NARX IDENTIFICATION OF HAMMERSTEIN MODELS USING LEAST SQUARES SUPPORT VECTOR MACHINES

Ivan Goethals [*,1] **Kristiaan Pelckmans** [*,3]
Johan A.K. Suykens [*,2] **Bart De Moor** [*,4]

[*] *K.U.Leuven, ESAT-SCD-SISTA*
Kasteelpark Arenberg 10, B-3001 Leuven-Heverlee, Belgium
email: ivan.goethals@esat.kuleuven.ac.be

Abstract: In this paper we propose a new technique for the identification of NARX Hammerstein systems. The new technique is based on the theory of Least Squares Support Vector Machines function-approximation and allows to determine the memoryless static nonlinearity as well as the linear model parameters. As the technique is non-parametric by nature, no assumptions about the static nonlinearity need to be made.

Keywords: Hammerstein models, NARX, LS-SVM

1. INTRODUCTION

Linear identification techniques have become a standard tool in many engineering areas such as the chemical process industry and the study of vibrating systems. Throughout the last few decades, the field of linear modelling has been explored to the level that most linear identification problems can be solved with fairly standard and well known tools.

The estimation of nonlinear systems is in general far more complex than that of linear models, and it is mostly due to the advent of more powerfull computers and theoretical breakthroughs in the area of splines, neural networks (Bishop, 1995), regularization networks (Poggio and Girosi, 1990) and support vector machines (SVMs (Vapnik, 1998)) that the area of nonlinear system identification has received ever increasing attention over the last few year.

Although estimating a complex nonlinear system is sometimes necessary, in many situations Hammerstein models may be a good approximation for nonlinear plants. Hammerstein models are composed of a memoryless static nonlinearity followed by a linear dynamical system. Many techniques have been proposed for the estimation of Hammerstein models. Based on an approximation of the nonlinear function by a finite order polynomial (the so-called parametric methods (Billings and Fakhouri, 1979; Lang, 1993)), the approximation of the nonlinear function by a finite sum of orthogonal basis functions (the non-parameteric methods (Greblicki and Pawlak, 1991; Pawlak, 1991)), or the assumption that the model is almost linear so that the static nonlinearity can be treated as a nonlinear distortion (Schoukens *et al.*, 2003).

In this paper we propose a nonlinear identification method based on Least Squares Support Vector Machines. Support Vector Machines (Vapnik, 1998) and related methods (see e.g. (Hastie *et*

[1] I. Goethals is a Research Assistant with the Fund for Scientific Research-Flanders (FWO-Vlaanderen).
[2] J. Suykens is an Associate Professor with KULeuven.
[3] K. Pelckmans is a Research Assistent with KULeuven.
[4] B. De Moor is a Full Professor with KULeuven.

al., 2001)) are a powerful methodology for solving problems in nonlinear classification, function estimation and density estimation. SVMs and related methods have been introduced within the context of statistical learning theory and structural risk minimization. In the methods one solves convex optimization problems, typically quadratic programs. Least Squares Support Vector Machines (LS-SVMs) (Suykens *et al.*, 2002) are reformulations of standard SVMs which lead to solving linear KKT systems for classification tasks as well as regression estimation. As a result, methods based on Least Squares Support Vector Machines are fast and can also be robustified. LS-SVM's have been proposed as a class of kernel machines with primal-dual formulations to static nonlinear regression, kernel FDA,PCA,CCA,PLS, recurrent networks and control. The Hammerstein identification method proposed in this paper will be shown to result in good estimates for the static nonlinearity, as well as the linear model parameters.

The outline of this paper is as follows: in Section 2, the general concepts of the theory of Least Squares Support Vector Machines, applied to static function estimation are reviewed. In Section 3 and 4 a new method for the identification of nonlinear Hammerstein systems is proposed. In Section 5, the new method is tested by means of an example.

