Incorporating Best Linear Approximation within LS-SVM-Based Hammerstein System Identification

Ricardo Castro-Garcia, Koen Tiels, Johan Schoukens, Johan A. K. Suykens

***Abstract*— Hammerstein systems represent the coupling of a static nonlinearity and a linear time invariant (LTI) system. The identification problem of such systems has been a focus of research during a long time as it is not a trivial task. In this paper a methodology for identifying Hammerstein systems is proposed. To achieve this, a combination of two powerful techniques is used, namely, we combine Least Squares Support Vector Machines (LS-SVM) and the Best Linear Approximation (BLA). First, an approximation to the LTI block is obtained through the BLA method. Then, the estimated coefficients of the transfer function from the LTI block are included in a LS-SVM formulation for modeling the system. The results indicate that a good estimation of the underlying nonlinear system can be obtained up to a scaling factor.**

I. INTRODUCTION

As an extension to the field of linear modeling, nonlinear block structured models have been introduced [1]. It is the case that nonlinear models, even simple ones, often result in better approximations to process dynamics than the linear ones. The Hammerstein system [2] is often used. It consists of a static part $f(\cdot)$ containing the nonlinearity followed by a linear part $G_0(q)$ representing the dynamics of the process (see Fig. 1). In this paper, the q-notation, which is frequently used in system identification literature and software, will be used. The operator q is a time shift operator of the form $q^{-1}x(t) = x(t-1)$.

Several identification methods for Hammerstein systems have been reported in the literature. An overview of previous works can be found in [3] and different classifications of these methods can be found in [4], [5] and [6].

The objective of this paper is to incorporate the techniques of the Best Linear Approximation (BLA) [7] within Least Squares Support Vector Machines (LS-SVM) [8]. In the method it is possible to clearly separate the steps for identification of the linear and nonlinear parts.

The research leading to these results has received funding from: European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC AdG A-DATADRIVE-B (290923). This paper reflects only the authors' views, the Union is not liable for any use that may be made of the contained information. Research Council KUL: GOA/10/09 MaNet, CoE PFV/10/002 (OPTEC), BIL12/11T; PhD/Postdoc grants. Flemish Government: FWO: projects: G.0377.12 (Structured systems), G.088114N (Tensor based data similarity); PhD/Postdoc grants. IWT: projects: SBO POM (100031); PhD/Postdoc grants. iMinds Medical Information Technologies SBO 2014. Belgian Federal Science Policy Office: IUAP P7/19 (DYSCO, Dynamical systems, control and optimization, 2012-2017). Fund for Scientic Research (FWO-Vlaanderen), by the Flemish Government (Methusalem), the Belgian Government through the Inter university Poles of Attraction (IAP VII) Program, and by the ERC advanced grant SNLSID, under contract 320378.

Ricardo Castro-Garcia and Johan A.K. Suykens are with KU Leuven, ESAT-STADIUS, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium. (ricardo.castro@esat.kuleuven.be, johan.suykens@esat.kuleuven.be)

Koen Tiels and Johan Schoukens are with the Vrije Universiteit Brussel, Faculty of Engineering, Department of Fundamental Electricity and Instrumentation, Pleinlaan 2 1050 Brussels, Belgium (koen.tiels@vub.ac.be, johan.schoukens@vub.ac.be)



Fig. 1. A Hammerstein system. $G_0(q)$ is a linear dynamical system and $f(u(t))$ is a static nonlinearity.

In the framework of System Identification, LS-SVM has been applied before. Examples on well known benchmark data sets like the Wiener-Hammerstein data set [9] are available (e.g. [10] and [11]). Given the black box nature of LS-SVM, a natural improvement would be the ability to incorporate information about the structure of the system into the LS-SVM itself. This has been somewhat explored (e.g. [12], [13]). In the specific case of Hammerstein systems, in [14] LS-SVM is used in combination with overparametrization [15], [16]. Under the proposed methodology, it will be shown that the solution of the model follows from solving a linear system of equations. By itself, this already constitutes an advantage over other methods like overparametrization in the sense that the proposed method is much more simple and easy to implement.

Incorporating information of the system's structure into a LS-SVM model can be difficult. To do that, in this paper we use the BLA approach to model the linear block and use the results to help LS-SVM modeling the nonlinear part. To achieve this, the primal formulation of LS-SVM is modified to include the information of the structure of the system and the approximation to the linear block obtained through the BLA.

The proposed methodology can be separated in two stages:

- The system's BLA is calculated and used as an approximation to the linear block.
- A modified LS-SVM model is trained including the information given by the BLA of the system.

Note then that the full Hammerstein model consists of a nonlinear block given by the resulting LS-SVM model and a linear part coming from the BLA.

In this paper, the method is applied to two simulation examples and the results are presented. There, the output of the Hammerstein system is measured in the presence of

white Gaussian additive noise (i.e. $v(t)$ in Fig. 1).

