

GENERALIZED KERNEL REGRESSION ESTIMATE FOR THE IDENTIFICATION OF HAMMERSTEIN SYSTEMS

GRZEGORZ MZYK

Institute of Computer Engineering, Control and Robotics
 Wrocław University of Technology
 Janiszewskiego 11/17, 50–372 Wrocław, Poland
 e-mail: grzegorz.mzyk@pwr.wroc.pl

A modified version of the classical kernel nonparametric identification algorithm for nonlinearity recovering in a Hammerstein system under the existence of random noise is proposed. The assumptions imposed on the unknown characteristic are weak. The generalized kernel method proposed in the paper provides more accurate results in comparison with the classical kernel nonparametric estimate, regardless of the number of measurements. The convergence in probability of the proposed estimate to the unknown characteristic is proved and the question of the convergence rate is discussed. Illustrative simulation examples are included.

Keywords: Hammerstein system, nonparametric regression, kernel estimation

1. Introduction

The problem of elaborating a universal approach to modeling stochastic nonlinear systems has been intensively discussed in the literature in recent years. In particular, the idea of a block-oriented system representation as an interconnection of nonlinear static elements and linear dynamics (Billings and Fakhouri, 1982; Greblicki, 2001; Janczak, 1999) has been developed and widely accepted in the system identification community. The Hammerstein system (Fig. 1) is the simplest structure of this kind and hence it is mostly considered in the system identification literature (see, e.g., (Giannakis and Serpedin, 2001; Haber and Keviczky, 1999) for the bibliography). It is built of a static nonlinearity $\mu(\cdot)$ and linear dynamics, with the impulse response $\{\gamma_i\}_{i=0}^{\infty}$, connected in a cascade and described by the following set of equations: $y_k = v_k + z_k$, $v_k = \sum_{i=0}^{\infty} \gamma_i w_{k-i}$, $w_k = \mu(u_k)$ or, equivalently,

$$y_k = \sum_{i=0}^{\infty} \gamma_i \mu(u_{k-i}) + z_k, \quad (1)$$

where u_k and y_k denote the system input and output at time k , respectively, and z_k is the output noise. Since the structure shown in Fig. 1 has good approximation capabilities regarding many real processes, the identification problem of Hammerstein systems has fundamental

meaning in practice, particularly in domains such as automatic control (Haber and Keviczky, 1999), signal processing (Giunta *et al.*, 1991), telecommunications (Jang and Kim, 1994), biocybernetics, artificial neural networks, the modeling of distillation, fermentation (Gomez and Basualdo, 2000) and heat exchange processes (Haber and Zeirfuss, 1988), geology (Zhang and Bai, 1996), etc.

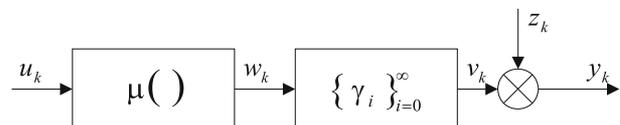


Fig. 1. Hammerstein system.

The Hammerstein system was first considered in the system identification framework by Narendra and Gallman (1966) and since then a large number of research papers have appeared in the literature. Basically, they differ in the requirements for nonlinear characteristics and the noise process. One can specify two mutually complementing kinds of approaches, leading to parametric and nonparametric identification methods. The first of them (Bai and Li, 2004; Chang and Luus, 1971; Chen, 2005; Haber and Keviczky, 1999; Ljung, 1987; Söderström and Stoica, 1982; Söderström and Stoica, 1989; Vörös, 1999) provides estimates with fast convergence and the model is given in closed form, but this requires rich *a-priori* knowl-

edge about the identified system. The system characteristics must have a known analytic form (e.g., polynomial) with a finite number of unknown parameters. Moreover, in most papers only white output noise is admissible. If the true system characteristic does not belong to the class of assumed parametric models, then a systematic approximation error appears. Recently, a number of methods with various applicability conditions have been offered, e.g., frequency domain identification or inverse OBF modeling (see, e.g., (Bai, 2003; Zhu, 2000; Latawiec, 2004)).

The nonparametric approach to block-oriented system identification was proposed in the 1980s by Greblicki, and it is based on regression function estimation by kernel methods (Greblicki and Pawlak, 1986; Greblicki and Pawlak, 1989; Greblicki and Pawlak, 1994; Härdle, 1990; Wand and Jones, 1995) or orthogonal series methods (Greblicki, 1989; Van den Hof *et al.*, 1995). The algorithms involve only a learning sequence and are free of the risk of improper *a-priori* knowledge. They are recommended when the nonlinear characteristic cannot be expressed in a closed form. The cost paid for it is the fact that the rate of convergence is a bit slower. The convergence conditions of the estimates were relaxed in (Greblicki *et al.*, 1984; Krzyżak, 1990; Krzyżak *et al.*, 2001), where the existence of the input probability density function is not required. In recent papers (Hasiewicz *et al.*, 2005; Pawlak and Hasiewicz, 1998), wavelet bases have been used. In (Hasiewicz and Mzyk, 2004a; Hasiewicz and Mzyk, 2004b), it was shown that nonparametric methods enable full decentralization of the block-oriented system identification task, i.e., independent parameter identification of a static nonlinearity and linear dynamics in a completely decomposed manner.

As the linear dynamics in a Hammerstein system can be simply identified, e.g., by the standard correlation method (Greblicki and Pawlak, 1986), we concentrate in the paper on effective recovering of the nonlinearity $\mu(\cdot)$. In Section 2, the problem is formulated in detail, and the standard kernel approach is reminded. The idea of averaging kernel estimates for various time-lags is introduced in Section 3 and analyzed in Section 4. Finally, the most important special cases are considered and some illustrative simulation examples are given.

2. Problem Statement

2.1. Assumptions. We make the following assumptions:

- (A1) The characteristic $\mu(u)$ of a static subsystem fulfils the condition $|\mu(u)| \leq p_1 + p_2 |u|$, where p_1 , and p_2 are some finite unknown constants.
- (A2) The linear dynamics with the unknown impulse response $\{\gamma_i\}_{i=0}^\infty$ is stable, i.e., $\sum_{i=0}^\infty |\gamma_i| < \infty$.
- (A3) The input $\{u_k\}$ and the noise $\{z_k\}$ are random, mutually independent i.i.d. processes, and $Ez_k = 0$.

