Computational algorithms for wavelet identification of nonlinearities in Hammerstein systems with random inputs

Przemysław Śliwiński* and Zygmunt Hasiewicz*

The Institute of Computer Engineering, Control and Robotics
Wrocław University of Technology, Wrocław, Poland

Abstract— **Simple and efficient computational algorithms for nonparametric wavelet-based identification of nonlinearities in Hammerstein systems driven by random signals are proposed. They exploit binary grid interpolations of compactly supported wavelet functions. The main contribution consists in showing how to use the wavelet values from the binary grid together with the fast wavelet algorithms to obtain the practical counterparts of the wavelet-based estimates for irregularly and randomly spaced data, without any loss of the asymptotic accuracy. The convergence and the rates of convergence are examined for the new algorithms and, in particular, conditions for the optimal convergence speed are presented. Efficiency of the algorithms for a finite number of data is also illustrated by means of the computer simulations.**

I. INTRODUCTION AND PRELIMINARIES

Growing popularity of compactly supported wavelets and their successful applications in different areas seem to have two reasons: (*i*) an effective (parsimonious) wavelet representation of a broad class of functions, [1], accompanied by (*ii*) the existence of fast algorithms for wavelet computations, [2]. These distinguishing features are achieved in spite of the lack of the explicit representations of the wavelet functions, which in general are defined pointwise by numerical procedures and in fact can be efficiently computed for dyadic arguments only [3], [4]. Obviously, this is not a relevant obstacle in applications which operate on equidistantly spaced data like, *e.g.*, time series denoising [5]. However, in other applications, especially like considered in this note nonparametric system identification tasks, where the data are usually distributed at random (*cf.* [6], [7]), obtaining the practical and effective computational counterparts of existing wavelet algorithms is a nontrivial due to: (*i*) the need of computing the empirical wavelet coefficients from random data and (*ii*) the necessity of evaluation of the values of wavelet estimates at arbitrary points. In the literature the former problem is solved mainly by data-dependent preprocessing, *e.g.* by interpolating the input-output data in order to obtain their approximated values at binary grid points [8], [9], [10], or by adapting the second generation wavelets to an irregular (*e.g.* random) grid at hand [11], [12]. In this work we solve both issues exploiting a simple and direct approach in which standard first generation wavelets are interpolated instead. This idea was already proposed in [13] and applied there to a class of *scaling function estimates* recovering a nonlinearity in Hammerstein systems [14] – a similar approach, assuming random input was also applied independently in [15, p. 47], however in a context of static systems. Here, we extend the idea to a wider class of *wavelet estimates*, elaborated in [7], decompose and refine the interpolation error bounds established in [13], and show how the influence of the interpolation error can be neutralized with a growing number of data. As a consequence, the beneficial features of the theoretical wavelet identification algorithms (like *e.g.* the best possible rate of convergence established in [7]) can automatically be conveyed to their easy-to-compute counterparts. Furthermore, by combining the proposed interpolators with a widely

* Przemysław Śliwiński and Zygmunt Hasiewicz are with The Institute of Computer Engineering, Control and Robotics, Wrocław University of Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland. tel. +4871 3203277, fax +4871 3212677, e-mail: {Przemyslaw.Sliwinski,Zygmunt.Hasiewicz}@pwr.wroc.pl

used fast wavelet transform (FWT), our algorithm processes random data with a numerical complexity of order $\mathcal{O}(N)$, where N is a number of measurements, *i.e.*, with the order typical for standard wavelet algorithms based on FWT with equidistant input data.

Hammerstein system. The system, being a cascade of a static nonlinearity and linear dynamics (*cf.* Fig. 1a), is an *archetype* for the ample class of block-oriented nonlinear dynamic systems which have gained popularity in various applications (*cf.* [16]), and is a standard 'example' system considered in the literature; see *e.g.* [6], [7], [17], [18].
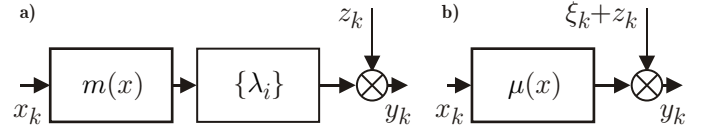

Fig. 1. **a)** The Hammerstein system **b)** The same system seen by identification algorithm

The following assumptions (typical for nonparametric system identification tasks; *cf.* [6], [7], [13], [18]) hold in the paper: *1)* the input signal, $\{x_k\}$, and the external additive noise, $\{z_k\}$, are zero-mean random processes with finite variances; they are mutually independent and $\{x_k\}$ is an i.i.d. process with a density function $f(x)$, *2)* the density, $f(x)$, and the static nonlinearity, $m(x)$, are bounded and continuous with some Hölder exponents $\nu_f, \nu_m > 0$ (in particular, they do not have to be invertible), *3)* the linear dynamic subsystem is asymptotically stable and its impulse response, $\{\lambda_i\}$, is unknown, *4)* only a set of input-output measurements $\{(x_k, y_k)\}$, $k = 1, \ldots, N$, is available.

From an input-output data point of view, the Hammerstein system in Fig. 1a can be described by the equation, $y_k = \sum_{i=0}^{\infty} \lambda_i m(x_{k-i}) + z_k$ and rewritten to the equivalent form, $y_k = \mu(x_k) + \xi_k + z_k$, in which past observations $\{x_{k-i}\}$ induce an *additive* stationary 'system noise' $\xi_k = \sum_{i=1}^{\infty} \lambda_i [m(x_{k-i}) - Em(x_1)]$, correlated because of the own system dynamic, and disturbing together with the external one, $\{z_k\}$, the output of a nonlinearity $\mu(x) = \lambda_0 m(x) + \zeta$, where $\zeta = Em(x_1)\sum_{i=1}^{\infty} \lambda_i$ is a system dependent constant; see Fig. 1b. This leads eventually to the observation that $\mu(x) = E(y_k|x_k = x)$, *i.e.* that $\mu(x)$ is in fact a regression function of y_k on x_k; *cf.* [6], [18].