It will be shown that in the presence of noise, the method can very effectively calculate an approximation to the nonlinear model (up to a scaling factor) and to the system as a whole. It is important to highlight that this scaling factor is not identifiable [17].

In this work scalars are represented in lower case, lower case followed by (t) is used for signals in the time domain and capital letters followed by (k) stand for signals in the frequency domain in the discrete time framework. e.g. x is a scalar, $x(t)$ is a time domain signal and $X(k)$ is the representation of $x(t)$ in the frequency domain (i.e. the discrete Fourier transform [18] is used). Also, vectors are represented as bold lower case variables, while matrices are bold upper case e.g. $\boldsymbol{x}$ is a vector and $\boldsymbol{X}$ is a matrix.

This work is organized as follows: In Section II the methods employed are explained. First LS-SVM for function estimation is presented and then, the BLA concept is described. The proposed method is presented in Section III where it is explained how the BLA and LS-SVM were used together. Section IV illustrates the results found when applying the described methodology on two simulation examples. Finally, in Section V, the conclusions and ideas for future work are presented.

II. Background and Problem Statement

To represent a linear dynamic block, an ARX model can be used [19]:

$$\hat{y}_t = \sum_{j=0}^{m} b_j u_{t-j} - \sum_{i=1}^{n} a_i y_{t-i}. \tag{1}$$

Here, $\hat{y}_t$ is the currently estimated value of the output, while y_{t-i} are past outputs and u_{t-j} represents the past and present inputs. Note that b_j and a_i represent the coefficients of the numerator and denominator of the linear block respectively.

In the Hammerstein case, the input u_t goes trough a nonlinear block first. This nonlinear block is represented as $f(u(t))$ in Fig. 1, therefore, the model is expressed as:

$$y_t = \sum_{j=0}^{m} b_j f(u_{t-j}) - \sum_{i=1}^{n} a_i y_{t-i} + e_t. \tag{2}$$

To obtain a representation of $f(u(t))$, LS-SVM will be employed and to obtain an approximation to the coefficients a_i and b_j, the BLA approach will be used.

A. Function estimation using LS-SVM

LS-SVM [8] is given in the framework of a primal-dual formulation. Having a data set $\{\boldsymbol{u}_i, x_i\}_{i=1}^{N}$, the objective is to find a model

$$\hat{x} = \boldsymbol{w}^T \boldsymbol{\varphi}(\boldsymbol{u}) + b. \tag{3}$$

Here $\boldsymbol{u} \in \mathbb{R}^n$, $\hat{x} \in \mathbb{R}$ represents the estimated output value, and $\boldsymbol{\varphi}(\cdot) : \mathbb{R}^n \to \mathbb{R}^{n_h}$ is the feature map to a high dimensional (possibly infinite) space.

A constrained optimization problem is then formulated [8]:

$$\begin{aligned} \min_{\boldsymbol{w},b,\boldsymbol{e}} \quad & \tfrac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + \tfrac{\gamma}{2}\sum_{i=1}^{N} e_i^2 \\ \text{subject to} \quad & x_i = \boldsymbol{w}^T\boldsymbol{\varphi}(\boldsymbol{u}_i) + b + e_i,\ i = 1, ..., N. \end{aligned} \tag{4}$$

Using Mercer's theorem [20], the kernel matrix $\boldsymbol{\Omega}_{i,j}$ can be represented by the kernel function $K(\boldsymbol{u}_i, \boldsymbol{u}_j) = \boldsymbol{\varphi}(\boldsymbol{u}_i)^T\boldsymbol{\varphi}(\boldsymbol{u}_j)$ with $i, j = 1, ..., N$. It is important to note that in this representation $\boldsymbol{\varphi}$ does not have to be explicitly known as this is implicitly used through the positive definite kernel function. In this paper the radial basis function kernel (i.e. RBF kernel) is used:

$$K(\boldsymbol{u}_i, \boldsymbol{u}_j) = \exp\left(\frac{-\|\boldsymbol{u}_i - \boldsymbol{u}_j\|_2^2}{\sigma^2}\right), \tag{5}$$

where σ is the tuning parameter.

From the Lagrangian $\mathcal{L}(\boldsymbol{w}, b, e; \alpha) = \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + \gamma\frac{1}{2}\sum_{i=1}^{N} e_i^2 - \sum_{i=1}^{N}\alpha_i(\boldsymbol{w}^T\boldsymbol{\varphi}(\boldsymbol{u}_i) + b + e_i - x_i)$ with $\alpha_i \in \mathbb{R}$ the Lagrange multipliers, the optimality conditions are derived:

