Automatica 93 (2018) 282–289

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Brief paper

A two-experiment approach to Wiener system identification☆

Giulio Bottegal *, Ricardo Castro-Garcia, Johan A.K. Suykens

ESAT-Stadius, KU Leuven, Kastelpaark Arenberg 10, B-3001 Leuven (Heverlee), Belgium

ARTICLE INFO

Article history:
Received 22 February 2017
Received in revised form 17 November 2017
Accepted 15 January 2018
Available online 1 April 2018

Keywords:
System identification
Wiener systems
Experiment design
Least squares support vector machines

ABSTRACT

We propose a new methodology for identifying Wiener systems using the data acquired from two separate experiments. In the first experiment, we feed the system with a sinusoid at a prescribed frequency and use the steady state response of the system to estimate the static nonlinearity. In the second experiment, the estimated nonlinearity is used to identify a model of the linear block, feeding the system with a persistently exciting input. We discuss both parametric and nonparametric approaches to estimate the static nonlinearity. In the parametric case, we show that modeling the static nonlinearity as a polynomial results into a fast least-squares based estimation procedure. In the nonparametric case, least squares support vector machines (LS-SVM) are employed to obtain a flexible model. The effectiveness of the method is demonstrated through numerical experiments.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

Block-oriented system identification aims at representing systems as interconnected linear and nonlinear blocks. The Wiener system is one of such representations, where a linear time-invariant (LTI) block representing the dynamics of the process is followed by a static nonlinear function. Wiener models have proved to be useful to represent nonlinear systems in many application areas, e.g., chemical processes (Kalafatis, Arifin, Wang, & Cluett, 1995; Zhu, 1999), and biological systems (Hunter & Korenberg, 1986). Identification of Wiener systems has been object of intensive research for many years; an overview of previous works can be found in Giri and Bai (2010).

Maximum likelihood/prediction error techniques are discussed in Hagenblad, Ljung, and Wills (2008). The main issue with maximum likelihood is that estimation of the parameters requires the solution of a nonlinear optimization problem. A possible approach to reduce the dimensionality of the problem is to use separable least squares (Bruls, Chou, Haverkamp, & Verhaegen, 1999), or build recursive identification schemes (Wigren, 1993). In Westwick and Verhaegen (1996), a subspace-based method is proposed. Nonparametric methods based on a weighted kernel regression are discussed in several contributions (Greblicki, 1997; Mzyk, 2007), where, however, it is required that the input is an i.i.d. sequence. Semi-parametric techniques relying on a Bayesian nonparametric model of the static nonlinearity and a parametric model of the LTI block are proposed in Lindsten, Schön, and Jordan (2013) and Pillonetto (2013). Other approaches based on instrumental variables and the Wiener G-functionals are found in Janczak (2007) and Tiels and Schoukens (2011), respectively. Experiment design techniques specifically tailored for Wiener system identification are discussed in Mahata, Schoukens, and De Cock (2016).

In this paper, we discuss a novel method for Wiener systems that separates the estimation of the nonlinearity from the identification of the LTI block, facilitating the identification process and reducing the computational burden of maximum likelihood/prediction error techniques. To do so, it is required that the user has the freedom to design the input to the system. Under this setup, a direct approach to identify the Wiener structure is to design a Gaussian input and exploit Bussgang's theorem to separate the linear and the nonlinear blocks (see e.g. Enqvist & Ljung, 2005). This strategy however, being based on computing second order moments, may require collecting long data sets. Furthermore, it is well known that even nonlinearities lead to system estimates equal to zero (Schoukens & Tiels, 2017).

Our approach is based on designing two separate experiments, each consisting of feeding the system with a specific input. In the first experiment, the system is driven by a simple sinusoidal signal with prefixed frequency and phase. Using this signal, we show that we can easily reconstruct the static nonlinearity as a function of the unknown phase delay introduced by the LTI block. We discuss three possible modeling approaches for the nonlinearity. Depending on the adopted approach, we show how to fully recover the nonlinear function (up to a scaling factor), that is, how to

☆ The material in this paper was partially presented at the 56th IEEE Conference on Decision and Control, December 12–15, 2017, Melbourne, Australia. This paper was recommended for publication in revised form by Associate Editor Er-Wei Bai under the direction of Editor Torsten Söderström.

* Corresponding author.
E-mail addresses: g.bottegal@tue.nl (G. Bottegal), ricardo.castro@esat.kuleuven.be (R. Castro-Garcia), johan.suykens@esat.kuleuven.be (J.A.K. Suykens).

https://doi.org/10.1016/j.automatica.2018.03.069
0005-1098/© 2018 Elsevier Ltd. All rights reserved.

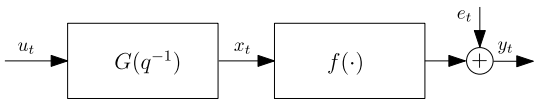


Fig. 1. Block scheme representation of a Wiener system.

remove the ambiguity introduced by the unknown phase delay. The first modeling approach relies on a parametric description of the nonlinearity as a linear combination of a number of basis functions. Here, the phase delay is recovered using a special instance of separable least squares. The second approach is a special case of the first, where the basis functions are monomials. In this case, the function can be fully estimated via a simple procedure involving least-squares estimation. The third approach is a nonparametric one; it relies on the least squares support vector machines (LS-SVM) framework (Suykens, Van Gestel, De Brabanter, De Moor, & Vandewalle, 2002), under the assumption that the nonlinearity is a smooth function. In this case, the phase delay is estimated along with the hyperparameters characterizing the kernel used in the LS-SVM estimation procedure.

We note that the idea of feeding a Wiener system with a sine signal was also explored in previous work. In Bai (2003), the phase delay introduced by the LTI block is estimated by comparing the frequency content of the output and the input. In Giri, Rochdi, and Chaoui (2009), the phase delay is estimated using a geometric approach. In this paper we use different approaches to phase estimation, depending on the model adopted for the static nonlinearity.

Using the estimated model of the static nonlinearity, we perform a second experiment where the system is fed with a persistently exciting input. In this way, we can identify the LTI block by means of a modified version of the standard prediction error method (PEM) for linear output-error (OE) systems (Ljung, 1999). The computational burden of this second step reduces essentially to the one of PEM for OE systems.

2. Wiener system identification using a two-experiment approach

We consider the following SISO system, also called a Wiener system (see Fig. 1 for a schematic representation):

$$\begin{aligned} x_t &= G(q^{-1})u_t \\ y_t &= f(x_t) + e_t\,. \end{aligned} \tag{1}$$

In the former equation, $G(q^{-1})$ represents the transfer function of a causal, asymptotically stable, LTI subsystem, driven by the input u_t, where q^{-1} denotes the time shift operator, namely $q^{-1}u_t = u_{t-1}$. In the latter equation, y_t is the result of a static nonlinear transformation, denoted by $f(\cdot)$, of the signal x_t, and e_t is white noise with (finite) unknown variance σ^2 and finite higher-order moments. Furthermore, we assume that the noise is independent of the input and that $f \in \mathcal{C}^0$, the set of continuous and pointwise defined functions. The problem under study is to estimate the LTI subsystem and the nonlinear function from a set of input and output measurements.