Remark 1: Using the data $\{(x_k, y_k)\}$ we can only identify $\mu(x)$, a scaled and translated version of $m(x)$. This is an inevitable consequence of the composite structure of the Hammerstein system and the inaccessibility of the interconnecting inner signal (see [6], [7], [18]). Moreover, that the input $\{x_k\}$ is a white process, and hence a *persistently exciting signal* of infinite order [19], makes possible identification of the system nonlinearity by *a direct* estimation of regression function $\mu(x)$, regardless the structure of the dynamics. Otherwise, one needs to apply other approaches like *e.g.* those for Wiener systems, [20].

The reference algorithm. Wavelet estimates of regression functions have been primarily applied to nonparametric system identification in [18] and then further studied in *e.g.* [7], [21]. As it was proven in [7] – due to superior approximation abilities of the compactly supported wavelets – they outperform *classical* nonparametric orthogonal series estimates (*e.g.* polynomial or trigonometric, *cf.* [6]). The following estimate, proposed and examined in [7] (*cf.* also [12, Sec. 3.1] or [6], [18]), will be referred to as the *reference algorithm*:

$$\hat{\mu}_K(x) = \hat{g}_K(x) / \hat{f}_K(x) \qquad (1)$$

where $\hat{g}_K(x)$ and $\hat{f}_K(x)$ are the estimates of the wavelet approximations at some scale K acting as a smoothing parameter, *cf.* [7],

[14, Remark 5.1], of a product $g(x) = \mu(x) \cdot f(x)$, and of the input signal density $f(x)$. They are presented below in a vector-like form for conciseness and to emphasize similarities of the computations needed by $\hat{g}_K(x)$ and $\hat{f}_K(x)$:

$$\begin{bmatrix} \hat{g}_K(x) \\ \hat{f}_K(x) \end{bmatrix} = \sum_{n=\lceil 2^M x - s_2 \rceil}^{\lfloor 2^M x - s_1 \rfloor} \varphi_{Mn}(x) \cdot \begin{bmatrix} \hat{\alpha}_{Mn} \\ \hat{a}_{Mn} \end{bmatrix} \qquad (2)$$
$$+ \sum_{m=M}^{K-1} \sum_{n=\lceil 2^m x - t_2 \rceil}^{\lfloor 2^m x - t_1 \rfloor} \psi_{mn}(x) \cdot \begin{bmatrix} \hat{\beta}_{mn} \\ \hat{b}_{mn} \end{bmatrix}$$

with the empirical coefficients, $\hat{\alpha}_{Mn}, \hat{a}_{Mn}, \hat{\beta}_{mn}$, and $\hat{b}_{mn}$, calculated as

$$\begin{bmatrix} \hat{\alpha}_{Mn} \\ \hat{a}_{Mn} \end{bmatrix} = \frac{1}{N} \sum_{k=1}^{N} \varphi_{Mn}(x_k) \cdot \begin{bmatrix} y_k \\ 1 \end{bmatrix}$$
$$\begin{bmatrix} \hat{\beta}_{mn} \\ \hat{b}_{mn} \end{bmatrix} = \frac{1}{N} \sum_{k=1}^{N} \psi_{mn}(x_k) \cdot \begin{bmatrix} y_k \\ 1 \end{bmatrix} \qquad (3)$$

where $[s_1, s_2]$ and $[t_1, t_2]$ are supports of the *father* and *mother* wavelets, φ and ψ, respectively. We assume that they are continuous with a Hölder exponent $\nu_\varphi > 0$, and that $\{\varphi_{Mn}\}, \{\psi_{mn}\}$, for $m = M, M+1, \ldots$, some fixed M, and $n = \ldots, -1, 0, 1, \ldots$, constitute an orthogonal basis of $L^2(R)$ space. The properties of the reference algorithm are established by the following lemma, being a 'decomposed' version of the Theorem 2 in [7] with respect to the numerator of the estimate, and pointing out the behavior of the mean square error components (see proof of Th. 2 in [7]).

Lemma 1: Let the wavelet family in the estimate $\hat{\mu}_K(x)$ in (1) have p vanishing moments. Since the following error bounds hold

$$\text{bias } \hat{g}_K(x) = \mathcal{O}(2^{-\gamma K}), \quad \text{where } \gamma = \min\{\nu_m, \nu_f, p\} \quad (4)$$
$$\text{var } \hat{g}_K(x) \leq c_\varphi \cdot 2^K / N, \quad \text{where } c_\varphi \sim \sup_x \varphi^2(x) \quad (5)$$

for any x, then selecting the estimate scale K according to the rule

$$K = \left\lfloor \frac{1}{2\gamma + 1} \log_2 N \right\rfloor \qquad (6)$$

makes the reference estimate $\hat{\mu}_K(x)$ converge to $\mu(x)$ as $N \to \infty$ at the rate

$$\hat{\mu}_K(x) = \mu(x) + \mathcal{O}(N^{-\frac{\gamma}{2\gamma+1}}) \text{ in probability,} \qquad (7)$$

for arbitrary x at which $f(x) > 0$.

Note that the convergence given in (7) does not depend on a structure of the dynamic subsystem $\{\lambda_i\}$; *cf.* discussion in [7], [18]. Moreover, provided that $p \geq \min\{\nu_m, \nu_f\}$, this rate is asymptotically optimal, *i.e.*, the best attainable by any nonparametric estimate of a nonlinearity for which the Hölder exponent is only known (see [22]). Clearly, these properties are worthy to be maintained in an implementation.

II. THE COMPUTATIONAL ALGORITHM

Since the input signal has a density and its realizations $\{x_k\}$ can take any values in the regions where $f(x) > 0$, a practical use of the reference estimate (1)-(3) is strongly hindered by the need to compute the wavelet values in arbitrary (random) points. To overcome this difficulty we simply replace the wavelet functions in the reference algorithm by their piecewise-constant interpolators.

