A two-experiment approach to Wiener system identification

Giulio Bottegal [a], Ricardo Castro-Garcia [a], Johan A. K. Suykens [a]

[a] *ESAT-Stadius, KU Leuven, Kastelpaark Arenberg 10, B-3001 Leuven (Heverlee), Belgium*

Abstract

Wiener systems consist of the coupling of a linear time-invariant (LTI) system and a static nonlinearity. In this paper 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.

Key words: system identification; Wiener systems; experiment design; least squares support vector machines

1 Introduction

Block-oriented system identification aims at representing systems as interconnected linear and nonlinear blocks. The Wiener system [1] 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.

Any nonlinear system can be approximated by a Wiener model with arbitrarily high accuracy, as shown in [2]. Moreover, Wiener models have proved to be useful to represent nonlinear systems in many application areas, e.g., chemical processes [3], [4], and biological systems [5]. For these reasons, identification of Wiener systems has been object of intensive research for many years; an overview of previous works can be found in [6].

Maximum likelihood/prediction error techniques are discussed in [7]. 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 [8], or build recursive identification schemes [9]. In [10], a subspace-based method is proposed. Nonparametric methods based on a weighted kernel regression are discussed in several contributions [11,12,13,14], 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 [15] and [16]. In [15] the LTI block and the static nonlinearity are estimated together using a joint maximum-a-posteriori/maximum-likelihood criterion, which requires the solution of a nonlinear optimization problem. On the other hand, the method proposed in [16] provides a minimum mean square estimate of the system, but uses a computationally demanding integration scheme based on Markov Chain Monte Carlo techniques. A frequency-domain approach is proposed in [17], while other approaches based

⋆ 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).

Email addresses: giulio.bottegal@esat.kuleuven.be (Giulio Bottegal), ricardo.castro@esat.kuleuven.be (Ricardo Castro-Garcia), johan.suykens@esat.kuleuven.be (Johan A. K. Suykens).

on instrumental variables and the Wiener G-functionals are found in [18] and [19], respectively. Experiment design techniques specifically tailored for Wiener system identification are discussed in [20] and [21].

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 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 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 [22], 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.

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 identifiy the LTI block by means of a modified version of the standard prediction error method (PEM) for linear output-error (OE) systems [23]. The computational burden of this second step reduces essentially to the one of PEM for OE systems.

The proposed framework is tested via numerical experiments showing its effectiveness compared to other identification techniques for Wiener systems.

The paper is organized as follows. In Section 2 we define the problem under study. Section 3 describes the proposed method including its parametric and nonparametric versions. Section 4 illustrates the results found when applying the described methodology on two simulation examples and compares the obtained results with other methods in the literature. Finally, in Section 5 some conclusions are presented.

2 Wiener system identification using a two-experiment approach

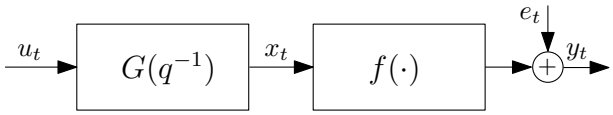

Figure 1. Block scheme representation of a Wiener system.

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 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 unknown variance σ^2. 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. Use the steady-state data to estimate the nonlinear function $f(\cdot)$.
(2) Feed a 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

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

so that the LTI subsystem is completely characterized by the parameter vector $\theta := \begin{bmatrix} b_0 & b_1 & \ldots & b_m & a_1 & \ldots & a_n \end{bmatrix}$. 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

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

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 optimization problem, and in both cases gradient-based methods can be easily applied [23]. Moreover, it has been shown in [24] that recursive PEM schemes for Wiener system identification with known nonlinearity are guaranteed to converge under mild assumptions.

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.

Remark 1 *The method can be easily extended to the case where the additive noise is colored. In that case, the parameters of the noise shaping filter can be estimated either in the first or in the second phase of the proposed procedure.*

3 Three 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 following input signal

$$u_t = \sin(\omega t + \phi_0) ,$$

where ω is an user-prescribed frequency and ϕ_0 is a known phase delay. Then, after the transient effect of $G(q^{-1})$ has vanished, we have that

$$x_t = A_\omega \sin(\omega t + \phi_0 + \phi_\omega) ,$$

where A_ω and ϕ_ω are the gain and the phase delay of the LTI subsystem $G(q^{-1})$ at the frequency ω [23, Ch. 2]. Due to the structural non-identifiability of Wiener systems, A_ω can not be determined (see Remark 2 below). We thus drop it and define a new signal

$$\bar{x}_t = \sin(\omega t + \phi_0 + \phi_\omega) ,$$

which is parameterized by the unknown quantity ϕ_ω. Accordingly, we write the output of the system as

$$y_t = f(\sin(\omega t + \phi_0 + \phi_\omega)) + e_t . \quad (4)$$

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 2 *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 [25]) 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.*

