

ON-LINE WAVELET ESTIMATION OF HAMMERSTEIN SYSTEM NONLINEARITY

PRZEMYSŁAW ŚLIWIŃSKI

Institute of Computer Engineering, Control and Robotics
Wrocław University of Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
e-mail: przemyslaw.sliwinski@pwr.wroc.pl

A new algorithm for nonparametric wavelet estimation of Hammerstein system nonlinearity is proposed. The algorithm works in the on-line regime (viz., past measurements are not available) and offers a convenient uniform routine for nonlinearity estimation at an arbitrary point and at any moment of the identification process. The pointwise convergence of the estimate to locally bounded nonlinearities and the rate of this convergence are both established.

Keywords: Hammerstein systems, on-line nonparametric identification, wavelet estimates, convergence analysis.

1. Introduction

In the paper we present a novel nonparametric identification algorithm based on orthogonal wavelet expansion and recovering Hammerstein system nonlinearity in an on-line fashion. In system identification, on-line algorithms are of interest for several well-known reasons, e.g.,

- *No need for measurements storing.* The number of measurements in nonparametric algorithms typically exceeds thousands and keeping them in memory may be a prohibitive overhead in some embedded system applications.
- *Recursive formulas.* Usually simpler and less computationally demanding than their closed form counterparts, recursive formulas are favored in applications with limited computational capabilities (and/or power constrained).

Various approaches to the problem are considered in the literature (Rutkowski, 1980; Greblicki and Pawlak, 2008, Chapters 4–5). Kernel and orthogonal series semirecursive algorithms were proposed by Greblicki and Pawlak (1989) (for Hammerstein systems) and Rutkowski (1980; 2004) (for memoryless nonlinear systems), respectively. These algorithms are of the quotient form and the term *semirecursive* is used to emphasize that both the numerator and the denominator are computed recursively, but not the quotient itself. Fully *recursive* algorithms, exploiting the stochastic approximation scheme, were in-

troduced by Greblicki (2002). The *recursive* order statistics algorithm based on the Haar kernel was recently proposed by Śliwiński and Hasiewicz (2009).

In all these algorithms, due to their kernel form, the estimation points need to be set at the very beginning of the identification routine. Otherwise, if the new estimation points are to be added later in the course of the identification experiment, then they need to be either computed from scratch or some indirect scheme like, e.g., an interpolation algorithm (see, e.g., the work of Unser (1999)) should be applied.

Our wavelet algorithm is also *semirecursive*. However, in contrast to the kernel-like orthogonal series algorithms presented, e.g., by Rutkowski (1980; 2004), we directly exploit the series expansion formula and the main idea of the algorithm consists in recursive estimation of the successively incorporated wavelet orthogonal expansion coefficients. The approach is inspired by the concept of transform coding (exploited intensively in data compression (Donoho *et al.*, 1998)), where the target nonlinearity is (globally) represented by the empirical orthogonal series coefficients rather than (locally) by individual points. As a result, the algorithm offers the following practical benefits:

- It allows effective *direct* evaluation of the estimate values in arbitrarily selected points at any moment of the identification process. This is the main advantage over kernel on-line algorithms, where, as has been mentioned, convenient computations can

only be performed at (separate) preselected points (Greblicki and Pawlak, 1989; Greblicki, 2002; Chen, 2004; Zhao and Chen, 2006).

- Wavelet bases employ only *two* functions, the father and mother wavelets. This in turn is an advantage over on-line polynomial and trigonometric algorithms, where the number of basis polynomial functions increases with the number of measurements (Rutkowski, 1980; Greblicki and Pawlak, 1989; Györfi *et al.*, 2002).

Moreover (Daubechies, 1992; Mallat, 1998),

- wavelet approximations offer the most parsimonious representations for many important classes of nonlinearities (Donoho, 1993), which yields fast convergence rates,
- multiscale (multiresolution) wavelet analysis finds an elegant and intuitive theoretical framework for on-line algorithms, and, finally,
- the compactness of wavelet function supports reduces the computation burden due to local processing of measurement data (Śliwiński and Hasiewicz, 2008).

The proposed type of on-line algorithm has not yet been explored in the literature (a similar, order statistics, algorithm used by Śliwiński *et al.* (2009) is also *semirecursive* and based on the estimation of the expansion coefficients but requires all measurements to be stored and thus remains inapplicable in on-line conditions) and can be seen as a direct on-line counterpart of off-line wavelet algorithms examined, e.g., by Śliwiński and Hasiewicz (2008) as well as Hasiewicz *et al.* (2005). Limit properties of the algorithm, viz. convergence conditions and convergence rates, are shown to be the same as those of off-line algorithms, cf. the results of Greblicki and Pawlak (1986) or Hsu *et al.* (2008).

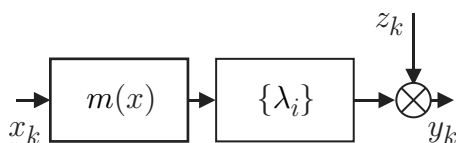


Fig. 1. Identified Hammerstein system.

The Hammerstein system (Fig. 1), being a cascade of a memoryless nonlinearity followed by a linear dynamics, is a popular nonlinear system modeling tool and has already found applications in several areas, e.g., automatic control, signal processing, economy, chemistry, and biomedical engineering, see the works of Coca and Billings (2001), Chen *et al.* (1989), Srinivasan *et al.* (2005), Nordso and Zetterberg (2001), Giannakis and Serpedin (2001) and the exhaustive reference set therein).

Example 1. An interesting application of the Hammerstein system in telecommunications can be found in the work of Kang *et al.* (1999), where the system is used in tandem with the Wiener one (cf. the results of Greblicki (2001)) for the recursive identification algorithm of Wiener systems) to linearize a high-power amplifier (HPA) in the orthogonal frequency division multiplexing (OFDM) system. Various applications in biomedical engineering, like, e.g., in sensory systems, reflex loops, organ systems, and tissue mechanics, are demonstrated by Westwick and Kearney (2003, Chapters 6–8). ♦

2. Problem statement

Our goal is to recover on-line the Hammerstein system nonlinearity from input-output data pairs (x_k, y_k) , $k = 1, 2, \dots$, arriving and being processed sequentially in time, under the assumptions typical for nonparametric identification tasks (Greblicki and Pawlak, 1989; Hasiewicz and Śliwiński, 2002; Lang, 1997; Mzyk, 2007; Pawlak and Hasiewicz, 1998; Śliwiński *et al.*, 2009):

1. The nonlinear characteristic, $m(x)$, is an arbitrary locally bounded function.
2. The dynamic subsystem is asymptotically stable. Its impulse response, $\{\lambda_i, i = 0, 1, \dots\}$, is unknown.
3. The system is driven by a random i.i.d. signal, $\{x_k, k = \dots, -1, 0, 1, \dots\}$, with a bounded probability density function $f(x)$.
4. The external disturbance, represented in the form $\{z_k, k = \dots, -1, 0, 1, \dots\}$, is a zero mean random (independent of the input) noise of an arbitrary distribution with a finite variance.