We assume that the user has the freedom to design the input signal u_t. In particular, we assume that the user has the possibility to run two separate experiments, each having a particular signal u_t as an input. The goal of this paper is to describe an identification technique for the system (1) that is linked to a particular choice of these experiments. It consists of the two following steps:

(1) Feed the system with a sinusoid at a prescribed frequency and use the steady-state data to estimate the nonlinear function $f(\cdot)$;

(2) Feed the system with a persistently exciting input signal and identify the LTI subsystem using the information gathered on the first step regarding the static nonlinearity.

Let us first briefly discuss the second step of the proposed procedure. Let the LTI subsystem be described through the parameterized transfer function

$$G(q^{-1},\,\theta) = \frac{b_0 + b_1 q^{-1} + \cdots + b_m q^{-m}}{1 + a_1 q^{-1} + \cdots + a_n q^{-n}}\,, \tag{2}$$

so that its dynamics are completely characterized by the parameter vector $\theta := [b_0 \quad b_1 \quad \dots \quad b_m \quad a_1 \quad \dots \quad a_n]$. Then, assuming that an estimate of the nonlinearity say, $\hat{f}(\cdot)$, is available after the first step of the procedure, we can set up a PEM-based identification criterion as follows

$$\hat{\theta} = \arg\min_{\theta} \frac{1}{N_2} \sum_{t=1}^{N_2} \left(y_t - \hat{f}(G(q^{-1},\,\theta)u_t)\right)^2\,, \tag{3}$$

where N_2 is the number of samples collected during the second experiment. Note that this is a mild generalization of the standard PEM, requiring only to account, in the optimization process, for the nonlinear transformation induced by $\hat{f}(\cdot)$. This does not make the solution of (3) harder than a standard PEM applied to an output-error model, because in both cases we have to face a nonlinear and non-convex optimization problem. Solution of PEM-type problems for OE systems has been object of intensive research for decades, and the field is now at a mature stage, with several developed methodologies that ensure achieving the global minimum of (3) (see, e.g., Eckhard, Bazanella, Rojas, & Hjalmarsson, 2017 for prefiltering techniques for OE identification, or Lacy, Erwin, and Bernstein (2001) and Wigren (1994) for Wiener system identification when the nonlinearity is known).

As opposed to the aforementioned second step, the first step can be more involved and requires a more thorough analysis. We shall focus on this step in the remainder of the paper.

3. Approaches to estimate the nonlinearity

In this section we discuss the first step of the procedure, proposing three estimation approaches for the static nonlinearity.

We consider the input signal $u_t = \sin(\omega t + \phi_0)$, where ω is a user-prescribed frequency and ϕ_0 is a known phase delay. Without loss of generality, in the remainder of the paper we shall consider $\phi_0 = 0$. Then, after the transient effect of $G(q^{-1})$ has vanished, we have that $x_t = A_\omega \sin(\omega t + \phi_\omega)$, where A_ω and ϕ_ω are the gain and the phase delay of the LTI subsystem $G(q^{-1})$ at the frequency ω (Ljung, 1999, Ch. 2). Due to the structural non-identifiability of Wiener systems, A_ω cannot be determined (see Remark 2). We thus drop it and define a new signal $\bar{x}_t := \sin(\omega t + \phi_\omega)$, which is parameterized by the unknown quantity ϕ_ω. Accordingly, we write the output of the system as

$$y_t = f(\sin(\omega t + \phi_\omega)) + e_t\,. \tag{4}$$

Then, the problem under study, that is to estimate $f(\cdot)$, is coupled with the problem of estimating ϕ_ω. In the following, we describe three approaches to this problem, assuming that the number of collected samples of y_t (at its steady state) is equal to N_1.

Remark 1. Since we are estimating the static nonlinearity using the signal $\bar{x}_t$ instead of x_t, we are obtaining a scaled (in the x-axis) version of $f(\cdot)$, that is, we are estimating $f(x/A_\omega)$ instead of $f(x)$. This scaling effect is compensated in the second phase of the method; in fact (3) will return the estimate $A_\omega G(q^{-1})$ instead of $G(q^{-1})$. Then, we need additional information (e.g., on the LTI system gain, see Bai, 1998) to uniquely recover $G(q^{-1})$ and $f(\cdot)$; this

lack of identifiability is a well known issue in block oriented system identification. However, if the focus is on output prediction (as is in the experiments of Section 4), rescaling of the two blocks is not required.

Remark 2. In practice, the quantity A_ω crucially determines the quality of the identified static nonlinearity. Because we are assuming additive white noise, the average signal-to-noise ratio (SNR) is $A_\omega^2/2\sigma^2$; then optimal results are obtained where A_ω is maximum. However, A_ω is not known in advance, and to determine its biggest value one would have to perform a preliminary experiment swiping the frequency spectrum to detect where $|G(\omega)|$ is maximum, using e.g. the technique proposed in Rallo, Formentin, Rojas, Oomen, and Savaresi (2017).

3.1. Parametric approach

We assume that there exist a set of known basis functions $h_0(\cdot),\ h_1(\cdot),\ \ldots,\ h_p(\cdot)$ such that

$$f(x) = \sum_{i=0}^{p} c_i h_i(x) \quad , \quad \forall x \in \mathbb{R}.$$

Then, the problem of estimating $f(\cdot)$ reduces to determining the coefficients $c_0,\ \ldots,\ c_p$. We rewrite (4) as

$$y_t = \sum_{i=0}^{p} c_i h_i(\sin(\omega t + \phi_\omega)) + e_t\,. \tag{5}$$

Let $H(\phi_\omega) \in \mathbb{R}^{N_1 \times p+1}$ be a matrix such that $[H(\phi_\omega)]_{t,i} = h_i(\sin(\omega t + \phi_\omega))$. Then we have the following regression model

$$y = H(\phi_\omega)c + e\,, \tag{6}$$

where y and e are N_1-dimensional vectors of the samples of y_t and e_t, and c is the vector of the coefficients of the basis functions. Because ϕ_ω is unknown, we treat it as an unknown parameter to be determined together with c; for sake of clarity we rewrite the regression model (6) replacing ϕ_ω with a generic ϕ:

$$y = H(\phi)c + e\,. \tag{7}$$

An unbiased estimate of c can be obtained via least squares (recall that e is white noise). This estimate is function of the unknown phase delay introduced by the LTI block, and corresponds to

$$\hat{c}(\phi) = \left(H^T(\phi)H(\phi)\right)^{-1} H^T(\phi)y\,. \tag{8}$$

To estimate ϕ (and hence c), we introduce the following criterion:

$$\hat{\phi}_\omega = \arg\min_{\phi \in \mathcal{I}} \frac{1}{N_1} \|y - \hat{f}_\phi\|^2\,, \tag{9}$$

where $\hat{f}_\phi = H(\phi)\hat{c}(\phi)$ is the estimate of $f(\cdot)$, evaluated at $\bar{x}_t$, $t = 1,\ \ldots,\ N_1$ and in vector notation, and $\mathcal{I}$ is a suitable subset of the negative real semi-axis (recall that the phase delay is always negative for causal systems). The question is under which conditions (9) admits a unique solution. To this end we introduce the following concept.

