A Censored Sample Mean Approach to Nonparametric Identification of Nonlinearities in Wiener Systems

Grzegorz Mzyk

Abstract—A new, censored sample mean nonparametric identification algorithm for estimation of a nonlinear characteristic in Wiener system using properly preselected input–output data is proposed. Conditions imposed on the unknown characteristic are weak. In particular, its invertibility and global continuity are not required. The algorithm is based on computation of local sample-mean of proper output measurements. The mean square consistency of the estimate is proved for each continuity point of the unknown characteristic and the issue of the asymptotic convergence rate is discussed. Computer simulations are included to illustrate efficiency of the method also for small and moderate number of data.

Index Terms—Nonparametric identification, sample mean, Wiener system.

I. INTRODUCTION

IN THIS BRIEF, the identification problem of a nonlinear characteristic $\mu()$ of the Wiener system (Fig. 1) with the input finite-impulse response (*FIR*) dynamics $\{\lambda_j\}_{j=0}^{S}$ ($S < \infty$) or the infinite-impulse response (*IIR*)($S = \infty$), in the presence of a random correlated output noise $\{z_k\}$ is considered. The problem of recovering true nonlinearities has fundamental meaning in practice [9], particularly in such domains as: signal processing [12] and automatic control [3], biocybernetics, artificial neural networks [13], medicine [2], chemistry [4], modeling of distillation and fermentation processes. Since the linear dynamics precedes the nonlinear element, the identification of Wiener systems is in general more difficult than identification of Hammerstein systems composed of the same blocks but connected in reverse order [8], [10], [11], [17]. This is because the linear dynamics produces, from the viewpoint of the static block, a correlated input "system noise" $\{\delta_k\}$ (Fig. 2), which is next transferred to the output. This causes that, from the nonlinear characteristic point of view, we obtain an errors-in-variables identification problem (see [16]) and hence standard approaches lead to biased estimates. The state of the art in Wiener and Wiener–Hammerstein (sandwich) system identification is still not satisfying. Correlation methods proposed in the 1980's [1] assumed polynomial form of nonlinear characteristic $\mu()$ and white Gaussian noise $\{z_k\}$ on the output. Nonparametric methods based on the kernel regression or orthogonal series expansion required in turn $\mu()$ to be globally invertible, and as-
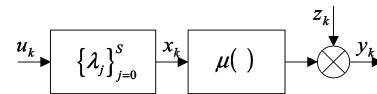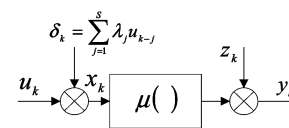
sumed Gaussian excitations of the Wiener systems. General advantages of nonparametric approach to nonlinear systems identification have been widely discussed in [5]–[8]. Nonparametric estimates are free of the systematic approximation error and converge to the true system nonlinearity. In particular, they are recommended when the nonlinear characteristic can not be expressed by a closed formula including finite number of unknown parameters and hence parametric approaches cannot be applied. In this paper we propose a new, nonparametric identification method, which imposes extremely weak *a priori* restrictions on the nonlinear subsystem, input excitation and the random output disturbance, in particular: 1) we do not assume invertibility of the nonlinear characteristic; 2) the method admits correlated output noise of almost arbitrary distribution; 3) the algorithm is computationally simple and only consists of proper data selection and averaging. Good statistical properties, i.e., convergence and satisfactory rate of convergence for a vast class of nonlinearities, assure wide applicability of the method. Practical utility of the method is further validated by numerical experiments.

Fig. 1. Wiener system.

Fig. 2. Equivalent form of Wiener system.

