Wiener-Hammerstein system identification with non-Gaussian input ⋆

Grzegorz Mzyk *

* *The Institute of Computer Engineering, Control and Robotics, Wrocław University of Technology, 50-370 Wrocław, Poland, tel: +48 71 320 32 77, e-mail: grzegorz.mzyk@pwr.wroc.pl*

Abstract: The paper addresses the problem of non-parametric estimation of the static characteristic in Wiener-Hammerstein (sandwich) system excited and disturbed by random processes. A new, kernel-like method is presented. The proposed estimate is consistent under small amount of a priori information. An IIR dynamics, non-invertible static non-linearity, and non-Gaussian excitations are admitted. The convergence of the estimate is proved for each continuity point of the static characteristic and the asymptotic rate of convergence is analysed. The results of computer simulation example are included to illustrate the behaviour of the estimate for moderate number of observations.

Keywords: Wiener-Hammerstein system, nonparametric identification, kernel estimate, convergence analysis.

1. INTRODUCTION

The problem of nonlinear dynamic systems modelling by means of block-oriented models has been strongly elaborated for the last four decades, due to vast variety of applications (see e.g. Giannakis and Serpedin [2001]). The conception of block-oriented models assumes that the real plant, as a whole, can be treated as a system of interconnected blocks, static nonlinearities (SN) and linear dynamics (LD), where the interaction signals cannot be measured. The most popular in this class are two-element cascade structures, i.e., Hammerstein-type (SN-LD), Wiener-type (LD-SN), and sandwich-type (LD-SN-LD) representations. Particularly, since in the Wiener system (Fig. 1) the nonlinear block is preceded by the linear dynamics, its identification under random excitation is much more difficult in comparison with the Hammerstein system. However the Wiener model allows for better approximation of many real processes (Celka, et al. [2001], Hunter and Korenberg [1986], Vanbeylen, et al. [2009], Vörös [2007], Westwick and Verhaegen [1996]). Such difficulties in theoretical analysis forced the authors to consider special cases, and to take restrictive assumptions on the input signal, impulse response and the shape of the nonlinear characteristic. In particular, for Gaussian input the problem of Wiener system identification becomes much easier. Since the internal signal $\{x_k\}$ is then also Gaussian, the linear block can be simply identified by the cross-correlation approach (Billings and Fakhouri [1977], and the static characteristic can be recovered e.g. by the non-parametric inverse regression approach Greblicki [1992]-Greblicki and Pawlak [2008]). Non-Gaussian random input is very rarely met in the literature. It is allowed e.g. in Pawlak, et al. [2007], but the algorithm presented there requires prior knowledge of the parametric representation of the linear subsystem. Most recent methods for Wiener system identification assume FIR linear dynamics, invertible nonlinearity, or require the use of specially designed input excitations (Bai and Rayland [2008], Bershad, et al. [2000], Hasiewicz [1987], Lacy and Bernstein [2003], Wigren [1994]).

The estimate proposed in this paper successfully recovers the unknown static nonlinear characteristic of the Wiener-Hammerstein system under poor a priori knowledge about the system. The paper is extension and generalization of the idea presented in Mzyk [2007], and the particular contribution is the following:

- In Sections 2-4, we propose a method of the identification of Wiener systems, and show that it works under mild assumptions on the subsystems and excitations. In particular, in contrast to most earlier papers: (1) the input sequence need not to be a Gaussian white noise, (2) the nonlinear characteristic is not assumed to be invertible, (3) the IIR linear dynamics is admitted, and (4) the algorithm is of nonparametric nature (see e.g. Greblicki and Pawlak [2008]), i.e. it is not assumed that the subsystems can be described with the use of finite number of parameters; in consequence the estimate is free of the possible approximation error. We provide strict convergence proof of our, kernel-based, estimate of the static characteristic, and, in addition to Mzyk [2007], analyze the asymptotic rate of convergence also for the case of IIR dynamic subsystem.
- In Section 5, the convergence proof is generalized for Wiener-Hammerstein system, and Hammerstein system as a particular case of the latter.
- In Section 6 we exploit the idea of the combined parametric-nonparametric approach (see Hasiewicz and Mzyk [2004], Hasiewicz and Mzyk [2009] and Mzyk [2009]) to system identification, and in this context we present the use of the proposed method

⋆ The paper is supported by the grant N N514 3160 33

as a preliminary step for parameter estimation of nonlinear subsystem, when its parametric description is a priori known, i.e., when we are given the closed formula describing the nonlinearity, which includes finite number of unknown parameters. This formula need not to be linear in the parameters.