Definition 3. A function $h(\cdot)$ is *phase-indistinguishable* in the set $\mathcal{I}$ if there exist ϕ_1 and ϕ_2 in $\mathcal{I}$ and $\alpha \in \mathbb{R}$ such that

$$h(\sin(\omega t + \phi_1)) = \alpha h(\sin(\omega t + \phi_2))\,, \tag{10}$$

for all $t \in \mathbb{N}$.

Therefore, if $f(\cdot)$ is parameterized through a set of phase-indistinguishable basis functions, criterion (9) may not admit a unique minimum. Quite fortunately, the following lemma clarifies that, with a suitable definition of $\mathcal{I}$, the set of phase-indistinguishable functions becomes trivial (the proof is direct and thus it is skipped).

Lemma 4. *Condition (10) is satisfied if and only if $\phi_2 = \phi_1 + k\pi$, $k \in \mathbb{Z}$, (and $\alpha = (-1)^k$) or h is constant (and $\alpha = 1$).*

Therefore, we have to restrict our phase search in the interval $\mathcal{I} = (-\pi,\ 0]$. Using the lemma, we can prove the following result (see the Appendix for a proof).

Proposition 5. *Let $\mathcal{I} = (-\pi,\ 0]$. Let e_t be a stochastic stationary ergodic process with finite variance. It holds that*

$$\phi_\omega = \arg\min_{\phi \in \mathcal{I}} \mathbb{E}\,\|y - \hat{f}_\phi\|^2.$$

Moreover, the asymptotic (in N_1) solution of the problem (9) is equal to ϕ_ω.

We summarize the procedure for estimation of parametric static nonlinearities in the following:

(1) Select a subset of values $\phi \in \mathcal{I}$;
(2) For every ϕ solve (8) and evaluate $\|y - \hat{f}_\phi\|^2$;
(3) Choose $\hat{c}(\phi)$ associated with ϕ solving (9).

Albeit (9) is still a nonlinear problem, it requires to search for the optimum of a scalar variable within a bounded interval. Therefore, it can be solved quickly by using e.g. a grid search over $\mathcal{I}$.

3.2. Polynomial nonlinearity

We now discuss the case where the nonlinear function $f(\cdot)$ can be expressed as a polynomial of known order, namely

$$f(x) = c_0 + c_1 x + \cdots + c_p x^p\,. \tag{11}$$

For ease of exposition, we assume that p is even (the case p odd runs along the same lines of reasoning). We recall that (see e.g. Bai, 2003), for $i \in \mathbb{N}$,

$$\sin^{2i}(x) = \frac{1}{2^{2i}}\binom{2i}{i} + \frac{(-1)^i}{2^{2i-1}}\sum_{k=0}^{i-1}(-1)^k\binom{2i}{k}\cos(2(i-k)x),$$

$$\sin^{2i+1}(x) = \frac{(-1)^i}{4^i}\sum_{k=0}^{i}(-1)^k\binom{2i+1}{k}\sin((2i+1-2k)x).$$

Combining these expressions with standard trigonometric addition formulas, we can rewrite the nonlinear function as

$$\begin{aligned} f(\bar{x}_t) = \sum_{i=0}^{p/2} c_{2i}\Big[&\frac{1}{2^{2i}}\binom{2i}{i} + \frac{(-1)^i}{2^{2i-1}}\sum_{k=0}^{i-1}(-1)^k\binom{2i}{k} \\ &\times \left[\cos(2(i-k)\phi_\omega)\cos(2(i-k)\omega t)\right. \\ &\left.- \sin(2(i-k)\phi_\omega)\sin(2(i-k)\omega t)\right]\Big] \\ + \sum_{i=0}^{p/2-1} c_{2i+1}&\frac{(-1)^i}{4^i}\sum_{k=0}^{i}(-1)^k\binom{2i+1}{k} \\ &\times \left[\cos((2i+1-2k)\phi_\omega)\sin((2i+1-2k)\omega t)\right. \\ &\left.+ \sin((2i+1-2k)\phi_\omega)\cos((2i+1-2k)\omega t)\right]. \end{aligned} \tag{12}$$

This equation is particularly interesting because it permits to express $f(\bar{x}_t)$ as a linear combination of sines and cosines with frequency $q\omega_t$, where $q = 0,\ \ldots,\ p$. The coefficients of this

linear combination are also a function of sines and cosines of the unknown phase delay ϕ_ω. Let

$$k_0 := \sum_{i=0}^{p/2} c_{2i} \frac{1}{2^{2i}} \binom{2i}{i}$$

$$k_{s,q} := \sum_{i=q/2}^{p/2} c_{2i} \frac{1}{2^{2i-1}} \binom{2i}{i-q/2} \sin(q\phi_\omega) \qquad (q \text{ even})$$

$$k_{c,q} := -\sum_{i=q/2}^{p/2} c_{2i} \frac{1}{2^{2i-1}} \binom{2i}{i-q/2} \cos(q\phi_\omega) \qquad (q \text{ even})$$

$$k_{s,q} := (-1)^{2i-(q-1)/2} \sum_{i=(q-1)/2}^{p/2-1} c_{2i+1} \frac{1}{4^i} \binom{2i+1}{i-(q-1)/2} \times \cos(q\phi_\omega) \qquad (q \text{ odd})$$

$$k_{c,q} := (-1)^{2i-(q-1)/2} \sum_{i=(q-1)/2}^{p/2-1} c_{2i+1} \frac{1}{4^i} \binom{2i+1}{i-(q-1)/2} \times \sin(q\phi_\omega) \qquad (q \text{ odd}).$$

Let also, for $q = 1, \ldots, p$,

$$k_q = \sum_{i=q/2}^{p/2} c_{2i} \frac{1}{2^{2i-1}} \binom{2i}{i-q/2}, \qquad (q \text{ even})$$

$$k_q = \sum_{i=(q-1)/2}^{p/2-1} c_{2i+1} \frac{1}{4^i} \binom{2i+1}{i-(q-1)/2}, \qquad (q \text{ odd})$$

so that there exists a matrix $M \in \mathbb{R}^{p+1 \times p+1}$ such that

$$\mathbf{k} = Mc, \quad \mathbf{k} := \begin{bmatrix} k_0 & \ldots & k_p \end{bmatrix}^T. \tag{13}$$