Construction. Denote by

$$\bar{\varphi}^H(x) = \varphi\left(\frac{\lfloor 2^H \cdot x + 1/2 \rfloor}{2^H}\right) \text{ and } \bar{\psi}^H(x) = \psi\left(\frac{\lfloor 2^H \cdot x + 1/2 \rfloor}{2^H}\right) \quad (8)$$

the interpolators of the original father and mother wavelet functions φ and ψ, where the *true values* of wavelet functions for arbitrary argument, x, are approximated by their values computed for the

nearest neighbor of x located on a dyadic grid 2^{-H}, $H = 0, 1, \ldots$; the parameter H plays the role of an interpolation scale. Their scaled and translated versions are defined as

$$\bar{\varphi}_{Mn}^H(x) = 2^{\frac{M}{2}} \varphi\left(\frac{\lfloor 2^{H+M} \cdot 2^M x + 1/2 \rfloor}{2^{H+M}} - n\right) \qquad (9)$$
$$\bar{\psi}_{mn}^H(x) = 2^{\frac{m}{2}} \psi\left(\frac{\lfloor 2^{H+m} \cdot 2^m x + 1/2 \rfloor}{2^{H+m}} - n\right)$$

or, equivalently, as $\bar{\varphi}_{Mn}^H(x) = 2^{\frac{M}{2}} \varphi(2^M \bar{x}_M^H - n)$ and $\bar{\psi}_{mn}^H(x) = 2^{\frac{m}{2}} \psi(2^m \bar{x}_m^H - n)$, where $\bar{x}_\bullet^H = \lfloor 2^{H+2\bullet} x + 1/2 \rfloor / 2^{H+2\bullet}$, and where $m = M, \ldots, K-1$. The needed values of wavelet functions at the grid points can easily be precomputed by using *e.g.* the Strang algorithm described in [4]; *cf.* also [3], [23] and [13, Appendix IV]. The computational counterpart of the estimate $\hat{\mu}_K(x)$, obtained by plugging $\bar{\varphi}_{Mn}^H$ and $\bar{\psi}_{mn}^H$ in place of φ_{Mn} and ψ_{mn} in (1)-(3), will be denoted as $\bar{\mu}_K^H(x)$, its nominator and denominator as $\bar{g}_K^H(x)$ and $\bar{f}_K^H(x)$, and the coefficients as $\bar{\alpha}_{Mn}^H, \bar{a}_{Mn}^H, \bar{\beta}_{mn}^H$, and $\bar{b}_{mn}^H$, respectively.

Calculations. Evaluating of the estimate $\bar{\mu}_K^H(x)$ can be split in two separate phases:

1) Computing the estimate of the coefficients from the measurement data set, and
2) Computing the resulting estimate value for a given point x.

In the first phase we compute the coefficients $\bar{\alpha}_{Mn}^H, \bar{a}_{Mn}^H, \bar{\beta}_{mn}^H$, and $\bar{b}_{mn}^H$ 'activated' by a given measurement set $\{(x_k, y_k)\}_{k=1}^N$, *i.e.* these with the translation factors n ranging (for $a = \min_k \{x_k\}$ and $b = \max_k \{x_k\}$) over

$$n = \lceil 2^M a - s_2 \rceil, \ldots, \lfloor 2^M b - s_1 \rfloor \text{ and } n = \lceil 2^m a - t_2 \rceil, \ldots, \lfloor 2^m b - t_1 \rfloor \qquad (10)$$

for $m = M, \ldots, K-1$. For this purpose we employ the standard Mallat's FWT transform (see [2, p. 255] and *cf.* (20) in Appendix). The transform's initial coefficients $\bar{\alpha}_{Kn}^H$ and $\bar{a}_{Kn}^H$, are calculated, at the scale K, from the raw (non-ordered) data $\{(x_k, y_k)\}$ by the recursive versions of the formulas in (3)

$$\begin{bmatrix} \bar{\alpha}_{Kn}^{H,(k)} \\ \bar{a}_{Kn}^{H,(k)} \end{bmatrix} = \frac{k-1}{k} \begin{bmatrix} \bar{\alpha}_{Kn}^{H,(k-1)} \\ \bar{a}_{Kn}^{H,(k-1)} \end{bmatrix} + \frac{1}{k} \bar{\varphi}_{Kn}^H(x_k) \cdot \begin{bmatrix} y_k \\ 1 \end{bmatrix} \quad (11)$$

for $k = 1, \ldots, N$ and $n = \lceil 2^K x_k - s_2 \rceil, \ldots, \lfloor 2^K x_k - s_1 \rfloor$, where $\bar{\alpha}_{Kn}^{H,(0)} = \bar{a}_{Kn}^{H,(0)} = 0$. To accomplish the second phase we simply reuse the reference estimate (1) with interpolators $\bar{\varphi}_{Mn}^H(x)$ and $\bar{\psi}_{mn}^H(x)$, $m = M, \ldots, K-1$, instead of $\varphi_{Mn}(x)$ and $\psi_{mn}(x)$, which, due to its 'local' pointwise character, incorporates into computations only terms which are in *a cone of influence* of x; [2, p. 175].

Convergence rate. Taking the interpolators in place of the original wavelet functions certainly introduces an additional error. This error, referred to as an *interpolation error*, is evaluated in the Appendix; see (19). The impact of the interpolation error on the properties of the computational algorithm $\bar{\mu}_K^H(x)$ (as compared to the reference estimate properties) is characterized in the proposition below. The proper rule for selection of the interpolation scale H that guarantees that the computational algorithm $\bar{\mu}_K^H(x)$ preserves the reference rate of convergence (7), is also introduced.