2. STATEMENT OF THE PROBLEM

We begin from considering a Wiener system, i.e., a tandem two-element connection shown in Fig. 1, where u_k and y_k is a measurable system input and output at time k respectively, z_k is a random noise, $\mu()$ is the unknown characteristic of the static output nonlinearity and $\{\lambda_j\}_{j=0}^{\infty}$ – the unknown impulse response of the linear input dynamics. By assumption, the interaction x_k is not available for measurements. Such a system can be described by the

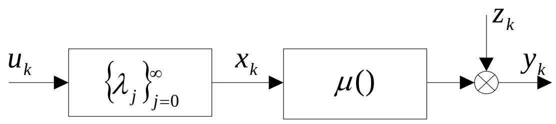

Fig. 1. The Wiener system.

following input-output equation

$$y_k = \mu\left(\sum_{j=0}^{\infty} \lambda_j u_{k-j}\right) + z_k. \qquad (1)$$

2.1 Assumptions

We assume that:

(A1) The input $\{u_k\}$ is an i.i.d., bounded ($|u_k| < u_{\max}$; unknown $u_{\max} < \infty$) random process. There exists a probability density of the input, $\vartheta_u(u_k)$ say, which is a continuous and strictly positive function around the estimation point x, i.e., $\vartheta_u(x) \geqslant \varepsilon > 0$.

(A2) The unknown impulse response $\{\lambda_j\}_{j=0}^{\infty}$ of the linear IIR filter is exponentially upper bounded, that is

$$|\lambda_j| \leqslant c_1 \lambda^j, \text{ some unknown } 0 < c_1 < \infty, \qquad (2)$$

where $0 < \lambda < 1$ is an a priori known constant.

(A3) The nonlinearity $\mu(x)$ is an arbitrary function, continuous almost everywhere on $x \in (-u_{\max}, u_{\max})$ (in the sense of Lebesgue measure).

(A4) The output noise $\{z_k\}$ is a zero-mean stationary and ergodic process, which is independent of the input $\{u_k\}$.

(A5) For simplicity of presentation we also let $L \triangleq \sum_{j=0}^{\infty} \lambda_j = 1$ and $u_{\max} = \frac{1}{2}$.

The goal is to estimate the unknown characteristic of the nonlinearity $\mu(x)$ on the interval $x \in (-u_{\max}, u_{\max})$ on the basis of M input-output measurements $\{(u_k, y_k)\}_{k=1}^{M}$ of the whole Wiener system.

2.2 Comments to assumptions

1) We emphasize, that in **(A2),** we do not assume parametric knowledge of the linear dynamics. In fact, the condition (2), with unknown c_1, is rather not restrictive, and characterizes the class of stable objects. Moreover, observe that, in particular case of FIR linear dynamics, Assumption **(A2)** is fulfilled for arbitrarily small $\lambda > 0$.

2) Assumption **(A5)** is of technical meaning only. We note that the members of the family of Wiener systems composed by series connection of linear filters with the impulse responses $\{\overline{\lambda}_j\} = \{\frac{\lambda_j}{c_2}\}_{j=0}^{\infty}$ and the nonlinearities $\overline{\mu}(x) = \mu(c_2 x)$ are, for $c_2 \neq 0$, indistinguishable from the input-output point of view. In consequence, from the input-output viewpoint, $\mu()$ can be recovered in general only up to some domain scaling factor c_2, independently of the applied identification method.

3) From **(A1)** and **(A2)** it holds that $|x_k| < x_{\max} < \infty$, where $x_{\max} \triangleq u_{\max} \sum_{j=0}^{\infty} |\lambda_j|$. Since $\sum_{j=0}^{\infty} |\lambda_j| \geqslant L$ and $L = 1$ (see **(A5)**), thus the support of the random variables x_k, i.e. $(-x_{\max}, x_{\max})$, is generally wider than the estimation interval $x \in (-u_{\max}, u_{\max})$. In Sections 3-5 we introduce and analyze the nonparametric estimate of the part of characteristic $\mu(x)$, for $x \in (-u_{\max}, u_{\max})$, and in Section 6 we expand the obtained results for $x \in (-x_{\max}, x_{\max})$, when the parametric knowledge of $\mu()$ is provided.

3. BACKGROUND OF THE APPROACH

Let x be a chosen estimation point of $\mu(\cdot)$. For a given x let us define a "weighted distance" between the measurements $u_k, u_{k-1}, u_{k-2}, ..., u_1$ and x as

$$\delta_k(x) \triangleq \sum_{j=0}^{k-1} |u_{k-j} - x| \lambda^j = |u_k - x| \lambda^0 + |u_{k-1} - x| \lambda^1 + ... \\ ... + |u_1 - x| \lambda^{k-1}, \qquad (3)$$

i.e. $\delta_1(x) = |u_1 - x|$, $\delta_2(x) = |u_2 - x| + |u_1 - x| \lambda$, $\delta_3(x) = |u_3 - x| + |u_2 - x| \lambda + |u_1 - x| \lambda^2$, etc., which can be computed recursively as follows