Then we can write

$$\begin{aligned} \bar{\mathbf{k}} &= \begin{bmatrix} k_0 & k_{s,1} & k_{c,1} & \ldots & k_{s,p} & k_{c,p} \end{bmatrix}^T \\ &= \begin{bmatrix} \gamma_0 k_0 & \gamma_{s,1} k_1 \sin(\phi_\omega) & \gamma_{c,1} k_1 \cos(\phi_\omega) & \ldots \\ \ldots & \gamma_{s,p} k_p \sin(p\phi_\omega) & \gamma_{c,p} k_p \cos(p\phi_\omega) \end{bmatrix}^T, \end{aligned} \tag{14}$$

where γ_0 and the $\gamma_{s,i}$, $\gamma_{c,i}$, $i = 1, \ldots, p$, are equal to -1 or 1 and are known in advance. Therefore, if we are able to estimate the vector $\bar{\mathbf{k}}$ and the phase delay ϕ_ω, we can also estimate $\mathbf{k}$ and consequently c. To this end, we define

$$\psi_t^T := [1 \ \cos(\omega t) \ \sin(\omega t) \ \ldots \ \sin(p\omega t) \ \cos(p\omega t)], \tag{15}$$

so that we can rewrite (12) as $f(\bar{x}_t) = \psi_t^T \bar{\mathbf{k}}$, and express the measurement equation via the linear regression model

$$y_t = \psi_t^T \bar{\mathbf{k}} + e_t,$$

which also represents the Fourier expansion of the output signal. We can then obtain the least-squares estimate

$$\widehat{\bar{\mathbf{k}}} = \left(\sum_{t=1}^{N_1} \psi_t \psi_t^T \right)^{-1} \sum_{t=1}^{N_1} \psi_t y_t. \tag{16}$$

Using this estimate, we now show how to recover the coefficients of the polynomial. From (14), the absolute value of each k_i, $i = 1, \ldots, p$, can be reconstructed via

$$|\hat{k}_i| = \sqrt{\hat{k}_{s,i}^2 + \hat{k}_{c,i}^2}. \tag{17}$$

Using the estimates $\hat{k}_{s,1}$ and $\hat{k}_{c,1}$, one can recover the phase delay ϕ_ω as

$$\hat{\phi}_\omega = \tan^{-1}\left(\frac{\hat{k}_{s,1}}{\hat{k}_{c,1}} \right), \tag{18}$$

where uniqueness of the solution is guaranteed in $\mathcal{I} = (-\pi, 0]$. Using $\hat{\phi}_\omega$, we can uniquely recover the sign of the coefficients k_i, exploiting the knowledge of the coefficients $\gamma_{s,i}$, $\gamma_{c,i}$ introduced in (14). Finally, using the matrix M defined in (13), the estimate $\hat{c}$ can be recovered via

$$\hat{c} = M^{-1} \hat{\mathbf{k}}. \tag{19}$$

The whole procedure for the polynomial case is summarized in the following:

(1) Construct the regressors (15) and compute the least-squares estimate (16);
(2) Recover the absolute value of the coefficients k_i, $i = 0, \ldots, p$ using (17);
(3) Estimate the phase shift via (18) and the sign of the coefficients k_i, $i = 0, \ldots, p$;
(4) Recover the coefficients c_i, $i = 0, \ldots, p$, using (19).

We observe that this procedure still requires to recover the phase delay ϕ_ω. However, in this case this can be done with one simple operation, namely (18), instead of requiring the solution of a dedicated optimization problem, like in the general parametric case discussed in Section 3.1. Furthermore, it is not required to have a highly accurate estimate of ϕ_ω. In fact, what we need is just that $\hat{\phi}_\omega$ lies in the same orthant of ϕ_ω, so that we can correctly determine the sign of the coefficients k_i, $i = 0, \ldots, p$. As for the asymptotic performance of this estimation procedure, we have the following result showing that the asymptotic variance of the procedure *does not depend on the choice of* ω. (see the Appendix for the proof).

Proposition 6. *The procedure outlined above gives consistent estimates of the coefficients c_i, $i = 0, \ldots, p$. Furthermore, the asymptotic covariance of their estimates is equal to*

$$\frac{\sigma^2}{N_1} M^{-1} D M^{-T}, \tag{20}$$

where $D = diag\{1, 2, \ldots, 2\}$.

Example 7. Consider a third-order polynomial nonlinearity. We have

$$\begin{aligned} f(\bar{x}_t) = c_0 &+ \frac{c_2}{2} + (c_1 + \frac{3}{4}c_3)\cos(\phi_\omega)\sin(\omega t) \\ &+ (c_1 + \frac{3}{4}c_3)\sin(\phi_\omega)\cos(\omega t) \\ &+ \frac{c_2}{2}\sin(2\phi_\omega)\sin(2\omega t) - \frac{c_2}{2}\cos(2\phi_\omega)\cos(2\omega t) \\ &- \frac{c_3}{4}\cos(3\phi_\omega)\sin(3\omega t) - \frac{c_3}{4}\sin(3\phi_\omega)\cos(3\omega t), \end{aligned}$$

so $[\gamma_0, \gamma_{s,1}, \gamma_{c,1}, \ldots, \gamma_{s,3}, \gamma_{c,3}] = [1, 1, 1, 1, -1, -1, -1]$. The matrix M and the asymptotic covariance of the coefficient estimates computed via (20) are, respectively,

$$\begin{bmatrix} 1 & 0 & 1/2 & 0 \\ 0 & 1 & 0 & 3/4 \\ 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1/4 \end{bmatrix} \quad \text{and} \quad \frac{\sigma^2}{N_1} \begin{bmatrix} 3 & 0 & -4 & 0 \\ 0 & 20 & 0 & -24 \\ -4 & 0 & 8 & 0 \\ 0 & -24 & 0 & 32 \end{bmatrix}.$$

This shows that, for the case $p = 3$, the odd coefficients are estimated with lower accuracy. Not surprisingly, the estimates of the even coefficients and the estimates of the odd coefficients are uncorrelated.

Remark 8. A similar approach to Wiener system identification with polynomial nonlinearity was proposed in Bai (2003, Sec. 4). The main difference between that approach and the one proposed

in this paper lies in the way the phase delay is estimated. In Bai (2003) the phase delay is estimated by comparing the DFTs of the output and the input, leading to a non-consistent phase estimate (Giri, Rochdi, & Chaoui, 2008). An additional advantage of the method proposed in this paper is that it is required to estimate only the orthant of the phase delay, arguably leading to estimates of nonlinearity coefficients with higher precision.