2. LEAST SQUARES SUPPORT VECTOR MACHINES FOR FUNCTION APPROXIMATION

Let $\{x_k, y_k\}_{i=1}^N \subset \mathbb{R}^d \times \mathbb{R}$ be the inputs and outputs of the training data. Consider the regression model $y_i = f(x_i) + e_i$ where $x_1, \ldots, x_N$ are deterministic points (fixed design), $f : \mathbb{R}^d \to \mathbb{R}$ is an unknown real-valued smooth function and $e_1, \ldots, e_N$ are uncorrelated random errors with $E[e_i] = 0$, $E\left[e_i^2\right] = \sigma_e^2 < \infty$. In recent years, Least Squares Support Vector Machines (LS-SVMs) have been used for the purpose of estimating the nonlinear f. The following model is assumed:

$$f(x) = w^T \varphi(x) + b, \tag{1}$$

where $\varphi(x) : \mathbb{R}^d \to \mathbb{R}^{n_f}$ denotes the potentially infinite ($n_f = \infty$) dimensional feature map. The regularized least squares cost function is given as

$$\min_{w,b,e} \mathcal{J}(w,e) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{k=1}^{n} e_k^2, \tag{2}$$

$$\text{s.t. } y_k = w^T \varphi(x_k) + b + e_k, k = 1, \ldots, N. \tag{3}$$

The relative importance between the smoothness of the solution and the data fitting is governed by the scalar γ referred to as the regularization constant. The optimization performed is known as a ridge regression (Golub and Van Loan, 1989). In order to solve the constrained optimization problem, a Lagrangian is constructed:

$$\mathcal{L}(w,b,e;\alpha) = \mathcal{J}(w,e) - \sum_{k=1}^{N} \alpha_k \{w^T \varphi(x_k) + b + e_k - y_k\}, \tag{4}$$

with α_k the Lagrange multipliers. The conditions for optimality are given by:

$$\frac{\partial \mathcal{L}}{\partial w} = 0 \to w = \sum_{k=1}^{N} \alpha_k \varphi(x_k), \tag{5}$$

$$\frac{\partial \mathcal{L}}{\partial b} = 0 \to \sum_{k=1}^{N} \alpha_k = 0, \tag{6}$$

$$\frac{\partial \mathcal{L}}{\partial e_k} = 0 \to \alpha_k = \gamma e_k, k = 1, \ldots, N, \tag{7}$$

$$\frac{\partial \mathcal{L}}{\partial \alpha_k} = 0 \to \text{Eq. (3)}, k = 1, \ldots, N, \tag{8}$$

wich can be summarized in matrix notation

$$\left[\begin{array}{c|c} 0 & 1_N{}^T \\ \hline 1_N & \Omega + \gamma^{-1} I_N \end{array}\right] \left[\begin{array}{c} b \\ \hline \alpha \end{array}\right] = \left[\begin{array}{c} 0 \\ \hline y \end{array}\right], \tag{9}$$

where $y = \begin{bmatrix} y_1 & \ldots & y_N \end{bmatrix}^T$, $1_N = \begin{bmatrix} 1 & \ldots & 1 \end{bmatrix}^T$, $\alpha = \begin{bmatrix} \alpha_1 & \ldots & \alpha_N \end{bmatrix}$, $\Omega_{kl} = \varphi(x_k)^T \varphi(x_l) = K(x_k, x_l)$ with application of the so-called kernel trick (positive definite kernel K). For the choice of the kernel $K(\cdot,\cdot)$, see e.g. (Schoelkopf and Smola, 2002). Typical examples are the use of a polynomial kernel $K(x_i, x_j) = (\tau + x_i^T x_j)^d$ of degree d or the RBF kernel $K(x_i, x_j) = \exp(-\|x_i - x_j\|_2^2/\sigma)$ where σ denotes the bandwidth of the kernel. The resulting LS-SVM model for function estimation can be evaluated in a new point x_* as

$$\hat{f}(x_*;\theta) = \sum_{k=1}^{N} \alpha_k K(x_*, x_k) + b, \tag{10}$$

where $\theta = (b, \alpha)$ is the solution to (9).