$$\delta_k(x) = \lambda \delta_{k-1}(x) + |u_k - x|. \qquad (4)$$

Making use of assumptions **(A5)** and **(A2)** we obtain

$$\begin{aligned} |x_k - x| &= \left|\sum_{j=0}^{\infty} \lambda_j u_{k-j} - \sum_{j=0}^{\infty} \lambda_j x\right| = \left|\sum_{j=0}^{\infty} \lambda_j (u_{k-j} - x)\right| = \\ &= \left|\sum_{j=0}^{k-1} \lambda_j (u_{k-j} - x) + \sum_{j=k}^{\infty} \lambda_j (u_{k-j} - x)\right| \leqslant \\ &\leqslant \sum_{j=0}^{k-1} |\lambda_j| |u_{k-j} - x| + 2u_{\max} \sum_{j=k}^{\infty} |\lambda_j| \leqslant \\ &\leqslant \delta_k(x) + \frac{\lambda^k}{1-\lambda} \triangleq \Delta_k(x). \end{aligned} \qquad (5)$$

Observe that if in turn

$$\Delta_k(x) \leqslant h(M), \qquad (6)$$

then the true (but unknown) interaction input x_k is located close to x, provided that $h(M)$ (further, a calibration parameter) is small. The distance given in (5) may be easily computed as the point x and the data $u_k, u_{k-1}, u_{k-2}, ..., u_1$ are each time at ones disposal. In turn, the condition (6) selects k's for which the input sequences $\{u_k, u_{k-1}, u_{k-2}, ..., u_1\}$ are such that the true nonlinearity inputs $\{x_k\}$ surely belong to the neighborhood of the estimation point x with the radius $h(M)$. Let us also notice that asymptotically, as $k \to \infty$, it holds that

$$\delta_k(x) = \Delta_k(x), \tag{7}$$

with probability 1.

Proposition 1. If, for each $j = 0, 1, ..., \infty$ and some $d > 0$, it holds that

$$|u_{k-j} - x| \leqslant \frac{d}{\lambda^j}, \tag{8}$$

then

$$|x_k - x| \leqslant d \log_\lambda d + d \frac{1}{1-\lambda}. \tag{9}$$

Proof. The condition (8) is fulfilled with probability 1 for each $j > j_0$, where $j_0 = \lfloor \log_\lambda d \rfloor$ is the solution of the following inequality

$$\frac{d}{\lambda^j} \geqslant 2u_{\max} = 1.$$

On the basis of assumption **(A2)**, analogously as in (5), we obtain

$$|x_k - x| \leqslant \sum_{j=0}^{j_0} \lambda^j \frac{d}{\lambda^j} + \frac{\lambda^{j_0+1}}{1-\lambda} = d\left(j_0 + 1 + \frac{\lambda}{1-\lambda}\right),$$

which yields (9). ■

4. ESTIMATION OF THE STATIC CHARACTERISTIC

4.1 The kernel-like estimate

We propose the following nonparametric kernel-like estimate of the nonlinear characteristic $\mu()$ at the given point x, exploiting the distance $\delta_k(x)$ between x_k and x, and having the form

$$\widehat{\mu}_M(x) = \frac{\sum_{k=1}^{M} y_k \cdot K\left(\frac{\delta_k(x)}{h(M)}\right)}{\sum_{k=1}^{M} K\left(\frac{\delta_k(x)}{h(M)}\right)}, \tag{10}$$

where $K()$ is a window kernel function of the form

$$K(v) = \begin{cases} 1, \text{ as } |v| \leqslant 1 \\ 0, \text{ elsewhere} \end{cases}. \tag{11}$$

Since the estimate (10) is of the ratio form we treat the case 0/0 as 0.

4.2 Limit properties

The convergence The following theorem holds.

Theorem 2. If $h(M) = d(M) \log_\lambda d(M)$, where $d(M) = M^{-\gamma(M)}$, and $\gamma(M) = \left(\log_{1/\lambda} M\right)^{-w}$, then for each $w \in \left(\frac{1}{2}, 1\right)$ the estimate (10) is consistent in the mean square sense, i.e., it holds that

$$\lim_{M\to\infty} E\left(\widehat{\mu}_M(x) - \mu(x)\right)^2 = 0. \tag{12}$$

Proof. Let us denote the probability of selection as $p(M) \triangleq P\left(\Delta_k(x) \leqslant h(M)\right)$. To prove (12) it suffices to show that (see (19) and (22) in Mzyk [2007])

$$h(M) \to 0, \tag{13}$$

$$Mp(M) \to \infty, \tag{14}$$

as $M \to \infty$. They assure vanishing of the bias and variance of $\widehat{\mu}_M(x)$, respectively. Since under assumptions of Theorem 2

$$d(M) \to 0 \Rightarrow h(M) \to 0, \tag{15}$$

in view of (9), the bias-condition (13) is obvious. For the variance-condition (14) we have

