Int. J. Appl. Math. Comput. Sci., 2001, Vol.11, No.4, 977–991

RECURSIVE IDENTIFICATION OF WIENER SYSTEMS[†]

WŁODZIMIERZ GREBLICKI[*]

A Wiener system, i.e. a cascade system consisting of a linear dynamic subsystem and a nonlinear memoryless subsystem is identified. The *a priori* information is nonparametric, i.e. neither the functional form of the nonlinear characteristic nor the order of the dynamic part are known. Both the input signal and the disturbance are Gaussian white random processes. Recursive algorithms to estimate the nonlinear characteristic are proposed and their convergence is shown. Results of numerical simulation are also given. A known algorithm recovering the impulse response of the dynamic part is presented in a recursive form.

Keywords: Wiener system, system identification, recursive identification, nonparametric identification

1. Introduction

In the present paper, we study the identification of a Wiener system, i.e. a system consisting of a linear dynamic subsystem followed by a nonlinear memoryless subsystem. The problem consists in estimating the nonlinear characteristic of the memoryless subsystem and recovering the impulse response of the linear one from input-output observations of the whole system. The main difficulty is caused by the fact that the inner system signal is not measured. First papers devoted to such a problem can be traced back to the late 1970s, see (Billings and Fakhouri, 1977; 1978; Brillinger, 1977), see also (Billings, 1980; Hunter and Korenberg, 1986; Westwick and Kearney, 1992; Westwick and Verhaegen, 1996). The fact that the number of cited papers is so low is undoubtedly caused by great theoretical difficulties arising while examining such systems.

All the authors mentioned above assumed that the *a priori* information about the system is parametric, i.e. comparatively large. As far as a memoryless subsystem is concerned, its nonlinear characteristic has parametric form, usually a polynomial one. In contrast, in this paper, the characteristic can be of any form, which, in terms of the *a priori* information, means that the problem is nonparametric. It seems that the nonparametric method responds better to demands coming from applications since in real life the knowledge we possess about the system before an experiment is usually

[†] The work was supported by the State Committee for Scientific Research under Grant No. 7 T11A 015 20.

[*] Institute of Engineering Cybernetics, Wrocław University of Technology, Wybrzeże S. Wyspiańskiego 27, 50–370 Wrocław, Poland, e-mail: wgre@ict.pwr.wroc.pl

rather limited and uncertain. As for the input signal, like in all the papers cited above, it is a Gaussian white random process. Such a problem was already studied by Greblicki (1992; 1997), as well as by Krzyżak and Partyka (1993), who examined off-line algorithms.

In the present paper, two recursive algorithms to recover the nonlinear characteristic are proposed. Their convergence is established and convergence rates are given. Results of numerical simulations are also presented. The novelty of this paper is that our algorithms are recursive, which means that they can be applied on-line. From the numerical viewpoint, the algorithms identifying nonlinear and dynamic subsystems are mutually independent. This means that each subsystem can be identified separately, which is an obvious advantage.

2. System in Context and Identification Problems

The problem is to identify a Wiener system with an input U_n and an output Y_n, see Fig. 1, i.e. a system consisting of two subsystems connected in a cascade. The first is linear, dynamic and asymptotically stable, and its state equations are of the following form:

$$\begin{cases} X_{n+1} = AX_n + bU_n \\ W_n = c^T X_n, \end{cases} \tag{1}$$

$n = \ldots, -1, 0, 1, \ldots$, where X_n is the state vector and W_n is the output vector. The dimension of the state vector, matrix A and vectors b, c are unknown. The output is disturbed by the noise Z_n, which means that the nonlinear subsystem is excited by V_n, where $V_n = W_n + Z_n$. The other subsystem is memoryless, nonlinear and has a characteristic m being a Lebesgue measurable function. Thus

$$Y_n = m(V_n). \tag{2}$$

The system is driven by a Gaussian stationary white random process $\{U_n;\ n = \ldots, -1, 0, 1, \ldots\}$ having zero mean and unknown variance σ_U^2. Owing to this and the fact that the dynamic subsystem is stable, both $\{X_n;\ n = \ldots, -1, 0, 1, \ldots\}$ and $\{W_n;\ n = \ldots, -1, 0, 1, \ldots\}$ are Gaussian, stationary with zero mean, but correlated. Moreover, it is assumed that $\{Z_n;\ n = \ldots, -1, 0, 1, \ldots\}$ is also Gaussian, independent of the input, white with zero mean and unknown variance σ_Z^2.

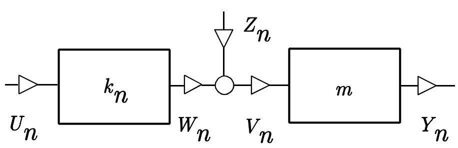

Fig. 1. The identified Wiener system.

Both the subsystems are identified from input-output observations of the whole system. This means that we estimate the characteristic m and the impulse response $\{k_i;\ i = 1, 2, \ldots\}$, where $k_i = c^T A^{i-1} b$, from $(U_0, Y_0), (U_1, Y_1), \ldots$. As we have already stated, the originality of the paper boils down to achieving these goals with recursive algorithms. We propose two algorithms recovering the nonlinearity. We also show that a known estimate of the impulse response can be calculated in a recursive way.

3. Preliminaries

The results presented in the lemmas below constitute a motivation for our identification algorithms. Despite the fact that they were already presented in (Greblicki, 1992), for the sake of completeness, we will repeat them here. In the lemmas below, as well as throughout the paper, $\rho = \sigma_U^2/\sigma_V^2$ with $\sigma_V^2 = \sigma_Z^2 + \sigma_U^2 \sum_{n=1}^{\infty} k_n^2$, $\alpha = \rho E\{V_0 m(V_0)\}$, and $\beta = \rho k_1$.