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial \boldsymbol{w}} = 0 \to \boldsymbol{w} = \sum_{i=1}^{N}\alpha_i\boldsymbol{\varphi}(\boldsymbol{u}_i) \\ \frac{\partial \mathcal{L}}{\partial b} = 0 \to \sum_{i=1}^{N}\alpha_i = 0 \\ \frac{\partial \mathcal{L}}{\partial e_i} = 0 \to \alpha_i = \gamma e_i, i = 1, ..., N \\ \frac{\partial \mathcal{L}}{\partial \alpha_i} = 0 \to x_i = \boldsymbol{w}^T\boldsymbol{\varphi}(\boldsymbol{u}_i) + b + e_i, i = 1, ..., N. \end{cases} \tag{6}$$

By elimination of $\boldsymbol{w}$ and e_i the following linear system is obtained:

$$\left[\begin{array}{c|c} 0 & \mathbf{1}_N^T \\ \hline \mathbf{1}_N & \boldsymbol{\Omega} + \frac{1}{\gamma}\boldsymbol{I}_N \end{array}\right]\left[\begin{array}{c} b \\ \hline \boldsymbol{\alpha} \end{array}\right] = \left[\begin{array}{c} 0 \\ \hline \boldsymbol{x} \end{array}\right] \tag{7}$$

with $\boldsymbol{x} = [x_1, ..., x_N]^T$ and $\boldsymbol{\alpha} = [\alpha_1, ..., \alpha_N]^T$. The resulting model is then:

$$\hat{x}(\boldsymbol{u}) = \sum_{i=1}^{N}\alpha_i K(\boldsymbol{u}, \boldsymbol{u}_i) + b. \tag{8}$$

B. Best Linear Approximation

A system with input $u(t)$ and output $y(t)$ has its Best Linear Approximation [7] (BLA) defined as the linear system whose output best approximates the system's output in mean-square sense:

$$G_{BLA}(k) := \underset{G(k)}{\arg\min}\, E_u\left\{\|\tilde{Y}(k) - G(k)\tilde{U}(k)\|_2^2\right\}, \tag{9}$$

with

$$\begin{cases} \tilde{u}(t) = u(t) - E\{u(t)\} \\ \tilde{y}(t) = y(t) - E\{y(t)\}. \end{cases}$$

Here, G_{BLA} represents the frequency response function (FRF) of the BLA. The expectation E_u in eq. (9) is taken with respect to the random point $u(t)$.

Through this work, the mean values (i.e. $E\{u(t)\}$ and $E\{y(t)\}$) are removed from the signals when the BLA is calculated. In consequence, the notation $\tilde{u}(t)$ and $\tilde{y}(t)$ will be dropped in favor of the simpler notation $u(t)$ and $y(t)$.

If the BLA exists, eq. (9) is reduced to

$$G_{BLA}(k) = \frac{S_{YU}(k)}{S_{UU}(k)}, \tag{10}$$

where the expectation in the cross-power and auto-power spectra is again taken with respect to the random point $u(t)$.

Following Bussang's theorem [21], for Gaussian distributed inputs $u(t)$ the BLA of a Hammerstein system is proportional to the underlying linear dynamic system. Even more, for periodic excitations, eq. (10) comes down to [22]:

$$G_{BLA}(k) = E_u \left\{ \frac{Y(k)}{U(k)} \right\}. \tag{11}$$

For a random-phase multisine excitation [7], which is asymptotically Gaussian distributed, the BLA can be estimated by averaging eq. (11) over phase realizations of the multisine. The robust method [22] uses this concept to obtain nonparametric estimates of the BLA as well as the noise variance, the nonlinear variance, and the total (i.e. noise plus nonlinear) variance $\sigma_G^2(k)$.

In addition to the nonparametric estimate, it is possible to obtain a parametric transfer function model through a weighted least-squares estimator [23]:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} J_N(\boldsymbol{\theta}), \tag{12a}$$

where the cost function $J_N(\boldsymbol{\theta})$ is

$$J_N(\boldsymbol{\theta}) = \frac{1}{N} \sum_{k=1}^{N} W(k) \left| \hat{G}_{BLA}(k) - G_M(k, \boldsymbol{\theta}) \right|^2. \tag{12b}$$

In this representation, $W(k) \in \mathbb{R}^+$ is a deterministic, $\boldsymbol{\theta}$-independent weighting sequence (e.g. $\frac{1}{\sigma_G^2(k)}$), $\hat{G}_{BLA}(k)$ is an approximation to the actual $G_{BLA}(k)$ as it is limited to a finite number of realizations of $U(k)$ and $Y(k)$, and $G_M(k, \boldsymbol{\theta})$ is a parametric transfer function model

$$\begin{aligned} G_M(k, \boldsymbol{\theta}) &= \frac{\sum_{l=0}^{n_b} b_l e^{-j2\pi \frac{k}{N} l}}{\sum_{l=0}^{n_a} a_l e^{-j2\pi \frac{k}{N} l}} \\ &= \frac{B_\theta(k)}{A_\theta(k)}, \\ \boldsymbol{\theta} &= \begin{bmatrix} a_0 & \cdots & a_{n_a} & b_0 & \cdots & b_{n_b} \end{bmatrix}^T, \end{aligned} \tag{12c}$$

with the constraint $\|\boldsymbol{\theta}\|_2 = 1$ and the first non-zero element of $\boldsymbol{\theta}$ positive to obtain a unique parameterization.

