

# Semiparametric Approach to Hammerstein System Identification

Włodzimierz Greblicki and Grzegorz Mzyk

\* The Institute of Computer Engineering, Control and Robotics,  
 Wrocław University of Technology, 50-370 Wrocław, Poland,  
 tel: +48 71 320 25 49, e-mail:  
 wlodzimierz.greblicki(grzegorz.mzyk)@pwr.wroc.pl

---

**Abstract:** Two semiparametric algorithms to recover a nonlinear characteristic in a Hammerstein system are proposed. Both are obtained by incorporating a parametric component into the kernel nonparametric algorithm. For small number of observations, their identification errors are smaller than that of the purely nonparametric algorithm. The same idea is also proposed for identification of linear dynamic component. Parametric instrumental variables estimate is elastically substituted by the nonparametric correlation-based method, when the number of observations tends to infinity.

---

## 1. INTRODUCTION

A problem of nonlinearity recovering in dynamic system with composite structure is extensively studied in the literature. Nonparametric identification algorithms are a proper tool if the *a priori* information about the system is small, less than parametric. It is however known that, due to small *a priori* information, nonparametric algorithms have rather large error for small number of observations. We show how this drawback can be eliminated or at least significantly reduced. We do it by incorporating a parametric component into a purely nonparametric algorithm. The resulting estimate, called semiparametric, has, thus, two terms. A parametric term is active for small number of observations while a nonparametric one dominates when the number is large.

## 2. THE IDENTIFICATION PROBLEM

A discrete-time Hammerstein system with input  $U_i$  and output  $Y_i$ , see Fig. 1, consists of a memoryless nonlinear subsystem followed by a linear dynamic one. The first part has a characteristic  $m$  which means that  $V_i = m(U_i)$  while the second one is described by the following state equation:

$$\begin{cases} X_{i+1} = AX_i + bV_i \\ W_i = c^T X_i, \end{cases}$$

where  $X_i$  is the state vector. Its output signal is disturbed by  $Z_i$  due to which  $Y_i = W_i + Z_i$  is output of the whole system.

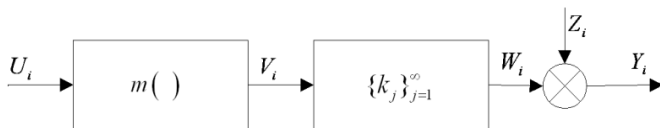


Fig. 1. The Hammerstein system.

We assume that  $\{\dots, U_{-1}, U_0, U_1, \dots\}$  is a stationary random process such that

$$EU_i^2 < \infty. \quad (1)$$

The probability density of  $U_i$ 's denoted by  $f$  is completely unknown. Since the characteristic  $m$  satisfies the following restriction:

$$|m(u)| \leq \alpha + \beta|u|, \quad (2)$$

where  $\alpha, \beta > 0$  are some unknown finite constants, we have  $EV_i^2 < \infty$ . Owing to this and the fact that, by assumption, the linear subsystem is stable,  $W_i$ 's are random variables and  $\{\dots, W_{-1}, W_0, W_1, \dots\}$  is a stationary random process. We assume that the disturbance  $\{\dots, Z_{-1}, Z_0, Z_1, \dots\}$  is also a stationary random process independent of the input signal and such that  $EZ_i = 0$ .

Both the nonlinear characteristic  $m$  and the impulse response  $\{k_j\}$  of the linear part are unknown and our goal is to recover them from input-output observations  $(U_0, Y_0), (U_1, Y_1), \dots, (U_n, Y_n)$  of the whole system. Since both  $U_i$  and  $Y_i$  are random variables,  $U_i$  by a straight assumption while  $Y_i$  due to (1), (2) and stability of the dynamic subsystem, our identification problem can be posed in probabilistic terms.

The class of all characteristics satisfying (2) is so wide that cannot be parameterized. Neither can be the family of all possible stable subsystems, since its order, i.e., the dimension of the state vector  $X_i$ , is unknown. Thus, the *a priori* information, and consequently our identification problem, about both subsystems is nonparametric.

Our additional restriction that  $\text{var } Z_i < \infty$  is rather standard. Not to confine generality of our further considerations but to simplify them, we assume, moreover, that

$$EU_i = 0, \quad (3)$$

$$f \text{ is an even function.} \quad (4)$$

$$m \text{ is an odd function,} \quad (5)$$

We need (3), (4), and (5) only to assure that  $EV_i = 0$ , and consequently  $EX_i = 0$ .

