NONPARAMETRIC RECOVERING OF NONLINEARITY IN WIENER-HAMMERSTEIN SYSTEMS

First Author Name, Second Author Name

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

Abstract

In the paper we recover the static characteristic of Wiener-Hammerstein (sandwich) system from input-output data. The system is excited and disturbed by random processes with arbitrary distribution. Two kernel-based estimates are proposed and compared. It is shown that they can successfully recover the system characteristic under small amount of a priori information about the static characteristic and the surrounding dynamic blocks. The identified nonlinear function is not parametrized and is not assumed to be invertible, which is common restriction in the literature. The orders of linear dynamic blocks are also unknown. The convergence of the estimates take place for the points in which the input probability density function in positive. The effectiveness of the algorithms is illustrated in simulation example.

1 INTRODUCTION

The paper addresses the problem of nonlinearity recovering in block-oriented system of the Wiener-Hammerstein structure (see Fig. 1). It consists of one static nonlinear block with the characteristic $\mu()$, surrounded by two linear dynamic components with the impulse responses $\{\lambda_j\}_{j=0}^{\infty}$ and $\{\gamma_j\}_{j=0}^{\infty}$, respectively. Such a structure, and its particular cases (Wiener systems and Hammerstein systems), are widely considered in the literature because of numerous potential applications in various domains of science and technology (see e.g. (Giannakis and Serpedin (2001))). The Wiener and Wiener-Hammerstein models allow for a good approximation of many real processes ((Celka, et al. (2001)), (Hunter and Korenberg (1986)), (Vanbeylen, et al. (2009)), (Vörös (2007)), (Westwick and Verhaegen (1996))). Nevertheless, serious difficulties in theoretical analysis force the authors to consider only special cases, and to take restrictive assumptions on the input signal, impulse response and the shape of the nonlinear characteristic. In particular, it is commonly assumed that (see e.g. (Billings and Fakhouri (1977)), (Greblicki (1992))-(Greblicki and Pawlak (2008)), (Pawlak, et al. (2007)), (Bai and Rayland (2008)), (Bershad, et al. (2000)), (Lacy and Bernstein (2003)), (Wigren (1994))): (i) the input is a random Gaussian process, a sine wave, or a binary signal, (ii) the static nonlinear block is invertible, (iii) the linear dynamic blocks have finite memory (FIR), and/or, (iv) the parametric representation of subsystems is given a priori.

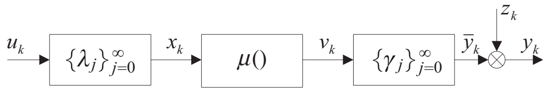

Figure 1: Wiener-Hammerstein (sandwich) system

It was noticed in the paper that the nonparametric algorithms proposed in (Greblicki (2010)) and (Mzyk (2010b)) for a Wiener system, can be adopted, without any modification, for a broad class of Wiener-Hammerstein (sandwich) systems. All the assumptions taken therein remain the same.

In Section 2, the problem is formulated in detail and the assumptions imposed on signals and system components are discussed. Then, in Section 3 we present two nonparametric kernel-based estimates of the nonlinearity, and analyse their properties. Finally, in Section 4, we illustrate their behaviour in simulation example, for various numbers of observations and values of tuning parameters.

We emphasize that in contrast to most earlier papers concerning sandwich system identification:

- the input sequence need not to be a Gaussian white noise,

- the nonlinear characteristic is not assumed to be invertible,
- the IIR linear dynamic blocks are admitted,
- 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 and known number of parameters. In consequence, the estimates are free of the possible approximation error, or this error can be made arbitrarily small by proper selection of tuning parameters.

2 ASSUMPTIONS

We consider a tandem three-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 nonlinearity and $\{\lambda_j\}_{j=0}^{\infty}$, $\{\gamma_j\}_{j=0}^{\infty}$ – the unknown impulse responses of the linear dynamic components. By assumption, the interaction signals x_k and v_k are not available for measurements.

