

International Journal of Control

ISSN: 0020-7179 (Print) 1366-5820 (Online) Journal homepage: http://www.tandfonline.com/loi/tcon20

Hammerstein system identification through best linear approximation inversion and regularisation

Ricardo Castro-Garcia, Koen Tiels, Oscar Mauricio Agudelo & Johan A. K. Suykens

To cite this article: Ricardo Castro-Garcia, Koen Tiels, Oscar Mauricio Agudelo & Johan A. K. Suykens (2017): Hammerstein system identification through best linear approximation inversion and regularisation, International Journal of Control, DOI: 10.1080/00207179.2017.1329550

To link to this article: http://dx.doi.org/10.1080/00207179.2017.1329550

Published online: 23 May 2017.

Submit your article to this journal

Article views: 27

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=tcon20

Download by: [KU Leuven Libraries] **Date:** 30 July 2017, At: 23:27

INTERNATIONAL JOURNAL OF CONTROL, 2017
https://doi.org/10.1080/00207179.2017.1329550

Taylor & Francis
Taylor & Francis Group

Check for updates

Hammerstein system identification through best linear approximation inversion and regularisation

Ricardo Castro-Garcia [a], Koen Tiels[b], Oscar Mauricio Agudelo[a] and Johan A. K. Suykens[a]

[a]ESAT-STADIUS, KU Leuven, Leuven, Belgium; [b]Department of Fundamental Electricity and Instrumentation, Faculty of Engineering, Vrije Universiteit Brussel, Brussels, Belgium

ABSTRACT

Hammerstein systems are composed by the cascading of a static nonlinearity and a linear system. In this paper, a methodology for identifying such systems using a combination of least squares support vector machines (LS-SVM) and best linear approximation (BLA) techniques is proposed. To do this, a novel method for estimating the intermediate variable is presented allowing a clear separation of the identification steps. First, an approximation to the linear block is obtained through the BLA of the system. Then, an approximation to the intermediate variable is obtained using the inversion of the estimated linear block and the known output. Afterwards, a nonlinear model is calculated through LS-SVM using the estimated intermediate variable and the known input. To do this, the regularisation capabilities of LS-SVM play a crucial role. Finally, a parametric re-estimation of the linear block is made. The method was tested in three examples, two of them with hard nonlinearities, and was compared with four other methods showing very good performance in all cases. The obtained results demonstrate that also in the presence of noise, the method can effectively identify Hammerstein systems. The relevance of these findings lies in the fact that it is shown how the regularisation allows to bypass the usual problems associated with the noise backpropagation when the inversion of the estimated linear block is used to compute the intermediate variable.

ARTICLE HISTORY
Received 20 September 2016
Accepted 7 May 2017

KEYWORDS
Best linear approximation; frequency domain; Hammerstein systems; least squares support vector machines; nonlinear system identification; linear block inversion

1. Introduction

Various block structured models have been introduced (Billings & Fakhouri, 1982) and it is known that even simple nonlinear models can often provide much better approximations to process dynamics than linear ones. A frequently used nonlinear model structure is the Hammerstein system (Narendra & Gallman, 1966). It consists of a static part $f(\cdot)$, which contains the nonlinearity, followed by a linear part $G_0(q)$, which contains the dynamics of the process (see Figure 1). Throughout this paper, the q-notation will be used. The operator q is a time shift operator of the form $q^{-1}x(t) = x(t-1)$.

Many identification methods of Hammerstein systems have been reported in the literature. An overview of previous works can be found in Giri and Bai (2010). Also, different classifications of these methods can be found in Bai (2003), Haber and Keviczky (1999) and Janczak (2005).

In this paper, the best linear approximation (BLA) technique (Pintelon & Schoukens, 2012) is used in order to find an initial estimate of the linear block. With this model the identification of the nonlinear model can proceed. It is important to use a technique that includes regularisation as will be shown later. For this task we use least squares support vector machines (LS-SVM) (Suykens, Van Gestel, De Brabanter, De Moor, & Vandewalle, 2002) given its well-known generalisation properties and its incorporated regularisation mechanisms. In order to link these two steps a novel method to obtain an estimation of the intermediate variable (i.e. $x(t)$ in Figure 1) is proposed. A mix of techniques is used then and they are applied where they excel, that is, we use the BLA to model the linear part and LS-SVM to model the nonlinear one.

LS-SVM has been applied to system identification before and results on well-known benchmark data sets like the Wiener–Hammerstein data set (Schoukens, Suykens, & Ljung, 2009) have been presented (e.g. De Brabanter et al., 2009 ;Espinoza, Pelckmans, Hoegaerts, Suykens, & De Moor, 2004). Some have attempted to include information about the structure of the system into the LS-SVM models (Falck, Pelckmans, Suykens, & De Moor, 2009; Falck et al., 2012) and some have applied them specifically to Hammerstein systems (Castro-Garcia, Tiels, Schoukens, & Suykens, 2015; Goethals, Pelckmans, Suykens, & De Moor, 2005). However, given that incorporating information of the system's structure into the LS-SVM model is difficult, in this paper we have opted for a different and much simpler approach.

CONTACT Ricardo Castro-Garcia rcastro@esat.kuleuven.be

© 2017 Informa UK Limited, trading as Taylor & Francis Group

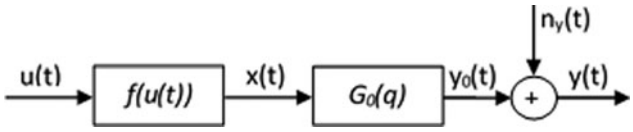

Figure 1. A Hammerstein system. $G_0(q)$ is a linear dynamical system, $f(u(t))$ is a static nonlinearity and $n_y(t)$ is the measurement noise.

The proposed methodology can be separated in four stages:

- The system's BLA is calculated and used as an approximation to the linear block.
- The intermediate variable is estimated using the inversion of the approximated linear block and the known output.
- An LS-SVM model is trained using the known input and the estimated intermediate variable.
- On an independent data set, the linear block is re-estimated using the result from applying the input to the estimated model of the nonlinear block (i.e. the estimated intermediate variable) and the known output.

The proposed method uses a backward approach as defined in Sun, Liu, and Sano (1999) where the linear dynamic part is identified first, then the intermediate variable is estimated and finally the nonlinear part is modelled. This type of approach is also used in other works like Bai and Fu (2002), Bai (1998) and Wang, Sano, Chen, and Huang (2009). Note, however, that our approach differs from the blind approach defined in Bai and Fu (2002) as the linear block is estimated using both input and output signals. Additionally, in this work we make a final refinement of the estimation of the linear block after having a model for the nonlinear block.

The idea of using the inversion of the estimated linear block is not new but is in general regarded as a bad idea. As explained in Crama and Schoukens (2001), among other problems, the inversion of the found model implies that at those frequencies where the amplitude of the transfer function is small, the noise will be amplified. Nevertheless, the concept has been employed, for example in Bai and Fu (2002) and Bai (2004) where the effect of noise is not explored deeply. In contrast, in the present work the output of the Hammerstein system is measured in the presence of high levels of white Gaussian additive noise $n_y(t)$ (see Figure 1). Additionally, some common assumptions like a known order of the linear system are not necessary here.

In addition to the problem posed by the noise, the possibility of the system being non-minimum phase is a serious concern to this type of approach. In the presented method, we offer a way to work around these problems.

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. For example, x is a scalar, $x(t)$ is a signal in the time domain and $X(k)$ is the representation of $x(t)$ in the frequency domain (i.e. the discrete Fourier transform (Brigham & Morrow, 1967) is used). Also, vectors are represented with bold lower case and matrices with bold upper case, e.g. $\boldsymbol{x}$ is a vector and $\boldsymbol{X}$ is a matrix.

The paper is organised as follows: In Section 2, the methods used to implement the proposed system identification technique are explained. This includes LS-SVM for function estimation and the BLA concept. In Section 3, the proposed method is presented. Section 4 illustrates the results found when applying the described methodology on three simulation examples, one of which is based on a real-life application. Finally, in Section 5, the conclusions are presented.

2. Background methods

2.1 Best linear approximation

The BLA (Pintelon & Schoukens, 2012) of a PISPO (i.e. Period In Same Period Out) system[1] with input $u(t)$ and output $y(t)$ is defined as the linear system whose output approximates the system's output best in mean-square sense, i.e.

$$G_{\mathrm{BLA}}(k) := \arg\min_{G(k)} E_u \left\{ \|\tilde{Y}(k) - G(k)\tilde{U}(k)\|_2^2 \right\}, \quad (1)$$

with

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

where G_{BLA} is the frequency response function of the BLA, and where the expectation in Equation (1) is taken with respect to the random input $u(t)$.

It is assumed that the mean values are removed from the signals when a BLA is calculated. The notations $u(t)$ and $y(t)$ will be used instead of $\tilde{u}(t)$ and $\tilde{y}(t)$.

If the BLA exists, the minimiser in Equation (1) can be found as

$$G_{\mathrm{BLA}}(k) = \frac{S_{YU}(k)}{S_{UU}(k)}, \quad (2)$$

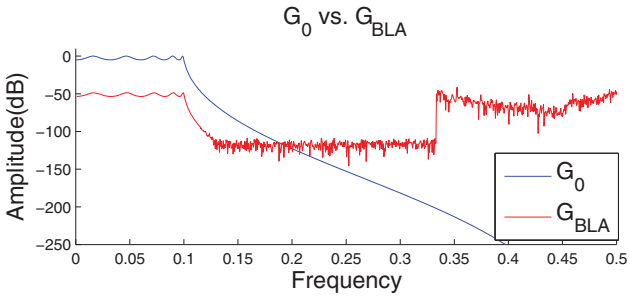

Figure 2. Example: comparison of the magnitudes of $G_{\mathrm{BLA}}(k)$ and the actual transfer function $G_0(k)$. $G_0(k)$ corresponds to a 10th order Chebyshev low-pass digital filter with normalised passband edge frequency 0.2 and 5 dB of peak-to-peak ripple in the passband.

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

Note that for periodic excitations with a fixed amplitude spectrum $|\boldsymbol{U}(k)|$ (such that $E_u\left\{|\boldsymbol{U}(k)|^2\right\} = |\boldsymbol{U}(k)|^2$), Equation (2) reduces to Schoukens, Pintelon, and Rolain (2012):

$$G_{\mathrm{BLA}}(k) = E_u\left\{\frac{Y(k)}{U(k)}\right\}. \tag{3}$$

In this work, random-phase multisine excitations (Pintelon & Schoukens, 2012) will be used for calculating the BLA. These asymptotically Gaussian-distributed signals are periodic and the BLA can thus be estimated by averaging (3) over a number of phase realisations of the multisine. This is the main idea in the robust method (Schoukens et al., 2012), which provides non-parametric estimates of the BLA, the noise variance, the nonlinear variance and the total (i.e. noise plus nonlinear) variance. Alternatively, the fast method (Pintelon & Schoukens, 2012) can be used to calculate the BLA from only one phase realisation, although some restrictions apply (Schoukens et al., 2012).

For Gaussian-distributed inputs $u(t)$, it follows from Bussgang's theorem (Bussgang, 1952) that the BLA of a Hammerstein system is proportional to the underlying linear dynamic system.

An example of a resulting $G_{\mathrm{BLA}}(k)$ compared to the actual transfer function $G_0(k)$ is shown in Figure 2. As it can be seen, $G_{\mathrm{BLA}}(k)$ resembles quite well the shape of $G_0(k)$ up to a certain frequency and up to a certain scaling factor.