Owing to the next lemma, we can construct our algorithms to identify the nonlinear subsystem. Denote by $m(\mathbb{R})$ the image of the real line $\mathbb{R}$ under the mapping m while m^{-1} is the inverse of m in the Cartesian product $\mathbb{R} \times m(\mathbb{R})$.

Lemma 1. *If the inverse m^{-1} exists, then $E\{U_0 \mid Y_n = y\} = \rho k_n m^{-1}(y)$. In particular,*

$$E\big\{U_0 \mid Y_1 = y\big\} = \beta m^{-1}(y).$$

Proof. Since the pair (U_0, V_n) has a Gaussian distribution with zero marginal means and covariance $\sigma_U^2 k_n$, we find the conditional density of U_0 conditioned on $V_n = v$ normal with mean $\rho k_n v$ and variance $(1 - \rho k_n^2)\sigma_U^2$. Hence $E\{U_0 \mid V_n\} = \rho k_n V_n$. Thus, using (2) and the fact that m^{-1} exists, we get $E\{U_0 \mid Y_n = y\} = E\{U_0 | V_n = m^{-1}(y)\} = \rho k_n m^{-1}(y)$ and the proof is complete. ■

Lemma 2. *Let $E|V_0 m(V_0)| < \infty$. Then $E\{U_0 Y_n\} = \alpha k_n$ for $n = 1, 2, \ldots$.*

Proof. Since $E\{U_0 | V_n\} = \rho k_n V_n$ (see the proof of Lemma 1), we obtain $E\{U_0 Y_n \mid V_n\} = m(V_n) E\{U_0 \mid V_n\} = \rho k_n m(V_n) V_n$ and the lemma follows. ■

4. Nonlinear Subsystem Identification

4.1. Algorithms

Lemma 1 suggests that recovering $\beta m^{-1}(y)$ is equivalent to estimating the regression function $E\{U_0 \mid Y_1 = y\}$. To estimate the regression, we apply the following two algorithms:

$$\tilde{\mu}_n(y) = \frac{\tilde{\xi}_n(u)}{\tilde{\eta}_n(u)}, \tag{3}$$

where $\tilde{\xi}_n(u)$ and $\tilde{\eta}_n(u)$ are given by the following recursive procedures:

$$\tilde{\xi}_n(u) = \tilde{\xi}_{n-1}(u) + U_n K\left(\frac{y - Y_{n+1}}{h_n}\right),$$

$$\tilde{\eta}_n(u) = \tilde{\eta}_{n-1}(u) + U_n K\left(\frac{y - Y_{n+1}}{h_n}\right),$$

with $\tilde{\xi}_0(u) = \tilde{\eta}_0(u) = 0$, and

$$\hat{\mu}_n(y) = \frac{\hat{\xi}_n(u)}{\hat{\eta}_n(u)}, \tag{4}$$

where

$$\hat{\xi}_n(u) = \hat{\xi}_{n-1}(u) + U_n \frac{1}{h_n} K\left(\frac{y - Y_{n+1}}{h_n}\right),$$

$$\hat{\eta}_n(u) = \hat{\eta}_{n-1}(u) + U_n \frac{1}{h_n} K\left(\frac{y - Y_{n+1}}{h_n}\right),$$

with $\hat{\xi}_0(u) = \hat{\eta}_0(u) = 0$. Here K is a kernel function and $\{h_n\}$ stands for a positive number sequence, both being suitably selected. In our algorithms and throughout the paper, $0/0$ is treated as zero.

The kernel K is a Lebesgue measurable function satisfying the following restrictions:

$$\sup_{-\infty < y < \infty} |K(y)| < \infty, \tag{5}$$

$$\int |K(y)|\,\mathrm{d}y < \infty, \tag{6}$$

$$yK(y) \to 0 \text{ as } |y| \to \infty, \tag{7}$$

$$|K(x) - K(y)| \le c_K\,|x - y| \tag{8}$$

with some c_K, for all $x, y \in \mathbb{R}$. As the kernel one can select a window kernel equal to 1 for $|y| \le 1$ and zero otherwise, or a parabolic one equal to $1 - y^2$ and zero for $|y| \le 1$ and $1 < |y|$, respectively. Other kernels include, for example, $1/(1+y^2)$, $1/(1+|y|)^2$, $\exp(-|y|)$, or $\exp(-y^2)$.

The positive number sequence $\{h_n\}$ is selected to satisfy, depending on the identification algorithm, some of the following conditions:

$$h_n \to 0 \quad \text{as } n \to \infty, \tag{9}$$

$$\frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{h_i^2} \to 0 \quad \text{as } n \to \infty, \tag{10}$$

$$h_n \sum_{i=1}^{n} h_i \to \infty \qquad \text{as } n \to \infty. \tag{11}$$

There certainly exist number sequences satisfying the above restrictions. For $h_n = \delta n^{-\gamma}$, $\delta > 0$, (9) is satisfied for $0 < \gamma$. In turn, both (10) and (11) are satisfied for $0 < \gamma < 1/2$.