There exists a probability density of the input $\vartheta_u(u_k)$.

Without any loss of generality and for the clarity of exposition, we also assume that $\mu(0)$ (for a discussion, see, e.g., (Hasiewicz and Mzyk, 2004a)).

2.2. Classical Kernel Estimate. The following dependence between the regression functions $R_c(u)$ and the true system characteristic $\mu(\cdot)$ is of paramount importance for the identification routine of the Hammerstein system developed in the paper:

$$\begin{aligned} R_c(u) &= E\{y_k | u_{k-c} = u\} \\ &= E\{\gamma_c \mu(u_{k-c}) + \sum_{i \neq c} \gamma_i \mu(u_{k-i}) + z_k | u_{k-c} = u\} \\ &= \gamma_c \mu(u) + \delta_c, \end{aligned} \tag{2}$$

where $\delta_c = E\mu(u) \sum_{i \neq c} \gamma_i$, and c is any time-lag between the input and output. Due to (2), the characteristic $\mu(\cdot)$ may be estimated only up to some scaling and shifting constants γ_c and δ_c , respectively, provided that we can estimate $R_c(u)$. This feature is, however, independent of the identification method and it is a simple consequence of the inaccessibility of the interior signal $\{w_k\}$ for a direct measurement (for a discussion, see (Hasiewicz and Mzyk, 2004a)).

In standard nonparametric methods, the Hammerstein system is treated in fact as a nonlinear static element corrupted by correlated noise. Indeed, one can specify three components of the output,

$$y_k = \underbrace{\gamma_c \mu(u_{k-c})}_{\text{useful}} + \underbrace{\sum_{i \neq c} \gamma_i \mu(u_{k-i})}_{\text{"system noise"}} + \underbrace{z_k}_{\text{output noise}}. \tag{3}$$

In such a description only the c -th term of the sum in (1) is privileged (see Fig. 2), which means that most of the signal y_k is in a sense ignored, although the "system noise"

$$\xi_k \triangleq \sum_{i \neq c} \gamma_i \mu(u_{k-i}) \tag{4}$$

also depends on the identified function $\mu(\cdot)$.

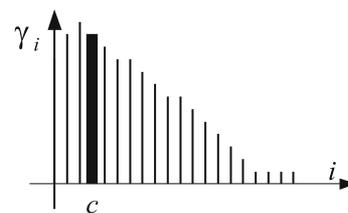


Fig. 2. Impulse response of the linear dynamics.

The classical kernel estimate for the identification of Hammerstein systems was introduced in the 1980s by

Greblicki and Pawlak. For the model (3), it has the following form:

$$\widehat{\mu}_{c,M}(u) = \widehat{R}_{c,M}(u) - \widehat{R}_{c,M}(0), \quad (5a)$$

where

$$\widehat{R}_{c,M}(u) = \frac{\sum_{k=1}^M \left\{ y_{k+c} K \left(\frac{u_k - u}{h(M)} \right) \right\}}{\sum_{k=1}^M K \left(\frac{u_k - u}{h(M)} \right)}, \quad (5b)$$

$K(\cdot)$ being a kernel function satisfying the conditions

$$K(x) > 0, \quad \int_{-\infty}^{\infty} K(x) dx < \infty, \quad K(x) = K(-x), \quad (6)$$

and $h(M)$ a proper bandwidth parameter. Standard examples are $K(x) = I_{[-0.5,0.5]}(x)$, $(1 - |x|)I_{[-1,1]}(x)$ or $(1/\sqrt{2\pi}) e^{-x^2/2}$ and $h(M) = \text{const} \cdot M^{-\alpha}$ with a positive constant and $0 < \alpha < 1$ (Wand and Jones, 1995). Observe that the estimate (5) requires additional c measurements $y_{M+1}, y_{M+2}, \dots, y_{M+c}$, but it has no influence on the limit properties, as the lag c is fixed. In (Greblicki and Pawlak, 1986), it was proved that if

$$h(M) \rightarrow 0 \quad \text{and} \quad Mh(M) \rightarrow \infty \quad (7)$$

as $M \rightarrow \infty$, then

$$\widehat{\mu}_{c,M}(u) \rightarrow \gamma_c \mu(u) \quad \text{in probability, as } M \rightarrow \infty \quad (8)$$

for each continuity point u of the characteristic $\mu(\cdot)$ and the input probability density function $\vartheta_u(\cdot)$. If, moreover, $\mu(\cdot)$ and $\vartheta_u(\cdot)$ are twice differentiable in u , then for $h(M) = O(M^{-1/5})$ we have

$$|\widehat{\mu}_{c,M}(u) - \mu(u)| = O(M^{-2/5}) \quad \text{in probability.} \quad (9)$$

The rate in (9) is only asymptotic, i.e., valid for large values of M . In practice (for $M \ll \infty$), the value of the mean square error (MSE):

$$\begin{aligned} MSE \widehat{\mu}_{c,M}(u) &= E(\widehat{\mu}_{c,M}(u) - \mu(u))^2 \\ &= \text{bias}^2 \widehat{\mu}_{c,M}(u) + \text{var} \widehat{\mu}_{c,M}(u) \end{aligned} \quad (10)$$

determines the quality of the method. We emphasize that even for the noise-free output case ($z_k \equiv 0$) the variance $\text{var} \widehat{\mu}_{c,M}(u)$ in (10) can be large, as it depends on the component ξ_k (see (4) and (3)). The main purpose of the paper is to modify the kernel estimate (5) so as to obtain lower variance using the same set of observations. First attempts to solve this problem were based on the estimation of the d -dimensional regression $\overline{R}_{c_1, c_2, \dots, c_d}(u) = E\{y_k | u_{k-c_1} = u, u_{k-c_2} = u, \dots, u_{k-c_d} = u\}$ instead of the simple regression $R_c(u)$. Unfortunately, the application of the estimate of $\overline{R}_{c_1, c_2, \dots, c_d}(u)$ leads to a very

restrictive convergence condition $Mh^d(M) \rightarrow \infty$. As we show in Sections 3 and 4, this disadvantage may be avoided by estimating simple (1-dimensional) regressions for different time-lags c_1, c_2, \dots, c_d . We limit ourselves to the estimates involving two lags ($d = 2$), since the presented methodology can be simply generalized for $d > 2$.

