8th IEEE International Conference on Methods and Models in Automation and Robotics
Szczecin, Poland, September 2-5, 2002
vol. 1, 451-456, 2002

NON-LINEARITY ESTIMATION IN HAMMERSTEIN SYSTEM BASED ON ORDERED OBSERVATIONS AND WAVELETS

WŁODZIMIERZ GREBLICKI*, PRZEMYSŁAW ŚLIWIŃSKI†

*†Institute of Engineering Cybernetics, Wrocław University of Technology,
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland,
*wgre@ict.pwr.wroc.pl, †slk@ict.pwr.wroc.pl

Abstract. The non-linear subsystem of a Hammerstein system is identified, i.e., its characteristic is estimated from input-output observations of the whole system. The input and a disturbance are white stochastic processes. The identified characteristic satisfies a Lipschitz condition only. Presented wavelet-based identification algorithm is calculated from ordered input-output observations, i.e., from pairs of observations arranged in a sequence in which input measurements increase in value. The mean integrated square error of the resulting estimate converges to zero as the number of observations tends to infinity. Convergence rate is insensitive to the shape of the probability density of the input signal.

Key Words. system identification, Hammerstein system, non-linearity estimation, order statistics, Daubechies wavelets.

1. INTRODUCTION

Motivations. The block oriented approach to the non-linear system identification has been receiving growing attention in the theoretical literature, see Bendat [1], Billings [2], Billings and Fakhouri [3], Priestley [22], and in applications in various fields, e.g., control, Zi-Qiang [24], chemistry, Eskinat *et al.* [6], biology, Emerson *et al.* [5], Hunter and Korenberg [16], Korenberg and Hunter [17]. The main idea of the approach is that the identified system consists of simple subsystems such as linear, dynamic and non-linear, memoryless. The main objective of the block-oriented identification is to recover descriptions of all subsystems from observations taken at input and output of the whole system. So far, the greatest attention has been paid to Hammerstein systems, i.e., cascade systems consisting of a non-linear memoryless element followed by a linear dynamic one, see e.g. Narendra and Gallman [20], Haist *et al.* [13] and Lang Zi-Qiang [24]. Authors mentioned above have assumed that the non-linearity is known up to a finite number of coefficients. The most common restriction imposed on the non-linearity confines considerations to polynomials which coefficients are estimated. Resulting identification problems are parametric.

Inspirations. Greblicki and Pawlak [10] and Greblicki [7] significantly enlarged the class of considered characteristics. Their restrictions are mild, which means that their a priori information concerning the non-linearity is extremely poor since they assume that the characteristic is, e.g., bounded or square integrable only. Owing to that, the family of all possible characteristics admitted by them is so ample that can not be represented in a parametric form. Therefore, their nonparametric identification problems are closer to real problems encountered in applications.

The non-linear characteristic in the Hammerstein system has been already recovered with various orthogonal series estimates by Greblicki and Pawlak in e.g. [11], Greblicki [7], Pawlak [21], Krzyżak [18], Hasiewicz [14, 15] and Śliwiński [23]. However, convergence rates of their algorithms are sensitive to irregularities of the input probability density, while ours is not. Algorithms with convergence rates independent of the shape of the input signal density have been proposed by Greblicki and Pawlak [12] and Greblicki [8].

Scope of the paper. In this paper, the problem of recovering the non-linearity in a Hammerstein system is also non-parametric. The characteristic is assumed only to satisfy a Lipschitz condition. We propose wavelet-based algorithms to estimate the non-linear characteristic of the memoryless subsystem. For the algorithm the mean integrated square error converges to zero as the number of observations tends to infinity. The convergence rate depends on the non-linearity. The smoother characteristic, the greater speed of convergence. The rate is, moreover, independent of the shape of the probability density of the input signal. This property is an important advantage of our algorithms over those mentioned above, since their rates get worse for irregular densities. We derive the identification algorithm not from the original but ordered sequence of observations. Ordering means that input-output pairs of observations are rearranged with respect to input observations. In the new sequence, input observations increase in value.

2. IDENTIFICATION PROBLEM

We deal with *Hammerstein system*, i.e., a system consisting of a non-linear memoryless subsystem followed by a linear dynamic one (see Fig. 1), The system is driven by stationary white random noise $\{U_n; n = \cdots, -1, 0, 1, 2\cdots\}$.