It is important to highlight that the reliability of $G_{\mathrm{BLA}}(k)$, the model found, decreases as the frequency grows apart from the band of interest. This can be seen, for instance, in the fluctuations present in Figure 2 from 12% and in Figure 3 from 30% of the sampling frequency (i.e. the signal-to-noise ratio is smaller around these frequencies than for lower ones). Note that the sudden perturbation at 33% of the sampling frequency in

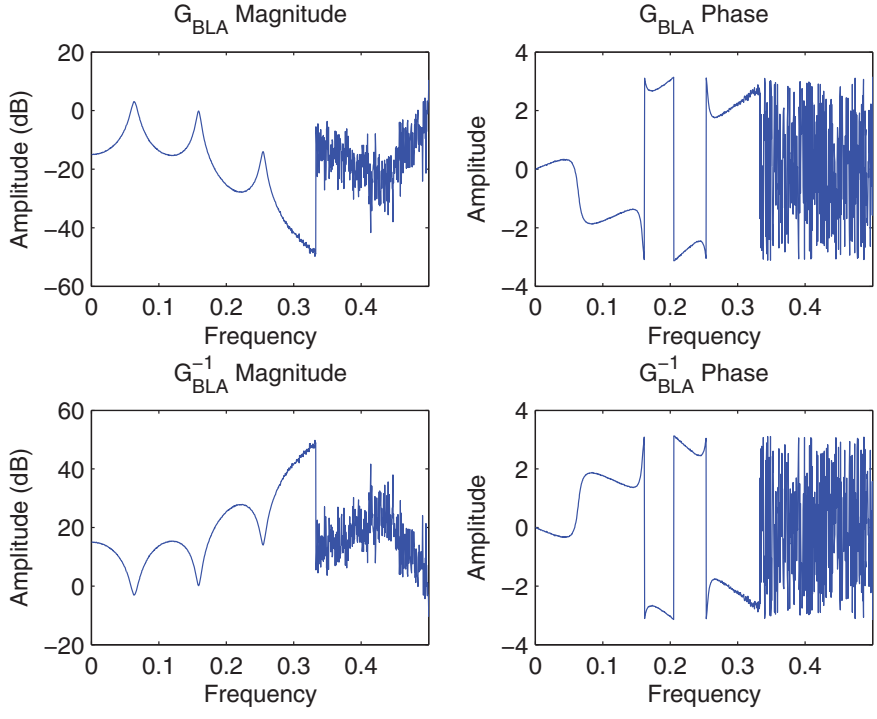

Figure 3. Example: Bode plot for $G_{\mathrm{BLA}}(k)$ (upper left and right) and its corresponding inverted result $G_{\mathrm{BLA}}^{-1}(k)$ (lower left and right). The shown system corresponds to $G_0(q) = \frac{q^6+0.8q^5+0.3q^4+0.4q^3}{q^6-2.789q^5+4.591q^4-5.229q^3+4.392q^2-2.553q+0.8679}$.

both graphs is due to the lack of excitation of the following frequencies and thus is nothing more than an artifact due to the chosen excitation signal.

On top of the non-parametric estimate, a parametric transfer function model could be estimated using a weighted least squares estimator (Schoukens, Dobrowiecki, & Pintelon, 1998):

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} J_N(\boldsymbol{\theta}), \tag{4a}$$

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

$$J_N(\boldsymbol{\theta}) = \frac{1}{N}\sum_{k=1}^{N} W(k)|\hat{G}_{\text{BLA}}(k) - G_M(k,\boldsymbol{\theta})|^2. \tag{4b}$$

Here, $W(k) \in \mathbb{R}^+$ is a deterministic, $\boldsymbol{\theta}$-independent weighting sequence, $\hat{G}_{\text{BLA}}(k)$ is an approximation to the actual $G_{\text{BLA}}(k)$ as it is limited to a finite number of realisations of $U(k)$ and $Y(k)$, and $G_M(k,\boldsymbol{\theta})$ is a parametric transfer function model

$$\begin{aligned} G_M(k,\boldsymbol{\theta}) &= \frac{\sum_{l=0}^{n_b} b_l \exp\left(-j2\pi\frac{k}{N}l\right)}{\sum_{l=0}^{n_a} a_l \exp\left(-j2\pi\frac{k}{N}l\right)} \\ &= \frac{B_\theta(k)}{A_\theta(k)}, \\ \boldsymbol{\theta} &= [\, a_0 \cdots a_{n_a}\ b_0 \cdots b_{n_b}\,]^T, \end{aligned} \tag{4c}$$

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

In this work, we are not initially interested in a parametric representation of the linear block. This is due to the fact that we will use directly the estimated values of magnitude and phase at each frequency as shown in Equation (3) (see Section 3). Later on, at the final stage, a parametric re-estimation will be carried out to further enhance the performance.

2.2 Function estimation using LS-SVM

The framework of LS-SVM (Suykens et al., 2002) consists of a primal-dual formulation. Given the data set $\{\boldsymbol{u}_i, x_i\}_{i=1}^N$, the objective is to find a model:

$$\tilde{x} = \boldsymbol{w}^T\boldsymbol{\varphi}(\boldsymbol{u}) + b \tag{5}$$

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

An optimisation problem is then formulated (Suykens et al., 2002):

$$\begin{aligned} \min_{\boldsymbol{w},b,e} \quad & \tfrac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + \tfrac{\gamma}{2}\sum_{i=1}^{N} e_i^2 \\ \text{subject to } & x_i = \boldsymbol{w}^T\boldsymbol{\varphi}(\boldsymbol{u}_i) + b + e_i,\ i = 1, \ldots, N \end{aligned} \tag{6}$$

with $\gamma > 0$ the regularisation 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 for this formulation are

$$\begin{cases} \frac{\partial\mathcal{L}}{\partial\boldsymbol{w}} = 0 \to \boldsymbol{w} = \sum_{i=1}^N \alpha_i\boldsymbol{\varphi}(\boldsymbol{u}_i) \\ \frac{\partial\mathcal{L}}{\partial b} = 0 \to \sum_{i=1}^N \alpha_i = 0 \\ \frac{\partial\mathcal{L}}{\partial e_i} = 0 \to \alpha_i = \gamma e_i,\ i = 1, \ldots, N \\ \frac{\partial\mathcal{L}}{\partial\alpha_i} = 0 \to x_i = \boldsymbol{w}^T\boldsymbol{\varphi}(\boldsymbol{u}_i) + b + e_i,\ i = 1, \ldots, N. \end{cases} \tag{7}$$

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

$$\left(\begin{array}{c|c} 0 & \mathbf{1}_N^T \\ \hline \mathbf{1}_N & \boldsymbol{\Omega} + \frac{1}{\gamma}\boldsymbol{I}_N \end{array}\right)\left(\begin{array}{c} b \\ \hline \boldsymbol{\alpha} \end{array}\right) = \left(\begin{array}{c} 0 \\ \hline \boldsymbol{x} \end{array}\right) \tag{8}$$

with $\boldsymbol{x} = [x_1, \ldots, x_N]^T$ and $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_N]^T$.

Through the use of Mercer's theorem (Mercer, 1909), the entries of the kernel matrix $\boldsymbol{\Omega}_{i,j}$ can be represented by $K(\boldsymbol{u}_i, \boldsymbol{u}_j) = \boldsymbol{\varphi}(\boldsymbol{u}_i)^T\boldsymbol{\varphi}(\boldsymbol{u}_j)$ with $i, j = 1, \ldots, N$. Note then that $\boldsymbol{\varphi}$ does not have to be explicitly known as this is done implicitly through the positive definite kernel function. In this paper, the Gaussian radial basis function kernel (RBF kernel) is used:

$$K(\boldsymbol{u}_i, \boldsymbol{u}_j) = \exp\left(\frac{-\|\boldsymbol{u}_i - \boldsymbol{u}_j\|_2^2}{\sigma^2}\right), \tag{9}$$

where σ is the kernel parameter.

The resulting model is then

$$\tilde{x}(\boldsymbol{u}) = \sum_{i=1}^{N} \alpha_i K(\boldsymbol{u}, \boldsymbol{u}_i) + b. \tag{10}$$

Note that as stated LS-SVM uses two parameters (i.e. γ and σ). In order to tune such parameters, Coupled Simulated Annealing (Xavier-de Souza, Suykens, Vandewalle, & Bollé, 2010) was used followed by a simplex approach to fine tune them. All of this is executed under a 10-fold cross-validation scheme (i.e. see LS-SVMlab v1.8[2]).

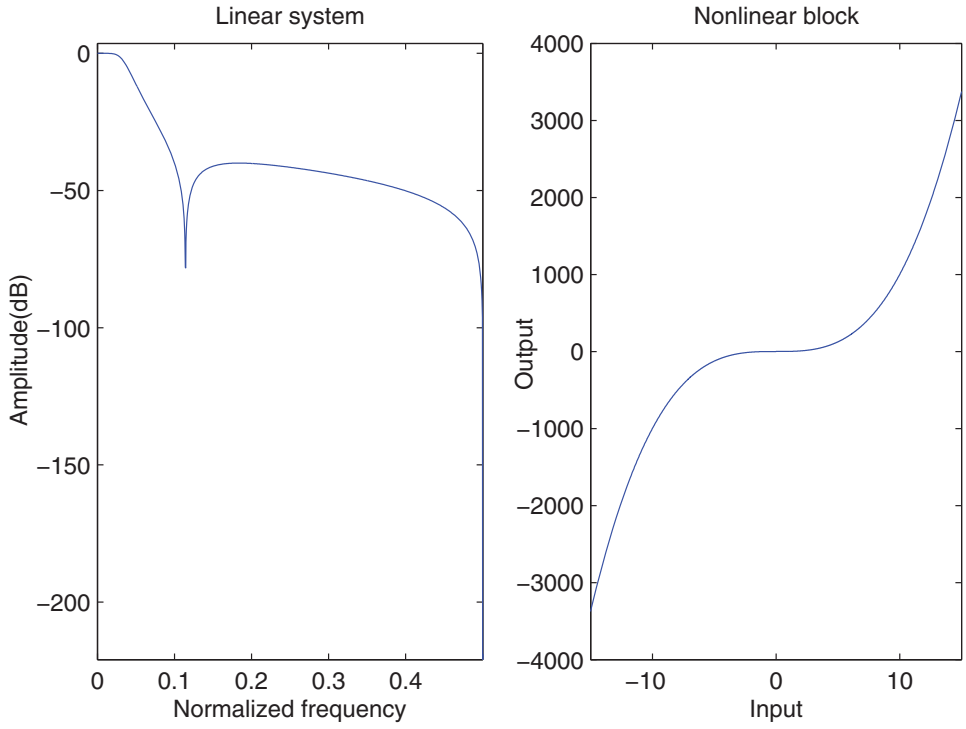

Figure 4. (Left) Linear system in normalised frequency and dB. (Right) Nonlinear block.

3. Proposed method

3.1 Inversion in the frequency domain

When dealing with system identification of Hammerstein systems it would be desirable to be able to use $y(t)$, the output of the system, and the inverse of the estimated linear block to obtain an approximation to the intermediate variable, i.e. $\hat{x}(t)$. However, it is often the case that this inversion is not straightforward. For instance, if the identified linear system is parametric and has zeros outside the unit circle (i.e. for the discrete-time case), when the system is inverted, those zeros would become poles and the resulting system would be unstable, thus making the operation unfeasible.