3.3. Nonparametric approach

In the LS-SVM framework, the unknown function $f(\cdot)$ is represented as follows

$$f(\bar{x}_t) = w^T \varphi(\bar{x}_t) + b\,, \tag{21}$$

where $\varphi(\cdot) : \mathbb{R}^n \to \mathbb{R}^{n_h}$ is a function mapping its argument to a high dimensional (possibly infinite) space, w is an unknown vector of coefficients, and b an unknown bias term. We shall show that an explicit expression of $\varphi(\bar{x}_t)$ is not required; all we need is a *kernel* representation of the type $K(\bar{x}_i,\ \bar{x}_j) = \varphi(\bar{x}_i)^T \varphi(\bar{x}_j)$. In this paper, we make use of the radial basis function (RBF) kernel, defined as

$$K(\bar{x}_i,\ \bar{x}_j) = \exp\left(-\eta^{-1}(\bar{x}_i - \bar{x}_j)^2\right)\,, \tag{22}$$

where $\eta \geq 0$ is usually called a tuning hyperparameter. The problem of estimating $f(\cdot)$ can then be formulated as

$$\begin{aligned} \min_{w,b,\varepsilon} \quad & \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_{t=1}^{N_1} \varepsilon_t^2 \\ \text{subject to} \quad & y_t = w^T \varphi(\bar{x}_t) + b + \varepsilon_t, t = 1,\ \ldots,\ N_1\,, \end{aligned} \tag{23}$$

where γ is a trade-off parameter regulating the balance between adherence to data and smoothing of the estimated function (this latter characteristic is given by the w-dependent term). The term ε_t denotes the prediction error; ideally, one has $\varepsilon_t = e_t$ when (21) describes the nonlinear function perfectly. Computing the Lagrangian associated with this problem we get

$$\begin{aligned} \mathcal{L}(w, b, \varepsilon; \alpha) = & \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_{t=1}^{N_1} \varepsilon_t^2 \\ & - \sum_{t=1}^{N_1} \alpha_t (w^T \varphi(\bar{x}_t) + b + \varepsilon_t - y_t)\,, \end{aligned}$$

where $\alpha_t \in \mathbb{R}$ are the Lagrange multipliers. It can be shown (see e.g. Suykens et al., 2002) that the optimality conditions are obtained by solving the following linear system

$$\begin{bmatrix} 0 & \mathbf{1}_{N_1}^T \\ \mathbf{1}_{N_1} & \Omega + \dfrac{1}{\gamma} I_{N_1} \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix}\,, \tag{24}$$

where $\alpha = [\alpha_1\ \ldots\ \alpha_N]^T$, and $\Omega \in \mathbb{R}^{N_1 \times N_1}$ is such that $[\Omega]_{i,j} = K(\bar{x}_i,\ \bar{x}_j)$. Computing α we can obtain an estimate of $f(\cdot)$ at any point $\tilde{x}$ (not necessarily corresponding to one of the $\bar{x}_t$, $t = 1,\ \ldots,\ N_1$, used to estimate the model) as

$$\hat{f}(\tilde{x}) = \sum_{t=1}^{N_1} \alpha_t K(\bar{x}_t, \tilde{x}) + b\,. \tag{25}$$

The obtained estimate can be shown to be consistent (Christmann & Steinwart, 2007; Steinwart, 2005), and depends on the regularization parameter γ entering (23), the kernel shaping hyperparameter η in (22), and also on the unknown phase delay ϕ_ω. Tuning suitable values for the first two quantities can be done using standard techniques, such as marginal likelihood, SURE, or cross validation (Friedman, Hastie, & Tibshirani, 2001). As for ϕ_ω, it is natural to ask to which interval we should restrict our search, and if we can treat it as an additional hyperparameter, to be tuned along with the other model hyperparameters. The following lemma answers the first question (the proof is straightforward and skipped for space limitation).

Lemma 9. *Let $K(\cdot,\ \cdot)$ be an isotropic kernel, i.e. a kernel such that $K(\bar{x}_i,\ \bar{x}_j) = K(|\bar{x}_i - \bar{x}_j|)$, for every $\bar{x}_i,\ \bar{x}_j$. Denote by $\hat{f}_{\phi_1}(\tilde{x})$ and $\hat{f}_{\phi_2}(\tilde{x})$ the estimates of $f(\cdot)$ at the point $\tilde{x}$ obtained using $\bar{x}_t = \sin(\omega t + \phi_1)$ and $\bar{x}_t = \sin(\omega t + \phi_2)$ as input locations, respectively. Then $\hat{f}_{\phi_1}(\tilde{x}) = \hat{f}_{\phi_2}(\tilde{x})$ for all $\tilde{x} \in \mathbb{R}$ if $\phi_1 = \phi_2 + k\pi$, $k \in \mathbb{Z}$.*

Then, also in this case the search of ϕ_ω has to be restricted to $\mathcal{I} = (-\pi,\ 0]$. The following example clarifies that ϕ_ω can be seen as an additional kernel hyperparameter.

Example 10. Consider the RBF kernel (22). We have that $(\bar{x}_i - \bar{x}_j)^2 = [z_i^s - z_j^s\ \ z_i^c - z_j^c]\Psi[z_i^s - z_j^s\ \ z_i^c - z_j^c]^T$, where $z_i^s = \sin(\omega i)$, $z_i^c = \cos(\omega i)$ (the same notation holds for j) and

$$\Psi = \begin{bmatrix} \cos^2\phi_\omega & \dfrac{1}{2}\sin 2\phi_\omega \\ \dfrac{1}{2}\sin 2\phi_\omega & \sin^2\phi_\omega \end{bmatrix}.$$

Then we can write $K(\bar{x}_i,\ \bar{x}_j) = \exp(-\eta^{-1}\|z_i - z_j\|_\Psi^2)$, with $z_i = [z_i^s\ z_i^c]^T$, which shows that the RBF kernel with the input $\bar{x}_t = \sin(\omega t + \phi_\omega)$ can be seen as a RBF kernel with a bi-dimensional known input $z_t = [\sin(\omega t)\ \ \cos(\omega t)]$, weighted by the Mahalanobis distance Ψ, which depends on the hyperparameter ϕ_ω. Therefore, we can treat the tuning of ϕ_ω in the same way we treat the tuning of the other kernel hyperparameter η.

In the following we summarize the procedure for nonparametric estimation of $f(\cdot)$:

(1) Tune the hyperparameters γ, η, ϕ_ω;
(2) Compute $f(\tilde{x})$ using (25).

4. Numerical experiments

The proposed methodology is tested on two simulated Wiener systems, which we refer to as S1 and S2. The LTI block of S1 is obtained using the Matlab command `cheby2(3,5,0.2)`, which returns a third-order system with stopband edge frequency 0.2π and 5 dB of stopband attenuation at the passband value. The static nonlinearity is the third-order polynomial $f(x) = x^3$. The parameters characterizing S1 are then $a = [2.46\ \ 2.26\ \ -0.77]$, $b = 10^{-2}[0.47\ \ 1.42\ \ 1.42\ \ 0.47]$, $c = [0\ \ 0\ \ 0\ \ 1]$. As for S2, the LTI part is obtained using the Matlab command `cheby2(4,18,0.2)`, while the nonlinear part is the third-order polynomial $f(x) = x + 5x^2 - 0.5x^3$. Thus, S2 is described by the parameters $a = [-2.49\ \ 2.53\ \ -1.18\ \ 0.22]$, $b = [0.11\ \ -0.21\ \ 0.28\ \ -0.21\ \ 0.11]$, $c = [0\ \ 1\ \ 5\ \ -0.5]$. The variance of the output noise is obtained by matching different values of the SNR. In particular, we test the proposed methods for SNR = 10, 20, 40 dB. Thus, we have in total 6 experimental conditions. For each experimental condition we generate 100 Monte Carlo, each with a different noise realization. The performance of the methods is evaluated by assessing the accuracy in tracking the output of a noiseless test set $y_{t,\text{test}}$ of length $N = 500$, obtained by feeding the systems with a i.i.d. Gaussian sequence. We use the normalized mean absolute error (NMAE), defined as $\%\text{NMAE} = \frac{100}{N}\sum_{t=1}^{N} \left|y_{t,\text{test}} - \hat{y}_{t,\text{test}}\right| / \left|\max(y_{t,\text{test}}) - \min(y_{t,\text{test}})\right|$, and the normalized mean square error (NMSE), defined as

$$\%\text{NMSE} = 100\frac{\sqrt{\sum_{t=1}^{N}\left(y_{t,\text{test}} - \hat{y}_{t,\text{test}}\right)^2}}{\sqrt{\sum_{t=1}^{N}\left(y_{t,\text{test}} - \text{mean}(y_{t,\text{test}})\right)^2}}.$$

Table 1
Median values of %NMAE and %NMSE over 100 Monte Carlo runs obtained by the three parametric methods on the 6 tested experimental conditions. The best results are in boldface.