Note that, under the assumptions above, virtually any nonlinearity in the Hammerstein system is admissible. For instance, the characteristic $m(x)$ can be differentiable or not, continuous or piecewise-continuous with jumps (in particular, it can be a polynomial of any order). The impulse response, $\{\lambda_i\}$, can be finite or not (it can have, e.g., damped oscillation terms, cf. Assumption 2). Similarly, the input x_k can be of arbitrary density while the noise, z_k , of any distribution (with a variance). The latter can also be white or correlated.

We emphasize that, due to the *on-line* design restrictions, the past measurements are not stored and only the current measurement pair, (x_k, y_k) , is available for processing.

Remark 1. Under the fully nonparametric assumptions 1–4, not the genuine characteristic $m(x)$ but a nonlinearity $\mu(x) = \lambda_0 m(x) + \zeta$ (where $\zeta = Em(x_1) \cdot \sum_{i>0} \lambda_i$ is a system dependent constant) can at most be identified

from input-output data (Greblicki and Pawlak, 1986). Indeed, since the following holds:

$$\begin{aligned} E(y_k | x_k = x) &= \lambda_0 m(x) + Ez_k + E \sum_{i>0} \lambda_i m(x_{k-i}) \\ &= \lambda_0 m(x) + \zeta, \end{aligned}$$

the true nonlinearity $m(x)$ can only be recovered with additional *a priori* information, e.g., the parametric knowledge about the subsystems (Chen, 2005; Zhao and Chen, 2006; Mzyk, 2009), or with the help of the *active experiment* approach, viz. with a controlled input signal x_k (Chen, 2004, Remark 3). Specifically, if it is known that $m(0) = 0$ (which is often the case in practice), then $\zeta = \mu(0)$, and, to eliminate this additive constant (and to recover scaled-only version $\lambda_0 m(x)$), it suffices to estimate $\mu(x) - \mu(0)$.

3. On-line algorithm

Let $\varphi(x)$ and $\psi(x)$ be compactly supported scaling and wavelet functions whose dilations and translations, $\varphi_{Mn}(x) = 2^{M/2} \varphi(2^M x - n)$ and $\psi_{mn}(x) = 2^{m/2} \psi(2^m x - n)$ for $m = M, M + 1, \dots$, and $n = \dots, -1, 0, 1, \dots$, constitute an orthogonal basis of the space of square integrable functions (Daubechies, 1992, Chapter 6). Both $\varphi(x)$ and of $\psi(x)$ have the support $[0, s]$, for some natural s . Denote by $g(x)$ the product $\mu(x) \cdot f(x)$, and by α_{Mn} and β_{mn} its expansion coefficients associated with basis functions $\varphi_{Mn}(x)$ and $\psi_{mn}(x)$. Similarly, by a_{Mn} and b_{mn} , we denote the corresponding coefficients of the input density function $f(x)$.

The identification algorithm is comprised of two computationally independent *subroutines*:

1. The estimation of expansion coefficients, and
2. The estimation of the nonlinearity.

The *first subroutine* is performed after each arrival of the new measurement pair (x_k, y_k) and is based on the observation that, under Assumptions 1–4, the coefficients α_{Mn} , a_{Mn} , β_{mn} and b_{mn} are equal to the respective expectations (Greblicki, 1989; Hasiewicz *et al.*, 2005):

$$\begin{aligned} \alpha_{Mn} &= E \{ \varphi_{Mn}(x_1) \times y_1 \}, \\ a_{Mn} &= E \{ \varphi_{Mn}(x_1) \}, \end{aligned} \tag{1a}$$

and

$$\begin{aligned} \beta_{mn} &= E \{ \psi_{mn}(x_1) \times y_1 \}, \\ b_{mn} &= E \{ \psi_{mn}(x_1) \}, \end{aligned} \tag{1b}$$

and can be recursively estimated by the formulas proposed below—referred further to as the *empirical coefficients*

(we will use a vector-like convention for conciseness of presentation):

$$\underbrace{\begin{bmatrix} \hat{\alpha}_{Mn}^{(k)} \\ \hat{a}_{Mn}^{(k)} \end{bmatrix}}_{\text{for } k>0} = \frac{k-1}{k} \underbrace{\begin{bmatrix} \hat{\alpha}_{Mn}^{(k-1)} \\ \hat{a}_{Mn}^{(k-1)} \end{bmatrix}}_{\text{for } k>0} + \frac{1}{k} \begin{bmatrix} \varphi_{Mn}(x_k) \cdot y_k \\ \varphi_{Mn}(x_k) \end{bmatrix}, \tag{2a}$$

and

$$\underbrace{\begin{bmatrix} \hat{\beta}_{mn}^{(k)} \\ \hat{b}_{mn}^{(k)} \end{bmatrix}}_{\text{for } k>\kappa_m} = \frac{k-1}{k} \underbrace{\begin{bmatrix} \hat{\beta}_{mn}^{(k-1)} \\ \hat{b}_{mn}^{(k-1)} \end{bmatrix}}_{\text{for } k>\kappa_m} + \frac{1}{k} \begin{bmatrix} \psi_{mn}(x_k) \cdot y_k \\ \psi_{mn}(x_k) \end{bmatrix}, \tag{2b}$$

where $\{\kappa_m\}$ is an increasing number sequence.

At the beginning, the estimate consists of only scaling function coefficients $\hat{\alpha}_{Mn}^{(k)}$, $\hat{a}_{Mn}^{(k)}$. In the course of identification, when the number of processed measurements k exceeds the consecutive threshold values κ_m associated with the scale m , the new wavelet empirical coefficients, $\hat{\beta}_{mn}^{(k)}$, $\hat{b}_{mn}^{(k)}$, are successively added to the estimate. Initially, all the empirical coefficients are zero (inactive), i.e., $\hat{\alpha}_{Mn}^{(0)} = \hat{a}_{Mn}^{(0)} = \hat{\beta}_{mn}^{(0)} = \hat{b}_{mn}^{(0)} = 0$, for all $m = M, M + 1, \dots$

Note that a faster growth of elements of $\{\kappa_m\}$ implies a slower pace of the wavelet coefficients incorporation (since the coefficients are added after a larger number of the processed measurements k) and one can expect poor approximation properties of the estimate. Conversely, a slowly growing sequence admits earlier coefficients inclusion and results in a larger estimate variance. A properly balanced (and asymptotically optimal) rate of the growth of the thresholding sequence $\{\kappa_m\}$ is formally established in Section 3.1.

Remark 2. From the asymptotic properties viewpoint (Appendix), the initial value of M can be an arbitrary natural number and in applications one can set it using the off-line scale selection rule, $M = \frac{1}{2\nu+1} \log_2 k$ (Hasiewicz *et al.*, 2005). If, for instance, the estimation routine starts with the first measurement pair (i.e., for $k = 1$), the rule yields $M = 0$.

