Wiener System Identification using Best Linear Approximation within the LS-SVM framework

Ricardo Castro-Garcia, Johan A. K. Suykens

***Abstract*— Wiener systems represent a linear time invariant (LTI) system followed by a static nonlinearity. The identification of these systems has been a research problem for a long time as it is not a trivial task. A new methodology for identifying Wiener systems is proposed in this paper. The proposed method is a combination of well known techniques, namely the Best Linear Approximation (BLA) from the system identification field and Least Squares Support Vector Machines (LS-SVM). Through the BLA a non-parametric approximation to the LTI block is obtained. Next, the coefficients of the transfer function from the LTI block are estimated. Finally the calculated coefficients are included in an LS-SVM formulation for modeling the system. The results indicate that a good estimation of the underlying linear and nonlinear parts can be obtained up to a scaling factor.**

I. INTRODUCTION

When using a Block Oriented System Identification approach, systems are represented as interconnected linear and nonlinear blocks [1]. It is known that even simple nonlinear models often result in better approximations to process dynamics than the linear ones.

The Wiener system is a nonlinear system representation often used. It consists of a linear part $G_0(q)$ representing the dynamics of the process followed by a static part $f(\cdot)$ containing the nonlinearity (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 $q^{-1}x(t) = x(t-1)$. Many identification methods for Wiener systems have been reported in the literature. An overview of previous works can be found in [2].

The objective of this paper is to incorporate the techniques of the Best Linear Approximation (BLA) [3] within Least Squares Support Vector Machines (LS-SVM) [4] for the identification of Wiener Systems. In this paper, it is assumed that the intermediate variable between the two blocks is unknown, this is: only the input and output can be sampled.

EU: The research leading to these results has received funding from the European Research Council under the European Unions Seventh Framework Programme (FP7/2007-2013) / ERC AdG A-DATADRIVE-B (290923). This paper reflects only the authors views and the Union is not liable for any use that may be made of the contained information. Research Council KUL: 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 grant iMinds Medical Information Technologies SBO 2015 IWT: POM II SBO 100031 Belgian Federal Science Policy Office: IUAP P7/19 (DYSCO, Dynamical systems, control and optimization, 2012-2017)

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)



Fig. 1. A Wiener system. $G_0(q)$ is a linear dynamical system, $f(x(t))$ is a static nonlinearity and $v(t)$ represents the measurement noise.

The steps for identification of the linear and nonlinear parts can be clearly separated in the proposed method.

LS-SVM has been applied before in the System Identification community. Works like [5] and [6] have been tested on well known benchmark data sets like the Wiener-Hammerstein data set [7]. However, the black box nature of LS-SVM appears to be a limitation when other types of information are available. For instance, if the structure of the system is known, including it in the model should improve the performance. This has been explored e.g. in [8], [9], [10]. In fact the method proposed in this paper mirrors the work in [10], by the same authors. There, the Hammerstein case was explored. Note that even while the concept is similar, the mathematical development is completely different.

The incorporation of additional information regarding the structure of the system into an LS-SVM model can be difficult. In this paper the BLA approach is used to model the linear block and these results are used to help LS-SVM modeling the nonlinear part. For the proposed method 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 given the simplicity and easiness of implementation while offering a very good performance.

The proposed methodology can be separated in four stages:

- The system's BLA is calculated.
- A parametric version of the BLA is estimated and used as an approximation to the linear block.
- An approximation to the intermediate variable $\hat{x}(t)$ is obtained using the parametric BLA and the known input $u(t)$.
- An LS-SVM model is trained using $\hat{x}(t)$ and the known $y(t)$.

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

In this paper, the method is applied to two simulation

examples and the results are presented. There, the output of the Wiener system is measured in the presence of white Gaussian additive noise (i.e. $v(t)$ in Fig. 1).

It will be shown that also in the presence of noise, the method can very effectively calculate an approximation to the system as a whole.

Through 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 [11] 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

In the Wiener case, the input $u(t)$ goes through a linear block first. To represent a linear dynamic block, an ARX model can be used [12]:

$$x(t) = \sum_{j=0}^{m} b_j u(t-j) - \sum_{i=1}^{n} a_i x(t-i). \tag{1}$$

Here $x(t)$ is the intermediate variable at time t, while $x(t-i)$ are past outputs of such model and $u(t-j)$ the past and present inputs.