Rewriting our algorithms in the following compact forms:

$$\tilde{\mu}_n(y) = \sum_{i=1}^{n} U_i K\left(\frac{y - Y_{i+1}}{h_i}\right) \Bigg/ \sum_{i=1}^{n} K\left(\frac{y - Y_{i+1}}{h_i}\right) \tag{12}$$

and

$$\hat{\mu}_n(y) = \sum_{i=1}^{n} U_i \frac{1}{h_i} K\left(\frac{y - Y_{i+1}}{h_i}\right) \Bigg/ \sum_{i=1}^{n} \frac{1}{h_i} K\left(\frac{y - Y_{i+1}}{h_i}\right), \tag{13}$$

we can observe that they are recursive modifications of the kernel regression estimate

$$\bar{\mu}_n(y) = \sum_{i=1}^{n} U_i K\left(\frac{y - Y_{i+1}}{h_n}\right) \Bigg/ \sum_{i=1}^{n} K\left(\frac{y - Y_{i+1}}{h_n}\right), \tag{14}$$

proposed in (Nadaraya, 1964; Watson, 1964). We deal with its recursive versions (3) and (4) introduced in (Ahmad and Lin, 1976; Collomb, 1977) and then examined in (Devroye and Wagner, 1980; Krzyżak and Pawlak, 1984) as well as (Greblicki and Pawlak, 1987). In all those papers, however, pairs (U_i, Y_{i+1}) are independent. Thus, all the above-mentioned papers can be applied when $A = 0$, i.e. when the dynamic part is a simple delay. Therefore, in the context of dynamic system identification, all those results are useless. Nevertheless, in (Greblicki, 1992; 1997) it was shown that the off-line algorithm (14) successfully recovers the nonlinearity in Wiener systems. In the present paper, we demonstrate that recursive estimates (3) and (4) can also be employed to recover the nonlinearity in Wiener systems. We do it owing to our crucial Lemma 6 in Appendix A.

4.2. Motivation

An intuitive motivation for (3), i.e. for (12), can be explained in the following way: Writing

$$\tilde{g}_n(y) = \frac{1}{\sum_{i=1}^{n} h_i} \sum_{i=1}^{n} U_i K\left(\frac{y - Y_{i+1}}{h_i}\right), \tag{15}$$

$$\tilde{f}_n(y) = \frac{1}{\sum_{i=1}^{n} h_i} \sum_{i=1}^{n} K\left(\frac{y - Y_{i+1}}{h_i}\right), \tag{16}$$

we get $\tilde{\mu}_n(y) = \tilde{g}_n(y)/\tilde{f}_n(y)$. From Lemma 1, it follows that

$$E\tilde{g}_n(y) = \frac{\beta}{\sum_{i=1}^{n} h_i} \sum_{i=1}^{n} h_i \mu(y; h_i), \quad E\tilde{f}_n(y) = \frac{1}{\sum_{i=1}^{n} h_i} \sum_{i=1}^{n} h_i f(y; h_i),$$

with

$$\begin{aligned} \mu(y;h) &= \frac{1}{h} E\left\{ U_0 K\left(\frac{y - Y_1}{h}\right)\right\} = \frac{1}{h} E\left\{ E\left\{U_0 | Y_1\right\} K\left(\frac{y - Y_1}{h}\right)\right\} \\ &= \frac{1}{h} E\left\{ m^{-1}(Y_1) K\left(\frac{y - Y_1}{h}\right)\right\} \\ &= \int \frac{1}{h} K\left(\frac{y-\zeta}{h}\right) m^{-1}(\zeta) f(\zeta)\, \mathrm{d}\zeta, \end{aligned} \tag{17}$$

and

$$f(y;h) = \frac{1}{h} E\left\{ K\left(\frac{y - Y_0}{h}\right)\right\} = \int \frac{1}{h} K\left(\frac{y-\zeta}{h}\right) f(\zeta)\, \mathrm{d}\zeta,$$

where f is the density of Y_n, assumed here to exist. Since, owing to (5)–(7),

$$\frac{1}{h} K\left(\frac{y-\zeta}{h}\right)$$

gets close to the Dirac impulse located at the point $\zeta = y$ as $h \to 0$, one can expect that $\mu(y;h) \to m^{-1}(y) f(y)$ as $h \to 0$ and, consequently, that $E\tilde{g}_n(y) \to \beta\mu^{-1}(y) f(y)$ as $h_n \to 0$. For the same reasons, one can hope that $f(y;h) \to f(y)$ as $h \to 0$. Finally, one can expect that $E\tilde{\xi}_n(y)/E\tilde{\eta}_n(y)$ converges to $\beta m^{-1}(y)$ as $h_n \to 0$, i.e. that $\tilde{\mu}_n(y) = \tilde{\xi}_n(y)/\tilde{\eta}_n(y)$ recovers $\beta m^{-1}(y)$. In the paper, we go further and present proofs of the convergence for both the algorithms.

4.3. Convergence

We assume that

$$m \text{ is continuous, strictly monotone,} \tag{18}$$

$$m' \text{ is bounded, continuous, nonzero.} \tag{19}$$

The lemma below shows that the density f of Y_n does exist.