Remark 3. The wavelet coefficients, $\hat{\beta}_{mn}^{(k)}$, $\hat{b}_{mn}^{(k)}$, when appearing in the course of identification, cannot be initialized with past measurements (as they are missing in the on-line régime) and are simply set to zero. This makes them *biased estimates* (see the formula (12) in Appendix) but simultaneously contributes to *computational simplicity* of the algorithm (in particular, the same current value k is used at the same time in all active empirical coefficients calculations). Zero-initialization can also be motivated by

the observation that wavelet coefficients decay with the growing scale m as fast as $\mathcal{O}(2^{-m/2})$ for any nonlinearities (and even faster for smoother ones, e.g., for differentiable nonlinearities the vanishing rate raises to the order $\mathcal{O}(2^{-3m/2})$) (Hasiewicz *et al.*, 2005; Mallat, 1998, Chapter 9).

The *other subroutine* evaluates the nonlinearity estimate (Greblicki, 1989; Hasiewicz *et al.*, 2005):

$$\hat{\mu}^{(k)}(x) = \frac{\hat{g}^{(k)}(x)}{\hat{f}^{(k)}(x)}, \quad (3)$$

where the numerator, $\hat{g}^{(k)}(x)$, and the denominator, $\hat{f}^{(k)}(x)$, are respectively wavelet on-line estimates of $g(x)$ and $f(x)$ comprising only an already active scaling function and wavelet empirical coefficients:

$$\begin{aligned} \begin{bmatrix} \hat{g}^{(k)}(x) \\ \hat{f}^{(k)}(x) \end{bmatrix} &= \underbrace{\sum_n \varphi_{Mn}(x) \cdot \begin{bmatrix} \hat{\alpha}_{Mn}^{(k)} \\ \hat{a}_{Mn}^{(k)} \end{bmatrix}}_{2^M x - s < n < 2^M x} \\ &+ \sum_{m \geq M} \sum_n \underbrace{\psi_{mn}(x) \cdot \begin{bmatrix} \hat{\beta}_{mn}^{(k)} \\ \hat{b}_{mn}^{(k)} \end{bmatrix}}_{2^m x - s < n < 2^m x}. \end{aligned} \quad (4)$$

The algorithm *subroutines* are computationally independent and can be intertwined. Hence, $\hat{\mu}^{(k)}(x)$ can be directly and easily evaluated from (3)–(4) at an arbitrary point x at any moment of identification (i.e., after any number k of measurements processed by (2)). This distinct property distinguishes our algorithm from kernel on-line algorithms, where $\hat{\mu}^{(k)}(x)$ can effectively be estimated only in fixed *a priori* known points (Greblicki, 1989; 2002; Rutkowski, 1980).

Remark 4. Although this property comes at the cost of keeping the empirical coefficients in memory, that cost seems to be moderate: in a practical situation, where the measurements are from a finite interval, the number of active coefficients is merely of order $\mathcal{O}(2^{\nu+1}\sqrt{k})$ (Hasiewicz and Śliwiński, 2002, Eqn. (63)), that is, even for differentiable nonlinearity (viz., for $\nu = 1$) this number is of order $\mathcal{O}(\sqrt[3]{k})$, which is significantly smaller than the number of all measurements k (that would have been stored by kernel algorithms to offer the same functionality).

If one is interested in the estimation of nonlinearity only at preselected points, then, for each of such points, the stochastic approximation algorithm can be employed. The algorithm, in the wavelet version, is of the following form (Greblicki, 2002; Greblicki and Pawlak, 2008, Chapter 5):

$$\begin{aligned} \bar{\mu}^{(k)}(x) &= \bar{\mu}^{(k-1)}(x) \\ &+ \gamma_k \vartheta_K(x, x_k) \cdot [y_k - \bar{\mu}^{(k-1)}(x)], \end{aligned} \quad (5)$$

where $\gamma_k \vartheta_K(x, u)$ is a random gain, and $\vartheta_K(x, u)$ is the kernel of the wavelet orthogonal series (Walter and Shen, 2001, Chapter III):

$$\vartheta_K(x, u) = \sum_{2^K x - s < n < 2^K x} \varphi_{Kn}(x) \varphi_{Kn}(u).$$

In view of the results presented by Greblicki (2002) as well as Greblicki and Pawlak (2008, Chapter 4), one can, however, expect $\bar{\mu}^{(k)}(x)$ to have a convergence rate slightly worse than ours.

Furthermore, observing that the recursive formula for $\hat{\alpha}_{Mn}^{(k)}$ in (2) can be written in the equivalent form (cf. (5))

$$\hat{\alpha}_{Mn}^{(k)} = \hat{\alpha}_{Mn}^{(k-1)} + \frac{1}{k} [\varphi_{Mn}(x_k) \cdot y_k - \hat{\alpha}_{Mn}^{(k-1)}],$$

one can still consider the estimation of the expansion coefficients with the help of the stochastic approximation scheme (Chen, 2004):

$$\hat{\alpha}_{Mn}^{(k)} = \hat{\alpha}_{Mn}^{(k-1)} + \gamma_k [\varphi_{Mn}(x_k) \cdot y_k - \hat{\alpha}_{Mn}^{(k-1)}].$$

In our algorithm we have used the simple case, $\gamma_k = 1/k$, mainly due to the direct correspondence to its off-line prototype and because of some practical reasons as well: the main advantage of the application of the stochastic approximation algorithm in, e.g., the works of Greblicki (2002) or Greblicki and Pawlak (2008, Chapter 4), is that the resulting algorithm has no quotient form, and—*a fortiori*—there are no random variables in its denominator (see (5)). The application of the stochastic approximation formula to expansion coefficients estimation would rather not eliminate the quotient form of our algorithm.

3.1. Algorithm properties. Before the presentation of the limit properties, we describe and briefly discuss some of the algorithm features resulting from the application of the wavelet basis:

- *Compact support and local data processing:* In the first subroutine of the algorithm, the wavelet empirical coefficients, $\hat{\beta}_{mn}^{(k)}, \hat{b}_{mn}^{(k)}$, $m = M, M + 1, \dots$, are effectively zero and remain inactive (i.e., are not taken into account during actual computations) until the following two conditions are simultaneously satisfied: (i) the number of overall measurements, k becomes greater than the threshold value κ_m , corresponding to their scale m and (ii) at least one of the incoming inputs, x_k , $k > \kappa_m$, falls in the supports of the associated wavelet functions $\psi_{mn}(x)$, see (2). Similarly, the evaluation of the nonlinearity estimate, $\hat{\mu}^{(k)}(x)$, for a given x , in the algorithm's second subroutine involves only the non-zero coefficients whose supports of respective scaling functions, $\varphi_{Mn}(x)$, and wavelets, $\psi_{mn}(x)$, include x , see the summation limits in (4) and the work Mallat (1998, p. 174).

- **Multiresolution (zooming) property:** The scaling empirical coefficients, $\hat{\alpha}_{Mn}^{(k)}, \hat{a}_{Mn}^{(k)}$, convey information about a crude, initial approximation of the identified nonlinearity, while their wavelet counterparts, $\hat{\beta}_{mn}^{(k)}, \hat{b}_{mn}^{(k)}$, successively created and updated with each new measurement pair, (x_k, y_k) , bring to the estimate the details located at finer-and-finer resolutions (scales) $m = M, M + 1, \dots$, see (2) and (4) and the work of Mallat (1998, Chapter VI).

