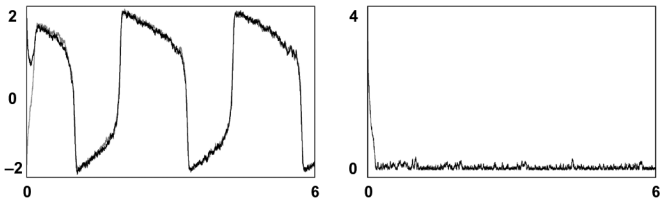

Fig. 1. Synchronization of two noisy FitzHugh–Nagumo oscillators. Left plot: membrane potentials of two coupled noisy FN oscillators. Right plot: absolute difference between the two membrane potentials.

where $i = 1,2$. Let $\mathbf{x} = (v_1, w_1, v_2, w_2)^T$ and $\mathbf{V} = 1/\sqrt{2}\begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix}$. The Jacobian matrix of the projected noise-free system is then given by

$$\begin{pmatrix} c - \frac{c(v_1^2+v_2^2)}{2} - k & c \\ -1/c & -b/c \end{pmatrix}$$

Thus, if the coupling strength verifies $k > c$ then the projected system will be stochastically contracting in the diagonal metric $\boldsymbol{M} = \mathrm{diag}(1, c)$ with rate $\min(k-c, b/c)$ and bound σ^2. Hence, the average absolute difference between the two membrane potentials $|v_1 - v_2|$ will be upper-bounded by $\sigma/\sqrt{\min(1,c)\min(k-c, b/c)}$ after exponential transients (see Fig. 1 for a numerical simulation).

ACKNOWLEDGMENT

The authors wish to thank Dr S. Darses, Dr. D. Bennequin, and Dr. M. Yor, for stimulating discussions, and the Editors and the Reviewers for their helpful comments.

REFERENCES

[1] J.-J. Slotine and W. Li, *Applied Nonlinear Control*. Englewood Cliffs, NJ: Prentice-Hall, 1991.

[2] D. Lewis, "Metric properties of differential equations," *Amer. J. Math.*, vol. 71, pp. 294–312, 1949.

[3] B. Demidovich, "Dissipativity of a nonlinear system of differential equations," *Ser. Mat. Mekh.*, pp. 19–27, 1961.

[4] P. Hartmann, *Ordinary Differential Equations*. New York: Wiley, 1964.

[5] W. Lohmiller and J.-J. Slotine, "Contraction analysis of nonlinear distributed systems," *Int. J. Control*, vol. 78, pp. 678–688, 2005.

[6] J. Jouffroy, "Some ancestors of contraction analysis," in *Proc. IEEE Conf. Decision Control*, 2005, pp. 5450–5455.

[7] E. Sontag and Y. Wang, "Output-to-state stability and detectability of nonlinear systems," *Syst. Control Lett.*, vol. 29, pp. 279–290, 1997.

[8] W. Lohmiller and J.-J. Slotine, "On contraction analysis for nonlinear systems," *Automatica*, vol. 34, pp. 671–682, 1998.

[9] V. Fromion, G. Scorletti, and G. Ferreres, "Nonlinear performance of a PI controlled missile: An explanation," *Int. J. Robust Nonlin. Control*, vol. 9, pp. 485–518, 1999.

[10] D. Angeli, "A Lyapunov approach to incremental stability properties," *IEEE Trans. Automat. Control*, vol. 47, no. 3, pp. 410–421, Mar. 2002.

[11] H. Kushner, *Stochastic Stability and Control*. New York: Academic, 1967.

[12] R. Has'minskii, *Stochastic Stability of Differential Equations*. Rockville, MD: Sijthoff and Nordhoff, 1980.

[13] P. Florchinger, "Lyapunov-like techniques for stochastic stability," *SIAM J. Control Optim.*, vol. 33, pp. 1151–1169, 1995.

[14] X. Mao, *Stability of Stochastic Differential Equations With Respect to Semimartingales*. White Plains, NY: Longman, 1991.

[15] P. Florchinger, "Feedback stabilization of affine in the control stochastic differential systems by the control Lyapunov function method," *SIAM J. Control Optim.*, vol. 35, pp. 500–511, 1997.

[16] J. Tsinias, "The concept of "exponential ISS" for stochastic systems and applications to feedback stabilization," *Syst. Control Lett.*, vol. 36, pp. 221–229, 1999.