Lemma 3. *Let m satisfy (18) and (19). In $m(\mathbb{R})$, m^{-1} is continuous while f is both positive and continuous.*

Proof. The fact that m^{-1} is continuous in $m(\mathbb{R})$ is obvious. Supposing that m is an increasing function, we find that m^{-1} is increasing, too. Thus $f(y)$ equals $f_V(m^{-1}(y))[m^{-1}(y)]'$, and zero for $y \in m(\mathbb{R})$ and $y \notin m(\mathbb{R})$, respectively, where f_V is a density of V_n. Since f_V is normal, $f_V(m^{-1}(\cdot))$ is continuous and nonzero in $m(\mathbb{R})$. Observe now that $m'(m^{-1}(\cdot))$ is continuous, nonzero in $m(\mathbb{R})$, and that $[m^{-1}(y)]' = 1/m'(m^{-1}(y))$. Thus f is continuous and nonzero in $m(\mathbb{R})$. Since, for a decreasing m, the proof is similar, the lemma follows. ∎

Theorem 1. *Let m satisfy (18) and (19). Assume that K fulfils (5)–(8). Let a monotone positive number sequence $\{h_n\}$ satisfy (9) and (11). Then, at every point $y \in m(\mathbb{R})$,*

$$\tilde{\mu}_n(y) \to \beta m^{-1}(y) \text{ as } n \to \infty.$$

Proof. First of all, we apply Lemma 3 to observe that both m^{-1} and f are continuous in $m(\mathbb{R})$ whereas f is positive. We begin our reasoning by observing that $\tilde{\mu}_n(y) = \tilde{g}_n(y)/\tilde{f}_n(y)$, where $\tilde{g}_n(y)$ and $\tilde{f}_n(y)$ are as in (15) and (16), respectively. Let $y \in m(\mathbb{R})$. From Lemma 1, it follows that

$$E\tilde{g}_n(y) = \beta \frac{1}{\sum_{i=1}^{n} h_i} \sum_{i=1}^{n} h_i \mu(y; h_i),$$

with $\mu(y;h)$ as in (17). Using Lemma 4 in Appendix A and (9), we deduce that $\mu(y;h_n) \to \beta f(y) m^{-1}(y) \int K(\xi)\,\mathrm{d}\xi$ as $n \to \infty$. Therefore, observing that (9) and (11) imply $\sum_{n=1}^{\infty} h_n = \infty$, we get $E\tilde{g}_n(y) \to \beta f(y) m^{-1}(y) \int K(v)dv$ as $n \to \infty$.

Furthermore, $\mathrm{var}[\tilde{g}_n(y)] = R_n(y) + S_n(y)$ with

$$R_n(y) = \frac{1}{\left(\sum_{i=1}^{n} h_i\right)^2} \sum_{i=1}^{n} \mathrm{var}\left[U_0 K\left(\frac{y - Y_1}{h_i}\right)\right],$$

$$S_n(y) = \frac{1}{\left(\sum_{i=1}^{n} h_i\right)^2} \left\{ \sum_{i=1}^{n} \sum_{j=1}^{i-1} \mathrm{cov}\left[U_i K\left(\frac{y - Y_{i+1}}{h_i}\right), U_j K\left(\frac{y - Y_{j+1}}{h_j}\right)\right] \right.$$
$$\left. + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \mathrm{cov}\left[U_i K\left(\frac{y - Y_{i+1}}{h_i}\right), U_j K\left(\frac{y - Y_{j+1}}{h_j}\right)\right] \right\}. \quad (20)$$

Observe that

$$R_n(y) \le \frac{1}{\left(\sum_{i=1}^{n} h_i\right)^2} \sum_{i=1}^{n} h_i \psi(y; h_i) \le \frac{\psi(y)}{\sum_{i=1}^{n} h_i},$$

with

$$\psi(y;h) = \frac{1}{h} E\left\{U_0^2 K^2\left(\frac{y - Y_1}{h}\right)\right\}, \quad (21)$$

and $\psi(y) = \sup_{0<h} \psi(y;h)$ being finite, by virtue of Lemma 5. In turn, owing to Lemma 6 and the fact that $\{h_n\}$ is monotone, the first term in (20) is bounded in the absolute value by

$$\frac{\eta(y)}{\left(\sum_{i=1}^{n} h_i\right)^2} \sum_{i=1}^{n} \sum_{j=1}^{i-1} \frac{h_j}{h_i} \left\|A^{i-j}\right\| \le \frac{\eta(y)}{h_n \sum_{i=1}^{n} h_i} \sum_{j=1}^{i-1} \left\|A^{i-j}\right\| \le \frac{\eta(y)}{h_n \sum_{i=1}^{n} h_i} \sum_{j=1}^{\infty} \left\|A^{j}\right\|,$$

while the other by

$$\frac{\eta(y)}{\left(\sum\limits_{i=1}^{n} h_i\right)^2} \sum_{i=1}^{n} \sum_{j=i+1}^{n} \frac{h_i}{h_j} \|A^{j-i}\| \le \frac{\eta(y)}{h_n \sum\limits_{i=1}^{n} h_i} \sum_{j=1}^{\infty} \|A^j\|.$$

Observe that, due to the stability of the linear subsystem, $\sum_{j=1}^{\infty} \|A^j\| < \infty$. Hence

$$|S_n(y)| \le \frac{2\eta(y)}{h_n \sum\limits_{i=1}^{n} h_i} \sum_{j=1}^{\infty} \|A^j\|.$$

Thus there exists a finite $\phi(u)$ such that

$$\operatorname{var}[\tilde{g}_n(y)] \le \frac{1}{h_n \sum\limits_{i=1}^{n} h_i} \phi(u). \tag{22}$$

In this way, we have shown that $\tilde{g}_n(y) \to \beta f(y) m^{-1}(y) \int K(\xi)\,\mathrm{d}\xi$ in probability as $n \to \infty$. Using similar arguments, we easily verify that $\tilde{f}_n(y) \to f(y) \int K(\xi)\,\mathrm{d}\xi$ in probability as $n \to \infty$, and the proof is complete. ■

Using similar arguments, one can verify our next result.