An example of a resulting parametric transfer function $G_{BLA}(k)$ compared to the actual transfer function $G_0(k)$ is shown in Fig. 2. As can be seen, $G_{BLA}(k)$ resembles quite well the shape of $G_0(k)$ up to a certain scaling factor. This particular illustration corresponds to the second example presented in Section IV.

III. Proposed Method

Given that an approximation to the a_i and b_j coefficients of the transfer function can be estimated from the BLA, the aim is to incorporate this approximation in the formulation of LS-SVM to exploit the knowledge of the structure of the system.

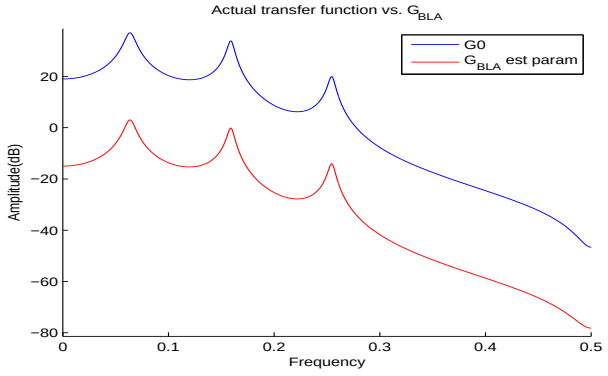

Fig. 2. Comparison of the magnitudes of $G_{BLA}(k)$ and the actual transfer function $G_0(k)$. This corresponds to example number 2.

It is important to note again that there is a scaling factor that differentiates G_0 and G_{BLA} as evidenced in Fig. 2 and mentioned in Section II-B. This of course affects the final results in the calculations, so in order to compensate for this, an additional constant k_{bla} is introduced. This constant will be tuned as an additional parameter in the model selection level. This gives the following model to be identified:

$$y_t = k_{bla} \sum_{j=0}^{m} b_j (\boldsymbol{w}^T \boldsymbol{\phi}(u_{t-j}) + d_0) - \sum_{i=1}^{n} a_i y_{t-i} + e_t. \tag{13}$$

For this model, one formulates the following constrained optimization problem:

$$\min_{\boldsymbol{w}, d_0, \boldsymbol{e}} J = \frac{1}{2} \boldsymbol{w}^T \boldsymbol{w} + \frac{\gamma}{2} \sum_{t=r}^{N} e_t^2 \tag{14}$$

s.t. eq. (13) holds for all $t = r, ..., N$. Here $r = \max(n, m) + 1$.

Given these elements, one has the following Lagrangian:

$$\begin{aligned} \mathcal{L}(\boldsymbol{w}, d_0, \boldsymbol{e}, \boldsymbol{\alpha}) = J - \sum_{t=r}^{N} \boldsymbol{\alpha}_t \Bigg(k_{bla} \sum_{j=0}^{m} b_j (\boldsymbol{w}^T \boldsymbol{\phi}(u_{t-j}) \\ + d_0) - \sum_{i=1}^{n} a_i y_{t-i} + e_t - y_t \Bigg). \end{aligned} \tag{15}$$

The optimality conditions become:

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial \boldsymbol{w}} = 0 & \rightarrow \quad \boldsymbol{w} = \sum_{t=r}^{N} \alpha_t k_{bla} \sum_{j=0}^{m} b_j \boldsymbol{\phi}(u_{t-j}) \\ \frac{\partial \mathcal{L}}{\partial d_0} = 0 & \rightarrow \quad \sum_{t=r}^{N} \alpha_t k_{bla} \sum_{j=0}^{m} b_j = 0 \\ \frac{\partial \mathcal{L}}{\partial e_t} = 0 & \rightarrow \quad \alpha_t = \gamma e_t \text{ for } t = r, ..., N \\ \frac{\partial \mathcal{L}}{\partial \alpha_t} = 0 & \rightarrow \quad y_t = k_{bla} \sum_{j=0}^{m} b_j (\boldsymbol{w}^T \boldsymbol{\phi}(u_{t-j}) + d_0) \\ & \qquad - \sum_{i=1}^{n} a_i y_{t-i} + e_t \text{ for } t = r, ..., N. \end{cases} \tag{16}$$

By replacing the first and third conditions (i.e. $\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}} = 0$ and $\frac{\partial \mathcal{L}}{\partial e_t} = 0$) into the last one (i.e. $\frac{\partial \mathcal{L}}{\partial \alpha_t} = 0$) one gets for

