



**HAL**  
open science

## Recursive subspace identification of Hammerstein models based on LS-SVM

Laurent Bako, Guillaume Mercère, Stéphane Lecoeuche, Marco Lovera

► **To cite this version:**

Laurent Bako, Guillaume Mercère, Stéphane Lecoeuche, Marco Lovera. Recursive subspace identification of Hammerstein models based on LS-SVM. Adaptation and Learning in Control and Signal Processing 2007, ALCOSP'07, Aug 2007, Saint Petersburg, Russia. pp.CDROM. hal-00180151

**HAL Id: hal-00180151**

**<https://hal.science/hal-00180151>**

Submitted on 17 Oct 2007

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# RECURSIVE SUBSPACE IDENTIFICATION OF HAMMERSTEIN MODELS BASED ON LS-SVM

Laurent Bako<sup>\*,\*\*\*</sup> Guillaume Mercère<sup>\*\*</sup>  
Stéphane Lecoeuche<sup>\*,\*\*\*</sup> Marco Lovera<sup>\*\*\*\*</sup>

*\* Ecole des Mines de Douai, Département Informatique et  
Automatique, 59508, Douai, France*

*\*\* Laboratoire d'Automatique et d'Informatique  
Industrielle, Poitiers, France*

*\*\*\* Laboratoire d'Automatique, Génie Informatique et  
Signal - UMR 8146, Villeneuve d'Ascq 59655, France*

*\*\*\*\* Dipartimento di Elettronica e Informazione,  
Politecnico di Milano, Italy*

Abstract: This paper presents a recursive scheme for the identification of Hammerstein MIMO models. The Markov parameters of the system are determined first by a Least Squares Support Vector Machines (LS-SVM) regression through an over-parameterization technique. Then, a state space realization of the system is retrieved using an adapted online subspace identification method. Simulation results are provided to demonstrate the effectiveness of the algorithm in the presence of white output noise. *Copyright © 2007 IFAC*

Keywords: LS-SVM, identification, subspace methods, recursive algorithms, nonlinear systems

## 1. INTRODUCTION

System identification based on input-output data measurements is a necessary step in most applications such as control design, plant diagnosis and monitoring. A number of methods have been developed for the identification of linear systems as well in input-output form as in state space representation. Some of the proposed techniques have even been extended to slowly time varying systems (Lovera et al., 2000), (Mercère et al., 2003), (Oku and Kimura, 2002). But nonlinear non-stationary systems identification in state space form still remains a very challenging research area. To that purpose the identification of a Hammerstein system is considered in this paper. This model consists of a nonlinear memoryless function  $f(\cdot)$  followed by a linear dynamic system. The difficulty in separating the nonlinear part

from the linear one is related to the fact that their respective terms are present only in cross-product forms in the system equations. A typical way to proceed is to treat each of these cross-product terms as an unknown. That results in an increased number of unknowns and the procedure is referred to as the over-parameterization method (Bai, 1998), (Goethals et al., 2005b). Generally, the static nonlinearity  $f(\cdot)$  is expanded on a basis of known elementary functions (Radial Basis Functions, polynomials, B-splines, ...) and then, the coordinates associated to it are estimated. (Dempsey and Westwick, 2004), (Ramos and Durand, 1999), (Vörös, 2003). Many proposed methods are based on input-output models and very few are based on state space models which may be sometimes more convenient for describing systems and especially multivariable systems. The features

shared by most of these methods are that they often identify an input-output behavior of the system and use either iterative optimization or over-parameterization or basis expansion technique to deal with the non-linearity. The present paper is inspired by (Goethals et al., 2005a) which firstly demonstrated the efficiency of componentwise LS-SVM in Hammerstein systems identification. Note that the proposed approach is very similar in the results to the well-known technique of expansion the nonlinearity on gaussian functions basis. Reliable results have been obtained for Linear Time Invariant systems. However, in actual problems there exist only a few number of such systems; almost any real life system is subject to some parameters variation during its operating time. In order to make up for this evolution, we extend in this paper the work presented in (Goethals et al., 2005a) to time-varying systems by using LS-SVM to estimate the nonlinear part of the system and ordinary least squares for recovering the linear part in state space form.