II. STATEMENT OF THE PROBLEM

We consider a Wiener system, i.e., a tandem 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 $\{\lambda_j\}_{j=0}^{S}$—the unknown impulse response of the linear input dynamics. By assumption, the interaction x_k is not available for measurements. Such a system is described by the following input–output equation:

$$y_k = \mu\left(\sum_{j=0}^{S} \lambda_j u_{k-j}\right) + z_k. \tag{1}$$
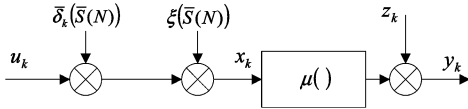
We assume the following.

Fig. 3.   Equivalent form of Wiener system with IIR filter.

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

A2) The unknown impulse response $\{\lambda_j\}_{j=0}^{S}$ of the linear *FIR* filter is normalized, i.e., $\lambda_0 = 1$, and the impulse response length S is *a priori* known.

A3) The nonlinearity $\mu(x)$ is an arbitrary function, which is 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 ergodic process, which is independent of the input $\{u_k\}$.

The aim is to estimate the unknown characteristic of the nonlinearity $\mu(x)$ only on the basis of $N + S$ input–output measurements $\{(u_k, y_k)\}_{k=1-S}^{N}$ of the whole Wiener system. Fig. 3 shows the equivalent form of the Wiener system with IIR filter.

The input probability density function $\vartheta_u(u)$ is assumed to be positive at the particular point $u = 0$ [see A1)] only for clarity of exposition. In fact, instead of $u = 0$ it may be any point $u_0 \in (-u_{\max}, u_{\max})$. Assumption A2) is also of technical meaning. It follows from the observation that the members of the family of Wiener systems composed by series connection of linear filters with the impulse responses $\{\overline{\lambda}_j\} = \{(\lambda_j/c)\}_{j=0}^{S}$ and the nonlinearities $\overline{\mu}(x) = \mu(cx)$ for $c \neq 0$ are indistinguishable from the input–output point of view. However, for the case of $c = \lambda_0 = 1$ the true nonlinearity can be potentially recovered. In consequence $\mu()$ can be estimated in general only up to some domain scaling factor c, independently of the applied identification method. Under Assumption A2) the system description (1) can be rewritten as follows (see Fig. 2):

$$y_k = \mu(u_k + \delta_k) + z_k, \text{ where } \delta_k = \sum_{j=1}^{S} \lambda_j u_{k-j} \quad (2)$$

which evidences that the nonlinearity recovering task is in fact the errors-in-variables identification problem. Let us emphasize in particular that A4) does not involve $\{z_k\}$ to be a white noise.

III. BACKGROUND OF THE APPROACH

We see that $x_k = u_k + \delta_k$. Let x be a chosen estimation point of $\mu(\cdot)$. For a given x let us define the "blurred" distance between the measurement u_k and x as follows

$$\Delta_k(x) \triangleq |u_k - x| + \varkappa(|u_{k-1}| + |u_{k-2}| + \cdots + |u_{k-S}|). \quad (3)$$

Observe that if in turn

$$\Delta_k(x) \leqslant h(N) \quad (4)$$

then the true (but unknown) interaction input x_k is located close to x provided that $h(N)$ (further, a calibration parameter) is small. Clearly it holds that

$$|x_k - x| \leqslant |x_k - u_k| + |u_k - x|$$
$$\leqslant |\delta_k| + |u_k - x| \leqslant L\Delta_k(x) + \Delta_k(x) \quad (5)$$

and hence

$$|x_k - x| \leqslant (L+1)h(N) \quad (6)$$

where $L = (1/\varkappa)\sum_{j=1}^{S} |\lambda_j|$, and $\varkappa$ can be any positive constant. For clarity in the presentation let further $\varkappa = 1$. The distance given in (3) may be easily computed as the point x and the data $u_k, \ldots, u_{k-S}$ are each time at ones disposal. In turn, the condition (4) selects the input sequences $\{u_k, u_{k-1}, \ldots, u_{k-s}\}$, for which true nonlinearity inputs $\{x_k\}$ surely belong to the neighborhood of the estimation point x with the radius $(L+1)h(N)$. Though the condition (4) look quite restrictive, this is not an impossible event, and its occurrence can be expected with probability (see Lemma in Appendix A)