We emphasize that only a single pair of wavelet functions, $\varphi(x)$ and $\psi(x)$, is used in computations. This property results in uniformed formulas (2) of empirical coefficients for all M, m and n , and significantly simplifies algorithm implementation, see Remark 3.

The following theorems add to the comments above the formal characterization of the algorithm. The first one describes the algorithm convergence conditions.

Theorem 1. *Let Assumptions 1–4 hold. For any $c > 1$, the threshold sequence*

$$\{\kappa_m = 2^{cm}, m = M, M + 1, \dots\} \quad (6)$$

makes the algorithm (2)–(4) converge in probability to the nonlinearity $\mu(x)$ with the increasing number of processed measurements k . The convergence holds in all points where $g(x)$ and $f(x) > 0$ are continuous.

The next theorem shows how to select the threshold sequence $\{\kappa_m\}$ depending on the smoothness of the nonlinearity $\mu(x)$, the input probability density function $f(x)$, and the employed wavelet series

Theorem 2. *Let Assumptions 1–4 hold. The threshold sequence selected as*

$$\{\kappa_m = 2^{(2\nu+1)m}, m = M, M + 1, \dots\}, \quad (7)$$

i.e., for $c = 2\nu + 1$, it yields the following rate of convergence:

$$\hat{\mu}^{(k)}(x) = \mu(x) + \mathcal{O}(k^{-\nu/(2\nu+1)}) \quad (8)$$

in probability, where $\nu = \min\{\nu_\mu, \nu_f, p\}$, while ν_μ, ν_f are the Hölder smoothness exponents of $\mu(x)$ and $f(x)$, respectively (Mallat, 1998, Chapter IX), and p is the number of vanishing moments of the mother wavelet $\psi(x)$ used in the algorithm.

Proof. See Appendix. ■

Remark 5. In Theorem 1 we require local boundedness of the nonlinearity $\mu(x)$. The same convergence condition as in (6) can also be obtained when it is known that the nonlinearity (along with the input density functions) is square integrable. On the other hand, if the ν_μ -th and ν_f -th derivatives of $\mu(x)$ and $f(x)$ are not locally bounded (as in

Theorem 2) but square integrable, then we can expect the convergence rate of order $\mathcal{O}(k^{-(\nu+1/2)/2\nu})$, that is, somewhat slower (cf. the results established for similar kernel on-line algorithms by Greblicki and Pawlak (2008, Chapter 4, Table 6.1)).

Additional properties of the algorithm can further be concluded from the theorems:

- **Local smoothness and convergence rate.** The convergence speeds up with growing $\nu_g = \min\{\nu_\mu, \nu_f\}$, i.e., with growing regularity of the product $g(x)$ (as $\lceil \nu_g - 1 \rceil$ is the number of continuous derivatives of $g(x)$, see the work of Mallat (1998, Chapter 9)). That is, the smoother the nonlinearity $\mu(x)$ and the input density $f(x)$ in the neighborhood of x , the faster the convergence there can be.
- **Vanishing moments and the asymptotically optimal rate of convergence.** For given $\mu(x)$ and $f(x)$, the convergence rate can be controlled by wavelets with an adequate number of vanishing moments. For instance, if $\nu_g = 3$, then for Haar wavelets (with only $p = 1$ vanishing moments), $\nu = 1$, and the rate does not exceed $\mathcal{O}(k^{-1/3})$, cf. (8). For Daubechies wavelets with $p = 3$, it reads $\nu = 3$, and the rate increases to $\mathcal{O}(k^{-3/7})$ but does not further accelerate for $p > 3$, cf. Fig. 2(b). In general, if $p \geq \nu_g$, then the asymptotic convergence rate is optimal, i.e., the fastest possible amongst all nonparametric algorithms (Stone, 1980).
- **Robustness of the rate.** Neither the structure of the dynamic subsystem nor the amplitude, distribution, or correlation of the external disturbances influence the convergence rate.

3.2. Experimental illustration. In order to illustrate the properties of the proposed algorithm, two representative nonlinearities, the quantizer $m(x) = \lfloor 8x - 4 \rfloor / 4 + 1/8$ and the polynomial $m(x) = 10(2x^3 - 3x + x)$, were estimated by four algorithms based on Daubechies wavelet families with $p = 1, 2, 3$ and 4 vanishing moments. The input x_k was of the uniform distribution over the unit interval and the dynamic system had the infinite impulse response, $\lambda_i = 2^{-i}$, $i = 0, 1, \dots$ (note that $\mu(x) = m(x)$ with this setting, see the work of Greblicki and Pawlak (1986)). The noise had the uniform distribution with the amplitude set such that $\text{NSR} \{m/z\} \triangleq \max_k |z_k| / \max_x |m(x)| = 0.1$.

The diagrams show the absolute error, $|\mu(x) - \hat{\mu}^{(k)}(x)|$, cf. (8), averaged over 500 points $x_i = i/500$, $i = 1, \dots, 500$, in 100 independent runs of the experiment. The diagram in Fig. 2(a) reveals that, for the quantizer nonlinearity, increasing the number of vanishing moments can decrease the performance of the algorithm

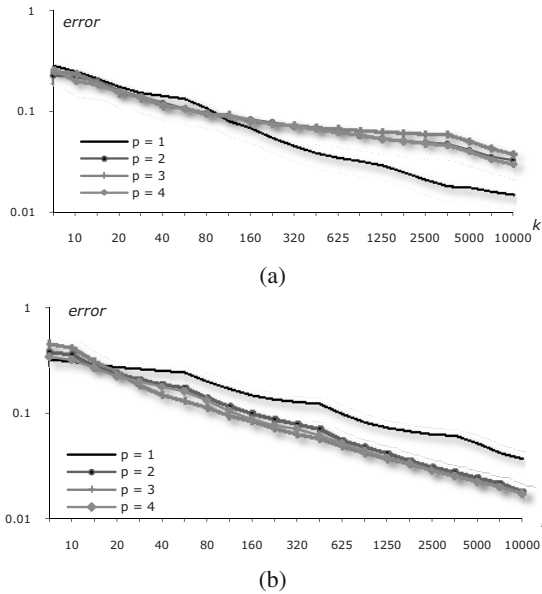


Fig. 2. Influence of the number of vanishing moments on the convergence rate for the quantizer (a) and polynomial nonlinearities (b).

(which is particularly visible in the example since wavelets with $p = 1$ are Haar wavelets and they recover well quantizer-like nonlinearities). In contrast, (see Fig. 2(b)), for the smooth polynomial nonlinearity (for which $\nu_g > p$, any p), the growth of the number of wavelet vanishing p moments improves the convergence rate (cf. the comments following Theorem 1 in the previous section).