---

\* The paper is supported by the grant N N514 3160 33

### 3. NONPARAMETRIC KERNEL ALGORITHM

The motivation for nonparametric identification algorithms is the following fact:

$$E\{Y_{i+1}|U_i = u\} = c^T A E X_i + c^T b m(u) = \mu(u),$$

where  $\mu(u) = \alpha m(u)$ , and where  $\alpha = c^T b$  is unknown. Therefore to recover  $\mu(u)$ , i.e.  $m(u)$  up to some unknown constant  $\alpha$ , it suffices to estimate the regression function  $E\{Y_{i+1}|U_i = u\}$  in a nonparametric way. The kernel nonparametric algorithm has the following form (see Nadaraya [1964] and Watson [1964]):

$$\hat{\mu}(u) = \frac{\sum_{i=1}^n Y_{i+1} K\left(\frac{u - U_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{u - U_i}{h_n}\right)}. \quad (6)$$

The kernel  $K$  is such, that

$$\begin{aligned} \sup_u |K(u)| &< \infty, \\ \int |K(u)| du &< \infty, \\ uK(u) &\rightarrow 0 \text{ as } |u| \rightarrow \infty \end{aligned}$$

while the number sequence  $\{h_1, h_2, \dots\}$  is selected to satisfy the following restrictions:

$$\begin{aligned} h_n &\rightarrow 0 \text{ as } n \rightarrow \infty, \\ nh_n &\rightarrow \infty \text{ as } n \rightarrow \infty. \end{aligned}$$

It is known that the estimate is consistent, i.e., that

$$\hat{\mu}(u) \rightarrow \mu(u) \text{ as } n \rightarrow \infty \text{ in probability} \quad (7)$$

at every point  $u$  at which both  $m$  and  $f$  are continuous and, moreover,  $f(u) > 0$ , see Greblicki and Pawlak [1986], Greblicki and Pawlak [1987], and also Greblicki and Pawlak [1989], Greblicki and Pawlak [1994], Krzyżak [1990], Krzyżak and Partyka [1993]. Moreover, if both  $m$  and  $f$  are differentiable at a point  $u$  and  $h_n \sim n^{-1/2}$ ,

$$|\hat{\mu}(u) - \mu(u)| = O(n^{-1/4}) \text{ in probability.} \quad (8)$$

Numerical simulation unveils, however, some unwanted property of the estimate. For small number of observations, the Mean Integrated Squared Error

$$\text{MISE} = \int_{-\infty}^{\infty} E(\hat{\mu}(u) - \mu(u))^2 du$$

is large and diminishes rather slowly. To overcome this disadvantage, we modify (6) and present two new algorithms. For small  $n$ , both behave much better than (6), for large  $n$  are close to (6).

For small  $n$  the algorithms pay main attention to the linear part of in  $m$ , for large  $n$  they are targeted at the nonlinearity. By the linear part we mean the best linear approximation  $\rho u$  of  $\mu(u)$ , i.e.,  $\rho u$  with  $\rho$  minimizing  $E(\mu(U_0) - \rho U_0)^2$ , or, equivalently,  $E(Y_1 - \rho U_0)^2$ . The optimal approximation, i.e., the linear part of the nonlinearity is certainly  $\rho^* u$  with  $\rho^* = E\{Y_1 U_0\} / E U_0^2$ .

Each our algorithm has two components, one parametric and one nonparametric. The parametric one estimates the linear part while the other recovers the nonlinearity. The parametric one decreases MISE for small number of observations while the nonparametric component assures consistency.

Both algorithms are still nonparametric in nature. Nevertheless they are called semiparametric like all various nonparametric estimates with incorporated parametric components studied in the statistical literature, see, e.g., Härdle [1990]. In general, the parametric component does not change asymptotic properties, but is expected to improve the behavior of the estimate for small and moderate number of observations.

### 4. SEMIPARAMETRIC ALGORITHMS

Estimating  $\rho^*$  in the following way:

$$\hat{\rho} = \frac{\sum_{i=1}^n Y_{i+1} U_i}{\sum_{i=1}^n U_i^2},$$

we define

$$\hat{p}(u) = \hat{\rho} u \quad (9)$$

as an estimate of  $\rho^* u$ . The remaining part of  $\mu(u)$ , i.e.,  $q(u) = \mu(u) - \rho^* u$ , we estimate with the kernel estimate, see (6), in the following way:

$$\hat{q}(u) = \frac{\sum_{i=1}^n (Y_{i+1} - \hat{\rho} U_i) K\left(\frac{u - U_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{u - U_i}{h_n}\right)}.$$

Therefore, our first semiparametric estimate has the following form:

$$\hat{\mu}_1(u) = \hat{p}(u) + \hat{q}(u). \quad (10)$$

Its first component recovers  $\rho^* u$  in a parametric way, the other estimates the remaining part  $q(u) = \mu(u) - \rho^* u$  in a nonparametric way. Advantages of the first dominate for small  $n$ , of the other for large.