3. Improved Kernel Estimate

Let us define the average of two simple regression functions,

$$\begin{aligned} R_{c_1, c_2}(u) &\triangleq \frac{1}{2} [R_{c_1}(u) + R_{c_2}(u)] \\ &= E \left\{ \frac{1}{2} [y_{k+c_1} + y_{k+c_2}] \mid u_k = u \right\} \\ &= \frac{\gamma_{c_1} + \gamma_{c_2}}{2} \mu(u) + \frac{\delta_{c_1} + \delta_{c_2}}{2} \end{aligned} \quad (11)$$

and introduce the averaged kernel estimate of the following form (cf. (5)):

$$\begin{aligned} \widehat{\mu}_{c_1, c_2, M}(u) &= \frac{1}{2} [\widehat{\mu}_{c_1, M}(u) + \widehat{\mu}_{c_2, M}(u)] \\ &= \frac{1}{2} [\widehat{R}_{c_1, M}(u) - \widehat{R}_{c_1, M}(0) + \widehat{R}_{c_2, M}(u) - \widehat{R}_{c_2, M}(0)] \\ &= \widehat{R}_{c_1, c_2, M}(u) - \widehat{R}_{c_1, c_2, M}(0), \end{aligned} \quad (12)$$

where

$$\begin{aligned} \widehat{R}_{c_1, c_2, M}(u) &= \frac{1}{2} [\widehat{R}_{c_1, M}(u) + \widehat{R}_{c_2, M}(u)] \\ &= \frac{\sum_{k=1}^M \left\{ \frac{y_{k+c_1} + y_{k+c_2}}{2} K \left(\frac{u_k - u}{h(M)} \right) \right\}}{\sum_{k=1}^M K \left(\frac{u_k - u}{h(M)} \right)}. \end{aligned} \quad (13)$$

As we can conclude from (13) and Fig. 3, the method is in fact based on implementing an additional output preprocessing block using the output averaging linear fil-

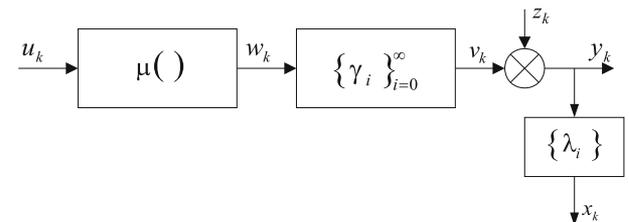


Fig. 3. Graphical interpretation of the method.

ter $x_k = \sum_{i=0}^{\infty} \lambda_i y_{k-i}$ with the impulse response

$$\lambda_i = \begin{cases} \frac{1}{2} & \text{if } i = c_1 \text{ or } i = c_2, \\ 0 & \text{otherwise,} \end{cases}$$

i.e., another ‘fictitious’ system with the input u_k and the output x_k is actually considered. Obviously, the system with the input u_k and the output x_k also belongs to the Hammerstein system class, has the same nonlinear characteristic $\mu(u)$ as the original one, and a proper kernel estimate of the adequate regression function is

$$\widehat{R}_M(u) = \frac{\sum_{k=1}^M \left\{ x_k K \left(\frac{u_k - u}{h(M)} \right) \right\}}{\sum_{k=1}^M K \left(\frac{u_k - u}{h(M)} \right)}.$$

4. Statistical Properties

4.1. Limit Properties. Since the generalized estimate (12) is a linear combination of the simple estimates (5), it obviously has identical convergence properties (see (Greblicki and Pawlak, 1994) and Remarks 1 to 3).

Remark 1. If $h(M) \rightarrow 0$ and $Mh(M) \rightarrow \infty$ as $M \rightarrow \infty$, then $\widehat{\mu}_{c_1, c_2, M}(u) \rightarrow \frac{1}{2}(\gamma_{c_1} + \gamma_{c_2})\mu(u)$ in probability as $M \rightarrow \infty$ for each continuity point u of the characteristic $\mu(\cdot)$ and the input probability density function $\vartheta_u(\cdot)$.

The consistency of the classical kernel estimate $\widehat{\mu}_{c, M}(u)$ (see (5)) was also proved in the case when the input $\{u_k\}$ has a discrete distribution at the points u for which $P(u_k = u) > 0$, under mild restrictions imposed on $h(M)$. This property is obviously inherited by the proposed estimate $\widehat{\mu}_{c_1, c_2, M}(u)$.

Remark 2. If

$$|\widehat{\mu}_{c, M}(u) - \gamma_c \mu(u)| = O(M^{-\tau})$$

in probability, then

$$|\widehat{\mu}_{c_1, c_2, M}(u) - \frac{\gamma_{c_1} + \gamma_{c_2}}{2} \mu(u)| = O(M^{-\tau})$$

in probability.

Remark 3. If $h(M) = O(M^{-1/5})$, then $\text{var } \widehat{\mu}_{c, M}(u) = O(M^{-4/5})$ and $\text{var } \widehat{\mu}_{c_1, c_2, M}(u) = O(M^{-4/5})$.

Therefore, the proposed estimate has asymptotically the same rate of convergence, which means that asymptotically (as $M \rightarrow \infty$) it provides results of comparable accuracy. Nevertheless, in practice we have at our disposal a fixed finite number of observations and we wish to obtain the smallest possible estimation error. The evaluation of the degree of variance reduction in a general case is difficult. It can be performed in some popular special cases (see Section 6). Simulation results show that the percentage reduction in the variance compared with the classical kernel estimate does not significantly depend on M (the number of data). The superiority of (12) over (5) in the

sense of the absolute error is obviously more visible for small and moderate M .