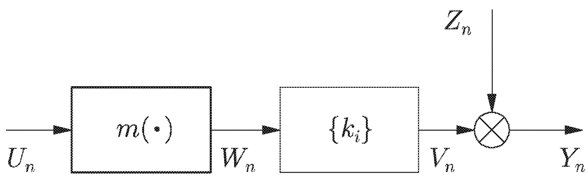

Fig. 1. Identified Hammerstein system with $k_0 = d$ and $k_i = c^T A^{i-1} b$, $i = 1, 2, \cdots$

We assume that

$$0 \le U_n \le 1.$$

The probability density f of U_n's is unknown and satisfies the following restriction:

$$0 < \delta \le f(u) \tag{1}$$

with all $u \in [0, 1]$, some unknown δ. The non-linear memoryless subsystem has a characteristic denoted by m, which means that

$$W_n = m(U_n).$$

The characteristic m is a Borel measurable function and satisfies a Lipschitz condition, i.e.

$$|m(u) - m(v)| \le L\,|u - v| \tag{2.2}$$

with some $L > 0$, and all u, v.

The dynamic subsystem is described by a state space equation

$$\left.\begin{aligned} X_{n+1} &= AX_n + bW_n \\ V_n &= c^T X_n + dW_n \end{aligned}\right\} \tag{2.4}$$

where X_n is the state vector at time n, and where the matrix A, vectors b, c, and the number d are all unknown. So is the dimension of the state vector. The matrix A is asymptotically stable, which means that all its eigenvalues lie in the unit circle. The output of the system is disturbed by stationary white random noise $\{Z_n; n = \cdots, -1, 0, 1, 2\cdots\}$. Therefore,

$$Y_n = V_n + Z_n.$$

The noise is independent of the input signal, has zero mean and unknown variance σ_Z^2. Owing to all that, $\{Y_n; n = \cdots, -1, 0, 1, 2\cdots\}$ is a sequence of *dependent* identically distributed random variables. The sequence is a stationary ARMA stochastic process.

The goal of the paper is to recover m from observations $(U_1, Y_1), (U_2, Y_2), \cdots, (U_n, Y_n)$ taken at the input and output of the whole system. Observe that

$$Y_n = \mu(U_n) + Z_n + \xi_n,$$

where

$$\mu(u) = dm(u) + c^T \mathrm{E} X_0 \tag{2.5}$$

and where

$$\begin{aligned} \xi_n &= c^T (X_n - \mathrm{E}X_n) = \\ &= \sum_{i=1}^{\infty} c^T A^{i-1} b\,[m(U_{n-i}) - \mathrm{E}m(U_{n-i})]\,. \end{aligned} \tag{2.6}$$

Clearly $\mu(u) = \mathrm{E}\{Y_n\,|U_n = u\}$, which means that the regression μ is observed in the presence of noise $Z_n + \xi_n$. Its first component is white, while the other incurred by the dynamic subsystem is correlated and depends on both m and the input signal. Observing the input and output of the whole system, we estimate μ , i.e., the regression of Y_n on U_n. The fact that we are able to recover m only up to some unknown constants d, and is a simple consequence of the composite structure of the system.

3. IDENTIFICATION ALGORITHM

Preliminaries. To identify the non-linear part of the system, we rearrange the sequence $U_1, U_2, \cdots, U_n$ of input observations into a new one $U_{(1)}, U_{(2)}, \cdots, U_{(n)}$, in which $U_{(1)} < U_{(2)} < \cdots < U_{(n)}$. Ties, i.e., events that $U_{(j)} = U_{(i)}$, for $j \neq i$, have zero probability since U_n's possess a density. Moreover, we define $U_{(0)} = 0$ and $U_{(n+1)} = 1$. The sequence $U_{(1)}, U_{(2)}, \cdots, U_{(n)}$ is called the order statistics of $U_1, U_2, \cdots, U_n$. We then rearrange the sequence $(U_1, Y_1), (U_2, Y_2), \cdots, (U_n, Y_n)$ of input-output observations into the following one: $(U_{(1)}, Y_{[1]}), (U_{(2)}, Y_{[2]}), \cdots, (U_{(n)}, Y_{[n]})$. Observe that $Y_{[j]}$'s are not ordered, but just paired with $U_{(j)}$'s.

Daubechies functions $\varphi^p_{kl}(u) = 2^{k/2}\varphi^p(2^k u - l)$, $k, l = \cdots - 1, 0, 1, 2\cdots$, are, for fixed p, dilations and translations of the scaling function (father wavelet) $\varphi^p(u)$. The index $p = 1, 2, \cdots$, is the wavelet number, see [4].