SNR (dB)	Method	S1		S2	
		NMAE	NMSE	NMAE	NMSE
10	**2-E-P**	**0.19**	**4.82**	0.5	6.97
	2-E-Poly	0.2	5.1	**0.49**	**6.87**
	NLHW	**0.19**	5.28	3.3	39.63
20	**2-E-P**	**0.15**	4.13	0.34	4.74
	2-E-Poly	0.16	4.07	**0.33**	**4.67**
	NLHW	0.16	**3.71**	3.72	51.12
40	**2-E-P**	0.14	3.44	**0.25**	3.55
	2-E-Poly	0.13	3.33	**0.25**	**3.49**
	NLHW	**0.09**	**2.25**	3.61	44.35

4.1. Experiments using the parametric and the polynomial approaches

First, we test the methods presented in Section 3.1 and in Section 3.2. We refer to the two methods as **2-E-P** (two-experiment parametric) and **2-E-Poly** (two-experiment polynomial). We use the same input signals for both the methods. In the first stage of the approach, we feed the system with the signal $u_t^1 = \sin(0.05t)$, keeping $N_1 = 500$ transient-free samples. In the second stage, we use random white Gaussian noise with unit variance as input u_t^2, collecting $N_2 = 500$ samples. The two methods are compared with a single-stage PEM-based estimator, implemented through the Matlab command `nlhw` (see Ljung, Zhang, Lindskog, Iouditski, & Singh, 2006 for details); we refer to this method as **NLHW**. To get a fair comparison, the data used by **NLHW** are obtained by feeding the system with a random white Gaussian sequence of length $N_1 + N_2$ with variance equal to the variance of the sequence $[u_1^1 \ \dots \ u_{N_1}^1 \ u_1^2 \ \dots \ u_{N_2}^2]$.

In Table 1 we report the medians of the results obtained in the 6 experiments. As can be seen, on S1 the performance of the three methods are comparable to each other, while **NLHW** fails in providing a good model of S2. The motivation is that the method often falls into a local minimum of the cost function related to the PEM optimization problem. Separating the identification of the LTI block from the static nonlinearity avoids this issue; thus, the proposed methods **2-E-P** and **2-E-Poly** give reliable results for both S1 and S2.

4.2. Experiments using the nonparametric approach

For the nonparametric approach, the input used in the first stage is $u_t = 10\sin(t + \pi/2)$. We collect $N_1 = 500$ steady-state output samples. In the second stage the systems are excited with a uniformly distributed random signal, collecting $N_2 = 500$ samples.

In the proposed nonparametric approach, dubbed **2-E-NP**, $\hat{f}(\cdot)$ is estimated using the LS-SVM approach described in Section 3.3, under a 10-fold cross validation scheme. To solve the optimization problem associated with hyperparameter selection, we use coupled simulated annealing (Xavier-de Souza, Suykens, Vandewalle, & Bollé, 2009); an implementation of this technique is available in the toolbox LS-SVMlab v1.8.[1] The proposed method is compared with the method introduced in Castro-Garcia and Suykens (2016), where the Best Linear Approximation (BLA) (Pintelon & Schoukens, 2012) is used together with LS-SVM to model the linear and nonlinear blocks, respectively. This method is referred to as **BLA+LS-SVM**. This method uses 10 random phase multisine (see Pintelon & Schoukens, 2012) inputs $u_{t,\mathrm{BLA}}^{(i)}$ with their corresponding outputs $y_{t,\mathrm{BLA}}^{(i)}$. Here $t = 1, \dots, 2500$ and $i = 1, \dots, 10$. Using those signals, the BLA is calculated and as an approximation to the LTI block. Using a second data set with a new input signal $u_{t,nl}$, the intermediate variable $\hat{x}_{t,nl}$ is estimated. With $\hat{x}_{t,nl}$ and the known output $y_{t,nl}$ a model of the nonlinear block is estimated using LS-SVM. In this case, $u_{t,nl}$ is the same sinusoidal signal described used by **2-E-NP**. To train the LS-SVM model, only the last 500 samples of $\hat{x}_{t,nl}$ and $y_{t,nl}$ are used.

[1] http://www.esat.kuleuven.be/sista/lssvmlab/.

Table 2
Median values of %NMAE over 100 Monte Carlo runs obtained by the two nonparametric methods on the 6 tested experimental conditions. The best results are in boldface.

SNR (dB)	Method	S1 NMAE	S2 NMAE
10	**2-E-NP**	**2.29**	**2.58**
	BLA+LS-SVM	16.7	53.81
20	**2-E-NP**	**1.73**	**1.45**
	BLA+LS-SVM	29.08	4.68
40	**2-E-NP**	**1.59**	**0.94**
	BLA+LS-SVM	13.15	4.72

In Table 2 we report the median values of %NMAE obtained using the two nonparametric methods. It can be seen that the proposed technique outperforms **BLA+LS-SVM**, despite the simplicity of the input employed in the first stage of the procedure, as compared with the random phase multisine signal required by BLA (note also that **BLA+LS-SVM** uses a 5 times longer input in this phase). It should be noted that the performance of **2-E-NP** is lower than those obtained by the parametric methods. This can be explained by the fact that the parametric approaches exploit more detailed prior information on the static nonlinearity, because it is known that the nonlinearity is a third-order polynomial, while when **2-E-NP** is used, it is only known that the nonlinearity is smooth.

5. Conclusions

We have proposed a new method for Wiener system identification. The main idea underlying the method is that we can separate the estimation of the static nonlinear function from the identification of the LTI block composing the Wiener system. To do so, we have to excite the system using two different inputs. The first input is a sinusoid, which, after the transient effect of the LTI system has vanished, permits to estimate the static nonlinearity as a function of the phase delay introduced by the LTI block. We have described three different approaches to nonlinearity estimation, namely a parametric, a polynomial, and a nonparametric approach. Depending on the adopted approach, the phase delay is also estimated so that the static nonlinearity is recovered. Using the information on the static nonlinearity, we use a persistently exciting input to identify the LTI block. The proposed method is shown to compare favorably with other techniques for Wiener system identification.