3. IDENTIFICATION OF NONLINEAR ARX HAMMERSTEIN MODELS

Hammerstein systems, in their most basic form, consist out of a static memoryless nonlinearity, followed by a linear dynamical system, as shown in Figure 1. The aim of Hammerstein modelling is to find estimates for the nonlinearity and the model parameters of the linear system from input/output measurements.

In this paper, we will restrict ourselves to SISO systems (single input - single output) although much of the reasoning can be transferred to the

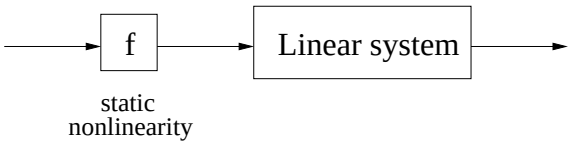

Fig. 1. A Hammerstein system consists of a memoryless static nonlinearity f followed by a linear dynamical system.

MIMO case. For the linear dynamical part, we will assume a model structure of the form:

$$y_t = \sum_{i=1}^{n} a_i y_{t-i} + \sum_{j=1}^{m} b_j u_{t-j} + e_t, \qquad (11)$$

with $u_t, y_t \in \mathbb{R}, t \in \mathbb{N}$ $\{u_t, y_t\}$ a set of input and output measurements and e_t the so-called equation error which will be assumed to be white and gaussian. The model structure (11) is generally known as the "AutoRegressive model with eXogeneous input" (ARX) and is one of the best known model structures in linear identification. Adding a static nonlinearity $f : \mathbb{R} \to \mathbb{R} : x \to f(x)$ to (11) leads to:

$$y_t = \sum_{i=1}^{n} a_i y_{t-i} + \sum_{j=1}^{m} b_j f\left(u_{t-j}\right) + e_t, \qquad (12)$$

which is the general model structure that will be assumed in this paper.

Applying LS-SVM function estimation outlined in the former section, we assume the following structure for the static nonlinearity f:

$$f(u) = w^T \varphi(u) + b_0. \qquad (13)$$

with

$$\Omega_{kl} = K(u_k, u_l) = \varphi(u_k)^T \varphi(u_l). \qquad (14)$$

a kernel of choice. Hence, equation (12) can be rewritten as follows:

$$y_t = \sum_{i=1}^{n} a_i y_{t-i} + \sum_{j=1}^{m} b_j \left(w^T \varphi\left(u_{t-j}\right) + b_0\right) + e_t. \qquad (15)$$

We focus on finding estimates for the linear parameters $a_i, i = 1, \ldots, n$ and $b_j, j = 1, \ldots, m$ and the static nonlinearity f. The Lagrangian of the resulting estimation problem is given by

$$\mathcal{L}(w, b, e, a; \alpha) = \mathcal{J}(w, e) - \sum_{t=r}^{N} \alpha_t \{ \sum_{i=1}^{n} a_i y_{t-i} + \sum_{j=1}^{m} b_j \left(w^T \varphi\left(u_{t-j}\right) + b_0\right)\big) + e_t - y_t \} \qquad (16)$$

with $r = \max(m, n) + 1$. The conditions for optimality are as follows:

$$\frac{\partial \mathcal{L}}{\partial w} = 0 \to w = \sum_{t=r}^{N} \sum_{j=1}^{m} \alpha_t b_j \varphi(u_{t-j}), \qquad (17)$$

$$\frac{\partial \mathcal{L}}{\partial b} = 0 \to \sum_{t=r}^{N} \sum_{j=1}^{m} \alpha_t b_j = 0, \qquad (18)$$

$$\frac{\partial \mathcal{L}}{\partial a_i} = 0 \to \sum_{t=r}^{N} \alpha_t y_{t-i} = 0, i = 1, \ldots, n, \qquad (19)$$

$$\frac{\partial \mathcal{L}}{\partial b_i} = 0 \to \sum_{t=r}^{N} \alpha_t \left(w^T \varphi(u_{t-i}) + b_0\right) = 0, \qquad (20)$$
$$i = 1, \ldots, m,$$

$$\frac{\partial \mathcal{L}}{\partial e_t} = 0 \to \alpha_t = \gamma e_t, t = r, \ldots, N, \qquad (21)$$

$$\frac{\partial \mathcal{L}}{\partial \alpha_t} = 0 \to \sum_{i=1}^{n} a_i y_{k-i} + e_t - y_t + \sum_{j=1}^{m} b_j \left(w^T \varphi\left(u_{t-j}\right) + b_0\right) = 0, t = r, \ldots, N, \qquad (22)$$

