FAST ALGORITHMS FOR NON-LINEARITY RECOVERING IN HAMMERSTEIN SYSTEMS WITH ORDERED OBSERVATIONS

PRZEMYSŁAW ŚLIWIŃSKI

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

Abstract. In the paper an equivalent form of some non-parametric algorithms based on ordered observations is proposed. In presented algorithms versions, time-consuming definite integration is replaced by appropriate subtraction. A special treatment of wavelet-based algorithms is also proposed.

Key Words. system identification, non-linearity estimation, identification algorithm, order statistics.

1. INTRODUCTION

In works [3, 5], non-parametric algorithms estimating a non-linearity in Hammerstein systems (Fig. 1) have been proposed and examined. The algorithms are based either on kernel functions

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

or on orthogonal series expansions

$$\tilde{\mu}(u) = \sum_{k=0}^{q(n)} \tilde{c}_k \varphi_k(u), \; \tilde{c}_k = \sum_{j=1}^{n} Y_{[j]} \int_{U_{(j-1)}}^{U_{(j)}} \varphi_k(v)\, dv \quad (2)$$

In both algorithms, n is a number of measurements and $\left\{\left(U_{(j)}, Y_{[j]}\right)\right\}$ is a set of measurement pairs, sorted increasingly according to the input values while $\{h(n)\}$ and $\{q(n)\}$ are appropriate number sequences. Since numerical integrating is time consuming operation, for both types of algorithms their simpler counterparts have been also presented there, i.e.

$$\check{\mu}(u) = \frac{1}{h(n)} \sum_{j=1}^{n} Y_{[j]} K\left(\frac{u-U_{(j)}}{h(n)}\right) \left(U_{(j)} - U_{(j-1)}\right) \quad (3)$$

and

$$\bar{\mu}(u) = \sum_{k=0}^{q(n)} \bar{c}_k \varphi_k(u) \quad (4)$$

with

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

in which definite integrals are replaced by subtraction operations. The latter algorithms are obviously computationally simpler, however, as we see in the next section, they converge slower than the former ones.

In the paper we show that, in some cases, calculations of integrals in (1) and (2) can be replaced by appropriate subtractions, and hence that the complexity of these algorithms can be reduced to the order characteristic for simpler algorithms (3) and (4).

1.1. Algorithms' convergence rates

We start with a brief recall of convergence rate of the algorithms (1), (2), (3) and (4). They are established for the following assumptions (see [3, 5] and cf e.g. [4, 11, 10]):

A. The input signal $\{U_j; j = \cdots -1, 0, 1, 2, \cdots\}$ is a stationary white random process with an unknown probability density function f. We assume that $-1 \le U_j \le 1$ and that $f(u) \ge \delta > 0$.

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

C. Noise $\{Z_j; j = \cdots -1, 0, 1, 2, \cdots\}$ is a stationary white random process independent of the input signal and has zero mean and finite variance.

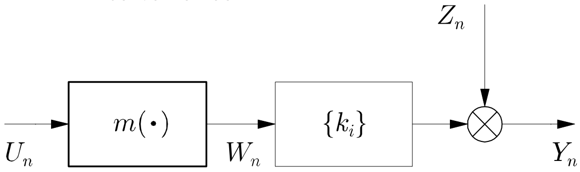

Fig. 1. Hammerstein system

Remark 1 *Since only input-output signals are available for measurements, one can identify the nonlinearity* $\mu(u) = k_0 m(u) + \mathrm{E}\, m(U_0) \sum_{i=1}^{\infty} k_i$ *rather than the genuine characteristic (cf. [4]).*

To compare convergence rates of both type of algorithms we recall the following theorems from the already cited papers. The first one characterizes convergence rate of algorithms (1) and (2) for p times differentiable non-linearities (with bounded last derivative) (cf. [3, 5])

Theorem 1 *Let K be a compactly supported kernel function with p vanishing moments and $\{\varphi_k\}$ be trigonometric series. If*

$$h(n) \sim n^{-\frac{1}{2p+1}} \text{ and } q(n) \sim n^{\frac{1}{2p+1}}$$

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

$$\mathsf{E} \int_{-(1-\epsilon)}^{1-\epsilon} [\mu(u) - \hat{\mu}(u)]^2 \, du = O\left(n^{-\frac{2p}{2p+1}}\right)$$

$$\mathsf{E} \int_{-(1-\epsilon)}^{1-\epsilon} [\mu(u) - \tilde{\mu}(u)]^2 \, du = O\left(n^{-\frac{2p}{2p+1}}\right)$$

respectively.