$$\begin{aligned}
p(M) &\geqslant P\left\{\bigcap_{j=0}^{\min(k,j_0)} \left(|u_{k-j} - x| < \frac{d(M)}{\lambda^j}\right)\right\} \geqslant \\
&\geqslant P\left\{\bigcap_{j=0}^{\min(k,j_0)} \left(|u_{k-j} - x| < \frac{d(M)}{\lambda^j}\right)\right\} = \\
&= \prod_{j=0}^{j_0} P\left(|u_{k-j} - x| < \frac{d(M)}{\lambda^j}\right) \geqslant \\
&\geqslant \varepsilon \frac{d(M)}{\lambda^0} \cdot \varepsilon \frac{d(M)}{\lambda^1} \cdot ... \cdot \varepsilon \frac{d(M)}{\lambda^{j_0}} = \\
&= \frac{(\varepsilon d(M))^{j_0+1}}{\lambda^{\frac{j_0(j_0+1)}{2}}} = \left(\frac{\varepsilon d(M)}{\lambda^{\frac{j_0}{2}}}\right)^{j_0+1} = \\
&= \left(\varepsilon \sqrt{d(M)}\right)^{j_0+1} = \varepsilon \cdot d(M)^{\frac{1}{2}\log_\lambda d(M) + \log_\lambda \varepsilon + \frac{1}{2}}
\end{aligned} \tag{16}$$

By inserting $d(M) = M^{-\gamma(M)} = (1/\lambda)^{-\gamma(M)\log_{1/\lambda} M}$ to (16) we obtain

$$M \cdot p(M) = \varepsilon \cdot M^{1-\gamma(M)\left(\frac{1}{2}\gamma(M)\log_{1/\lambda} M + \log_\lambda \varepsilon + \frac{1}{2}\right)}. \tag{17}$$

For $\gamma(M) = \left(\log_{1/\lambda} M\right)^{-w}$ and $w \in \left(\frac{1}{2}, 1\right)$ from (17) we simply conclude (14) and consequently (12). ■

The rate of convergence To establish the asymptotic rate of convergence we additionally assume that:

(A6) The nonlinear characteristic $\mu(x)$ is a Lipschitz function, i.e., it exists a positive constant $l < \infty$, such that for each $x_a, x_b \in R$ it holds that $|\mu(x_a) - \mu(x_b)| \leqslant l\,|x_a - x_b|$.

For a window kernel (11) we can rewrite (10) as $\widehat{\mu}_M(x) = \frac{1}{S_0}\sum_{i=1}^{S_0} y_{[i]}$, where $[i]$'s are indexes, for which $K\left(\frac{\delta_{[i]}(x)}{h(M)}\right) = 1$, and S_0 is a random number of selected output measurements. For each $y_{[i]}$, $i = 1, 2, ..., S_0$, respective $x_{[i]}$ is such that $\left|x_{[i]} - x\right| \leqslant h(M)$. On the basis of **(A6)** we obtain

$$\left|\mu(x_{[i]}) - \mu(x)\right| \leqslant lh(M),$$

which for $Ez_k = 0$ (see **(A4)**) leads to

$$|\text{bias}\widehat{\mu}_M(x)| = \left|Ey_{[i]} - \mu(x)\right| = \left|E\mu(x_{[i]}) - \mu(x)\right| \leqslant lh(M),$$

$$\text{bias}^2\widehat{\mu}_M(x) = O\left(h^2(M)\right). \tag{18}$$

For the variance we have

$$\begin{aligned}
\text{var}\widehat{\mu}_M(x) &= \sum_{n=0}^{M} P(S_0 = n) \cdot \text{var}\left(\widehat{\mu}_M(x) | S_0 = n\right) = \\
&= \sum_{n=1}^{M} P(S_0 = n) \cdot \text{var}\left(\frac{1}{n}\sum_{i=1}^{n} y_{[i]}\right).
\end{aligned}$$

Since, under strong law of large numbers and Chebychev inequality, it holds that $\lim_{M\to\infty} P(S_0 > \alpha E S_0) = 1$ for each $0 < \alpha < 1$ (see Mzyk [2007]), we obtain asymptotically

$$\mathrm{var}\widehat{\mu}_M(x) = \sum_{n>\alpha ES_0} P(S_0 = n)\cdot \mathrm{var}\left(\frac{1}{n}\sum_{i=1}^{n} y_{[i]}\right) \tag{19}$$

with probability 1. Taking into account that $y_{[i]} = \overline{y}_{[i]} + z_{[i]}$, where $\overline{y}_{[i]}$ and $z_{[i]}$ are independent random variables we obtain

$$\mathrm{var}\left(\frac{1}{n}\sum_{i=1}^{n} y_{[i]}\right) = \mathrm{var}\left(\frac{1}{n}\sum_{i=1}^{n} \overline{y}_{[i]}\right) + \mathrm{var}\left(\frac{1}{n}\sum_{i=1}^{n} z_{[i]}\right). \tag{20}$$