[17] H. Deng, M. Krstic, and R. Williams, "Stabilization of stochastic nonlinear systems driven by noise of unknown covariance," *IEEE Trans. Automat. Control*, vol. 46, no. 8, pp. 1237–1253, Aug. 2001.

[18] T. Caraballo, P. Kloeden, and B. Schmalfuss, "Exponentially stable stationary solutions for stochastic evolution equations and their perturbation," *Appl. Math. Optim.*, vol. 50, pp. 183–207, 2004.

[19] N. Aghannan and P. Rouchon, "An intrinsic observer for a class of Lagrangian systems," *IEEE Trans. Automat. Control*, vol. 48, no. 6, pp. 936–945, Jun. 2003.

[20] J. Jouffroy and T. Fossen, "On the combination of nonlinear contracting observers and UGES controllers for output feedback," in *Proc. IEEE Conf. Decision Control*, 2004, pp. 4933–4939.

[21] W. Wang and J.-J. E. Slotine, "On partial contraction analysis for coupled nonlinear oscillators," *Biol Cybern*, vol. 92, no. 1, pp. 38–53, Jan. 2005.

[22] Q.-C. Pham and J.-J. Slotine, "Stable concurrent synchronization in dynamic system networks," *Neural Networks*, vol. 20, no. 1, pp. 62–77, Jan. 2007.

[23] B. Girard, N. Tabareau, Q.-C. Pham, A. Berthoz, and J.-J. Slotine, "Where neuroscience and dynamic system theory meet autonomous robotics: A contracting basal ganglia model for action selection," *Neural Networks*, vol. 21, pp. 628–641, 2008.

[24] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge University Press, 1985.

[25] L. Arnold, *Stochastic Differential Equations: Theory and Applications*. New York: Wiley, 1974.

[26] A. Pavlov, N. van de Wouw, and H. Nijmeijer, *Uniform Output Regulation of Nonlinear Systems*. Boston, MA: Birkhauser, 2006.

[27] Q.-C. Pham, 2007, A Variation of Gronwall's Lemma [Online]. Available: http://arxiv.org/abs/0704.0922

[28] A. Pavlov, A. Pogromsky, N. van de Wouw, and H. Nijmeijer, "Convergent dynamics, a tribute to Boris Pavlovich Demidovich," *Syst. Control Lett.*, vol. 52, pp. 257–261, Jul. 2004.

[29] V. Popov, *Hyperstability of Control Systems*. New York: Springer-Verlag, 1973.

[30] N. Tabareau and J.-J. Slotine, Notes on Contraction Theory MIT, Cambridge, MA, NSL Rep., 2005.

[31] E. Dickmanns, *Dynamic Vision for Intelligent Vehicles*. Cambridge, MA: MIT, 1998.

[32] N. Tabareau, J.-J. Slotine, and Q.-C. Pham, How Synchronization Protects From Noise 2008 [Online]. Available: http://arxiv.org/abs/0801.0011

Wavelet Amendment of Polynomial Models in Hammerstein Systems Identification

Przemysław Śliwiński, Jerzy Rozenblit, Michael W. Marcellin, and Ryszard Klempous

***Abstract*—A new wavelet algorithm for *on-line* improvement of an existing polynomial model of nonlinearity in a Hammerstein system is proposed and its properties are examined. The algorithm employs wavelet bases on interval. Convergence of the resulting assembly, comprising the parametric polynomial model and a nonparametric wavelet add-on, to the system nonlinearity is shown. Rates of convergence for uniformly smooth and piecewise smooth nonlinearities with discontinuities are both established.**

***Index Terms*—Hammerstein system, nonlinear system identification, order statistics, polynomial models, semiparametric approach, wavelet bypass, wavelet regression estimate.**

I. INTRODUCTION

MANY existing models of nonlinear dynamic systems derived from a block-oriented methodology (where models are composed of interconnected static nonlinear and linear dynamic blocks; cf.

Manuscript received July 15, 2008; revised July 16, 2008, September 22, 2008, and October 06, 2008. Current version published April 08, 2009. Recommended by Associate Editor J.-F. Zhang.

P. Śliwiński and R. Klempous are with the The Institute of Computer Engineering, Control and Robotics, Wrocław University of Technology, Wrocław 50-370, Poland (e-mail: przemyslaw.sliwinski@pwr.wroc.pl; ryszard.klempous@pwr.wroc.pl).