## 2. PROBLEM STATEMENT

Consider a discrete-time SIMO Hammerstein state space system represented as follows

$$\begin{cases} x_{t+1} = Ax_t + Bf(u_t) + w_t \\ y_t = Cx_t + Df(u_t) + v_t \end{cases} \quad (1)$$

where  $y_t \in \mathbb{R}^{n_y}$  and  $x_t \in \mathbb{R}^{n_x}$  are the output and state vectors respectively. The input  $u_t \in \mathbb{R}$  is a white gaussian sequence and  $w_t \in \mathbb{R}^{n_x}$ ,  $v_t \in \mathbb{R}^{n_y}$  are some zero-mean noises. Moreover, the system under consideration is assumed to be stable and of known order  $n_x$ .  $f(\cdot)$  is a scalar-valued smooth nonlinear function. Given  $N$  consecutive input-output data measurements  $\{u_t, y_t\}_{t=1}^N$ , the following work aims at determining the parameters  $(A, B, C, D)$  and  $f$  of the system in (1) up to a similarity transformation.

## 3. IDENTIFICATION

Let us write  $y_t$  as a  $d$ -step ahead predictive model using equation (1). The equation obtained is very similar to a Finite Impulse Response (FIR) model

$$y_t = CA^d x_{t-d} + \sum_{i=0}^d H_i f(u_{t-i}) + \underbrace{\sum_{i=1}^d CA^{i-1} w_{t-i}}_{e_t} + v_t \quad (2)$$

$\forall t > d$ , where  $d$  is a positive integer,  $e_t$  is the equation error and  $H_i$ , defined as

$$H_i = \begin{cases} D & \text{if } i = 0 \\ CA^{i-1}B & \text{if } i \geq 1 \end{cases} \quad (3)$$

are the so-called Markov parameters and represent the impulse responses of the system. These parameters do not depend on the basis of the state

space representation in (1). Equation (2) amounts to

$$y_t(s) = C_s A^d x_{t-d} + \sum_{i=0}^d H_i(s) f(u_{t-i}) + e_t(s) \quad (4)$$

$s = 1, \dots, n_y$ , where  $y_t(s)$  is the  $s$ th output,  $C_s$  the  $s$ -th row of the observation matrix  $C$ .

In (Goethals et al., 2005a), a method for off-line identification of Hammerstein models using LS-SVM regression techniques was proposed. This method produces reliable results but is computationally heavy since it requires costly singular value decomposition (SVD) steps (for computing the states sequences and for extracting the nonlinearity). Here, we would like to investigate a way of avoiding the SVD computation in this method for online application purposes. Although it is feasible to estimate a fully parameterized MIMO system by applying twice the LS-SVM regression procedure as in (Goethals et al., 2005a), only the case of SIMO system is considered in this paper for the sake of simplicity. Anyway, the system may be multi-input; the only requirement concerns the nonlinearity  $f(\cdot)$  which has to be a scalar-valued function.

We will first estimate the terms  $H_i f(\cdot)$  and then the extended observability matrix. In equation (4), we choose the integer  $d$  at least larger than the order of the system and sufficiently large so that the term  $CA^d x_{t-d}$  may be considered negligible with respect to the noise term. This approximation is all the more acceptable as the system has rapid dynamics or  $d$  is relatively large. Then, equation (4) becomes

$$y_t(s) \approx \sum_{i=0}^d H_i(s) f(u_{t-i}) + e_t(s) \quad (5)$$

for  $s = 1, \dots, n_y$ . In the case that the system does not present rapid dynamics it may be safer to use the observer-augmenting procedure described in (Juang et al., 1993) instead. This latter algorithm introduces an observer gain to set the poles of the resulting system as close to zero as possible so that the accuracy of the previous equation is noticeably improved even for small values of  $d$ .

Recall that neither the parameters  $H_i$  nor the nonlinear function  $f(\cdot)$  are known. The LS-SVM regression procedure developed by Suykens (Suykens et al., 2002) consists in modeling the nonlinearity as  $f(\cdot) = \omega^\top \varphi(\cdot) + \delta$  where  $\varphi : \mathbb{R}^{n_u} \rightarrow \mathbb{R}^{n_h}$  is a nonlinear mapping from the input space toward a possibly high dimensional feature space  $\mathbb{R}^{n_h}$ ,  $\delta$  is a bias compensator and  $\omega \in \mathbb{R}^{n_h}$ . However, direct application of this idea results here in a hard non-convex optimization problem (Goethals et al., 2005b). Fortunately, this difficulty can be circumvented by combin-

ing over-parameterization and componentwise LS-SVM (Pelckmans et al., 2004). Letting

$$g_{i,s}(\cdot) = H_i(s)f(\cdot) = \omega_{i,s}^\top \varphi(\cdot) + \delta_{i,s} \quad (6)$$