Since the process $\{z_{[i]}\}$ is ergodic, under strong law of large numbers, it holds that

$$\mathrm{var}\left(\frac{1}{n}\sum_{i=1}^{n} z_{[i]}\right) = O\left(\frac{1}{Mp(M)}\right) = O\left(\frac{1}{M}\right). \tag{21}$$

The process $\{\overline{y}_{[i]}\}$ is in general not ergodic, but in consequence of (6) it has compact support $[\mu(x)-lh(M), \mu(x)+lh(M)]$ and the following inequality holds

$$\mathrm{var}\left(\frac{1}{n}\sum_{i=1}^{n} \overline{y}_{[i]}\right) \leqslant \mathrm{var}\overline{y}_{[i]} \leqslant (2lh(M))^2. \tag{22}$$

From (19), (20), (21) and (22) we conclude that

$$\mathrm{var}\widehat{\mu}_M(x) = O(h^2(M)), \tag{23}$$

which in view of (18) leads to

$$|\widehat{\mu}_M(x) - \mu(x)| = O(h^2(M)) \tag{24}$$

in the mean square sense. A relatively slow rate of convergence, guaranteed in a general case, for $h(M)$ as in Theorem 2, is a consequence of small amount of a priori information. Emphasize that for, e.g., often met in applications piecewise constant functions $\mu(x)$, it exists $M_0 < \infty$, such that $\mathrm{bias}^2\widehat{\mu}_M(x) = 0$ and $\mathrm{var}\left(\frac{1}{n}\sum_{i=1}^{n}\overline{y}_{[i]}\right) = 0$ for $M > M_0$, and consequently $|\widehat{\mu}_M(x) - \mu(x)| = O(\frac{1}{M})$ as $M \to \infty$ (see (21)).

5. OTHER BLOCK-ORIENTED STRUCTURES

In this section we show that under **(A6)** the estimate (10) converges to the true characteristic $\mu(x)$, without any modification, also when applied for Hammerstein systems and for Wiener-Hammerstein systems.

5.1 Hammerstein system

For the Hammerstein system (Fig. 2) described by

$$y_k = \sum_{j=0}^{\infty} \gamma_j \mu(x_{k-j}) + z_k, \tag{25}$$

we assume that the unknown impulse response $\{\gamma_j\}_{j=0}^{\infty}$ fulfils conditions analogous to **(A2)**, and **(A5)**, i.e., $|\gamma_j| \leqslant c_1\lambda^j$, and $G = \sum_{j=0}^{\infty}\gamma_j = 1$.

For Lipschitz function $\mu()$ we simply get

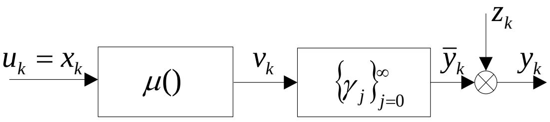

Fig. 2. The Hammerstein system.

$$|\overline{y}_k - \mu(x)| = \left|\sum_{j=0}^{\infty}\gamma_j\mu(x_{k-j}) - \sum_{j=0}^{\infty}\gamma_j\mu(x)\right| =$$

$$= \left|\sum_{j=0}^{\infty}\gamma_j\left(\mu(x_{k-j}) - \mu(x)\right)\right| =$$

$$= \left|\sum_{j=0}^{k-1}\gamma_j\left(\mu(x_{k-j}) - \mu(x)\right) + \sum_{j=k}^{\infty}\gamma_j\left(\mu(x_{k-j}) - \mu(x)\right)\right| \leqslant$$

$$\leqslant \sum_{j=0}^{k-1}|\gamma_j|\,|\mu(x_{k-j}) - \mu(x)| + 2u_{\max}l\sum_{j=k}^{\infty}|\lambda_j| \leqslant$$

$$\leqslant l\delta_k(x) + \frac{l\lambda^k}{1-\lambda} = l\Delta_k(x). \tag{26}$$

If, for a given x, the selection condition (6) is fulfilled, then the noise-free output $\overline{y}_k$ is located close to $\mu(x)$ Under (26), the convergence (12) given in Theorem 2 can be proved for Hammerstein system, with the use of the same technique.

5.2 Wiener-Hammerstein system

Similarly, for the Wiener-Hammerstein (sandwich) system, presented in Fig. 3, we have

$$|\overline{y}_k - \mu(x)| = \left|\sum_{i=0}^{\infty}\gamma_i\mu(x_{k-i}) - \sum_{i=0}^{\infty}\gamma_i\mu(x)\right| =$$

$$= \left|\sum_{i=0}^{\infty}\gamma_i\mu\left(\sum_{j=0}^{\infty}\lambda_j u_{k-i-j}\right) - \sum_{i=0}^{\infty}\gamma_i\mu\left(\sum_{j=0}^{\infty}\lambda_j x\right)\right| =$$

$$= \left|\sum_{i=0}^{\infty}\gamma_i\left[\mu\left(\sum_{j=0}^{\infty}\lambda_j u_{k-i-j}\right) - \mu\left(\sum_{j=0}^{\infty}\lambda_j x\right)\right]\right| \leqslant$$