The system is described as follows

$$\begin{aligned} v_k &= \mu\left(\sum_{j=0}^{\infty}\lambda_j u_{k-j}\right), \quad (1) \\ y_k &= \sum_{j=0}^{\infty}\gamma_j v_{k-j} + z_k. \end{aligned}$$

We assume that:

(A1) The input $\{u_k\}$ is an i.i.d., bounded ($|u_k| < u_{\max}$; unknown $u_{\max} < \infty$) random process, and there exists a probability density of the input, say $\vartheta_u(u_k)$, 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 responses $\{\lambda_j\}_{j=0}^{\infty}$ and $\{\gamma_j\}_{j=0}^{\infty}$ of the linear IIR filters are exponentially upper bounded, that is

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

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

(A3) 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

$$|\mu(x_a) - \mu(x_b)| \leqslant l\,|x_a - x_b|\,.$$

(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 $L \triangleq \sum_{j=0}^{\infty}\lambda_j = 1$, $G \triangleq \sum_{j=0}^{\infty}\gamma_j = 1$, 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 N input-output measurements $\{(u_k, y_k)\}_{k=1}^{N}$ of the whole Wiener-Hammerstein system.

From **(A1)** and **(A2)** it holds that $|x_k| < x_{\max} < \infty$, where $x_{\max} \triangleq u_{\max}\sum_{j=0}^{\infty}\left|\lambda_j\right|$.

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.

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 $\lambda > 0$.

3 THE ALGORITHMS

In the paper we propose and compare the following two nonparametric kernel-based estimates of the nonlinear characteristic $\mu()$

$$\widehat{\mu}_N^{(1)}(x) = \frac{\sum_{k=1}^{N} y_k \cdot K\left(\frac{\sum_{j=0}^{k}|u_{k-j}-x|\lambda^j}{h(N)}\right)}{\sum_{k=1}^{N} K\left(\frac{\sum_{j=0}^{k}|u_{k-j}-x|\lambda^j}{h(N)}\right)}, \quad (3)$$

$$\widehat{\mu}_N^{(2)}(x) = \frac{\sum_{k=1}^{N} y_k \prod_{i=0}^{p} K\left(\frac{x-u_{k-i}}{h(N)}\right)}{\sum_{k=1}^{N} \prod_{i=0}^{p} K\left(\frac{x-u_{k-i}}{h(N)}\right)}. \quad (4)$$

In (3) and (4) $K()$ is a bounded kernel function with compact support, i.e., it fulfills the following conditions

$$\begin{aligned} \int_{-\infty}^{\infty} K(x)dx &= 1, \\ \sup_x |K(x)| &< \infty, \quad (5) \\ K(x) &= 0 \text{ for } |x| > x_0, \text{ some } x_0 < \infty. \end{aligned}$$

The sequence $h(N)$ (bandwidth parameter) is such that

$$h(N) \to 0, \text{ as } N \to \infty.$$

The following theorem holds.

Theorem 1. If $h(N) = d(N)\log_{\lambda} d(N)$, where $d(N) = N^{-\gamma(N)}$, and $\gamma(N) = \left(\log_{1/\lambda} N\right)^{-w}$, then for each $w \in \left(\frac{1}{2}, 1\right)$ the estimate (3) is consistent in the mean square sense, i.e., it holds that

$$\lim_{N\to\infty} E\left(\widehat{\mu}_N^{(1)}(x) - \mu(x)\right)^2 = 0. \qquad (6)$$

Proof. 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

$$\delta_k(x) \triangleq \sum_{j=0}^{k-1} \left|u_{k-j} - x\right| \lambda^j = |u_k - x|\lambda^0 + + |u_{k-1} - x|\lambda^1 + ... + |u_1 - x|\lambda^{k-1}, \qquad (7)$$

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

$$\delta_k(x) = \lambda\delta_{k-1}(x) + |u_k - x|. \qquad (8)$$

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

$$\begin{aligned} |x_k - x| &= \left|\sum_{j=0}^{\infty} \lambda_j u_{k-j} - \sum_{j=0}^{\infty} \lambda_j x\right| = \\ &= \left|\sum_{j=0}^{\infty} \lambda_j \left(u_{k-j} - x\right)\right| = \\ &= \left|\sum_{j=0}^{k-1} \lambda_j \left(u_{k-j} - x\right) + \sum_{j=k}^{\infty} \lambda_j \left(u_{k-j} - x\right)\right| \leqslant \\ &\leqslant \sum_{j=0}^{k-1} \left|\lambda_j\right| \left|u_{k-j} - x\right| + 2u_{\max} \sum_{j=k}^{\infty} \left|\lambda_j\right| \leqslant \\ &\leqslant \delta_k(x) + \frac{\lambda^k}{1-\lambda} \triangleq \Delta_k(x). \end{aligned} \qquad (9)$$