Substituting (17) and (21) in (22) leads to:

$$\sum_{j=1}^{m} \sum_{q=r}^{N} \sum_{p=1}^{m} b_j \left(b_p \alpha_q \varphi\left(u_{q-p}\right)^T \varphi\left(u_{t-j}\right) + b_0\right) + \sum_{i=1}^{n} a_i y_{t-i} + e_t - y_t = 0, t = r, \ldots, N, \qquad (23)$$

If the b_j values were known, the resulting problem would be linear in the unknowns and easy to solve through:

$$\left[\begin{array}{c|c|c} 0 & 0 & \tilde{b} 1_{N-r+1}^T \\ \hline 0 & 0 & Y \\ \hline \tilde{b} 1_{N-r+1} & Y^T & \overline{K} + \gamma^{-1} I \end{array}\right] \left[\begin{array}{c} b_0 \\ \hline a \\ \hline \alpha \end{array}\right] = \left[\begin{array}{c} 0 \\ \hline 0 \\ \hline y_f \end{array}\right],$$

with

$$\beta = \left[\beta_1 \ \ldots \ \beta_m\right]^T, \qquad (24)$$

$$\alpha = \left[\alpha_r \ \ldots \ \alpha_N\right]^T, \qquad (25)$$

$$\tilde{b} = \sum_{j=1}^{m} b_j, \qquad (26)$$

$$a = \left[a_1 \ \ldots a_n\right]^T, \qquad (27)$$

$$Y = \begin{bmatrix} y_{r-1} & y_r & \cdots & y_{N-1} \\ y_{r-2} & y_{r-1} & \cdots & y_{N-2} \\ \vdots & \vdots & & \vdots \\ y_{r-n} & y_{r-n+1} & \cdots & y_{N-n} \end{bmatrix}, \qquad (28)$$

$$\overline{K}_{p,q} = \sum_{j=1}^{m} b_j b_l \sum_{l=1}^{m} \Omega_{p+r-j, q+r-l}, \qquad (29)$$

$$y_f = \left[y_{r+1} \ \ldots \ y_N\right]^T. \qquad (30)$$

Since the b_j are in general not known and the solution to the resulting third order estimation problem (23) is by no means trivial, we will use an approximative method to obtain models of the form (12).

4. AN APPROXIMATIVE METHOD

To avoid having to solve the problem (23), we propose to rewrite (15) as follows:

$$y_t = \sum_{i=1}^{n} a_i y_{t-i} + \sum_{j=1}^{m} w_j^T \varphi(u_{t-j}) + d + e_t, \quad (31)$$

which, as will be shown shortly, can conveniently be solved using LS-SVM's. Note, however, that the resulting model class is wider than (15) due to the replacement of one w by several $w_j, j = 1, \ldots, m$. The model class (31) is therefore not necessarily limited to the description of Hammerstein systems. A sufficient condition for the estimated model to belong to this class of systems is for the obtained w_j to be collinear in which case w_j is seen as a replacement for $b_j w$. Taking this into account during the estimation leads to extra constraints requiring the angles between any pair $\{w_j, w_k\}, j, k, \in 1, \ldots, m$ to be zero, or $\left(w_j^T w_k\right)^2 = \sqrt{w_j^T w_j}\sqrt{w_k^T w_k}$. Alternatively, the collinearity constraint can be written as:

$$\operatorname{rank}\begin{bmatrix} w_1 & \ldots & w_m \end{bmatrix} = 1, \quad (32)$$

which is equivalent to ensuring that a set of $\frac{m(m-1)n_f(n_f-1)}{4}$ 2x2 determinants are zero. As n_h, the dimension of w is unknown and possibly very high, it is obvious that including such constraints in the Lagrangian would again lead to a non-trivial high-order estimation problem.