Proposition 2: The following bound for the bias error of $\bar{g}_K^H(x)$ holds (*cf.* (4) and see Fig. 2)

$$\text{bias } \bar{g}_K^H(x) = \underbrace{\mathcal{O}\left(2^{-\gamma K}\right)}_{\text{bias } \hat{g}_K(x)} + \underbrace{\mathcal{O}\left(2^{-\eta(H+M+1)}\right)}_{\overline{\text{bias}}\, \bar{g}_K^H(x)} \qquad (12)$$

where $\eta = \min\{\nu_\varphi, 1\}$, and comprises the bias error present in the reference algorithm, bias $\hat{g}_K(x)$, and additional bias introduced by

the interpolation inaccuracy error, $\overline{\text{bias}}\,\bar{g}_K^H(x)$. The latter includes in turn the following terms (see Appendix for definitions)

$$\overline{\text{bias}}\,\bar{g}_K^H(x) = \underbrace{\mathcal{O}(2^{-\eta(H+K+1)-2\eta(K-M)})}_{\overline{\text{bias}}_\alpha\,\bar{g}_K^H(x)} + \underbrace{\mathcal{O}(2^{-\eta(H+K+1)})}_{\overline{\text{bias}}_\beta\,\bar{g}_K^H(x)} \quad (13)$$

$$+ \underbrace{\mathcal{O}(2^{-\eta(H+M+1)})}_{\overline{\text{bias}}_\varphi\,\bar{g}_K^H(x)} + \underbrace{\mathcal{O}(2^{-\gamma M-\eta(H+M+1)})}_{\overline{\text{bias}}_\psi\,\bar{g}_K^H(x)} \quad (14)$$

being the upper bounds of the errors of the first phase of computations ($\overline{\text{bias}}_\alpha\,\bar{g}_K^H(x)$ and $\overline{\text{bias}}_\beta\,\bar{g}_K^H(x)$), and of the errors introduced in the second phase of computing $\bar{g}_K^H(x)$ (i.e. $\overline{\text{bias}}_\varphi\,\bar{g}_K^H(x)$ and $\overline{\text{bias}}_\psi\,\bar{g}_K^H(x)$), respectively. The bound of the variance error remains the same as in the reference algorithm, *cf.* (5)

$$\text{var}\,\bar{g}_K^H(x) \leq c_\varphi \cdot 2^K/N \quad (15)$$

for any x. If the scale K in the estimate is governed by the rule (6) and the interpolation scale H is selected as

$$H = \left\lfloor \frac{\gamma}{\eta} K \right\rfloor, \quad (16)$$

then the estimate $\bar{\mu}_K^H(x)$ converges to the nonlinearity $\mu(x)$ as $N \to \infty$ with the rate

$$\bar{\mu}_K^H(x) = \mu(x) + \mathcal{O}\left(N^{-\frac{\gamma}{2\gamma+1}}\right) \text{ in probability}, \quad (17)$$

for each point x such that $f(x) > 0$, *i.e.* in a similar way as the reference estimate $\hat{\mu}_K(x)$.

Proof: See Appendix. ∎

Remark 2: The decomposition of the interpolation error in (13)-(14) shows that if in the second phase the estimate $\bar{\mu}_K^H(x)$ is computed in points placed on binary grid $2^{-(H+M)}$ exclusively, then the summands in (14), i.e., the terms $\overline{\text{bias}}_\varphi\,\bar{g}_K^H(x)$ and $\overline{\text{bias}}_\psi\,\bar{g}_K^H(x)$, are both zero and therefore the order of interpolation part of the overall bias error reduces to $\overline{\text{bias}}\,\bar{g}_K^H(x) = \mathcal{O}(2^{-\eta(H+K+1)})$; *cf.* Fig. 2b. This means, amongst others, that in such a case the rule (16) can be weakened, *e.g.* to the form (*cf.* (22))

$$H = \left\lfloor \frac{\gamma-\eta}{\eta} K \right\rfloor \quad (18)$$

without a deterioration of the convergence rate in (17).

To make the weakened rule (18) valid for arbitrary arguments, a biorthogonal wavelet family with spline wavelets on a synthesis side (*cf.* [3, pp. 271-278]) can be used instead of the orthogonal one. The 'second phase' error components in (14), $\overline{\text{bias}}_\varphi\,\bar{g}_K^H(x)$ and $\overline{\text{bias}}_\psi\,\bar{g}_K^H(x)$, annihilate in this case since spline wavelets are given in closed forms and hence do not need interpolation.

Computational complexity. We shall separately establish the complexity of each phase of computations and will focus only on the numerator $\bar{g}_K^H(x)$ since the computations needed in the denominator are similar. We assume that the scale K is selected according to the rule (6) and the scale $M < K$ is positive. We also assume that the values of wavelet functions at the required grid points are precomputed and therefore do not influence the complexity of the procedure.

Computing initial (finest scale) coefficients $\bar{\alpha}_{Kn}^H$ for FWT requires constant number (dependent merely on the length, $s_2 - s_1$, of the scaling function support; *cf.* (11)) of operations for each data pair (x_k, y_k) and thus is of complexity $\mathcal{O}(N)$ only. In turn, performing the FWT algorithm itself, requires (*cf.* (10))

$$\mathcal{O}\left(\lfloor 2^M b - s_1 \rfloor - \lceil 2^M a - s_2 \rceil + 1\right)$$
$$+ \mathcal{O}\left(\sum_{m=M}^{K-1} (\lfloor 2^m b - t_1 \rfloor - \lceil 2^m a - t_2 \rceil + 1)\right) \text{ ops.}$$

Hence, the complexity of these steps is of order $\mathcal{O}(N) + \mathcal{O}(2^K) = \mathcal{O}(N) + \mathcal{O}(N^{1/(2\gamma+1)})$. This means that the first phase of computations needs only $\mathcal{O}(N)$ operations, regardless of the actual value of the global smoothness index γ in (6) (*i.e.* independently of the nonlinearity $m(x)$, input density function $f(x)$, and the applied wavelet family). In turn, the second phase, *i.e.* evaluation of the estimate value, requires for a given point x merely $(K - M + 1)(s_2 - s_1)$ operations and therefore has (since M is a constant) the complexity of order $\mathcal{O}(K) = \mathcal{O}(\log N)$.