Observe that if in turn

$$\Delta_k(x) \leqslant h(N), \qquad (10)$$

then the true (but unknown) interaction input x_k is located close to x, provided that $h(N)$ (further, a calibration parameter) is small. If, for each $j = 0, 1, ..., \infty$ and some $d > 0$, it holds that

$$\left|u_{k-j} - x\right| \leqslant \frac{d}{\lambda^j}, \qquad (11)$$

then

$$|x_k - x| \leqslant d\log_{\lambda} d + d\frac{1}{1-\lambda}. \qquad (12)$$

The condition (11) 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 (9), 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 (12). For the Wiener-Hammerstein (sandwich) system we have

$$\begin{aligned} |\bar{y}_k - \mu(x)| &= \left|\sum_{i=0}^{\infty} \gamma_i \mu(x_{k-i}) - \sum_{i=0}^{\infty} \gamma_i \mu(x)\right| = \\ &= \left|\sum_{i=0}^{\infty} \gamma_i \mu\left(\sum_{j=0}^{\infty} \lambda_j u_{k-i-j}\right) - \sum_{i=0}^{\infty} \gamma_i \mu\left(\sum_{j=0}^{\infty} \lambda_j x\right)\right| = \\ &= \left|\sum_{i=0}^{\infty} \gamma_i \left[\mu\left(\sum_{j=0}^{\infty} \lambda_j u_{k-i-j}\right) - \mu\left(\sum_{j=0}^{\infty} \lambda_j x\right)\right]\right| \leqslant \\ &\leqslant l\sum_{i=0}^{\infty} |\gamma_i| \left|\sum_{j=0}^{\infty} \lambda_j \left(u_{k-i-j} - x\right)\right| \leqslant \\ &\leqslant l\sum_{i=0}^{\infty} |\gamma_i| \sum_{j=0}^{\infty} \left|\lambda_j\right| \left|u_{k-i-j} - x\right| = l\sum_{i=0}^{\infty} \varkappa_i \left|u_{k-i} - x\right| \end{aligned} \qquad (13)$$

where the sequence $\{\varkappa_i\}_{i=0}^{\infty}$ obviously fulfills the condition $|\varkappa_i| \leqslant \lambda^i$. Let us denote the probability of selection as $p(N) \triangleq P(\Delta_k(x) \leqslant h(N))$. To prove (6) it suffices to show that (see (19) and (22) in (Mzyk (2007)))

$$h(N) \rightarrow 0, \qquad (14)$$

$$Np(N) \rightarrow \infty, \qquad (15)$$

as $N \to \infty$. The conditions (14) and (15) assure vanishing of the bias and variance of $\widehat{\mu}_N(x)$, respectively. Since under assumptions of Theorem 3

$$d(N) \to 0 \Rightarrow h(N) \to 0, \qquad (16)$$

in view of (12), the bias-condition (14) is obvious. For the variance-condition (15) we have

$$\begin{aligned} p(N) &\geqslant P\left\{\bigcap_{j=0}^{\min(k,j_0)} \left(\left|u_{k-j} - x\right| < \frac{d(N)}{\lambda^j}\right)\right\} \geqslant \\ &\geqslant P\left\{\bigcap_{j=0}^{\min(k,j_0)} \left(\left|u_{k-j} - x\right| < \frac{d(N)}{\lambda^j}\right)\right\} = \\ &= \prod_{j=0}^{j_0} P\left(\left|u_{k-j} - x\right| < \frac{d(N)}{\lambda^j}\right) \geqslant \\ &\geqslant \varepsilon\frac{d(N)}{\lambda^0} \cdot \varepsilon\frac{d(N)}{\lambda^1} \cdot ... \cdot \varepsilon\frac{d(N)}{\lambda^{j_0}} = \\ &= \frac{(\varepsilon d(N))^{j_0+1}}{\lambda^{\frac{j_0(j_0+1)}{2}}} = \left(\frac{\varepsilon d(N)}{\lambda^{\frac{j_0}{2}}}\right)^{j_0+1} = \\ &= \left(\varepsilon\sqrt{d(N)}\right)^{j_0+1} = \varepsilon \cdot d(N)^{\frac{1}{2}\log_{\lambda} d(N) + \log_{\lambda}\varepsilon + \frac{1}{2}}. \end{aligned} \qquad (17)$$