Theorem 2. *Let m satisfy (18) and (19). Assume that K fulfils (5)–(8). Let a nonnegative number sequence $\{h_n\}$ satisfy (9) and (10). Then, at every point $y \in m(\mathbb{R})$,*

$$\hat{\mu}_n(y) \to \beta m^{-1}(y) \text{ as } n \to \infty.$$

To examine the convergence rate, we assume additionally that $m^{-1}(\cdot)$ has three bounded derivatives. Supposing also that K is odd and $\int y^2 K(y)\,\mathrm{d}y < \infty$, we get

$$\mu(y;h) - \beta m^{-1}(y) f(y) \int K(\xi)\,\mathrm{d}\xi$$
$$= \beta \int \left[m^{-1}(y+h\xi) f(y+h\xi) - m^{-1}(y) f(y)\right] K(\xi)\,\mathrm{d}\xi,$$

where $\mu(y;h)$ is as in (17). Expanding $m^{-1}(y+h\xi)$ and $f(y+h\xi)$ in a Taylor series and taking into account the fact that $\int yK(y)\,\mathrm{d}y = 0$, we can write $|\mu(y;h) - m^{-1}(y)f(y)\int K(\xi)\,\mathrm{d}\xi| \le \varphi(y)h^2$ with some function φ finite in $m(\mathbb{R})$. Thus $|E\tilde{g}_n(y) - \beta m^{-1}(y) f(y)|$ equals or does not exceed

$$\beta \frac{1}{\sum\limits_{i=1}^{n} h_i} \sum_{i=1}^{n} h_i \left(\mu(y;h_i) - m^{-1}(y) f(y) \int K(\xi)\,\mathrm{d}\xi\right) \le \beta\varphi(y) \frac{\sum\limits_{i=1}^{n} h_i^3}{\sum\limits_{i=1}^{n} h_i}$$

which, for $h_n \sim n^{-1/6}$, is of order $n^{-1/3}$ as $n \to \infty$. Recalling (22), we find $\mathrm{var}[\tilde{g}_n(y)] = O(n^{-2/3})$. Thus $E(\tilde{g}_n(y) - \beta m^{-1}(y)f(y))^2 = O(n^{-2/3})$. For similar reasons, $E(\tilde{f}_n(y) - f(y))^2 = O(n^{-2/3})$ and, finally,

$$\tilde{\mu}_n(y) - \beta m^{-1}(y)f(y) = O(n^{-1/3}) \text{ in probability as } n \to \infty.$$

Applying similar arguments, we obtain

$$\hat{\mu}_n(y) - \beta m^{-1}(y)f(y) = O(n^{-1/3}) \text{ in probability as } n \to \infty.$$

The result is the same as that obtained in (Greblicki, 1992) for the off-line estimate (14).

4.4. Simulation Example

In the example, the state vector was just a scalar and $X_{n+1} = aX_n + U_n$ with $a = 0.6$, $\sigma_U^2 = 1$ and $\sigma_Z^2 = 0.1$. The nonlinearity was of the form $m(v) = v$ for $0 \le v$, and $m(v) = 2v$ for $v < 0$. In algorithms (3) and (4), the kernel was parabolic while $h_n = \delta n^{-\gamma}$. The quality of $\tilde{\mu}_n$ and $\hat{\mu}_n$ was measured with MISE$=$ $E\int_{-3}^{3}(\mu_n(y) - \beta m^{-1}(y))^2\,\mathrm{d}y$, where μ_n is the proper estimate.

For the estimate (3) with $\gamma = 1/5$ and n varying from 10 to 1280, the MISE is shown in Fig. 2. In Fig. 3, we have $\delta = 1$. Results for the algorithm (4) are very similar and so they are not presented. The simulation suggests that too small h and too large γ should be avoided. In other words, too small h_n produces a large error and is not recommended.

For $\delta = 1$ and $\gamma = 1/5$, the error for both the estimates is depicted in Fig. 4. In the figure, the MISE for the off-line estimate (14) is also shown. The error for the latter algorithm is somewhat smaller. The fact that the MISE is greater for on-line estimates is the price paid for the fact that they are recursive.

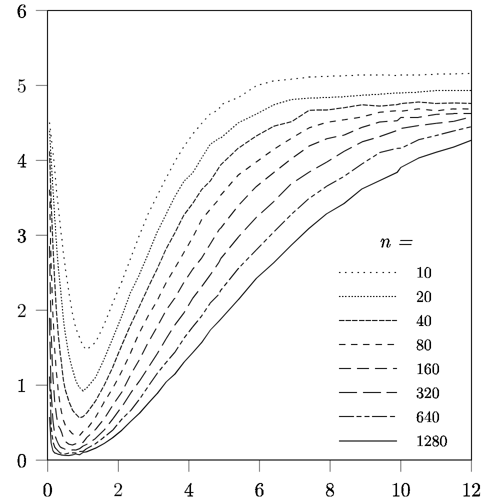

Fig. 2. MISE versus δ.

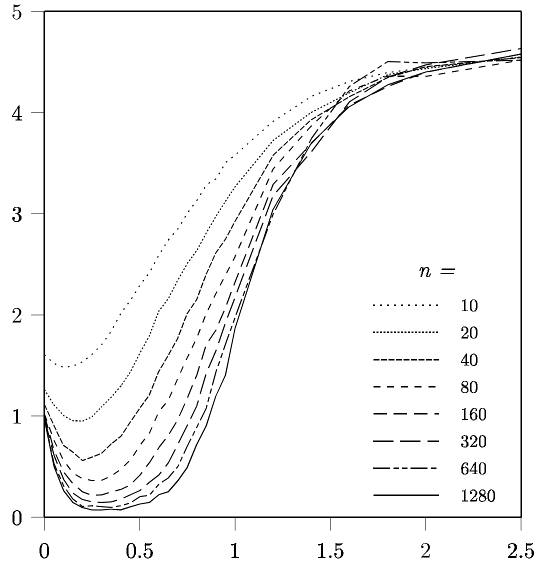

Fig. 3. MISE versus γ.