After this, the intermediate variable goes through a nonlinear block. This block is represented as $f(x(t))$ in Fig. 1, therefore, the output $\hat{y}(t)$ can be represented as:

$$y(t) = f(x(t)) = f(\sum_{j=0}^{m} b_j u(t-j) - \sum_{i=1}^{n} a_i x(t-i)). \tag{2}$$

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

A. Function estimation using LS-SVM

LS-SVM [4] is given in the framework of a primal-dual formulation. Having a data set $\{\boldsymbol{x}_i, y_i\}_{i=1}^{N}$, the objective is to find a model

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

Here $\boldsymbol{x} \in \mathbb{R}^n$ is the input, $\hat{y} \in \mathbb{R}$ 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 [4]:

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

This is the primal problem formulation and γ is the regularization term.

From the Lagrangian the optimality conditions can be derived:

$$\begin{aligned} \mathcal{L}(\boldsymbol{w}, b, e, \alpha) = \quad & \tfrac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + \gamma\tfrac{1}{2}\sum_{i=1}^{N} e_i^2 \\ & - \sum_{i=1}^{N} \alpha_i(\boldsymbol{w}^T\boldsymbol{\varphi}(\boldsymbol{x}_i) + d + e_i - y_i) \end{aligned} \tag{5}$$

with $\alpha_i \in \mathbb{R}$ the Lagrange multipliers and so:

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

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

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

with $\boldsymbol{\Omega}_{i,j} = \boldsymbol{\varphi}(\boldsymbol{x}_i)^T\boldsymbol{\varphi}(\boldsymbol{x}_j)$, $\boldsymbol{y} = [y_1, ..., y_N]^T$ and $\boldsymbol{\alpha} = [\alpha_1, ..., \alpha_N]^T$. This constitutes the dual problem formulation.

Using Mercer's theorem [13], the kernel matrix $\boldsymbol{\Omega}_{i,j}$ can be represented by the kernel function $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{\varphi}(\boldsymbol{x}_i)^T\boldsymbol{\varphi}(\boldsymbol{x}_j)$ with $i, j = 1, ..., N$. It is important to note that in this representation $\boldsymbol{\varphi}(\cdot)$ does not have to be explicitly known as it is implicitly used through the positive definite kernel function. In this paper the radial basis function kernel (i.e. RBF kernel) is used:

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(\frac{-\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2}{\sigma^2}\right), \tag{8}$$

where σ is the tuning parameter.

The resulting model is then:

$$\hat{y}(\boldsymbol{x}) = \sum_{i=1}^{N} \alpha_i K(\boldsymbol{x}, \boldsymbol{x}_i) + d. \tag{9}$$

B. Best Linear Approximation

For any system with input $u(t)$ and output $y(t)$, the Best Linear Approximation (BLA) is defined as the linear system whose output best approximates the system's output in mean-square sense [3] :

$$G_{BLA}(k) := \arg\min_{G(k)} E_u \left\{ \|\tilde{Y}(k) - G(k)\tilde{U}(k)\|_2^2 \right\}, \tag{10}$$

with

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

The G_{BLA} represents the frequency response function of the BLA. The expectation E_u in eq. (10) is taken with respect to the random point $u(t)$.

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

If the BLA exists, eq. (10) is reduced to

$$G_{BLA}(k) = \frac{S_{YU}(k)}{S_{UU}(k)}, \tag{11}$$

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

From Bussgang's theorem [14], the BLA of a Wiener system is proportional to the underlying linear dynamic system for Gaussian distributed inputs $u(t)$. Even more, for periodic excitations, eq. (11) can be rewritten as [15]:

$$G_{BLA}(k) = E_u\left\{\frac{Y(k)}{U(k)}\right\}. \tag{12}$$

For a random-phase multisine excitation [3], [16], which is asymptotically Gaussian distributed, the BLA can be estimated by averaging eq. (12) over phase realizations of the multisine.