J. Rozenblit and M. W. Marcellin are with the Department of Electrical and Computer Engineering, University of Arizona, Tucson, AZ 85721 USA (e-mail: jr@ece.arizona.edu; marcellin@ece.arizona.edu).

Color versions of one or more of the figures in this technical note are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2008.2009620

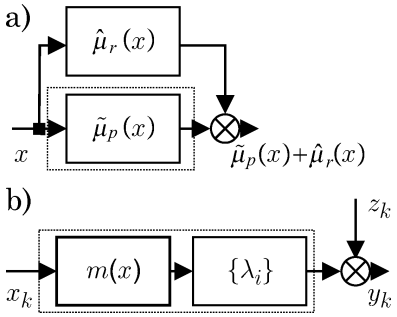

Fig. 1. (a) Scheme of the bypass amendment of the polynomial model of nonlinearity. (b) Identified Hammerstein system.

[1], [2]) employ parametric representations of the system nonlinearities; cf. e.g. [3]–[7]. Such models, yet simple, can offer only a crude approximation if, for instance, they are based on polynomials and the genuine nonlinearity turns out to be e.g. a piecewise smooth function with discontinuities, cf. [8]. To eliminate this inaccuracy, we propose an algorithm which amends the already existing polynomial model by means of a *nonparametric wavelet bypass* module [as depicted in Fig. 1(a)]

- without affecting this model (being, for instance, hard-wired or implemented as a software library; see e.g. System Identification Toolbox for Matlab, [9]);
- in a convenient *on-line* fashion.

The algorithm, inspired by a general concept of additive semiparametric regression, see e.g. [10]–[13], employs nonparametric regression estimates based on order statistics (see e.g. [14], [15]) and orthogonal wavelet expansions on intervals (see e.g. [16]–[18]). A similar approach, which can be found in [19], can only be applied to recover a nonlinearity (*off-line*) in static systems with a deterministic input.

II. Assumptions and Preliminaries

We consider a Hammerstein system [viz. a cascade of static nonlinearity followed by a linear dynamics; see Fig. 1(b)], being one of the most prevalent example of block-oriented nonlinear dynamic models found in the literature (cf. Remark 2).

The goal of the algorithm is to recover a nonlinearity in the Hammerstein system described by the input-output equation

$$y_k = \sum_{i=0}^{\infty} \lambda_i m(x_{k-i}) + z_k = \lambda_0 m(x_k) + \sum_{i=1}^{\infty} \lambda_i m(x_{k-i}) + z_k \quad (1)$$

with the following assumptions:

1) An input signal, $\{x_k\}$, and an external additive noise, $\{z_k\}$, are zero-mean second-order random stationary processes. They are mutually independent. The input $\{x_k\}$ is white and has a density, $f(x)$, which is strictly positive in the identification interval, say [0,1].
2) A nonlinear characteristic of the static system, $m(x)$, has a Hölder continuity exponent $\nu > 0$; cf. e.g. [18, Ch. VI].
3) A linear dynamic subsystem is asymptotically stable. Its impulse response, $\{\lambda_i\}$, $i = 0, 1, \ldots$, is unknown.
4) A set, $\{(x_l, y_l)\}$, $l = 1, 2, \ldots, k$, of the system input and output measurements is available; cf. Remark 1. Moreover:
5) There exists a (pre-)model of the nonlinearity, $\tilde{\mu}_p(x)$, based on a polynomial of order up to $p-1$, for some known p.

The above assumptions are of mixed nature. The first four are typical for nonparametric identification tasks (see [14], [15] and the references cited therein), and impose only weak constraints on system's characteristics and signals. In particular, Assumption 2 admits, for $\nu > 1$, nonlinearities with $\lceil \nu \rceil - 1$ continuous derivatives; for $0 < \nu < 1$, continuous nonlinearities, but also discontinuous nonlinearities with separate jumps as well; cf. [18, Ch. VI]. Therefore, virtually all nonlinearities bounded on an interval are taken into account (the unit interval in Assumption 1 is used only for convenience). Moreover, the impulse response $\{\lambda_i\}$ can be finite or not, and the external noise $\{z_k\}$ can be white or correlated and of any probability distribution with a finite variance. In turn, Assumption 5 supposing the polynomial model of the nonlinearity occurs in several parametric identification algorithms, cf. e.g.: [5], [20], [21]. Here, the model $\tilde{\mu}_p(x)$ can be a standard polynomial, $\tilde{a}_0 + \tilde{a}_1 x + \cdots + \tilde{a}_{p-1} x^{p-1}$ or can, for example, be composed of the first p terms of Legendre polynomial series.