Note that the resulting rate is the same for both kernel and orthogonal algorithms and that it is the optimal one for non-parametric algorithms (see [14]). The next theorem provides with convergence rate of the simpler algorithms (cf. [3, 5]).

Theorem 2 *Let K be a compactly supported kernel function with p vanishing moments and $\{\varphi_k\}$ be trigonometric series as in the theorem above. If*

$$h(n) \sim n^{-\frac{1}{2p+3}} \text{ or } q(n) \sim n^{\frac{1}{2p+3}}$$

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

$$\mathsf{E} \int_{-(1-\epsilon)}^{1-\epsilon} [\mu(u) - \check{\mu}(u)]^2 \, du = O\left(n^{-\frac{2p}{2p+3}}\right)$$

$$\mathsf{E} \int_{-(1-\epsilon)}^{1-\epsilon} [\mu(u) - \bar{\mu}(u)]^2 \, du = O\left(n^{-\frac{2p}{2p+3}}\right)$$

respectively.

The theorems reveal in particular that the price we pay for simplicity of the latter algorithms is their slower convergence rate.

Example 1 *To compare the convergence rates assume that the non-linear characteristic m has one bounded derivative ($p = 1$). Then, applying the algorithms (1) or (2) we obtain the optimal rate $O\left(n^{-2/3}\right)$, and the worse rate $O\left(n^{-2/5}\right)$ is got for algorithms (3) or (4).*

2. MAIN RESULT

It is a well-know fact of calculus that if $G(u)$ is the indefinite integral of $g(u)$ being continuous on $[a, b]$, then

$$\int_a^b g(u)\, du = G(b) - G(a) \tag{5}$$

Hence the algorithms in (1) and (2) can be rewritten in the following forms

$$\hat{\mu}(u) = \frac{1}{h(n)} \sum_{j=1}^{n} Y_{[j]} \left[\kappa\left(\frac{u - U_{(j)}}{h(n)}\right) - \kappa\left(\frac{u - U_{(j-1)}}{h(n)}\right) \right] \tag{6}$$

and

$$\tilde{\mu}(u) = \sum_{k=0}^{q(n)} \tilde{c}_k \varphi_k(u) \tag{7}$$

with

$$\tilde{c}_k = \sum_{j=1}^{n} Y_{[j]} \left(\phi_k\left(U_{(j)}\right) - \phi_k\left(U_{(j-1)}\right)\right)$$

where $\kappa(u)$ and $\phi_k(u)$ are the indefinite integrals of functions $K(u)$ and $\varphi_k(u)$, respectively. Obviously, both algorithms (6) and (7) achieve now the numerical complexity of algorithms (3) and (4), preserving simultaneously the original (optimal) convergence rate of algorithms (1) and (2).
According to this observation, selected examples for kernel functions and orthogonal functions are given in the Table I in Appendix. Nevertheless, our simple solution cannot be directly applied to algorithms with:

1. Gaussian kernel (since its primitive is not given by finite formula),
2. Window kernel and Haar wavelet functions (because they are discontinuous).
3. Compactly supported orthogonal wavelet functions (due to the lack of explicit formulas),

Gaussian kernels are not considered in the paper. A special treatment for wavelet-based algorithm only is necessary since integration in algorithm (1) with window kernel and algorithm (2) with Haar functions can be done trivially.

2.1. WAVELET-BASED ALGORITHM

Application of wavelet-based versions of algorithms (1) and (2) has been proposed and investigated in [6]

$$\tilde{\mu}(u) = \sum_{l=l_{\min}(u)}^{l_{\max}(u)} \tilde{c}_{q(n),l}\varphi_{q(n),l}(u) \tag{8}$$

with

$$\tilde{c}_{q(n),l} = \sum_{j=1}^{n} Y_{[j]} \int_{U_{(j-1)}}^{U_{(j)}} \varphi_{q(n),l}(v)\, dv$$

where $\varphi_{ql}(u) = 2^{q/2}\varphi(2^q u - l)$ are dyadic dilations and integer translations of wavelet scaling function $\varphi(u)$, and the $l_{\min}(u)$ and $l_{\max}(u)$ are appropriate limits (see Table II in Appendix). It was shown in [6] that:

Theorem 3 *Let $\varphi(u)$ be Daubechies scaling function with wavelet number p. If*

$$2^{q(n)} \sim n^{\frac{1}{2p+1}}$$

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

$$\mathsf{E} \int_{-(1-\epsilon)}^{1-\epsilon} [\mu(u) - \tilde{\mu}(u)]^2\, du = O\left(n^{-\frac{2p}{2p+1}}\right)$$

That is, the convergence rate of the wavelet algorithm also attains the optimal rate. Observe that algorithm (8) is of theoretical value only due to following properties of Daubechies wavelet functions:

1. They are not given explicitly but as the limits of recursive formulas and, moreover,
2. Wavelet functions can be known exactly only in binary mesh points.

Hence, as it was pointed out in e.g. [2, 7, 9, 12, 13], wavelet-based algorithms cannot be applied directly when input signals (and hence wavelet function arguments) are randomly distributed. In [12, 13], a solution based on replacement of wavelet functions by their piecewise constant interpolations has been proposed and examined. Basing on the same approach, we obtain the following counterpart of the algorithm (8):

$$\breve{\mu}(u) = \sum_{l=l_{\min}(u)}^{l_{\max}(u)} \breve{c}_{q(n),l}\bar{\varphi}^{H}_{q(n),l}(u) \tag{9}$$

with

$$\breve{c}_{q(n),l} = \sum_{j=1}^{n} Y_{[j]} \int_{U_{(j-1)}}^{U_{(j)}} \bar{\varphi}^{H}_{q(n),l}(v)\, dv$$

where $\bar{\varphi}^{H}(u)$ is piecewise constant interpolation of wavelet scaling function $\varphi(u)$, sampled at the mesh 2^{-H} $(H \in \mathbf{Z})$. The integrations in (9) can be now replaced by following subtractions

$$\breve{c}_{q(n),l} = \sum_{j=1}^{n} Y_{[j]} \left(\bar{\phi}^{H}_{ql}\left(U_{(j)}\right) - \bar{\phi}^{H}_{ql}\left(U_{(j-1)}\right)\right)$$

with $\bar{\phi}^{H}_{ql}(u) = 2^{q/2}\bar{\phi}(2^q u - l)$ and $\bar{\phi}^{H}(u)$ defined as

$$\bar{\phi}^{H}(u) = \int_{0}^{2^H u} \bar{\varphi}^{H}(v)\, dv.$$

Note that the latter is easy to compute due to piecewise constant nature of $\bar{\varphi}^{H}(u)$.

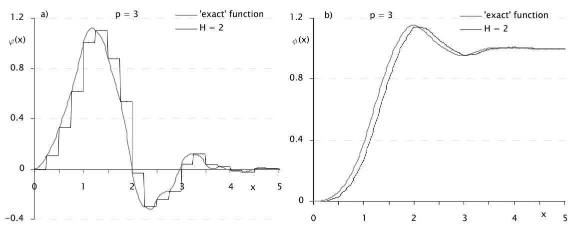

Fig. 2 a) Third Daubechies scaling function $\varphi(u)$ and its interpolation $\bar{\varphi}^2(u)$. b) Appropriate integrals $\phi(u)$ and $\bar{\phi}^2(u)$

Remark 2 *Substituting genuine wavelet functions by their piecewise constant counterparts induces an additional error of the resulting algorithm (9) in comparison with the original wavelet-based one (which is clearly revealed in Fig. 2). In [12, 13], the following condition has been imposed on sampling resolution H to guarantee that the interpolated algorithm preserves the properties of the one with 'exact' wavelet functions*

$$H(n) = \left\lceil \frac{p}{\eta} q(n) \right\rceil$$

where p is wavelet number of applied scaling function $\varphi(u)$ and η is its Hölder exponent (see Table II in Appendix).

3. CONCLUSIONS

We observed that identification algorithms (1) and (2) proposed in [3, 5, 6] for different orthogonal series and kernel functions, can be rewritten in the equivalent form in which definite integrals are replaced by appropriate subtractions. These new versions of the algorithms preserve the same fastest possible convergence rate and moreover have numerical complexity of their counterparts (3) and (4) characterized by slower convergence rate. For Daubechies wavelet-based algorithms an approximate solution has been proposed and its properties have been shortly commented.