By inserting $d(N) = N^{-\gamma(N)} = (1/\lambda)^{-\gamma(N)\log_{1/\lambda} N}$ to (17) we obtain

$$N \cdot p(N) = \varepsilon \cdot N^{1-\gamma(N)\left(\frac{1}{2}\gamma(N)\log_{1/\lambda} N + \log_\lambda \varepsilon + \frac{1}{2}\right)}. \quad (18)$$

For $\gamma(N) = \left(\log_{1/\lambda} N\right)^{-w}$ and $w \in \left(\frac{1}{2}, 1\right)$ from (18) we simply conclude (15) and consequently (6). ■

In contrast to $\widehat{\mu}_N^{(1)}(x)$, the estimate $\widehat{\mu}_N^{(2)}(x)$ uses the FIR(p) approximation of the linear subsystems. We will show that since the linear blocks are asymptotically stable, the approximation of $\mu()$ can be made with arbitrary accuracy, i.e., by selecting p large enough. Let us introduce the following regression-based approximation of the true characteristic $\mu()$

$$m_p(x) = E\{y_k | u_k = u_{k-1} = ... = u_{k-2p+1} = x\} \quad (19)$$

and the constants

$$g_p = \sum_{i=0}^{p-1} \gamma_i, \qquad l_p = \sum_{j=0}^{p-1} \lambda_j.$$

The following theorem holds.

Theorem 2. If K() satisfy (5) then it holds that

$$\widehat{\mu}_N^{(2)}(x) \to m_p(l_p x) \text{ in probability}, \quad (20)$$

as $N \to \infty$, at every point x, for which $\vartheta_u(x) > 0$ provided that

$$Nh^{2p}(N) \to \infty, \text{ as } N \to \infty.$$

Proof. The proof is a consequence of (13) and the proof of Theorem 1 in (Greblicki (2010)). ■

From (19) we obtain that

$$m_p(x) =$$
$$= E\left\{\sum_{i=0}^{p-1} \gamma_i \mu(x_{k-i}) + \varsigma | u_k = ... = u_{k-2p+1} = x\right\}$$

where $\varsigma = \sum_{i=p}^{\infty} \gamma_i \mu(x_{k-i})$. Moreover, since $x_k = \sum_{j=0}^{p-1} \lambda_j u_{k-j} + \xi$, where $\xi = \sum_{j=p}^{\infty} \lambda_j u_{k-j}$ it holds that

$$|m_p(l_p x) - \mu(l_p x)| =$$
$$|E\{g_p \mu(l_p x + \xi) + \varsigma\} - \mu(l_p x)| \leq$$
$$\leq E|\{g_p \mu(l_p x + \xi) + \varsigma\} - \mu(l_p x)| \leq$$
$$\leq |g_p - 1| (lEu_k + E\mu(x_k)),$$

and under stability of linear components (see **(A2)** and **(A5)**) we have

$$|g_p - 1| \leq c_0^p, \text{ some } |c_0| < 1.$$

Consequently,

$$\widehat{\mu}_N^{(2)}(x) \to \mu(l_p x) + \varepsilon_p$$

in probability, as $N \to \infty$, where $\varepsilon_p = c_0^p (lu_{\max} + v_{\max}) \phi(x)$, and $|\phi(x)| \leq 1$. Since $\lim_{p\to\infty} l_p = 1$, and $\lim_{p\to\infty} \varepsilon_p = 0$ we conclude that (20) is constructive in the sense that the approximation model of $\mu()$ can have arbitrary accuracy by proper selection of p.