Considering that Hammerstein models are a subset of models of the form (31), however, if a good model is obtained for (31), it will automatically be close to the set of models of the form (15). Hence, in order to obtain a practical estimation algorithm, we propose to remove the collinearity constraints from the Lagrangian alltogether, solve the more general problem (31), and project the obtained model onto the model-set (15) later. Although this approach may seem ad-hoc at first, many existing Hammerstein identification schemes implement a similar technique, also known as the Bai's relaxation method (Bai, 1998). In essence, most existing approaches can be summarized as writing the static nonlinearity as a linear combination of general nonlinear basis-functions f_i, each with a certain weight c_i, eg.

$$f(u_t) = \begin{bmatrix} c_1 & \ldots & c_p \end{bmatrix} \begin{bmatrix} f_1(u_t) \\ \vdots \\ f_p(u_t) \end{bmatrix}, \quad (33)$$

where f_1, f_2, and f_p are chosen beforehand. This substitution leads to a classical linear identification algorithm where linear model parameters $p_1, p_2, \ldots$ are replaced by vectors $p_1 \begin{bmatrix} c_1 & \ldots & c_p \end{bmatrix}$. Afterwards, collinearity of these vectors is imposed, e.g. by applying an SVD, and the original model parameters $p_1, p_2, \ldots$ are recovered. Some examples of recently proposed methods employing this approach are found in (Pawlak, 1991; McKelvey and Hanner, 2003).

Taking all of the above into account, the optimization problem that is ultimately solved is the following (with $e = e_r, \ldots, e_N$):

$$\min_{w_j, a, d, e} \mathcal{J}(w_j, e) = \frac{1}{2}\sum_{j=1}^{m} w_j^T w_j + \gamma \frac{1}{2}\sum_{t=r}^{N} e_t^2, \quad (34)$$

subject to

$$\sum_{j=1}^{m} w_j^T \varphi(u_{t-j}) + \sum_{i=1}^{n} a_i y_{t-i} + d + e_t - y_t = 0, t = r, \ldots, N \quad (35)$$

$$\sum_{k=1}^{N} w_j^T \varphi(u_k) = 0, j = 1, \ldots, m \quad (36)$$

Note the extra constraints (36) to center the nonlinear functions $w_j^T \varphi(\cdot), j = 1, \ldots, m$ around their average over the training set. This to remove the uncertainty resulting from the fact that any set of constants can be added to the terms of an additive nonlinear function, as long as the sum of the constants is zero. Removing this uncertainty will facilitate the extraction of the parameters b_j in (12) later. Furthermore, this constraint enables us to give a clear meaning to the bias parameter d, namely $d = \sum_{j=1}^{m} b_j \left(\frac{1}{N}\sum_{k=1}^{N} f(u_k)\right)$.

The resulting Lagrangian is:

$$\mathcal{L}(w_j, d, a, e; \alpha, \beta) = \mathcal{J}(w_j, e) - \sum_{t=r}^{N} \alpha_t \{\sum_{i=1}^{n} a_i y_{t-i} + \sum_{j=1}^{m} w_j^T \varphi(u_{t-j}) + d + e_t - y_t\} + \sum_{j=1}^{m} \beta_j \{\sum_{k=1}^{N} w_j^T \varphi(u_k)\}. \quad (37)$$