$$\leqslant l\sum_{i=0}^{\infty}|\gamma_i|\left|\sum_{j=0}^{\infty}\lambda_j\left(u_{k-i-j} - x\right)\right| \leqslant$$

$$\leqslant l\sum_{i=0}^{\infty}|\gamma_i|\sum_{j=0}^{\infty}|\lambda_j|\,|u_{k-i-j} - x| = l\sum_{i=0}^{\infty}\varkappa_i\,|u_{k-i} - x|$$

where the sequence $\{\varkappa_i\}_{i=0}^{\infty}$ is the convolution of $\{|\gamma_i|\}_{i=0}^{\infty}$ with $\{|\lambda_j|\}_{i=0}^{\infty}$, which obviously fulfills the condition $|\varkappa_i| \leqslant \lambda^i$.

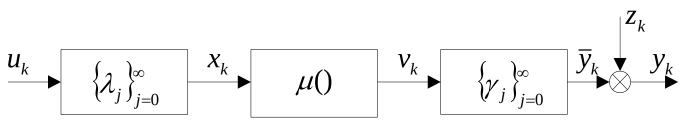

Fig. 3. The Wiener-Hammerstein (sandwich) system.

Remark 3. If the technical assumption **(A5)** is not fulfilled, i.e. the gains $L = \sum_{j=0}^{\infty} \lambda_j$ or $G = \sum_{j=0}^{\infty} \gamma_j$ are not unit, then the estimate (10) converges to the scaled and dilated version $G\mu(Lx)$ of the true system characteristic $\mu(x)$. The constants G and L are not identifiable, since the internal signals x_k and v_k, respectively, cannot be measured.

6. ESTIMATION UNDER PARAMETRIC PRIOR KNOWLEDGE

The presented kernel-type algorithm is applied in this section to support estimation of parameters, when our prior knowledge about the system is large, and in particular, the parametric model of the characteristic is known.

Assume that we are given the class $\mu(x, c)$, such that $\mu(x) \subset \mu(x, c)$, where $c = (c_1, c_2, ..., c_m)^T$ and let us denote by $c^* = (c_1^*, c_2^*, ..., c_m^*)^T$ the vector of true parameters, i.e., $\mu(x, c^*) = \mu(x)$. Let moreover the function $\mu(x, c)$ be by assumption differentiable with respect to c, and the gradient $\bigtriangledown_c \mu(x, c)$ be bounded in some convex neighbourhood of c^* for each x. We assume that c^* is identifiable, i.e., there exists a sequence $x^{(1)}, x^{(2)}, ..., x^{(N_0)}$ of estimation points, such that

$$\mu(x^{(i)}, c) = \mu(x^{(i)}),\ i = 1, 2, ..., N_0 \Longrightarrow c = c^*.$$

The proposed estimate has two steps.

Step 1. For the sequence $x^{(1)}, x^{(2)}, ..., x^{(N_0)}$ compute N_0 pairs

$$\left\{ \left(x^{(i)}, \widehat{\mu}_M(x^{(i)}) \right) \right\}_{i=1}^{N_0},$$

using the estimate (10).

Step 2. Perform the minimization of the cost-function

$$Q_{N_0,M}(c) = \sum_{i=1}^{N_0} \left(\widehat{\mu}_M(x^{(i)}) - \mu(x^{(i)}, c) \right)^2,$$

with respect to variable vector c, and take

$$\widehat{c}_{N_0,M} = \arg\min_c Q_{N_0,M}(c) \tag{27}$$

as the estimate of c^*.

Theorem 4. Since in *Step 1* (nonparametric) for the estimate (10) it holds that $\widehat{\mu}_M(x^{(i)}) \to \mu(x^{(i)})$ in probability as $M \to \infty$ for each $i = 1, 2, ..., N_0$, thus

$$\widehat{c}_{N_0,M} \to c^*$$

in probability, as $M \to \infty$.

Proof. See the proof of Theorem 1 in Hasiewicz and Mzyk [2009]. ■

7. SIMULATION EXAMPLE

In the computer experiment we generated uniformly distributed i.i.d. input sequence $u_k \sim U[-1, 1]$ and the output noise $z_k \sim U[-0.1, 0.1]$. We simulated the IIR linear dynamic subsystems $x_k = 0.5x_{k-1} + 0.5u_k$ and $\overline{y}_k = 0.5\overline{y}_{k-1} + 0.5v_k$, i.e. $\lambda_j = \gamma_j = 0.5^{j+1}$, $j = 0, 1, ..., \infty$, sandwiched with the not invertible and not linear in the parameters static nonlinear characteristic $\mu(x_k) = c_1^* x_k + c_2^* + c_3^* \sin(c_4^* x_k)$, with $c_1^* = 1$, $c_2^* = 0$, $c_3^* = 0.2$ and $c_4^* = 2\pi$. The nonparametric estimate (10) was computed in the $N_0 = 21$ equispaced points $x^{(i)} = -1 + \frac{i-1}{10}$, $i = 1, 2, ..., N_0$.