In Section 4.2 we construct an upper bound for $\text{var } \widehat{\mu}_{c_1, c_2, M}(u)$, and then optimize it with respect to c_1 and c_2 . We emphasize that the class of the analyzed estimates $\{\widehat{\mu}_{c_1, c_2, M}(u)\}$ also includes the subset $\{\widehat{\mu}_{c, M}(u)\}$, as $c_1 = c_2$ reduces the generalized kernel estimate to the classical one.

4.2. Optimal Choice of c_1 and c_2 . As has been pointed out (see (11), (12) and Remark 1) $\widehat{\mu}_{c_1, c_2, M}(u)$ is (merely) an estimate of $\frac{1}{2}(\gamma_{c_1} + \gamma_{c_2})\mu(u)$, and hence its variance obviously does depend on the scale factor $\frac{1}{2}(\gamma_{c_1} + \gamma_{c_2})$. As a consequence, the variance of $\widehat{\mu}_{c_1, c_2, M}(u)$ does not determine unambiguously the quality of the identification algorithm. Thus the scale $\frac{1}{2}(\gamma_{c_1} + \gamma_{c_2})$ should also be taken into consideration during comparisons of the efficiency of the estimates $\widehat{\mu}_{c_1, c_2, M}(u)$ for various c_1 and c_2 . To this end we introduce the following definition:

Definition 1. Two estimates $\widehat{\theta}_1$ and $\widehat{\theta}_2$ of scaled θ^* (i.e., $E\widehat{\theta}_1 = a\theta^*$ and $E\widehat{\theta}_2 = b\theta^*$) are said to have the *same relative variance* if

$$\frac{\text{var } \widehat{\theta}_1}{\text{var } \widehat{\theta}_2} = \frac{a^2}{b^2}.$$

The minimization of the relative variances of the estimates $\widehat{\mu}_{c_1, c_2, M}(u)$ with respect to c_1 and c_2 is equivalent to the minimization of the variances (in the classical sense) of the normalized estimates

$$\overline{\mu}_{c_1, c_2, M}(u) = \frac{2}{\gamma_{c_1} + \gamma_{c_2}} \widehat{\mu}_{c_1, c_2, M}(u), \quad (14)$$

where γ_{c_1} and γ_{c_2} are unknown. Introducing the auxiliary symbol

$$s_k(c_1, c_2) = \frac{2}{\gamma_{c_1} + \gamma_{c_2}} \frac{y_{k+c_1} + y_{k+c_2}}{2} = \frac{y_{k+c_1} + y_{k+c_2}}{\gamma_{c_1} + \gamma_{c_2}} \quad (15)$$

and writing, for simplicity s_k instead of $s_k(c_1, c_2)$, they can be presented in a compact form as

$$\overline{\mu}_{c_1, c_2, M}(u) = \frac{\sum_{k=1}^M \left\{ s_k K \left(\frac{u_k - u}{h(M)} \right) \right\}}{\sum_{k=1}^M K \left(\frac{u_k - u}{h(M)} \right)}. \quad (16)$$

In (Greblicki and Pawlak, 1994), it was proved that the following convergence holds as $M \rightarrow \infty$:

$$Mh(M) \text{var } \widehat{\mu}_{c, M}(u) \rightarrow \frac{G \int_{-\infty}^{\infty} K^2(v) dv}{f(u)},$$

where

$$G = \sum_{i \neq c} \gamma_i^2 \text{var } w_k.$$

Since

$$G \leq \sum_{i=0}^{\infty} \gamma_i^2 \text{var } w_k = \text{var } y_k,$$

we conclude that, asymptotically,

$$\text{var } \widehat{\mu}_{c,M}(u) \leq \frac{1}{Mh(M)} \text{var } y_k,$$

and, by analogy,

$$\text{var } \overline{\mu}_{c_1, c_2, M}(u) \leq \frac{1}{Mh(M)} \text{var } s_k.$$

Hence we obtain the following criterion:

$$\text{var } s_k \rightarrow \min_{c_1, c_2}.$$

Using the well-known fact concerning the variance of the sum of random variables $\text{var}(X + Y) = \text{var } X + \text{var } Y + 2\text{cov}(X, Y)$, from (15) we get

$$\begin{aligned} \text{var } s_k &= \frac{1}{(\gamma_{c_1} + \gamma_{c_2})^2} \text{var}(y_{k+c_1} + y_{k+c_2}) \\ &= \frac{\sigma_y^2 + \sigma_y^2 + 2\sigma_y^2 \xi_y(c_2 - c_1)}{(\gamma_{c_1} + \gamma_{c_2})^2} \\ &= 2\sigma_y^2 \frac{1 + \xi_y(c_2 - c_1)}{(\gamma_{c_1} + \gamma_{c_2})^2}, \end{aligned}$$

where $\xi_y(\cdot)$ is the autocorrelation function of the output process $\{y_k\}$ ($-1 \leq \xi_y(\tau) \leq 1$), and $\sigma_y^2 = \text{var } y_k = \text{const}$. Thus, the selection of c_1 and c_2 should be conducted according the rule

$$Q(c_1, c_2) = \frac{1 + \xi_y(c_2 - c_1)}{(\gamma_{c_1} + \gamma_{c_2})^2} \rightarrow \min_{c_1, c_2}. \quad (17)$$

Since the autocorrelation of the output and the terms of the impulse response are unknown, direct application of (17) in the identification routine is not possible. Therefore, $\xi_y(\tau)$, and γ_τ for $\tau = 0, 1, \dots$ should be estimated in the first place (for details, see section 5). Let us emphasize that the form of $Q(c_1, c_2)$ is consistent with intuition. We postulate c_1 and c_2 , which guarantee a small output autocorrelation for $\tau = c_2 - c_1$ and, simultaneously, possibly large values of γ_{c_1} and γ_{c_2} .