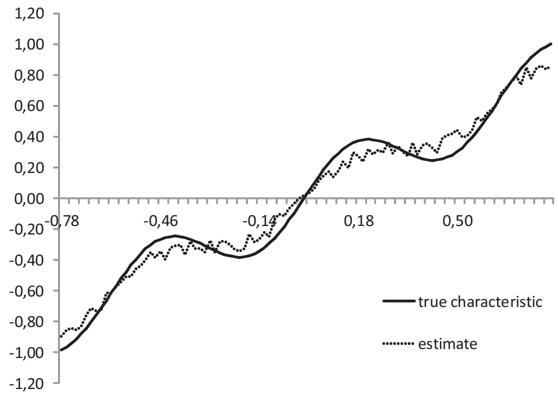

Figure 2: The true characteristic $\mu(x) = x + 0.2\sin(10x)$ and its nonparametric estimate $\widehat{\mu}_N^{(1)}(x)$.

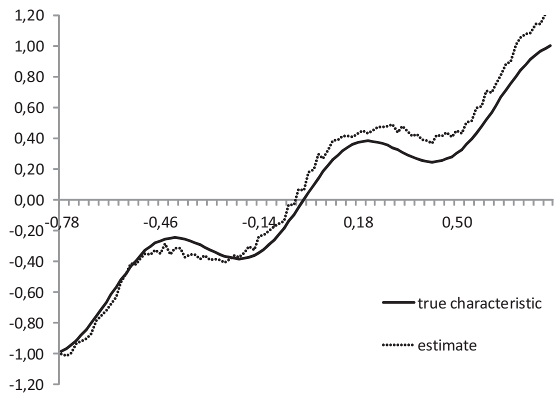

Figure 3: The true characteristic $\mu(x) = x + 0.2\sin(10x)$ and its nonparametric estimate $\widehat{\mu}_N^{(2)}(x)$.

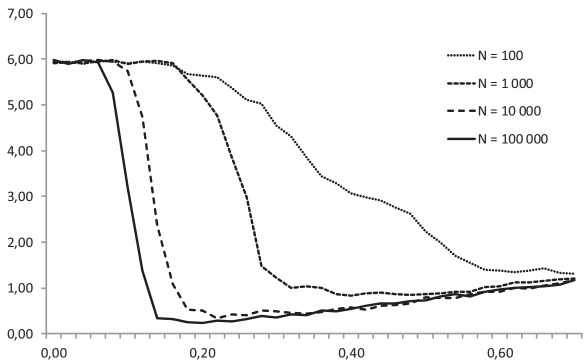

Figure 4: Relationship between the estimation error $ERR\left(\widehat{\mu}_N^{(1)}(x)\right)$ and the bandwidth parameter h.

Table 1: The errors of the estimates (3) and (4) versus N

N	10^2	10^3	10^4	10^5	10^6
$ERR\left(\widehat{\mu}_N^1(x)\right)$	6.1	4.9	0.8	0.5	0.3
$ERR\left(\widehat{\mu}_N^2(x)\right)$	9.8	8.1	4.4	1.1	0.8

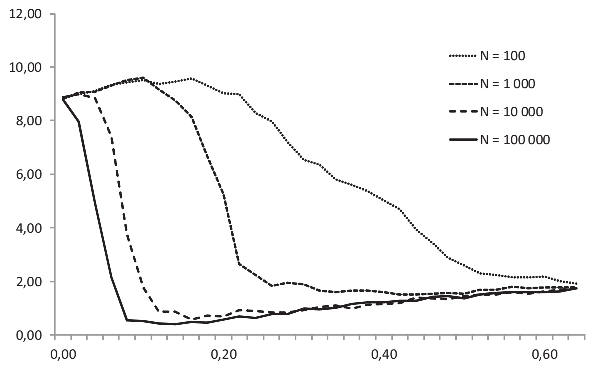

Figure 5: Relationship between the estimation error $ERR\left(\widehat{\mu}_N^{(2)}(x)\right)$ and the bandwidth parameter h.