$$P\{\Delta_k(x) \leqslant h(N)\}$$
$$\geqslant P\left\{\bigcap_{i=1}^{S}\left(|u_{k-i}| < \frac{h(N)}{S+1}\right) \cap \left(|u_k - x| < \frac{h(N)}{S+1}\right)\right\}$$
$$= \prod_{i=1}^{S} P\left(|u_{k-i}| < \frac{h(N)}{S+1}\right) \cdot P\left(|u_k - x| < \frac{h(N)}{S+1}\right)$$
$$= p^S q \quad (7)$$

where

$$p = \int_{-(h(N)/S+1)}^{(h(N)/S+1)} \vartheta_u(u)\,du$$
$$\text{and } q = \int_{x-(h(N)/S+1)}^{x+(h(N)/S+1)} \vartheta_u(u)\,du \quad (8)$$

due to Assumption A1). The average number of such sequences in a set of N data is thus of order $p^S q N$. The random event (4) is thus the most probable if the input probability density function $\vartheta_u(u)$ concentrates around the points $u = 0$ and $u = x$.

IV. IDENTIFICATION ALGORITHM
A. The Algorithm

We propose the following class of two step identification procedures, in which $\alpha \in (0, (1/S + 1))$ plays the role of the tuning parameter.

Step 1. Set $h(N) \sim N^{-\alpha}$, and for the estimation point x determine the set

$$S_h(x) \triangleq \{y_k : \Delta_k(x) \leqslant h(N)\}. \quad (9)$$

Step 2. Compute the average (a censored sample mean) of the elements $y_{[1]}, y_{[2]}, \ldots, y_{[N_0]} \in S_h(x)$ selected in Step 1 and take this mean as the estimate of the nonlinearity $\mu(\cdot)$ at the point x, i.e.,

$$\widehat{\mu_N}(x) = \text{Avg}S_h(x) \triangleq \begin{cases} \frac{1}{N_0}\sum_{i=1}^{N_0} y_{[i]}, & \text{as } N_0 > 0 \\ 0, & \text{elsewhere} \end{cases} \quad (10)$$

where N_0 is random cardinality of the set $S_h(x)$.

B. Limit Properties

Although the algorithm involves only learning sequence it is consistent and free of the risk of taking improper parametric *a priori* knowledge. The following theorem holds.

Theorem 1: If $h(N) \sim N^{-\alpha}$ and $\alpha \in (0, (1/S+1))$ then under Assumptions A1) $\div$ A4) it holds that

$$E\left[\widehat{\mu_N}(x) - \mu(x)\right]^2 \to 0, \text{ as } N \to \infty \qquad (11)$$

in each point x, such that $x \in \text{cont}(\mu(), \vartheta_u())$, where $\text{cont}(\mu(), \vartheta_u())$ is a set of all the points of continuity of $\mu()$ and $\vartheta_u()$.

(for the proof—see Appendix B)

The interesting open question is optimization of the convergence rate with respect to the selection of $h(N)$. Notice that small α ($\alpha \approx 0$) causes slow narrowing of the neighborhood of x given by (6). In consequence, the cardinality of the set $S_h(x)$ may be high, which is favorable for reducing the variance of the estimate (10) (influence of the output noise $\{z_k\}$ on the estimation error). However α close to $(1/S+1)$ may in turn improve the accuracy (reducing the bias) in case of irregular nonlinearities. Thus, the best choice of α needs to be a compromise dependent on the type of $\mu()$ and the variance of the noise $\{z_k\}$. For Lipschitz characteristics $\mu(\cdot)$ the asymptotically (as $N \to \infty$) optimal value of α is given by the following theorem.

