NON-PARAMETRIC IDENTIFICATION OF NON-LINEARITY IN HAMMERSTEIN SYSTEMS

W. Greblicki and P. Śliwiński
Institute of Engineering Cybernetics, Wroclaw University of Technology, Wroclaw, Poland

printed in: IFAC Proceedings Volumes, Vol. 36, No. 16, 825-830, 2003

13 IFAC Symposium on System Identification
Rotterdam, The Netherlands, 27-29 August 2003

Abstract. In the paper the non-parametric algorithms identifying the nonlinearity in Hammerstein system are presented. About the non-linearity it is only assumed that it is piecewise Lipschitz function. The linear dynamic element of the system can be unknown. It is shown that the algorithms converge to any of such non-linearity with growing number of measurements n. For Lipschitz and smoother non-linearities the algorithms attain the rate $O(n^{-2p/(2p+1)})$, the best possible for non-parametric algorithms. The convergence and its rate are independent of the regularity of the input probability density function.

Key Words. Hammerstein system, non-parametric identification, kernel algorithm, orthogonal series algorithm, algorithm accuracy

1. Introduction

In statistical inference we deal with two kind of information:

- *a priori* information possessed before experiment,
- empirical information obtained from measurements.

If the former is reach enough to build a reliable parametric model — the parametric approach is applied. However, it is often the case that this information is not available and we can solely rely on the empirical data–then a non-parametric approach appears to be more suitable.

In the paper we present non-parametric algorithms identifying the non-linearity in Hammerstein system. About the non-linearity we only assume that it is piecewise Lipschitz (i.e. it can have discontinuities). From practical point of view — we assume nothing about the non-linearity – thus we admit in algorithms virtually all non-linearities. It is clear that there is no parametric model able to handle all possible non-linearities from such a class. Thus the non-parametric approach may be considered. We show that in spite of the poor a priori information our algorithms converge the genuine non-linearity – i.e. they are able to recover non-linearity basing on the empirical knowledge only.We also examine the rates of algorithms convergence and show that they are not much slower than the ones achieved by the parametric algorithm, thus the nonparametric algorithms accuracy is not much worse than their parametric counterparts. The problem of identification of Hammerstein system has been widely investigated; see e.g. Narendra and Gallman (1966), Chang and Luus (1971), Haist et al. (1973), Thathachar and Ramaswamy (1973) for parametric algorithms and e.g. Greblicki and Pawlak (1986), Krzyżak (1989), Pawlak (1991), Hasiewicz (1999) or Śliwiński (2000) for non-parametric ones.

Hammerstein systems have been also applied in various fields: e.g. control, Zi-Qiang (1993), chemistry, Eskinat et al. (1991) or biology Emerson et al. (1992), Hunter and Korenberg (1986), Korenberg and Hunter (1986).

2. Identification problem

Consider the Hammerstein system, i.e. the cascade system with non-linear static element followed by dynamic linear one (Fig. 1).

Denote by m the non-linear characteristic of the

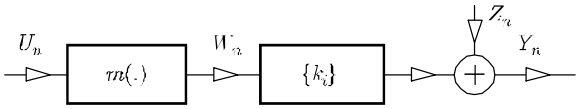

Figure 1: Identified Hammerstein system.

former and by $\{k_i\}_{i=0}^{\infty}$ the discrete impulse response of the latter. The goal is to recover the non-linearity from the random input-output measurements $(U_1, Y_1), (U_2, Y_2), \ldots, (U_n, Y_n)$ of the whole system (the internal signal W_n is not available for measurements) under the following assumptions:

1. The input signal $\{U_n; n = \ldots -1, 0, 1, 2, \ldots\}$ is a stationary white random process with unknown probability density function f. We assume that $-1 \leq U_n \leq 1$ and that f satisfies the restriction
$$|f(u) - f(u+h)| \geq ch \tag{1}$$
some $c > 0$, for each u where $f(u) = 0$.

2. The non-linearity m is piecewise Lipschitz function (see Appendix A).

3. The dynamic subsystem is asymptotically stable, i.e. $\sum_{i=0}^{\infty} |k_i| < \infty$.

4. The external noise $\{Z_n; n = \ldots -1, 0, 1, 2, \ldots\}$ is a stationary white process with zero mean and finite variance.