we get from (5)

$$\begin{aligned} y_t(s) &= \sum_{i=0}^d g_{i,s}(u_{t-i}) + e_t(s) = \\ &= \sum_{i=0}^d \omega_{i,s}^\top \varphi(u_{t-i}) + \delta_s + e_t(s) \end{aligned} \quad (7)$$

with  $\delta_s = \sum_{i=0}^d \delta_{i,s}$ ,  $s = 1, \dots, n_y$ ,  $t = d + 1, \dots, N$ . Since only input-output data  $(u_t, y_t)$  are available, we would like to compute the nonlinear components  $g_{i,s}$  from the sample values of their sum along  $i$ . That corresponds to the so-called component-wise LS-SVM problem. One can then notice that the solution of equation (7) is not unique. In fact, for any solution  $\{g_{i,s}(\cdot)\}_{i=0}^d$  of (7),  $\{g_{i,s}(\cdot) + \varepsilon_{i,s}(\cdot)\}_{i=0}^d$  is also solution for all arbitrary functions  $\{\varepsilon_{i,s}(\cdot)\}_{i=0}^d$ , satisfying  $\sum_{i=0}^d \varepsilon_{i,s}(u_{t-i}) = 0, \forall t = d + 1, \dots, N$ . Hence, each function  $g_{i,s}$  can only be estimated up to an additive undetermined function term. A typical way to overcome that difficulty is to introduce some additional constraint on the  $\omega_{i,s}^\top \varphi(\cdot)$  to be obtained by solving (7) so that the redundancy is removed. In (Goethals et al., 2005a), this constraint is chosen as a centering relation over the time window of the measurement points. An alternative way, which is quite more practical for recursive estimation, is used in the following to deal with this problem.

The LS-SVM constrained optimization problem is based on the cost function

$$\begin{aligned} \mathcal{J}(w_{i,s}, e_t(s)) &= \\ &= \frac{1}{2} \sum_{s=1}^{n_y} \sum_{i=0}^d \omega_{i,s}^\top \omega_{i,s} + \frac{\gamma}{2} \sum_{s=1}^{n_y} \sum_{t=d+1}^N e_t(s)^2 \end{aligned} \quad (8)$$

subject to (7), where  $\gamma$  is a regularization parameter and serves as a weight in the cost function to be minimized. In order to apply the Lagrange multipliers method, we formulate the associated Lagrangian as

$$\begin{aligned} \mathcal{L}(w_{i,s}, \delta_s, \alpha_{t,s}, e_t(s)) &= \\ &= \mathcal{J}(w_{i,s}, e_t(s)) - \sum_{s=1}^{n_y} \sum_{t=d+1}^N \alpha_{t,s} \left\{ \sum_{i=0}^d \omega_{i,s}^\top \varphi(u_{t-i}) \right. \\ &\quad \left. + \delta_s + e_t(s) - y_t(s) \right\} \end{aligned}$$

where the  $\alpha_{t,s}$  stand for the Lagrange multipliers. The associated Karush-Kuhn-Tucker optimality conditions are given by

$$\frac{\partial \mathcal{L}}{\partial \omega_{i,s}} = 0 \rightarrow \omega_{i,s} = \sum_{t=d+1}^N \alpha_{t,s} \varphi(u_{t-i}) \quad (9)$$

$$\forall i = 0, \dots, d \quad \forall s = 1, \dots, n_y$$

$$\frac{\partial \mathcal{L}}{\partial \delta_s} = 0 \rightarrow \sum_{t=d+1}^N \alpha_{t,s} = 0 \quad \forall s = 1, \dots, n_y \quad (10)$$

$$\frac{\partial \mathcal{L}}{\partial \alpha_{t,s}} = 0, \forall t = d + 1, \dots, N \quad \forall s = 1, \dots, n_y$$

$$\rightarrow y_t(s) = \sum_{i=0}^d \omega_{i,s}^\top \varphi(u_{t-i}) + \delta_s + e_t(s) \quad (11)$$

$$\frac{\partial \mathcal{L}}{\partial e_t(s)} = 0 \rightarrow e_t(s) = \gamma^{-1} \alpha_{t,s}, \quad (12)$$

$$\forall t = d + 1, \dots, N \quad \forall s = 1, \dots, n_y$$

Taking into account (9) and (12) in (11) gives

$$\begin{aligned} y_t(s) &= \sum_{i=0}^d \left( \sum_{q=d+1}^N \alpha_{q,s} \varphi(u_{q-i})^\top \varphi(u_{t-i}) \right) \\ &\quad + \delta_s + \gamma^{-1} \alpha_{t,s} \end{aligned} \quad (13)$$