5. Algorithm

In light of (17), we propose the following steps:

Step 1. Nonparametric estimation of the impulse response $\{\widehat{\gamma}_\tau\}$ of the linear dynamics (Greblicki and Pawlak, 1986):

$$\widehat{\gamma}_\tau = \frac{\widehat{w}_\tau}{\widehat{w}_0},$$

where

$$\widehat{w}_\tau = \frac{1}{N} \sum_{k=1}^{N-\tau} (y_{k+\tau} - \bar{y})(u_k - \bar{u})$$

$$\bar{y} = \frac{1}{N} \sum_{k=1}^N y_k, \quad \bar{u} = \frac{1}{N} \sum_{k=1}^N u_k.$$

Step 2. Nonparametric estimation of the output autocorrelation $\widehat{\xi}_y(\tau)$ (standard):

$$\widehat{\xi}_y(\tau) = \frac{\widehat{\text{cov}}_y(\tau)}{\widehat{\text{var}}_y} = \frac{\frac{1}{N-\tau} \sum_{k=1}^{N-\tau} (y_{k+\tau} - \bar{y})(y_k - \bar{y})}{\frac{1}{N-1} \sum_{k=1}^N (y_k - \bar{y})^2}$$

Step 3. Selection of \tilde{c}_1 , and \tilde{c}_2 by the minimization of the empirical criterion:

$$\widehat{Q}(c_1, c_2) = \frac{1 + \widehat{\xi}_y(c_2 - c_1)}{(\widehat{\gamma}_{c_1} + \widehat{\gamma}_{c_2})^2}.$$

Step 4. Nonparametric (improved) estimation of the non-linear characteristic using $\widehat{\mu}_{\tilde{c}_1, \tilde{c}_2, M}(u)$.

6. Special Cases

6.1. Two-Element Impulse Response. Let us consider the linear *FIR* filter with the following impulse response:

$$\gamma_i \begin{cases} \neq 0 & \text{if } i = d_1 \text{ or } i = d_2, \\ = 0 & \text{otherwise.} \end{cases}$$

Assuming, for clarity, that $z_k \equiv 0$, after simple computations we obtain

$$\xi_y(\tau) = \begin{cases} 1 & \text{if } \tau = 0, \\ \frac{\gamma_{d_1} \gamma_{d_2}}{\gamma_{d_1}^2 + \gamma_{d_2}^2} & \text{if } \tau = d_1 - d_2 \text{ or } \tau = d_2 - d_1, \\ 0 & \text{otherwise.} \end{cases}$$

Thus three variants of the identification routine are admissible:

(a) classical kernel regression based on the time-lag d_1 , i.e., $c_1 = c_2 = d_1$, with

$$Q(d_1, d_1) = \frac{1 + 1}{(\gamma_{d_1} + \gamma_{d_1})^2} = \frac{1}{2\gamma_{d_1}^2}, \quad (18)$$

(b) classical kernel regression based on the time-lag d_2 , i.e., $c_1 = c_2 = d_2$, with

$$Q(d_2, d_2) = \frac{1 + 1}{(\gamma_{d_2} + \gamma_{d_2})^2} = \frac{1}{2\gamma_{d_2}^2}, \quad (19)$$

(c) an improved estimate based on the combination $c_1 = d_1$ and $c_2 = d_2$ (or, equivalently, $c_1 = d_2$ and $c_2 = d_1$), with

$$Q(d_1, d_2) = \frac{1 + \frac{\gamma_{d_1}\gamma_{d_2}}{\gamma_{d_1}^2 + \gamma_{d_2}^2}}{(\gamma_{d_1} + \gamma_{d_2})^2}. \quad (20)$$

Corollary 1. *The estimate (a) is better than (b), i.e., $Q(d_1, d_1) < Q(d_2, d_2)$ if, and only if,*

$$|\gamma_{d_1}| > |\gamma_{d_2}|. \quad (21)$$

Proof. The condition (21) is a simple consequence of (18) and (19). ■

Corollary 2. *The estimate (c) is the best one, i.e., $Q(d_1, d_2) < \min(Q(d_1, d_1), Q(d_2, d_2))$ if, and only if,*

$$\left| \frac{\gamma_{d_1}}{\gamma_{d_2}} \right| \in \left(\frac{1}{\eta_0}, \eta_0 \right), \quad (22)$$

where η_0 is the largest real solution of the equation $\eta(\eta^3 - 2) = 1$, i.e., $1/\eta_0 \approx 0.716$ and $\eta_0 \approx 1.396$.

Proof. Observe that the relations between indices Q given by (18)–(20) are not sensitive to the scaling of γ_{d_1} and γ_{d_2} . Replacing γ_{d_1} and γ_{d_2} in (18)–(20) by $\bar{\gamma}_{d_1} = \alpha\gamma_{d_1}$ and $\bar{\gamma}_{d_2} = \alpha\gamma_{d_2}$, respectively, has the same influence on Q for each of the variants (a)–(c) (i.e., $\bar{Q} = Q/\alpha^2$). Therefore, without any loss of generality, we can set $\alpha = 1/\gamma_{d_1}$ and compare \bar{Q} s for $\bar{\gamma}_{d_1} = 1$ and $\bar{\gamma}_{d_2} = \gamma_{d_2}/\gamma_{d_1}$. The condition (22) thus results from the solution of the following system of inequalities:

$$\begin{cases} \frac{1 + \frac{\bar{\gamma}_{d_2}}{1 + \bar{\gamma}_{d_2}^2}}{(1 + \bar{\gamma}_{d_2})^2} < \frac{1}{2}, \\ \frac{1 + \frac{\bar{\gamma}_{d_2}}{1 + \bar{\gamma}_{d_2}^2}}{(1 + \bar{\gamma}_{d_2})^2} < \frac{1}{2\bar{\gamma}_{d_2}^2}, \end{cases}$$

with respect to $\bar{\gamma}_{d_2}$. ■

Observe, in particular, that if $\gamma_{d_1} = \gamma_{d_2}$, we obtain the best possible (25%) reduction of Q , i.e., $Q(d_1, d_2) = \frac{3}{4}Q(d_1, d_1)$.