2.1. Comments upon assumptions

By assumption 1, the class of admitted input signals is confined to those which have smooth distribution function (i.e. for which exists the probability density function). For identification algorithms converging for any distribution of U (i.e. distribution-free or universally consistent algorithms) we refer to the paper by Greblicki and Pawlak (1989); see also Györfi et al. (2002) for the up-to-date results in statistics. This assumptions also limits the range of input signals to white noise processes – for algorithms applicable for correlated input we refer in turn to the paper by Greblicki and Pawlak (1994b). The restriction in (1) admits densities which are not bounded from 0 (e.g. triangle and parabolic densities) (cf. Greblicki (1996), Greblicki and Pawlak (1994a)).

Assumption 2 admits, in general, discontinuous non-linearities and hence allows to identify virtually all non-linearities met in practice. It is clear that the class of non-linearities satisfying the assumption is too ample to be represented in the closed parametric form. This makes our problem non-parametric.

Due to assumption 3 the structure of the dynamic part can be arbitrary ARMA system with unknown order, and therefore, it can remain unknown during identification routine. Similarly, assumption 4 imposes no restrictions on the distribution of the external noise.

3. Algorithms of identification

The input-output equation of the system can be written as
$$Y_n = \mu(U_n) + \xi_n + Z_n,$$
where μ is scaled and shifted version of the genuine characteristic m, i.e. $\mu(u) = k_0 m(u) + d$ with $d = Em(U_1) \sum_{i=1}^{\infty} k_i$. The signal xi_n is defined as
$$\xi_n = \sum_{i=1}^{\infty} k_i [m(U_{n-i}) - Em(U_1)]$$
and is an additional distortion acting, together with noise Z_n, on the output of μ.

Construction of identification algorithms starts from the observation made by Greblicki and Pawlak (1986), that for Hammerstein systems holds:
$$E\{Y_n | U_n = u\} = \mu(u)$$
i.e. that the regression of Y_n on U_n is equal to the scaled and shifted non-linear characteristic m. Hence, estimating the regression function we recover the non-linearity μ. Below we present two non-parametric algorithms identifying the non-linearity μ. The kernel algorithm and the orthogonal one.

Remark 1 The fact that the non-linearity can be recovered up to some constants is a consequence of the assembled structure of the system and of the inaccessibility of the interconnecting signal W_n – it is independent

of the identification algorithms. Furthermore, the constants cannot be estimated without additional a priori information (e.g. known values of m in some points).

3.1. Preprocesing

Both algorithms operate on the measurement data rearranged into a new sequence $U_{(1)}, Y_{[1]}, U_{(2)}, Y_{[2]}, \cdots, U_{(n)}, Y_{[n]}$ for which $U_{(1)} < U_{(2)} < \cdots < U_{(n)}$, that is, on the measurement pairs sorted increasingly according to the values of U_i's.

3.2. Algorithms

The first algorithm is the kernel algorithm of the following form (see Greblicki (1996)):

$$\hat{\mu}(u) = \frac{1}{h(n)} \sum_{j=1}^{n} Y_{[j]} \int_{U_{(j-1)}}^{U_{(j)}} K\left(\frac{u-v}{h(n)}\right) dv, \qquad (2)$$

where K is a kernel function (see Appendix A) and $\{h(n)\}$ is a positive number sequence. The second is the orthogonal algorithm of the form (see Greblicki and Pawlak (1994a)):

$$\tilde{\mu}(u) = \sum_{k=0}^{q(n)} \tilde{c}_k \varphi_k(u), \qquad (3)$$

with

$$\tilde{c}_k = \sum_{j=1}^{n} Y_{[j]} \int_{U_{(j-1)}}^{U_{(j)}} \varphi_k(v) dv, \qquad (4)$$

where $\{\varphi_k; k = 0, 1, 2, \ldots\}$ is a set of complete orthogonal functions on $[-1, 1]$ (e.g. trigonometric or Legendre set) and $\{q(n)\}$ is a positive number sequence.