$t = r, ..., N$:

$$\begin{aligned} y_t &= k_{bla} \sum_{j=0}^{m} b_j \left(\left(\sum_{q=r}^{N} \alpha_q k_{bla} \sum_{p=0}^{m} b_p \boldsymbol{\phi}(u_{q-p}) \right)^T \right. \\ &\quad \left. \boldsymbol{\phi}(u_{t-j}) + d_0 \right) - \sum_{i=1}^{n} a_i y_{t-i} + \frac{\alpha_t}{\gamma} \\ &= k_{bla}^2 \sum_{j=0}^{m} \sum_{q=r}^{N} \sum_{p=0}^{m} b_j b_p \alpha_q \boldsymbol{\phi}(u_{q-p})^T \boldsymbol{\phi}(u_{t-j}) \\ &\quad + k_{bla} \sum_{j=0}^{m} b_j d_0 - \sum_{i=1}^{n} a_i y_{t-i} + \frac{\alpha_t}{\gamma}. \end{aligned} \tag{17}$$

Let us define:

$$\eta = N - r + 1 \tag{18}$$

$$\tilde{b} = \sum_{j=0}^{m} b_j \tag{19}$$

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_r & \cdots & \alpha_N \end{bmatrix}^T \quad \in \mathbb{R}^{\eta} \tag{20}$$

$$\boldsymbol{a} = -\begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix}^T \quad \in \mathbb{R}^{n} \tag{21}$$

$$\boldsymbol{y}_f = \begin{bmatrix} y_r & \cdots & y_N \end{bmatrix}^T \quad \in \mathbb{R}^{\eta} \tag{22}$$

$$\Omega_{k,l} = \boldsymbol{\phi}(u_k)^T \boldsymbol{\phi}(u_l) \quad \text{for } k, l = 1, ..., N \tag{23}$$

$$M(q,t) = \sum_{j=0}^{m} \sum_{p=0}^{m} b_j b_p (\Omega_{(q-p,t-j)}) \quad \text{for } t, q = r, ..., N \tag{24}$$

$$\boldsymbol{Y}_p = \begin{bmatrix} y_{r-1} & y_r & \cdots & y_{N-1} \\ y_{r-2} & y_{r-1} & \cdots & y_{N-2} \\ \vdots & \vdots & \ddots & \vdots \\ y_{r-n} & y_{r-n+1} & \cdots & y_{N-n} \end{bmatrix} \quad \in \mathbb{R}^{n \times \eta} \tag{25}$$

From $\frac{\partial \mathcal{L}}{\partial d_0} = 0$ one gets

$$k_{bla} \sum_{t=r}^{N} \alpha_t \tilde{b} = k_{bla} \tilde{b} \mathbf{1}^T \boldsymbol{\alpha} = 0 \tag{26}$$

and from $\frac{\partial \mathcal{L}}{\partial \alpha_t} = 0$:

$$\begin{aligned} \boldsymbol{y}_f &= k_{bla}^2 \sum_{j=0}^{m} \sum_{p=0}^{m} b_j b_p \boldsymbol{\Omega}_{(q-p,t-j)} \boldsymbol{\alpha} + \boldsymbol{Y}_p^T \boldsymbol{a} \\ &\quad + k_{bla} \tilde{b} \mathbf{1}^T d_0 + \gamma^{-1} \boldsymbol{I} \boldsymbol{\alpha} \quad \text{for } t, q = r, ..., N. \\ &= k_{bla}^2 \boldsymbol{M} \boldsymbol{\alpha} + \boldsymbol{Y}_p^T \boldsymbol{a} + k_{bla} \tilde{b} \mathbf{1}^T d_0 + \gamma^{-1} \boldsymbol{I} \boldsymbol{\alpha}. \end{aligned} \tag{27}$$

The obtained linear system can now be written as:

$$\begin{bmatrix} 0 & k_{bla} \tilde{b} \mathbf{1}_{\eta}^T \\ k_{bla} \tilde{b} \mathbf{1}_{\eta} & (k_{bla}^2 \boldsymbol{M} + \frac{\boldsymbol{I}}{\gamma}) \end{bmatrix} \begin{bmatrix} d_0 \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ \boldsymbol{y}_f - \boldsymbol{Y}_p{}^T \boldsymbol{a} \end{bmatrix}. \tag{28}$$

Under this representation, the model is linear in the unknowns, therefore, it can be solved directly given that the value of k_{bla} is obtained from the model selection.

Note that once $\boldsymbol{\alpha}$ and d_0 are known, it is possible to directly apply the model to new data points.

