Recursive identification of continuous-time Wiener systems

WŁODZIMIERZ GREBLICKI

Institute of Engineering Cybernetics, Wrocław University of Technology,
50-370 Wrocław, Poland.

Recursive algorithms to identify both subsystems of a continuous-time Wiener system are presented. The system is driven and disturbed by Gaussian white random signals. The impulse response of the linear dynamic subsystem is recovered with a correlation method. It is shown that the inverse of the non-linear characteristic of the other subsystem is a regression function. Then, to recover the inverse, two estimates are presented. The algorithms converge to the unknown impulse response, and the inverse of the characteristic, respectively. Convergence rates are presented. Moreover, results of simulation examples are given.

printed in:
International Journal of Control
Vol. 72, No. 11, pp. 981–989, 1999

1 Introduction

Compared to simple systems, the identification of composite ones creates specific problems. A great difficulty is produced by the fact that signals interconnecting the subsystems are not measured and we have to identify them from observations taken at input and output of the whole system. As a result, when the *a priori* information is small, which is rather typical in real situations, we can recover characteristics of particular subsystems only up to some constants. Another problem is the fact that algorithms identifying subsystems should be numerically independent one of another. This goal is not always easy to achieve.

Systems with the cascade structure are most often investigated. Quite a lot work have been devoted to Hammerstein systems, i.e., systems consisting of a non-linear system followed by linear and dynamic systems (see, e.g., Billings 1980, Bendat 1990, as well as papers cited therein). When the subsystems are connected in reverse order, we deal with so called Wiener systems. Their identification is much tougher and, therefore, much fewer papers on this topic can be found in the literature (see Brillinger 1977, Billings 1980, Hunter and Korenberg 1986, Hasiewicz 1987, Bendat 1990, Westwick and Kearney 1992, Wigren 1994, for parametric problems). Most of the mentioned authors have confined their works to the analysis of statistical properties of Wiener systems. They have only noticed that, e.g., some correlation functions depend on unknown system parameters and have then proposed some algorithms to estimate them. Almost all of those works, however, lack proofs of convergence, probably, because the type of dependence between consecutive output observations in Wiener systems is complicated and not easy to analyze. The nonparametric approach to the discrete-time Wiener system has been proposed and examined by Greblicki (1992, 1994, 1997). He has presented some classes of estimators to recover the non-linear characteristic in discrete-time systems and, then, shown their local and global convergence.

In this paper, we identify a continuous-time Wiener system. The *a priori* information about the system, i.e., about its subsystems is so small that problems of the identification of both the linear and non-linear parts are nonparametric. The order of the dynamic subsystem is unknown. The non-linear characteristic is only assumed to be invertible and satisfy a Lipschitz condition. These and the cascade structure of the system result in the fact that we can recover both the impulse response and the non-linear characteristic only up to some unknown constants. The impulse response is estimated recursively with the help of correlation methods. As far as the non-linear subsystem is concerned, the inverse of its characteristic can be represented as a regression function. We propose two recursive algorithms to estimate the regression. We show their pointwise convergence and examine convergence rate. The algorithms are continuous-time versions of those examined in the statistical literature for independent discrete-time observations (see, e.g., Härdle 1990 or Prakasa Rao 1983). Our problem is, however, more complicated because our observations come from a dynamic system and, therefore, are dependent. Theoretical results presented in the paper are asymptotic in nature and, therefore, to have a look into their behavior for moderate and small length of the observation interval, we present results of a simulation example.

We want to mention here that also statisticians estimate a regression from dependent data. They use, however, a model of dependence which is quite different from ours. They impose so called mixing restrictions on observations (see, e.g. Györfi *et al.* 1989). Restrictions of this kind are not easy to verify and, therefore, not very useful in system identification.

2. Identification Problem

The input of a Wiener system shown in figure 1 is a Gaussian stationary white random process $\{U(t); t \in (-\infty, \infty)\}$ with zero mean, and autocovariance function $\sigma_U^2 \delta(.)$. By δ, we denote the Dirac impulse. The system consists of two subsystems connected in a cascade. A linear dynamic one is followed by a non-linear memoryless one. The dynamic subsystem is described by the following state-space equation:

$$\begin{aligned} \dot{X}(t) &= AX(t) + bU(t) \\ W(t) &= c^T X(t), \end{aligned}$$

where A, b, and c are all unknown. The subsystem is asymptotically stable and, therefore, $\int_0^\infty k^2(t)dt < \infty$, where $k(t) = c^T e^{At} b$ is the impulse response of the subsystem. Thus, both $\{X(t); t \in (-\infty, \infty)\}$ and $\{W(t); t \in (-\infty, \infty)\}$ are stationary Gaussian processes. We also assume that

$$k(t) = 0, \tag{1}$$

i.e. that $c^T b = 0$. The restriction is satisfied by a wide class of subsystems, i.e., by those whose transfer function $c^T(sI - A)^{-1}b = L(s)/M(s)$, where L and M are polynomials, is such that the degree of M is greater than that of L by 2 or more. The output of the subsystem is disturbed by additive noise $\{Z(t); t \in (-\infty, \infty)\}$, i.e.,

$$V(t) = W(t) + Z(t).$$

The noise $\{Z(t)\}$ is a zero-mean stationary white Gaussian random process with autocovariance function $\sigma_Z^2 \delta(.)$. The process is independent of $\{U(t)\}; t \in (-\infty, \infty)\}$.

The non-linear subsystem has a characteristic m, which means that

$$Y(t) = m(V(t)).$$

The characteristic satisfies the following Lipschitz condition:

$$|m(x) - m(y)| \le c_m |x - y|, \tag{2}$$

some c_m, all x, y in R. We assume that the characteristic m is an invertible Borel measurable function and denote its inverse by m^{-1}. Since, moreover, m^{-1} is differentiable, $Y(t)$ has a probability density denoted by f.

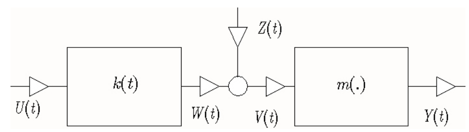

Figure 1. The identified Wiener system.

The aim of the paper is to identify both subsystems, i.e., to estimate the characteristic m of the non-linear part, and the impulse response k of the linear subsystem from observations taken at input and output of the whole system, i.e., from $\{U(t), Y(t); t \in [0, \infty)\}$.