Once an approximation to the linear block $G_{\text{BLA}}(k)$ is obtained (see Section 2.1), it is possible to use it to obtain an approximation to the intermediate variable $\hat{X}(k)$. To do this, first $G_{\text{BLA}}(k)$ is inverted at each frequency:

$$G_{\text{BLA}}^{-1}(k) = \frac{1}{|G_{\text{BLA}}(k)|} \exp\left(-\Theta_{G_{\text{BLA}}(k)} j\right) \tag{11}$$

with $|G_{\text{BLA}}(k)|$ the magnitude and $\Theta_{G_{\text{BLA}}(k)}$ the phase of $G_{\text{BLA}}(k)$ at each frequency. From this representation it is clear that the proposed method will be able to invert linear blocks even if they would result in unstable systems if inverted in the time domain. This nice property comes from the fact that this inversion is done in the frequency domain.

In a Bode plot, this would look as the example presented in Figure 3 where the product of the magnitudes is one and the sum of the phases is zero. Note that the sudden perturbation from 33% of the sampling frequency is nothing more than an artifact due to the chosen excitation signal.

Once $G_{\text{BLA}}^{-1}(k)$ is obtained, and having $Y(k)$ (i.e. the representation of the known $y(t)$ in the frequency domain), their product will result in $\hat{X}(k)$

$$\hat{X}(k) = \frac{|Y(k)|}{|G_{\text{BLA}}(k)|} \exp\left(\left(\Theta_{Y(k)} - \Theta_{G_{\text{BLA}}(k)}\right) j\right) \tag{12}$$

with $|Y(k)|$ the magnitude and $\Theta_{Y(k)}$ the phase of $Y(k)$ at each frequency k.

From $\hat{X}(k)$, its corresponding time domain representation $\hat{x}(t)$ could be recovered and with it, the identification of the nonlinear block, through LS-SVM in this case, could proceed. Note that this would be possible due to the fact that the input to the nonlinear block $u(t)$ is known and an approximation to its output $\hat{x}(t)$ would be available.

It is evident from the multiplication in Equation (12) that whatever noise is contained in $Y(k)$ will be propagated into this intermediate variable estimation. Normally this would be a problem and obviously is an undesired effect. In fact, $\hat{x}(t)$ will be in general a poor approximation to the actual $x(t)$ since beside the back-propagated noise it is only estimated in the frequency

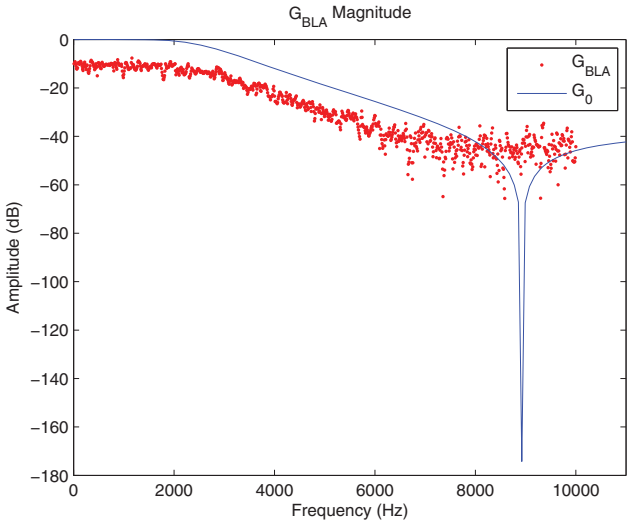

Figure 5. Non-parametric BLA.

band where $G_{\text{BLA}}(k)$ was obtained. However, it is of paramount importance to highlight the fact that $\hat{x}(t)$ is used here exclusively to train the LS-SVM model and that due to the regularisation properties of LS-SVM, the aforementioned issues can be overcome (i.e. see Section 3.2 and Figure 6).

3.2 Role of regularisation

To illustrate the effect of the regularisation, it will be shown how the changes in the noise level affect the resulting model.

Given that $\hat{x}_T(t)$ (i.e. the estimated intermediate variable of the training data) is defined as a multiplication in the frequency domain, it can be seen naturally as a convolution in the time domain. To express this, we will define $\boldsymbol{M}_{G_{\text{BLA}}^{-1}}$ as a Toeplitz matrix containing the time domain representation of the $G_{\text{BLA}}^{-1}(k)$:

$$\hat{\boldsymbol{x}}_T = \boldsymbol{M}_{G_{\text{BLA}}^{-1}} \boldsymbol{y}. \tag{13}$$

Note that here $\boldsymbol{M}_{G_{\text{BLA}}^{-1}} \in \mathbb{R}^{N\times N}$ and $\hat{\boldsymbol{x}}_T, \boldsymbol{y} \in \mathbb{R}^N$ with $\boldsymbol{y}$ the measured output and $\hat{\boldsymbol{x}}_T$ the estimated intermediate variable of the training data (i.e. it comes from $Y(k)$ and the inversion of $G_{\text{BLA}}(k)$).

Here, $\hat{\boldsymbol{x}}_T$ is used in Equation (8) instead of $\boldsymbol{x}$, therefore, from the LS-SVM formulation, we have

$$\hat{\boldsymbol{x}}_T = \left(\boldsymbol{\Omega} + \frac{\boldsymbol{I}}{\gamma}\right)\boldsymbol{\alpha} + \boldsymbol{1}_N b, \tag{14}$$

$$\tilde{\boldsymbol{x}} = \boldsymbol{\Omega}\boldsymbol{\alpha} + \boldsymbol{1}_N b. \tag{15}$$

Note that $\tilde{\boldsymbol{x}}$ corresponds to the predicted values of the intermediate variable through the evaluation of the found model on the training data.

From Equations (13) and (14) into Equation (15), we can rewrite $\tilde{\boldsymbol{x}}$ as

$$\begin{aligned}\tilde{\boldsymbol{x}} &= \hat{\boldsymbol{x}}_T - \frac{\boldsymbol{\alpha}}{\gamma} \\ &= \hat{\boldsymbol{x}}_T - \frac{1}{\gamma}\left(\boldsymbol{\Omega} + \frac{\boldsymbol{I}}{\gamma}\right)^{-1}(\hat{\boldsymbol{x}}_T - \boldsymbol{1}_N b) \\ &= \boldsymbol{M}_{G_{\text{BLA}}^{-1}}\boldsymbol{y} - \frac{1}{\gamma}\left(\boldsymbol{\Omega} + \frac{\boldsymbol{I}}{\gamma}\right)^{-1}(\boldsymbol{M}_{G_{\text{BLA}}^{-1}}\boldsymbol{y} - \boldsymbol{1}_N b).\end{aligned} \tag{16}$$

Now, if we want to see how $\tilde{\boldsymbol{x}}$ changes if the measurement noise $\boldsymbol{n}_y$ changes, we have

$$\begin{aligned}\tilde{\boldsymbol{x}} + \boldsymbol{\Delta}_{\tilde{\boldsymbol{x}}} =& \boldsymbol{M}_{G_{\text{BLA}}^{-1}}(\boldsymbol{y} + \boldsymbol{\Delta}_{n_y}) \\ &- \frac{1}{\gamma}\left(\boldsymbol{\Omega} + \frac{\boldsymbol{I}}{\gamma}\right)^{-1}(\boldsymbol{M}_{G_{\text{BLA}}^{-1}}(\boldsymbol{y} + \boldsymbol{\Delta}_{n_y}) - \boldsymbol{1}_N b) \\ =& \boldsymbol{M}_{G_{\text{BLA}}^{-1}}\boldsymbol{y} - \frac{1}{\gamma}\left(\boldsymbol{\Omega} + \frac{\boldsymbol{I}}{\gamma}\right)^{-1}(\boldsymbol{M}_{G_{\text{BLA}}^{-1}}\boldsymbol{y} - \boldsymbol{1}_N b) \\ &+ \boldsymbol{M}_{G_{\text{BLA}}^{-1}}\boldsymbol{\Delta}_{n_y} - \frac{1}{\gamma}\left(\boldsymbol{\Omega} + \frac{\boldsymbol{I}}{\gamma}\right)^{-1}\boldsymbol{M}_{G_{\text{BLA}}^{-1}}\boldsymbol{\Delta}_{n_y} \\ =& \tilde{\boldsymbol{x}} + \boldsymbol{M}_{G_{\text{BLA}}^{-1}}\boldsymbol{\Delta}_{n_y} - \frac{1}{\gamma}\left(\boldsymbol{\Omega} + \frac{\boldsymbol{I}}{\gamma}\right)^{-1}\boldsymbol{M}_{G_{\text{BLA}}^{-1}}\boldsymbol{\Delta}_{n_y}.\end{aligned} \tag{17}$$

From Equation (17) we get then

$$\boldsymbol{\Delta}_{\tilde{\boldsymbol{x}}} = \left(\boldsymbol{I} - \frac{1}{\gamma}\left(\boldsymbol{\Omega} + \frac{\boldsymbol{I}}{\gamma}\right)^{-1}\right)\boldsymbol{M}_{G_{\text{BLA}}^{-1}}\boldsymbol{\Delta}_{n_y}. \tag{18}$$

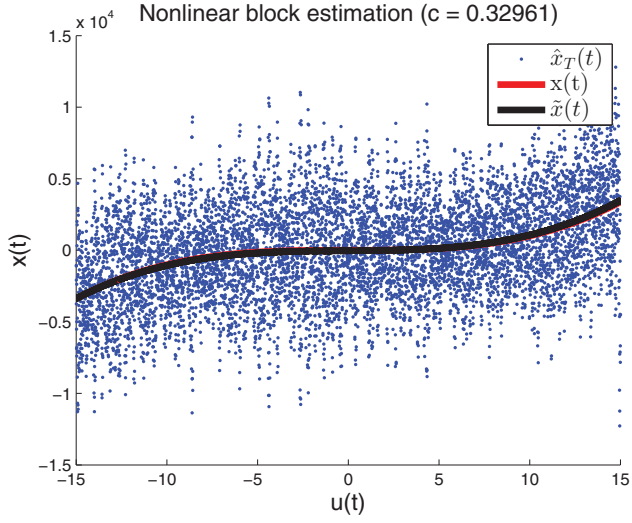

Figure 6. Nonlinear block estimation. Here, *c* corresponds to a rescaling factor. An interpolation to increase the number of points was performed and thus the extended number of points.

Now, from Equation (18) let us define

$$\boldsymbol{W} = \left(\boldsymbol{I} - \frac{1}{\gamma}\left(\boldsymbol{\Omega} + \frac{\boldsymbol{I}}{\gamma}\right)^{-1}\right). \tag{19}$$

Hence

$$\boldsymbol{\Delta}_{\tilde{\boldsymbol{x}}} = \boldsymbol{W}\boldsymbol{M}_{\boldsymbol{G}_{\mathrm{BLA}}^{-1}}\boldsymbol{\Delta}_{\boldsymbol{n}_y}. \tag{20}$$

Here, $\boldsymbol{W}$ is the matrix determining the effect of the change of the measurement noise in the model. Note also that it can be rewritten as

$$\begin{aligned}\boldsymbol{W} &= \boldsymbol{I} - (\gamma\boldsymbol{\Omega} + \boldsymbol{I})^{-1} \\ &= \boldsymbol{\Omega}\left(\boldsymbol{\Omega} + \tfrac{\boldsymbol{I}}{\gamma}\right)^{-1}.\end{aligned} \tag{21}$$

Theorem 3.1: *In Equation (20), $\boldsymbol{\Delta}_{\tilde{\boldsymbol{x}}}$ is upper bounded as follows:* $\|\boldsymbol{\Delta}_{\tilde{\boldsymbol{x}}}\|_2 \leq \left\|\boldsymbol{M}_{\boldsymbol{G}_{BLA}^{-1}}\boldsymbol{\Delta}_{\boldsymbol{n}_y}\right\|_2$

Proof: Let λ_i be an eigenvalue of $\boldsymbol{\Omega}$ which is a symmetric positive semi-definite matrix. The eigenvalues of $\gamma\boldsymbol{\Omega} + \boldsymbol{I}$ are equal to $\gamma\lambda_i + 1$.