4. Final remarks

The algorithm introduced in the paper enables convenient wavelet nonparametric on-line estimation of nonlinearities in Hammerstein systems. The proposed estimate converges pointwise to the nonlinearity with growing number of processed measurements. The rate of convergence depends on the regularity of the identified nonlinearity $\mu(x)$ and of the input density function $f(x)$, and on the wavelets applied. The rate is asymptotically optimal for properly selected wavelet functions.

Therefore, the algorithm replicates the asymptotic behavior of the aforementioned on-line nonparametric algorithms (and hence possesses the same limit properties as their *off-line* prototypes, cf. the results of Greblicki and Pawlak (2008), Hasiewicz *et al.* (2005)), offering additionally—practically important—advantages (collected in Table 1), cf. the works of Greblicki (1989; 2008) and Śliwiński *et al.* (2007).

Observe the following (cf. the detailed discussion in Section 3.1):

- (1) Recall that because of the lack of the past measurements, non-preselected estimation points have to be

Table 1. Comparison of selected properties of on-line algorithms.

Algorithm	kernel	polynomial	wavelet
Effective estimation in arbitrary x and k	$\pm^{(1)}$	+	+
Local processing	$\pm^{(2)}$	–	+
Fixed set of basis functions	$\pm^{(3)}$	–	+

either computed from the very beginning or, e.g., interpolated using the existing points as the interpolation knots.

- (2) The kernel functions without the compact support, (e.g., Gaussian kernel) require non-local processing of the measurements, that is, all the estimation points are updated with the arrival of the new measurement pair.
- (3) Kernels of the orthogonal series, like, e.g., the Dirichlet kernel of the trigonometric expansion (Rutkowski, 1980; Györfi *et al.*, 2002; Greblicki and Pawlak, 2008, Appendix B.2):

$$D_m(x, u) = \frac{\sin\left(m + \frac{1}{2}\right)(x - u)}{2\pi \sin\frac{1}{2}(x - u)},$$

depend on the parameter m (which increases during the identification process).

Remark 6. One of the consequences of the quotient form of the algorithm considered (and the subsequent presence of the estimates of the input signal probability density function $f(x)$ in its denominator, and of the product $\mu(x)f(x)$ in the numerator cf. (4)) is that the algorithm convergence and the convergence rate are both dependent on the smoothness of the density function. This feature of quotient-form algorithms is well recognized in the literature (Greblicki and Pawlak, 1985; 2008, Appendix C), cf. Appendix for the formal analysis). Nevertheless, there is a class of distribution-free quotient algorithms (including, in particular, those based on the rectangular kernel or on the Haar series kernel), whose asymptotic behavior is not affected by the smoothness of the density function Györfi *et al.* (2002, Chapter 4.1), Greblicki and Pawlak (2008, Chapter 3), Rutkowski (2004). Furthermore, algorithms based on order statistics possess density-independent asymptotic properties; (Śliwiński and Hasiewicz, 2009; Śliwiński *et al.*, 2009; Greblicki and Pawlak, 2008, Chapter 7), however, they need all the measurements to be stored during identification.

The algorithm can be easily adopted for the estimation of nonlinearities in time-varying systems, (Rutkowski, 1982; Vörös, 2005) and applied to systems from the class of the block-oriented nonlinear dynamic

systems, like parallel or serial-parallel, or Uryson systems (Hasiewicz and Śliwiński, 2002; Billings and Fakhouri, 1978; Greblicki and Pawlak, 1994).

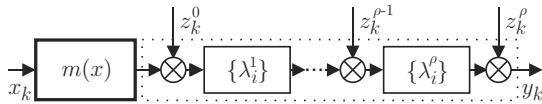


Fig. 3. Hammerstein system with a distributed dynamic subsystem as a model of the transmission line with the input nonlinearity

Example 2. Consider the Hammerstein model of a nonlinear transmission line (i.e., the one with an input nonlinearity—see Fig. 3). Assuming that all the noises, z_k^r , $r = 0, \dots, \rho$, are zero-mean second order stationary processes, and that all dynamic subsystems $\{\lambda_i^r\}$, $r = 1, \dots, \rho$ are stable, then the system in Fig. 3 satisfies Assumptions 1–4 and, from the algorithmic point of view, is equivalent to the *canonical* Hammerstein system structure in Fig. 1. ♦

Acknowledgment

The author wishes to thank the reviewers for their insightful comments.

References

- Billings, S.A. and Fakhouri, S.Y. (1978). Theory of separable processes with application to the identification of nonlinear systems, *Proceedings of IEE* **125**(10): 1051–1058.
- Capobianco, E. (2002). Hammerstein system representation of financial volatility processes, *The European Physical Journal B—Condensed Matter* **27**(2): 201–211.
- Chen, H.-F. (2004). Pathwise convergence of recursive identification algorithms for Hammerstein systems, *IEEE Transactions on Automatic Control* **49**(10): 1641–1649.
- Chen, H.-F. (2005). Strong consistency of recursive identification for Hammerstein systems with discontinuous piecewise-linear memoryless block, *IEEE Transactions on Automatic Control* **50**(10): 1612–1617.
- Chen, S., Billings, S.A. and Luo, W. (1989). Orthogonal least squares methods and their application to non-linear system identification, *International Journal of Control* **50**(5): 1873–1896.
- Coca, D. and Billings, S. A. (2001). Non-linear system identification using wavelet multiresolution models, *International Journal of Control* **74**(18): 1718–1736.
- Daubechies, I. (1992). *Ten Lectures on Wavelets*, SIAM, Philadelphia, PA.
- Donoho, D.L. (1993). Unconditional bases are optimal bases for data compression and for statistical estimation, *Applied and Computational Harmonic Analysis* **1**(1): 100–115.
- Donoho, D.L., Vetterli, M., DeVore, R.A. and Daubechies, I. (1998). Data compression and harmonic analysis, *IEEE Transactions on Information Theory* **44**(6): 2435–2476.
- Giannakis, G.B. and Serpedin, E. (2001). A bibliography on nonlinear system identification, *Signal Processing* **81**(3): 533–580.
- Greblicki, W. (1989). Nonparametric orthogonal series identification of Hammerstein systems, *International Journal of Systems Science* **20**(12): 2355–2367.
- Greblicki, W. (2001). Recursive identification of Wiener systems, *International Journal of Applied Mathematics and Computer Science* **11**(4): 977–991.
- Greblicki, W. (2002). Stochastic approximation in nonparametric identification of Hammerstein systems, *IEEE Transactions on Automatic Control* **47**(11): 1800–1810.
- Greblicki, W. and Pawlak, M. (1985). Fourier and Hermite series estimates of regression function, *Annals of the Institute of Statistical Mathematics* **37**(3): 443–454.
- Greblicki, W. and Pawlak, M. (1986). Identification of discrete Hammerstein system using kernel regression estimates, *IEEE Transactions on Automatic Control* **31**(1): 74–77.
- Greblicki, W. and Pawlak, M. (1989). Recursive nonparametric identification of Hammerstein systems, *Journal of the Franklin Institute* **326**(4): 461–481.
- Greblicki, W. and Pawlak, M. (1994). Nonparametric recovering nonlinearities in block oriented systems with the help of Laguerre polynomials, *Control—Theory and Advanced Technology* **10**(4): 771–791.
- Greblicki, W. and Pawlak, M. (2008). *Nonparametric System Identification*, Cambridge University Press, New York, NY.
- Györfi, L., Kohler, M., Krzyzak, A. and Walk, H. (2002). *A Distribution-Free Theory of Nonparametric Regression*, Springer-Verlag, New York, NY.
- Haber, R. and Keviczky, L. (1999). *Nonlinear System Parameter Identification*, Kluwer Academic Publishers, Dordrecht/Boston/London.
- Hasiewicz, Z., Pawlak, M. and Śliwiński, P. (2005). Nonparametric identification of non-linearities in block-oriented complex systems by orthogonal wavelets with compact support, *IEEE Transactions on Circuits and Systems I: Regular Papers* **52**(1): 427–442.
- Hasiewicz, Z. and Śliwiński, P. (2002). Identification of nonlinear characteristics of a class of block-oriented non-linear systems via Daubechies wavelet-based models, *International Journal of Systems Science* **33**(14): 1121–1144.
- Hsu, K., Poolla, K. and Vincent, T.L. (2008). Identification of structured nonlinear systems, *IEEE Transactions on Automatic Control* **53**(11): 2497–2513.
- Jang, W. and Kim, G. (1994). Identification of loudspeaker nonlinearities using the NARMAX modelling technique, *Journal of Audio Engineering Society* **42**(1/2): 50–59.
- Jyothi, S. N. and Chidambaram, M. (2000). Identification of Hammerstein model for bioreactors with input multiplicities, *Bioprocess Engineering* **23**(4): 323–326.