3. Nonlinear Subsystem Identification Algorithms

Obviously, $V(t)$ has zero mean and variance $\sigma_V^2 = \sigma_U^2 \int_0^\infty k^2(t)dt + \sigma_Z^2$. Observe now that the pair $(U(t-\lambda), V(t))$ has a Gaussian distribution with marginal variances σ_U^2, σ_V^2, and the correlation coefficient $(\sigma_U/\sigma_V)k(\lambda)$. Therefore, $E\{U(t-\lambda)|V(t)\} = (\sigma_U^2/\sigma_V^2)k(\lambda)V(t)$. Denoting $\alpha = (\sigma_U^2/\sigma_V^2)k(\lambda)$, we now obtain

Lemma 1 *Let* $\lambda > 0$. *In the system,*

$$E\{U(t-\lambda)|Y(t) = y\} = \alpha m^{-1}(y).$$

Due to Lemma 1 recovering $\alpha m^{-1}(y)$ is equivalent to the estimation of the regression $E\{U(t-\lambda)|Y(t) = y\}$. To estimate the regression, we propose the following algorithms:

$$\tilde{\mu}_t(y) = \frac{\displaystyle\int_0^t \frac{1}{h(\tau)} U(\tau - \lambda) K\left(\frac{y - Y(\tau)}{h(\tau)}\right) d\tau}{\displaystyle\int_0^t \frac{1}{h(\tau)} K\left(\frac{y - Y(\tau)}{h(\tau)}\right) d\tau}, \tag{3}$$

and

$$\hat{\mu}_t(y) = \frac{\displaystyle\int_0^t U(\tau - \lambda) K\left(\frac{y - Y(\tau)}{h(\tau)}\right) d\tau}{\displaystyle\int_0^t K\left(\frac{y - Y(\tau)}{h(\tau)}\right) d\tau}, \tag{4}$$

where K is a suitably selected Borel measurable kernel function and h is a bandwidth function, respectively. From the statistical viewpoint, the algorithms estimate the regression in Lemma 1 in a nonparametric way (see e.g., Härdle 1990 or Prakasa Rao 1983).

On the nonnegative Borel measurable kernel K, we impose the following constraints:

$$\sup_{y\in R}|K(y)| = \kappa < \infty, \tag{5}$$

$$\int_{-\infty}^{\infty} K(y)dy = 1, \tag{6}$$

$$K(y)y^{1+\delta} \to 0 \text{ as } |y| \to \infty, \tag{7}$$

some nonnegative δ. Moreover,

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

all x, y in R, some c_K. Depending on context, the positive Borel measurable function h satisfies some of the following conditions:

$$h(.) \text{ is monotonous}, \tag{9}$$

$$h(t) \to 0 \text{ as } t \to \infty, \tag{10}$$

$$th^2(t) \to \infty \text{ as } t \to \infty, \tag{11}$$

$$h(t)\int_0^t h(\tau)d\tau \to \infty \text{ as } t \to \infty. \tag{12}$$

Estimate (3) can be rewritten in the following form:

$$\tilde{\mu}_t(y) = \frac{\tilde{\xi}_t(y)}{\tilde{\eta}_t(y)},$$

where $\tilde{\xi}_t$ and $\tilde{\eta}_t$ are defined by:

$$\tilde{\xi}_t(y) = \frac{1}{t}\int_0^t U(\tau - \lambda)\frac{1}{h(\tau)}K\left(\frac{y - Y(\tau)}{h(\tau)}\right)d\tau, \tag{13}$$

and

$$\tilde{\eta}_t(y) = \frac{1}{t}\int_0^t \frac{1}{h(\tau)}K\left(\frac{y - Y(\tau)}{h(\tau)}\right)d\tau, \tag{14}$$

respectively. Now, observe that $\tilde{\xi}_t(y)$ and $\tilde{\eta}_t(y)$ can be calculated with the following recursive formulas:

$$\frac{d}{dt}\tilde{\xi}_t(y) = -\frac{1}{t}\left[\tilde{\xi}_t(y) - \frac{1}{h(t)}U(t-\tau)K\left(\frac{y - Y(t)}{h(t)}\right)\right],$$

where $\tilde{\xi}_0(y) = 0$, and

$$\frac{d}{dt}\tilde{\eta}_t(y) = -\frac{1}{t}\left[\tilde{\eta}_t(y) - \frac{1}{h(t)}K\left(\frac{y - Y(t)}{h(t)}\right)\right],$$

where $\hat{\eta}_0(y) = 0$, respectively. The mean speed at which $\tilde{\xi}_t(y)$ converges to zero is proportional to

$$E\{\tilde{\xi}_t(y)\} - \frac{1}{h(t)}E\left\{U(t-\lambda)K\left(\frac{y - Y(t)}{h(t)}\right)\right\}$$

which, in view of Lemma 1, and Lemma 9 in Appendix, converges to $\alpha m^{-1}(y)f(y)$, provided that both h, and K have been suitably selected. Thus it is reasonable to expect that $\tilde{\xi}_t(y)$ converges to $\alpha m^{-1}(y)f(y)$. For similar reasons, one can expect $\tilde{\eta}_t(y)$ to converge to $f(y)$. It finally suggests that (3) converges to $\alpha m^{-1}(y)$.

For estimate (4), we have

$$\hat{\mu}_t(y) = \frac{\hat{\xi}_t(y)}{\hat{\eta}_t(y)},$$

where $\hat{\xi}_t$ and $\hat{\eta}_t$ are defined by

$$\hat{\xi}_t(y) = \frac{\displaystyle\int_0^t U(\tau-\lambda)K\left(\frac{y - Y(\tau)}{h(\tau)}\right)d\tau}{\displaystyle\int_0^t h(\tau)d\tau}, \tag{15}$$

and

$$\hat{\eta}_t(y) = \frac{\displaystyle\int_0^t K\left(\frac{y - Y(\tau)}{h(\tau)}\right)d\tau}{\displaystyle\int_0^t h(\tau)d\tau}. \tag{16}$$

respectively. Observe that

$$\frac{d}{dt}\hat{\xi}_t(y) = -\gamma(t)\left[\hat{\xi}_t(y) - \frac{1}{h(t)}U(t-\lambda)K\left(\frac{y - Y(t)}{h(t)}\right)\right] \tag{17}$$

where

$$\gamma(t) = \frac{h(t)}{\displaystyle\int_0^t h(\tau)d\tau},$$

and where $\hat{\xi}_0(y) = 0$. Similarly

$$\frac{d}{dt}\hat{\eta}_t(y) = -\gamma(t)\left[\hat{\eta}_t(y) - \frac{1}{h(t)}K\left(\frac{y - Y(t)}{h(t)}\right)\right], \tag{18}$$

where $\hat{\eta}_0(y) = 0$. Using formulas (17) and (18), we can calculate estimate (4) in a recursive way.