Acknowledgments

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).

Appendix

Proof of Proposition 5. We have

$$\begin{aligned}\mathbb{E}\,\|y-\hat{f}_\phi\|^2 &= \mathbb{E}\,\|y-H(\phi)\hat{c}(\phi)\|^2\\ &= \mathbb{E}\,\left\|\left(I-H(\phi)\left(H^T(\phi)H(\phi)\right)^{-1}H^T(\phi)\right)y\right\|^2\\ &= \left\|\left(I-H(\phi)\left(H^T(\phi)H(\phi)\right)^{-1}H^T(\phi)\right)H(\phi_\omega)c\right\|^2\\ &+ \sigma^2\mathrm{tr}\left[\left(I-H(\phi)\left(H^T(\phi)H(\phi)\right)^{-1}H^T(\phi)\right)^2\right].\end{aligned}$$

Since $I-H(\phi)(H^T(\phi)H(\phi))^{-1}H^T(\phi)$ is a projection matrix, its trace is constant (and equal to its rank), independently of the value assumed by ϕ. Hence, it does not affect the value of $\mathbb{E}\,\|y-\hat{f}_\phi\|^2$. As for the first term of the above equality, we have $\|(I-H(\phi)(H^T(\phi)H(\phi))^{-1}H^T(\phi))H(\phi_\omega)c\|^2 \geq 0$. Equality to 0 holds only when $H(\phi)=H(\phi_\omega)$, because in $[0,\ \pi)$ there always exist at least one of the basis functions that is not phase-indistinguishable. This proves the first statement of Proposition 5. By multiplying the objective function of (9) by $1/N_1$, and letting $N_1\to\infty$ we have $\frac{1}{N_1}\sum_{t=1}^{N_1}(y_t-\hat{f}_\phi(x_t))^2 \to \mathbb{E}\,\|y_t-\hat{f}_\phi(x_t)\|^2$, where the expectation is taken with respect to the distribution of e_t. The second statement of Proposition 5 follows as a special case of the first.

Proof of Proposition 6. Let

$$\Gamma=\begin{bmatrix}\gamma_0 & 0 & \dots & 0\\ 0 & \gamma_{s,1}\sin(\phi_\omega) & \dots & 0\\ 0 & \gamma_{c,1}\cos(\phi_\omega) & \dots & 0\\ \vdots & \vdots & & \vdots\\ 0 & \dots & 0 & \gamma_{s,p}\sin(p\phi_\omega)\\ 0 & \dots & 0 & \gamma_{c,p}\cos(p\phi_\omega)\end{bmatrix},$$

so that (14) can be rewritten as $\bar{\mathbf{k}}=\Gamma\mathbf{k}$. We note that the Moore–Penrose pseudoinverse of Γ is $\Gamma^{\#}=(\Gamma^T\Gamma)^{-1}\Gamma^T=\Gamma^T$, so that $\mathbf{k}=\Gamma^T\bar{\mathbf{k}}$. Since we do not need to estimate ϕ_ω but only its sign, we can conclude that the covariance matrix of the estimated coefficients c_i, $i=0,\ \dots,\ p$ can be calculated as the same as the one obtained via least squares, that is

$$\begin{aligned}\mathbb{E}\,[(\hat{c}-c)(\hat{c}-c)^T] &= M^{-1}\mathbb{E}\,[(\hat{\mathbf{k}}-\mathbf{k})(\hat{\mathbf{k}}-\mathbf{k})^T]M^{-T}\\ &= M^{-1}\Gamma^T\mathbb{E}\,[(\widehat{\bar{\mathbf{k}}}-\bar{\mathbf{k}})(\widehat{\bar{\mathbf{k}}}-\bar{\mathbf{k}})^T]\Gamma M^{-T}\\ &= \sigma^2M^{-1}\Gamma^T\left(\sum_{t=1}^{N_1}\psi_t\psi_t^T\right)^{-1}\Gamma M^{-T}.\end{aligned}$$

Recalling the structure of ψ_t given in (15), it is straightforward to check that, as N_1 grows large,

$$\frac{1}{N_1}\sum_{t=1}^{N_1}\psi_t\psi_t^T \longrightarrow \mathrm{diag}\left\{1,\ \frac{1}{2},\ \dots,\ \frac{1}{2}\right\}:=\bar{D}^{-1},$$

where $\bar{D}$ has size $2p+1\times 2p+1$. Eq. (20) follows from the fact that $\Gamma^T\bar{D}\Gamma=\Gamma^T\Gamma D=D$. The consistency of the estimates follows from the consistency of the least-squares estimates (16). This concludes the proof.

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 Wiener models. *Automatica*, *39*(9), 1521–1530.

Bruls, J., Chou, C., Haverkamp, B., & Verhaegen, M. (1999). Linear and non-linear system identification using separable least-squares. *European Journal of Control*, *5*(1), 116–128.

Castro-Garcia, R., & Suykens, J. A. K. (2016). Wiener system identification using best linear approximation within the LS-SVM framework. In *Proc. of the 3rd Latin American conference on computational intelligence*.

Christmann, A., & Steinwart, I. (2007). Consistency and robustness of kernel-based regression in convex risk minimization. *Bernoulli*, 799–819.

Eckhard, D., Bazanella, A. S., Rojas, C. R., & Hjalmarsson, H. (2017). Cost function shaping of the output error criterion. *Automatica*, *76*, 53–60.

Enqvist, M., & Ljung, L. (2005). Linear approximations of nonlinear FIR systems for separable input processes. *Automatica*, *41*(3), 459–473.

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *Springer series in statistics*: *Vol. 1. The elements of statistical learning*. Berlin: Springer.

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

Giri, F., Rochdi, Y., & Chaoui, F. (2008). Comment on Frequency domain identification of Wiener models, by EW Bai, Automatica 39 (2003), 1521–1530. *Automatica*, *44*(5), 1451–1455.

Giri, F., Rochdi, Y., & Chaoui, F.-Z. (2009). An analytic geometry approach to Wiener system frequency identification. *IEEE Transactions on Automatic Control*, *54*(4), 683–696.

Greblicki, W. (1997). Nonparametric approach to Wiener system identification. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, *44*(6), 538–545.

Hagenblad, A., Ljung, L., & Wills, A. (2008). Maximum likelihood identification of Wiener models. *Automatica*, *44*(11), 2697–2705.

Hunter, I., & Korenberg, M. (1986). The identification of nonlinear biological systems: Wiener and Hammerstein cascade models. *Biological Cybernetics*, *55*(2–3), 135–144.

Janczak, A. (2007). Instrumental variables approach to identification of a class of MIMO Wiener systems. *Nonlinear Dynamics*, *48*(3), 275–284.

Kalafatis, A., Arifin, N., Wang, L., & Cluett, W. (1995). A new approach to the identification of pH processes based on the Wiener model. *Chemical Engineering Science*, *50*(23), 3693–3701.

Lacy, S. L., Erwin, R. S., & Bernstein, D. S. (2001). Identification of Wiener systems with known noninvertible nonlinearities. In *American control conference, 2001. Proceedings of the 2001, Vol. 6* (pp. 4890–4895). IEEE.