TABLE I
SELECTED PARAMETERS

	σ	γ	k_{bla}
Ex. 1	9.8220×10^{-4}	1.4846×10^{11}	4.8697×10^{4}
Ex. 2	0.002434	1.0224×10^{11}	4.2170×10^{5}

IV. RESULTS

The proposed methodology was applied to two systems in the discrete time framework. The first system was generated through a nonlinear block:

$$x(t) = u(t)^3 \tag{29}$$

and a linear block:

$$y(t) = \frac{B_1(q)}{A_1(q)} x(t) \tag{30}$$

where

$$\begin{aligned} B_1(q) &= 0.004728q^3 + 0.01418q^2 \\ &\quad + 0.01418q + 0.004728 \\ A_1(q) &= q^3 - 2.458q^2 + 2.262q - 0.7654. \end{aligned} \tag{31}$$

The second system corresponds to the one used in [14] and was generated through a nonlinear block:

$$x(t) = \text{sinc}(u(t))u(t)^2 \tag{32}$$

and a linear block:

$$y(t) = \frac{B_2(q)}{A_2(q)} x(t) \tag{33}$$

where

$$\begin{aligned} B_2(q) &= q^6 + 0.8q^5 + 0.3q^4 + 0.4q^3 \\ A_2(q) &= q^6 - 2.789q^5 + 4.591q^4 - 5.229q^3 \\ &\quad + 4.392q^2 - 2.553q + 0.8679. \end{aligned} \tag{34}$$

Note that this linear block and its corresponding BLA are shown in Fig. 2.

Both systems were excited using Multi Level Pseudo Random Signals (MLPRS) with an amplitude $\in \{-10, 10\}$. Grid search was used for the tuning of the parameters (i.e. σ, γ and k_{bla}). The corresponding selected values are shown in Table I. The training data set consisted of 2000 points, while the validation set consisted of 2500 data points.

The results for the first example can be seen in Figs. 3 and 4 while the results for the second example are shown in Figs. 5 and 6. In Figs. 3 and 5 the mean values of $y(t)$ and $\hat{y}(t)$ were extracted, their Mean Absolute Error (MAE) is calculated and the corresponding scaling factor used is presented. Figs. 4 and 6 show the comparison between the estimated model and the real ones. It is evident that even though they have very different magnitudes, their shape is quite similar. Note that this difference in scaling points to a factor appearing between the two blocks of the system. This factor is unidentifiable [17] and can be distributed between the two blocks.

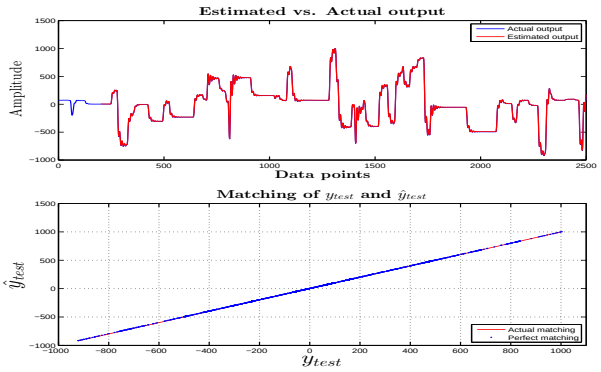

Fig. 3. Example 1: Overlapping of the actual output variable y and the estimation $\hat{y}$. Means extracted, MAE $= 2.6125$, $k_{bla} = 48696.7525$

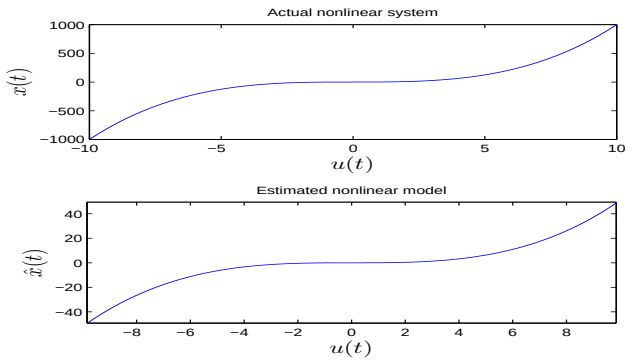

Fig. 4. Example 1: Comparison between the actual nonlinear system and the estimated model

The systems corresponding to both examples were affected with white Gaussian noise. However, just as in the calculation of the BLA, multiple realizations are used here, which diminishes the effect of the noise considerably.

The method was able to retrieve good approximations nevertheless. Figs. 7 and 8 show the evolution of the distributions of deviations from the actual output as the SNR changes.

As can be seen, the distribution of deviations from the actual output broadens as the SNR decreases. However, even with a large presence of noise, smaller deviations are more frequent which is in line with the type and magnitude of the noise introduced in the measurements.

Given that the proposed method takes the underlying structure of the system into account, it should better model the system than purely black box methods. Table II shows the results of the comparison between the proposed method and a LS-SVM NARX model in a validation set. In the table, the resulting error of LS-SVM NARX in One Step Ahead (1-ahead) mode is shown beside those of the proposed method (BLA+LS-SVM) in Simulation (sim) and one step ahead modes. These results were obtained when applying a SNR of 20dB and 100dB to examples 1 and 2 respectively. It can be seen that the proposed method clearly outperforms the purely black box approach of LS-SVM NARX.