Algorithm. We propose the following wavelet estimate of $\mu(u)$ based on the ordered sequence of observations:

$$\hat{\mu}(u) = \sum_{l=-2p+2}^{2^k-1} \hat{\alpha}^p_{kl}\varphi^p_{kl}(u) \tag{2}$$

where

$$\hat{\alpha}^p_{kl} = \sum_{j=1}^{n} Y_{[j]}\left(U_{(j)} - U_{(j-1)}\right)\varphi^p_{kl}\left(U_{(j)}\right) \tag{3}$$

Note that for each p we obtain a distinct wavelet-based estimate. We will show that for each p, and for suitably selected scale parameter k, the algorithm converges to $\mu(u)$ as the number of observations increases to infinity.

Convergence and its rate. The following two theorems characterize the estimate convergence and its rate. The first deals with conditions of mean integrated square convergence.

Theorem 1 *If*

$$k \to \infty \quad \text{such that} \quad 2^{2k}/n \to 0 \quad \text{as} \quad n \to \infty \tag{4}$$

then

$$\text{MISE}\,\hat{\mu} \to 0 \quad \text{as} \quad n \to \infty$$

where

$$\text{MISE}\,\hat{\mu} \overset{\text{def}}{=} \text{E}\int_0^1 [\mu(u) - \hat{\mu}(u)]^2\, du$$

for any Lipschitz non-linearity m.

Second theorem establishes the MISE-convergence rate:

Theorem 2 *If conditions in (4) hold and the scale parameter k is selected according to the rule*

$$k = k(n) = \frac{1}{4}\log_2 n \tag{5}$$

then

$$\text{MISE}\,\hat{\mu} \leq cn^{-1/2} \tag{6}$$

for some c, independent of n.

The sketchy proofs of these theorems can be found in appendices. Observe that the rate (6) is worse than the asymptotically optimal rate $cn^{-2/3}$ obtained for Fourier and for kernel estimates by Greblicki and Pawlak in [12] and by Greblicki in [8], respectively.

3.1. Haar algorithm

As an example, we present the algorithm based on Haar functions, i.e. on Daubechies function with $p = 1$. Due to simplicity of these functions which are, in fact, scaled and translated version of the indicator function

$$\mathbf{I}_{[0,1)}(u) = \begin{cases} 1 & \text{if } u \in [0,1) \\ 0 & \text{otherwise} \end{cases},$$

the algorithm is of extremely simple form

$$\hat{\mu}_H(u) = \sum_{l=0}^{2^k-1} \hat{\alpha}^1_{kl}\mathbf{I}_{\left[\frac{l}{2^k}, \frac{l+1}{2^k}\right)}(u)$$

where

$$\hat{\alpha}^1_{kl} = \sum_{j=1}^{n} Y_{[j]}\left(U_{(j)} - U_{(j-1)}\right)\mathbf{I}_{\left[\frac{l}{2^k}, \frac{l+1}{2^k}\right)}\left(U_{(j)}\right)$$

4. FINAL REMARKS

Mixed algorithm. The proposed algorithm consists of Daubechies scaling functions only. One can also consider the estimate which additionally incorporates Daubechies wavelet functions $\psi^p_{kl}(u) = 2^{k/2}\psi^p(2^k u - l)$, $k, l = \cdots - 1, 0, 1, 2\cdots$, where $\psi^p(u)$ is pth mother function. The resulting algorithm takes a bit more complicated form

$$\hat{\mu}(u) = \sum_{l=-2p+2}^{2^M-1} \hat{\alpha}^p_{Ml}\varphi^p_{Ml}(u) + \sum_{m=M}^{k-1}\sum_{l=-p+1}^{2^m+p-2} \hat{\beta}^p_{ml}\psi^p_{ml}(u) \tag{7}$$

some $M < k$, where $\hat{\alpha}^p_{Ml}$ are calculated as in (3) while the wavelet coefficients are calculated as follows

$$\hat{\beta}^p_{ml} = \sum_{j=1}^{n} Y_{[j]}\left(U_{(j)} - U_{(j-1)}\right)\psi^p_{ml}\left(U_{(j)}\right)$$

One can show that the asymptotic properties of this algorithm are the same as for the algorithm (2) – i.e. the Theorems 1 and 2 hold for algorithm (7) under the same, respective, conditions. Its main advantage over the former consists in ability of expanding without necessity of recalculation of the existing part. Moreover, applying *fast wavelet transform* (see e.g. [19]) to calculate the wavelet coefficients $\hat{\beta}^p_{kl}$, the numerical complexities of both algorithms are of the same order (cf [23]).