The kernel function is defined from  $\mathbb{R}^{n_u} \times \mathbb{R}^{n_u}$  to  $\mathbb{R}$  as  $K(u_q, u_t) = \varphi(u_q)^\top \varphi(u_t)$  and selected here to be a gaussian Radial Basis Function  $K(u_q, u_t) = \exp(-\|u_q - u_t\|_2^2 / \sigma^2)$ , with  $\sigma$  a constant. Equations (10), (13) can be summarized in the following linear equation with as unknowns the dual parameters of the optimization problem

$$\begin{bmatrix} 0 & \mathbf{1}_M^\top \\ \mathbf{1}_M & \Omega + \gamma^{-1} I_M \end{bmatrix} \begin{bmatrix} \delta_s \\ \bar{\alpha}_s \end{bmatrix} = \begin{bmatrix} 0 \\ \bar{\mathcal{Y}}_s \end{bmatrix} \quad \forall s = 1, \dots, n_y \quad (14)$$

where  $M = N - d$ ,  $I_M$  is the identity matrix of order  $M$  and  $\Omega = (\Omega_{ij}) \in \mathbb{R}^{M \times M}$ ,  $\mathbf{1}_M$ ,  $\bar{\alpha}_s$  and  $\bar{\mathcal{Y}}_s \in \mathbb{R}^M$  are defined as

$$\begin{aligned} \Omega_{ij} &= \sum_{m=0}^d K(u_{i+d-m}, u_{j+d-m}), \\ \mathbf{1}_M &= [1 \dots 1]^\top, \\ \bar{\alpha}_s &= [\alpha_{d+1,s} \dots \alpha_{N,s}]^\top, \\ \bar{\mathcal{Y}}_s &= [y_{d+1}(s) \dots y_N(s)]^\top. \end{aligned}$$

The solution of equation (14) is obtained as

$$\begin{aligned} \delta_s &= \mathbf{1}_M^\top \mathcal{T}^{-1} \bar{\mathcal{Y}}_s / (\mathbf{1}_M^\top \mathcal{T}^{-1} \mathbf{1}_M) \\ \bar{\alpha}_s &= \mathcal{T}^{-1} (\bar{\mathcal{Y}}_s - \mathbf{1}_M \delta_s), \quad s = 1, \dots, n_y \end{aligned} \quad (15)$$

where  $\mathcal{T} = \Omega + \gamma^{-1} I_M$ . Note that matrix  $\mathcal{T}$  is the same for all the outputs elements  $s$ . It depends only on the input  $u$  through the kernel function  $K$ . Obviously, the properties of that matrix (e.g., its conditioning) are directly related to those of the input.

Once the dual parameters  $\alpha_{t,s}$  and  $\delta_s$  are known, the searched nonlinear components from (6) can be expressed as:

$$\hat{\omega}_{i,s}^\top \varphi(\cdot) = \sum_{t=d+1}^N \alpha_{t,s} K(u_{t-i}, \cdot) = \bar{\alpha}_s^\top \bar{k}_i(\cdot), \quad (16)$$

$s = 1, \dots, n_y$  with

$$\bar{k}_i(\cdot) = [K(u_{d+1-i}, \cdot) \dots K(u_{N-i}, \cdot)]^\top.$$

It should be noticed that, depending on the selected kernel, the LS-SVM solution to the nonlinear regression problem is very similar to that obtained using the standard techniques of expanding the nonlinearity on polynomials or RBF bases. The solution (16) is known to be nonsparse, that is, some of the terms  $\alpha_{t,s}K(u_{t-i}, \cdot)$  have a too small contribution in the general error reduction and can therefore be removed without a great impact. Some pruning techniques have been proposed to render it sparse (see, e.g., (de Kruijff and de Vries, 2003)) but our presentation does not address this problem.