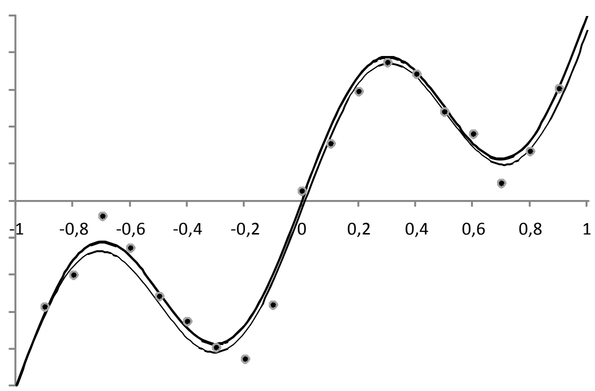

Fig. 4. The true characteristic $\mu(x) = x + 0.2 \sin 2\pi x$ (thick line), its nonparametric estimates $\widehat{\mu}_M(x^{(i)})$ (points), and the parametric model $\mu(x, \widehat{c}_{N_0,M})$ (thin line).

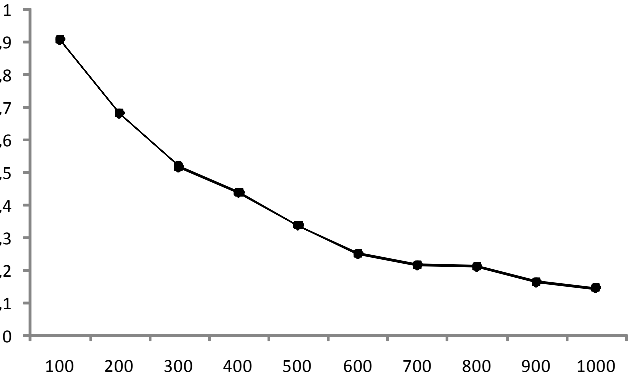

Fig. 5. Estimation error $ERR(\widehat{\mu}_M(x))$ depending on the number of measurements M.

In Assumption **(A2)** we took $\lambda = 0.8$. The estimation error was computed according the rule

$$ERR(\widehat{\mu}_M(x)) = \sum_{i=1}^{N_0} \left(\widehat{\mu}_M(x^{(i)}) - \mu(x^{(i)}) \right)^2.$$

The result of estimation for $M = 300$ is shown in Fig. 4. The criterion in (27) was minimized with the use of classical Levenberg-Marquardt algorithm. Figure 5 illustrates the consistency property.

In the experiment, the characteristic of the static block was changed for $\mu(x) = \sqrt[3]{x}$, which is not Lipschitz at $x = 0$ (cf. (**A6**)). The effect of slower convergence in the neighbourhood of $x = 0$ can be seen in Fig. 6. Next, the routine was repeated for various values of the tuning parameter h. As can be seen in Fig. 7, according to intuition, improper selection of h results in variance or bias augmentation.

8. CONCLUSIONS

The nonlinear characteristic of Wiener system is successfully recovered from the input-output data under small amount of a priori information. The proposed estimate is consistent under IIR dynamics, non-Gaussian input and non-invertible functions. The estimate is universal in the sense that it can be applied, under quite mild conditions,

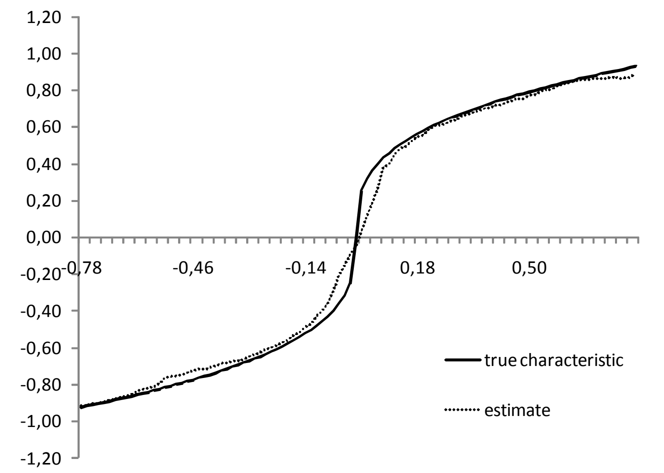

Fig. 6. The true characteristic $\mu(x) = \sqrt[3]{x}$ and its nonparametric estimates $\widehat{\mu}_M(x^{(i)})$.