Remark 1: It is well known that due to a composite structure of Hammerstein systems and the lack of measurements of the interconnecting signal (Assumption 4), a nonlinearity, $\mu(x) = am(x) + b$, where $a = \lambda_0$, $b = Em(x_1)\sum_{i=1}^{\infty} \lambda_i$, that is, a scaled and shifted version of the characteristic $m(x)$ of the static block, can at most be recovered from input-output measurements. We emphasize that this inability is a property of the system and hence occurs for any identification algorithm—either parametric (e.g. that leading to the model $\tilde{\mu}_p(x)$) or nonparametric (e.g. ours); cf. [22, S. VI] or [7, S. III]. We also assume—for simplicity of the presentation—that $\lambda_0 \neq 0$; any other $\lambda_i \neq 0$ can be used as well, cf. Assumption 3.

Remark 2: Other structures to which the proposed algorithm can *directly* be applied, like a parallel system, Uryson and MISO systems, etc., are demonstrated in [23]–[25]).

A. Wavelets on Intervals

The most common wavelets (and hence the most often ones present in the literature and applications) are the compactly supported wavelets invented by Daubechies, [26]. However, they constitute orthogonal bases on the real line only and cannot directly be used on intervals. Thus we employ more specific wavelet bases on interval, being proposed in [16], [17], i.e. *CDJV* wavelets. They preserve orthogonality, multiresolution property, fast computational algorithms and, in particular, the number, $p = 2, 3, \ldots$, of vanishing moments of the original Daubechies wavelets.

Wavelet bases on intervals consist of three types of functions: the left and the right end, and the internal ones (which factually are the Daubechies functions). Namely, for a given p, the basis consists of a single set of scaling functions, $\{\varphi_{Mn}(x)\} = \{\varphi^{left}_{Mn}(x), \varphi^{int}_{Mn}(x), \varphi^{right}_{Mn}(x)\}$, for $0 \leq n < p$, $p \leq n < 2^M - p$, and $2^M - p \leq n < 2^M$, respectively, and sets of wavelets, $\{\psi_{mn}(x)\} = \{\psi^{left}_{mn}(x), \psi^{int}_{mn}(x), \psi^{right}_{mn}(x)\}$, for the same ranges of n and increasing scales $m = M, M+1, \ldots$. The initial scale M needs (by design) to be sufficiently large in order to avoid the boundary functions on the one end to intersect with the other. Specifically, the supports of the functions at the left end are 'staggered', i.e. equal to $[0, 2^{-M}(n+p)]$, $0 \leq n < p$, (similarly at the right end), and this implies $M \geq \log_2 2p$ for a unit interval, [17, Th. 4.4].

Remark 3: The vanishing moments property is pivotal for our algorithm. Defined as

$$\int_0^1 x^i \cdot \psi_{mn}(x)dx = 0 \text{ for } i = 0, \ldots, p-1 \quad (2)$$

it means that wavelets with p vanishing moments are *orthogonal* to monomials of order up to $p-1$, and hence to *any polynomial model* $\tilde{\mu}_p(x)$, cf. Assumption 5

$$\langle \tilde{\mu}_p, \psi_{mn} \rangle = 0. \tag{3}$$

Any residual nonlinearity of the form $\mu_r(x) = \mu(x) - \tilde{\mu}_p(x)$, which is square integrable in the interval [0,1], has therefore a *wavelet expansion*

$$\mu_r(x) = \sum_{n=0}^{2^M-1} \alpha_{Mn}\varphi_{Mn}(x) + \sum_{m=M}^{\infty}\sum_{n=0}^{2^m-1} \beta_{mn}\psi_{mn}(x) \tag{4}$$

with the coefficients ($\alpha^p_{Mn} = \langle \tilde{\mu}_p, \varphi_{Mn} \rangle$)

$$\begin{aligned} \alpha_{Mn} &= \langle \mu - \tilde{\mu}_p, \varphi_{Mn} \rangle = \langle \mu, \varphi_{Mn} \rangle - \alpha^p_{Mn} \\ \beta_{mn} &= \langle \mu - \tilde{\mu}_p, \psi_{mn} \rangle = \langle \mu, \psi_{mn} \rangle. \end{aligned} \tag{5}$$