3.3. Convergence

The following theorems characterize the conditions of the global convergence of the algorithms for the Assumptions 1-4:

Theorem 1 *If*

$$h(n) \to 0 \text{ and } nh(n) \to \infty \text{ as } n \to \infty \qquad (5)$$

then

$$E \int_{-1}^{1} [\hat{\mu}(u) - \mu(u)]^2 du \to 0 \text{ as } n \to \infty.$$

Theorem 2 *If*

$$q(n) \to \infty \text{ and } q(n)/n \to 0 \text{ as } n \to \infty, \qquad (6)$$

then

$$E \int_{-1}^{1} [\tilde{\mu}(u) - \mu(u)]^2 du \to 0 \text{ as } n \to \infty.$$

Since (5) and (6) hold for $h(n) = n^{-\alpha}$ and $q(n) = n^{\alpha}$, respectively, where α is arbitrarily chosen from the interval $(0, 1)$, then the theorems say that under very mild conditions imposed on the sequences $\{h(n)\}$ and $\{q(n)\}$, the respective algorithms converge globally, in the mean square sense, to the nonlinearity μ with growing measurement number n.

3.4. Convergence rates

It is clear that the rate the algorithms converge with, corresponds to the their accuracy: the faster rate the better accuracy can be expected from the same measurement number. The theorems below unveil the convergence rates for exemplary algorithms:

Theorem 3 *Let m have p derivatives and let pth derivative be bounded in $[-1, 1]$. Let K in (2) be compactly supported and have p vanishing moments (see Appendix B). If*

$$h(n) \sim n^{-1/(2p+3)},$$

then, for any $0 < \varepsilon < 1$

$$E \int_{-(1-\varepsilon)}^{1-\varepsilon} [\hat{\mu}(u) - \mu(u)]^2 du = O\left(n^{-2p/(2p+3)}\right). \qquad (7)$$

Theorem 4 *Let m have p derivatives and let pth derivative be bounded in $[-1, 1]$. Let $\{\varphi_k\}$ in (3)-(4) be a trigonometric series (i.e.$\{\varphi_k(u) = e^{iku}; k = 0, \pm 1, \pm 2, ...\}$). If*

$$q(n) \sim n^{1/(2p+3)},$$

then, for any $0 < \varepsilon < 1$

$$E \int_{-(1-\varepsilon)}^{1-\varepsilon} [\tilde{\mu}(u) - \mu(u)]^2 du = O\left(n^{-2p/(2p+3)}\right). \qquad (8)$$

From the theorems one can infer that:

- Convergence rate is related to the smoothness of the non-linear characteristic: the smoother m (and hence μ) the faster algorithm's convergence rate and the higher the algorithm accuracy for the same number of measurement. For instance, for $p = 1$ the error is of order $O(n^{-2/3})$ and decreases to $O(n^{-6/7})$ for $p = 3$.
- For regular m with large p the rates approach the order $O(n^{-1})$ which is typical for parametric inference (i.e. for situation where the a priori information is very rich).
- The convergence rate is independent of the irregularity of the input probability density f. This property is the advantage of the presented algorithms over other known in the literature (see e.g. Greblicki and Pawlak (1986, 1987)) which convergence rates are worsened by irregular input probability density functions.
- The convergence rate $O(n^{-2p/(2p+1)})$ is asymptotically optimal, i.e. the best possible for algorithms with nonparametric a priori knowledge (see Stone (1982) or Härdle (1990)). It should be emphasized that for our algorithms this rate holds true in the presence of the correlated noise $\{\xi_n\}$.

