\mathcal{F}_1 \cap \mathcal{E}^{-1}_1 \subset \mathcal{E}^{-1}_2 \subset \mathcal{E}^{-1}_3$, this corresponds to proposition a--. Using assumption (16) and the fact that $\delta \geq \delta_1$, proposition $b$ is a consequence of proposition $a$--.

- From (8) and (9) it can be shown there exist scalars $s_{sup}$ and $s_{inf} \neq 0$ such that $s_{sup}P_1 \geq P_1^{-1} \geq s_{inf}$, then, proposition $c$-- is a consequence of the fact that $P_1^\parallel$ is bounded.

B- PROOF OF THEOREM 2

For lack of space we do not detail all the proof, we will simply state the essential points. Consider the following Lyapunov function $V_f = \tilde{T}_f^T P_1^{-1} \tilde{T}_f$ with $\tilde{T} = \tilde{t}_s - \tilde{t}_t$.

- $V_f$ can be rewritten as $V_f = (\tilde{\theta}_s - \tilde{\theta}_t)^T \left( \frac{1}{\lambda_1^2} \tilde{\theta}_s + \tilde{\theta}_t \right)^{-1} (\tilde{\theta}_s - \tilde{\theta}_t)$.

We have $\tilde{\theta}_s^T T_1^{-1} \tilde{\theta}_s \leq 1$ and $W_2^T W_2^{-1} W_1 \leq 1$ this gives

$$V_f \leq (1-\gamma)V_f^s \frac{\lambda}{\lambda_1^2} W_1^{-1} W_2$$

(26)

- Using the fact that $\| \tilde{\theta}_s - \tilde{\theta}_t \| \leq \delta$ and $\delta \leq \delta$ it can be shown that

$$V_f^s \leq \lambda V_f^{-1}$$

(27)

From (26) and (27) we obtain $V_f^s \leq (1-\gamma)\lambda V_f^s + \delta \Sigma_{\theta_0}^{-1}(1-\gamma)\lambda^T$. We know that $0 < s_{sup} \leq P_1^{-1} \leq s_{sup}$, then $s_{sup} / \lambda \leq (1-\gamma)\lambda V_f^s + \delta \Sigma_{\theta_0}^{-1}(1-\gamma)\lambda^T$. Proposition $a$ is a consequence of this inequality.

If there exists a time $t_0$ such that for all $t \geq t_0$, $w_1 = 0$, it can be shown that $V_f^s \leq (1-\gamma)\lambda V_f^s + \delta \Sigma_{\theta_0}^{-1}(1-\gamma)\lambda^T$, for $t \geq t_0$. Proposition $b$ is a consequence of this inequality with $\gamma = V_f^{-1} / s_{inf}$. With $0 < \lambda \leq 1$ and $0 < \gamma < 1$, proposition $c$ follows from proposition $b$. 

4924