As $\gamma\boldsymbol{\Omega} + \boldsymbol{I}$ is a square non-singular matrix, the eigenvalues of the matrix $(\gamma\boldsymbol{\Omega} + \boldsymbol{I})^{-1}$ are $\frac{1}{(\gamma\lambda_i+1)}$ and the eigenvalues of $\boldsymbol{W} = \boldsymbol{I} - (\gamma\boldsymbol{\Omega} + \boldsymbol{I})^{-1}$ are then $\left(1 - \frac{1}{(\gamma\lambda_i+1)}\right) = \frac{\gamma\lambda_i}{(\gamma\lambda_i+1)} = \frac{\lambda_i}{(\lambda_i+\frac{1}{\gamma})}$.

Given that $\boldsymbol{W}$ is symmetric, $\left(\frac{\lambda_i}{(\lambda_i+1)}\right)^2$ is an eigenvalue of $\boldsymbol{W}^T\boldsymbol{W}$.

Note that $0 \leq \frac{\lambda_i}{(\lambda_i+\frac{1}{\gamma})} \leq 1$ as by definition $\gamma > 0$ and $\lambda_i \geq 0$. Therefore, $0 \leq \left(\frac{\lambda_i}{(\lambda_i+\frac{1}{\gamma})}\right)^2 \leq 1$.

Finally, note that $\|\boldsymbol{W}\|_2 = \sqrt{\max_i\left(\frac{\lambda_i}{(\lambda_i+\frac{1}{\gamma})}\right)^2}$.

From Equation (20) we can write the following inequality: $\|\boldsymbol{\Delta}_{\tilde{\boldsymbol{x}}}\|_2 \leq \|\boldsymbol{W}\|_2\left\|\boldsymbol{M}_{\boldsymbol{G}_{\mathrm{BLA}}^{-1}}\boldsymbol{\Delta}_{\boldsymbol{n}_y}\right\|_2 \leq \left\|\boldsymbol{M}_{\boldsymbol{G}_{\mathrm{BLA}}^{-1}}\boldsymbol{\Delta}_{\boldsymbol{n}_y}\right\|_2$ as $0 \leq \|\boldsymbol{W}\|_2 \leq 1$. ■

Note that given the properties of $\boldsymbol{W}$ and the expression in Equation (20) we can expect that the effect of $\boldsymbol{\Delta}_{\boldsymbol{n}_y}$ will be generally dampened by $\boldsymbol{W}$. This is very relevant as $\boldsymbol{W}$ is heavily dependent on the regularisation term γ. See Figure 8 in Section 4.2 for an example of the behaviour of $\|\boldsymbol{W}\|_2$ and the effect of γ on it.

Let us consider the case where $\gamma \to 0$ while bearing in mind that the values of the kernel matrix $\boldsymbol{\Omega}_{i,j} \in [0, 1]\ \forall i, j$ when using the Gaussian RBF kernel, where i and j denote the row and column evaluated. Note that in this case, the weight given to the errors between the output of the training data set and the points obtained with the model is very small (see Equation (6)):

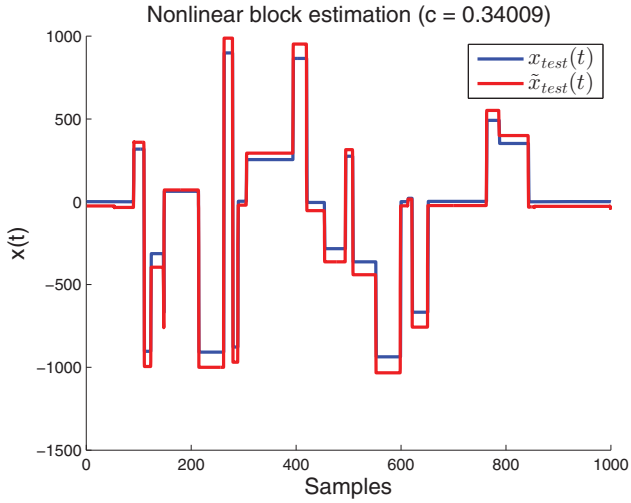

Figure 7. Comparison between the actual $x_{\text{test}}(t)$ and the estimated $\hat{x}_{\text{test}}(t)$. Here, c corresponds to a rescaling factor.

$$\begin{aligned}\boldsymbol{W} =& \boldsymbol{I} - \frac{1}{\gamma}\left(\boldsymbol{\Omega} + \frac{\boldsymbol{I}}{\gamma}\right)^{-1} \\ =& \boldsymbol{I} - \frac{1}{\gamma}\left(\frac{\gamma\boldsymbol{\Omega} + \boldsymbol{I}}{\gamma}\right)^{-1} \\ \approx& \boldsymbol{I} - \frac{1}{\gamma}\left(\frac{\boldsymbol{I}}{\gamma}\right)^{-1} = \boldsymbol{0}.\end{aligned} \tag{22}$$

Then

$$\boldsymbol{\Delta}_{\tilde{\boldsymbol{x}}} \approx \boldsymbol{0}. \tag{23}$$

Consider now the case where $\gamma \to \infty$. The errors between the training points and the points obtained with the model are extremely important (again, see Equation (6)). In other words, the estimated model would try to follow the output of the training data set as well as possible.

$$\boldsymbol{W} = \boldsymbol{I} - \frac{1}{\gamma}\left(\boldsymbol{\Omega} + \frac{\boldsymbol{I}}{\gamma}\right)^{-1} \approx \boldsymbol{I} - \frac{\boldsymbol{\Omega}^{-1}}{\gamma} \approx \boldsymbol{I}. \tag{24}$$

Then

$$\boldsymbol{\Delta}_{\tilde{\boldsymbol{x}}} \approx \boldsymbol{M}_{\boldsymbol{G}_{\mathrm{BLA}}^{-1}}\boldsymbol{\Delta}_{\boldsymbol{n}_y}. \tag{25}$$

This result was to be expected since the model will try to follow the output of the training data set. Any change in the training points will result in a direct change in the behaviour of the model as there is no regularisation at all.

It is important to remember that the Toeplitz matrix $\boldsymbol{M}_{\boldsymbol{G}_{\mathrm{BLA}}^{-1}}$ is used such that a convolution in the time domain

can be represented. If the expression in Equation (25) is taken to the frequency domain, it simply becomes a multiplication:

$$\Delta_{\tilde{X}}(k) \approx G_{\mathrm{BLA}}^{-1}(k)\Delta_{N_Y}(k). \tag{26}$$

From Equation (26) it is evident that the way $\Delta_{N_Y}(k)$ is going to affect $\Delta_{\tilde{X}}(k)$, or equivalently how $\boldsymbol{\Delta}_{\boldsymbol{n_y}}$ affects $\boldsymbol{\Delta}_{\tilde{\boldsymbol{x}}}$, depends directly on the frequency power distributions of $G_{\mathrm{BLA}}^{-1}(k)$ and $\Delta_{N_Y}(k)$ as there would be no regularisation in use. This means that in the worst case scenario, most of the power of $\Delta_{N_Y}(k)$ would be in the passband of $G_{\mathrm{BLA}}^{-1}(k)$. In this case, the effect of $\Delta_{N_Y}(k)$ would be amplified. It is possible, on the other hand, that most of the power of $\Delta_{N_Y}(k)$ is out of the passband of $G_{\mathrm{BLA}}^{-1}(k)$. In this case, the effect of $\Delta_{N_Y}(k)$ would be dampened.

3.3 Method summary

The proposed method finally consists of four parts. First, the BLA of the system (i.e. a non-parametric $G_{\mathrm{BLA}}(k)$) is calculated. Second, using $G_{\mathrm{BLA}}(k)^{-1}$ and the frequency domain representation of $y(t)$ (i.e. $Y(k)$) an approximation to the intermediate variable $\hat{X}_T(k)$ is obtained and from it, its corresponding time domain representation $\hat{x}_T(t)$. Next, a LS-SVM model is trained using $u(t)$ as input and $\hat{x}_T(t)$ as output. Finally, the linear block is re-estimated using a newly calculated intermediate variable (i.e. the result from applying an input to the estimated nonlinear block) and the corresponding known output. A summary of the method presenting the main steps is shown in Algorithm 1. Note that for each of the data sets mentioned, several realisations are used (see Section 4.2).

It is important to highlight that in Algorithm 1, steps 1–6 correspond to the system identification part, while steps 7– 9 correspond to the evaluation part.

4. Experimental results

In this section, a practical illustration of the proposed method is presented through synthetic examples. First, the specific characteristics of the employed signals will be described. Then, a didactical example will be offered where the steps of the proposed method described in Algorithm 1 are shown. Afterwards, the effects of the noise on the proposed method are illustrated in the previous example and in a new example which includes a hard to model nonlinearity. Finally, we offer a third example based on a real-life application and provide a performance comparison with different methods for the three examples.

Algorithm 1. Hammerstein system identification through BLA inversion and LS-SVM techniques.

Input: Random-phase multisine signal $u_1(t)$ and its corresponding output $y_1(t)$; ramp signal $u_2(t)$ and its corresponding output $y_2(t)$. Multilevel Pseudo Random Signal $u_3(t)$ and its corresponding output $y_3(t)$. Multilevel Pseudo Random Signal $u_{\mathrm{test}}(t)$.

Output: Evaluation of the test output signal $y_{\mathrm{test}}(t)$;

1: Eliminate the first data points in $u_1(t)$ and $y_1(t)$ to reduce the effect of the transients.
2: Get the frequency representation of the transient free signals (i.e. $U_1(k)$ and $Y_1(k)$).
3: Obtain $G_{\mathrm{BLA}}(k)$ and $G_{\mathrm{BLA}}^{-1}(k)$ from several realisations of $U_1(k)$ and $Y_1(k)$.
4: Get an approximation to the intermediate variable in the frequency domain: $\hat{X}_T(k) = G_{\mathrm{BLA}}^{-1}(k)Y_2(k)$.
5: Get $\hat{x}_T(t)$, the time domain representation of $\hat{X}_T(k)$, and together with $u_2(t)$ train a LS-SVM model to represent the nonlinearity;
6: Apply $u_3(t)$ to the estimated nonlinear model. Using the resulting $\tilde{x}_3(t)$ and the known output $y_3(t)$, estimate the linear block $G_{\mathrm{final}}(k)$;
7: Apply $u_{\mathrm{test}}(t)$ to the estimated nonlinear model. From the result $\tilde{x}_{\mathrm{test}}(t)$ obtain its frequency representation. Apply $\tilde{X}_{\mathrm{test}}(k)$ to $G_{\mathrm{final}}(k)$ to obtain the estimation of the output in the frequency domain $\hat{Y}_{\mathrm{test}}(k)$.
8: Finally, from $\hat{Y}_{\mathrm{test}}(k)$ obtain the corresponding time domain representation $\hat{y}_{\mathrm{test}}(t)$;
9: **return** $\hat{y}_{\mathrm{test}}(t)$;

Note that the particulars (e.g. amplitudes, frequency bands, sampling frequencies, etc.) of the signals used were picked for the corresponding examples and are not intended as general recommendations for the proposed method.

4.1 Signals description

In this section, the signals mentioned in Algorithm 1 are further explained.