It is known that the over-parameterization technique may result in an estimate for  $g_{i,s}$  which does not satisfy any longer the relation (6). That means that the estimates for the components  $g_{i,s}$  may not be collinear as they are expected to be. More precisely, the estimate may be corrupted by an unknown additive component  $\varepsilon_{i,s}(\cdot)$  which obeys to  $\sum_{i=0}^d \varepsilon_{i,s}(\cdot) = 0^1$ . Hence, we have

$$\hat{g}_{i,s} = \hat{\omega}_{i,s}^\top \varphi(\cdot) + \hat{\delta}_{i,s} = H_i(s)f(\cdot) + \varepsilon_{i,s}(\cdot).$$

Summing the above equation along the subscript  $i$  results in

$$\sum_{i=0}^d \hat{\omega}_{i,s}^\top \varphi(\cdot) + \hat{\delta}_s = \mu_s f(\cdot)$$

where  $\mu_s = \sum_i H_i(s)$  is supposed to be nonzero. Note that for any nonzero scalar  $\lambda$ ,  $H_i(s)f(\cdot) = (H_i(s)\lambda) \left(\frac{1}{\lambda}f(\cdot)\right)$ . In other words  $f(\cdot)$  can only be estimated up to a scalar factor just as the matrices  $(A, B, C, D)$  can be determined only up to a similar transformation. In the light of this remark, the static nonlinearity  $f$  can arbitrarily set to be  $\hat{f}(\cdot) \leftarrow \mu_s f(\cdot)$

$$\hat{f}(\cdot) = \sum_{i=0}^d \hat{\omega}_{i,s}^\top \varphi(\cdot) + \hat{\delta}_s, \quad s = 1, \dots, n_y \quad (17)$$

Since the nonlinearity  $f(\cdot)$  does not depend on a particular output, it is not worth repeating this estimation for every output as that may increase the computational cost. The nonlinear function  $f(\cdot)$  can be extracted from the measurements of the input and the ones of a single arbitrary output  $s$  or a combination of all the outputs.

From now on, given that the nonlinearity is known, the system matrices can be computed using any subspace identification method. The new linear system to be considered is that of input  $f(u)$  and of output the whole output vector  $y$  (instead of one component as above). We propose to

<sup>1</sup> Indeed,  $\bar{\varepsilon}(\zeta_t) = \varepsilon_0(\zeta_t^1) + \dots + \varepsilon_d(\zeta_t^{d+1}) = 0$ ,  $\{\zeta_t\}$  being an i.i.d. vector sequence. We assume then that  $\varepsilon_0(u) + \dots + \varepsilon_d(u) = 0$  for any scalar  $u$

estimate first the Markov parameters  $H_i$  from the FIR model (4) using least squares regression and then compute the system matrices. The regression equation is:

$$y_t = [H_0 \dots H_d] \begin{bmatrix} f(u_t) \\ \vdots \\ f(u_{t-d}) \end{bmatrix} + e_t, \quad (18)$$

$$t = d+1, \dots, N$$

When this step is completed, all the terms  $H_i$  i.e.,  $D, CB, CAB, \dots, CA^{d-1}B$  and  $f$  are known. From these parameters the system matrices can be directly computed using an SVD but we are interested here in developing a recursive scheme which avoids this SVD process (which is known to be computationally very demanding).

#### 4. RECURSIVE ALGORITHM

In the following, a recursive version of the identification algorithm presented in the previous section will be derived. While most recursive identification algorithms are developed by using a forgetting factor, in the present case we opt for a sliding window approach. For, assume that at a given instant  $N$  observations  $\{u_t, y_t\}_{t=k}^{k+N-1}$  are available. When a new pair of data is acquired, the window is slid to be  $\{u_t, y_t\}_{t=k+1}^{k+N}$ , that is, the oldest data is removed whenever a new sample becomes available. Then, matrix  $\mathcal{T}$  changes in such a way that its dimensions remain fixed. A similar approach has been used in (Liu et al., 2003) to implement an online LS-SVM based classifier.

We assume that the nonlinearity  $f(\cdot)$  may undergo some changes as well in its parameters as in its form during time. It is then necessary to update its estimate. At times  $k$  and  $k+1$  the matrix  $\mathcal{T}$  in (15) is formally given by

$$\mathcal{T}(k) = \begin{bmatrix} \eta_k & \underline{\varphi}_k^\top \\ \underline{\varphi}_k & \Delta_k \end{bmatrix}, \quad \mathcal{T}(k+1) = \begin{bmatrix} \Delta_k & \underline{\theta}_{k+1} \\ \underline{\theta}_{k+1} & \nu_{k+1} \end{bmatrix}$$

where the underlined elements indicate column vectors. Making use of the matrix block inversion identity we get:

$$\mathcal{T}(k)^{-1} = g_k^{-1} \begin{bmatrix} 1 & -\underline{\varphi}_k^\top \Delta_k^{-1} \\ -\Delta_k^{-1} \underline{\varphi}_k & (g_k I + \Delta_k^{-1} \underline{\varphi}_k \underline{\varphi}_k^\top) \Delta_k^{-1} \end{bmatrix} \quad (19)$$

with  $g_k = \eta_k - \underline{\varphi}_k^\top \Delta_k^{-1} \underline{\varphi}_k \in \mathbb{R}$ . Assume now that  $\mathcal{T}(k)^{-1}$  is known. Then, if we partition it as

$$\mathcal{T}(k)^{-1} = \begin{bmatrix} a_k & \underline{q}_k^\top \\ \underline{q}_k & Q_k \end{bmatrix}$$