3.1 Parametric approach

We assume that there exist a set of known basis functions $h_0(\cdot)$, $h_1(\cdot)$, ..., $h_p(\cdot)$ 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, ..., c_p. We rewrite (4) as

$$y_t = \sum_{i=0}^{p} c_i h_i(\sin(\omega t + \phi_0 + \phi_\omega)) + e_t . \quad (5)$$

Let $H(\phi_\omega) \in \mathbb{R}^{N_1 \times p+1}$ be a matrix such that

$$[H(\phi_\omega)]_{t,i} = h_i(\sin(\omega t + \phi_0 + \phi_\omega)) .$$

Then we have the following regression model

$$y = H(\phi_\omega)c + e , \quad (6)$$

where

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_{N_1} \end{bmatrix} \quad , \quad e = \begin{bmatrix} e_1 \\ \vdots \\ e_{N_1} \end{bmatrix} \quad , \quad c = \begin{bmatrix} c_0 \\ \vdots \\ c_p \end{bmatrix} . \quad (7)$$

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:

$$y = H(\phi)c + e\,. \tag{8}$$

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

$$\hat{c}(\phi) = \left(H^T(\phi)H(\phi)\right)^{-1} H^T(\phi)y\,. \tag{9}$$

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

$$\hat{\phi}_\omega = \arg\min_{\phi\in\mathcal{I}} \frac{1}{N_1}\|y - \hat{f}_\phi\|^2\,, \tag{10}$$

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 (10) admits an 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*

$$h(\sin(\omega t + \phi_1)) = \alpha h(\sin(\omega t + \phi_2))\,, \tag{11}$$

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

Therefore, if $f(\cdot)$ is parameterized through a set of phase-indistinguishable basis functions, criterion (10) 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* (11) *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.

Proposition 5 *Let $\mathcal{I} = (-\pi, 0]$. It holds that*

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

Similarly, we show that the phase estimation is consistent.

Proposition 6 *Let e_t be a stochastic stationary ergodic process with finite variance. Then the asymptotic (in N_1) solution of the problem* (10) *is equal to ϕ_ω.*

Algorithm 1 summarizes the first step of the proposed procedure, under a parametric modeling assumption of the static nonlinearity. Albeit (10) 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 $\mathcal{I}$.

Algorithm 1 Parametric static nonlinearity estimation

Require: $\{y_t\}_{t=1}^{N_1}$, ω, ϕ_0
Ensure: $\hat{c}_0, \ldots, \hat{c}_p$
1: Select a subset of values $\phi \in \mathcal{I}$.
2: For every ϕ solve (9) and evaluate $\|y - \hat{f}_\phi\|^2$
3: Choose $\hat{c}(\phi)$ associated with ϕ solving (10)

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

$$f(x) = c_0 + c_1 x + \ldots + c_p x^p\,. \tag{12}$$

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. [26], [27]), for $i \in \mathbb{N}$,

$$\begin{aligned}
\sin^{2i}(x) &= \frac{1}{2^{2i}}\binom{2i}{i} \\
&\quad + \frac{(-1)^i}{2^{2i-1}}\sum_{k=0}^{i-1}(-1)^k\binom{2i}{k}\cos(2(i-k)x)\,, \\
\sin^{2i+1}(x) &= \frac{(-1)^i}{4^i}\sum_{k=0}^{i}(-1)^k\binom{2i+1}{k} \\
&\quad \times \sin((2i+1-2k)x)\,,
\end{aligned}$$

and that, for any $q \in \mathbb{Z}$,

$$\begin{aligned}
\sin(q(wt+\phi_0+\phi_\omega)) &= \cos(q\phi_\omega)\sin(q(wt+\phi_0)) \\
&\quad + \sin(q\phi_\omega)\cos(q(wt+\phi_0)) \\
\cos(q(wt+\phi_0+\phi_\omega)) &= -\sin(q\phi_\omega)\sin(q(wt+\phi_0)) \\
&\quad + \cos(q\phi_\omega)\cos(q(wt+\phi_0))\,.
\end{aligned}$$