Lindsten, F., Schön, T. B., & Jordan, M. I. (2013). Bayesian semiparametric Wiener system identification. *Automatica*, *49*(7), 2053–2063.

Ljung, L. (1999). *Prentice Hall information and system sciences series. System identification : theory for the user*. Upper Saddle River (NJ): Prentice Hall PTR.

Ljung, L., Zhang, Q., Lindskog, P., Iouditski, A., & Singh, R. (2006). An integrated system identification toolbox for linear and nonlinear models. In *Proceedings of the 4th IFAC symposium on system identification*, Newcastle, Australia.

Mahata, K., Schoukens, J., & De Cock, A. (2016). Information matrix and D-optimal design with Gaussian inputs for Wiener model identification. *Automatica*, *69*, 65–77.

Mzyk, G. (2007). A censored sample mean approach to nonparametric identification of nonlinearities in Wiener systems. *IEEE Transactions on Circuits and Systems II: Express Briefs*, *54*(10), 897–901.

Pillonetto, G. (2013). Consistent identification of Wiener systems: A machine learning viewpoint. *Automatica*, *49*(9), 2704–2712.

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

Rallo, G., Formentin, S., Rojas, C. R., Oomen, T., & Savaresi, S. M. (2017). Data-drive $\mathcal{H}_\infty$-norm estimation via expert advice. In *Proceedings of the 56th IEEE conference on decision and control* (pp. 1560–1565). IEEE.

Schoukens, M., & Tiels, K. (2017). Identification of block-oriented nonlinear systems starting from linear approximations: A survey. *Automatica*, *85*, 272–292.

Steinwart, I. (2005). Consistency of support vector machines and other regularized kernel classifiers. *IEEE Transactions on Information Theory*, *51*(1), 128–142.

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

Tiels, K., & Schoukens, J. (2011). Identifying a Wiener system using a variant of the Wiener G-functionals. In *2011 50th IEEE conference on decision and control and European control conference* (pp. 5780–5785). IEEE.

Westwick, D., & Verhaegen, M. (1996). Identifying MIMO Wiener systems using subspace model identification methods. *Signal Processing*, *52*(2), 235–258.

Wigren, T. (1993). Recursive prediction error identification using the nonlinear Wiener model. *Automatica*, *29*(4), 1011–1025.

Wigren, T. (1994). Convergence analysis of recursive identification algorithms based on the nonlinear Wiener model. *IEEE Transactions on Automatic Control*, *39*(11), 2191–2206.

Xavier-de Souza, S., Suykens, J. A. K., Vandewalle, J., & Bollé, D. (2009). Coupled simulated annealing. *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, *40*(2), 320–335.

Zhu, Y. (1999). Distillation column identification for control using Wiener model. In *American control conference, 1999. Proceedings of the 1999, Vol. 5* (pp. 3462–3466). IEEE.



Giulio Bottegal was born in 1984 in Italy. He received the M.S. degree in Automation Engineering (cum laude) in 2009 and the Ph.D. degree in Information Engineering in 2013 from the University of Padova. From 2013 to 2016 he was a postdoctoral researcher at the School of Electrical Engineering, KTH Royal Institute of Technology, Sweden. From 2016 to 2017 he was a postdoctoral researcher at ESAT, KU Leuven, Belgium. He held visiting research positions at the Australian National University and the University of Melbourne. Currently, he holds a position as postdoctoral researcher at the Department of Electrical Engineering, TU Eindhoven, The Netherlands. His research interests include system identification and machine learning.



Ricardo Castro-Garcia was born in Medellín, Colombia. He received a B.Sc. in Electronic Engineering from Universidad Pontificia Bolivariana in 2004. In 2010, he obtained a post graduate degree in industrial maintenance at EAFIT University and another one in project management from Universidad Pontificia Bolivariana in 2012. From 2003 to 2009 he occupied various positions as engineer at Crystal Hosiery group, and since 2009 until 2012 he joined Ross International group as the maintenance director of Andes International Tooling and Moldes Medellin companies.

In 2012 Ricardo moved to Leuven, Belgium, where he obtained a M.Sc. in artificial intelligence, from KU Leuven. Since 2013 to 2017 he worked on his Ph.D. degree in the Department of Electrical Engineering (ESAT-STADIUS) of KU Leuven under the supervision of professors Johan A.K. Suykens and Johan Schoukens (VUB). Currently, Ricardo works at SILO.AI in Helsinki, Finland, as an artificial intelligence researcher.

Ricardo's research focus has been on Machine Learning applied to nonlinear system identification. In particular, he has used kernel methods to develop new and more powerful methodologies for the identification of Block Oriented Nonlinear systems. These methodologies include different reformulations of existing black-box algorithms to incorporate additional available information into models that otherwise would be solely driven by the relationship of the input and output signals.



Johan A.K. Suykens was born in Willebroek Belgium, May 18 1966. He received the master degree in Electro-Mechanical Engineering and the Ph.D. degree in Applied Sciences from the Katholieke Universiteit Leuven, in 1989 and 1995, respectively. In 1996 he has been a Visiting Postdoctoral Researcher at the University of California, Berkeley. He has been a Postdoctoral Researcher with the Fund for Scientific Research FWO Flanders and is currently a full Professor with KU Leuven. He is author of the books "Artificial Neural Networks for Modelling and Control of Non-linear Systems" (Kluwer Academic Publishers) and "Least Squares Support Vector Machines" (World Scientific), co-author of the book "Cellular Neural Networks, Multi-Scroll Chaos and Synchronization" (World Scientific) and editor of the books "Nonlinear Modeling: Advanced Black-Box Techniques" (Kluwer Academic Publishers),"Advances in Learning Theory: Methods, Models and Applications" (IOS Press) and "Regularization, Optimization, Kernels, and Support Vector Machines" (Chapman & Hall/CRC). In 1998 he organized an International Workshop on Nonlinear Modelling with Time-series Prediction Competition. He has served as associate editor for the IEEE Transactions on Circuits and Systems (1997–1999 and 2004–2007), the IEEE Transactions on Neural Networks (1998–2009) and the IEEE Transactions on Neural Networks and Learning Systems (from 2017). He received an IEEE Signal Processing Society 1999 Best Paper (Senior) Award and several Best Paper Awards at International Conferences. He is a recipient of the International Neural Networks Society INNS 2000 Young Investigator Award for significant contributions in the field of neural networks. He has served as a Director and Organizer of the NATO Advanced Study Institute on Learning Theory and Practice (Leuven 2002), as a program co-chair for the International Joint Conference on Neural Networks 2004 and the International Symposium on Nonlinear Theory and its Applications 2005, as an organizer of the International Symposium on Synchronization in Complex Networks 2007, a co-organizer of the NIPS 2010 workshop on Tensors, Kernels and Machine Learning, and chair of ROKS 2013. He has been awarded an ERC Advanced Grant 2011 and has been elevated IEEE Fellow 2015 for developing least squares support vector machines.