Remark 1 *Convergence rates hold in the narrower interval* $(-1+\varepsilon, 1-\varepsilon)$ *rather than the entire one due to the well known (for nonparametric inference) boundary effects (see e.g. Härdle (1990)).*

Remark 2 *The conclusions reported above hold for large measurements sets. To the best of authors knowledge there are no formal results concerning their behaviour for small sample sets.*

4. Modifications of the algorithms

To calculate algorithms (2) and (3)-(4) one should compute definite integrals which can be time-consuming. Below we present their computationally simpler counterparts. The kernel algorithm has now the following form (see Greblicki (1996)):

$$\check{\mu}(u) = \frac{1}{h(n)} \sum_{j=1}^{n} Y_{[j]} (U_{(j)} - U_{(j-1)}) K\left(\frac{u - U_{(j)}}{h(n)}\right), \tag{9}$$

while the orthogonal algorithm takes the form (see Greblicki and Pawlak (1994a)):

$$\bar{\mu}(u) = \sum_{k=0}^{q(n)} \bar{c}_k \varphi_k(u), \tag{10}$$

with

$$\bar{c}_k = \sum_{j=1}^{n} Y_{[j][j]} (U_{(j)} - U_{(j-1)}) \varphi_k(U_{(j)}). \tag{11}$$

4.1. Convergence

For these algorithms the following theorems hold:

Theorem 5 *If*

$$h(n) \to 0 \text{ and } n^2 h^3(n) \to \infty \text{ as } n \to \infty \tag{12}$$

then

$$E \int_{-1}^{1} [\check{\mu}(u) - \mu(u)]^2 du \to 0 \text{ as } n \to \infty.$$

Theorem 6 *If*

$$q(n) \to \infty \text{ and } q^3(n)/n^2 \to 0 \text{ as } n \to \infty, \tag{13}$$

then

$$E \int_{-1}^{1} [\bar{\mu}(u) - \mu(u)]^2 du \to 0 \text{ as } n \to \infty.$$

Observe that the conditions imposed in (12) and (13) are more restrictive then they coun-terparts in (5) and (6), respectively. The algorithms are now consistent for $h(n) = n^{-\alpha}$ and $q(n) = n^{\alpha}$ where a is from the interval $(0, 2/3)$ only.

4.2. Convergence rates

As it can be expected (since these new algorithms are 'more rough'), the obtained convergence rate are slower than their origins:

Theorem 7 *Let m have p derivatives and let pth derivative be bounded in $[-1, 1]$. Let K in (2) be compactly supported and have p vanishing moments. If*

$$h(n) \sim n^{-1/(2p+3)},$$

then, for any $0 < \varepsilon < 1$,

$$E \int_{-(1-\varepsilon)}^{1-\varepsilon} [\check{\mu}(u) - \mu(u)]^2 du = O\left(n^{-2p/(2p+3)}\right). \tag{14}$$

Theorem 8 *Let m have p derivatives and let pth derivative be bounded in $[-1,1]$. Let $\{\varphi_k\}$ in (3)-(4) be a trigonometric series. If*

$$q(n) \sim n^{1/(2p+3)},$$

then, for any $0 < \varepsilon < 1$,

$$E\int_{-(1-\varepsilon)}^{1-\varepsilon} [\bar{\mu}(u) - \mu(u)]^2 du = O\left(n^{-2p/(2p+3)}\right). \quad (15)$$

To compare the convergence rates of the algorithms (9) and (10)-(11) with (2) and (3)-(4) assume that m has one bounded derivative. We get $O(n^{-2/5})$ versus the previously obtained optimal $O(n^{-2/3})$. However, for more regular m, for instance for m with $p = 3$, we get $O(n^{-2/3}) \sim O(n^{-0.67})$ and $O(n - 6/7) \sim O(n^{-0.86})$, respectively. In general, the difference between accuracy of these algorithms diminishes with increasing regularity of m.

5. Conclusions – accuracy of algorithms

Taking into consideration presented proper-ties, the following conclusions about the ac-curacy of algorithms can be drawn from the convergence conditions:

- The algorithms accuracy grows with increasing number of measurements for virtually all nonlinearities which can be met in practice.
- Accuracy can be arbitrarily high (the algorithm error can be made arbitrarily small) with growing n.
- It is not affected either by the regularity of the input density function f or the presence of the unknown dynamics (neither f nor $\{k_i\}$ has to be known to recover μ).