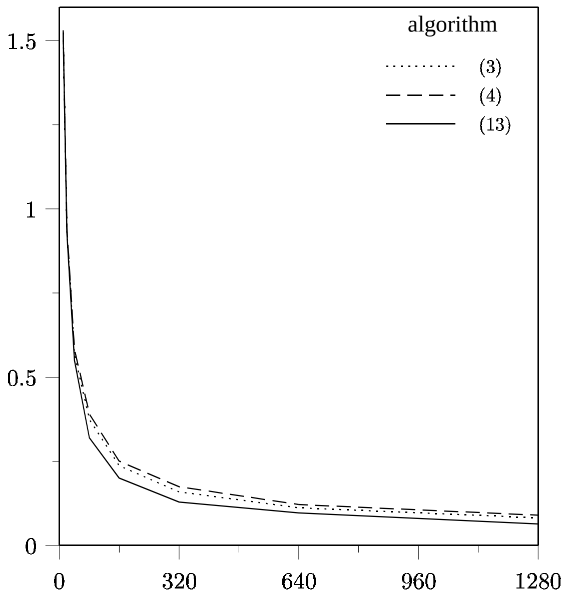

Fig. 4. MISE versus n.

5. Dynamic Subsystem Identification

Lemma 2 suggests the following algorithm to estimate αk_i:

$$\hat{\kappa}_{i,n} = n^{-1} \sum_{j=1}^{n} U_j Y_{j+i}.$$

The algorithm was already examined in (Greblicki, 1997). Setting $\hat{\kappa}_{i,0} = U_0 Y_i$, we can rewrite it in recursive form:

$$\hat{\kappa}_{i,n} = \hat{\kappa}_{i,n-1} - \frac{1}{n} \left(\hat{\kappa}_{i,n-1} - U_n Y_{n+i} \right).$$

The next result establishes its convergence.

Theorem 3. (Greblicki, 1997) *Let m satisfy (19). Then $E(\hat{\kappa}_{i,n} - \alpha k_i)^2 \to 0$ as $n \to \infty$, $i = 1, 2, \ldots$.*

We do not examine the algorithm but solely show that it can be calculated recursively. We want to mention that a recursive algorithm based on a stochastic approximation framework to estimate coefficients of the difference equation corresponding to (1) was proposed in (Wigren, 1993). In that paper, the input signal distribution may not be Gaussian, which is an obvious advantage, but the nonlinear characteristic must be known, which constitutes a severe limitation.

6. Final Remarks

In the present paper, the input signal has a Gaussian distribution which is a typical assumption in both parametric and nonparametric problems of recovering the nonlinearity in Wiener systems and has been applied in all relevant references. We must mention in this context the results of (Krzyżak and Partyka, 1993) concerning the Wiener system identification, i.e. their Theorem 7, seem to lack the restriction that the input is Gaussian.

We assume that both the functional form of the characteristic of the nonlinear subsystem and the order of the linear one are unknown. This means that our *a priori* information about the identified system is nonparametric, which, in the author's opinion, is rather typical in applications.

In turn, the nonlinear characteristic is invertible and differentiable, but its functional form is completely unknown. To see these restrictions in a proper way, we want to point out that in all the references concerning the parametric identification, the characteristic is a polynomial of a known degree (Billings, 1980). Thus, in those papers it is also differentiable but may not be invertible. Nevertheless, its functional form is fixed. In this context, we want to mention that one can verify that also our algorithms work properly when the characteristic is not invertible. To achieve this goal, it is sufficient to apply the method used in (Greblicki, 1997).

Finally, we emphasize again that the algorithms identifying the nonlinear subsystem and those identifying the dynamic one are mutually independent from the numerical viewpoint. This advantage is not taken into consideration very often. As regards the system structure, it has already been applied to model a visual system (den Brinker, 1989), a pH process (Kalafatis *et al.*, 1995), as well as a fluid flow (Wigren, 1993).

All those features make our algorithms interesting not only for researchers, but for engineers too. Therefore, despite theoretical difficulties, the problem of the Wiener system identification is worth further studies.

References

Ahmad I.P. and Lin N. (1976): *Nonparametric sequential estimation of multiple regression function.* — Bull. Math. Statist., Vol.17, pp.63–75.

Billings S. (1980): *Identification of nonlinear systems—A survey.* — IEE Proc., Vol.127, pp.272–285.

Billings S. and Fakhouri S. (1977): *Identification of nonlinear systems using the Wiener model.* — Electron. Lett., Vol.13, pp.502–504.

Billings S. and Fakhouri S. (1978): *Theory of separable processes with applications to the identification of nonlinear systems.* — IEE Proc., Vol.125, pp.1051–1058.

Brillinger D. (1977): *The identification of a particular nonlinear time series system.* — Biometrica, Vol.64, pp.509–515.

Collomb M. (1977): *Quelques proprietés de la méthode du noyau pour l'estimation non paramétrique de la régression en un point fixé.* — C. R. Acad. Sc. Paris, Vol.285, pp.289–292 (in French).

den Brinker A. (1989): *A comparison of results from parameter estimations of impulse responses of the transient visual system.* — Biol. Cybern., Vol.61, pp.139–151.