Other algorithms. A superiority of the presented estimates (and of their prototypes introduced by Greblicki and Pawlak in [12]) over others, known in literature, comes from the fact that its convergence rate is independent of the shape of the density of the input signal, i.e., is insensitive to its irregularities. The rate depends only on the smoothness of m and holds for any f bounded from zero while for estimates of quotient form, as for instance the following (considered by Hasiewicz in [15]):

$$\tilde{\mu}(u) = \frac{\sum_{l=-2p+2}^{2^k-1} \tilde{\alpha}^p_{kl}\varphi^p_{kl}(u)}{\sum_{l=-2p+2}^{2^k-1} \tilde{a}^p_{kl}\varphi^p_{kl}(u)}$$

with

$$\tilde{\alpha}^p_{kl} = \frac{1}{n}\sum_{j=1}^{n} Y_j \varphi^p_{kl}(U_j)$$

and

$$\tilde{a}^p_{kl} = \frac{1}{n}\sum_{n=1}^{n} \varphi^p_{kl}(U_j),$$

it depends on the smoothness of both m and f and hence, get worse for rough f.

It should be also emphasized that our estimate is of simpler form.

The toll we pay for insensitivity of the convergence rate to irregularities of the probability density of the input signal is an increased computational complexity of the proposed algorithms. It seems, however, that this extra effort is well compensated.

Future works. In further studies the conditions of convergence and its rate for the proposed algorithm will be studied for characteristics being discontinuous or continuous and multiple differentiable (cf the approaches in [8] and [23]). An adoption of the algorithm to the identification of the non-linear characteristic in continuous-time Hammerstein system (see [9]) will be also developed.

APPENDIX

Ordered observations in Hammerstein systems. The following lemma holds for Hammerstein system (see [12]):

Lemma 1 *Let f satisfy (1). Then for any real $q > 0$ and any $n \geq 1$,*

$$\mathrm{E}\left(U_{(j)} - U_{(j-1)}\right)^q = O\left(n^{-q}\right)$$

any $j = 1, 2, \ldots, n+1$,

$$\mathrm{E}\left\{\sum_{j=1}^{n+1}\left(U_{(j)} - U_{(j-1)}\right)^{q+1}\right\}^2 = O\left(n^{-2q}\right).$$

Moreover

$$\mathrm{E}\left\{\sum_{j=1}^{n+1}\xi_{[j]}\left(U_{(j)} - U_{(j-1)}\right)\right\}^2 \leq O\left(n^{-1}\right).$$

Wavelet expansions. Denote the wavelet coefficients of expansion of μ in the pth wavelet series as below

$$\beta^p_{ml} = \int_0^1 \mu(u)\,\psi^p_{ml}(u)\,du$$

Lemma 2 *For any characteristic μ bounded in the interval $[0, 1]$ it holds that*

$$\sum_{m=k}^{\infty}\sum_{l=-p+1}^{2^m+p-2}\left(\beta^p_{ml}\right)^2 \to 0 \quad as \quad k \to \infty \qquad (8)$$

Moreover, for μ being Lipschitz

$$\sum_{m=k}^{\infty}\sum_{l=-p+1}^{2^m+p-2}\left(\beta^p_{ml}\right)^2 \leq c2^{-2k} \qquad (9)$$

some c, independent of k.

Proof of the Theorem 1. We have

$$\mathrm{MISE}\,\hat{\mu} = \sum_{l=-2p+2}^{2^k-1}\mathrm{E}\left(\alpha^p_{kl} - \hat{\alpha}^p_{kl}\right)^2 + \sum_{m=k}^{\infty}\sum_{l=-p+1}^{2^m+p-2}\left(\beta^p_{ml}\right)^2 \qquad (10)$$

In the proof we exploit the following identities

$$\begin{aligned}\alpha^p_{kl} &= \int_0^1 \mu(u)\,\varphi^p_{kl}(u)\,\mathrm{d}u \\ &= \sum_{j=1}^{n}\int_{U_{(j-1)}}^{U_{(j)}} \mu(u)\,\varphi^p_{kl}(u)\,\mathrm{d}u \\ &\quad + \int_{U_{(n)}}^{U_{(n+1)}} \mu(u)\,\varphi^p_{kl}(u)\,\mathrm{d}u\end{aligned}$$

and

$$\hat{\alpha}^p_{kl} = \sum_{j=1}^{n} Y_{[j]} \int_{U_{(j-1)}}^{U_{(j)}} \varphi^p_{kl}\left(U_{(j)}\right) \mathrm{d}u$$