4. Linear Subsystem Identification Algorithm

To introduce the algorithm recovering the impulse response, observe

$$E\{U(t-\lambda)Y(t)\} = E\{Y(t)E\{U(t-\lambda)|Y(t)\}\}$$

which, by virtue of Lemma 1 equals $\alpha E\{m^{-1}(Y(t))Y(t)\}$. In this way we have verified

Lemma 2 *Let $\lambda > 0$. In the system,*

$$E\{U(t-\lambda)Y(t)\} = \beta k(\lambda),$$

where $\beta = (\sigma_U^2/\sigma_V^2)E\{V(0)Y(0)\}$.

The lemma suggests the following estimate of $k(\lambda)$:

$$\tilde{k}_t(\lambda) = \frac{1}{t}\int_0^t U(\tau-\lambda)Y(\tau)d\tau \quad (19)$$

which can be rewritten in the following recursive form:

$$\frac{d}{dt}\tilde{k}_t(\lambda) = -\frac{1}{t}\left[\tilde{k}_t(y) - U(t-\lambda)Y(t)\right],$$

where $\tilde{k}_0(\lambda) = 0$.

Theorem 1 *Let (1) and (2) hold. Then*

$$E(\tilde{k}_t(\lambda) - \beta k(\lambda))^2 \to 0 \text{ as } t \to \infty,$$

any $0 < \lambda$.

Since $E\tilde{k}_t(\lambda) = \beta k(\lambda)$, to prove the theorem it suffices to verify that $\operatorname{var}[\tilde{k}_t(\lambda)]$ vanishes as t increases to infinity. The variance equals

$$\frac{1}{t^2}\int_0^t\int_0^t \operatorname{cov}\left[U(\eta-\lambda)Y(\eta), U(\xi-\lambda)Y(\xi)\right] d\xi d\eta$$

which, by virtue of Lemma 7, is bounded in absolute value by

$$\frac{c_2}{t^2}\int_0^t\int_0^\eta \|e^{A(\eta-\xi)}\| d\xi d\eta + \frac{c_1}{t^2}\iint_{|\eta-\xi|<\lambda} d\xi d\eta$$
$$= O(t^{-1}).$$

Therefore

$$E(\tilde{k}_t(\lambda) - \beta k(\lambda))^2 \le \frac{c}{t},$$

some constant c. The proof has been completed. ■

5. Convergence of Nonlinear Subsystem Identification Algorithms

We shall now show that our algorithms are consistent estimates of $\alpha m^{-1}(y)$. Here and in the whole paper, almost everywhere means almost everywhere with respect to the Lebesgue measure.

Theorem 2 *Let the impulse response satisfy (1) and let m satisfy (2). Let the positive bandwidth function h satisfy (9)–(11). Let the nonnegative Borel kernel K satisfy (5), (6), (8). If (7) holds with $\delta = 0$, then*

$$\tilde{\mu}_t(y) \to \alpha m^{-1}(y) \text{ as } t \to \infty \text{ in probability,}$$

every $y \in R$ at which both m^{-1} and f are continuous, and $f(y) > 0$. If (7) holds with $0 < \delta$, then the convergence takes place also at almost every $y \in R$ at which $f(y) > 0$.

Proof. Let (7) hold with $0 < \delta$. We have $\tilde{\mu}_t(y) = \tilde{\xi}_t(y)/\tilde{\eta}_t(y)$, where $\tilde{\xi}_t(y)$ and $\tilde{\eta}_t(y)$ are defined by (13) and (14), respectively. Observe

$$E\tilde{\xi}_t(y) = \frac{1}{t}\int_0^t E\left\{\alpha m^{-1}(Y(0))K\left(\frac{y-Y(0)}{h(\tau)}\right)\right\}d\tau$$

and apply Lemma 10 in Appendix to find

$$E\tilde{\xi}_t(y) \to \alpha m^{-1}(y)f(y)\int_{-\infty}^{\infty} K(x)dx \text{ as } t \to \infty, \quad (20)$$

almost every $y \in R$. In turn, $\operatorname{var}[\tilde{\xi}_t(y)]$ equals

$$\frac{1}{t^2}\int_0^t\int_0^t \operatorname{cov}\left[U(\xi-\lambda)K\left(\frac{y-Y(\xi)}{h(\xi)}\right),\right.$$
$$\left. U(\eta-\lambda)K\left(\frac{y-Y(\eta)}{h(\eta)}\right)\right] d\xi d\eta.$$

Applying Lemma 6 in Appendix, we find the above expression not greater than

$$4\kappa^2\lambda\sigma_U^2\frac{1}{th^2(t)}$$
$$+\frac{2c_K d}{th^2(t)}\frac{1}{t}\int_0^t E\left\{\phi_\tau(Y(0))K\left(\frac{y-Y(0)}{h(\tau)}\right)\right\}d\tau$$

Invoking Lemma 9 in Appendix, we find the quantity in square brackets converging as $t \to \infty$ at almost every $y \in R$. Thus,

$$\operatorname{var}[\tilde{\xi}_t(y)] = O\left(\frac{1}{th^2(t)}\right), \quad (21)$$

almost every $y \in R$. Recalling now (11) and (20), we get

$$\tilde{\xi}_t(y) \to \alpha m^{-1}(y)\int_{-\infty}^{\infty} K(x)dx \text{ as } t \to \infty$$

in probability, almost every $y \in R$. Since, using similar arguments, we can verify

$$\tilde{\eta}_t(y) \to f(y)\int_{-\infty}^{\infty} K(x)dx \text{ as } t \to \infty$$

in probability, almost every $y \in R$, the first part of the theorem has been completed. To verify convergence at continuity points, we apply similar arguments and complete the proof. ■

Theorem 3 *Let the impulse response satisfy (1) and let m satisfy (2). Let the positive bandwidth function h satisfy (9), (10), (12). Let the nonnegative Borel kernel K satisfy (5), (6), (8). If (7) holds with $\delta = 0$, then,*

$$\hat{\mu}_t(y) \to \alpha m^{-1}(y) \text{ as } t \to \infty \text{ in probability,}$$

every $y \in R$ at which both m^{-1} and f are continuous, and $f(y) > 0$. If (7) holds with $0 < \delta$, then the convergence takes place also at almost every $y \in R$ at which $f(y) > 0$.