Signal $u_1(t) = \sum_{k=1}^{F} |U_k| \cos(2\pi k \frac{f_s}{N} t + \phi_k)$ is a full random-phase multisine with a random harmonic grid in the excited frequency range {0, 10, 000}Hz. Note that here, F stands for the excited frequencies, N is the number of samples and $|U_k|$ denotes the amplitude used at each frequency. The sampling frequency f_s is equal to 78, 125Hz. In this case, all $|U_k| \neq 0$ are chosen equal to each other such that $u_1(t)$ has an rms value of 0.3. The phases ϕ_k are uniformly and randomly distributed between 0 and 2π. A ramp signal with 45 degrees slope was used to generate $u_2(t)$. Finally, Multilevel Pseudo Random Signals with 2% switching probability and amplitude values

Table 1. Summary of input signals used.

Name	Points used	Max. Value	Min. Value	Description
$u_1(t)$	5000	–	–	Random-phase multisine signal.
$u_2(t)$	1000	15	−15	Ramp signal with 45 degrees slope.
$u_3(t)$	1000	10	−10	Multilevel Pseudo Random Signal.
$u_{test}(t)$	1000	10	−10	Multilevel Pseudo Random Signal.

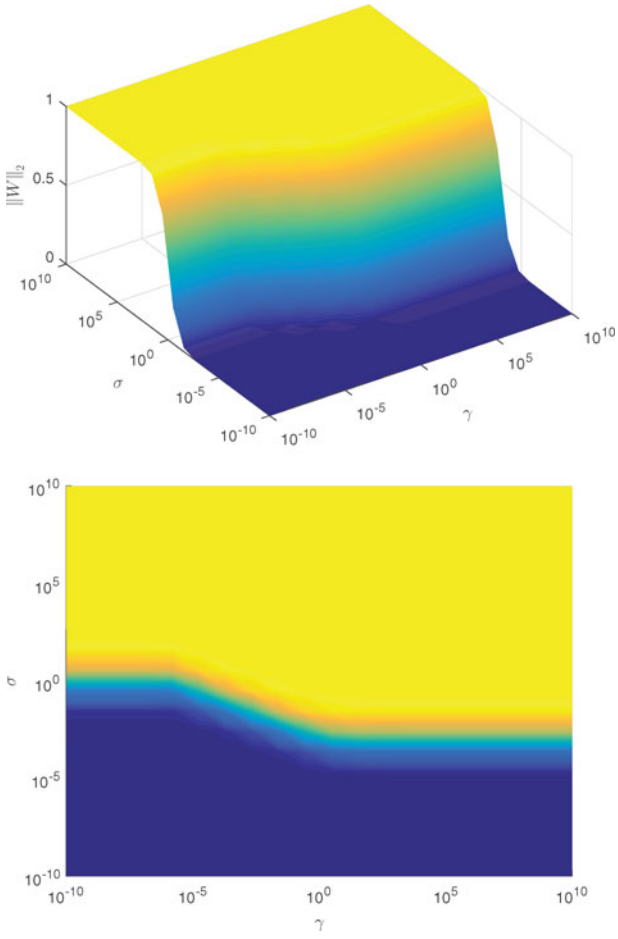

Figure 8. $\|W\|_2$ is calculated for a wide range of σ and γ. (Top) As can be seen, the values are bounded between 0 and 1. (Bottom) Upper view of the top figure.

drawn from a uniform distribution were used to generate $u_3(t)$ and $u_{\text{test}}(t)$.

A summary and further details describing the used signals can be found in Table 1.

It is important to note that even though the signals have the number of points noted in Table 1, it is possible to perform an interpolation of such signals to have more points.

4.2 Method steps

In this section, the proposed methodology was applied to one system in the discrete-time domain. This system is of didactical nature and is used to illustrate the method steps. It was generated through a nonlinear block:

$$x(t) = u(t)^3 \text{ with } u(t) \in [-15, 15], \tag{27}$$

and a linear block:

$$y(t) = \frac{B_1(q)}{A_1(q)} x(t) \tag{28}$$

where

$$B_1(q) = 0.008935q^3 - 0.004525q^2 - 0.004525q + 0.008935 \tag{29}$$

and

$$A_1(q) = q^3 - 2.564q^2 + 2.218q - 0.6456. \tag{30}$$

The linear and nonlinear blocks in this example (later on referred to as Example 1) are depicted in Figure 4. The Linear Time Invariant (LTI) block was chosen with a sharp zero in order to illustrate how this affects the method.

White Gaussian noise with zero mean was applied to the output of the system with a Signal to Noise Ratio (SNR) of 10 dB. For each of the stages in the procedure, two periods and five realisations of the signals were used.

First, a non-parametric $G_{\text{BLA}}(k)$ was estimated as shown in Section 2.1 through the use of a random- phase multisine input and its corresponding output (see Figure 5 for the results of this step in the example). Afterwards, $G_{\text{BLA}}^{-1}(k)$ was calculated as explained in Section 3. Next, the estimation of the intermediate variable $\hat{x}_T(t)$ was carried out. Then, the nonlinear block was estimated through the use of LS-SVM using $u_2(t)$ as input and $\hat{x}_T(t)$ as output. Note that this $u_2(t)$ still corresponds to the same signal used to estimate $\hat{x}_T(t)$. Figure 6 displays the results of this step.

Note that there is a rescaling of the approximated nonlinear block before overlapping it to the actual one. The actual difference in scaling has no effect on the input–output behaviour of the Hammerstein system (i.e. any pair of $\{f(u(t))/\eta, \eta G(q)\}$ with $\eta \neq 0$ would yield identical input and output measurements). Other than this rescaling, it is evident that the reproduction of the nonlinear block is quite accurate. In fact, in Figure 7, the comparison between the actual $x_{\text{test}}(t)$ and the estimated $\tilde{x}_{\text{test}}(t)$ is shown. This comparison is carried out for illustrative purposes and does not have any role at all in the procedure. As can be seen, the reconstruction is accurate up to a scaling factor.

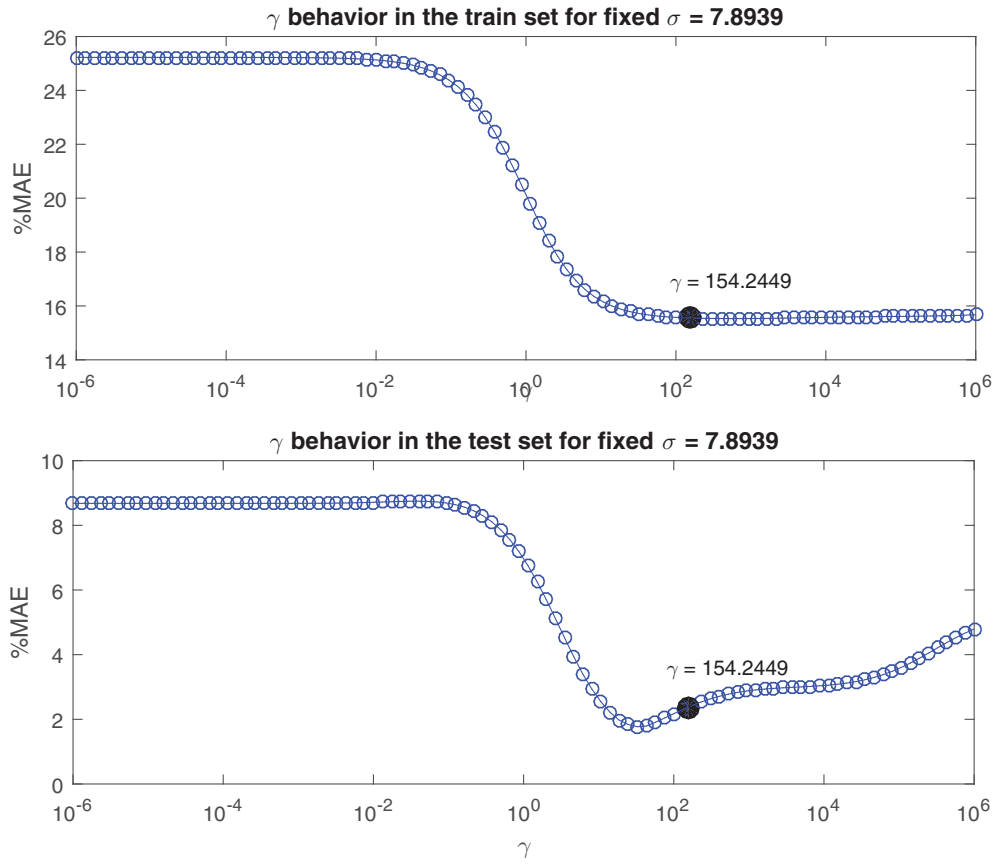

Figure 9. σ is fixed while a wide range of γ values is evaluated. The black point corresponds to the chosen value of γ.

In Figure 8, the behaviour of $\|\boldsymbol{W}\|_2$ is displayed for different values of σ and γ. As can be seen, $\|\boldsymbol{W}\|_2$ is bounded between 0 and 1 as explained in Theorem 3.1.

Also, in order to show the effect of parameter tuning during the modelling of the nonlinear block, Figures 9 and 10 are presented. In these figures, alternatively one of the parameters (i.e. σ or γ) is fixed while the result of different values for the other are tried at the training and test set. The fixed values correspond to the selected parameters through coupled simulated annealing (Xavier-de Souza et al., 2010) followed by a simplex approach for fine tuning under a 10-fold cross-validation scheme (i.e. see LS-SVMlab v1.8).

As can be seen, neither of the parameters is selected at the corresponding lowest error in the training set, however, in the test set these parameters prove to be very effective. Note that even though it is not possible to guarantee the absolute minimum error in the test set, a very good result is obtained.

Having performed the previous steps, the last remaining thing to do is to re-estimate the linear block. To do this, a least squares (LS) approach is used. The result can be seen in Figure 11, and in Figure 12 the final estimation for the test set is shown.

4.3 Noise effect analysis

For evaluating how the noise affects the performance of the method, 100 Monte Carlo simulations were carried out in a test set for each of four different levels of SNR.

In order to be able to make a comparison between the results, let us have the normalised Mean Absolute Error (MAE) defined as shown in Equation (31) for a signal with N measurements. Note that the normalised MAE uses the noise-free signal $y_{\text{test}}(t)$:

$$\%\text{MAE} = \frac{100}{N}\frac{\sum_{t=1}^{N}\left|y_{\text{test}}(t)-\hat{y}_{\text{test}}(t)\right|}{\left|\max(y_{\text{test}}(t))-\min(y_{\text{test}}(t))\right|}. \quad (31)$$

The results of the 100 Monte Carlo simulations are summarised in Figure 13.