6.2. Finite-Memory Integrator. For the Hammerstein system with the linear dynamics of the form

$$\gamma_i = \begin{cases} 1 & \text{if } i < d, \\ 0 & \text{if } i \geq d, \end{cases}$$

the output autocorrelation function is as follows:

$$\xi_y(\tau) = \begin{cases} \frac{d - |\tau|}{d} & \text{if } |\tau| < d, \\ 0 & \text{if } |\tau| \geq d. \end{cases}$$

Since for $|c_1 - c_2| \geq d$ we obtain $\gamma_{c_1}\gamma_{c_2} = 0$, the tags c_1 and c_2 must not be shifted more than d , i.e., $|c_1 - c_2| < d$, and then

$$Q(c_1, c_2) = \frac{1 + \frac{d - |c_1 - c_2|}{d}}{4} = \frac{2 - \frac{|c_1 - c_2|}{d}}{4}.$$

Observe, in particular, that for d large and $|c_1 - c_2| = d - 1$ we obtain a 50% reduction of Q , i.e., $Q(0, d - 1) = \frac{1}{2}Q(0, 0)$.

6.3. ARX(1) Linear Dynamics. We shall show that, in some cases, the application of the four-step procedure from Section 5 as the classical kernel regression estimate cannot be improved, independently of the c_1 and c_2 selection. Let us examine the Hammerstein system with stable ARX(1) linear dynamics (Hannan and Deistler, 1998), i.e., the one described by the difference equation $v_k = \lambda v_{k-1} + w_k$, where $0 < \lambda < 1$. Its impulse response is of the form

$$\gamma_i = \lambda^i, \quad i = 0, 1, 2, \dots, \infty.$$

For simplicity of presentation, assume that $z_k = 0$, $Ew_k = 0$ and $\text{var } w_k = 1$. In this case, we obtain the following two families of methods:

- (a) $c_1 = c_2 = d$, i.e., classical kernel regression estimates based on the time-lag $d = 0, 1, \dots, \infty$, with $Q(d, d) = 1/2\gamma_d^2$;
- (b) $c_1 \neq c_2$, where $c_1, c_2 \in \{0, 1, \dots\}$, i.e., generalized kernel estimates.

Corollary 3. *The optimal choice of the time-lag in (a) with respect to Q is $d = 0$.*

Proof. It is sufficient to observe that $Q(d, d) = 1/2\lambda^{2d}$ and $\arg \min_d Q(d, d) = 0$. ■

Corollary 4. *For each $c_1 \leq c_2$, we have*

$$Q(c_1, c_2) \geq Q(c_1, c_1). \quad (23)$$

Proof. We shall show that $\arg \min_{c_2} Q(c_1, c_2) = c_1$. For $c_1 = 0$, we have

$$Q(0, c_2) = \frac{1 + \xi(c_2)}{(1 + \lambda^{c_2})^2}.$$

Since

$$\text{var } y_k = \sum_{i=0}^{\infty} \lambda^{2i} = \frac{1}{1 - \lambda^2}$$

and the output autocovariance function has the form

$$A_y(\tau) = \sum_{i=0}^{\infty} \lambda^{2i+\tau} = \frac{\lambda^\tau}{1-\lambda^2},$$

we obtain

$$\xi(c_2) = \frac{A_y(c_2)}{\text{var } y_k} = \frac{\frac{\lambda^{c_2}}{1-\lambda^2}}{\frac{1}{1-\lambda^2}} = \lambda^{c_2}$$

and further

$$Q(0, c_2) = \frac{1 + \lambda^{c_2}}{(1 + \lambda^{c_2})^2} = \frac{1}{1 + \lambda^{c_2}}.$$

Consequently, we conclude that $\arg \min_{c_2} Q(0, c_2) = 0 = c_1$. The proof of (23) for $c_1 \neq 0$ can be conducted following the same steps. ■

7. Simulation Examples

7.1. Example 1. For illustration, let us analyze the Hammerstein system with $\{\gamma_i\} = \{1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, \dots\}$, $\mu(u) = \text{atan } u$ and $u_k \sim U[-1, 1]$. The system output was disturbed by white random noise $z_k \sim U[-0.1, 0.1]$. Simple computations lead to the following conclusions:

$$\xi_y(\tau) = \begin{cases} \frac{10 - |\tau|}{10} & \text{if } |\tau| < 10, \\ 0 & \text{if } |\tau| \geq 10, \end{cases}$$

$$Q(c_1, c_2) = \frac{1 + \xi_y(c_2 - c_1)}{(\gamma_{c_1} + \gamma_{c_2})^2} = \frac{2 - \frac{|\tau|}{10}}{4} \rightarrow \min_{c_1, c_2}$$

$$\arg \min_{|\tau|=0,1,2,\dots,9} Q(c_1, c_2) = 9 \rightarrow c_1^* = 0, c_2^* = 9.$$

We tested the performance of the four-step algorithm in practice using $N = 100$ simulated data pairs $\{u_k, y_k\}_{k=1}^N$, and obtained the same results, i.e., $\tilde{c}_1 = c_1^* = 0, \tilde{c}_2 = c_2^* = 9$. The sample mean square error (*MSE*) obtained in the experiment (see Tab. 1) shows the superiority of the combined estimate $\hat{\mu}_{0,9}(\cdot)$ over $\hat{\mu}_0$ and $\hat{\mu}_9(\cdot)$.

Table 1. Sample mean square errors.

$\widehat{MSE} \hat{\mu}_0(u=0)$	$\widehat{MSE} \hat{\mu}_9(u=0)$	$\widehat{MSE} \hat{\mu}_{0,9}(u=0)$
2.13	2.16	1.20

7.2. Example 2. Replacing the linear filter from Example 1 with

$$\gamma = \left\{ 1, \frac{1}{100}, 0, 0, 0, \dots \right\},$$

we obtain $c_1^* = c_2^* = 0$. The domination of the term γ_0 over the others reduces the improved estimate $\hat{\mu}_{c_1, c_2}(\cdot)$ to its classical version $\hat{\mu}_c(\cdot)$ (since the estimate $\hat{\mu}_{0,0}(\cdot)$ is equivalent to $\hat{\mu}_0(\cdot)$).