$\Delta_k^{-1}$  can be deduced with respect to (19) as follows

$$\Delta_k^{-1} = Q_k - \frac{1}{a_k} \underline{q}_k \underline{q}_k^\top.$$

After that  $\mathcal{T}(k+1)^{-1}$  may be computed as:

$$\mathcal{T}(k+1)^{-1} = h_{k+1}^{-1} \begin{bmatrix} (h_{k+1}I + \Delta_k^{-1}\underline{\theta}_{k+1}\underline{\theta}_{k+1}^\top)\Delta_k^{-1} & -\Delta_k^{-1}\underline{\theta}_{k+1} \\ -\underline{\theta}_{k+1}^\top\Delta_k^{-1} & 1 \end{bmatrix} \quad (20)$$

with  $h_{k+1} = v_{k+1} - \underline{\theta}_{k+1}^\top \Delta_k^{-1} \underline{\theta}_{k+1} \in \mathbb{R}$

The dual parameters  $\bar{\alpha}_s$  and  $\delta_s$  follow immediately from  $\mathcal{T}(k+1)^{-1}$  and then, the nonlinear function  $f(\cdot)$ . Using the obtained estimate of the nonlinearity, the parameters  $H_i$  are computed in a recursive least squares fashion from equation (18).

In order to recursively update the estimates of the system matrices, results from the field of recursive subspace identification (see, e.g., (Lovera et al., 2000; Mercère et al., 2003) and the references therein) will be exploited. In particular, the so-called Propagator algorithm for system identification will be adopted (Mercère et al., 2003), which is briefly summarized in the following. Consider the data vectors

$$\begin{aligned} y_{\bar{n}}(t) &= [y(t)^\top \cdots y(t + \bar{n} - 1)^\top]^\top \\ u_{\bar{n}}(t) &= [u(t)^\top \cdots u(t + \bar{n} - 1)^\top]^\top, \end{aligned}$$

where  $\bar{n}$  is chosen such that  $n_x \leq \bar{n} \leq d$ . From (1), it is possible to derive this following relation

$$z_{\bar{n}}(t) = \Gamma_{\bar{n}} x(t) + n_{\bar{n}}(t) \quad (21)$$

where  $z_{\bar{n}} = y_{\bar{n}}(t) - H_{\bar{n}} f(u_{\bar{n}}(t))$ ,  $\Gamma_{\bar{n}}$  is the extended observability matrix,  $H_{\bar{n}}$  is the block Toeplitz matrix of the impulse responses from  $u$  to  $y$  and  $n_{\bar{n}}$  a sum of all the noise terms.

$$H_{\bar{n}} = \begin{bmatrix} D & 0 & \cdots & 0 \\ CB & D & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ CA^{\bar{n}-2}B & \cdots & CB & D \end{bmatrix}$$

$$\Gamma_{\bar{n}} = [(C)^\top \ (CA)^\top \ \cdots \ (CA^{\bar{n}-1})^\top]^\top$$

The so-called observation vector is explicitly computed as  $z_{\bar{n}}(t) = y_{\bar{n}}(t) - \hat{H}_{\bar{n}} \hat{f}(u_{\bar{n}}(t))$  thanks to the estimates of  $H_{\bar{n}}$  and  $f(\cdot)$  obtained from the previous subsection. Under the assumption that the system in (1) is observable,  $\Gamma_{\bar{n}}$  has at least  $n_x$  linearly independent rows which can be gathered in a submatrix  $\Gamma_1$ . Then, based on rows permutation,  $\Gamma_{\bar{n}}$  can be partitioned as

$$\Gamma_{\bar{n}} = \begin{bmatrix} \Gamma_1 \\ P_{\bar{n}} \Gamma_1 \end{bmatrix} = \begin{bmatrix} I_{n_x} \\ P_{\bar{n}} \end{bmatrix} \Gamma_1 \quad (22)$$

where the matrix  $P_{\bar{n}}$  is the so-called propagator (Mercère et al., 2003). Thus, since  $\Gamma_1$  is nonsingular, a basis of the observability matrix can be estimated by estimating the propagator. To that purpose the following criterion is considered

$$\mathcal{J}(P_{\bar{n}}) = \|R_{z_2\eta} - P_{\bar{n}} R_{z_1\eta}\|_F^2 \quad (23)$$

where  $\eta$  is an instrumental variable supposed to be asymptotically uncorrelated with the noise but sufficiently correlated with the state vector,