III. Identification Algorithm

We begin with the observation made in [22] that for Hammerstein systems it holds that [cf. (1)]:

$$E\,(y_k | x_k = x) = \mu(x) \tag{6}$$

i.e., the identified nonlinearity $\mu(x)$ is a regression function of the output y_k on the input x_k. Having, by Assumption 5, its polynomial model $\tilde{\mu}_p(x)$, we are interested in the remaining part

$$\mu_r(x) = \mu(x) - \tilde{\mu}_p(x) \tag{7}$$

referred further to as a *residual nonlinearity*, cf. [11, Ch. 9]. This function is square integrable in [0,1] (cf. Assumptions 2 and 5) and for its recovery we propose the following wavelet estimate (see (4) and cf. [27]):

$$\hat{\mu}_r(x) = \sum_{n=0}^{2^M-1} \hat{\alpha}_{Mn}\varphi_{Mn}(x) + \sum_{m=M}^{K-1}\sum_{n=0}^{2^m-1} \hat{\beta}_{mn}\psi_{mn}(x) \tag{8}$$

where K is the estimate scale increasing with a growing number of measurements k. The expansion coefficients estimates, $\hat{\alpha}_{Mn}$ and $\hat{\beta}_{mn}$, are computed from ordered observations, i.e. from the measurements set, $\{(x_l, y_l)\}$, $l = 1, \ldots, k$, sorted pairwise according to increasing values of x_l and supplemented with an extra pair $(x_0 = 0, y_0 = 0)$ (see (5) and cf. [27]), according to the rules:

$$\begin{aligned} \hat{\alpha}_{Mn} &= \sum_{l=1}^{k} y_l \cdot \int_{x_{l-1}}^{x_l} \varphi_{Mn}(x)dx - \alpha^p_{Mn} \\ \hat{\beta}_{mn} &= \sum_{l=1}^{k} y_l \cdot \int_{x_{l-1}}^{x_l} \psi_{mn}(x)dx. \end{aligned} \tag{9}$$

Note that k, being the number of measurements (cf. Assumption 4) is also interchangeably used as an index of the last measurement pair in the ordered set.

A. Computational Issues

Since explicit integrations in (9) are not possible (as the wavelet functions are not given in the explicit form, cf. [26], [28]), we introduce the equivalent *recursive* versions of the coefficients estimates $\hat{\alpha}_{Mn}$ and $\hat{\beta}_{mn}$, which enable (numerical) implementation of the algorithm in a convenient *on-line* way.

Given the ordered sequence, $\{(x_1, y_1), \ldots, (x_l, y_l), (x_{l+1}, y_{l+1}), \ldots, (x_k, y_k)\}$, assume that for the new, $(k+1)$th measurement pair, (x_{k+1}, y_{k+1}), it holds that $x_l < x_{k+1} < x_{l+1}$. Then, *(i)* the new pair is inserted between (x_l, y_l) and (x_{l+1}, y_{l+1}) to maintain the ascending order of the updated measurement set, and *(ii)* the following recurrence formulas are applied to compute the coefficients estimates (a vector-like notation is used for shortness):

$$\begin{bmatrix} \hat{\alpha}^{(k+1)}_{Mn} \\ \hat{\beta}^{(k+1)}_{mn} \end{bmatrix} = \begin{bmatrix} \hat{\alpha}^{(k)}_{Mn} \\ \hat{\beta}^{(k)}_{mn} \end{bmatrix} + (y_{k+1} - y_{l+1}) \begin{bmatrix} \Phi_{Mn}(x_{k+1}) - \Phi_{Mn}(x_l) \\ \Psi_{mn}(x_{k+1}) - \Psi_{mn}(x_l) \end{bmatrix} \tag{10}$$

for $k = 1, 2, \ldots$, where $\Phi_{Mn}(x)$ and $\Psi_{mn}(x)$ are *formally* defined as the definite integrals

$$\Phi_{Mn}(x) = \int_0^x \varphi_{Mn}(v)dv \text{ and } \Psi_{mn}(x) = \int_0^x \psi_{mn}(v)dv \tag{11}$$

and can easily be evaluated numerically (see e.g. [29, Ch. 4] for numeric integration algorithms; the fast routines computing wavelet function values in binary grid points are presented in, e.g., [28]).