Beside the nonparametric estimate, it is possible to obtain a parametric transfer function model through a weighted least-squares estimator [17]:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} J_N(\boldsymbol{\theta}), \tag{13a}$$

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

$$J_N(\boldsymbol{\theta}) = \frac{1}{N}\sum_{k=1}^{N} W(k)\left|\hat{G}_{BLA}(k) - G_M(k,\boldsymbol{\theta})\right|^2. \tag{13b}$$

Here, $W(k) \in \mathbb{R}^+$ is a deterministic, $\boldsymbol{\theta}$-independent weighting sequence, $\hat{G}_{BLA}(k)$ 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

$$\begin{aligned} G_M(k,\boldsymbol{\theta}) &= \frac{\sum_{l=0}^{m} b_l e^{-j2\pi\frac{k}{N}l}}{\sum_{l=0}^{n} a_l e^{-j2\pi\frac{k}{N}l}} \\ &= \frac{B_\theta(k)}{A_\theta(k)}, \\ \boldsymbol{\theta} &= \begin{bmatrix}\hat{a}_0 & \cdots & \hat{a}_n & \hat{b}_0 & \cdots & \hat{b}_m\end{bmatrix}^T, \end{aligned} \tag{13c}$$

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

III. Proposed Method

The goal of this work is to incorporate the coefficients estimated through the BLA (i.e. $\hat{a}_i$ and $\hat{b}_j$) into an LS-SVM model to exploit the knowledge of the structure of the system. With these coefficients, obtaining an approximation to the intermediate variable $\hat{x}(t)$ is straightforward:

$$\hat{x}(t) = \sum_{j=0}^{m}\hat{b}_j u(t-j) - \sum_{i=1}^{n}\hat{a}_i\hat{x}(t-i). \tag{14}$$

Replacing (14) into (2) we get $\hat{y}(t) = f(\hat{x}(t))$ and using (3) to model the non linear block, we can estimate an approximation to the output signal $\hat{y}(t)$:

$$\hat{y}(t) = \boldsymbol{w}^T\boldsymbol{\varphi}(\hat{x}(t)) + d. \tag{15}$$

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

$$\min_{\boldsymbol{w},d_0,\boldsymbol{e}} J = \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + \frac{\gamma}{2}\sum_{t=1}^{N} e_t^2 \tag{16}$$

s.t. eq. (15) holds for all $t = 1, ..., N$.

Given these elements, one has the following Lagrangian:

$$\begin{aligned}\mathcal{L}(\boldsymbol{w},d,\boldsymbol{e},\boldsymbol{\alpha}) = \;& \tfrac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + \tfrac{\gamma}{2}\sum_{t=1}^{N} e_t^2 \\ & - \sum_{t=r}^{N}\boldsymbol{\alpha}_t(\boldsymbol{w}^T\boldsymbol{\varphi}(\hat{x}(t)) + d + e(t) - y(t)).\end{aligned} \tag{17}$$

From here, it is evident that the formulation becomes exactly as described in Section II-A. This is very convenient as the problem becomes then a standard LS-SVM one after we obtain the coefficients $\hat{a}_i$ and $\hat{b}_i$.

Note that the order of the transfer function representing the linear block is unknown, this means that different values have to be tried until a fit combination is found.

It is important to note that there is a scaling factor that differentiates the actual $G_0(q)$ and the actual $G_{BLA}(q)$. This scaling difference implies that the nonlinear model will have to compensate for the difference so it has no effect on the input-output behavior of the estimated Wiener model (i.e. any pair of $\{G(q)/\eta,\ \eta f(x(t))\}$ with $\eta \neq 0$ would yield identical input and output measurements). However, this factor between the blocks is unidentifiable [18].

The accuracy of the method will depend then on how well the parameters of the linear block are estimated as will be shown in Section IV-B.

IV. Simulation Results

A. Examples

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

$$y(t) = x(t)^3 \tag{18}$$

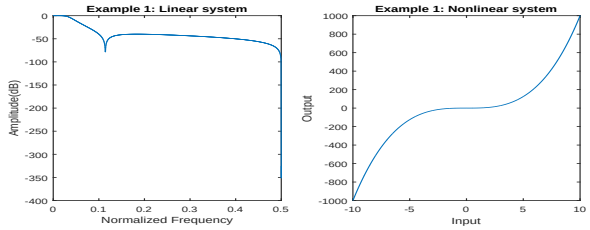

Fig. 2. Example 1. (Left) Linear system. (Right) Nonlinear system.