As can be seen, the performance of the model is consistent independently of the level of noise (i.e. the medians of the different simulations oscillate in a small range between 2.8101% and 3.4928%). This phenomenon can be explained by the particular shape of the linear block in this system. As can be seen in Figures 5 and 11, the zero at 8926 Hz is not modelled properly due to the high level of

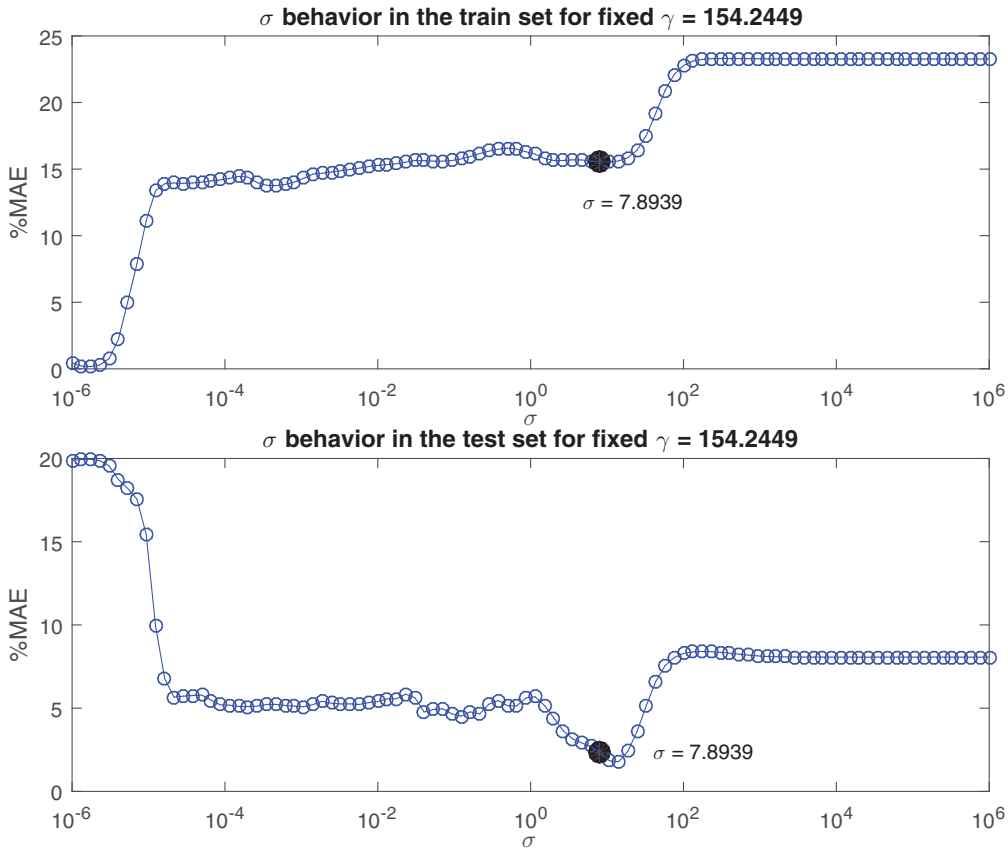

Figure 10. γ is fixed while a wide range of σ values is evaluated. The black point corresponds to the chosen value of σ.

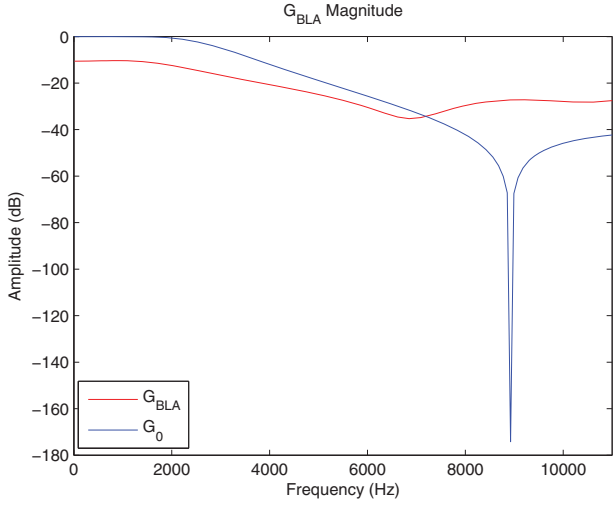

Figure 11. Estimated parametric BLA.

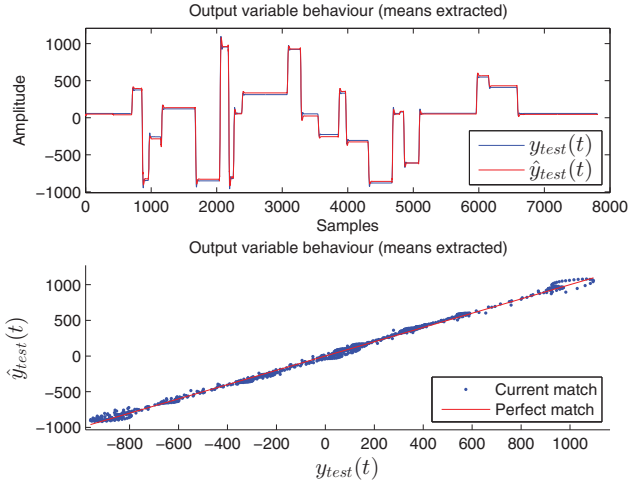

Figure 12. (Top) $\hat{y}_{\text{test}}(t)$ (rescaled) is superimposed to the actual $y_{\text{test}}(t)$. (Bottom) A scatter plot between $\hat{y}_{\text{test}}(t)$ and $y_{\text{test}}(t)$.

noise present. In Figure 14, the same modelling is shown for different SNR levels.

It is clear that as the level of noise decreases, the modelling of the zero improves. This better modelling has a drawback for the proposed methodology: when the model is inverted, a pole will appear at the same frequency where the original zero was. In other words there is a trade-off in this example: on one hand, the higher the noise the more difficult it is to obtain a good modelling. On the other hand, the lower the noise, the more problematic the zero at 8926 Hz becomes. Surprisingly, the

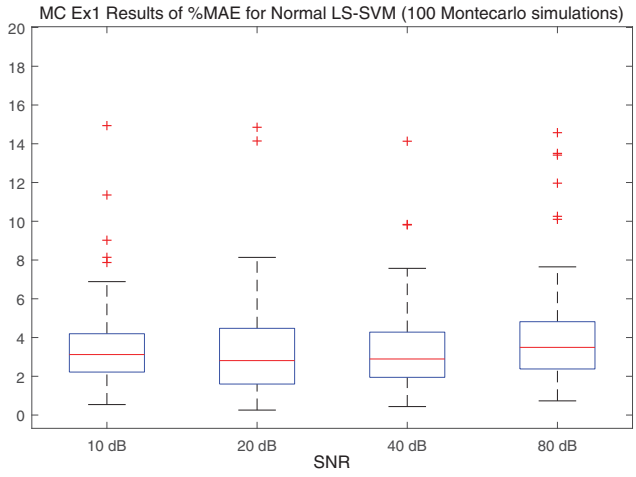

Figure 13. Normalised mean absolute error for a 100 Monte Carlo simulation for different levels of noise in Example 1.

noise itself acts as a sort of protection when modelling systems with this type of zeros.

In order to offer another perspective on the way the proposed methodology works, the following example (i.e. Example 2) was used with nonlinear block:

$$x(t) = u(t)\cos(u(t)) \text{ with } u(t) \in [-15, 15], \quad (32)$$

and linear block

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

where

$$B_2(q) = 0.004728q^3 + 0.01418q^2 + 0.01418q + 0.004728 \quad (34)$$

and

$$A_2(q) = q^3 - 2.458q^2 + 2.262q - 0.7654. \quad (35)$$

The linear and nonlinear blocks of Example 2 are depicted in Figure 15. Note that the nonlinearity in this example is particularly difficult to model with polynomial basis function approaches.

For this example, again 100 Monte Carlo simulations were run and the corresponding results are presented in Figure 16.

It can be seen that the results here are more intuitive in the sense that as the SNR increases, the normalised MAE decreases.

4.4 Methods comparison

In order to compare the current method, four additional methodologies were considered. The compared methods include:

- Inversion + LS-SVM (i.e. the presented method).
- NARX LS-SVM (Suykens et al., 2002) with 10 lags of input and 10 lags of output.
- The Hammerstein and Wiener identification procedure (in this paper denoted by WHIP) presented in Schoukens (2015).
- The iterative method (in this paper denoted by IM) presented in Bai and Li (2010).
- The state-dependent parameter (SDP) method in combination with the RIVBJ routine contained in the CAPTAIN toolbox (Young, 2000; Young, McKenna, & Bruun, 2001). This will be referred to as the Captain method.

To carry out the comparison, a third example of a more realistic nature is introduced (from now on referred to as Example 3). This example models a push–pull type B amplifier as depicted in Figure 17. To model the speaker, its electrical and mechanical dynamics were considered. The way the speaker is modelled is represented in Figure 18 where the input is the input voltage from the amplifier (i.e. $v(t)$) and the output is the displacement of the cone of the speaker $x_d(t)$. For this example, the number of realisations used was 50.

The system is represented as shown in Figure 19 where the nonlinear block is a piecewise function with saturations given by the positive and negative supply rails (i.e. see Figure 17) and a deadzone between -0.55 and 0.55 V generated by the transistors. The speaker, on the other hand, is modelled through a transfer function (in continuous time) as shown in Equation (36) (Ravaud, Lemarquand, & Roussel, 2009):

$$\frac{X_d(s)}{V(s)} = \frac{k_m}{(Ls+R)(ms^2+bs+k) - k_v k_m s}, \quad (36)$$

where $k_m = 3.14$ N/A is the constant from the Lorentz force, $k_v = 3.14$ Vs/m is the constant of the electromotive force (i.e. EMF), $k = 20000$ N/m is the spring constant, $b = 50$ N/m/s is the dampener constant, $m = 0.004$Kg is the mass of the cone and the coil, $R = 5\Omega$ is the resistance and $L = 50\mu$H the inductance of the speaker. The transfer function was then discretised with the zero-order

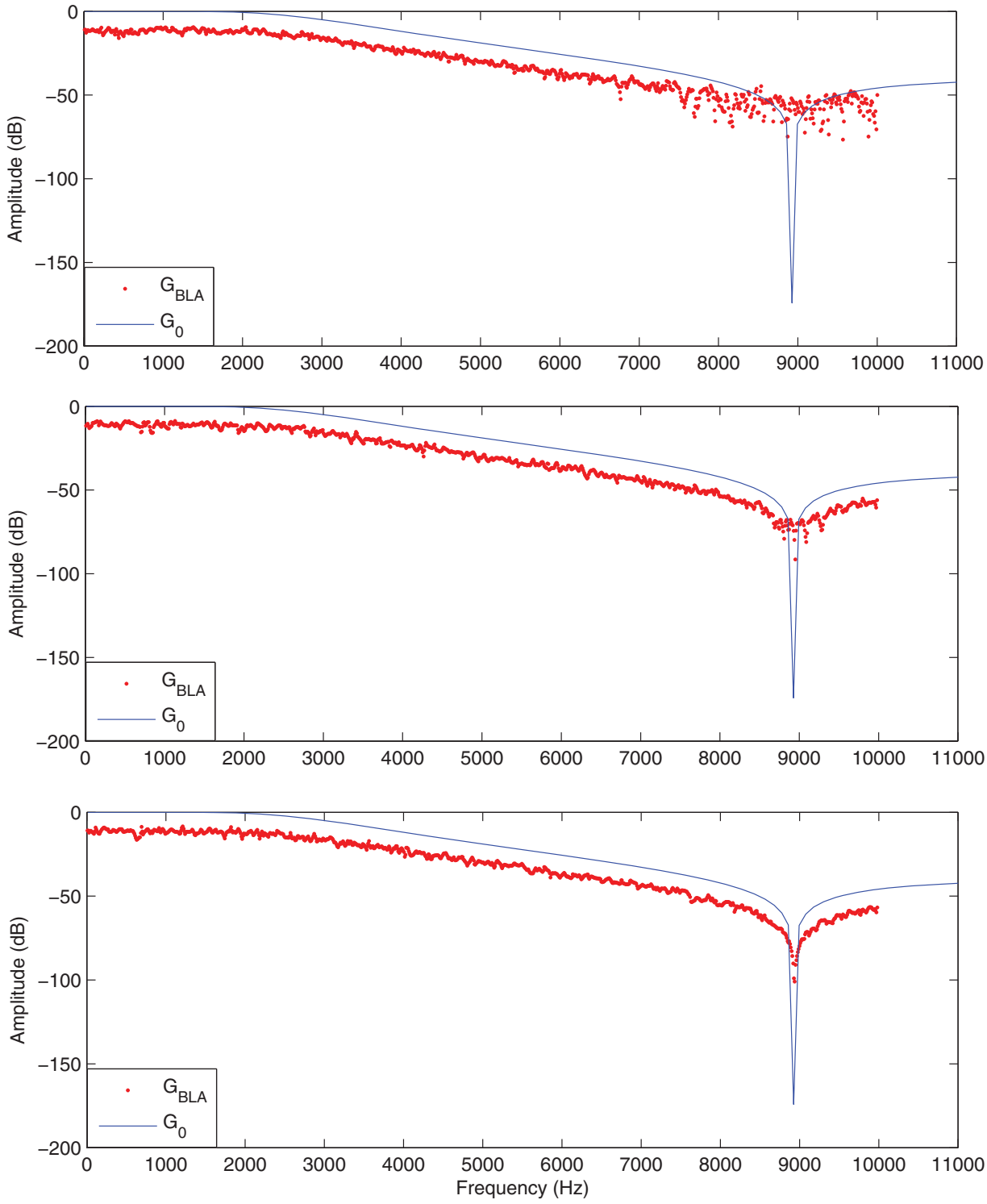

Figure 14. Non-parametric BLA for different noise levels: (Top) SNR = 20 dB. (Centre) SNR = 40 dB. (Bottom) SNR = 80 dB.