Then, we can rewrite the nonlinear function as

$$\begin{aligned}
f(\bar{x}_t) = \sum_{i=0}^{p/2} c_{2i}\Big[&\frac{1}{2^{2i}}\binom{2i}{i} + \frac{(-1)^i}{2^{2i-1}}\sum_{k=0}^{i-1}(-1)^k\binom{2i}{k} \\
&\times [\cos(2(i-k)\phi_\omega)\cos(2(i-k)(wt+\phi_0)) \\
&- \sin(2(i-k)\phi_\omega)\sin(2(i-k)(wt+\phi_0))]\Big] \\
+ \sum_{i=0}^{p/2-1} c_{2i+1}&\frac{(-1)^i}{4^i}\sum_{k=0}^{i}(-1)^k\binom{2i+1}{k} \\
&\times [\cos((2i+1-2k)\phi_\omega)\sin((2i+1-2k)(wt+\phi_0))
\end{aligned}$$

$$+\sin((2i+1-2k)\phi_\omega)\cos((2i+1-2k)(wt+\phi_0))]\,. \tag{13}$$

This equation is particularly interesting because it permits to express $f(\bar{x}_t)$ as a linear combination of sines and cosines with frequency $q\omega_t$ and known phase delay $q\phi_0$, 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

$$\mathbf{k} = Mc\,, \quad \mathbf{k} := \begin{bmatrix} k_0 & \ldots & k_p \end{bmatrix}^T . \tag{14}$$

Then we can write

$$\bar{\mathbf{k}} = \begin{bmatrix} k_0 \\ k_{s,1} \\ k_{c,1} \\ \vdots \\ k_{s,p} \\ k_{c,p} \end{bmatrix} = \begin{bmatrix} \gamma_0 k_0 \\ \gamma_{s,\,1} k_1 \sin(\phi_\omega) \\ \gamma_{c,\,1} k_1 \cos(\phi_\omega) \\ \vdots \\ \gamma_{s,\,p} k_p \sin(p\phi_\omega) \\ \gamma_{c,\,p} k_p \cos(p\phi_\omega) \end{bmatrix}, \tag{15}$$

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

$$\psi_t^T := [1 \;\; \cos(\omega t + \phi_0) \;\; \sin(\omega t + \phi_0) \;\; \ldots \;\; \ldots \;\; \sin(p(\omega t + \phi_0)) \;\; \cos(p(\omega t + \phi_0))]\,, \tag{16}$$

so that we can rewrite (13) as $f(\bar{x}_t) = \psi_t^T \bar{\mathbf{k}}$, and express the measurement equation via the linear regression model

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

We can then obtain the least-squares estimate

$$\widehat{\bar{\mathbf{k}}} = \left(\sum_{t=1}^{N_1} \psi_t \psi_t^T \right)^{-1} \sum_{t=1}^{N_1} \psi_t y_t\,. \tag{17}$$

Using this estimate, we now show how to recover the coefficients of the polynomial. From (15), the absolute value of each k_i, $i = 1, \ldots, p$, can be reconstructed via

$$|\hat{k}_i| = \sqrt{\hat{k}_{s,i}^2 + \hat{k}_{c,i}^2}\,. \tag{18}$$

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

$$\hat{\phi}_\omega = \tan^{-1}\left(\frac{\hat{k}_{s,1}}{\hat{k}_{c,1}} \right), \tag{19}$$

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 (15). Finally, the estimate $\hat{c}$ can be recovered via

$$\hat{c} = M^{-1}\hat{\mathbf{k}}\,, \tag{20}$$

where M is defined in (14).

We summarize the procedure for the polynomial case in Algorithm 2.

Algorithm 2 Polynomial static nonlinearity estimation

Require: $\{y_t\}_{t=1}^{N_1}$, ω, ϕ_0
Ensure: $\hat{c}_0, \ldots, \hat{c}_p$

1: Construct the regressors (16) and compute the least-squares estimate (17)
2: Recover the absolute value of the coefficients k_i, $i = 0, \ldots, p$ using (18)
3: Estimate the phase shift via (19) and the sign of the coefficients k_i, $i = 0, \ldots, p$
4: Recover the coefficients c_i, $i = 0, \ldots, p$, using (20)

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 (19), 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.