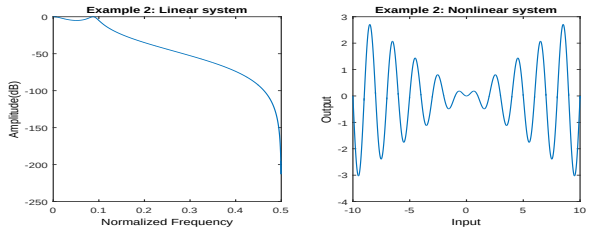

Fig. 3. Example 2. (Left) Linear system. (Right) Nonlinear system.

and a linear block:

$$x(t) = \frac{B_1(q)}{A_1(q)} u(t) \tag{19}$$

where

$$\begin{aligned} B_1(q) = &\quad 0.0089q^3 - 0.0045q^2 - 0.0045q + 0.0089 \\ A_1(q) = &\quad q^3 - 2.5641q^2 + 2.2185q - 0.6456. \end{aligned} \tag{20}$$

The second system was generated through a nonlinear block:

$$y(t) = \text{sinc}(x(t))x(t)^2 \tag{21}$$

and a linear block:

$$x(t) = \frac{B_2(q)}{A_2(q)} u(t) \tag{22}$$

where

$$\begin{aligned} B_2(q) = &\quad 0.0047q^3 + 0.0142q^2 + 0.0142q^1 + 0.0047 \\ A_2(q) = &\quad q^3 - 2.458q^2 + 2.262q - 0.7654 \end{aligned} \tag{23}$$

Figures 2 and 3 illustrate examples 1 and 2 respectively. Both systems were trained using a ramp signal from -15 to 15 (45 degrees slope). Also, in both cases the test sets were Multi Level Pseudo Random Signals (MLPRS) with an amplitude $\in \{-10, 10\}$. A Coupled Simulated Annealing algorithm was used to tune the parameters (i.e. σ and γ) using a 10-fold Cross-Validation scheme (e.g. LS-SVMlab v1.8). The training set for the nonlinear block and the test data set consisted each of 1000 points.

For the estimation of the orders of the transfer function, values from $n \in \{1, 5\}$ and $m \in \{1, 5\}$ with $n \geq m$ were tried (i.e. see (1)). At each iteration, the combination of n and m giving the best accuracy was selected.

In order to be able to compare between the results of both examples, let us have the Normalized MAE defined as shown in (24) for a signal with N measurements. Note that the

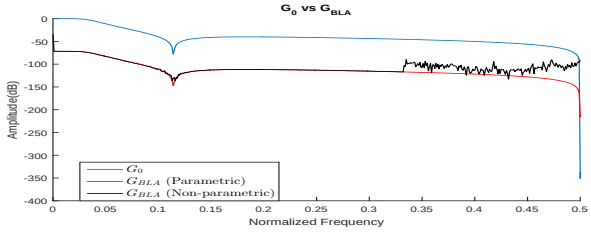

Fig. 4. Example 1. Linear block estimation.

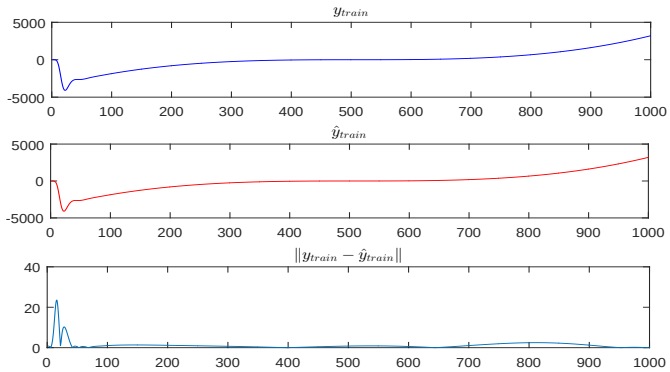

Fig. 5. Example 1. Non-Linear block behavior in the training set. Horizontal axes are the samples. Vertical axes are amplitude. (Top) Actual training output. (Middle) Estimated train output. (Bottom) Difference between actual and estimated outputs.