hold method with a sampling time $Ts = 0.0004$ s producing finally

$$x_d(t) = \frac{0.003639q^2 + 0.0009408q + 9.87 \times 10^{-08}}{q^3 - 0.8595q^2 + 0.005371q - 4.302 \times 10^{-20}} \upsilon(t). \tag{37}$$

For Examples 1 and 2 the inputs of the test set range between -10 and 10. For Example 3, the test set range of the inputs goes from -20 to 20V.

To train the NARX LS-SVM method, uniformly distributed white noise inputs were used. In Examples

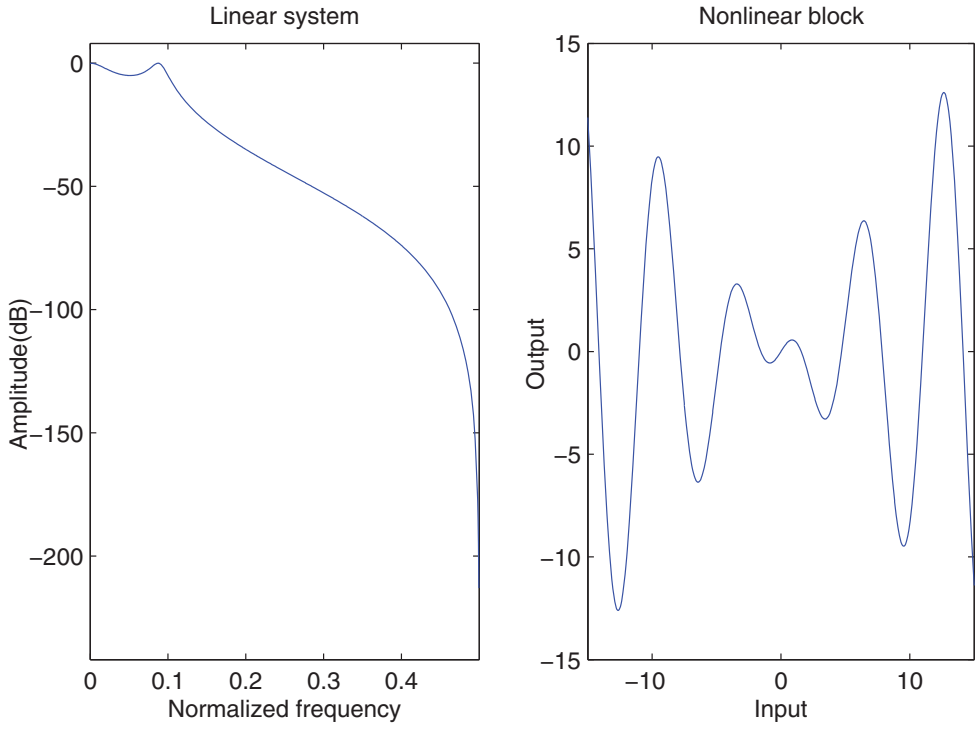

Figure 15. (Left) Linear system in normalised frequency and dB. (Right) Nonlinear block.

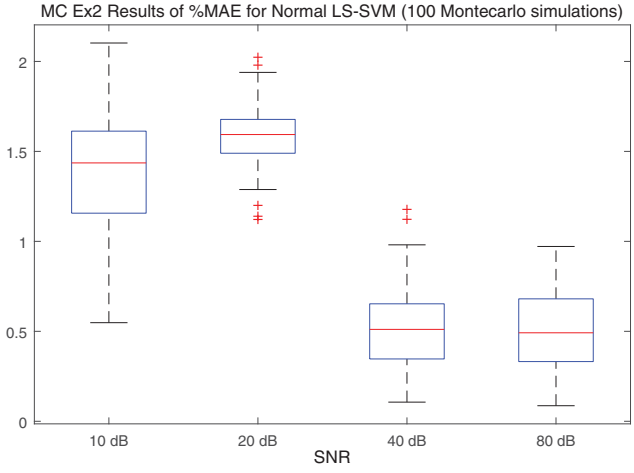

Figure 16. Normalised mean absolute error for a 100 Monte Carlo simulation for different levels of noise in Example 2.

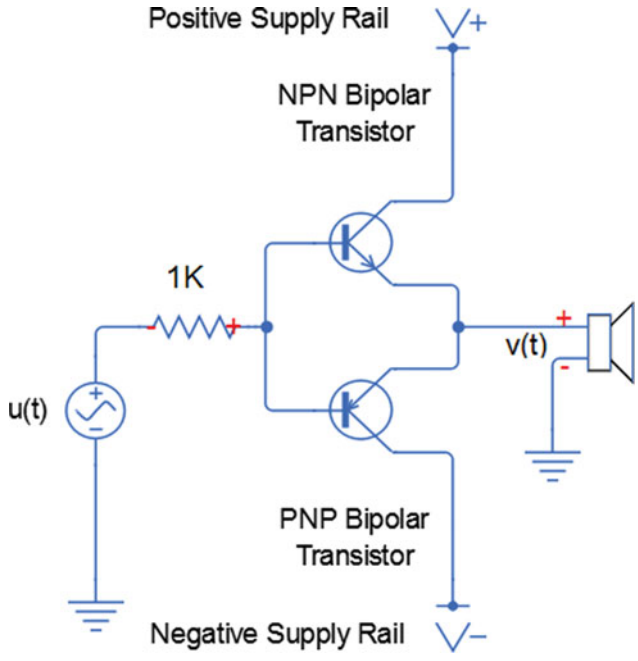

Figure 17. Example 3. Push–pull type B amplifier.

1 and 2, $u_{\mathrm{NARX}}(t)$ covers at least $[-15, 15]$ while in Example 3, the amplitude of the input $u_{\mathrm{NARX}}(t)$ covered $[-20, 20]$.

For the training of the Captain method, uniformly distributed white noise inputs were used. In Example 2, $u_{\mathrm{CAP}}(t)$ was scaled so that it covered at least $[-15, 15]$ while in Example 3, the amplitude of the input $u_{\mathrm{CAP}}(t)$ covered $[-20, 20]$. Linear model orders between 2 and 4 were scanned and nonlinear degrees between 3 and 25 for the first and second examples and between 3 and 15 for the third one were scanned. The specific functions employed were `sdp` and `rivbj`.

To train the IM method, Gaussian noise inputs were used. In the first example, $u_{\mathrm{IM}}(t)$ had a Root Mean Square (RMS) value of 0.3. In the second example, the amplitude of the Gaussian noise input $u_{\mathrm{IM}}(t)$ was rescaled so that it covers at least $[-15, 15]$. Similarly, in Example 3 $u_{\mathrm{IM}}(t)$

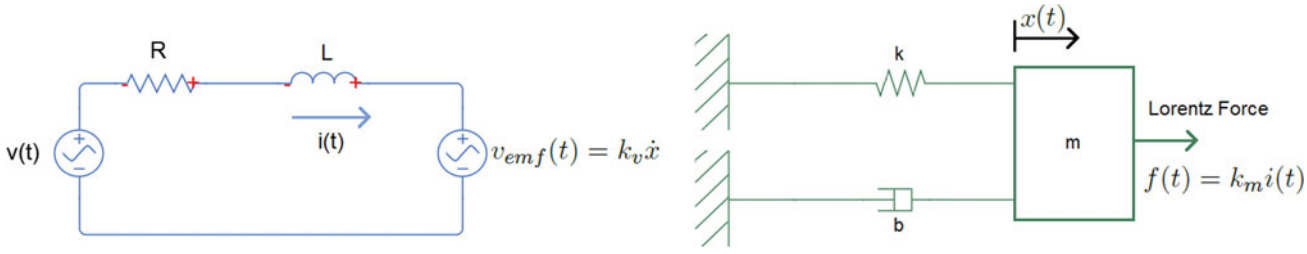

Figure 18. Example 3. Speaker modelling. (Left) Electrical dynamics. (Right) Mechanical dynamics.

covers at least $[-20, 20]$. This was done to avoid extrapolation issues as the nonlinearity in Equation (32) is not in the polynomial model class. These signals were used to estimate iteratively the linear and the nonlinear blocks. Linear model orders between 2 and 4 were scanned. Also, nonlinear degrees between 3 and 5 for the first example, between 3 and 25 for the second one and between 3 and 21 for Example 3 were scanned. The linear model order and the nonlinear degree are chosen simultaneously using cross-validation with the Multilevel Pseudo Random validation signal $u_3(t)$.

For the training of the WHIP case, for Example 1 multisine signals with an RMS value of 0.3 were used as inputs. For Example 2, $u_{\text{WHIP}}(t)$ covered at least $[-10, 10]$. For Example 3, $u_{\text{WHIP}}(t)$ covered $[-20, 20]$. Two steady-state periods of four out of the five phase realisations of the multisine were used for estimation of the linear and the nonlinear blocks. One steady-state period of the fifth realisation of the multisine was used for model order selection of the linear block (i.e. the model order was chosen between 2/2, 2/3, 2/4, 3/3, 3/4 and 4/4). A Multilevel Pseudo Random Validation signal $u_3(t)$ was used for selecting the nonlinear degree (i.e. between 3 and 5 for the first example and between 3 and 25 for the second and third ones). After following the steps above, the obtained model was optimised using the four phase realisations of the multisine that were used earlier to estimate the linear and the nonlinear block.

In Table 2, the results of the comparison in normalised MAE form (see (31)) are presented. Each of the presented results corresponds to an average over 10 runs.

The IM method seems to be very sensitive to the application of noise. Given that the selected examples contain a low SNR, the performance of the method is severely affected. However, this method obtains better results if the noise is much smaller (e.g. for 80 dB SNR a normalised MAE of 0.00063% is obtained in Example 1).