Further, using Lemma 1 one can show that (cf [8])

$$\mathrm{E}\left(\alpha^p_{kl} - \hat{\alpha}^p_{kl}\right)^2 \leq 4\left(V_1 + V_2 + V_3 + V_4\right)$$

where

$$\begin{aligned} V_1 &\leq \delta_1 2^k n^{-1} \\ V_2 &\leq \delta_2 2^k n^{-1} \\ V_3 &\leq \delta_3 2^k n^{-1} \\ V_4 &\leq \delta_4^2 n^{-2} \end{aligned}$$

for some $\delta_1, \delta_2, \delta_3$ and δ_4. Finally we get

$$\begin{aligned} \mathrm{MISE}\,\hat{\mu} \;\leq\; & 4\left(2^k + 2p\right) \times \\ & \left(\delta_1 2^k n^{-1} + \delta_2 2^k n^{-1} + \right. \\ & \left. \delta_3 2^k n^{-1} + \delta_4^2 n^{-2}\right) \\ & + \sum_{m=k}^{\infty} \sum_{l=-p+1}^{2^m+p-2} \left(\beta^p_{ml}\right)^2 \end{aligned} \quad (11)$$

which completes the proof due to (8). ■

Proof of the Theorem 2. The convergence rate in (6) is the simple consequence of (9), (10) and (11). We have

$$\mathrm{MISE}\,\hat{\mu} \leq c\left(2^{2k}/n + 2^{-2k}\right)$$

for some c, independent of k. Hence, applying to the inequality above the rule from (5) finishes the proof. ■

REFERENCES

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

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

[3] S. A. Billings and S. Y. Fakhouri. Identification of systems containing linear dynamic and static non-linear elements. *Automatica*, 18:15–26, 1982.

[4] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Communication on Pure and Applied Mathematics*, 42:909–996, 1992.

[5] R. C. Emerson, M. J. Korenberg, and M. C. Citron. Identification of complex-cell intensive nonlinearities in a cascade model of cat visual cortex. *Biological Cybernetics*, 66:291–300, 1992.

[6] E. Eskinat, S. H. Johnson, and W. L. Luyben. Use of hammerstein models in identification of non-linear systems. *American Institute of Chemical Engineers Journal*, 37:255–268, 1991.

[7] W. Greblicki. Nonparametric orthogonal series identification of Hammerstein systems. *International Journal of Systems Science*, 20:2355–2367, 1989.

[8] W. Greblicki. Nonlinearity estimation in Hammerstein systems based on ordered statistics. *IEEE Transactions on Signal Processing*, 44:1224–1233, 1996.

[9] W. Greblicki. Continuous-time Hammerstein system identification. *IIEEE Transactions on Automatic Control*, 45:1232–1236, 2000.

[10] W. Greblicki and M. Pawlak. Identification of discrete Hammerstein system using kernel regression estimates. *IEEE Transactions on Automatic Control*, 31:74–77, 1986.

[11] W. Greblicki and M. Pawlak. Nonparametric identification of a particular nonlinear time series system. *IEEE Transactions on Signal Processing*, 40:985–989, 1992.

[12] W. Greblicki and M. Pawlak. Dynamic system identification with order statistics. *IEEE Transactions on Information Theory*, 1994.

[13] N. D. Haist, F. H. L. Chang, and R. Luus. Non-linear identification in the presence of the correlated noise using hammerstein model. *IEEE Transactions on Automatic Control*, 18:552–555, 1973.

[14] Z. Hasiewicz. Hammerstein system identification by the Haar multiresolution approximation. *International Journal of Adaptive Control and Signal Processing*, 13:697–717, 1999.

[15] Z. Hasiewicz. Non-parametric estimation of non-linearity in a cascade time series system by multiscale approximation. *Signal Processing*, 81:791–807, 2001.

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

[17] M. J. Korenberg and I. W. Hunter. The identification of nonlinear biological system: Lnl cascade models. *Biological Cybernetics*, 55:125–134, 1986.

[18] A. Krzyżak. Identification of discrete Hammerstein systems by the Fourier series regression estimate. *International Journal of System Science*, 20:1729–1744, 1989.

[19] S. G. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, San Diego, 1998.

[20] K. S. Narendra and P. G. Gallman. An iterative method for the identification of nonlinear systems using the hammerstein model. *IEEE Transactions on Automatic Control*, 11:546–550, 1966.

[21] M. Pawlak. On the series expansion approach to the identification of Hammerstein systems. *IEEE Transactions on Automatic Control*, 36:763–767, 1991.

[22] M. B. Priestley. *Non-linear and Non-stationary Time Series Analysis*. Academic Press, London, 1991.

[23] P. Śliwiński. *Non-linear system identification by wavelets*. PhD dissertation, Wrocław University of Technology, Institute of Engineering Cybernetics, 2000.

[24] L. Zi-Qiang. Controller design oriented model identification method for Hammerstein system. *Automatica*, 29:767–771, 1993.