Normalized MAE uses the noise free signal $y_{test}(t)$ and its estimated counterpart $\hat{y}_{test}(t)$.

$$\%\text{MAE} = \frac{100}{N} \frac{\sum_{t=1}^{N} |y_{test}(t) - \hat{y}_{test}(t)|}{|\max(y_{test}(t)) - \min(y_{test}(t))|}. \tag{24}$$

Results for the first example can be seen in Figs. 4, 5 and 6 and results for the second example are shown in Figs. 7, 8 and 9. The systems corresponding to both examples were affected with white Gaussian noise (i.e. A Signal to Noise Rartio of 40dB was used in Figures 4 to 9).

Figures 4 and 7 show the estimated model of the linear blocks from the BLA for both examples. Note that the perturbation in the non-paranetric G_{BLA} after 33% of the frequency is due to the lack of excitation from the used signals. Figures 5 and 8 show the behavior of the estimated model of the nonlinear block for the training set of both examples. Finally, Figures 6 and 9 show the behavior of the estimated model of the whole system in the test set for each example.

Note that even though the models of the linear an nonlinear parts have different magnitudes than their corresponding actual blocks, their shape is very similar. The difference in scaling in the linear and nonlinear blocks points to a factor appearing between the two blocks of the system. This factor is unidentifiable and can be distributed between the two blocks as mentioned in Section III. Nonetheless, the resulting model is very accurate.

For the estimation of the BLA multiple realizations were used (i.e. 5000 realizations of 1000 points each for each example). This diminishes the effect of the noise considerably

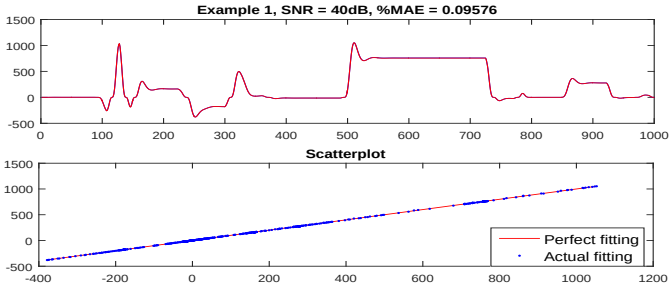

Fig. 6. Example 1. Model behavior in the test set. (Top) Overlapping of actual $y_{test}(t)$ and $\hat{y}_{test}(t)$. (Bottom) Scatterplot comparing the ideal and actual outputs.

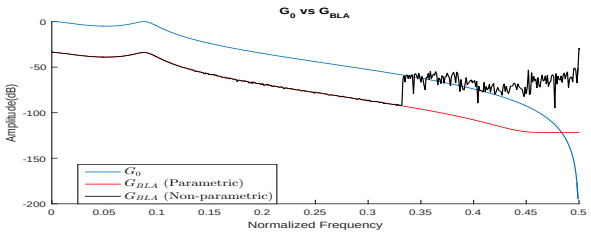

Fig. 7. Example 2. Linear block estimation.

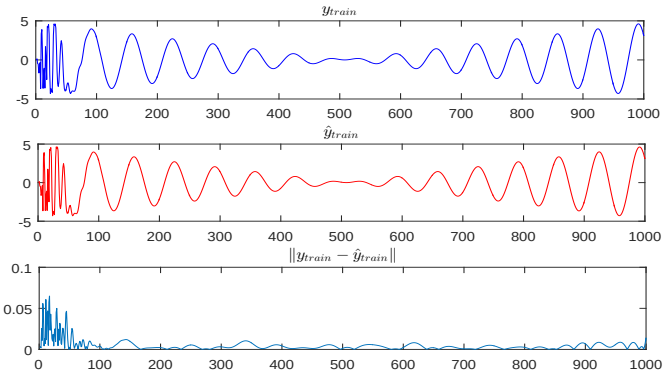

Fig. 8. Example 2. Non-Linear block behavior in the training set. Horizontal axes are the samples. Vertical axes are amplitude. (Top) Actual training output. (Middle) Estimated train output. (Bottom) Difference between actual and estimated outputs.