From (7), it follows that  $\hat{q}(u) \rightarrow q(u)$  as  $n \rightarrow \infty$  in probability at every point  $u$  at which both  $m$  and  $f$  are continuous and, moreover,  $f(u) > 0$ . Since  $\hat{p}(u) \rightarrow \rho^* u$  as  $n \rightarrow \infty$  in probability at every point  $u$ , our semiparametric estimate converges to the nonlinear characteristic, i.e.,

$$\hat{\mu}_1(u) \rightarrow \mu(u) \text{ as } n \rightarrow \infty \text{ in probability}$$

at every point  $u$  at which both  $m$  and  $f$  are continuous and  $f(u) > 0$ .

Our second semiparametric algorithm is of the following form:

$$\hat{\mu}_2(u) = \lambda_n \hat{p}(u) + (1 - \lambda_n) \hat{\mu}(u), \quad (11)$$

with  $\lambda_n \rightarrow 0$  as  $n \rightarrow \infty$ . It is just a linear combination of  $\hat{p}(u)$  recovering the linear part and  $\hat{\mu}(u)$  estimating  $\mu(u)$ . It is obvious that consistency of  $\hat{\mu}(u)$  implies consistency of  $\hat{\mu}_2(u)$ , which means that

$$\hat{\mu}_2(u) \rightarrow \mu(u) \text{ as } n \rightarrow \infty \text{ in probability}$$

at every point  $u$  at which both  $m$  and  $f$  are continuous and  $f(u) > 0$ . If the balance in the combination is proper, for small  $n$ , MISE should decrease at the rate typical for parametric inference.

It is clear that the parametric component  $\hat{p}(u)$  in (10) converges to  $\rho^* u$  in probability at the rate  $n^{-1/2}$  while the nonparametric one  $\hat{q}(u)$  to  $q(u)$  at the rate typical for nonparametric inference being slower than  $n^{-1/2}$ , see, e.g., (8). The same is with (11),  $\hat{p}(u)$  converges to  $\rho^* u$  faster than  $\hat{\mu}(u)$  to  $\mu(u)$ . Therefore for small  $n$ , where parametric components dominate nonparametric and errors of both (10) and (11) are expected to be smaller than that of (6).

## 5. IDENTIFICATION OF THE LINEAR DYNAMIC SUBSYSTEM

In general, the linear dynamic component with the impulse response  $\{k_j\}_{j=1}^\infty$ , i.e.,

$$W_i = \sum_{j=1}^{\infty} k_j V_{i-j}, \quad k_1 = 1 \quad (12)$$

cannot be described by the model, which includes finite number of parameters. Therefore, almost each purely parametric method is characterized by systematic approximation error. We will show that this error can be eliminated with the help of additional nonparametric term in the estimate. Let us consider the class of 2nd-order models with the transmittances

$$K(z, a_1, a_2) = \frac{z}{z^2 - a_1 z - a_2}, \quad a_1, a_2 \in R \quad (13)$$

and let the vector  $\mathbf{a}^* = (a_1^*, a_2^*)^T$  includes parameters of the best approximation of the true dynamic object (12) in the class (13), i.e.,

$$\mathbf{a}_1^*, \mathbf{a}_2^* = \arg \min_{a_1, a_2} E(W_i - a_1 W_{i-1} - a_2 W_{i-2})^2$$

and

$$\{\hat{k}_j^*\} = \mathcal{Z}^{-1}(K(z, \mathbf{a}_1^*, \mathbf{a}_2^*))$$

be respective impulse response of this model. Traditionally, parameters  $a_1^*$ , and  $a_2^*$  are identified by the instrumental variables (*iv*) method (see Wong and Polak [1967] and Stoica and Söderström [2002])

$$\hat{\mathbf{a}}^{(iv)} = (\hat{a}_1^{(iv)}, \hat{a}_2^{(iv)}) = (\Psi_n^T \Phi_n)^{-1} \Psi_n^T \mathbf{Y}_n, \quad (14)$$

where  $\Phi_n = (\phi_1, \phi_2, \dots, \phi_n)^T$ ,  $\phi_i = (Y_{i-1}, Y_{i-2})^T$ ,  $\mathbf{Y}_n = (Y_1, Y_2, \dots, Y_n)^T$  and  $\Psi_n = (\psi_1, \psi_2, \dots, \psi_n)^T$  is the matrix of appropriately selected instruments, such that the following two conditions are fulfilled:

**(C1)**  $\text{Plim} \Psi_n^T \Phi_n$  exists and is not singular,

**(C2)**  $\text{Plim} \Psi_n^T \tilde{\mathbf{Z}}_n = 0$ , where  $\tilde{\mathbf{Z}}_n = (\tilde{Z}_1, \tilde{Z}_2, \dots, \tilde{Z}_n)^T$  and  $\tilde{Z}_i = Z_i - a_1 Z_{i-1} - a_2 Z_{i-2}$ .