Proof. Let (7) hold with $\delta > 0$. We have $\hat{\mu}_t(y) = \hat{\xi}_t(y)/\hat{\eta}_t(y)$, where $\hat{\xi}_t(y)$ and $\hat{\eta}_t(y)$ are defined by (15) and (16), respectively. Observing

$$E\hat{\xi}_t(y) = \frac{\displaystyle\int_0^t E\left\{\alpha m^{-1}(Y(0))K\left(\frac{y-Y(0)}{h(\tau)}\right)\right\}d\tau}{\displaystyle\int_0^t h(\tau)d\tau}$$

and applying Lemma 10 in Appendix, we obtain

$$E\hat{\xi}_t(y) \to \alpha m^{-1}(y)f(y)\int_{-\infty}^{\infty}K(x)dx \text{ as } t\to\infty, \quad (22)$$

almost every $y \in R$. In turn, $\mathrm{var}[\hat{\xi}_t(y)]$ equals

$$\frac{1}{\left[\int_0^t h(\tau)d\tau\right]^2}\int_0^t\int_0^t \mathrm{cov}\left[U(\xi-\lambda)K\left(\frac{y-Y(\xi)}{h(\xi)}\right),\right.$$
$$\left.U(\eta-\lambda)K\left(\frac{y-Y(\eta)}{h(\eta)}\right)\right]d\xi d\eta.$$

An application of Lemma 5 in Appendix, leads to the following upper bound for the above quantity

$$\frac{4\kappa^2\lambda\sigma_U^2}{h(t)\displaystyle\int_0^t h(\tau)d\tau}\frac{th(t)}{\displaystyle\int_0^t h(\tau)d\tau} + \frac{2c_K d}{h(t)\displaystyle\int_0^t h(\tau)d\tau}$$
$$\times\frac{\displaystyle\int_0^t E\left\{\phi_\tau(Y(0))K\left(\frac{y-Y(0)}{h(\tau)}\right)\right\}d\tau}{\displaystyle\int_0^t h(\tau)d\tau}.$$

Since (9) and (0) hold, $th(t) \le \int_0^t h(\tau)d\tau$ and, therefore, the first term is not greater than $4\kappa^2\lambda\sigma_U^2/h(t)\int_0^t h(\tau)d\tau$. Invoking Lemma 10 in Appendix, we find

$$\frac{\displaystyle\int_0^t E\left\{\phi_\tau(Y(0))K\left(\frac{y-Y(0)}{h(\tau)}\right)\right\}d\tau}{\displaystyle\int_0^t h(\tau)d\tau}$$

converging to $\phi(y)f(y)\int_{-\infty}^{\infty}K(x)dx$ as $t\to\infty$ at almost every $y \in R$. Hence,

$$\mathrm{var}[\hat{\xi}_t(y)] = O\left(\frac{1}{h(t)\int_0^t h(\tau)d\tau}\right), \quad (23)$$

almost every $y \in R$. Recalling now (22), we get

$$\hat{\xi}_t(y) \to \alpha m^{-1}(y)\int_{-\infty}^{\infty}K(x)dx \text{ as } t\to\infty$$

in probability, almost every $y \in R$. Since, using similar arguments, we can verify

$$\hat{\eta}_t(y) \to f(y)\int_{-\infty}^{\infty}K(x)dx \text{ as } t\to\infty$$

in probability, almost every $y \in R$, the almost every version of the lemma has been verified. Since the verification of the continuous version goes in a similar way, the proof of the theorem has been completed. ∎

6. Convergence Rate

As the kernel we can apply, e.g., $1/(1+y^2)$, $\exp(-y)$, $\exp(-y^2)$. Kernels with bounded support can be also employed. Examples of such are, e.g., a triangle kernel equal to $1-|y|$ or 0, for $|y| \le 1$ or $|y| > 1$, respectively, and a parabolic kernel equal $1-y^2$, for $|y| \le 1$ and 0 otherwise, respectively. As far as the bandwidth function h is concerned, we can use, e.g., $h(t) = ct^{-\alpha}$, any positive c. In such a case, (10)–(11) as well as (10)–(12) required by Theorems 1 and 2 are is satisfied for $0 < \alpha < 1/2$.

Theorems 1 and 2 establish convergence of our algorithms. To compare them assume that m^{-1} is differentiable. Therefore, $Y(t)$, has a density $f(y) = f_V(|m^{-1}(y)|)[m^{-1}(y)]'$, where f_V is the density of $V(t)$. Observe that f exists at each point at which the derivative of m^{-1} exists. Suppose, moreover, that $\int_{-\infty}^{\infty}xK(x)dx = 0$, $\int_{-\infty}^{\infty}x^2K(x)dx < \infty$ and that $h(t) \sim ct^{-\alpha}$, some positive c, and α. The last assumption means that $h(t)/ct^{-\alpha}$ has a nonzero limit as $t\to\infty$. Applying now Lemma 11 in Appendix, we find

$$E\tilde{g}_t(y) - \alpha m^{-1}(y)f(y) = O(t^{-\alpha}),$$

where $\tilde{g}(y)$ is defined by (13). Recalling (21), and selecting $\alpha = -1/6$, we get

$$E\left[\tilde{g}_t(y) - \alpha m^{-1}(y)f(y)\int_{-\infty}^{\infty}K(x)dx\right]^2 = O(t^{-2/3}).$$

Since, for similar reasons, $\tilde{f}_t(y)$ converges to $f(y)$ at the same rate, we obtain

$$\tilde{\mu}_t(y) - \alpha m^{-1}(y) = O(t^{-1/3}) \text{ as } t\to\infty$$

in probability. Using similar arguments, one can verify

$$\hat{\mu}_t(y) - \alpha m^{-1}(y) = O(t^{-1/3}) \text{ as } t\to\infty$$

in probability. Therefore, asymptotic properties of our algorithms are similar. Our rate is not much worse than $t^{-1/2}$, i.e., with the rate usually obtained for parametric inference.