APPENDIX: EXAMPLES OF KERNEL FUNCTIONS AND ORTHOGONAL SERIES

In Table I, popular kernel functions, and selected orthogonal sets are presented together with their integrals. Table II contains basic properties of the Daubechies functions or symmlets (see e.g. [1]) which can be applied in wavelet version (8) of the generic orthogonal series algorithm (2).

	kernel function $K(u)$	corresponding integral $\kappa(u)$
Triangle kernel	$(1-\|u\|)\, I_{[\|u\|\le 1]}(u)$	$(2u - u\,\|u\| + 1)\, I_{[\|u\|\le 1]}(u)$
Epanechnikov kernel	$\frac{3}{4}(1-u^2) I_{[\|u\|\le 1]}(u)$	$\frac{1}{4}\left(-u^3+3u+2\right) I_{[\|u\|\le 1]}(u)$
Kernel from [8]	$\frac{1}{8}\left(9-15u^2\right) I_{[\|u\|\le 1]}(u)$	$\frac{1}{8}\left(-5u^3+9u+4\right) I_{[\|u\|\le 1]}(u)$
Cauchy kernel	$\pi^{-1}\left(1+u^2\right)^{-1}$	$\frac{1}{2}+\pi^{-1}\arctan(u)$
	orthogonal set $\{\varphi_k(u)\}$	corresponding integrals $\{\phi_k(u)\}$
Fourier series	$e^{-ik\pi u}$	$\frac{i}{k\pi} e^{-ik\pi u}$
Legendre polynomials	$\frac{\sqrt{2k+1}}{2\sqrt{2}k!} \frac{d^k}{du^k}\left(u^2-1\right)^k$	$\frac{\sqrt{2k+1}}{2\sqrt{2}k!} \frac{d^{k-1}}{du^{k-1}}\left(u^2-1\right)^k$
cosine series	$\cos k\pi u$	$\frac{1}{k\pi}\sin k\pi u$
sine series	$\sin k\pi u$	$-\frac{1}{k\pi}\cos k\pi u$

Table I. Popular kernels and orthogonal series and their integrals

	pth Daubechies function/symmlet
$l_{\min}(u)$	$\lfloor 2^k u \rfloor - 2p + 2$
$l_{\max}(u)$	$\lceil 2^k u \rceil - 1$
supports of $\varphi(u)$	$[0, 2p-1]$
Hölder exponents of $\varphi(u)$	~ 0.55 for $p=1$ and 1 for $p>1$

Table II. Basic properties of exemplary wavelet scaling functions (p – wavelet number)

REFERENCES

[1] A. Cohen. *Numerical analysis of wavelets methods.* Studies in Mathematics and Its Applications. Elsevier, Amsterdam, 2003.

[2] B. Delyon and A. Juditsky. Estimating wavelet coefficients. In A. Antoniadis and G. Oppenheim, editors, *Wavelets and Statistics*, pages 151–168. Springer-Verlag, New York, 1995.

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

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

[5] W. Greblicki and M. Pawlak. Dynamic system identification with order statistics. *IEEE Transactions on Information Theory*, 40:1474–1489, 1994.

[6] W. Greblicki and P. Śliwiński. Nonparametric estimation of non-linearities in Hammerstein systems. Submitted to *Automatica*, 2004.

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

[8] W. Härdle. *Applied Nonparametric Regression.* Cambridge University Press, Cambridge, 1990.

[9] W. Härdle, G. Kerkyacharian, D. Picard, and A. Tsybakov. *Wavelets, Approximation, and Statistical Applications.* Springer-Verlag, New York, 1998.

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

[11] M. Pawlak and Z. Hasiewicz. Nonlinear system identification by the Haar multiresolution analysis. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 45:945–961, 1998.

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

[13] P. Śliwiński and Z. Hasiewicz. Computational algorithms for multiscale identification of nonlinearities in Hammerstein systems with random inputs. Accepted for publication in: *IEEE Transactions on Signal Processing*, 2004.

[14] C. J. Stone. Optimal global rates of convergence for nonparametric regression. *Annals of Statistics*, 10:1040–1053, 1982.