The impulse response can be estimated indirectly as

$$\{\hat{k}_j^{(iv)}\} = \mathcal{Z}^{-1} \left( K(z, \hat{a}_1^{(iv)}, \hat{a}_2^{(iv)}) \right). \quad (15)$$

Since for  $n \rightarrow \infty$  it holds that (see Wong and Polak [1967] and Hasiewicz and Mzyk [2009])

$$\hat{\mathbf{a}}^{(iv)} \rightarrow \mathbf{a}^* \quad \text{in probability}, \quad (16)$$

under continuity of the transform  $\mathcal{Z}^{-1}(\cdot)$  we get also

$$\hat{k}_j^{(iv)} \rightarrow k_j^* \quad \text{in probability} \quad (17)$$

for each  $j = 1, 2, \dots$ . Let us however emphasize that  $k_j^* \neq k_j$ , whenever the parametric model is not correct. For  $n$  large, the true elements  $\{k_j\}$  of the impulse response of the linear dynamics can be recovered successfully by the nonparametric way, using input-output cross-correlation analysis (Greblicki and Pawlak [1986]). The nonparametric estimate has the form

$$\hat{k}_j^{(corr)} = \frac{\hat{\chi}_j}{\hat{\chi}_1}, \quad \text{where } \hat{\chi}_j = \frac{1}{n-j} \sum_{i=1}^{n-j} Y_{i+j} U_i. \quad (18)$$

Although the estimate (18) is consistent, i.e.,

$$\hat{k}_j^{(corr)} \rightarrow k_j \quad \text{in probability},$$

as  $n \rightarrow \infty$ , it has huge variance-error if  $n - j$  is small. Therefore, analogously to (11) we propose to combine (15) with (18) as follows

$$\hat{k}_j^{(semi)} = \lambda_n \hat{k}_j^{(iv)} + (1 - \lambda_n) \hat{k}_j^{(corr)} \quad (19)$$

with  $\lambda_n \rightarrow 0$  as  $n \rightarrow \infty$ .

## 6. TECHNICAL ASPECTS

In the algorithms presented in Sections 3 and 4 the nonlinear characteristic is identified completely independently of the linear dynamic part. Similarly, in Section 5 the linear dynamics is identified independently of the static nonlinear block. Here we describe the ideas of the combined estimates (see Hasiewicz and Mzyk [2004]), in which the results of identification of one part of Hammerstein system support identification of the other.

### 6.1 Nonparametric instrumental variables

Most of results concerning strategies of generation of instrumental variables cannot be directly adopted from the linear system theory. The classical instruments (Wong and Polak [1967]) of the form

$$\psi_i^{(1)} = (\alpha V_{i-2}, \beta V_{i-3})^T \quad (20)$$

fulfill conditions **(C1)** and **(C2)** for each constants  $\alpha, \beta \neq 0$ , but they cannot be realized because the signal  $V_i$  is not accessible for measurement in Hammerstein system. On the other hand, the instruments based on the system input, e.g.,

$$\psi_i^{(2)} = (\alpha U_{i-2}, \beta U_{i-3})^T \quad (21)$$

generally does not assure fulfillment of **(C1)** and **(C2)**. One of the constructive solutions is

$$\psi_i^{(3)} = (\hat{V}_{i-2}, \hat{V}_{i-3})^T, \quad (22)$$

where  $\hat{V}_i$ 's ( $= \hat{\mu}(U_i)$ ) are nonparametric, kernel estimates of  $V_i$ 's. The consistency of the estimate (14) with the instrumental variables as in (22) has been strictly proved in Hasiewicz and Mzyk [2009].

### 6.2 Kernel regression estimation with filtered output

In kernel method (6), the Hammerstein system is treated in fact as a nonlinear static element

$$Y_i = m(U_{i-1}) + \xi_i$$

with the correlated output noise

$$\xi_i \triangleq \sum_{j>1} k_j m(U_{i-j}) + Z_i. \quad (23)$$

Since the component  $\sum_{j>1} k_j m(U_{i-j})$ , connected with dynamic part of the system usually dominates  $Z_i$ , and has harmful influence on the kernel estimate for small  $n$ , let us introduce the generalized version of (6)

$$\hat{\mu}^F(u) = \frac{\sum_{i=1}^n Y_{i+1}^F K\left(\frac{u - U_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{u - U_i}{h_n}\right)}, \quad (24)$$

where  $Y_i^F$  is filtered output (see Fig. 2) and  $F(z)$  is additional linear filter. Resulting system, with the input

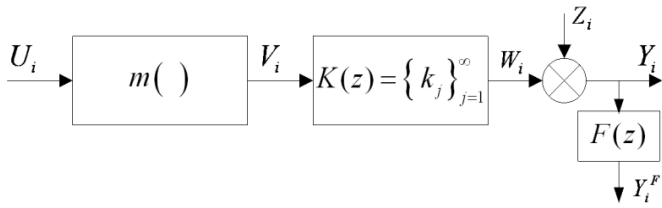


Fig. 2. The Hammerstein system with filtered output.