The procedure starts with the following initial values and initial measurement set, cf. (2), (3) and (5):

$$\begin{bmatrix} \hat{\alpha}^{(1)}_{Mn} \\ \hat{\beta}^{(1)}_{mn} \end{bmatrix} = \begin{bmatrix} -\alpha^p_{Mn} \\ 0 \end{bmatrix} \text{ and } \left\{ \begin{matrix} (x_0 = 0, y_0 = 0), \\ (x_1 = 1, y_1 = 0) \end{matrix} \right\}. \tag{12}$$

Remark 4: Both versions of the algorithm, the *off-line* in (9) and the *on-line* one in (10)–(12), require the measurement set to be kept (preferably—in the ordered form).

IV. Algorithm Properties

Before we pass to the formal part of the algorithm analysis, we will briefly discuss the meaning of the wavelet expansion coefficients estimates in (9). First, notice that regardless of its actual type, any polynomial model $\tilde{\mu}_p(x)$ has the following equivalent representations (cf. (5) and Assumption 5):

$$\tilde{\mu}_p(x) = \sum_{i=0}^{p-1} \tilde{a}_i x^i = \sum_{n=0}^{2^M-1} \alpha^p_{Mn}\varphi_{Mn}(x) \tag{13}$$

for any parameters set $\{\tilde{a}_i\}$, $i = 0, \ldots, p-1$. The estimates $\hat{\alpha}_{Mn}$, being initiated with $-\alpha^p_{Mn}$ [cf. (9) and (12)], are therefore able to accommodate, in a course of identification process, the *contingent differences* between the model $\tilde{\mu}_p(x)$ and the nonlinearity $\mu(x)$, in case when $\mu(x)$ is also a polynomial of order $p-1$. In turn, if $\mu(x)$ is either a polynomial of higher order or not a polynomial but, e.g., a piecewise-smooth function, then all estimates $\hat{\alpha}_{Mn}$ and $\hat{\beta}_{mn}$ are needed to recover the part of $\mu(x)$ not represented by polynomial model $\tilde{\mu}_p(x)$; cf. (2) and (3).

Remark 5: The estimates $\hat{\alpha}_{Mn}$ and $\hat{\beta}_{mn}$ can prospectively be used in a two-stage polynomial model validation procedure: the statistical insignificance of the wavelet coefficients estimates $\hat{\beta}_{mn}$ would testify, for instance, to the sufficiency of the model order. If additionally all $\hat{\alpha}_{Mn}$'s remained insignificant, then the complete stand-alone model $\tilde{\mu}_p(x)$ could be assumed to be sufficient.

A. Convergence

We will now show that the estimate $\hat{\mu}_r(x)$ converges in the mean integrated square error (MISE) sense, to the residual nonlinearity $\mu_r(x)$ (and, *a fortiori*, the assembly $\tilde{\mu}_p(x) + \hat{\mu}_r(x)$ to the genuine nonlinearity $\mu(x)$) with a growing number of the measurement data. The first theorem provides with conditions of the convergence.

Theorem 1: Let the assumptions (1)–(5) hold. If the estimate scale K increases with the measurements number k so that

$$K(k) \to \infty \text{ and } 2^{K(k)}/k \to 0 \text{ as } k \to \infty \tag{14}$$

then

$$\mathrm{MISE}\hat{\mu}_r = E\int_0^1 [\mu_r(x) - \hat{\mu}_r(x)]^2\,dx \to 0 \text{ as } k \to \infty.$$

Proof: See Appendix A. ∎

The following conclusions can immediately be drawn from the theorem:

- In order to assure the convergence of $\hat{\mu}_r(x)$ to $\mu_r(x)$, the scale K needs *only* to increase, however, the rate of this growth should mutually allow the term $2^K/k$ to vanish as $k \to \infty$.
- The convergence of the assembly $\tilde{\mu}_p(x) + \hat{\mu}_r(x)$ to $\mu(x)$ holds for any estimate scales selected as $K(k) = \lceil c\log_2 k\rceil$, where $0 < c < 1$.

We emphasize that neither the structure of the dynamic part $\{\lambda_i\}$ nor the presence of a correlated noise $\{z_k\}$ affects the convergence. Similar conditions have earlier been established for wavelet-based ratio estimates of nonlinearities; see [23]–[25]—here, moreover, the convergence holds for arbitrary model $\tilde{\mu}_p(x)$ compliant with Assumption 5, i.e. for any polynomial of order not greater than $p-1$, see (13).

B. Convergence Rate

In Theorem 1 the convergence conditions are characterized. It is however also interesting to establish the rate the estimate converges with—as it seems to be intuitively clear that convergence to smoother nonlinearities should be faster than to the irregular ones. Our second theorem provides a formal support of this intuition and shows how to select the constant c (according to the Hölder exponent ν of the nonlinearity m and to the order of the polynomial pre-model p, to make the convergence rate the fastest.