7.3. Example 3. In the experiment described in Example 1, the linear dynamics were replaced by the *IIR* filter with the transfer function

$$K(z) = \frac{z^5 + 1}{z^5 - 0.5z^4}$$

(see the impulse response in Fig. 4(a)). For various numbers of data, the following aggregated estimation error was computed:

$$AEE(M) = \frac{1}{P} \sum_{p=1}^P \left(\hat{\mu}_M(u^{[i]}) - \mu(u^{[i]}) \right)^2,$$

where $u^{[1]}, \dots, u^{[P]}$ were deterministic equispaced estimation points in the domain of $\mu(\cdot)$. As \hat{Q} is minimal for $c_1 = 0$ and $c_2 = 5$ (see Figs. 4(b)–(c)), we compared the classical kernel regression estimate $\hat{\mu}_5(\cdot)$ with $\hat{\mu}_{0,5}(\cdot)$. The estimation errors are presented in Fig. 4(d). It illustrates the advantage of using the generalized approach also for *IIR* systems.

8. Conclusions

A simple modification of the classical kernel regression estimate in Hammerstein system identification makes it possible to limit the influence of the “system noise” produced by the linear dynamics, and reduces the variance of the estimate. The convergence conditions and asymptotic convergence rate are not changed. The only cost paid for an increased accuracy is the fact that the computational complexity is a bit greater. In general, nonparametric methods are recommended when the nonlinear characteristic cannot be expressed in a closed form including a finite number of unknown parameters and hence parametric approaches cannot be applied. However, their good properties appear asymptotically (for large M). Hence, to obtain a satisfactory accuracy a large number of collected data is required. The main advantage of the strategy presented in the paper is the fact that, under the expected accuracy, it allows us to apply nonparametric methods for relatively smaller M in comparison with the classical kernel regression. The generalized estimate is recommended instead of the classical one independently of the number of collected data. In addition, observe that the idea of averaging nonparametric estimates with various time-lags can be also applied to variance reduction in the orthogonal series approach.

References

Bai E.W. (2003): *Frequency domain identification of Hammerstein models.* — IEEE Trans. Automat. Contr., Vol. 48, No. 4, pp. 530–542.

Bai E.W. and Li D. (2004): *Convergence of the iterative Hammerstein system identification algorithm.* — IEEE Trans. Automat. Contr., Vol. 49, No. 11, pp. 1929–1940.

Billings S.A. and Fakhouri S.Y. (1982): *Identification of systems containing linear dynamic and static nonlinear elements.* — Automat., Vol. 18, No. 1, pp. 15–26.

Chang F.H.I. and Luus R. (1971): *A non-iterative method for identification using Hammerstein model.* — IEEE Trans. Automat. Contr., Vol. AC-16, No. 4, pp. 464–468.

Chen H.F. (2005): *Strong consistency of recursive identification for Hammerstein systems with discontinuous piecewise-linear memoryless block.* — IEEE Trans. Automat. Contr., Vol. 50, No. 10, pp. 1612–1617.

Giannakis G.B. and Serpedin E. (2001): *A bibliography on nonlinear system identification.* — Signal Process., Vol. 81, No. 3, pp. 533–580.

Giunta G., Jacovitti G. and Neri A. (1991): *Bandpass nonlinear system identification by higher cross correlation.* — IEEE Trans. Signal Process., Vol. 39, No. 9, pp. 2092–2095.

Gomez J.C. and Basualdo M. (2000): *Nonlinear model identification of batch distillation process.* — Proc. Int. IFAC Symp. Advanced Control of Chemical Processes, AD-CHEM, Pisa, Italy, pp. 953–959.

Greblicki W. (1989): *Nonparametric orthogonal series identification of Hammerstein systems.* — Int. J. Syst. Sci., Vol. 20, No. 12, pp. 2355–2367.

Greblicki W. (2001): *Recursive identification of Wiener systems.* — Int. J. Appl. Math. Comp. Sci., Vol. 11, No. 4, pp. 977–991.

Greblicki W., Krzyżak A. and Pawlak M. (1984): *Distribution-free pointwise consistency of kernel regression estimate.* — Ann. Stat., Vol. 12, No. 4, pp. 1570–1575.

Greblicki W. and Pawlak M. (1986): *Identification of discrete Hammerstein systems using kernel regression estimates.* — IEEE Trans. Automat. Contr., Vol. 31, No. 1, pp. 74–77.

Greblicki W. and Pawlak M. (1989): *Nonparametric identification of Hammerstein systems.* — IEEE Trans. Inf. Theory, Vol. 35, No. 2, pp. 409–418.

Greblicki W. and Pawlak M. (1994): *Cascade non-linear system identification by a non-parametric method.* — Int. J. Syst. Sci., Vol. 25, No. 1, pp. 129–153.

Haber M. and Keviczky L. (1999): *Nonlinear System Identification – Input-Output Modeling Approach.* — Dordrecht: Kluwer.

Haber R. and Zeirfuss P. (1988): *Identification of an electrically heated heat exchanger by several nonlinear models using different structures and parameter estimation methods.* — Tech. Rep., Inst. Machine and Process Automation, Technical University of Vienna, Austria.