$U_i$  and the output  $Y_i^F$  obviously belongs also to the class of Hammerstein systems, and has the same static characteristic  $m$  as the original one. It was shown in Mzyk [2007] and Mzyk [2009] that appropriate selection of the filter  $F(z)$ , e.g.  $F(z) = K^{-1}(z)$  for invertible dynamics, allows to decrease the variance of (24) in comparison with (6). Since  $K(z)$  is unknown, we recommend to apply the inverse of its AR(2) approximation, obtained in parametric way, i.e.,

$$Y_i^F = Y_i - \hat{a}_1^{(iv)} Y_{i-1} - \hat{a}_2^{(iv)} Y_{i-2}.$$

## 7. SIMULATION EXAMPLES

### 7.1 Recovering of the nonlinear characteristic

In the simulation example,  $U_i$  is distributed uniformly on  $[-1, 1]$ ,  $Z_i$  has a Gaussian distribution with zero mean and variance 0.25. The linear subsystem is described by the following equations:  $X_{i+1} = aX_i + V_i$ ,  $W_i = X_i$ , with  $a = 0.6$ . Moreover, see Fig. 3,

$$m(u) = \begin{cases} u + 0.5 \sin \frac{10\pi u^2}{1+u^2} - 0.5, & \text{for } u \leq 0 \\ u + 0.5 \sin \frac{10\pi u^2}{1+u^2} + 0.5, & \text{otherwise.} \end{cases}$$

The kernel is a rectangular function equal to 1 for  $|u| \leq 1$

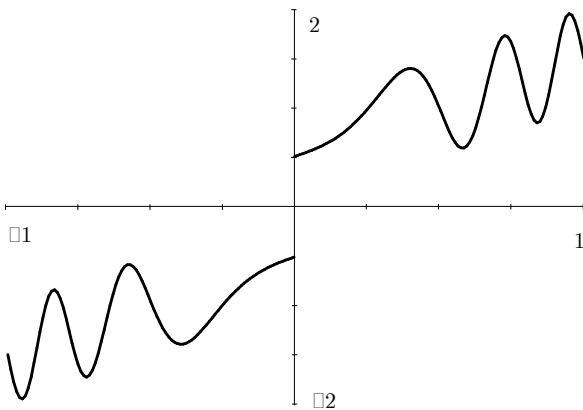


Fig. 3. The nonlinear characteristic  $m$ .

and 0 otherwise. Moreover, in estimate (11),

$$\lambda_n = \frac{1}{1 + (n/200)^6}.$$

The quality of each examined algorithm is measured with MISE, e.g., for (6)  $\text{MISE} = \int_{-1}^1 E(\hat{\mu}(u) - \mu(u))^2 du$ .

It is well known that purely nonparametric estimate (6) is sensitive to improper selection of  $h_n$ . This phenomenon is shown in Fig. 4;  $h_n$  is replaced with  $h$ . It is obvious that

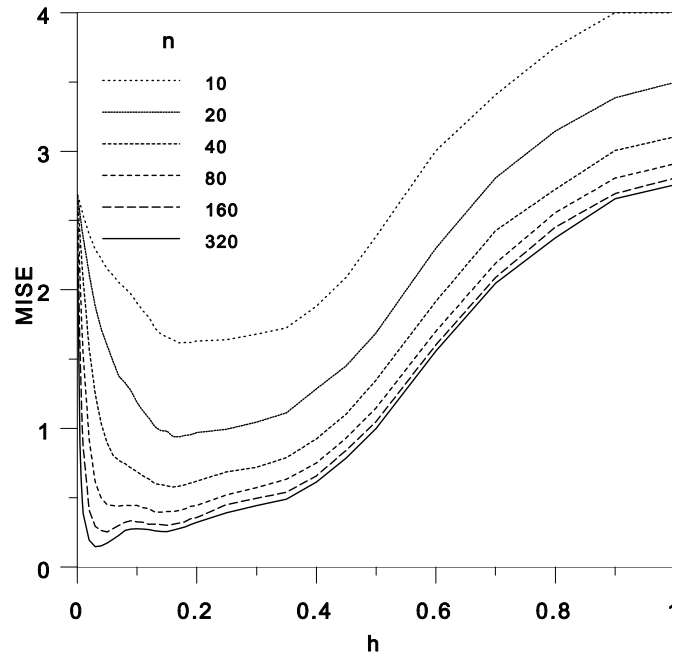


Fig. 4. MISE for nonparametric estimate (6).