Proposition 7 *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*

$$\frac{\sigma^2}{N_1} M^{-1} D M^{-T}, \tag{21}$$

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

Example 8 *Consider the third-order polynomial nonlinearity*

$$f(x) = c_0 + c_1 x + c_2 x^2 + c_3 x^3 .$$

We have

$$\begin{aligned} f(\bar{x}_t) = c_0 + \frac{c_2}{2} &+ (c_1 + \frac{3}{4}c_3)\cos(\phi_\omega)\sin(\omega t + \phi_0) \\ &+ (c_1 + \frac{3}{4}c_3)\sin(\phi_\omega)\cos(\omega t + \phi_0) \\ &+ \frac{c_2}{2}\sin(2\phi_\omega)\sin(2(\omega t + \phi_0)) \\ &- \frac{c_2}{2}\cos(2\phi_\omega)\cos(2(\omega t + \phi_0)) \\ &- \frac{c_3}{4}\cos(3\phi_\omega)\sin(3(\omega t + \phi_0)) \\ &- \frac{c_3}{4}\sin(3\phi_\omega)\cos(3(\omega t + \phi_0)) , \end{aligned}$$

and

$$\begin{bmatrix} \gamma_0 \\ \gamma_{s,1} \\ \gamma_{c,1} \\ \gamma_{s,2} \\ \gamma_{c,2} \\ \gamma_{s,3} \\ \gamma_{c,3} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ -1 \\ -1 \\ -1 \end{bmatrix}, \; M = \begin{bmatrix} 1 & 0 & 1/2 & 0 \\ 0 & 1 & 0 & 3/4 \\ 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1/4 \end{bmatrix} .$$

The asymptotic covariance of the coefficient estimates, computed via (21)*, is then*

$$\frac{\sigma^2}{N_1} \begin{bmatrix} 3 & 0 & -4 & 0 \\ 0 & 20 & 0 & -24 \\ -4 & 0 & 8 & 0 \\ 0 & -24 & 0 & 32 \end{bmatrix},$$

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

3.3 Nonparametric approach

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

$$f(\bar{x}_t) = w^T \varphi(\bar{x}_t) + b , \tag{22}$$

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

$$K(\bar{x}_i, \bar{x}_j) = \exp\left(\frac{-(\bar{x}_i - \bar{x}_j)^2}{\eta}\right), \tag{23}$$

where $\eta \geq 0$ is usually called a tuning hyperparameter.

The problem of estimating $f(\cdot)$ can then be formulated in the primal space as

$$\begin{aligned} \min_{w,b,e} \quad & \tfrac{1}{2} w^T w + \tfrac{\gamma}{2} \sum_{t=1}^{N_1} e_t^2 \\ \text{subject to } & y_t = w^T \varphi(\bar{x}_t) + b + e_t, \; t = 1, \ldots, N_1 , \end{aligned} \tag{24}$$

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). Computing the Lagrangian associated with this problem we get

$$\begin{aligned} \mathcal{L}(w, b, e; \alpha) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{t=1}^{N_1} e_t^2 \\ - \sum_{t=1}^{N_1} \alpha_t (w^T \varphi(\bar{x}_t) + b + e_t - y_t) , \end{aligned} \tag{25}$$

where $\alpha_t \in \mathbb{R}$ are the Lagrange multipliers. We can derive the following optimality conditions:

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial w} = 0 \to w = \sum_{t=1}^{N_1} \alpha_t \varphi(\bar{x}_t) \\ \frac{\partial \mathcal{L}}{\partial b} = 0 \to \sum_{t=1}^{N_1} \alpha_t = 0 \\ \frac{\partial \mathcal{L}}{\partial e_t} = 0 \to \alpha_t = \gamma e_t,\ t = 1, \ldots, N_1 \\ \frac{\partial \mathcal{L}}{\partial \alpha_t} = 0 \to y_t = w^T \varphi(\bar{x}_t) + b + e_t,\ t = 1, \ldots, N_1\,. \end{cases} \tag{26}$$

By eliminating w and e_t we obtain the following linear system

$$\left[\begin{array}{c|c} 0 & \mathbf{1}_{N_1}^T \\ \hline \mathbf{1}_{N_1} & \Omega + \frac{1}{\gamma} I_{N_1} \end{array}\right] \left[\begin{array}{c} b \\ \hline \alpha \end{array}\right] = \left[\begin{array}{c} 0 \\ \hline y \end{array}\right], \tag{27}$$

where y is defined in (7), $\alpha = [\alpha_1 \ \ldots \ \alpha_N]^T$, 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 and recalling from (26) that $w = \sum_{t=1}^{N_1} \alpha_t \varphi(\bar{x}_t)$, 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

$$\hat{f}(\tilde{x}) = \sum_{t=1}^{N_1} \alpha_t K(\bar{x}_t, \tilde{x}) + b\,. \tag{28}$$

The obtained estimate depends on the regularization parameter γ entering (24), the kernel shaping hyperparameter η in (23), and also 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 [28]. 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.