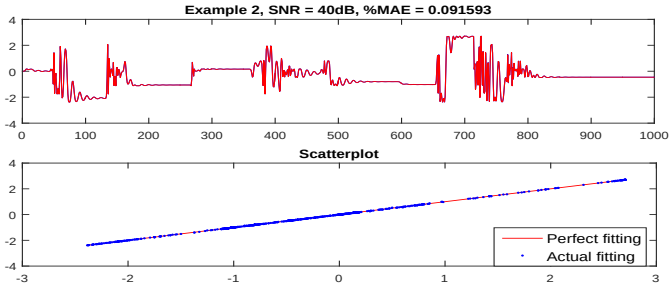

Fig. 9. Example 2. Model behavior in the test set. (Top) Overlapping of actual $y_{test}(t)$ and $\hat{y}_{test}(t)$. (Bottom) Scatterplot comparing the ideal and actual outputs.

		Example 1 Realizations				Example 2 Realizations			
SNR	Points	10	100	1000	5000	10	100	1000	5000
Inf	100	24.1393	6.8191	0.24781	0.22055	74.3324	13.1627	5.0186	0.71159
	500	102.4168	0.34096	0.067456	0.046364	15.0309	4.3071	0.35229	0.15282
	1000	63.2599	0.41402	0.041593	0.028606	15.5399	0.91661	0.20108	0.080908
	2000	376.8908	0.4435	0.036867	0.01633	8.1452	0.45106	0.11077	0.037773
40dB	100	134.3097	14.6472	0.2178	0.22376	20.3352	12.2842	3.0184	0.57317
	500	18.8368	0.44293	0.081862	0.062664	15.0489	3.6523	0.31716	0.19719
	1000	48.1492	0.46629	0.07963	0.059179	15.3069	0.95028	0.18943	0.089143
	2000	240.6889	0.50775	0.077693	0.049547	5.7503	0.50761	0.12734	0.089663
10dB	100	1627.2739	1785.4266	1.8413	1.3841	124.9963	11.6749	3.7194	2.599
	500	13.4591	2.4265	1.3931	1.2077	13.3474	4.464	2.1215	1.7422
	1000	12.4578	2.4247	1.241	1.0101	12.8389	2.821	1.7309	1.7859
	2000	11.8653	2.3946	1.1708	1.0846	11.9016	2.1782	1.7085	1.6505

TABLE I

EFFECT OF THE NUMBER OF POINTS USED IN THE BLA ESTIMATION OVER THE $\%MAE$ OF THE RESULTING MODEL. EACH VALUE REFLECTS THE MEAN OF THE $\%MAE$ OVER 20 MONTE CARLO SIMULATIONS.

in the linear block modeling. Further study of the impact of the number of points used during the BLA estimation will be presented in Section IV-B. The effect of different levels of noise will be considered in Section IV-C.

B. Impact of number of realizations for the BLA

In order to determine the effect of the number of realizations and the number of points per realization for the estimation of the BLA and subsequently for the accuracy of the model, a series of Monte Carlo simulations were run. For every different number of realizations and points per realizations used, 20 Monte Carlo simulations were carried out and the average of their Normalized MAEs is presented in Table I. Three different levels of Signal to Noise Ratio were used to offer a view of how the relevance of this options vary with the level of noise present.

It is clear that the more points that are used for the estimation of the BLA, the more accurate the final result of the model will be. This is not particularly surprising, however, it is interesting to note, that even when not using the maximum number of points, the results can still be very good as long as enough points are used as shown in Table I. This means that a tradeoff between the number of points used and the desired accuracy is present in the method and is particularly relevant for the BLA part.

C. Noise impact and methods comparison

Once more, in order to consider the effect of noise in the results given by the method, a series of Monte Carlo simulations were run. In Fig. 10 the results of 100 Monte Carlo simulations are presented for examples 1 and 2. For each of the examples, different levels of noise were considered. Other than the noise, the same type of signals of Section IV-A were used. In addition, in Fig. 11 an equivalent series of Monte Carlo simulations were run using the NARX-LSSVM approach [4]. These results allow a comparison between the proposed method and a black box approach to take place. For the NARX-LSSVM cases, the same number of realizations as used in the proposed method were averaged, thus diminishing the effective noise considerably. Note that in the proposed case this is only done for the linear block estimation.