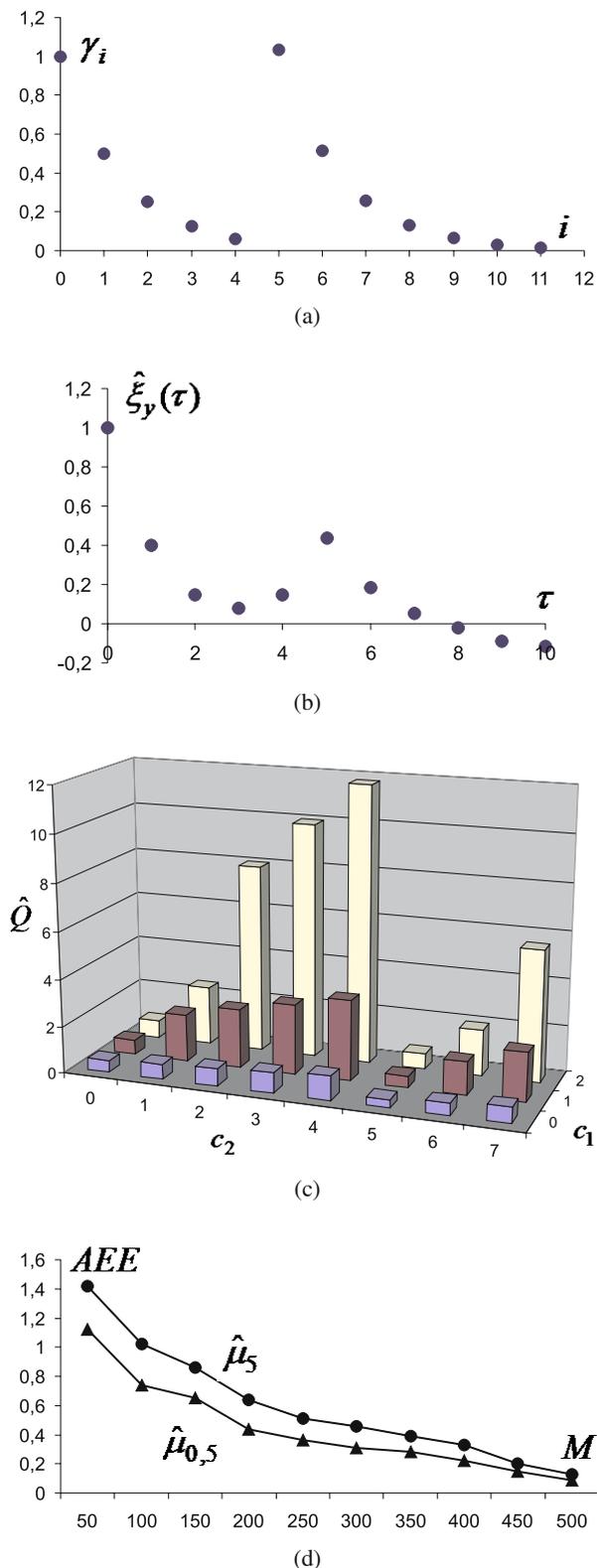


Fig. 4. (a) Impulse response of the IIR filter, (b) estimated output autocorrelation, (c) sample \hat{Q} versus c_1 and c_2 , (d) estimation error versus M .

- Hannan E.J. and Deistler M. (1998): *The Statistical Theory of Linear Systems*. — New York: Wiley.
- Hasiewicz Z. and Mzyk G. (2004a): *Combined parametric-nonparametric identification of Hammerstein systems*. — IEEE Trans. Automat. Contr., Vol. 49, No. 8, pp. 1370–1376.
- Hasiewicz Z. and Mzyk G. (2004b): *Nonparametric instrumental variables for Hammerstein system identification*. — Int. J. Contr., (submitted).
- Hasiewicz Z., Pawlak M. and Śliwiński P. (2005): *Nonparametric identification of nonlinearities in block-oriented systems by orthogonal wavelets with compact support*. — IEEE Trans. Circ. Syst. I: Fund. Theory Applic., Vol. 52, No. 2, pp. 427–442.
- Härdle W. (1990): *Applied Nonparametric Regression*. — Cambridge: Cambridge University Press.
- Janczak A. (1999): *Parameter estimation based fault detection and isolation in Wiener and Hammerstein systems*. — Int. J. Appl. Math. Comput. Sci., Vol. 9, No. 3, pp. 711–735.
- Jang W. and Kim G. (1994): *Identification of loudspeaker nonlinearities using the NARMAX modelling technique*. — J. Audio Eng. Soc., Vol. 42, No. 1/2, pp. 50–59.
- Krzyżak A. (1990): *On estimation of a class of nonlinear systems by the kernel regression estimate*. — IEEE Trans. Inf. Theory, Vol. IT-36, No. 1, pp. 141–152.
- Krzyżak A., Sasiadek J. and Kégl B. (2001): *Identification of dynamic nonlinear systems using the Hermite series approach*. — Int. J. Syst. Sci., Vol. 32, No. 10, pp. 1261–1285.
- Latawiec K.J. (2004): *The Power of Inverse Systems in Linear and Nonlinear Modeling and Control*. — Opole: Opole University of Technology Press.
- Ljung L. (1987): *System Identification: Theory for the User*. — Englewood Cliffs, NJ: Prentice Hall.
- Narendra K.S. and Gallman P.G. (1966): *An iterative method for the identification of nonlinear systems using the Hammerstein model*. — IEEE Trans. Automat. Contr., Vol. 11, No. 3, pp. 546–550.
- Pawlak M. and Hasiewicz Z. (1998): *Nonlinear system identification by the Haar multiresolution analysis*. — IEEE Trans. Circ. Syst., Vol. 45, No. 9, pp. 945–961.
- Söderström T. and Stoica P. (1982): *Instrumental-variable methods for identification of Hammerstein systems*. — Int. J. Contr., Vol. 35, No. 3, pp. 459–476.
- Söderström T. and Stoica P. (1989): *System Identification*. — Englewood Cliffs, NJ: Prentice Hall.
- Van den Hof P., Heuberger P. and Bokor J. (1995): *System identification with generalized orthonormal basis functions*. — Automatica, Vol. 31, No. 12, pp. 1821–1834.
- Vörös J. (1999): *Iterative algorithm for identification of Hammerstein systems with two-segment nonlinearities*. — IEEE Trans. Automat. Contr., Vol. 44, No. 11, pp. 2145–2149.
- Wand M.P. and Jones H.C. (1995): *Kernel Smoothing*. — London: Chapman and Hall.
- Zhang Y.K. and Bai E.W. (1996): *Simulation of spring discharge from a limestone aquifer in Iowa*. — Hydrogeol. J., Vol. 4, No. 4, pp. 41–54.
- Zhu Y. (2000): *Identification of Hammerstein models for control using ASYM*. — Int. J. Contr., Vol. 73, No. 18, pp. 1692–1702.

Received: 6 November 2006

Revised: 22 February 2007

Re-revised: 2 April 2007