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 $\bar{x}_t = \sin(\omega t + \phi_0 + \phi_1)$ and $\bar{x}_t = \sin(\omega t + \phi_0 + \phi_2)$ 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 $\phi_1 = \phi_2 + k\pi$, $k \in \mathbb{Z}$.*

Proof Follows from the fact that

$$\sin(\omega t + \phi_0 + \phi_1) = \begin{cases} \sin(\omega t + \phi_0 + \phi_1 + k\pi) & \text{k even} \\ -\sin(\omega t + \phi_0 + \phi_1 + k\pi) & \text{k odd} \end{cases} \tag{29}$$

and the fact that the kernel is isotropic. □

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* (23). *We have that*

$$\begin{aligned} (\bar{x}_i - \bar{x}_j)^2 &= (\sin(\omega i + \phi_0 + \phi_\omega) - \sin(\omega j + \phi_0 + \phi_\omega))^2 \\ &= (\sin\phi_\omega \cos(\omega i + \phi_0) + \cos\phi_\omega \sin(\omega i + \phi_0) \\ &\quad - \sin\phi_\omega \cos(\omega j + \phi_0) + \cos\phi_\omega \sin(\omega j + \phi_0))^2 \\ &= (\sin\phi_\omega(\cos(\omega i + \phi_0) - \cos(\omega j + \phi_0)) \\ &\quad + \cos\phi_\omega(\sin(\omega i + \phi_0) - \sin(\omega j + \phi_0)))^2 \\ &= [z_i^s - z_j^s \ \ z_i^c - z_j^c]\Psi \left[\begin{array}{c} z_i^s - z_j^s \\ z_i^c - z_j^c \end{array}\right], \end{aligned} \tag{30}$$

where $z_i^s = \sin(\omega i + \phi_0)$, $z_i^c = \cos(\omega i + \phi_0)$ (the same notation holds for j) and

$$\Psi = \left[\begin{array}{cc} \cos^2\phi_\omega & \frac{1}{2}\sin 2\phi_\omega \\ \frac{1}{2}\sin 2\phi_\omega & \sin^2\phi_\omega \end{array}\right].$$

Then we can write

$$K(\bar{x}_i, \bar{x}_j) = \exp\left(\frac{-\|z_i - z_j\|_\Psi^2}{\eta}\right),$$

with $z_i = [z_i^s \ z_i^c]^T$, which shows that the RBF kernel with the input $\bar{x}_t = \sin(\omega t + \phi_0 + \phi_\omega)$ can be seen as a RBF kernel with a bi-dimensional known input

$$z_t = [\sin(\omega t + \phi_0) \ \ \cos(\omega t + \phi_0)]\,,$$

weighted by the Mahalanobis distance Ψ [29], 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.

We summarize the procedure for nonparametric estimation of $f(\cdot)$ in Algorithm 3.

Algorithm 3 Nonparametric static nonlinearity estimation

Require: $\{y_t\}_{t=1}^{N_1}$, ω, ϕ_0
Ensure: $f(\tilde{x})$, for any $\tilde{x} \in \mathbb{R}$
1: Tune the hyperparameters γ, η, ϕ_ω
2: Compute $f(\tilde{x})$ using (28)

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

$$a = \begin{bmatrix} 2.46 & 2.26 & -0.77 \end{bmatrix},$$
$$b = 10^{-2} \begin{bmatrix} 0.47 & 1.42 & 1.42 & 0.47 \end{bmatrix},$$
$$c = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}.$$

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

$$a = \begin{bmatrix} -2.49 & 2.53 & -1.18 & 0.22 \end{bmatrix},$$
$$b = \begin{bmatrix} 0.11 & -0.21 & 0.28 & -0.21 & 0.11 \end{bmatrix},$$
$$c = \begin{bmatrix} 0 & 1 & 5 & -0.5 \end{bmatrix}.$$

The two systems are depicted in Fig. 2. The variance of the output noise is obtained by matching different signal-to-noise ratios (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

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

and the normalized mean square error (NMSE), defined as

$$\%\text{NMSE} = 100 \frac{\sqrt{\sum_{t=1}^{N} (y_{t,\text{test}} - \hat{y}_{t,\text{test}})^2}}{\sqrt{\sum_{t=1}^{N} (y_{t,\text{test}} - \text{mean}(y_{t,\text{test}}))^2}}. \quad (32)$$

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 [30] 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 $[u_1^1 \ \dots \ u_{N_1}^1 \ u_1^2 \ \dots \ u_{N_2}^2]$.

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.