Theorem 2: If the characteristic $\mu()$ fulfils the Lipschitz condition in the neighborhood of the estimation point x, i.e., there exist $\delta_{\max} > 0$ and $0 \leqslant d < \infty$ such that

$$|\mu(x+\delta) - \mu(x-\delta)| \leqslant d\delta, \text{ for each } 0 \leqslant \delta < \delta_{\max} \quad (12)$$

then for $N \to \infty$ the best value of α and the guaranteed corresponding rate of convergence are

$$\alpha_{\text{opt}} = \frac{1}{S+3}, \quad \text{MSE}(\widehat{\mu_N}(x)) = O(N^{-2/(S+3)}) \qquad (13)$$

where $\text{MSE}(\widehat{\mu_N}(x))$ is the mean-square error of the estimate (10).

(for the proof—see Appendix C)

In accordance with intuition, high order S of the *FIR* filter causes slower convergence of the estimate. For the purely static system ($S = 0$) we obtain the guaranteed rate $O(N^{-2/3})$, which is a bit worse than $O(N^{-1})$ typical for parametric approach. The slower convergence is here the cost payed for the poor *a priori* knowledge of the system.

Let us also emphasize that the algorithm is of linear computational complexity and only elementary operations are performed. Moreover $\Delta_k(x^{(i)})$'s may be computed recursively

$$\Delta_{k+1}(x) = \Delta_k(x)$$
$$+ (|u_k| - |u_{k-S}| + |u_{k+1} - x| - |u_k - x|) \quad (14)$$

and the complexity of (14) does not depend on the length S of the impulse response.

V. GENERALIZATION FOR THE *IIR* LINEAR FILTER

Assume that the system is described by the equation $y_k = \mu\left(\sum_{j=0}^{\infty} \lambda_j u_{k-j}\right) + z_k$ and the *IIR* linear filter is asymptotically stable. The impulse response of linear dynamics fulfils the following condition

$$|\lambda_j| \leq \zeta \eta^j \qquad (15)$$

where $\zeta > 0$ and $0 \leq \eta < 1$ are some constants.

We propose to approximate the true *IIR* filter by the *FIR* model of the order $S(N)$, dependent on the number of collected data N, with the property $S(N) \to \infty$ as $N \to \infty$. One can thus write (compare (2))

$$y_k = \mu(u_k + \delta_k(S(N)) + \xi(S(N))) + z_k \qquad (16)$$

where $\delta_k(S(N)) = \sum_{j=1}^{S(N)} \lambda_j u_{k-j}$ and $\xi(S(N)) = \sum_{j=S(N)+1}^{\infty} \lambda_j u_{k-j}$. The approximation error (tail) ξ, produced by the terms $\{\lambda_i\}_{i=S(N)+1}^{\infty}$ can be treated as the additional system noise.

In the case of (15) we obtain $|\xi(S(N))| = \left|\sum_{j=S(N)+1}^{\infty} \lambda_j u_{k-j}\right| \leq u_{\max} \sum_{j=S(N)+1}^{\infty} |\lambda_j| \leq u_{\max} \sum_{j=S(N)+1}^{\infty} \zeta \eta^j = (\zeta u_{\max}/1 - \eta) \cdot \eta^{S(N)+1}$ and hence

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

Introducing (see (3))

$$\overline{\Delta}_k(x) \triangleq \begin{cases} |u_k - x| + \zeta \sum_{i=1}^{S(N)} \eta^i |u_{k-i}|, & \text{known } \zeta, \eta \\ |u_k - x| + \sum_{i=1}^{S(N)} |u_{k-i}|, & \text{elsewhere} \end{cases}$$

we conclude that if (compare (4) and (5)) $\overline{\Delta}_k(x) \leq \overline{h}(N)$, then $|x_k - x| \leq |x_k - u_k| + |u_k - x| \leq |\delta_k(S(N))| + |\xi(S(N))| + |u_k - x| \leq \overline{\Delta}_k(x) L + \overline{\Delta}_k(x) + |\xi(S(N))| \leq (L+1)\overline{h}(N) + |\xi(S(N))|$ where $L = \sum_{j=1}^{S(N)} |\lambda_j|$. Now, we select the input sequences $\{u_k, u_{k-1}, \ldots, u_{k-S(N)}\}$, for which the true nonlinearity input x_k lies in the neighborhood of the estimation point x of radius $(L+1)\overline{h}(N) + |\xi(S(N))|$. The adequate identification algorithm, similar to that in Section IV is thus the following.