Remark 3: The resulting complexity of the proposed wavelet coefficients procedure is equal in order to the complexity of the standard FWT algorithm. We would like to emphasize again that in our case the input signal is neither sorted nor deterministic and equidistantly spaced.

III. NUMERICAL EXPERIMENTS

All presented results concerning statistical properties of the computation algorithm $\bar{\mu}_K^H(x)$ are asymptotic in nature, *i.e.* refer to large quantities of data. By means of numerical experiments we briefly illustrate a behavior of $\bar{\mu}_K^H(x)$ for small and moderate number of measurements N and scales K, H. To make the experiments close to practical conditions, the scale K was governed by the *practical selection rule*, $K = \lfloor 1/3 \cdot \log_2 N \rfloor$, in which the factor, $\gamma = 1$, is set independently of the actual (usually unknown) smoothness of $m(x)$ and $f(x)$, yielding a *robust wavelet estimate* (*cf.* [12, p. 652] and the appropriate discussions in [7, Sec. V.C]); the initial scale M was set to zero; *cf.* [24]. Daubechies wavelets with a wavelet number $p = 3$ (having a smoothness index $\nu_\varphi \simeq 1.018$, *cf.* [23, p. 1570]) were employed (thus $\eta = 1$ and $H = K$, see (16)). The input signal was uniformly distributed in the interval $[0, 1]$. The nonlinearity was a polynomial, $m(x) = 10(2x^3 - 3x^2 + x)$, *i.e.* a non-invertible function. The dynamic subsystem had infinite impulse response, $\lambda_i = (0.9)^{-i}$, $i = 0, 1, \ldots$. The results, presented in Figs 2a,b,d, show that the behavior of the computational algorithm, $\bar{\mu}_K^H(x)$, is consistent with the theoretical findings presented in the paper. In particular:

- The graph in Fig. 2a confirms the necessity of adjusting the interpolation scale H to the growing scale of the estimate K; the fixed H makes the interpolation error component non-decreasing (and the computational estimate diverging) in spite of the growth of K; *cf.* also (12) and (16).
- The slope of the bias error (Fig. 2b) for $H = 8$ illustrates the actual vanishing of the error summands, $\overline{\text{bias}}_\varphi\,\bar{g}_K^H(x)$ and $\overline{\text{bias}}_\psi\,\bar{g}_K^H(x)$ at grid points; *cf.* Proposition 2 and the conclusion of Section II.
- Finally, Fig. 2d testifies that the behavior of the algorithm $\bar{\mu}_K^H(x)$ fully resembles that of the much more computationally demanding origin, $\hat{\mu}_K(x)$. (For the qualitative and critical discussion of the properties of $\hat{\mu}_K(x)$ we refer to [7, Sec. VI]).

IV. FINAL REMARKS

It is well known that wavelet functions can also be approximated by using so called *cascade algorithm* or *subdivision scheme*, [23], [3, p. 202]. The error bound for such a method is of order $\mathcal{O}(2^{m/2}2^{-\eta(H+m)})$ with $\eta = \min\{\nu_\varphi, 1\}$, [3, Proposition 6.5.2], and thus the performance of the computational algorithm based on these approximations is comparable with ours (*cf.* (19) in Appendix), but appears a bit worse in practice; see Fig. 2c. Noting further that $\bar{\varphi}^H$ is tantamount to the first order spline interpolation of φ, one can expect improvement for higher order interpolation schemes. However, from Strang-Fix theorem (*cf. e.g.* eq. (9) in [25]) we get that for continuous φ with a Sobolev exponent $\nu_S > 1$ and for
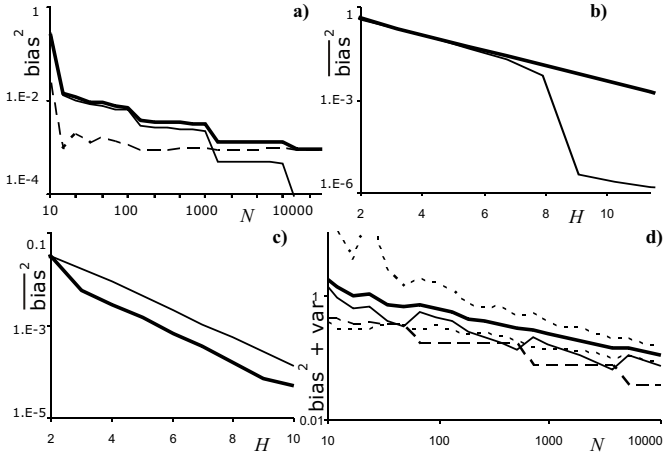
Fig. 2. **a)** Bias error (thick line) and its components induced by approximation (thin line) and interpolation error (dashed line); fixed H and growing K. **b)** Interpolation errors evaluated at the grid 2^{-8} (thin line) and at arbitrary points (thick line); $M = 0$. **c)** Interpolation errors of the estimate $\bar{\mu}_K^H(x)$ with wavelet functions interpolated (thick line) and approximated by cascade algorithm (thin line) **d)** An averaged (50 experiment runs) error of the computational algorithm (thick line) and its components: bias (dashed line) and variance (thin line), for the practical selection rule of K. The overall error graph is 'enveloped' by the best and the worst performances of the estimate during the experiment (thin dotted lines).