The conditions for optimality are:

$$\frac{\partial \mathcal{L}}{\partial w_j} = 0 \rightarrow w_j = \sum_{t=r}^{N} \alpha_t \varphi(u_{t-j}) + \beta_j \sum_{k=1}^{N} \varphi(u_k), j = 1, \ldots, m, \quad (38)$$

$$\frac{\partial \mathcal{L}}{\partial a_i} = 0 \rightarrow \sum_{t=r}^{N} \alpha_t y_{t-i} = 0, i = 1, \ldots, n, \quad (39)$$

$$\frac{\partial \mathcal{L}}{\partial d} = 0 \rightarrow \sum_{t=r}^{N} \alpha_t = 0 \quad (40)$$

$$\frac{\partial \mathcal{L}}{\partial e_t} = 0 \rightarrow \alpha_t = \gamma e_t, t = r, \ldots, N, \quad (41)$$

$$\frac{\partial \mathcal{L}}{\partial \alpha_t} = 0 \rightarrow Eq.(35), \quad (42)$$

$$\frac{\partial \mathcal{L}}{\partial \beta_j} = 0 \rightarrow Eq.(36), \quad (43)$$

with solution:

$$\left[\begin{array}{c|c|c|c} 0 & 0 & 1^T & 0 \\ \hline 0 & 0 & Y & 0 \\ \hline 1 & Y^T & \tilde{K}+\gamma^{-1}I & K^0 \\ \hline 0 & 0 & {K^0}^T & (1^T\Omega 1)\cdot I_m \end{array}\right] \left[\begin{array}{c} d \\ \hline a \\ \hline \alpha \\ \hline \beta \end{array}\right] = \left[\begin{array}{c} 0 \\ \hline 0 \\ \hline y_f \\ \hline 0 \end{array}\right], \tag{44}$$

where

$$\tilde{K}_{p,q} = \sum_{j=1}^{m} \Omega_{p+r-j,q+r-j}, \tag{45}$$

$$K^0_{p,q} = \sum_{k=1}^{n} \Omega_{k,p-q}. \tag{46}$$

The projection of the obtained model onto (12) goes as follows; Estimates for the autoregressive parameters $a_i, i = 1, \ldots, n$ are directly obtained from (44). Furthermore, for the training input sequence $\begin{bmatrix} u_1 & \ldots & u_N \end{bmatrix}$, we have:

$$\begin{aligned} \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} & \begin{bmatrix} \underline{\hat{f}}(u_1) & \ldots & \underline{\hat{f}}(u_N) \end{bmatrix} \\ &= \begin{bmatrix} \alpha_N & \ldots & \alpha_r & & 0 \\ & \alpha_N & \ldots & \alpha_r & \\ & & \ddots & & \ddots \\ 0 & & & \alpha_N & \ldots \ \alpha_r \end{bmatrix} \\ &\times \begin{bmatrix} \Omega_{N-1,1} & \Omega_{N-1,2} & \ldots & \Omega_{N-1,N} \\ \Omega_{N-2,1} & \Omega_{N-2,2} & \ldots & \Omega_{N-2,N} \\ \vdots & \vdots & & \vdots \\ \Omega_{r-m,1} & \Omega_{r-m,2} & \ldots & \Omega_{r-m,N} \end{bmatrix} \\ &+ \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_m \end{bmatrix} \sum_{k=1}^{N} \begin{bmatrix} \Omega_{k,1} & \ldots & \Omega_{k,N} \end{bmatrix}, \end{aligned} \tag{47}$$

with $\underline{\hat{f}}(u)$ an estimate for

$$\underline{f}(u) = f(u) - \frac{1}{N}\sum_{k=1}^{N} f(u_k). \tag{48}$$

Hence, estimates for b_j and the static nonlinearity f can be obtained from a rank 1 decomposition of the right hand side of (47), for instance using a singular value decomposition. Once all the b_j are known, $\sum_{k=1}^{N} f(u_k)$ can be obtained as $\sum_{k=1}^{N} f(u_k) = \frac{Nd}{\sum_{j=1}^{m} b_j}$.