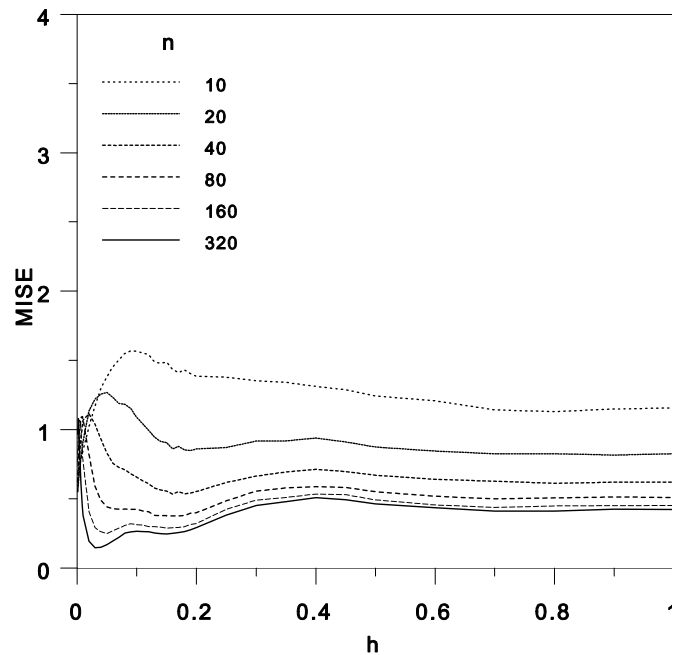


Fig. 5. MISE for semiparametric estimate (10).

too small  $h$  must be avoided. This disadvantageous effect is much weaker in semiparametric algorithms, see Figs. 5 and 6.

To demonstrate what can be achieved by our algorithms,  $h_n$  has been selected optimally in nonparametric (6) as well as in semiparametric (10) and (11). Results are shown in Fig. 7. To get more comprehensive view, we have shown also the error for purely parametric algorithm (9). Superiority of semiparametric algorithms for  $n$  smaller than 80 is obvious.

In real situations selecting optimal  $h_n$  is clearly impossible. Therefore, for  $h_n$  selected rather arbitrarily as  $0.8n^{-0.3}$ ,

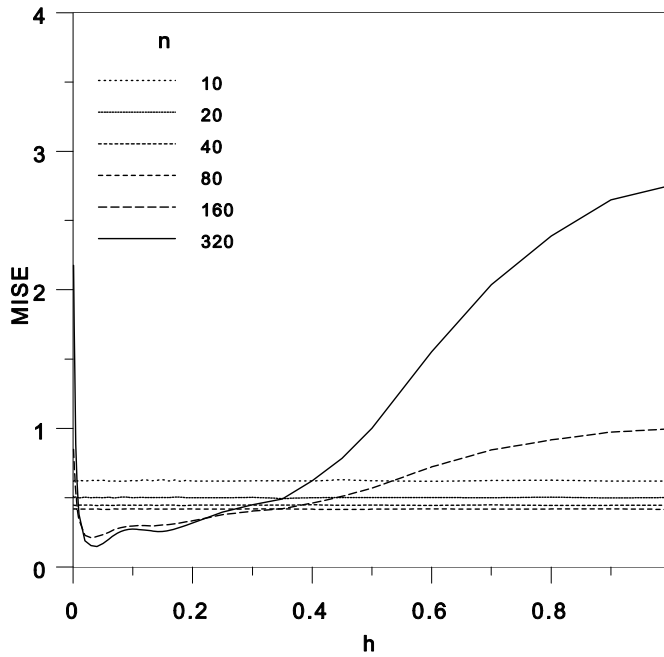


Fig. 6. MISE for semiparametric estimate (11).