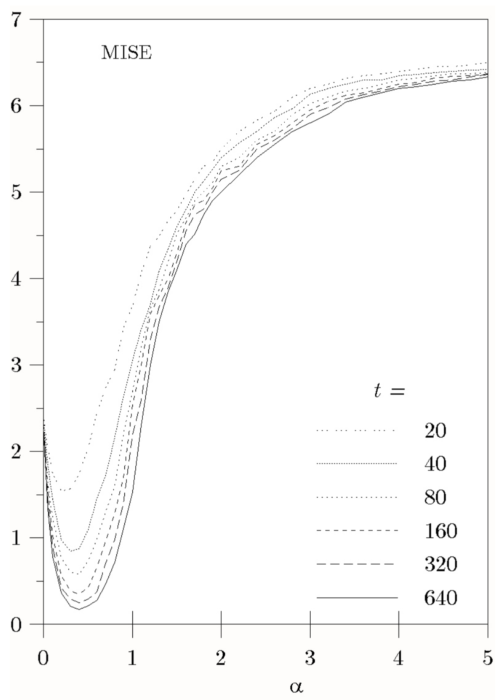

Figure 1: MISE versus α; $h(\tau) = 2\tau^{-\alpha}$, various t.

7. Simulation Example

Results concerning estimates of the non-linear characteristic given in the paper are asymptotic, i.e., hold for large t. From a practical viewpoint, it is interesting to examine our estimates for small or moderate t. We present results of a simulation example in which the transfer function of the linear part is $12/(s+1)(s+4)$ while $m(v) = (1/4)v + v^2\text{sign}(v)$. In the example, $\sigma_U^2 = 1$, and $\sigma_Z^2 = 0.1$. In both estimates, $\lambda = 0.7$, and the kernel is parabolic, i.e. equals 0 or $1 - v^2$ for $|v| > 1$ or $|v| \le 1$, respectively. In the paper, we have shown pointwise convergence of the estimates, nevertheless, in the example, we have numerically calculated the mean integrated square error (MISE for short) defined as $\int_{-2}^{2}(\mu_t(y) - \alpha m^{-1}(y))^2 dy$, where μ_t is $\hat{\mu}_t$ or $\tilde{\mu}_t$, respectively.

At first, we have selected $h(\tau) = c\tau^{-1/6}$ and estimated the MISE for c varying from 0 to 12. For t equal to 20, 40, $\cdots$, 640, results are shown in figure 2. In turn, figure 3 shows the MISE for $h(\tau) = 2\tau^{-\alpha}$, and α varying in the interval $[0, 5]$. For $h(\tau) = c\tau^{-1/6}$, various c, figure 4 shows the MISE versus t. All results obtained for our two estimates differ by a few percents only and, therefore, each curve represents both estimates. From presented results, we conclude that too small $h(\tau)$, i.e., too small c, and too large α, should be avoided.

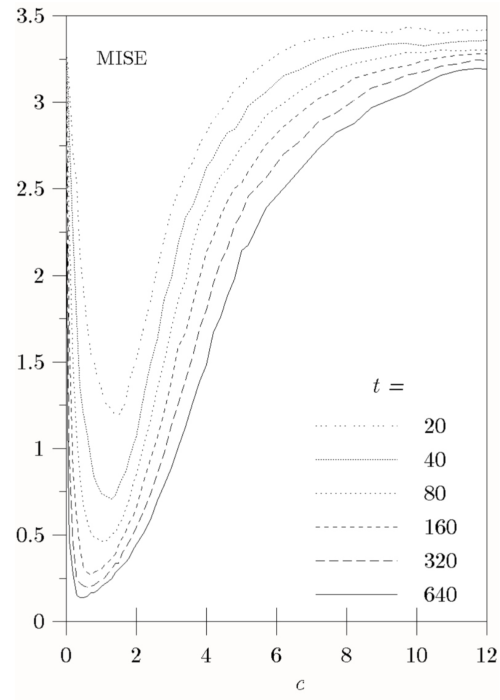

Figure 2: MISE versus c, $h(\tau) = c\tau^{-1/6}$, various t.

8. Final Remarks

We want to mention here that, for discrete-time systems, the result given in Lemma 2 has been already known (Brillinger 1977). The same result, but for continuous-time systems has been presented in Billings, and Fakhouri (1978), however, for the non-linear characteristic of a polynomial of a finite order, only. As far as the assumption that the input signal is Gaussian, it is crucial in their papers as well as in ours.

Observe that our *a priori* information about both subsystems is very small. Since the order of the linear dynamic one is unknown, the problem of recovering its impulse response is nonparametric. So is that of the nonlinear part of the system since its characteristic can't be represented in a parametric form.

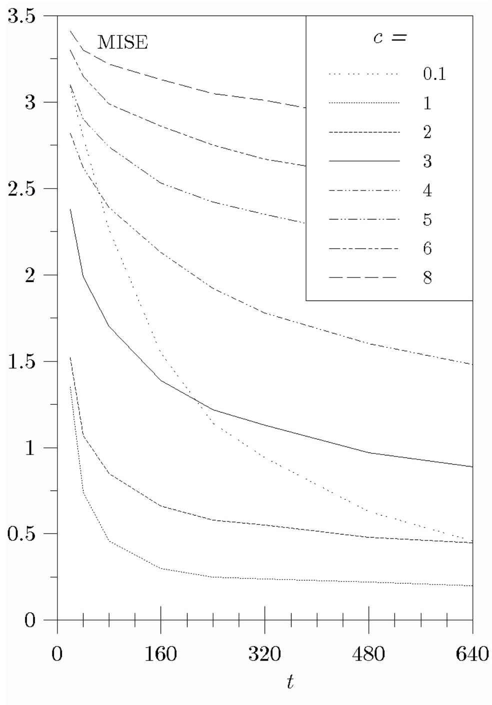

Figure 3: MISE versus t, $h(\tau) = \tau^{-1/6}$, various c.