Theorem 2: If the estimate scale K is increased according to the following rule

$$K(k) = \lceil c\log_2 k\rceil \text{ where } c = \frac{1}{2\gamma+1} \tag{15}$$

with $\gamma = \min\{\nu, p\}$ or $\gamma = \min\{\nu, 1/2\}$ for continuous or piecewise-continuous nonlinearities $\mu(x)$, respectively, then the estimate $\hat{\mu}_r(x)$ converges to the residuum $\mu_r(x)$ in the MISE error sense with the asymptotic rate

$$\mathrm{MISE}\hat{\mu}_r = \mathcal{O}\left(k^{-\frac{2\gamma}{2\gamma+1}}\right)$$

and the assembly $\tilde{\mu}_p(x) + \hat{\mu}_r(x)$ converges to the nonlinearity $\mu(x)$ with the same rate.

Proof: See Appendix B. ∎

Note that the best possible convergence rate (amongst all nonparametric estimates for a given ν; see e.g. [11], [30]) is achieved by our estimate. Moreover:

- the rate $\mathcal{O}(k^{-2\gamma/(2\gamma+1)})$ approaches, for large γ, i.e., for smooth nonlinearities and for wavelets with $p \geq \nu$ vanishing moments, the rate $\mathcal{O}(k^{-1})$, which is the best attainable for *parametric* estimates, [31].
- the rate $\mathcal{O}(k^{-1/2})$ is obtained for discontinuous, piecewise smooth nonlinearities, i.e. having an arbitrary finite number of jumps. This rate is also the best possible for that nonlinearities and, furthermore, it can only be achieved by those estimates (amongst *all* orthogonal expansion-based) which employ wavelets with compact support (cf. [32]).

Remark 6: If we apply wavelets with a number of vanishing moments lower than the order of $\tilde{\mu}_p(x)$ or we relax Assumption 5 and admit arbitrary models (like e.g. those considered in [33]), then the representation in (13) (and hence the algorithm convergence) does not hold. Nevertheless, the convergence will be maintained if we subtract $\beta^p_{mn} = \langle\tilde{\mu}_p, \psi_{mn}\rangle$ from every new coefficient $\hat{\beta}_{mn}$, cf. (5) and (9). The rate of the algorithm convergence will then become dependent also on the model smoothness (in the analogous way it depends now on the smoothness of the nonlinearity) and e.g. for a model with Hölder exponent ν' we will get $\gamma' = \min\{\nu', \gamma\}$; cf. Appendix B.

V. Numerical Illustration

The properties presented in Sections IV-A, IV-B characterize asymptotic behavior of our algorithm. To get an insight into its performance for small and moderate number of measurements, several numerical experiments were made. Specifically, to illustrate Theorem 2 for continuous and discontinuous nonlinearities, the following characteristics:

$$m_1(x) = 5(x^5 - x^3) \text{ and } m_2(x) = \begin{cases} -1 & \text{if } x < 1/2 \\ 1 & \text{if } 1/2 \leq x \end{cases}$$

were coupled with wavelet estimates, $\hat{\mu}_r(x)$, employing CDJV wavelets with $p = 5$ vanishing moments, and with the scale governed by *the practical selection rule*, $K(k) = \lceil 1/3 \cdot \log_2 k\rceil$; cf. (15) and see [25].

The input $\{x_l\}$ was uniformly distributed in [0,1], and the (infinite) impulse response of the dynamic part was $\lambda_i = 2^{-i}$, $i = 0, 1, \ldots$ (thus we have either $\mu(x) = m_1(x)$ or $\mu(x) = m_2(x)$, cf. Remark 1); the external uniform noise was set to make $\max|z_l|/\max|m(x)| = 10\%$.

A preliminary model based on Legendre polynomials was used: $\tilde{\mu}_p(x) = \sum_{i=0}^4 \tilde{\alpha}_i p_i(x)$, where $p_i(x) = \sqrt{2i+1}\cdot P_i(x)$, and where $P_i(x) = (2i-1)xP_{i-1}(x) + (i-1)P_{i-2}(x)$ [with $P_0(x) = 1$ and $P_1(x) = \sqrt{3}(2x-1)$] are Legendre polynomials orthogonal on [0,1]; cf. [34]. For each characteristic, the best (in the mean square sense) polynomial model parameters were evaluated numerically, i.e. $\tilde{\alpha}_i = \sum_{l=1}^L m(x)_l \int_{(l-1)/L}^{l/L} p_i(x)dx$, $L = 500$; cf. (9).