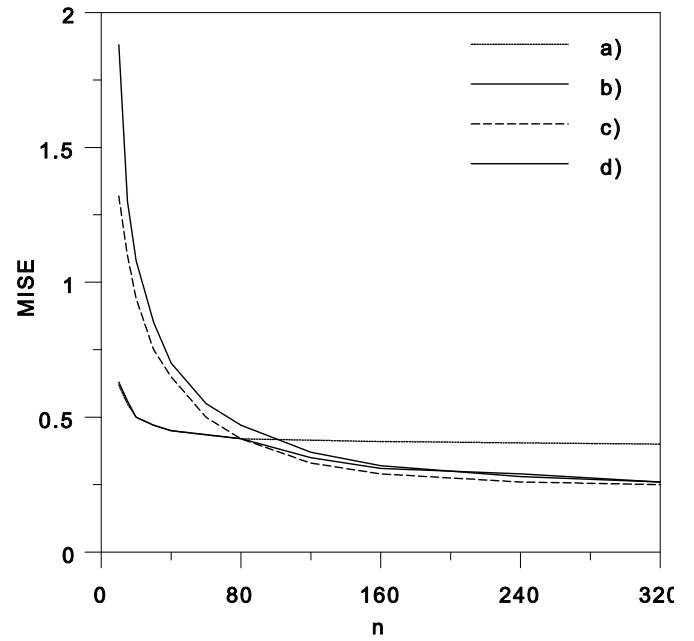


Fig. 8. MISE versus  $n$ ;  $h_n = 0.8n^{-0.3}$ ; a) – (9), b) – (6), c) – (10), d) – (11).

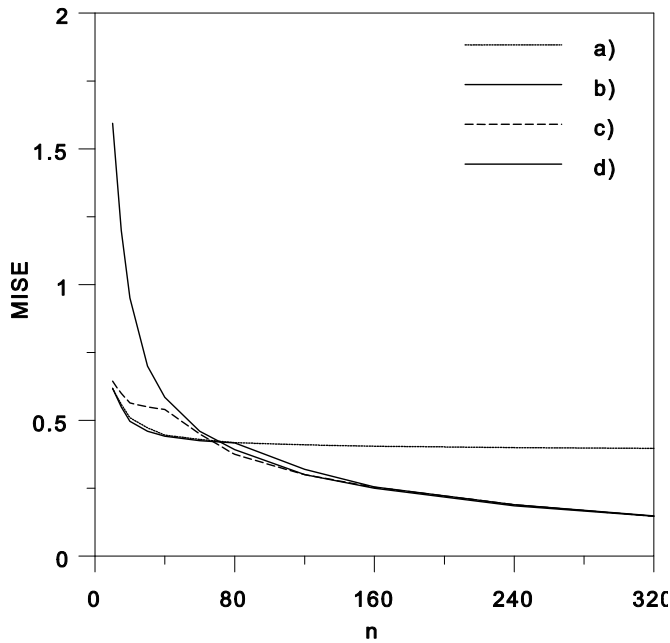


Fig. 7. MISE versus  $n$ ;  $h$  optimal; a) – (9), b) – (6), c) – (10), d) – (11).

results are shown in Fig. 8. Semiparametric estimates are also superior.

### 7.2 Nonlinearity recovering with data pre-filtering

For the system simulated in Section 7.1 we compared the estimates (6) and (24) for various  $n$ . We used the rectangular kernel and set  $h_n = 0.8n^{-0.3}$  in both algorithms. The results are presented in Table 1. Thanks to data pre-filtering the MISE has been reduced significantly.

Table 1. MISE of the kernel estimate with filtration and without filtration.

$N$	$MISE_{\hat{\mu}^F}(u)$	$MISE_{\hat{\mu}}(u)$
100	0.08	0.54
300	0.05	0.32
500	0.03	0.16
1000	0.02	0.11
3000	0.01	0.06
5000	0.00	0.03

### 7.3 Estimation of the impulse response

In the simulation presented in Section 7.1 we replaced the linear dynamic component with the FIR filter, including four nonzero elements of the impulse response, i.e., we set  $k_1 = 1$ ,  $k_2 = 0.75$ ,  $k_3 = 0.5$ ,  $k_4 = 0.25$ , and  $k_j = 0$  for  $j > 4$ . Since  $\rho^* \neq 0$ , one can show that

$$\text{cov}(U_i, V_i) \neq 0$$

and consequently the instruments generated according the rule

$$\psi_i = (U_{i-2}, U_{i-3})^T$$

fulfill conditions **(C1)** and **(C2)**. The estimation error of both  $\hat{k}_j^{(iv)}$  and  $\hat{k}_j^{(corr)}$  has been computed as follows

$$ERR\{\hat{k}_j\} = \frac{\sum_{j=1}^{10} (\hat{k}_j - k_j)^2}{\sum_{j=1}^{10} k_j^2} \cdot 100[\%]$$

Obviously, an FIR object does not belong to the class of AR(2) models defined in (13), which allows to demonstrate the asymptotic approximation error of the parametric estimate and consistency of the nonparametric one (see Fig. 9). However, for  $n$  small the parametric approach can be better, thanks to the rich *a priori* knowledge. It is because the class of admissible structures of the linear dynamics is narrowed at the beginning, and only two parameters are estimated. Appropriately selection of  $\lambda_n$  in semiparametric approach (19), e.g.,