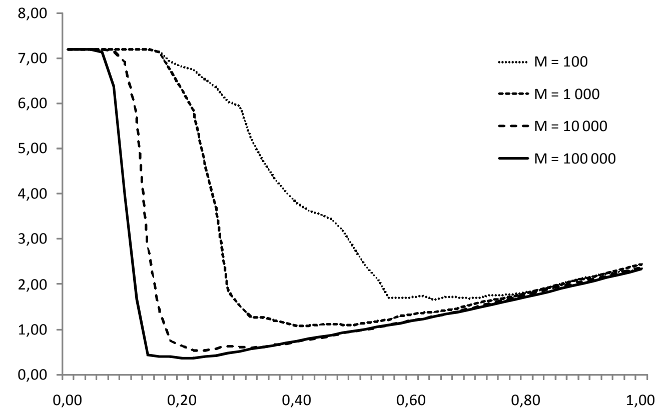

Fig. 7. Relationship between the estimation error $ERR(\widehat{\mu}_M(x))$ and the bandwidth parameter h.

for Hammerstein systems and for Wiener-Hammerstein systems. The strategy allows for decomposition of the identification task of block-oriented system and can support estimation of parameters. Computing of both the estimate $\widehat{\mu}_M(x)$ and the distance $\delta_k(x)$ has the numerical complexity $O(M)$, and can be performed in recursive or semi-recursive version (see Greblicki and Pawlak [2008]). The main limitation is assumed knowledge of λ, i.e., the upper bound of the impulse response. The issue of proper selection of λ is open for further studies. Potential generalizations of the algorithm for unbounded-input case and for other kernel functions seem to be promising.

REFERENCES

E. W. Bai, J. Reyland, "Towards identification of Wiener systems with the least amount of a priori information on the nonlinearity", *Automatica*, vol. 44, No. 4, pp. 910-919, 2008.

N. J. Bershad, P. Celka, J. M. Vesin, "Analysis of stochastic gradient tracking of time-varying polynomial Wiener systems", *IEEE Transactions on Signal Processing*, vol. 48, No. 6, pp. 1676-1686, 2000.

S. A. Billings, S.Y. Fakhouri, "Identification of nonlinear systems using the Wiener model", *Automatica*, vol. 13, No. 17, pp. 502-504, 1977.

P. Celka, N. J. Bershad, J.M. Vesin, "Stochastic gradient identification of polynomial Wiener systems: analysis and application", *IEEE Transactions on Signal Processing*, vol. 49, No. 2, pp. 301-313, 2001.

G. B. Giannakis, E. Serpedin, "A bibliography on nonlinear system identification", *Signal Processing*, vol. 81, pp. 533-580, 2001.

W. Greblicki, "Nonparametric identification of Wiener systems", *IEEE Transactions on Information Theory*, vol. 38, pp. 1487-1493, 1992.

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.

W. Greblicki, M. Pawlak, *Nonparametric System Identification*, Cambridge University Press, 2008.

Z. Hasiewicz, "Identification of a linear system observed through zero-memory non-linearity", *International Journal of Systems Science*, vol. 18, pp. 1595-1607, 1987.

Z. Hasiewicz, G. Mzyk, "Combined parametric-nonparametric identification of Hammerstein systems", *IEEE Transactions on Automatic Control*, vol. 49, pp. 1370-1376, 2004.

Z. Hasiewicz, G. Mzyk, "Hammerstein system identification by non-parametric instrumental variables", *International Journal of Control,* vol. 82, No. 3, pp. 440-455, 2009.

I. W. Hunter, M. J. Korenberg, "The identification of nonlinear biological systems: Wiener and Hammerstein cascade models", *Biological Cybernetics*, vol. 55, pp. 135-144, 1986.

S. L. Lacy, D. S. Bernstein, "Identification of FIR Wiener systems with unknown, non-invertible, polynomial nonlinearities", *International Journal of Control*, vol. 76, No. 15, pp. 1500-1507, 2003.

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.

G. Mzyk, "Nonlinearity recovering in Hammerstein system from short measurement sequence", *IEEE Signal Processing Letters*, vol. 16, No. 9, pp. 762-765, 2009.

M. Pawlak, Z. Hasiewicz, P. Wachel, "On nonparametric identification of Wiener systems", *IEEE Transactions on Signal Processing*, vol. 55, No. 2, pp. 482-492, 2007.

L. Vanbeylen, R. Pintelon, J. Schoukens, "Blind maximum-likelihood identification of Wiener systems", *IEEE Transactions on Signal Processing*, vol. 57, No. 8, pp. 3017-3029, 2009.

J. Vörös, "Parameter identification of Wiener systems with multisegment piecewise-linear nonlinearities", *Systems and Control Letters*, vol. 56, pp. 99-105, 2007.

D. Westwick, M. Verhaegen, "Identifying MIMO Wiener systems using subspace model identification methods", *Signal Processing*, vol. 52, pp. 235-258, 1996.

T. Wigren, "Convergence analysis of recursive identification algorithms based on the nonlinear Wiener model", *IEEE Transactions on Automatic Control*, vol. 39, pp. 2191-2206, 1994.