- Kang, H.W., Cho, Y.S. and Youn, D. H. (1999). On compensating nonlinear distortions of an OFDM system using an efficient adaptive predistorter, *IEEE Transactions on Communications* **47**(4): 522–526.
- Kukreja, S., Kearney, R. and Galiana, H. (2005). A least-squares parameter estimation algorithm for switched Hammerstein systems with applications to the VOR, *IEEE Transactions on Biomedical Engineering* **52**(3): 431–444.
- Lang, Z.Q. (1997). A nonparametric polynomial identification algorithm for the Hammerstein system, *IEEE Transactions on Automatic Control* **42**(2): 1435–1441.
- Lortie, M. and Kearney, R.E. (2001). Identification of time-varying Hammerstein systems from ensemble data, *Annals of Biomedical Engineering* **29**(2): 619–635.
- Mallat, S.G. (1998). *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, CA.
- Marmarelis, V.Z. (2004). *Nonlinear Dynamic Modeling of Physiological Systems*, IEEE Press Series on Biomedical Engineering, Wiley IEEE Press, Piscataway, NJ.
- Mzyk, G. (2007). Generalized kernel regression estimate for Hammerstein system identification, *International Journal of Applied Mathematics and Computer Science* **17**(2): 101–109, DOI: 10.2478/v10006-007-0018-z.
- Mzyk, G. (2009). Nonlinearity recovering in Hammerstein system from short measurement sequence, *IEEE Signal Processing Letters* **16**(9): 762–765.
- Nešić, D. and Mareels, I. M.Y. (1998). Dead-beat control of simple Hammerstein models, *IEEE Transactions on Automatic Control* **43**(8): 1184–1188.
- Niedźwiecki, M. (1988). Functional series modeling approach to identification of nonstationary stochastic systems, *IEEE Transactions on Automatic Control* **33**(10): 955–961.
- Nordsjo, A. and Zetterberg, L. (2001). Identification of certain time-varying nonlinear Wiener and Hammerstein systems, *IEEE Transactions on Signal Processing* **49**(3): 577–592.
- Pawlak, M. and Hasiewicz, Z. (1998). Nonlinear system identification by the Haar multiresolution analysis, *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications* **45**(9): 945–961.
- Rutkowski, L. (1980). Sequential estimates of probability densities by orthogonal series and their application in pattern classification, *IEEE Transactions on System, Man, and Cybernetics* **10**(12): 918–920.
- Rutkowski, L. (1982). On-line identification of time-varying systems by nonparametric techniques, *IEEE Transactions on Automatic Control* **27**(1): 228–230.
- Rutkowski, L. (2004). Generalized regression neural networks in time-varying environment, *IEEE Transactions on Neural Networks* **15**(3): 576–596.
- Śliwiński, P. and Hasiewicz, Z. (2008). Computational algorithms for wavelet identification of nonlinearities in Hammerstein systems with random inputs, *IEEE Transactions on Signal Processing* **56**(2): 846–851.
- Śliwiński, P. and Hasiewicz, Z. (2009). Recursive nonparametric estimation of Hammerstein system nonlinearity by Haar wavelet series, *IEEE Transactions on Automatic Control*, (submitted for review).
- Śliwiński, P., Hasiewicz, Z. and Wachel, P. (2007). On-line polynomial series estimates of Hammerstein system nonlinearities, *Proceedings of the 13th IEEE IFAC International Conference on Methods and Models in Automation and Robotics MMAR 2007*, Szczecin, Poland.
- Śliwiński, P., Rozenblit, J., Marcellin, M.W. and Klempous, R. (2009). Wavelet amendment of polynomial models in nonlinear system identification, *IEEE Transactions on Automatic Control* **54**(4): 820–825.
- Srinivasan, R., Rengaswamy, R., Narasimhan, S. and Miller, R. (2005). Control loop performance assessment: 2. Hammerstein model approach for stiction diagnosis, *Industrial & Engineering Chemistry Research* **44**(17): 6719–6728.
- Stone, C.J. (1980). Optimal rates of convergence for nonparametric regression, *Annals of Statistics* **8**(6): 1348–1360.
- Sureshbabu, N. and Farrell, J.A. (1999). Wavelet based system identification for non-linear control, *IEEE Transactions on Automatic Control* **44**(2): 412–417.
- Unser, M. (1999). Splines. A perfect fit for signal and image processing, *IEEE Signal Processing Magazine* **16**(6): 22–38.
- Vörös, J. (2005). Identification of Hammerstein systems with time-varying piecewise-linear characteristics, *IEEE Transactions on Circuits and Systems II: Express Briefs* **52**(12): 865–869.
- Walter, G.G. and Shen, X. (2001). *Wavelets and other orthogonal systems with applications*, 2nd Edn., Chapman & Hall, Boca Raton, FL.
- Westwick, D.T. and Kearney, R.E. (2001). Separable least squares identification of nonlinear Hammerstein models: Application to stretch reflex dynamics, *Annals of Biomedical Engineering* **29**(8): 707–718.
- Westwick, D.T. and Kearney, R.E. (2003). *Identification of nonlinear physiological systems*, IEEE Press Series on Biomedical Engineering, Wiley/IEEE Press, Piscataway, NJ.
- Zhao, W.-X. and Chen, H.-F. (2006). Recursive identification for Hammerstein system with ARX subsystem, *IEEE Transactions on Automatic Control* **51**(12): 1966–1974.