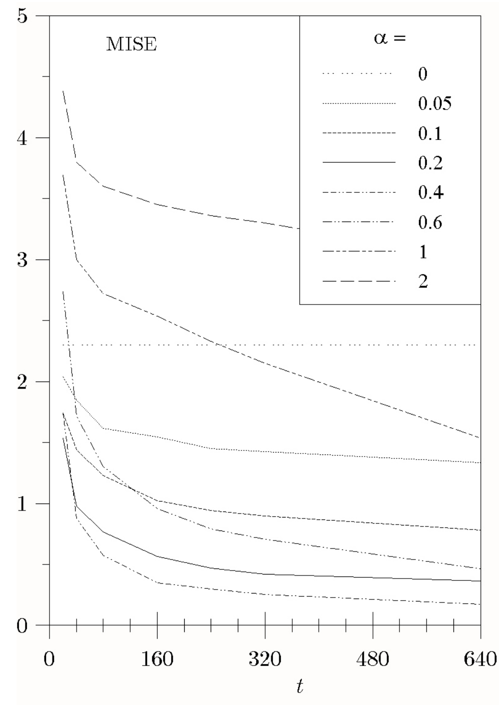

Figure 4: MISE versus t, $h(\tau) = 2\tau^{-2}$, various α.

Finally, we want to mention that our estimates are special cases of more general algorithms given by (17) and (18). Wide classes of recursive estimates can be obtained by various selections of $\gamma(.)$. The problem of their convergence has been, however, left for future works.

Acknowledgement

This work was supported by the Polish State Committee for Scientific Research.

Appendix

The Wiener System

Lemma 3 *Let (1), and (2) hold. Let φ and ψ be nonnegative Borel functions and let ψ satisfy the following Lipschitz condition:*

$$|\psi(x) - \psi(y)| \leq c_\psi |x - y|, \tag{24}$$

some c_ψ, all x, y in R. Let $0 < \lambda$. Then,

$$|\text{cov}\left[U(\tau-\lambda)\varphi(Y(\tau)), U(t-\lambda)\psi(Y(t))\right]|$$

$$\leq \begin{cases} 2d_\varphi d_\psi \sigma_U^2, & \text{for all } t,\ \tau, \\ c_\psi \|e^{A(t-\tau)}\| E\{\rho_\lambda(Y(0))\varphi(Y(0))\}, & \text{for } \lambda + \tau \leq t, \end{cases}$$

where $d_\varphi = \sup_y \varphi(y)$, $d_\psi = \sup_y \psi(y)$, some function ρ_λ independent of φ, ψ, t, and τ. Moreover, $E|\rho_\lambda(Y(0))| < \infty$.

Proof. First of all, observe that the covariance in the assertion equals $S_1 - S_2$, where

$$S_1 = E\left\{U(\tau-\lambda)U(t-\lambda)\varphi(Y(\tau))\psi(Y(t))\right\},$$

and

$$S_2 = E\left\{U(t-\lambda)\varphi(Y(t))\right\} E\left\{U(t-\lambda)\psi(Y(t))\right\}.$$

Observing $|S_1| \leq d_\varphi d_\psi \sigma_U^2$ and $|S_2| \leq d_\varphi d_\psi \sigma_U^2$, we complete the proof of the first part of the inequality in the assertion.

To verify second, suppose

$$\lambda + \tau \leq t. \tag{25}$$

Denote $\xi(t) = \int_\tau^t k(t-v)U(v)dv + Z(t)$ and observe that

$$V(t) = c^T e^{A(t-\tau)} X(\tau) + \xi(t).$$

Thus, using (2) and (24), we get

$$|\psi(m(V(t))) - \psi(m(\xi(t)))| \leq c_\psi |m(V(t)) - m(\xi(t))|$$

$$\leq c_\psi c_m |c^T e^{A(t-\tau)} X(\tau)|. \tag{26}$$

Obviously, $\xi(t)$ depends on $U(\upsilon)$, $\upsilon \in [\tau, t]$. Since (1) holds, $Y(\tau)$ is independent of $U(\upsilon)$, $\upsilon \in [\tau, \infty)$. Thus, we have found $\xi(t)$ and $Y(\tau)$ independent. Recalling (25) and $0 < \lambda$, we conclude that pairs $(U(\tau-\lambda), Y(\tau))$ and $(U(t-\lambda), \xi(t))$ are independent. Therefore,

$$\operatorname{cov}\left[U(\tau-\lambda)\varphi(Y(\tau)), U(t-\lambda)\psi(m(\xi(t)))\right] = 0. \tag{27}$$

Having (26) and (27), we begin the main part of the proof. We apply (27) to find the covariance in the assertion equal to

$$\operatorname{cov}\left[U(\tau-\lambda)\varphi(Y(\tau)),\right.$$
$$\left. U(t-\lambda)\left[\psi(m(V(t))) - \psi(m(\xi(t)))\right]\right] = Q_1 - Q_2,$$

where

$$Q_1 = E\{U(\tau-\lambda)U(t-\lambda)\varphi(Y(\tau)) \times \left[\psi(m(V(t))) - \psi(m(\xi(t)))\right]\},$$

and

$$Q_2 = E\left\{U(\tau-\lambda)\varphi(Y(\tau))\right\} \times E\left\{U(t-\lambda)\left[\psi(m(V(t))) - \psi(m(\xi(t)))\right]\right\}.$$

Using (26), we find $|Q_1|$ not greater than

$$c_\psi c_m |c| \|e^{A(t-\tau)}\| \times E\left\{|U(\tau-\lambda)U(t-\lambda)|\varphi(Y(\tau))\|X(\tau)\|\right\}.$$

Owing to (1), both $Y(\tau)$ and $X(\tau)$ are independent of $U(\upsilon)$, $\upsilon \in (-\infty, \tau]$. Thus, since (25) holds, $U(t-\lambda)$ is independent of both $Y(\tau)$, and $X(\tau)$. Therefore,

$$|Q_1| \leq c_\psi c_m |c| \|e^{A(t-\tau)}\| \times E\left\{|U(0)|\right\} E\left\{\rho_1(Y(0))\varphi(Y(0))\right\},$$

where $\rho_1(Y(0)) = E\{|U(-\lambda)|\|X(0)\|\,|U(0)\}$. For similar reasons $|Q_2|$ is bounded by

$$c_\psi c_m |c| \|e^{A(t-\tau)}\| \times E\left\{|U(0)|\right\} E\left\{\|X(0)\|\right\} E\left\{\rho_2(Y(0))\varphi(Y(0))\right\},$$

where $\rho_2(Y(0)) = E\{|U(-\lambda)|\,|Y(0)\}$. We have thus verified the first inequality in the assertion and, therefore, completed the proof. ■