ηth order interpolation, $\eta = \lceil \nu_S - 1 \rceil$, the interpolation error order is $\| \varphi_{mn}(x) - \bar{\varphi}_{mn}^H(x) \|_{L^2} = \mathcal{O}(2^{m/2} 2^{-\eta H})$, which means (*cf.* (19) in Appendix) that the theoretical advantage in accuracy of the higher-than-first order interpolations appears not before $\nu_S > 2$ (*i.e.* for Daubechies families with wavelet number $p \geq 7$; see *e.g.* [23, p. 1570]). The adequate counterpart of the rule (16) takes then the form $H = \lfloor (\gamma + 1)/\eta \cdot K \rfloor$. Simultaneously, for smoother wavelets (*e.g.* for the third or of higher number) Daubechies wavelets families for which $\nu_\varphi > 1$ [23, p. 1570]) the rule in (16) reduces to $H = \lfloor \gamma K \rfloor$. Thus, in order to maintain the best possible convergence rate of the reference algorithm (7), the computational algorithms based on ηth order interpolations of (sufficiently smooth) wavelet functions require approximately η times smaller interpolation scales H, however, at the cost of increased computational burden (by nearly the same factor η, [25, p. 754]).

Note also that while selection of both scales K and H controls the limit properties of the algorithm, setting the scale M remains somehow arbitrary, since the rules in (6) and in (16) guarantee that the algorithm converge with the reference rate (17) for any positive M; see (22). Nevertheless, if the estimate values are to be evaluated only at the points of binary grid 2^{-M_H}, for some M_H, then by setting the scale M so that $M + H \geq M_H$, one can remove the second phase computing errors in (14) and enable the weakened rule (18), *cf.* Remark 2.

We emphasize that the particular construction of the proposed computational algorithm has mainly been driven by a possibility of employing of the standard fast routines and the simplicity of the prospective implementation (software or hardware). For instance, the values of the select wavelet function at the required dyadic grid can easily be calculated (and tabulated for further use), and therefore, an implementation of our algorithm seems not to be conceptually more intricate than a one involving *e.g.* trigonometric functions and FFT. We finally note that in contrast to the alternative approaches mentioned in the Introduction, the measurement data are not preprocessed (*e.g.* binned, interpolated or sorted) and thus no extra processing resources are needed.

APPENDIX – PROOF OF THE PROPOSITION

Supports of interpolators and interpolation errors bounds. From the interpolation formulas in (8) one can immediately see that supports of the interpolators are versions of the original wavelet functions supports symmetrically shrunk by $2^{-(H+m+1)}$, that is

$$\text{if supp } \varphi_{mn} = \left[\frac{s_1+n}{2^m}, \frac{s_2+n}{2^m} \right] \text{ then}$$
$$\text{supp } \bar{\varphi}_{mn}^H = \left[\frac{s_1+n}{2^m} + \frac{1}{2^{H+m+1}}, \frac{s_2+n}{2^m} - \frac{1}{2^{H+m+1}} \right].$$

A similar relation holds for ψ_{mn} and $\bar{\psi}_{mn}^H$. These properties, in particular, allow computational algorithms to maintain the general formula of the original estimates in (1)-(3). Recalling that $\varphi \in C^{\nu_\varphi}$ we have

$$\left| \varphi_{mn}(x) - \bar{\varphi}_{mn}^H(x) \right|$$
$$= 2^{\frac{m}{2}} \left| \varphi(2^m x - n) - \varphi\left(\frac{\lfloor 2^{H+m} 2^m x + 1/2 \rfloor}{2^{H+m}} - n \right) \right|$$
$$\leq 2^{\frac{m}{2}} \left| \frac{2^{H+m} 2^m x - \lfloor 2^{H+m} 2^m x + 1/2 \rfloor}{2^{H+m}} \right|^\eta$$

for $\eta = \min\{\nu_\varphi, 1\}$. Clearly $|2^{H+m} 2^m x - \lfloor 2^{H+m} 2^m x + 1/2 \rfloor| \leq 1/2$ and thus

$$\left| \varphi_{mn}(x) - \bar{\varphi}_{mn}^H(x) \right| = \mathcal{O}\left(2^{\frac{m}{2}} 2^{-\eta(H+m+1)} \right), \text{ for any } x. \quad (19)$$

The same bound remains valid for the corresponding pair, ψ_{mn} and $\bar{\psi}_{mn}^H$. Note that the interpolation error decays with the growth of both scales, H and m, simultaneously. In that sense an *effective* interpolation scale for $\bar{\varphi}_{mn}^H$'s (as well as for $\bar{\psi}_{mn}^H$'s) is therefore equal to $H + m$.

Bias error components. We can split the bias error of $\bar{g}_K^H(x)$ into two terms

$$\text{bias } \bar{g}_K^H(x) = g(x) - E\bar{g}_K^H(x)$$
$$= [g(x) - E\hat{g}_K(x)] + [E\hat{g}_K(x) - E\bar{g}_K^H(x)]$$
$$= \text{bias } \hat{g}_K(x) + \overline{\text{bias}} \, \bar{g}_K^H(x)$$

$$\text{bias } \hat{g}_K(x) = \sum_{m=K}^{\infty} \sum_{n=\lceil 2^m x - t_2 \rceil}^{\lfloor 2^m x - t_1 \rfloor} \beta_{mn} \psi_{mn}(x), \text{ and}$$

$$\overline{\text{bias}} \, \bar{g}_K^H(x) = \underbrace{\sum_{n=\lceil 2^M x - s_2 \rceil}^{\lfloor 2^M x - s_1 \rfloor} \alpha_{Mn} [\varphi_{Mn}(x) - \bar{\varphi}_{Mn}^H(x)]}_{\overline{\text{bias}}_\varphi \, \bar{g}_K^H(x)}$$
$$+ \underbrace{\sum_{n=\lceil 2^M x - s_2 \rceil}^{\lfloor 2^M x - s_1 \rfloor} \bar{\varphi}_{Mn}^H(x) [\alpha_{Mn} - E\bar{\alpha}_{Mn}^H]}_{\overline{\text{bias}}_\alpha \, \bar{g}_K^H(x)}$$
$$+ \underbrace{\sum_{m=M}^{K-1} \sum_{n=\lceil 2^m x - t_2 \rceil}^{\lfloor 2^m x - t_1 \rfloor} \beta_{mn} [\psi_{mn}(x) - \bar{\psi}_{mn}^H(x)]}_{\overline{\text{bias}}_\psi \, \bar{g}_K^H(x)}$$
$$+ \underbrace{\sum_{m=M}^{K-1} \sum_{n=\lceil 2^m x - t_2 \rceil}^{\lfloor 2^m x - t_1 \rfloor} \bar{\psi}_{mn}^H(x) [\beta_{mn} - E\bar{\beta}_{mn}^H]}_{\overline{\text{bias}}_\beta \, \bar{g}_K^H(x)}$$