$R_{z_2\eta} = E(z_2\eta^\top)$ ,  $R_{z_1\eta} = E(z_1\eta^\top)$  are cross-correlation matrices.  $[z_1^\top \ z_2^\top]^\top$  is a partition of  $z_{\bar{n}}$  consistent with the one for  $\Gamma_{\bar{n}}$ . The propagator is then obtained as  $P_{\bar{n}}(t) = R_{z_2\eta}(t)R_{z_1\eta}^{-1}(t)$  (see (Mercère et al., 2003) for details in the recursive estimation of  $P_{\bar{n}}(t)$ ) and the system matrices are finally recovered from

$$\begin{aligned} C &= \Gamma_{\bar{n}}(1:l) \\ A &= \Gamma_{\bar{n}}(1:l(\bar{n}-1),:)^\dagger \Gamma_{\bar{n}}(l+1:l\bar{n},:) \\ D &= H_0, \quad B = \Gamma_{\bar{n}}(1:l(\bar{n}-1),:)^\dagger \begin{bmatrix} H_1 \\ \vdots \\ H_{\bar{n}-1} \end{bmatrix} \end{aligned}$$

where  $\dagger$  denotes the Moore-Penrose pseudo-inverse.

Finally, the complete algorithm can be summarised as follows. Whenever a new pair  $(u_t, y_t)$  of data is measured:

- (1) Compute the nonlinear function by computing the parameters  $\bar{\alpha}$  and  $\delta$ . That is realized in a recursive way by using (20) and (15)
- (2) Update the Markov parameters from (18) using recursive least squares
- (3) Update the extended observability matrix  $\Gamma_{\bar{n}}$  by adapting recursively the propagator
- (4) Compute the system matrices  $A, B, C, D$

## 5. SIMULATION RESULTS

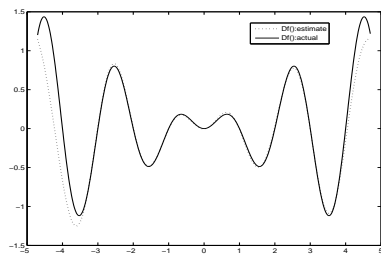
To evaluate the proposed algorithm the following SISO example is considered:

$$\begin{aligned} x_{t+1} &= \begin{bmatrix} -0.15 & 0.325 \\ 1.0 & 0 \end{bmatrix} x_t + \begin{bmatrix} 2 \\ 0 \end{bmatrix} f(u_t) \\ y_t &= [0.1750 \ 1.1625] x_t + f(u_t) + v_t \end{aligned} \quad (24)$$

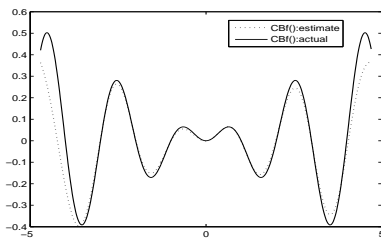
The static nonlinearity is chosen as  $f(u) = \text{sinc}(u)u^2$  and the excitation as a white gaussian noise sequence  $u(t) \sim N(0, 1)$ . The output noise  $v_t$  also white and gaussian is chosen such that the signal to noise ratio is 30 dB. The algorithm parameters are chosen as  $\sigma = 1, d = 10, \gamma^{-1} = 0.003, \bar{n} = 6$ . The simulation is driven with  $N = 300$  over a time interval of 3000 points. Figure 1 illustrates the products between the nonlinearity and the two first Markov parameters on the interval  $[-5, 5]$ . The actual functions and their respective estimates are almost undistinguishable, at least on the density interval of the input sequence. Figure 2 shows the poles of the system estimated recursively. Finally, Figure 3 shows the estimated poles of a time varying system derived from (24) by carrying out a slow and regular variation in the dynamics matrix as follows

$$A(t) = \left(1 - 10^{-2}\sqrt{t}\right) \begin{bmatrix} -0.15 & 0.325 \\ 1.0 & 0 \end{bmatrix} \quad (25)$$

The results presented show a quite satisfactory tracking of the system parameters.