Przemysław Śliwiński was born in 1971 in Poland. He received the M.Sc. and Ph.D. degrees in computer engineering from the Wrocław University of Technology, Poland, in 1996 and 2000, respectively. In 2000 he joined the Institute of Engineering Cybernetics of the same university, where he is currently an assistant professor. His research interests are wavelets and their application to nonparametric system identification, signal and image processing, analysis and compression. In 2006 he was a visiting research professor at the University of Arizona, Tucson, USA.

Appendix

Proofs of Theorems

The proof is split in two sections which correspond to the subroutines of the proposed identification algorithm. In the first part, the properties of the empirical coefficients are established. In the second, they are used to prove the convergence properties of the estimate $\hat{\mu}^{(k)}(x)$. In the following preparatory step, we will show that the fact that the nonlinearity $\mu(x)$ is, by virtue of Assumption 1, locally bounded and hence bounded in any finite interval, allows examining the pointwise properties of the algorithm with the help of the compactly supported wavelet orthogonal expansion apparatus (cf. Section 3).

Let $\Delta_M(x)$, referred further to as the *influence interval* (Mallat, 1998, Chapter 6), contain the supports of all wavelet series expansion functions intersecting with a given estimation point x (the rest of wavelet expansion functions, being zero in x , do not carry any information about the nonlinearity in this point), cf. (4):

$$\begin{aligned} \Delta_M(x) &\stackrel{\text{def}}{=} \underbrace{\bigcup_n \text{supp } \varphi_{Mn}(x)}_{2^M x - s < n < 2^M x} + \bigcup_{m \geq M} \underbrace{\bigcup_n \text{supp } \psi_{mn}(x)}_{2^m x - s < n < 2^m x} \\ &= [x - 2^{-M}s, x + 2^{-M}s]. \end{aligned}$$

Note that $\Delta_M(x)$ is finite. Let now

$$g_\Delta(x) \stackrel{\text{def}}{=} g(x) I_{\Delta_M}(x) = \mu(x) f(x) I_{\Delta_M}(x)$$

denote the original function $g(x)$ (cf. Section 3) truncated to the interval $\Delta_M(x)$ surrounding x . Clearly, $g_\Delta(x)$, being bounded for any x (cf. Assumptions 1 and 3), is square integrable and can be expanded into wavelet orthogonal series. Since, however, $g_\Delta(x) = g(x)$ in the influence interval $\Delta_M(x)$, we will—for notational simplicity—skip the subscript Δ in the following derivations (keeping in mind that, formally, $g(x)$ is not square integrable).

Bias error. Considering the empirical coefficients, $\hat{\alpha}_{Mn}^{(k)}$ and $\hat{\beta}_{mn}^{(k)}$, of the estimate numerator, $\hat{g}^{(k)}(x)$ (the analysis of the denominator, $\hat{f}^{(k)}(x)$, and its coefficients, $\hat{a}_{Mn}^{(k)}$ and $\hat{b}_{mn}^{(k)}$, is similar and thus omitted), it is convenient to convert them to the closed forms, cf. (2):

$$\hat{\alpha}_{Mn}^{(k)} = \frac{1}{k} \sum_{0 < i \leq k} \varphi_{Mn}(x_i) \cdot y_i, \quad (9)$$

$$\hat{\beta}_{mn}^{(k)} = \frac{1}{k} \sum_{\kappa_m < i \leq k} \psi_{mn}(x_i) \cdot y_i.$$

Under Assumptions 1–4, for the empirical scaling coeffi-

cients, $\hat{\alpha}_{Mn}^{(k)}$, we have, cf. (1),

$$\begin{aligned} E \hat{\alpha}_{Mn}^{(k)} &= E \left[\frac{1}{k} \sum_{0 < i \leq k} \varphi_{Mn}(x_i) \cdot y_i \right] \\ &= E [\varphi_{Mn}(x_1) \cdot y_1] = \alpha_{mn}, \end{aligned} \quad (10)$$

i.e., $\hat{\alpha}_{Mn}^{(k)}$'s are unbiased estimates of the scaling coefficients α_{mn} (cf. (1) and see also the results of Hasiewicz *et al.* (2005)). In turn, after noting that

$$\begin{aligned} E \hat{\beta}_{mn}^{(k)} &= E \left[\underbrace{\frac{1}{k} \sum_{0 < i \leq k} \psi_{mn}(x_i) \cdot y_i}_{\beta_{mn}} \right. \\ &\quad \left. - E \left[\underbrace{\frac{1}{k} \sum_{0 < i \leq \kappa_m} \psi_{mn}(x_i) \cdot y_i}_{\frac{\kappa_m}{k} \beta_{mn}} \right] \right], \end{aligned} \quad (11)$$

we find wavelet empirical coefficients, $\hat{\beta}_{mn}^{(k)}$, being biased estimates of the wavelet coefficients β_{mn} . The bias error is clearly caused by the fact that, in an on-line régime, the past measurements $\{(x_i, y_i)\}$, $0 < i \leq \kappa_m$, which are necessary for the proper initialization of the new wavelet coefficients $\hat{\beta}_{mn}^{(k)}$, are not available. Clearly, the bias error equals (cf. the second term in (11))

$$\text{bias } \hat{\beta}_{mn}^{(k)} = -\frac{\kappa_m}{k} \beta_{mn}. \quad (12)$$

Variance. For the coefficients $\hat{\alpha}_{Mn}^{(k)}$, the variance error order depends only on the measurement number k and equals (cf. (9) and see the work of Hasiewicz *et al.* (2005, pp. 438–441) for the proof of this property)

$$\text{var } \hat{\alpha}_{Mn}^{(k)} = \text{var} \left\{ \frac{1}{k} \sum_{0 < i \leq k} \varphi_{Mn}(x_i) \cdot y_i \right\} = \mathcal{O} \left(\frac{1}{k} \right).$$

Passing to the coefficients $\hat{\beta}_{mn}^{(k)}$, we observe that

$$\begin{aligned} \text{var } \hat{\beta}_{mn}^{(k)} &= \text{var} \left\{ \frac{k - \kappa_m}{k} \cdot \left[\frac{1}{k - \kappa_m} \sum_{\kappa_m < i \leq k} \psi_{mn}(x_i) \cdot y_i \right] \right\} \end{aligned} \quad (13)$$

and hence (cf. (9))

$$\text{var } \hat{\beta}_{mn}^{(k)} = \mathcal{O} \left(\frac{k - \kappa_m}{k} \cdot \frac{1}{k} \right) = \mathcal{O} \left(\frac{1}{k} \right), \quad (14)$$

since for a given scale m the threshold $\kappa_m < k$ is fixed and does not depend on k (cf. (2) and (9)).