Step 1. Set $\overline{h}(N) \sim N^{-\alpha_1(N)}$ and construct the set $T_h(x)$ of appropriate output measurements, such that

$$T_h(x) \triangleq \left\{y_k : \overline{\Delta}_k(x) \leq \overline{h}(N)\right\}. \qquad (17)$$

Step 2. Compute the average of the elements of $T_h(x)$ (a censored sample mean)

$$\widehat{\mu_N}(x) = \text{Avg} T_h(x) \qquad (18)$$

as the estimate of the nonlinearity $\mu(\cdot)$ at the point x.

One can prove the following theorem.

Theorem 3: If $\overline{h}(N) \sim N^{-\alpha_1(N)}$, where $\alpha_1 \sim (1/(\log N)^p)$ ($p \in (0,1)$), and $S(N) \sim (\log N)^q$, where $q \in (0, p)$ then

$$E\left[\widehat{\mu_N}(x) - \mu(x)\right]^2 \to 0, \text{ as } N \to \infty.$$

Theorem 3 may be proved similarly to Theorem 1, substituting $h(N)$, S, $\Delta_k(x)$ by $\overline{h}(N)$, $S(N)$, $\overline{\Delta}_k(x)$ respectively and exploiting that $S(N) \to \infty$ and $(N - S(N)) \to \infty$ as $N \to \infty$.

VI. SIMULATION EXAMPLE

We have established asymptotic properties of the algorithm, in particular its convergence when the number of data N goes
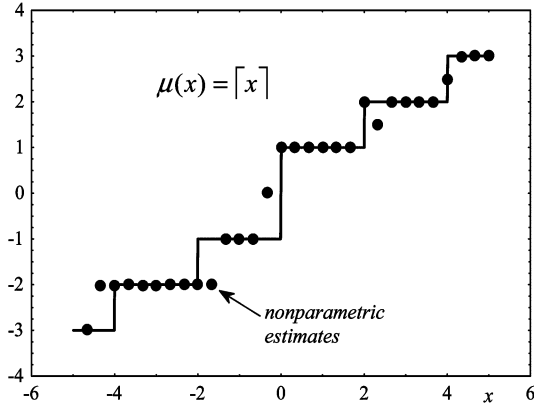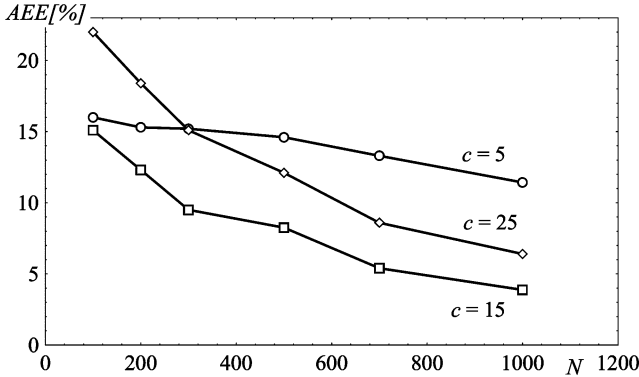
Fig. 4.   Resultant estimates.