SNR	**Method**	**S1**		**S2**	
(dB)		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

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 an 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 cross validation scheme. To solve the optimization problem associated with hyperparameter selection, we use coupled simulated annealing [31] and simplex [32]; an implementation of these

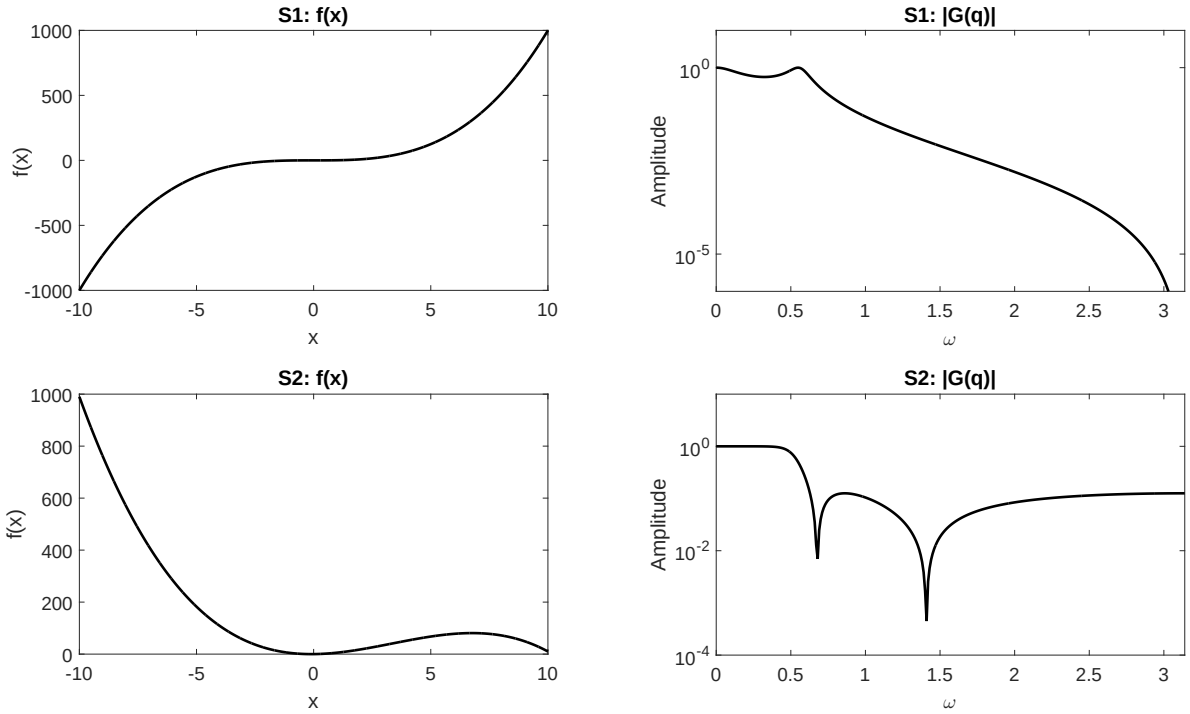

Figure 2. Blocks composing S1 and S2. (Up) S1. (Down) S2. (Left) Nonlinear blocks. (Right) Magnitude of the transfer functions of the LTI blocks.

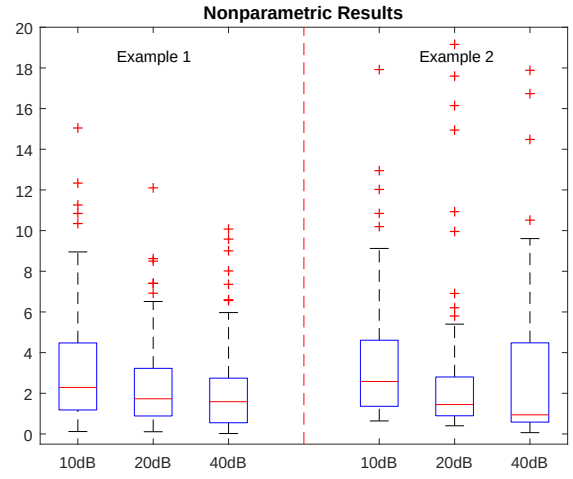

Figure 3. Box plots of the obtained %NMAE for the 100 Monte Carlo runs of the 6 experiments.

techniques is available in the toolbox LS-SVMlab v1.8 [1] . The results of the Monte Carlo simulations are displayed through the box plots of Fig. 3.