The IM, Captain and WHIP methods assume that the nonlinearity can be represented in a basis function expansion form with known basis functions. Note that

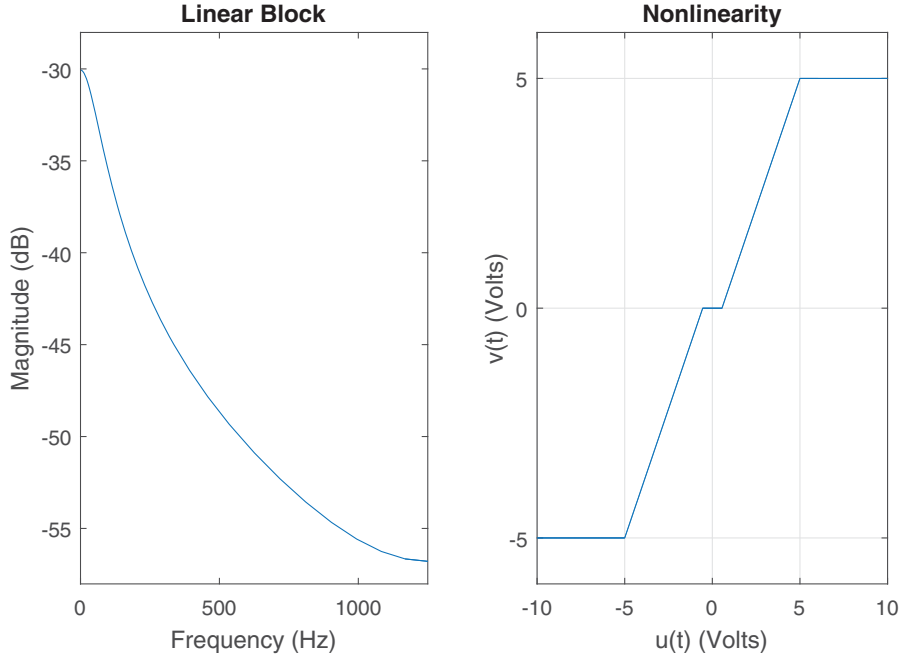

Figure 19. Example 3. (Left) Linear system in normalised frequency and dB. (Right) Nonlinear block. For visualisation purposes it is only displayed from -10 to 10V while the actual range is from -20 to 20V.

Table 2. Results comparison in Normalised MAE.

	Example 1		Example 2		Example 3	
SNR (dB)	10	20	10	20	10	20
Inversion + LS-SVM	4.3331	4.3735	**0.89992**	**0.76626**	**0.82356**	**0.31645**
NARX LS-SVM	6.2704	11.8499	20.8274	27.0158	3.9029	2.4138
IM	6.6991	6.7313	5.6814	6.1720	1.8026	2.6827
WHIP	**0.0676**	**0.0516**	5.0310	4.9834	1.5271	1.5004
Captain	3.8758	3.4821	12.815	12.1707	26.1013	25.4655

the nonlinearities in Equation (32) and Figures 15 and 19 are hard to model by a polynomial of reasonable degree for the input ranges used (i.e. see Table 1). Nevertheless, polynomials were used for these methods.

Given how sharp the zero at 8926 Hz is in the linear block of Example 1, it is clear that this is a particularly challenging problem for the proposed methodology due to the inversion step. Nonetheless as can be seen, the proposed method performs well in the three examples. Even more, for the second and third examples it achieves a better performance than the other methods. The results presented suggest that the proposed method is robust against the amount of noise used and also has a great generalisation capability when using different model classes, which clearly is a nice advantage.

It is important to highlight that in simpler cases, like the one presented in Example 1, even though the proposed method works well other methods could be better specially when they belong to the model class. For cases where the nonlinearity becomes challenging (i.e. hard nonlinearities), the proposed method becomes an attractive option due to its flexibility and good performance.

5. Conclusions

The described method presents a combination of techniques, namely the BLA and LS-SVM, for the identification of Hammerstein systems.

It is shown that the inversion of the estimated linear block can be used to make a preliminary estimation of the intermediate variable even in the presence of measurement noise. With this preliminary estimation and the known input the nonlinear block can be modelled. This is, as long as there is a mechanism to counter the influence of the back-propagated noise. In this paper, the regularisation provided by the LS-SVM methodology provides such a tool. Once this modelling is done, the intermediate variable can be re-estimated straightforwardly.

In Hammerstein systems the estimated intermediate variable, in combination with the known input, is enough to model the nonlinear block. Similarly, the intermediate variable in combination with the output variable can be used to model the linear block. Given this, the proposed method allows us to obtain a model for each of the composing blocks of the Hammerstein system (up to certain scaling factor) that can reproduce the input–output dynamics in an accurate way. This allows a deeper insight into the inner workings of the studied Hammerstein systems.

In this work, a method is offered which offers high flexibility with regard to the model class of the nonlinearity it can handle. Furthermore, when dealing with hard nonlinearities the presented method tends to perform better than other state of the art methods thanks to the generalisation and regularisation capabilities of LS-SVM.

Further extensions of the methodology could be achieved through its application to the Multiple-Input Multiple-Output (MIMO) case of Hammerstein systems and to Wiener–Hammerstein systems.

Notes

1. A PISPO system is a system that when excited with a periodic signal $u(t)$ produces a periodic output $y(t)$ (in steady-state) with the same period length as $u(t)$. Hammerstein systems are included in the class of PISPO systems.
2. http://www.esat.kuleuven.be/sista/lssvmlab/

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

European Research Council [290923, 320378]. FWO [G.0377.12, G.088114N, Methusalem 1]. IWT [SBO 100031]. Belgian Federal Science Policy Office [IUAP P7/19]. Research Council KU Leuven [GOA/10/09, CoE PFV/10/002, BIL12/11T].

ORCID

Ricardo Castro-Garcia http://orcid.org/0000-0002-2332-9381

References

Bai, E. W. (1998). An optimal two stage identification algorithm for Hammerstein-Wiener nonlinear systems. *Automatica, 34*, 333–338.

Bai, E. W. (2003). Frequency domain identification of Hammerstein models. *IEEE Transactions on Automatic Control*, *48* (4), 530–542.

Bai, E. W. (2004). Decoupling the linear and nonlinear parts in Hammerstein model identification. *Automatica, 40* (4), 671–676.

Bai, E. W., & Fu, M. (2002). A blind approach to Hammerstein model identification. *IEEE Transactions on Signal Processing, 50* (7), 1610–1619.

Bai, E. W., & Li, K. (2010). Convergence of the iterative algorithm for a general Hammerstein system identification. *Automatica, 46* (11), 1891–1896.

Billings, S. A., & Fakhouri, S. Y. (1982). Identification of systems containing linear dynamic and static nonlinear elements. *Automatica, 18* (1), 15–26.

Brigham, E., & Morrow, R. E. (1967). The fast Fourier transform. *IEEE Spectrum*, *4* (12), 63–70.

Bussgang, J. J. (1952). *Crosscorrelation functions of amplitude-distorted Gaussian signals* (Technical Report No. 216), Cambridge, MA: Research Laboratory of Electronics, Massachusetts Institute of Technology.

Castro-Garcia, R., , Tiels, K., Schoukens, J., & Suykens, J. A. K. (2015). Incorporating best linear approximation within LS-SVM-based Hammerstein system identification. In *Proceedings of the 54th IEEE Conference on Decision and Control*. IEEE, 10.1109/cdc.2015.7403387, (pp. 7392–7397).

Crama, P., & Schoukens, J. (2001). Initial estimates of Wiener and Hammerstein systems using multisine excitation. *IEEE transactions on Instrumentation and Measurement, 50* (6), 1791–1795.

De Brabanter, K., Dreesen, P., Karsmakers, P., Pelckmans, K., De Brabanter, J., Suykens, J. A. K., & De Moor, B. (2009). Fixed-size LS-SVM applied to the Wiener-Hammerstein benchmark. In *Proceedings of the 15th IFAC Symposium on System Identification*, (pp. 826–831). Saint-Malo, France: IFAC.

Espinoza, M., Pelckmans, K., Hoegaerts, L., Suykens, J. A. K., & De Moor, B. (2004). A comparative study of LS-SVMs applied to the Silverbox identification problem. In *Proceeding of the 6th IFAC Symposium on Nonlinear Control Systems (NOLCOS)*, (pp. 369–364). Stturgart, Germany: IFAC.

Falck, T., Dreesen, P., De Brabanter, K., Pelckmans, K., De Moor, B., & Suykens, J. A. K. (2012). Least- squares support vector machines for the identification of Wiener-Hammerstein systems. *Control Engineering Practice, 20* (11), 1165–1174.

Falck, T., Pelckmans, K., Suykens, J. A. K., & De Moor, B. (2009). Identification of Wiener-Hammerstein systems using LS-SVMs. In *Proceedings of the 15th IFAC Symposium on System Identification*, (pp. 820–825).Saint-Malo, France: IFAC.

Giri, F., & Bai, E.-W. (Eds.). (2010). *Block-oriented nonlinear system identification* (Vol. 1). Berlin: Heidelberg: Springer.

Goethals, I., Pelckmans, K., Suykens, J. A. K., & De Moor, B. (2005). Identification of MIMO Hammerstein models using least-squares support vector machines. *Automatica, 41* (7), 1263–1272.

Haber, R., & Keviczky, L. (1999). *Nonlinear system identification input-output modeling approach*. Dordrecht, The Netherlands: Springer.

Janczak, A. (2005). *Identification of nonlinear systems using neural networks and polynomial models: A block-oriented approach* (Vol. 310). Berlin Heidelberg: Springer-Verlag.

Mercer, J. (1909). Functions of positive and negative type, and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society A* , 209, 415–446. London.

Narendra, K. S., & Gallman, P. (1966). An iterative method for the identification of nonlinear systems using a Hammerstein model. *IEEE Transactions on Automatic Control, 11* (3), 546–550.

Pintelon, R., & Schoukens, J. (2012). *System identification: A frequency domain approach*. Hoboken, NJ: John Wiley & Sons.

Ravaud, R., Lemarquand, G., & Roussel, T. (2009). Time-varying non linear modeling of electrodynamic loudspeakers. *Applied Acoustics, 70* (3), 450–458.

Schoukens, J., Dobrowiecki, T., & Pintelon, R. (1998). Parametric and nonparametric identification of linear systems in the presence of nonlinear distortions-a frequency domain approach. *IEEE Transactions on Automatic Control, 43* (2), 176–190.

Schoukens, J., Pintelon, R., & Rolain, Y. (2012). *Mastering system identification in 100 exercises*. Hoboken, NJ: John Wiley & Sons.

Schoukens, J., Suykens, J. A. K., & Ljung, L. (2009). Wiener-Hammerstein benchmark. In *Proceedings of the 15th IFAC Symposium on System Identification.* Saint Malo, France: IFAC.

Schoukens, M. (2015). Identification of parallel block-oriented models starting from the best linear approximation. Brussels, Belgium: Vrije Universiteit Brussel (VUB).

Sun, L., Liu, W., & Sano, A. (1999). Identification of a dynamical system with input nonlinearity. In *IEE Proceedings of the Control Theory and Applications* (Vol. 146, pp. 41–51). United Kingdom: IET.

Suykens, J. A. K., Van Gestel, T., De Brabanter, J., De Moor, B., & Vandewalle, J. (2002). *Least squares support vector machines*, Singapore: World Scientific.

Wang, J., Sano, A., Chen, T., & Huang, B. (2009). Identification of Hammerstein systems without explicit parameterisation of non-linearity. *International Journal of Control, 82* (5), 937–952.

Xavier-de Souza, S., Suykens, J. A. K., Vandewalle, J., & Bollé, D. (2010). Coupled simulated annealing. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 40* (2), 320–335.

Young, P. C. (2000). Stochastic, dynamic modelling and signal processing: Time variable and state dependent parameter estimation, *Nonlinear and Nonstationary Signal Processing*. W. J. Fitzgerald, R. Smith, P. C. Young & A. Walden (Eds.). 74–114. Cambridge: Cambridge University Press.

Young, P. C., McKenna, P., & Bruun, J. (2001). Identification of non-linear stochastic systems by state dependent parameter estimation. *International Journal of Control, 74* (18), 1837–1857.