Fig. 5.   Estimation error versus N for various c.

to infinity. The aim of the simulation example is to illustrate the performance for moderate ("practical") number of data. We simulated the Wiener systems with the *FIR* dynamics of order $S = 2$ and $\lambda_0 = \lambda_1 = \lambda_2 = 1$. The input $\{u_k\}$ and the noise $\{z_k\}$ were generated from the uniform distributions $u_k \sim U(-u_{max}, u_{max})$, where $u_{max} = 5$ and $z_k \sim U(-0.1, 0.1)$. The tuning parameter was selected according to the rule $h(N) = cN^{-\alpha}$, where c and α were treated as the degrees of freedom. We used nonlinear characteristic $\mu(x) = \lceil x \rceil$, which is not invertible and is not continuous, and hence cannot be treated by the identification algorithms worked out up to now. The results of its identification by using our algorithm for $N = 300, c = 15$ and $\alpha = 0.2$ [see (13)] are shown in Fig. 4. To establish the impact of the choice of c in the tuning factor $h(N) = cN^{-\alpha}$ on the accuracy of the estimates we compared the values of the following aggregated estimation error:

$$\text{AEE}(N, c) = \frac{1}{P} \sum_{i=1}^{P} \frac{\left| \widehat{\mu_N}(x^{(i)}) - \mu(x^{(i)}) \right|}{\left| \mu(x^{(i)}) \right|} \cdot 100\%$$

obtained for various c, where $x^{(1)}, \ldots, x^{(P)}$ were deterministic equispaced estimation points in the domain of $\mu()$. The results presented in Fig. 5 show large sensitivity of efficiency of the algorithm to the tuning parameter and motivate further research of the issue.

VII.   CONCLUSION

The pointwise consistency of the proposed estimation routine of the nonlinearity in Wiener system has been proved, and the rate of convergence has been evaluated under random input, and random correlated noise. Sufficient identifiability conditions with reference to the nonlinear characteristic of the Wiener structure are not so restrictive as in previous papers. In particular, local continuity is needed instead of invertibility of the characteristic, and moreover, special excitation (i.e., sine wave or normally distributed random input) is not required. The 'price paid' is the necessity of data preselection in *Stage 1* of the algorithm. However it is worth it, owing to the wide range of applicability and extreme simplicity of the algorithm. Observe in addition that pointwise nonparametric estimates $\widehat{\mu_N}(x)$ can be for instance established (in *Stage 1*) only for a small set of preselected points and next the resulting data $(x, \widehat{\mu_N}(x))$ can be used (in *Stage 2*) as 'leading data' for determining some 'rational' parametric model of the nonlinearity in a Wiener system.

APPENDIX

Technical Lemma:

Lemma: For a sequence of independent and positive random variables $A_1, A_2, \ldots A_n$ it holds that

$$P \left(\sum_{i=1}^{n} A_i < \varepsilon \right) \geqslant \prod_{i=1}^{n} P \left(A_i < \frac{\varepsilon}{n} \right).$$

Sketch of the Proof of Theorem 1: Let $y_{[1]}, y_{[2]}, \ldots, y_{[N_0]}$ and $x_{[1]}, x_{[2]}, \ldots, x_{[N_0]}$ be the output measurements selected in *Step 1* and respective unknown inputs. For each $i = 1, 2, \ldots, N_0$ we have

$$\lim_{N \to \infty} E y_{[i]} = \lim_{N \to \infty} E \left[\mu(x_{[i]}) + z_{[i]} \right]$$
$$= \lim_{N \to \infty} E \mu(x + x_{[i]} - x) + \lim_{N \to \infty} E z_{[i]}.$$