The proposed method is compared with the method introduced in [33], where the Best Linear Approximation (BLA) [34] 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 [34]) inputs $u_{t,\mathrm{BLA}}^{(i)}$ with their corresponding outputs $y_{t,\mathrm{BLA}}^{(i)}$. Here $t = 1, \ldots, 2500$ and $i = 1, \ldots, 10$. 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.

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.

[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.

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

5 Conclusions

We have proposed a new method for Wiener system identification. Remarkably, the paper shows that for Wiener systems, a poorly exciting signal –such as a sinusoid– can help estimating (part of) the system by means of relatively simple least-squares based procedures. The main idea underlying the method is that we can separate the estimation of the static nonlinear function from the identificaiton 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.

Future challenges are to extend the two-experiment approach to more involved model structures, such as Wiener-Hammerstein and Hammerstein-Wiener systems.

A Appendix

A.1 Proof of Propositions 5 and 6

We have

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

Now, it is easy to see that, since

$$\left(I - H(\phi)\left(H^T(\phi)H(\phi)\right)^{-1}H^T(\phi)\right)$$

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

$$\left\|\left(I - H(\phi)\left(H^T(\phi)H(\phi)\right)^{-1}H^T(\phi)\right)H(\phi_\omega)c\right\|^2 \geq 0\,. \tag{A.1}$$

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 concludes the proof of Proposition 5. By multiplying the objective function of (10) by $1/N_1$, and letting $N_1 \to \infty$ we have

$$\frac{1}{N_1}\sum_{t=1}^{N_1}(y_t - \hat{f}_\phi(x_t))^2 \to \mathbb{E}\,\|y_t - \hat{f}_\phi(x_t)\|^2\,, \tag{A.2}$$

where the expectation is taken with respect to the distribution of e_t. The result of Proposition 6 follows as a special case of Proposition 5.

A.2 Proof of Proposition 7

Let

$$\Gamma = \begin{bmatrix} \gamma_0 & 0 & \dots & 0 \\ 0 & \gamma_{s,\,1}\sin(\phi_\omega) & 0 \dots & 0 \\ 0 & \gamma_{c,\,1}\cos(\phi_\omega) & 0 \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & \gamma_{s,\,p}\sin(p\phi_\omega) \\ 0 & \dots & 0 & \gamma_{c,\,p}\cos(p\phi_\omega) \end{bmatrix}, \tag{A.3}$$

so that (15) 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, $i = 0, \ldots, p$ can be calculated as the same as the one obtained via least-squares, that is

$$\begin{aligned}\mathbb{E}\left[(\hat{c}-c)(\hat{c}-c)^T\right] &= M^{-1}\mathbb{E}\left[(\hat{\mathbf{k}}-\mathbf{k})(\hat{\mathbf{k}}-\mathbf{k})^T\right]M^{-T} \\ &= M^{-1}\Gamma^T\mathbb{E}\left[(\hat{\bar{\mathbf{k}}}-\bar{\mathbf{k}})(\hat{\bar{\mathbf{k}}}-\bar{\mathbf{k}})^T\right]\Gamma M^{-T} \\ &= \sigma^2 M^{-1}\Gamma^T\left(\sum_{t=1}^{N_1}\psi_t\psi_t^T\right)^{-1}\Gamma M^{-T}.\end{aligned} \tag{A.4}$$

Recalling the structure of ψ_t given in (16), it is straightforward to check that, as N_1 grows large,

$$\frac{1}{N_1}\sum_{t=1}^{N_1}\psi_t\psi_t^T \longrightarrow \operatorname{diag}\left\{1, \frac{1}{2}, \ldots, \frac{1}{2}\right\} := \bar{D}^{-1}, \tag{A.5}$$

where $\bar{D}$ has size $2p+1 \times 2p+1$. Equation (21) 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 (17). This concludes the proof.

References

[1] N. Wiener, "Nonlinear problems in random theory," *Nonlinear Problems in Random Theory. Cambridge, Massachusetts, USA: The MIT Press, August 1966.(Paper)*, p. 142, 1966.

[2] S. Boyd and L. Chua, "Fading memory and the problem of approximating nonlinear operators with Volterra series," *IEEE Transactions on circuits and systems*, vol. 32, no. 11, pp. 1150–1161, 1985.

[3] Y. Zhu, "Distillation column identification for control using Wiener model," in *American Control Conference, 1999. Proceedings of the 1999*, vol. 5. IEEE, 1999, pp. 3462–3466.

[4] A. Kalafatis, N. Arifin, L. Wang, and W. Cluett, "A new approach to the identification of pH processes based on the Wiener model," *Chemical Engineering Science*, vol. 50, no. 23, pp. 3693–3701, 1995.