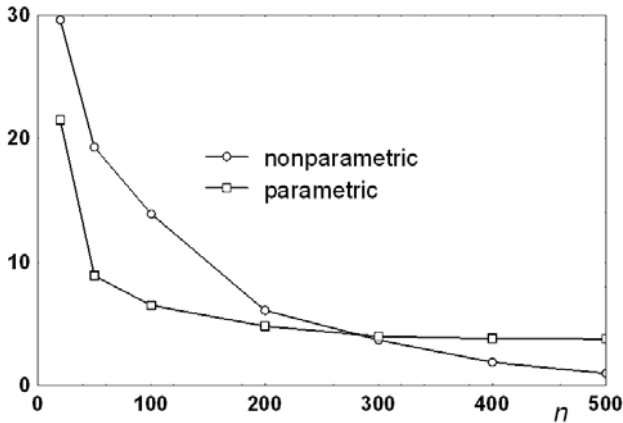


Fig. 9. Estimation error of the linear dynamics.

$$\lambda_n = \begin{cases} 1, & \text{as } n < 300 \\ 0, & \text{as } n \geq 300 \end{cases}$$

allows the estimate  $\hat{k}_j^{(semi)}$  to take advantages of both  $\hat{k}_j^{(iv)}$  and  $\hat{k}_j^{(corr)}$ .

## 8. CONCLUSIONS

Simulation examples confirm expectations resulting from theoretical considerations. For small number of observations, semiparametric algorithms behave better than purely nonparametric (6) and (18). Moreover, semiparametric algorithms are much less sensitive to not optimal selecting  $h_n$ . In the light of this, the semiparametric methodology is worth further studies.

## REFERENCES

- W. Greblicki and M. Pawlak. Identification of discrete Hammerstein systems using kernel regression estimates. *IEEE Transactions on Automatic Control*, AC-31(1):74–77, January 1986.
- W. Greblicki and M. Pawlak. Hammerstein system identification by non-parametric regression estimation. *International Journal of Control*, 45(1):343–354, 1987.
- W. Greblicki and M. Pawlak. Nonparametric identification of Hammerstein systems. *IEEE Transactions on Information Theory*, IT-35(2):409–418, March 1989.
- W. Greblicki and M. Pawlak. Cascade non-linear system identification by a non-parametric method. *International Journal of Systems Science*, 25(1):129–153, 1994.
- W. Greblicki, M. Pawlak, *Nonparametric system identification*, Cambridge University Press, 2008.
- Z. Hasiewicz, G. Mzyk, Combined parametric-nonparametric identification of Hammerstein systems, *IEEE Transactions on Automatic Control*, vol. 49, pp. 1370–1376, 2004.
- Z. Hasiewicz, G. Mzyk, Hammerstein system identification by non-parametric instrumental variables, *International Journal of Control*, vol. 82, No. 3, pp. 440–455, 2009.
- Z. Hasiewicz, P. Śliwiński, G. Mzyk, Nonlinear system identification under various prior knowledge, *Proceedings of the 17th World Congress The IFAC*, pp. 7849–7858, Seoul, Korea, 2008.
- W. Härdle. *Applied Nonparametric Regression*. Cambridge University Press, Cambridge, 1990.
- N. L. Hjort, I. K. Glad, Nonparametric density estimation with a parametric start, *Annals of Statistics*, vol. 23, No. 3, pp. 882–904, 1995.
- A. Krzyżak. On estimation of a class of nonlinear systems by the kernel regression estimate. *IEEE Transactions on Information Theory*, 36(1):141–152, February 1990.
- A. Krzyżak and M. A. Partyka. On identification of block orientated systems by non-parametric techniques. *International Journal of Systems Science*, 24(6):1049–1066, 1993.
- G. Mzyk, Generalized kernel regression estimate for the identification of Hammerstein systems, *International Journal of Applied Mathematics and Computer Science*, vol. 17, No. 2, pp. 189–197, 2007.
- G. Mzyk, Data pre-filtering for nonparametric identification of Hammerstein system, submitted to the *IEEE Signal Processing Letters*, 2009.
- E. A. Nadaraya, On estimating regression, *Theory of Probability and its Applications*, vol. 9, pp. 141–142. 1964.
- P. Stoica, T. Söderström, Instrumental variable methods for system identification, *Circuits Systems Signal Processing*, vol. 21, No. 1, pp. 1–9, 2002.
- P. Śliwiński, J. Rozenblit, M. W. Marcellin, R. Klempous, Wavelet amendment of polynomial models in nonlinear system identification, to appear in: *IEEE Transactions on Automatic Control*, 2009.
- G. S. Watson, Smooth regression analysis, *Sankhya, Series A*, vol. 26, pp. 355–372, 1964.
- K. Wong, E. Polak, Identification of linear discrete time systems using the instrumental variable method, *IEEE Transactions on Automatic Control*, vol. AC-12, No. 6, pp. 707–718, 1967.