(a)



(b)

Fig. 1. (a): true (solid line) and estimated (dotted line) Markov parameter  $Df()$ . (b): true (solid line) and estimated (dotted line) Markov parameter  $CBf()$ .

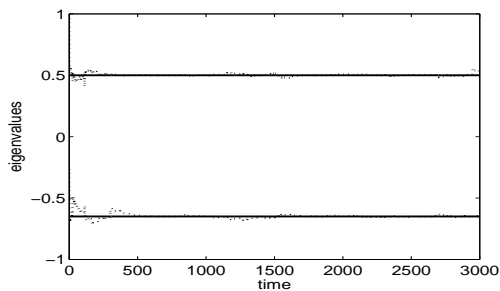


Fig. 2. Evolution of the eigenvalues of the estimated model (dotted) and of the actual system (24) (solid).

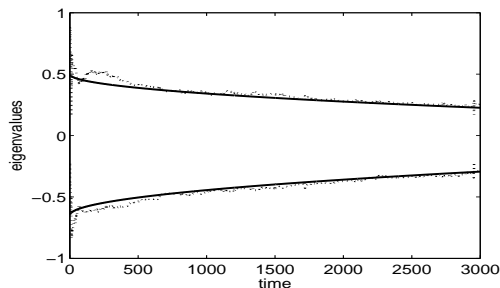


Fig. 3. Recursive estimates of the eigenvalues for the slowly time-varying system (25).

## 6. CONCLUDING REMARKS

A recursive algorithm combining LS-SVM regression and subspace identification techniques has been presented for the identification of SIMO Hammerstein systems. The whole algorithm works in a recursive fashion and is designed for slowly time-varying systems identification. Numerical simulations show quite good results. Future work will focus on the recursive identification of more general nonlinear state space models.

- E.-W. Bai. An optimal two-stage identification algorithm for Hammerstein-Wiener nonlinear systems. *Automatica*, 34(3):333–338, 1998.
- B. de Kruif and T. de Vries. Pruning error minimization in least squares support vector machines. *IEEE Transactions on Neural Networks*, 14(3):696–702, 2003.
- E. Dempsey and D. Westwick. Identification of Hammerstein models with cubic spline nonlinearities. *IEEE Transactions on Biomedical Engineering*, 51(2):237–245, 2004.
- I. Goethals, K. Pelckmans, J. Suykens, and B. De Moor. Subspace identification of Hammerstein systems using least squares support vector machines. *IEEE Transactions on Automatic Control*, 50(10):1509–1519, 2005a.
- I. Goethals, K. Pelckmans, J. Suykens, and B. De Moor. Identification of mimo Hammerstein models using least squares support vector machines. *Automatica*, 41(7):1263–1272, 2005b.
- J. N. Juang, M. Phan, L. G. Horta, and R. W. Longman. Identification of observer/Kalman filter Markov parameters : theory and experiments. *Journal of Guidance Control and Dynamics*, 16(2):320–329, 1993.
- J. Liu, J. Chen, S.n Jiang, and J. Cheng. Online LS-SVM for function estimation and classification. *Journal of University of Science and Technology Beijing*, 10(5):73–77, 2003.
- M. Lovera, T. Gustafsson, and M. Verhaegen. Recursive subspace identification of linear and non-linear Wiener state-space models. *Automatica*, 36(11):1639–1650, 2000.
- G. Mercère, S. Lecoeuche, and C. Vasseur. A new recursive method for subspace identification of noisy systems: EIVPM. In *13th IFAC Symposium on System Identification*, 2003.
- H. Oku and H. Kimura. Recursive 4SID algorithms using gradient type subspace tracking. *Automatica*, 38(6):1035–1043, 2002.
- K. Pelckmans, I. Goethals, J. De Brabanter, J.A.K. Suykens, and B. De Moor. Componentwise least squares support vector machines. Technical report, 04-75, ESAT-SISTA, K.U.Leuven, Belgium, 2004.
- J. Ramos and J.-F. Durand. Identification of nonlinear systems using a b-splines parametric subspace approach. In *American Control Conference, San Diego, California*, 1999.
- J. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle. *Least Squares Support Vector Machines*. World Scientific, 2002.
- Jozef Vörös. Modeling and identification of wiener systems with two-segment nonlinearities. *IEEE Transactions on Control Systems Technology*, 11, no. 2:253–257, 2003.