Devroye L. and Wagner T. (1980): *On the L1-convergence of kernel estimators of regression functions with application in discrimination.* — Z. Wahrsch. Verv. Gebiete, Vol.51, pp.15–25.

Greblicki W. (1992): *Nonparametric identification of Wiener systems.* — IEEE Trans. Inf. Theory, Vol.38, pp.1487–1493.

Greblicki W. (1997): *Nonparametric approach to Wiener system identification.* — IEEE Trans. Circuits and Systems I: Fundamental Theory and Applications, Vol.44, pp.538–545.

Greblicki W. and Pawlak M. (1987): *Necessary and sufficient consistency conditions for a recursive kernel regression estimate.* — J. Multivar. Anal., Vol.23, pp.67–76.

Hunter I. and Korenberg M. (1986): *The identification of nonlinear biological systems: Wiener and Hammerstein cascade models.* — Biol. Cybern., Vol.55, pp.135–144.

Kalafatis A., Afirin N., Wang L. and Cluett W. (1995): *A new approach to the identification of pH processes based on the Wiener model.* — Chem. Eng. Sci., Vol.23, pp.3693–3701.

Krzyzak A. and Pawlak M. (1984): *Almost everywhere convergence of a recursive regression estimate and classification.* — IEEE Trans. Inf. Theory, Vol.30, pp.91–93.

Krzyżak A. and Partyka A.M. (1993): *On identificcation of block oriented systems by nonparametric techniques.* — Int. J. Syst. Sci., Vol.24, pp.1049–1066.

Nadaraya E. (1964): *On regression estimators.* — Theory Prob. Appl., Vol.9, pp.157–159.

Watson G. (1964): *Smooth regression analysis.* — Sankhyā, Ser. A, Vol. 26, pp.359–372.

Westwick D. and Kearney R. (1992): *A new algorithm for the identification of multiple input Wiener systems.* — Biol. Cybern., Vol.68, pp.75–85.

Westwick D. and Verhaegen M. (1996): *Identifying MIMO Wiener systems using subspace model identification methods.* — Signal Process., Vol.52, pp.235–258.

Wheeden R. and Zygmund A. (1977): *Measure and Integral.* — New York: Dekker.

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

Appendix A. Wiener System

Lemma 4. *Let m satisfy (18) and (19). We have*

$$\mu(y;h) \to \beta f(y) m^{-1}(y) \int K(\xi)\,\mathrm{d}\xi \quad \text{as} \quad h \to 0,$$

where $\mu(y;h)$ is as (17). Moreover, $\sup_{0<h} |\mu(y;h)|$ is finite at every $y \in m(\mathbb{R})$.

Proof. The proof applies Lemmas 1, 3 and 7 in Appendix B. ■

Set

$$\omega(y;h) = \frac{1}{h} E\left\{ \|X_0\|^2 K^2\left(\frac{y - Y_0}{h}\right)\right\},$$

and recall that $\psi(y;h)$ was defined in (21).

Lemma 5. *Let m satisfy (18) and (19). We have*

$$\sup_{0<h} \omega(y;h) < \infty, \quad \sup_{0<h} \psi(y;h) < \infty,$$

at every $y \in m(\mathbb{R})$, where $\psi(y;h)$ is as in (21).

Proof. Inspecting the proof of Lemma 2, we deduce that $E\{U_0^2 \mid V_1 = v\} = a + bv^2$ with some a and b. Since $Y_1 = m(V_1)$, we get $E\{U_0^2 \mid Y_1 = y\} = p(y)$, where $p(y) = a + b[m^{-1}(y)]^2$. Applying now Lemma 7 in Appendix B, we see that $\psi(y;h) \to p(y) \int K^2(\xi)\,\mathrm{d}\xi$ as $h \to 0$ at every $y \in m(\mathbb{R})$. Thus $\sup_{0<h} \psi(y;h)$ is finite at every $y \in m(\mathbb{R})$.

To examine $\omega(y;h)$, observe that the pair (X_0, V_0) has a normal distribution with covariance matrix Σc, where $\Sigma = \mathrm{cov}[X_0, X_0]$, zero marginal expectations, and covariances Σ, and σ_V^2. Thus the conditional density of X_0 conditioned on V_0 is normal with expectation $\sigma_V^{-2}\Sigma c V_0$ and the covariance matrix $\Sigma - \sigma_V^{-2}\Sigma c^T c \Sigma$. Hence $E\{X_0 X_0^T \mid V_0\} = \Sigma - \sigma_V^{-2}\Sigma c^T c\Sigma + \sigma_V^{-4}\Sigma c c^T \Sigma V_0^2$, which yields $E\{X_0^T X_0 \mid V_0 = v\} = c + dv$ with $c = \mathrm{trace}(\Sigma - \sigma_V^{-2}\Sigma c^T c\Sigma)$ and $d = \sigma_V^{-4}\,\mathrm{trace}(\Sigma c c^T \Sigma)$. Hence $\omega(y;h) \to w(y) \int K^2(\xi)\,\mathrm{d}\xi$ as $h \to 0$ at every $y \in m(\mathbb{R})$, where $w(y) = c + d[m^{-1}(y)]^2$, see Lemma 7 again. Thus $\sup_{0<h} \omega(y;h) < \infty$ at every $y \in m(\mathbb{R})$. ■

The lemma given below is crucial since it makes it possible to analyze the variance of our algorithms. It modifies and generalizes Lemma A.2 in (Greblicki, 1997).