Since $E z_{[i]} = 0$ [see Assumption A4)], and $\left| x_{[i]} - x \right| \leqslant h(N)$ (see (4) and (5)) under continuity of $\mu()$ at the point x, and the fact that $h(N) \to 0$ as $N \to \infty$, we obtain

$$\lim_{N \to \infty} E y_{[i]} = \mu(x). \qquad (19)$$

Similarly, for independent processes $\{\mu(x_k)\}$ and $\{z_k\}$ [see Assumption A4)] one can write

$$\text{var} y_{[i]} = \text{var} \left[\mu(x_{[i]}) + z_{[i]} \right] = \text{var} \mu(x_{[i]}) + \sigma_z^2,$$

where for $\mu()$ continuous at the point x it holds that $\text{var} \mu(x_{[i]}) \to 0$ as $N \to \infty$. It yields to [see Assumption A4)]

$$\lim_{N \to \infty} \text{var} y_{[i]} = \sigma_z^2 < \infty. \qquad (20)$$

Assuming that in the neighbourhoods of $u = 0$ and $u = x$ it holds that $\vartheta_u(u) \geqslant c_0 > 0$, some constant c_0, we have [see (8) and Assumption **A1**)]

$$p, q > c_1 h(N), \quad \text{where } c_1 = \frac{2c_0}{S + 1}.$$

Form (7)

$$P(\Delta_k(x) < h(N)) > c_2 h(N)^{S+1}, \text{ where } c_2 = c_1^{S+1}. \quad (21)$$

As the number of successes, the cardinality N_0 of the set $S_h(u)$ is random and asymptotically (as $N \to \infty$) has the Poisson distribution with the expectation $EN_0 = N \cdot P(\Delta_k(x) < h(N)) > c_2 N h(N)^{S+1}$. Each Poisson-distributed random variable has the following property:

$$\lim_{N\to\infty} P(N_0 > cEN_0) = \begin{cases} 0, & \text{as } c > 1 \\ 1, & \text{as } c < 1. \end{cases} \quad (22)$$

The second case in (22) shows that if $\lim_{N\to\infty} EN_0 = \infty$ then $N_0 \to \infty$ in probability as $N \to \infty$. The following two conditions must be thus simultaneously fulfilled:

$$h(N) \to 0 \text{ and } EN_0 \to \infty, \text{ as } N \to \infty.$$

First of them leads to the requirement $\alpha > 0$ and the second one

$$Nh(N)^{S+1} = NN^{-\alpha(S+1)} = N^{1-\alpha(S+1)} \to \infty$$

means that $1 - \alpha(S+1) > 0$, i.e., $\alpha < (1/S+1)$ what ends the proof, as the convergence $\widehat{\mu_N}(x) \to \mu(x)$ is simple consequence of the law of large numbers (see (19) and (20)). ∎

Sketch of the Proof of Theorem 2: Under condition (12) and definition (9) we obtain

$$\text{bias}\widehat{\mu_N}(x) = E\left[\frac{1}{N_0}\sum_{i=1}^{N_0}\left(y_{[i]} - \mu(x)\right)\right]$$
$$= E\left[\frac{1}{N_0}\sum_{i=1}^{N_0}\left(\mu(x_{[i]}) + z_{[i]} - \mu(x)\right)\right]$$
$$\leqslant dh(N)$$

and hence

$$\text{bias}^2\widehat{\mu_N}(x) = O(h(N)^2). \quad (23)$$

The variance component of the mean-square error may be evaluated by using conditional variances as follows

$$\text{var}\widehat{\mu_N}(x) = \sum_{n=1}^{N} P(N_0 = n) \cdot \text{var}\frac{1}{n}\sum_{i=1}^{n} y_{[i]}$$
$$= \sum_{n=1}^{N} P(N_0 = n) \cdot \frac{\text{var} y_{[i]}}{n}$$

and for N large, making use of (22), we obtain

$$\text{var}\widehat{\mu_N}(x) = \sum_{n=1}^{N} P(N_0 = n) \cdot \frac{\sigma_z^2}{n} \leqslant \frac{\sigma_z^2}{CNh(N)^{S+1}}. \quad (24)$$

After substituting $h(N) = N^{-\alpha}$ to (23) and (24) we conclude that

$$\text{bias}^2\widehat{\mu(u)} = O(N^{-2\alpha}), \text{ and } \text{var}\widehat{\mu(u)} = O(N^{\alpha(S+1)-1}).$$