4 NUMERICAL 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 static nonlinear characteristic $\mu(x) = x + 0.2\sin(10x)$. The nonparametric estimates (3) and (4) were computed on the same simulated data $\{(u_k, y_k)\}_{k=1}^{N}$. In **(A2)** we assumed $\lambda = 0.8$ and in (19) we took $p = 3$. The estimation error was computed according the rule

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

where $\{x^{(i)}\}_{i=1}^{N_0}$ is the grid of equidistant estimation points. The result of estimation for $N = 1000$ are shown in Fig. 2 and Fig. 3. The routine was repeated for various values of the tuning parameter h. As can be seen in Fig. 4 and Fig. 5, according to intuition, improper selection of h increases the variance or bias of the estimate. The Table 1 shows asymptotic bias of $\widehat{\mu}_N^2(x)$ and consistency of $\widehat{\mu}_N^1(x)$.

5 FINAL REMARKS

In the paper, the nonlinear characteristic of Wiener-Hammerstein system is successfully recovered from the input-output data under small amount of a priori information. The estimates work under IIR dynamic blocks, non-Gaussian input and for non-invertible characteristics. Since the Hammerstein systems and the Wiener systems are special cases of the sandwich system, considered in the paper, the proposed approach is universal in the sense that it can be applied without the prior knowledge of the system structure.

As regards the limit properties, the estimates $\widehat{\mu}_N^{(1)}(x)$ and $\widehat{\mu}_N^{(2)}(x)$ are not equivalent. First of them has slower rate of convergence (logarithmic), but it converges to the true system characteristic, since the model becomes more complex as the number of observations tends to infinity. The main limitation is assumed knowledge of λ, i.e., the upper bound of the impulse response. On the other hand the convergence of the estimate $\widehat{\mu}_N^{(2)}(x)$ is faster (polynomial), but the estimate is biased, even asymptotically. However, the bias can be made arbitrarily small by selecting the cut-off parameter p large enough.

As it was shown in (Hasiewicz and Mzyk (2009)), the nonparametric methods allow for decomposition of the identification task of block-oriented system and can support estimation of its parameters. Computing of both estimates $\widehat{\mu}_N^{(1)}(x)$, $\widehat{\mu}_N^{(2)}(x)$ and the distance $\delta_k(x)$ has the numerical complexity $O(N)$, and can be performed in recursive or semi-recursive version (see (Greblicki and Pawlak (2008))).

The principal question in Wiener-Hammerstein system identification problem is selection of adequate method. The scope of application of each estimate is limited by a specific set of associated assumptions. Most of them requires a priori known parametric type of model, Gaussian input, FIR dynamics or invertible characteristic. Since the general Wiener-Hammerstein system identification problem includes many difficult aspects, existence of one universal algorithm cannot be expected. In the light of this, the nonparametric approach seems to be good tool, which allows for combining selected methods (see e.g. (Mzyk (2010b))), depending on specificity of the particular task. Moreover, pure nonparametric estimates are the only possible choice, when the prior knowledge of the system is poor or uncertain.

Nonparametric approach offers simple algorithms, which are asymptotically free of approximation error, i.e. they converge to the true system characteristics. However, the purely nonparametric methods are not commonly exploited in practice for the following reasons: (i) they depend on various tuning parameters and functions; in particular, proper selection of kernel and the bandwidth parameter or orthonormal basis and the scale factor are critical for the obtained results, (ii) the prior knowledge of subsystems is completely neglected; the estimates are based on measurements only, and the resulting model may be not satisfactory when the number of measurements is small, and (iii) bulk number of estimates must be computed when the model complexity grows large.

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.

E. W. Bai, "Frequency domain identification of Wiener models", *Automatica*, vol. 39, No. 9, pp. 1521–1530, 2003.

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, "Nonparametric input density-free estimation of the nonlinearity in Wiener systems", *IEEE Transactions on Information Theory*, vol. 56, No. 7, pp. 3575-3580, 2010.

W. Greblicki, G. Mzyk, "Semiparametric approach to Hammerstein system identification", *Proceedings of the 15th IFAC Symposium on System Identification*, pp. 1680-1685, Saint-Malo, France, July 6-8, 2009.

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 non-linearities”, *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.

G. Mzyk, "Parametric versus nonparametric approach to Wiener systems identification", Chap. 8., Vol. 404 of *Lecture Notes in Control and Information Sciences*, Springer, 2010.

G. Mzyk, "Wiener-Hammerstein system identification with non-gaussian input", In: *IFAC International Workshop on Adaptation and Learning in Control and Signal Processing*, Antalya, Turkey, 2010.

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.