Lemma 6. *Let m satisfy (18) and (19). Assume that the kernel K fulfils (5)–(8). Let both h and H be positive. Then, for $n = 1, 2, \ldots,$*

$$\left|\operatorname{cov}\left[U_n K\left(\frac{y - Y_{n+1}}{h}\right), U_n K\left(\frac{y - Y_1}{H}\right)\right]\right| \le \frac{H}{h}\left\|A^n\right\| \eta(y),$$

with some $\eta(y)$ finite for every $y \in m(\mathbb{R})$.

Proof. First of all, observe that, owing to Lemma 3, f is continuous at every point $y \in m(\mathbb{R})$. Thus, so is m^{-1}. Therefore all further consideration holds at every $y \in m(\mathbb{R})$. Moreover, owing to (19), m is a Lipschitz function, i.e. for all $v_1, v_2 \in \mathbb{R}$, and some c_m,

$$|m(v_1) - m(v_2)| \le c_m |v_1 - v_2|. \tag{23}$$

In the proof, we need the following inequality:

$$\left|K\left(\frac{y - Y_{n+1}}{h}\right) - K\left(\frac{y - m(\xi_{n+1})}{h}\right)\right| \le d\frac{1}{h}\left\|A^n\right\|\left\|X_1\right\|, \tag{24}$$

with $\xi_{n+1} = \sum_{i=1}^{n} c^T A^{n-i} b U_i + Z_{n+1}$ and $d = c_m c_K \|c\|$. Observing that $V_{n+1} = c^T A^n X_1 + \xi_{n+1}$ and using (8), we see that the quantity on the left-hand side is bounded by $c_K h^{-1}|Y_{n+1} - m(\xi_{n+1})| = c_K h^{-1}|m(V_{n+1}) - m(\xi_{n+1})|$, which, in view of (23), is not greater than $c_m c_K h^{-1}|V_{n+1} - \xi_{n+1}| = c_m c_K h^{-1}|cA^n X_1|$. In this way, (24) has been verified.

Having arrived at this point, observe that, since pairs (U_n, ξ_{n+1}) and (U_0, Y_1) are independent,

$$\operatorname{cov}\left[U_n K\left(\frac{y - m(\xi_{n+1})}{h}\right), U_0 K\left(\frac{y - Y_1}{H}\right)\right] = 0.$$

Therefore the examined covariance equals

$$\begin{aligned}\operatorname{cov}&\left[U_n\left(K\left(\frac{y - Y_{n+1}}{h}\right) - K\left(\frac{y - m(\xi_{n+1})}{h}\right)\right), U_0 K\left(\frac{y - Y_1}{H}\right)\right] \\ &= P(y) + Q(y),\end{aligned}$$

with

$$P(y) = E\left\{U_0 U_n K\left(\frac{y - Y_1}{H}\right)\left[K\left(\frac{y - Y_{n+1}}{h}\right) - K\left(\frac{y - m(\xi_{n+1})}{h}\right)\right]\right\},$$

$$Q(y) = E\left\{U_0 K\left(\frac{y - Y_1}{H}\right)\right\} E\left\{U_n\left[K\left(\frac{y - Y_{n+1}}{h}\right) - K\left(\frac{y - m(\xi_{n+1})}{h}\right)\right]\right\}.$$

Application of (24) yields

$$\begin{aligned}|P(y)| &\le d\frac{1}{h}\left\|A^n\right\| E\left\{\left\|X_1\right\| |U_0 U_n| \left|K\left(\frac{y - Y_1}{H}\right)\right|\right\} \\ &= d\frac{1}{h}\left\|A^n\right\| E\{|U_n|\} E\left\{\left\|X_1\right\| |U_0| \left|K\left(\frac{y - Y_1}{H}\right)\right|\right\}.\end{aligned}$$

Using the Schwartz inequality, we deduce that the second expectation in the above expression is bounded by $H\psi^{1/2}(y;H)\omega^{1/2}(y;H)$. Thus, by virtue of Lemma 5, $|P(y)| \le d(H/h)\|A^n\|\eta_1(y)E|U_0|$ with $\eta_1(y)$ finite.

To examine $Q(y)$, observe that (24) implies

$$\left| E\left\{ U_n \left[K\left(\frac{y-Y_{n+1}}{h}\right) - K\left(\frac{y-m(\xi_{n+1})}{h}\right)\right]\right\}\right| \le d\frac{1}{h}\left\|A^n\right\| E\left|U_n\right| E\left\|X_1\right\|.$$

Thus

$$|Q(y)| \le d\,(H/h)\left\|A^n\right\| E\left|U_0\right| E\left\|X_0\right\| \sup_{0<H} \mu(y;H),$$

with $\mu(y;H)$ as in (17). Application of Lemma 4 completes the proof. ■

Appendix B. General Results

Lemma 7. *Let X be a random variable with a probability density f, and let φ be a Lebesgue measurable function. Suppose that $E|\varphi(X)| < \infty$ and the Lebesgue measurable kernel K fulfils (5)–(7). Then*

$$\frac{1}{h}E\left\{\varphi(X)K\left(\frac{x-X}{h}\right)\right\} \to \varphi(x)f(x)\int K(\xi)\,\mathrm{d}\xi \text{ as } h \to 0$$

at every point $x \in \mathbb{R}$ at which both φ and f are continuous. Moreover,

$$\sup_{0<h}\frac{1}{h}E\left|\varphi(X)K\left(\frac{x-X}{h}\right)\right| < \infty$$

at the same points.

Proof. The first part of the lemma can be found in (Wheeden and Zygmund, 1997). The proof of the other is immediate. ■

Received: 16 February 2001
Revised: 7 August 2001