where $\alpha_{Mn} = \langle \varphi_{Mn}(x), g(x) \rangle = E\hat{\alpha}_{Mn}$ and $\beta_{mn} = \langle \psi_{mn}(x), g(x) \rangle = E\hat{\beta}_{mn}$ are appropriate coefficients of the wavelet expansion of $g(x)$; *cf.* [7]; while $E\bar{\alpha}_{Mn}^H = \langle \bar{\varphi}_{Mn}^H(x), g(x) \rangle$ and $E\bar{\beta}_{mn}^H = \langle \bar{\psi}_{mn}^H(x), g(x) \rangle$ are their computational counterparts; *cf.* [13]. For the former error term, bias $\hat{g}_K(x)$, the bound $\mathcal{O}(2^{-\gamma K})$ was derived in [2], [7], [13]. To establish the bound for the latter, $\overline{\text{bias}} \, \bar{g}_K^H(x)$, we need the following lemma.

Lemma 3: The effective interpolation scale of interpolations, $\bar{\varphi}_{mn}^H(x)$ and $\bar{\psi}_{mn}^H(x)$, in the computational coefficients, $\bar{\alpha}_{mn}^H$ and $\bar{\beta}_{mn}^H$, calculated by the FWT algorithm is increased by two with each algorithm step. Thus it holds that:

$$\bar{\alpha}_{mn}^H = \sum_{l=s_1}^{s_2} h_l \bar{\alpha}_{m+1,2n+l}^{H-2} \qquad (20)$$

$$\bar{\beta}_{mn}^H = \sum_{l=-s_2+1}^{-s_1+1} (-1)^n h_{-l+1} \bar{\alpha}_{m+1,2n+l}^{H-2}.$$

Proof: Recall that the FWT algorithm is based on *scaling* and *wavelet* equations, *cf.* [2, p. 255]:

$$\varphi_{mn}(x) = \sqrt{2} \sum_{l=s_1}^{s_2} h_l \varphi_{mn}(2x-l) \qquad (21)$$

$$\psi_{mn}(x) = \sqrt{2} \sum_{l=-s_2+1}^{-s_1+1} (-1)^n h_{-l+1} \varphi_{mn}(2x-l).$$

Applying the first equation to the scaling functions interpolations from (9) we obtain

$$\begin{aligned}
\bar{\varphi}_{mn}^H(x) &= \sqrt{2} \sum_{l=s_1}^{s_2} h_l 2^{\frac{m}{2}} \varphi\left(2\frac{\lfloor 2^{H+m}\cdot 2^m x+1/2\rfloor}{2^{H+m}} - 2n - l\right) \\
&= \sum_{l=s_1}^{s_2} h_l 2^{\frac{m+1}{2}} \varphi\left(\frac{\lfloor 2^{H-2+m+1}\cdot 2^{m+1} x+1/2\rfloor}{2^{H-2+m+1}} - (2n+l)\right) \\
&= \sum_{l=s_1}^{s_2} h_l \bar{\varphi}_{m+1,2n+l}^{H-2}(x)
\end{aligned}$$

and hence

$$\begin{aligned}
\bar{\alpha}_{mn}^H &= \frac{1}{N} \sum_{k=1}^N \bar{\varphi}_{mn}^H(x_k) \cdot y_k \\
&= \sum_{l=s_1}^{s_2} h_l \frac{1}{N} \sum_{k=1}^N \bar{\varphi}_{m+1,2n+l}^{H-2}(x_k) \cdot y_k \\
&= \sum_{l=s_1}^{s_2} h_l \bar{\alpha}_{m+1,2n+l}^{H-2}
\end{aligned}$$

With the use of the wavelet equation in (21), the proof of the lemma for coefficients $\bar{\beta}_{mn}^H$ can be completed in a similar fashion. ∎

Noting now that the initial coefficients for FWT, $\bar{\alpha}_{Kn}^H$, are computed with the effective interpolation scale $H+K$ yields

$$\begin{aligned}
|\alpha_{Mn} - E\bar{\alpha}_{Mn}^H| &= \mathcal{O}\left(2^{-\frac{M}{2}} 2^{-\eta(H+M+1)-2\eta(K-M)}\right) \\
|\beta_{mn} - E\bar{\beta}_{mn}^H| &= \mathcal{O}\left(2^{-\frac{m}{2}} 2^{-\eta(H+m+1)-2\eta(K-m)}\right).
\end{aligned}$$