In addition, in Table II, the corresponding medians of Figs. 10 and 11 are summarized for clarity.

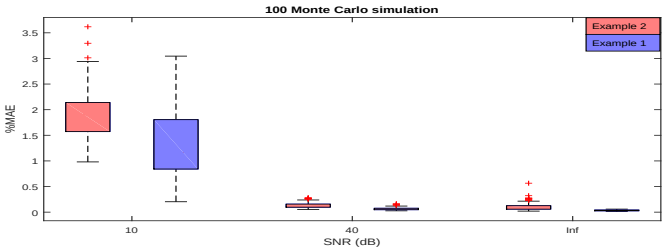

Fig. 10. Monte Carlo simulations for the proposed method.

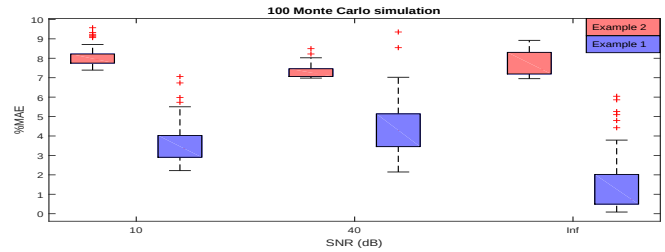

Fig. 11. Monte Carlo simulations for NARX-LSSVM.

V. CONCLUSIONS AND FUTURE WORK

A. Conclusions

The proposed method uses powerful techniques from two different fields. On one hand the BLA from the System Identification field and on the other, LS-SVM . When put together, these techniques are shown to be very effective for the identification of Wiener systems.

In this paper the LS-SVM formulation was modified to include further information from the system with the help of the BLA.

The results presented indicate that the method is very effective in the presence of zero mean, white Gaussian noise as long as enough samples can be measured. Indeed it can outperform powerful methods for black box modeling like NARX-LSSVM were the structure of the system is not considered.

Once all the parameters of the method (i.e. $\hat{a}_i$, $\hat{b}_j$, σ and γ) are estimated, new points can be easily evaluated.

The method can provide insight into the studied system as it allows to obtain models of the linear and nonlinear blocks that resemble the actual system quite accurately though in a rescaled manner.

Being able to draw the solution of the model from a linear system of equations is by itself an advantage over other methods like overparametrization.

		SNR (dB)		
		10	40	Inf
BLA + LS-SVM	EX 1	1.1675	0.062029	0.031687
	EX 2	1.8979	0.11666	0.082519
NARX-LSSVM	EX 1	3.3492	4.2889	0.74061
	EX 2	7.944	7.1879	7.4343

TABLE II

SUMMARY OF MEDIANS FOR THE MONTE CARLO SIMULATIONS OF FIGS. 10 AND 11

B. Future Work

An interesting extension would be combining the present work with the phase coupled multisine approach proposed in [19] and the Hammerstein System Identification presented in [10]. Such combination would be a natural extension for the identification of Wiener-Hammerstein systems. This would be possible thanks to the capabilities of the phase coupled multisine approach to give an estimation of each of the linear blocks of such a system.

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

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

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

[5] 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.

[6] 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.

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

[8] 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.

[9] 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.

[10] R. Castro-Garcia, K. Tiels, J. Schoukens, and J. A. K. Suykens, "Incorporating Best Linear Approximation within LS-SVM-Based Hammerstein System Identification," in *Proceedings of the 54th IEEE Conference on Decision and Control (CDC 2015)*. IEEE, 2015, pp. 7392–7397.

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

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

[13] 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.

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

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

[16] P. Crama and J. Schoukens, "Initial estimates of Wiener and Hammerstein systems using multisine excitation," *IEEE transactions on Instrumentation and Measurement*, vol. 50, no. 6, pp. 1791–1795, 2001.

[17] 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.

[18] 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.

[19] 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.