5. ILLUSTRATIVE EXAMPLES

The algorithm proposed in this paper was used for identification on the following Hammerstein system:

$$A(z)y = B(z)\text{sinc}(u) + e \tag{49}$$

with A and B polynomials in the forward shift operator z with $B(z) = z^5 + 0.8z^4 + 0.3z^3 + 0.4z^2$

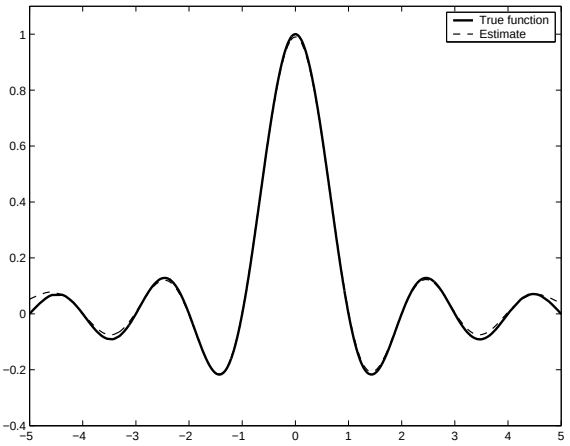

Fig. 2. True sinc function in the system $A(z)y = B(z)\text{sinc}(u)+e$ (solid line) and estimated sinc function for this system (dashed line).

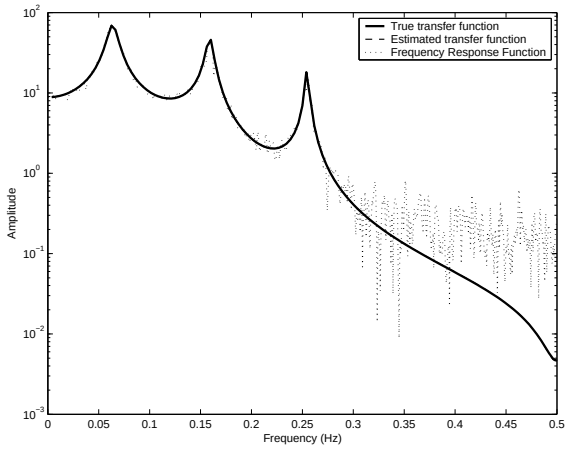

Fig. 3. True linear model (solid line), estimated linear model (dashed line), and frequency response function between y and sinc(u) (dotted line).

and $A(z) = (z - 0.98e^{\pm i})(z - 0.99e^{\pm 1.6i})(z - 0.97e^{\pm 0.4i})$ and e the so called equation noise. 2000 datapoints were generated using the system (49) and a white gaussian input sequence with zero mean and variance 2. The equation noise was chosen gaussian white with zero mean and as its standard deviation 10% of the standard deviation on the sequence sinc(u). Using the last 1000 datapoints for identification (so that all initial effects have died out), we obtain estimates for the sinc-function and the linear model. During the identification, an RBF kernel was used. The obtained estimate for the sinc function is displayed in Figure 2, together with the true sinc-function. As can be seen from the figure, the two are almost indistinguishable. The same holds for the linear model, of which the true and estimated transfer functions are displayed in Figure 3, together with the frequency response function between y and sinc(u). We conclude that for the given example a very close agreement between the true model and the estimated model was obtained.

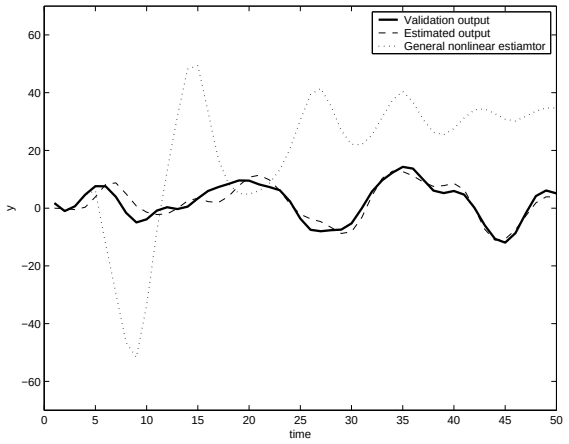

Fig. 4. Validation output (solid line), estimated output using NARX Hammerstein identification (dashed line), and the results from a general nonlinear identification (dotted line)

As a reference, the obtained model was tested by simulation on a noiseless validation set, together with a general nonlinear model of the form $\hat{y}_t = f(y_{t-1}, \ldots, y_{t-6}, u_{t-1}, \ldots, u_{t-4})$ obtained by a 10-fold cross-validation tuned LSSVM function estimation using an RBF kernel. As can be clearly seen from the first few samples of the simulated outputs, the output generated by the method proposed in this paper is in much closer agreement with the true output than that of the general nonlinear model (Figure 4). Once the initial conditions start to die out (right side of the plot), the proposed method follows the validation output very closely, while the general nonlinear model fails to even stay in the neighborhoud.