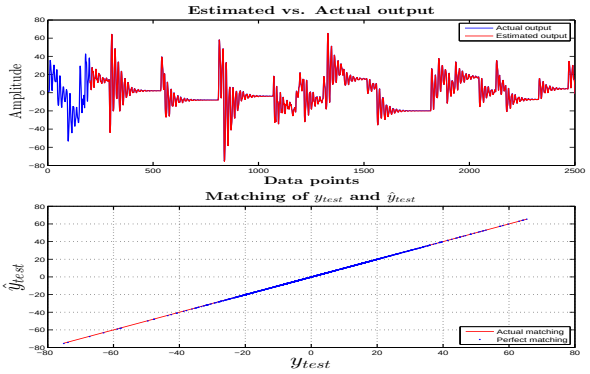

Fig. 5. Example 2: Overlapping of the actual output variable y and the estimation $\hat{y}$. Means extracted, MAE $= 0.18734$, $k_{bla} = 421696.5034$

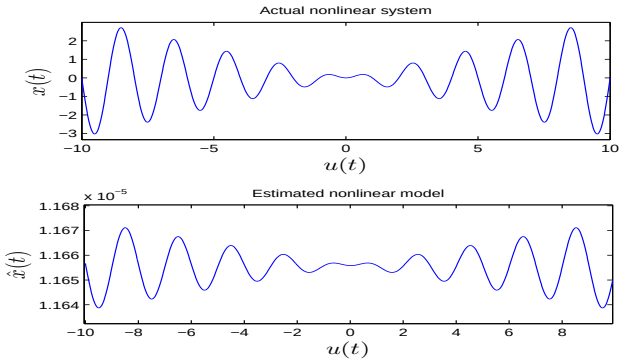

Fig. 6. Example 2: Comparison between the actual nonlinear system and the estimated model

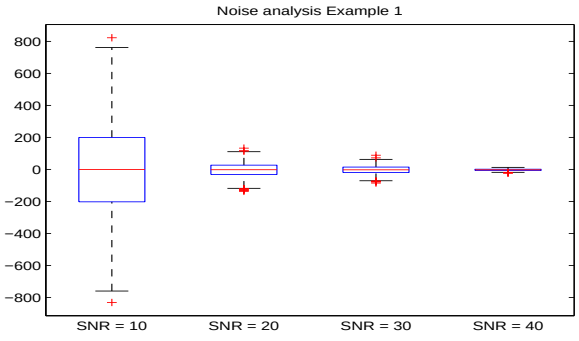

Fig. 7. Example 1: evolution of the distributions of deviations from the actual output as the SNR changes.

TABLE II
MEAN ABSOLUTE ERROR COMPARISON

Method	Ex 1	Ex 2
LSSVM NARX (1-ahead)	75.7453	0.48488
BLA+LS-SVM (sim)	34.7147	0.18133
BLA+LS-SVM (1-ahead)	30.7734	0.0031433

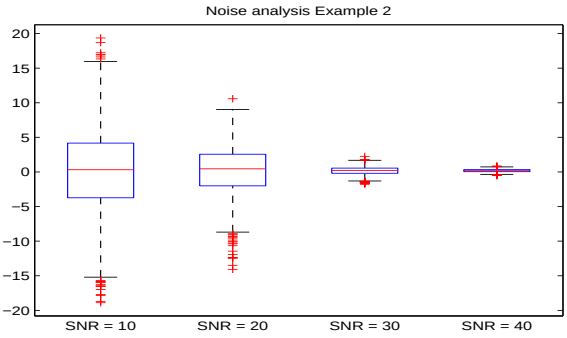

Fig. 8. Example 2: evolution of the distributions of deviations from the actual output as the SNR changes.

V. CONCLUSIONS AND FUTURE WORKS

A. Conclusions

The method presented combines two powerful techniques, namely LS-SVM and BLA, which when used in combination turn out to be quite effective for the identification of Hammerstein systems. In particular, the estimation of the linear block from BLA was used in the formulation of the dual representation for estimating the LS-SVM model.

The results presented indicate that the method is very effective in the presence of zero mean, white Gaussian noise.

For this method, the kernel parameter σ and the hyperparameter γ have to be calculated. However, once the model is learned, it can be easily applied.

It is important to highlight that the estimated nonlinear model is very close to the original one up to a scaling factor, which in a way allows great insight into the behavior of the system studied.

The solution of the model follows from solving a linear system of equations, which constitutes an advantage over other methods like the overparametrization presented in [14]. This is, in the sense of how easy it is to solve these equations and afterwards to apply the found model.

B. Future Works

Future work for the presented method includes the extension of the method to other block oriented structures like Wiener-Hammerstein systems where, after the identification of the estimated input and output linear blocks, the method could be applied. To separate these blocks, for example the phase coupled multisine approach [24] could be used.

Generalizing this method to express the cost function in the frequency domain would allow one to focus the fit of the model in a specifically needed part of the frequency band.

REFERENCES

[1] S. Billings and S. Fakhouri, "Identification of systems containing linear dynamic and static nonlinear elements," *Automatica*, vol. 18, no. 1, pp. 15–26, 1982.