As an immediate consequence of the lemma we get our next one

Lemma 4 *Let (1) and (2) hold. Let the nonnegative Borel kernel K satisfy (5), (8). Let $0 < \lambda$. Then, for any positive $h(\xi)$, $h(\eta)$,*

$$\left|\operatorname{cov}\left[U(\xi-\lambda)K\left(\frac{y-Y(\xi)}{h(\xi)}\right), U(\eta-\lambda)K\left(\frac{y-Y(\eta)}{h(\eta)}\right)\right]\right|$$
$$\leq \begin{cases} 2\kappa^2\sigma_U^2, & \text{for all } \eta,\ \xi, \\ \dfrac{c_K}{h(\eta)}\|e^{A(\eta-\xi)}\| E\left\{\phi_\lambda(Y(0))K\left(\frac{y-Y(0)}{h(\xi)}\right)\right\}, & \text{for } \lambda+\xi \leq \eta \end{cases}$$

with some function ϕ_λ independent of K, ξ, η, such that $E|\phi_\lambda(Y(0)| < \infty$.

From the lemma, we obtain

Lemma 5 *Let (1) and (2) hold. Let the nonnegative Borel kernel K satisfy (5), (8). Let $0 < \lambda$. Then, for any positive function h satisfying (9), (10),*

$$\left|\int_0^t \int_0^t \operatorname{cov}\left[U(\xi-\lambda)\frac{1}{h(\xi)}K\left(\frac{y-Y(\xi)}{h(\xi)}\right), U(\eta-\lambda)\frac{1}{h(\eta)}K\left(\frac{y-Y(\eta)}{h(\eta)}\right)\right] d\xi d\eta\right|$$
$$\leq 4\kappa^2\sigma_U^2\lambda\frac{t}{h^2(t)} + 2dc_K\frac{1}{h^2(t)} \times \int_0^t \frac{1}{h(\tau)}E\left\{\phi_\lambda(Y(0))K\left(\frac{y-Y(0)}{h(\tau)}\right)\right\} d\tau,$$

where ϕ_λ is a Borel function independent of both t and K, such that $E|\phi_\lambda(Y(0))| < \infty$, and where $d = \int_0^\infty \|e^{At}\|dt$.

Proof. By virtue of Lemma 4, the integral in the assertion is bounded in absolute value by

$$2c_K \int_0^t \int_0^\eta \frac{1}{h^2(\eta)h(\xi)}\|e^{A(\eta-\xi)}\| \times E\left\{\phi_\lambda(Y(0))K\left(\frac{y-Y(0)}{h(\xi)}\right)\right\} d\xi d\eta + 2\kappa^2\sigma_U^2 \iint_{|\eta-\xi|<\lambda} d\xi d\eta.$$

Since (9) holds, the quantity is not greater than

$$\frac{2dc_K}{h^2(t)}\int_0^t \frac{1}{h(\tau)}E\left\{\phi_\lambda(Y(0))K\left(\frac{y-Y(0)}{h(\tau)}\right)\right\} d\tau + 4\kappa^2\sigma_U^2\lambda\frac{t}{h^2(t)}.$$

The lemma follows. ■

In a similar way, we can verify

Lemma 6 *Let (1) and (2) hold. Let the nonnegative Borel kernel K satisfy (5), (8). Let $0 < \lambda$. Then, for any positive function h satisfying (9) and (10),*

$$\left|\int_0^t \int_0^t \operatorname{cov}\left[U(\xi-\lambda)K\left(\frac{y-Y(\xi)}{h(\xi)}\right), U(\eta-\lambda)K\left(\frac{y-Y(\eta)}{h(\eta)}\right) d\xi d\eta\right]\right|$$

$$\leq 4\kappa^2\sigma_U^2\lambda t + \frac{2dc_K}{h(t)} \times \int_0^t E\left\{\phi_\lambda(Y(0))K\left(\frac{y-Y(0)}{h(\tau)}\right)\right\} d\tau,$$

where ϕ_λ is a Borel function independent of both t and K, such that $E|\phi_\lambda(Y(0))| < \infty$, and where $d = \int_0^\infty \|e^{At}\|dt$.

Lemma 7 *Let (1) and (2) hold. Then*

$$|\operatorname{cov}[U(\tau-\lambda)Y(\tau), U(t-\lambda)Y(t)]| \leq \begin{cases} c_1, & \text{for all } t,\ \tau, \\ c_2\|e^{A(t-\tau)}\|, & \text{for } \lambda+\tau \leq t, \end{cases}$$

where c_1 and c_2 are independent of τ, t.

Proof. The first part of the inequality is obvious while the second can be verified as that in Lemma 3. ■

General results

Results given below are of general character.

Lemma 8 *Let a random variable X have a probability density f. Let ρ be a Borel measurable function such that $E|\rho(X)| < \infty$. Let K be a nonnegative Borel measurable function satisfying (5), (6). If (7) holds with $\delta = 0$, then*

$$\frac{1}{h}E\left\{\rho(X)K\left(\frac{x-X}{h}\right)\right\} \to \rho(x)f(x)\int_{-\infty}^{\infty} K(y)dy \text{ as } h \to 0 \tag{28}$$

at every $x \in R$ at which both f and ρ are continuous. If (7) holds with $0 < \delta$, then (28) takes place also at almost every $x \in R$.

The lemma can be found in Wheeden and Zygmund (1977).

Lemma 9 *Let a random variable X have a probability density f. Let ρ be a Borel measurable function such that $E|\rho(X)| < \infty$. Let K be a nonnegative Borel measurable function satisfying (5), (6). If (7) holds with $\delta = 0$, then*

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t \frac{1}{h(\tau)}E\left\{\rho(X)K\left(\frac{x-X}{h(\tau)}\right)\right\} d\tau = \rho(x)f(x)\int_{-\infty}^{\infty} K(y)dy \tag{29}$$

at every $x \in R$ at which both f and ρ are continuous. If (7) holds with $0 < \delta$, then (29) takes place also at almost every $x \in R$.