[5] I. Hunter and M. Korenberg, "The identification of nonlinear biological systems: Wiener and hammerstein cascade models," *Biological cybernetics*, vol. 55, no. 2-3, pp. 135–144, 1986.

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

[7] A. Hagenblad, L. Ljung, and A. Wills, "Maximum likelihood identification of Wiener models," *Automatica*, vol. 44, no. 11, pp. 2697–2705, 2008.

[8] J. Bruls, C. Chou, B. Haverkamp, and M. Verhaegen, "Linear and non-linear system identification using separable least-squares," *European Journal of Control*, vol. 5, no. 1, pp. 116–128, 1999.

[9] T. Wigren, "Recursive prediction error identification using the nonlinear Wiener model," *Automatica*, vol. 29, no. 4, pp. 1011–1025, 1993.

[10] D. Westwick and M. Verhaegen, "Identifying MIMO Wiener systems using subspace model identification methods," *Signal Processing*, vol. 52, no. 2, pp. 235–258, 1996.

[11] W. Greblicki, "Nonparametric approach to Wiener system identification," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 44, no. 6, pp. 538–545, 1997.

[12] ——, "Nonparametric identification of Wiener systems," *IEEE Transactions on information theory*, vol. 38, no. 5, pp. 1487–1493, 1992.

[13] P. Wachel and G. Mzyk, "Direct identification of the linear block in Wiener system," *International Journal of Adaptive Control and Signal Processing*, vol. 30, no. 1, pp. 93–105, 2016.

[14] G. Mzyk, "A censored sample mean approach to nonparametric identification of nonlinearities in Wiener systems," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 54, no. 10, pp. 897–901, 2007.

[15] G. Pillonetto, "Consistent identification of Wiener systems: A machine learning viewpoint," *Automatica*, vol. 49, no. 9, pp. 2704–2712, 2013.

[16] F. Lindsten, T. B. Schön, and M. I. Jordan, "Bayesian semiparametric Wiener system identification," *Automatica*, vol. 49, no. 7, pp. 2053–2063, 2013.

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

[18] A. Janczak, "Instrumental variables approach to identification of a class of MIMO Wiener systems," *Nonlinear Dynamics*, vol. 48, no. 3, pp. 275–284, 2007.

[19] K. Tiels and J. Schoukens, "Identifying a Wiener system using a variant of the Wiener G-functionals," in *Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on*. IEEE, 2011, pp. 5780–5785.

[20] M. Gevers, M. Caenepeel, and J. Schoukens, "Experiment design for the identification of a simple Wiener system," in *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*. IEEE, 2012, pp. 7333–7338.

[21] K. Mahata, J. Schoukens, and A. De Cock, "Information matrix and D-optimal design with Gaussian inputs for Wiener model identification," *Automatica*, vol. 69, pp. 65–77, 2016.

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

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

[24] T. Wigren, "Convergence analysis of recursive identification algorithms based on the nonlinear Wiener model," *IEEE Transactions on Automatic Control*, vol. 39, no. 11, pp. 2191–2206, 1994.

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

[26] W. H. Beyer, "Crc standard mathematical tables," *West Palm Beach, Fl.: Chemical Rubber Co., 1978, 25th ed., edited by Beyer, William H.*, 1978.

[27] A. Novak, L. Simon, F. Kadlec, and P. Lotton, "Nonlinear system identification using exponential swept-sine signal," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 8, pp. 2220–2229, 2010.

[28] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*. Springer series in statistics Springer, Berlin, 2001, vol. 1.

[29] K. Q. Weinberger and G. Tesauro, "Metric learning for kernel regression," in *International Conference on Artificial Intelligence and Statistics*, 2007, pp. 612–619.

[30] L. Ljung, Q. Zhang, P. Lindskog, A. Iouditski, and R. Singh, "An integrated system identification toolbox for linear and nonlinear models," in *Proceedings of the 4th IFAC Symposium on System Identification, Newcastle, Australia*, 2006.

[31] S. Xavier-de Souza, J. A. K. Suykens, J. Vandewalle, and D. Bollé, "Coupled simulated annealing," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 40, no. 2, pp. 320–335, 2009.

[32] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The computer journal*, vol. 7, no. 4, pp. 308–313, 1965.

[33] R. Castro-Garcia and J. A. K. Suykens, "Wiener System Identification using Best Linear Approximation within the LS-SVM framework," in *Proc. of the 3rd Latin American Conference on Computational Intelligence*, 2016.

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