6. CONCLUSIONS

In this paper, we have proposed a new technique for the identification of Hammerstein ARX models. The new method is based on the theory of Least Squares Support Vector Machines function-approximation and allows to determine the memoryless static nonlinearity as well as the linear model parameters. The method was tested on an illustrative example to illustrate its performance, and compared to results from a general nonlinear identification procedure to point out the advantages of exploiting the structure of nonlinear systems in an identification procedure.

ACKNOWLEDGEMENTS

Our research is supported by: Research Council KUL: GOA-Mefisto 666, several PhD/postdoc & fellow grants; Flemish Government: FWO: PhD/postdoc grants, projects, G.0240.99 (multilinear algebra), G.0407.02 (support vector machines), G.0197.02 (power islands), G.0141.03 (Identification and cryptography), G.0491.03 (control for intensive care glycemia), G.0120.03 (QIT), G.0800.01 (collective intelligence), research communities (ICCoS, ANMMM); AWI: Bil. Int. Collaboration Hungary/ Poland; IWT: PhD Grants, Soft4s (softsensors), Belgian Federal Government: DWTC (IUAP IV-02 (1996-2001) and IUAP V-22 (2002-2006), PODO-II (CP/40: TMS and Sustainability); EU: CAGE; ERNSI; Eureka 2063-IMPACT; Eureka 2419-FliTE; Contract Research / agreements: Data4s, Electrabel, Elia, LMS, IPCOS, VIB.

REFERENCES

Bai, E.W. (1998). An optimal two-stage identification algorithm for hammerstein-wiener nonlinear systems. *Automatica* **3**, 333–338.

Billings, S.A. and S.Y. Fakhouri (1979). Nonlinear system identification using the Hammerstein model. *International journal of Systems Sciences* **10**, 567–578.

Bishop, C. (1995). *Neural Networks for Pattern Recognition*. Oxford University Press.

Golub, G.H. and C.F. Van Loan (1989). *Matrix Computations*. John Hopkins University Press.

Greblicki, W and M. Pawlak (1991). Nonparametric identification of a cascade nonlinear time series system. *Signal processing* **22**, 61–75.

Hastie, T., R. Tibshirani and J. Friedman (2001). *The Elements of Statistical Learning*. Springer-Verlag. Heidelberg.

Lang, Z.H. (1993). Controller design oriented model identification method for Hammerstein systems. *Automatica* **29**, 767–771.

McKelvey, T. and C. Hanner (2003). On identification of Hammerstein systems using excitation with a finite number of levels. *Proceedings of the* 13th *International Symposiun on System Identification (SYSID2003)* pp. 57–60.

Pawlak, M. (1991). On the series expansion approach to the identification of Hammerstein systems. *IEEE Transactions on Automatic Control* **36**, 736–767.

Poggio, T. and F. Girosi (1990). Networks for approximation and learning. *Proceedings of the IEEE* **78**(9), 1481–1497.

Schoelkopf, B and A. Smola (2002). *Learning with Kernels*. MIT Press.

Schoukens, J., J. G. Nemeth, P. Crama, Y. Rolain and R. Pintelon (2003). Fast approximate identification of nonlinear systems. *Automatica* **39**, 1267–1274.

Suykens, J.A.K., T. Van Gestel, De Brabanter J., B. De Moor and J. Vandewalle (2002). *Least Squares Support Vector Machines*. World Scientific, Singapore.

Vapnik, V.N. (1998). *Statistical Learning Theory*. Wiley and Sons.