Putting together the bound in (19), the bounds above and the obvious bounds given below (with the last one shown in *e.g.* [2], [7], [13]):

$$\begin{aligned}
\left|\bar{\varphi}_{Mn}^H(x)\right| &= \mathcal{O}\left(2^{\frac{M}{2}}\right), \\
|\alpha_{Mn}| &= \mathcal{O}\left(2^{-\frac{M}{2}}\right) \\
\left|\bar{\psi}_{mn}^H(x)\right| &= \mathcal{O}\left(2^{\frac{m}{2}}\right), \\
|\beta_{mn}| &= \mathcal{O}\left(2^{-\left(\frac{m}{2}+\gamma m\right)}\right).
\end{aligned}$$

and including that $M < K$, after some elementary algebra, we get that (*cf.* (13)-(14))

$$\begin{aligned}
\overline{\text{bias}}_\alpha\, \bar{g}_K^H(x) &= \mathcal{O}\left(2^{-\eta(H+K+1)-2\eta(K-M)}\right), \\
\overline{\text{bias}}_\beta\, \bar{g}_K^H(x) &= \mathcal{O}\left(2^{-\eta(H+K+1)}\right), \\
\overline{\text{bias}}_\varphi\, \bar{g}_K^H(x) &= \mathcal{O}\left(2^{-\eta(H+M+1)}\right), \\
\overline{\text{bias}}_\psi\, \bar{g}_K^H(x) &= \mathcal{O}\left(2^{-\gamma M-\eta(H+M+1)}\right).
\end{aligned}$$

from which the aggregated bound of the interpolation error in (13)-(14) is directly obtained. To get the variance error bound in (15), it suffices to observe that $\sup_x |\bar{\varphi}^H(x)| \leq \sup_x |\varphi(x)|$ for any H, and apply the same arguments as for the corresponding variance error bound in (5), see [7, Apps II and III]. Finally, the rule in (16) is derived from the inequality $2^{-\eta(H+M+1)} \leq 2^{-\gamma K}$, *i.e.* from the observation that if the order of the interpolation error part of the overall bias error is equal or less than that of the approximation error component, *i.e.* if $\overline{\text{bias}}\, \bar{g}_K^H(x) \preceq \text{bias}\, \hat{g}_K(x)$, then bias $\hat{g}_K(x)$ and bias $\bar{g}_K^H(x)$ are of the same orders and (due to equality of variance errors in (5) and (15)) the convergence rate of the reference algorithm is conveyed without any loss to the computational counterpart. The subsequent relation guarantees therefore the asymptotic rate in (17) for any $M = 0, 1, \ldots$

$$H \geq \tfrac{\gamma}{\eta} K - (M+1) \qquad (22)$$

REFERENCES

[1] A. Cohen and J.-P. D'Ales, "Nonlinear approximation of random functions," *SIAM Journal of Applied Mathematics*, vol. 57, no. 2, pp. 518–540, 1997.

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

[3] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia: SIAM Edition, 1992.

[4] G. Strang, "Wavelet transforms vs. Fourier transforms," *Bulletin of American Mathematical Society*, vol. 28, pp. 288–305, 1993.

[5] D. L. Donoho, "De-noising by soft thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, 1995.

[6] W. Greblicki, "Nonparametric orthogonal series identification of Hammerstein systems," *International Journal of Systems Science*, vol. 20, pp. 2355–2367, 1989.

[7] Z. Hasiewicz, M. Pawlak, and P. Śliwiński, "Non-parametric identification of non-linearities in block-oriented complex systems by orthogonal wavelets with compact support," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 52, no. 1, pp. 427–442, 2005.

[8] A. Kovacz and B. W. Silverman, "Extending the scope of wavelet regression methods by coefficient-dependent thresholding," *Journal of the American Statistical Association*, vol. 95, no. 3, pp. 172–183, 2000.

[9] P. Hall and B. Turlach, "Interpolation methods for nonlinear wavelet regression with irregularly spaced design," *The Annals of Statistics*, vol. 25, no. 5, pp. 1912–1925, 1997.

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

[11] I. Daubechies, I. Guskov, P. Schröder, and W. Sweldens, "Wavelets on irregular point sets," *Phil. Trans. R. Soc. Lond. A*, vol. 357, pp. 2397–2413, 1999.

[12] V. Delouille, J. Simoens, and R. von Sachs, "Smooth design-adapted wavelets for nonparametric stochastic regression," *Journal of the American Statistical Association*, vol. 99, pp. 643–658, 2004.

[13] P. Śliwiński and Z. Hasiewicz, "Computational algorithms for multiscale identification of nonlinearities in Hammerstein systems with random inputs," *IEEE Transactions on Signal Processing*, vol. 53, no. 1, pp. 360–364, 2005.

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

[15] U. Amato, A. Antoniadis, and M. Pensky, "Wavelet kernel penalized estimation for non-equispaced design regression," *Statistics and Computations*, vol. 16, no. 1, pp. 37–55, 2006.

[16] G. B. Giannakis and E. Serpedin, "A bibliography on nonlinear system identification," *Signal Processing*, vol. 81, pp. 533–580, 2001.

[17] S. A. Billings and S. Y. Fakhouri, "Theory of separable processes with application to the identification of non-linear systems," *Proceedings of IEE*, vol. 125, pp. 1051–1058, 1978.

[18] M. Pawlak and Z. Hasiewicz, "Nonlinear system identification by the Haar multiresolution analysis," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 45, no. 9, pp. 945–961, 1998.

[19] T. Söderström and P. Stoica, *System Identification*. Prentice Hall, 1989.

[20] W. Greblicki, "Nonparametric identification of Wiener systems by orthogonal series," *IEEE Transactions on Automatic Control*, vol. 39, no. 10, pp. 2077–2086, 1994.

[21] D. Coca and S. A. Billings, "Non-linear system identification using wavelet multiresolution models," *International Journal of Control*, vol. 74, pp. 1718–1736, 2001.

[22] C. J. Stone, "Optimal rates of convergence for nonparametric regression," *Annals of Statistics*, vol. 8, no. 6, pp. 1348–1360, 1980.

[23] O. Rioul, "Simple regularity criteria for subdivision schemes," *SIAM Journal on Mathematical Analysis*, vol. 23, pp. 1544–1576, 1992.

[24] G. P. Nason, "Choice of wavelet smoothness, primary resolution and threshold in wavelet shrinkage," *Statistics and Computing*, no. 12, pp. 219–227, 2002.

[25] P. Thévenaz, T. Blu, and M. Unser, "Interpolation revisited," *IEEE Transactions on Medical Imaging*, vol. 19, no. 7, pp. 739–758, 2000.