From the properties of the convergence rates we conclude additionally that:

- Accuracy grows with growing regularity of the non-linearity m — the higher p the smaller algorithms' errors for the same size of measurement set.
- For large n, the accuracy of the algorithms (2) and (3)-(4) is the best possible amongst all algorithms with non-parametric a priori information (accuracy of (9) and (10)-(11) is only slightly worse)
- In spite of the poor a priori knowledge convergence rates are not so far from $O(n^{-1})$, which is typical for parametric inference. (The rates of non-parametric algorithms are slower than those achieved by the parametric one due to fact, that the poorer *a priori* knowledge has to be compensated by the richer empirical knowledge, that is, the greater number of measurements.)

It should be finally emphasized that the proposed algorithms (especially the versions (9) and (10)-(11)) are computationally simple (they involve only elementary arithmetic operations). The latter, together with the properties examined earlier in the paper, make them a valuable proposition from the theory in all situations in which the a priori knowledge does not allow to propose a reliable model.

Appendix A

A function g is piecewise Lipschitz if for some r there exists a partition $\{A_i; i = 1, ..., r\}$ such that $A_1 \cup A_2 \cup \cdots U A_r = R$ in which g is Lipschitz; i.e. $|g(u) - g(u)| \leq < \alpha_i |u - v|$, $u, v \in A_i$ some α_i's.

Appendix B

The following kernel functions with compact supports can be used in the algorithms:

$$\left.\begin{array}{ll} \cdot \text{ rectangle:} & 1/2 \\ \cdot \text{ triangle:} & 1 - |u| \\ \cdot \text{ parabolic:} & 1 - u^2 \end{array}\right\} \times I_{\{\|u\| \leq 1\}}(u),$$

where I is the indicator function. Observe that the latter has $p = 1$ vanishing moments. To get, e.g., $p = 3$ one should use $[-(63/32)u^4 - (45/16)u^2 + 27/32]I_{\{\|u\| \leq 1\}}(u)$.

References

[1] Chang, F. H. I. and R. Luus (1971). A noniterative method for identification using Hammerstein model. *IEEE Transactions on Automatic Control* 16, 464-468.

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

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

[4] Greblicki, W. (1996). Nonlinearity estimation in Hammerstein systems based on ordered statistics. *IEEE Transactions on Signal Processing* 44, 1224-1233.

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

[6] Greblicki, W. and M. Pawlak (1987). Hammerstein system identification by non-parametric regression estimation. International Journal of Control 45, 343-354.

[7] Greblicki, W. and M. Pawlak (1989). Nonparametric identification of Hammerstein systems. *IEEE Transactions on Information Theory* 35, 409-418.

[8] Greblicki, W. and M. Pawlak (1994a). Dynamic system identification with order statistics. *IEEE Transactions on Information Theory.* 40, 1474-1489.

[9] Greblicki, W. and M. Pawlak (1994b). Nonlinear system identification with nonparametric deconvolution. In: *Proceedings of IEEE International Symposium on Information Theory.* Trondheim. p. 124.

[10] Györfi, L., M. Kohler, A. Krzy2ak and H. Walk (2002). *A Distribution-Free Theory of Nonparametric Regression.* Springer-Verlag. New York.

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

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

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

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

[15] Korenberg, M. J. and I. W. Hunter (1986). The identification of nonlinear biological system: LNL cascade models. *Biological Cybernetics* 55, 125-134.

[16] Krzyżak, A. (1989). Identification of discrete Hammerstein systems by the Fourier series regression estimate. *International Journal of System Science* 20, 1729-1744.

[17] Narendra, K. S. and P. G. Gallman (1966). An iterative method for the identification of nonlinear systems using the Hammerstein model. IEEE Transactions on Auto-matic Control 11, 546-550.

[18] Pawlak, M. (1991). On the series expansion approach to the identification of Hammer-stein systems. *IEEE Transactions on Automatic Control* 36, 763-767.

[19] Śliwiński, P. (2000). Non-linear system identification by wavelets. PhD dissertation. Wroclaw University of Technology. Institute of Engineering Cybernetics.

[20] Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Annals of Statistics* 10, 1040-1053.

[21] Thathachar, M. A. L. and S. Ramaswamy (1973). Identification of a class of nonlinear systems. *International Journal of Control* 18, 741-752.

[22] Zi-Qiang, L. (1993). Controller design ori-ented model identification method for Hammerstein system. *Automatica* 29, 767-771.