Mean square error. In order to examine the mean-square error of the estimate $\hat{g}^{(k)}(x)$, we begin with its standard decomposition to the bias and the variance components

$$E \left[g(x) - \hat{g}^{(k)}(x) \right]^2 = \text{bias}^2 \hat{g}^{(k)}(x) + \text{var} \hat{g}^{(k)}(x). \tag{15}$$

Let K be the first scale for which the current measurement number k does not exceed the corresponding threshold κ_K , i.e., let it be the smallest scale at which the wavelet coefficients $\hat{\beta}_{Kn}^{(k)}$ are all zero and are not yet incorporated into the estimate $\hat{g}^{(k)}(x)$. Also, denote by $g_K(x)$ a wavelet approximation of the product $g(x)$ at the scale K . The bias error in (15) can now be expressed as (cf. (4) and (10))

$$\begin{aligned} \text{bias} \hat{g}^{(k)}(x) &= g(x) - E\hat{g}^{(k)}(x) \\ &= [g(x) - g_K(x)] + [g_K(x) - E\hat{g}^{(k)}(x)] \\ &= \underbrace{\sum_{m \geq K} \sum_n \psi_{mn}(x) \cdot \beta_{mn}}_{\widetilde{\text{bias}} \hat{g}^{(k)}(x)} \\ &\quad + \underbrace{\sum_{M \leq m < K} \sum_n \psi_{mn}(x) \cdot \text{bias} \hat{\beta}_{mn}^{(k)}}_{\overline{\text{bias}} \hat{g}^{(k)}(x)}, \end{aligned}$$

where $\widetilde{\text{bias}} \hat{g}^{(k)}(x)$ is the error of approximation of $g(x)$ by $g_K(x)$, and $\overline{\text{bias}} \hat{g}^{(k)}(x)$ is the error formed by the accumulation of the bias errors of empirical the wavelet coefficients (translation factors, n , run as in wavelet part of (4)).

Recalling that $\psi_{mn}(x) = \mathcal{O}(2^{m/2})$ and that for $g(x)$ with the Hölder smoothness index $\nu_g (= \min\{\nu_f, \nu_m\} > 0)$, and for wavelets with p vanishing moments, it holds that $\beta_{mn} = \mathcal{O}(2^{-(2\nu+1)m/2})$, where $\nu = \min\{\nu_g, p\}$, we obtain that the former error is of order (Hasiewicz *et al.*, 2005; Mallat, 1998, Chapter 9)

$$\widetilde{\text{bias}} \hat{g}^{(k)}(x) = \mathcal{O} \left(\sum_{m \geq K} 2^{-\frac{2\nu+1}{2}m} \cdot 2^{\frac{m}{2}} \right) = \mathcal{O}(2^{-\nu K}). \tag{16}$$

For the latter we get that (see (2), (12) and cf. (10))

$$\overline{\text{bias}} \hat{g}^{(k)}(x) = - \sum_{M \leq m < K} \sum_n \frac{\kappa_m}{k} \cdot \psi_{mn}(x) \cdot \beta_{mn}. \tag{17}$$

Finally, the variance is of order, cf. the results of Hasiewicz *et al.* (2005)

$$\text{var} \hat{g}^{(k)}(x) = \mathcal{O} \left(\frac{2^K}{k} \right). \tag{18}$$

To find the bound for the bias error in (17), we need to establish the formula for the thresholding sequence $\{\kappa_m\}$. Note, that the bounds (16) and (18) imply that the scale K needs to grow with the increasing number of measurement, k , and the pace of this growth should be $c^{-1} \log_2 k$, for $c > 1$, cf., e.g., the works of Pawlak and Hasiewicz (1998) as well as Hasiewicz *et al.* (2005). The sequence $\{\kappa_m\}$ which corresponds to such K 's is $\kappa_m = 2^{cm}$, cf. (6). Inserting it to (17) yields

$$\begin{aligned} \overline{\text{bias}} \hat{g}^{(k)}(x) &= \mathcal{O} \left(\sum_{M \leq m < K} \frac{2^{cm}}{k} \cdot 2^{\frac{m}{2}} \cdot 2^{-\frac{2\nu+1}{2}m} \right) \\ &= \mathcal{O} \left(\frac{2^{(c-\nu)K}}{k} \right) \end{aligned} \tag{19}$$

for $c \neq \nu$ and $\mathcal{O}(K/k)$ otherwise. Plugging altogether (16), (18) and (19) into (15) and applying $K = c^{-1} \cdot \log_2 k$, we obtain

$$\begin{aligned} E[g(x) - \hat{g}^{(k)}(x)]^2 &= \underbrace{\mathcal{O} \left(2^{-\frac{2\nu}{c} \log_2 k} \right)}_{\widetilde{\text{bias}}^2 \hat{g}^{(k)}(x)} + \underbrace{\mathcal{O} \left(k^{-2} \cdot 2^{\frac{2(c-\nu)}{c} \log_2 k} \right)}_{\overline{\text{bias}}^2 \hat{g}^{(k)}(x)} \\ &\quad + \underbrace{\mathcal{O} \left(k^{-1} \cdot 2^{\frac{1}{c} \log_2 k} \right)}_{\text{var} \hat{g}^{(k)}(x)} \\ &= \mathcal{O} \left(k^{-\frac{2\nu}{c}} + k^{-\frac{2\nu}{c}} + k^{\frac{1-c}{c}} \right) \\ &= \mathcal{O} \left(k^{-\frac{2\nu}{c}} + k^{\frac{1-c}{c}} \right), \end{aligned} \tag{20}$$

i.e., that, in particular, the first two terms, the approximation error, $\widetilde{\text{bias}}^2 \hat{g}^{(k)}(x)$, and the additional bias error, $\overline{\text{bias}}^2 \hat{g}^{(k)}(x)$, are of the same order. Hence the convergence condition in (6), although derived solely from approximation and variance errors, remains valid after taking into account the bias error. Using Lemma C.7 of Greblicki and Pawlak (2008, Appendix C.2) ends the proof of Theorem 1.

To prove the second theorem, it suffices now to observe that the direct optimization of (20), i.e., seeking for such c for which both bias and variance errors will vanish with equal orders as k grows, yields $c = 2\nu + 1$ (see (7)) and

$$E[g(x) - \hat{g}^{(k)}(x)]^2 = \mathcal{O}(k^{-2\nu/(2\nu+1)}). \tag{21}$$

Note that, for $c < 2\nu + 1$, the variance error dominates the bias one, and vice versa for $c > 2\nu + 1$. When $c = \nu$, we have $\overline{\text{bias}}^2 \hat{g}^{(k)}(x) = \mathcal{O}((K/k)^2)$ and this error—as of smaller order than $\mathcal{O}(2^K/k)$ for all positive K and k —is absorbed in (19) by the unit variance. Finally, as was

mentioned at the beginning of the Appendix, in a similar way one can show that for the same $c = 2\nu + 1$ we have

$$E[f(x) - \hat{f}^{(k)}(x)]^2 = \mathcal{O}(k^{-2\nu/(2\nu+1)}). \quad (22)$$

Applying Lemma C.8 of Greblicki and Pawlak (2008, Appendix C.2) to the bounds in (21) and (22) yields (8) and ends the proof.

Received: 30 July 2009

Revised: 3 March 2010