Proof. The examined quantity equals $(1/t)\int_0^t \phi_\tau(x)d\tau$, where

$$\phi_\tau(x) = \frac{1}{h(\tau)}E\left\{\rho(X)K\left(\frac{x-X}{h(\tau)}\right)\right\}.$$

Applying Lemma 8, we complete the proof. ■

Lemma 10 *Let a random variable X have a probability density f. Let ρ be a Borel measurable function such that $E|\rho(X)| \leq \infty$. Let (10), and (11) hold. Let K be a nonnegative Borel measurable function satisfying (5), (6). If (7) holds with $\delta = 0$, then*

$$\lim_{t\to\infty} \frac{\int_0^t E\left\{\rho(X)K\left(\frac{x-X}{h(\tau)}\right)\right\} d\lambda}{\int_0^t h(\tau)d\tau} = \rho(x)f(x)\int_{-\infty}^{\infty} K(y)dy \tag{30}$$

at every $x \in R$ at which both f and ρ are continuous. If (7) holds with $0 < \delta$, then (30) takes place also at almost every $x \in R$.

Proof. To verify the lemma, it suffices to show that

$$\lim_{t\to\infty} \frac{\int_0^t \psi_\tau(x)h(\tau)d\tau}{\int_0^t h(\tau)d\tau} = 0,$$

almost every $x \in R$, where

$$\psi_\tau(x) = \frac{1}{h(\tau)}E\left\{[\rho(X) - \rho(x)f(x)]\,K\left(\frac{x-X}{h(\tau)}\right)\right\}.$$

Owing to (10) and Lemma 8, $\lim_{\tau\to\infty}\psi_\tau(x) = 0$, at almost every $x \in R$ or at every point at which both ρ, and f are continuous, respectively. Observing now that (11) holds, we complete the proof. ■

In lemmas below, $\alpha(t) \sim \beta(t)$ means that $\alpha(t)/\beta(t)$ has a nonzero limit as t tends to infinity.

Lemma 11 *Let a random variable X have a probability density f twice differentiable a point x. Let a Borel function ρ be also twice differentiable at the point. Let*

$\int_{-\infty}^{\infty} xK(x)dx = 0$, $\int_{-\infty}^{\infty} x^2K(x)dx < \infty$. *If* $h(t) \sim t^{-\alpha}$, $0 < \alpha$, *then*

$$\frac{1}{t}\int_0^t \frac{1}{h(\tau)} E\left\{\rho(X)K\left(\frac{x-X}{h(\tau)}\right)\right\} d\tau$$
$$-\rho(x)f(x)\int_{-\infty}^{\infty} K(y)dy = O(t^{-2\alpha})$$

as $t \to \infty$.

Proof. The quantity in the assertion equals

$$\frac{1}{t}\int_0^t \int_{-\infty}^{\infty} [\rho(x-h(\tau)y)f(x-h(\tau)y) - \rho(x)f(x)]K(y)dyd\tau.$$

Expanding $\rho(x-h(\tau)y)$, and $f(x-h(\tau)y)$ in a Taylor series at the point x, we complete the proof. ■

Lemma 12 *Let a random variable* X *have a probability density* f*, twice differentiable a point* x*. Let a Borel function* ρ *be also twice differentiable at the point. Let* $\int_{-\infty}^{\infty} xK(x)dx = 0$, $\int_{-\infty}^{\infty} x^2K(x)dx < \infty$. *If* $h(t) \sim t^{-\alpha}$, $0 < \alpha$, *then*

$$\frac{\int_0^t E\left\{\rho(X)K\left(\frac{x-X}{h(\tau)}\right)\right\} d\tau}{\int_0^t h(\tau)d\tau}$$
$$-\rho(x)f(x)\int_{-\infty}^{\infty} K(y)dy = O(t^{-2\alpha})$$

as $t \to \infty$.

References

[1] Bendat, S., 1990, *Nonlinear System Analysis and Identification* (New York: Wiley).

[2] Billings, S., A., 1980, Identification of non-linear systems-a survey, *IEE Proceedings*, **127**, 272-285.

[3] Billings, S. A., and Fakhouri, S. Y., 1978, Identification of a class of non-linear systems using correlation analysis, *IEE Proceedings*, **125**, 691-697.

[4] Brillinger, D., 1977, The identification of a particular non-linear time series system, *Biometrica*, **61**, 509-515.

[5] Greblicki, W., 1992, Nonparametric identification of Wiener systems, *IEEE Transactions on Information Theory*, **38**, 1487-1493.

[6] Greblicki, W., 1994, Nonparametric identification of Wiener systems by orthogonal series, *IEEE Transactions on Automatic Control*, **39**, 2077-2086.

[7] Greblicki, W., 1997, Nonparametric approach to Wiener system identification, *IEEE Transactions on Circuits and Systems – I: Fundamental Theory and Applications*, **44**, 538-545.

[8] Györfi, L., Härdle, W., Sarda, P., and Vieu, P., 1989, *Nonparametric Curve Estimation from Time Series*, (Berlin: Springer).

[9] Härdle, W., 1990, *Applied Nonparametric Regression*, (Cambridge: Cambridge University Press).

[10] Hasiewicz, Z., 1987, Identification of a linear system observed through zero-memory non-linearity, *International Journal of Systems Science*, **18**, 595-1607.

[11] Hunter, I. W., and Korenberg, M., J., 1986, The identification of non-linear biological systems: Wiener and Hammerstein cascade models, *Biological Cybernetics*, **55**, 135-144.

[12] Prakasa Rao, B. L. S., 1983, *Nonparametric Functional Estimation*, (Orlando: Academic Press).

[13] Westwick, D. T., and Kearney, R. E., 1992, A new algorithm for the identification of multiple input Wiener systems, *Biological Cybernetics*, **68**, 75-85.

[14] Wheeden, R. L., and Zygmund, A., 1977, *Measure and Integral*, (New York: Dekker).

[15] Wigren, T., 1994, Convergence analysis of recursive identification algorithms based on the non-linear Wiener model, *IEEE Transactions on Automatic Control*, **39**, 2191-2106.