[2] K. Narendra and P. Gallman, "An iterative method for the identification of nonlinear systems using a Hammerstein model," *IEEE Transactions on Automatic Control*, vol. 11, no. 3, pp. 546–550, 1966.

[3] F. Giri and E.-W. Bai, Eds., *Block-oriented nonlinear system identification*. Springer, 2010, vol. 1.

[4] E.-W. Bai, "Frequency domain identification of Hammerstein models," *IEEE Transactions on Automatic Control*, vol. 48, no. 4, pp. 530–542, 2003.

[5] R. Haber and L. Keviczky, *Nonlinear System Identification Input-Output Modeling Approach*. Springer The Netherlands, 1999, vol. 1.

[6] A. Janczak, *Identification of Nonlinear Systems Using Neural Networks and Polynomial Models: A Block-Oriented Approach*. Springer-Verlag Berlin Heidelberg, 2005, vol. 310.

[7] R. Pintelon and J. Schoukens, *System identification: a frequency domain approach*. John Wiley & Sons, 2012.

[8] J. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle, *Least Squares Support Vector Machines*. World Scientific, 2002.

[9] J. Schoukens, J. A. K. Suykens, and L. Ljung, "Wiener-Hammerstein benchmark," in *Proceedings of the 15th IFAC Symposium on System Identification*, 2009.

[10] K. De Brabanter, P. Dreesen, P. Karsmakers, K. Pelckmans, J. De Brabanter, J. A. K. Suykens, and B. De Moor, "Fixed-size LS-SVM applied to the Wiener-Hammerstein benchmark," in *Proceedings of the 15th IFAC symposium on system identification (SYSID 2009)*, 2009, pp. 826–831.

[11] M. Espinoza, K. Pelckmans, L. Hoegaerts, J. A. K. Suykens, and B. De Moor, "A comparative study of LS-SVMs applied to the Silverbox identification problem," in *Proc. of the 6th IFAC Symposium on Nonlinear Control Systems (NOLCOS)*, 2004.

[12] T. Falck, K. Pelckmans, J. A. K. Suykens, and B. De Moor, "Identification of Wiener-Hammerstein systems using LS-SVMs," in *Proceedings of the 15th IFAC symposium on system identification (SYSID 2009)*, 2009, pp. 820–825.

[13] T. Falck, P. Dreesen, K. De Brabanter, K. Pelckmans, B. De Moor, and J. A. K. Suykens, "Least-Squares Support Vector Machines for the identification of Wiener-Hammerstein systems," *Control Engineering Practice*, vol. 20, no. 11, pp. 1165–1174, 2012.

[14] I. Goethals, K. Pelckmans, J. A. K. Suykens, and B. De Moor, "Identification of MIMO Hammerstein models using Least-Squares Support Vector Machines," *Automatica*, vol. 41, no. 7, pp. 1263–1272, 2005.

[15] F. H. I. Chang and R. Luus, "A noniterative method for identification using the hammerstein model," *IEEE Transactions on Automatic Control*, vol. 16, pp. 464–468, 1971.

[16] E.-W. Bai, "An optimal two stage identification algorithm for Hammerstein-Wiener nonlinear systems," *Automatica*, vol. 34, pp. 333–338, 1998.

[17] S. Boyd and L. O. Chua, "Uniqueness of a basic nonlinear structure," *IEEE Transactions on Circuits and Systems*, vol. 30, no. 9, pp. 648–651, 1983.

[18] E. Brigham and R. Morrow, "The fast Fourier transform," *Spectrum, IEEE*, vol. 4, no. 12, pp. 63–70, 1967.

[19] L. Ljung, *System Identification: Theory for the User*. Pearson Education, 1998.

[20] J. Mercer, "Functions of positive and negative type, and their connection with the theory of integral equations," *Philosophical Transactions of the royal society of London. Series A, containing papers of a mathematical or physical character*, pp. 415–446, 1909.

[21] J. J. Bussgang, "Crosscorrelation functions of amplitude-distorted Gaussian signals," Research Laboratory of Electronics, Massachusetts Institute of Technology, Tech. Rep. 216, 1952.

[22] J. Schoukens, R. Pintelon, and Y. Rolain, *Mastering system identification in 100 exercises*. John Wiley & Sons, 2012.

[23] J. Schoukens, T. Dobrowiecki, and R. Pintelon, "Parametric and nonparametric identification of linear systems in the presence of nonlinear distortions-a frequency domain approach," *IEEE Transactions on Automatic Control*, vol. 43, no. 2, pp. 176–190, 1998.

[24] J. Schoukens, K. Tiels, and M. Schoukens, "Generating initial estimates for Wiener-Hammerstein systems using phase coupled multisines," in *Proceedings of 19th IFAC World Congress*, 2014.

[25] V. N. Vapnik and V. Vapnik, *Statistical learning theory*. Wiley New York, 1998, vol. 1.

[26] G. Wahba, *Spline models for observational data*. Siam, 1990, vol. 59.