Increasing α reduces the bias, but simultaneously the variance is rising. Hence, for $N \to \infty$ the optimal choice of α is given by the equation $-2\alpha = \alpha(S+1) - 1$ what leads to the conclusion $\alpha_{\text{opt}} = (1/S + 3)$. ∎

REFERENCES

[1] S. A. Billings and S. Y. Fakhouri, "Identification of systems containing linear dynamic and static nonlinear elements," *Automatica*, vol. 18, no. 1, pp. 15–26, 1982.

[2] P. Celka and P. Colditz, "Nonlinear nonstationary Wiener model of infant eeg seizures," *IEEE Trans. Biomed. Eng.*, vol. 49, no. 3, pp. 556–564, Mar. 2002.

[3] G. B. Giannakis and E. Serpedin, "A bibliography on nonlinear system identification," *Signal Process.*, vol. 81, pp. 533–580, 2001.

[4] J. C. Gómez, A. Jutan, and E. Baeyens, "Wiener model identification and predictive control of pH neutralization process," *Cont. Theor. Appl. IEEE Proc.*, vol. 151, no. 3, pp. 329–338, 2004.

[5] W. Greblicki, "Nonparametric identification of Wiener systems by orthogonal series," *IEEE Trans. Autom. Contr.*, vol. 10, pp. 2077–2086, 1994.

[6] W. Greblicki, "Nonparametric approach to Wiener system identification," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 44, no. 6, pp. 538–545, Jun. 1997.

[7] W. Greblicki, "Continuous-time Wiener system identification," *IEEE Trans. Autom. Contr.*, vol. 43, no. 10, pp. 1488–1493, Oct. 1998.

[8] W. Greblicki and M. Pawlak, "Cascade non-linear system identification by a non-parametric method," *Int. J. Syst. Sci.*, vol. 25, no. 1, pp. 129–153, 1994.

[9] R. Haber and L. Keviczky, *Nonlinear System Identification Input–Output Modeling Approach*. Norwell, MA: Kluwer Academic, 1999.

[10] Z. Hasiewicz and G. Mzyk, "Combined parametric-nonparametric identification of Hammerstein systems," *IEEE Trans. Autom. Contr.*, vol. 49, no. 10, pp. 1370–1376, Oct. 2004.

[11] Z. Hasiewicz and G. Mzyk, "Nonparametric instrumental variables for Hammerstein system identification," *Int. J. Contr.*, submitted for publication.

[12] Z. Hasiewicz, M. Pawlak, and P.Śliwiński, "Nonparametric identification of nonlinearities in block-oriented systems by orthogonal wavelets with compact support," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 52, no. 2, pp. 427–442, Feb. 2005.

[13] J. d. J. Rubio and W. Yu, "Stability analysis of nonlinear system identification via delayed neural networks," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 54, no. 2, pp. 161–165, Feb. 2007.

[14] A. Krzyżak, J. Z. Sasiadek, and B. Kégl, "Nonparametric identification of dynamic nonlinear systems by a hermite series approach," *Int. J. Syst. Sci.*, vol. 32, no. 10, pp. 1261–1285, 2001.

[15] M. Pawlak and Z. Hasiewicz, "Nonlinear system identification by the Haar multiresolution analysis," *IEEE Trans. Circuits Syst.*, vol. 45, no. 9, pp. 945–961, Sep. 1998.

[16] E. Rafajlowicz, "Non-parametric identification with errors in independent variables," *Int. J. Syst. Sci.*, vol. 25, no. 9, pp. 1513–1520, 1994.

[17] J. Vörös, "Identification of Hammerstein systems with time-varying piecewise-linear characteristics," *IEEE Trans. Circuits Syst. II, Reg. Papers*, vol. 52, no. 12, pp. 865–869, Dec. 2005.