The results are shown in Fig. 2. The solid lines represent the systematic (approximation) error of the initial polynomial models, $\tilde{\mu}_p(x)$, while the dotted ones exhibit vanishing of the MISE error of the assemblies $\tilde{\mu}_p(x) + \hat{\mu}_r(x)$, validating the proposed approach for both smooth and discontinuous nonlinearities.

VI. Conclusion

In the technical note we introduced the wavelet algorithm improving accuracy of polynomial models of nonlinearities in Hammerstein system. The proposed bypass-like solution offers a (non-intrusive) adjustment/calibration of existing pre-models based on various types of polynomials. Application of wavelets offers several advantages:

- Wavelets approximate irregular nonlinearities better than polynomials due to compactness of their supports and subsequent good localization properties.
- There exist fast implementations of wavelet algorithms (`Matlab`, `C/C++/C#/Java`, etc. see, e.g., [17], [18], [29] for wavelet transform algorithms and [28] for routines computing wavelet values).

The proposed idea of improvement of the existing parametric polynomial models seems to be of interest not only from a theoretical vantage point, cf. [10]–[13], but also important in practice as the aforementioned polynomial models have been already utilized in modelling financial processes and data transmission channels, organs like eyes, skin and muscles, or distillation columns (see [35]–[40]).

Our algorithm can therefore furnish a better understanding of an investigated phenomena (e.g. to verify or enhance established laws, cf. [11, p. 313]) or a refined nonlinearity compensation (e.g. to enable application of linear optimization algorithms). We also point out that the recursive implementation of the algorithm makes it suitable for identification and tracking time-varying nonlinear systems, [41], [42].

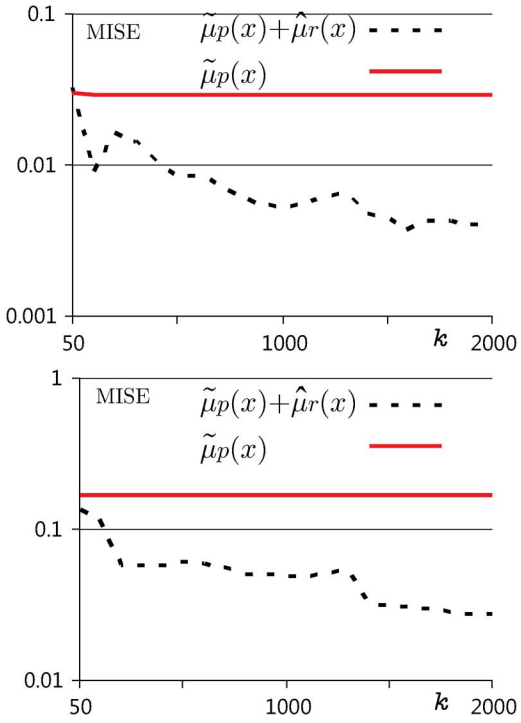

Fig. 2. Errors for (a) polynomial and (b) step-function nonlinearities.

Appendix A
Proof of Theorem 1

A mean integrated squared error of the estimate $\hat{\mu}_r(x)$ is defined in a standard way as

$$\mathrm{MISE}\hat{\mu}_r = E\int_0^1 (\mu_r(x) - \hat{\mu}_r(x))^2\, dx.$$

Using (4) and (8), the error can conveniently be expressed in terms of bias and variance errors of wavelet coefficients estimates, viz.

$$\begin{aligned}\mathrm{MISE}\hat{\mu}_r = &\underbrace{\sum_{m=K}^{\infty}\sum_{n=0}^{2^m-1}\beta_{mn}^2}_{\mathrm{approx}^2\mu_r^K}\\ &+\underbrace{\sum_{n=0}^{2^M-1}\mathrm{bias}^2\hat{\alpha}_{Mn}+\sum_{m=M}^{K-1}\sum_{n=0}^{2^m-1}\mathrm{bias}^2\hat{\beta}_{mn}}_{\mathrm{bias}^2\hat{\mu}_r}\\ &+\underbrace{\sum_{n=0}^{2^M-1}\mathrm{var}\hat{\alpha}_{Mn}+\sum_{m=M}^{K-1}\sum_{n=0}^{2^m-1}\mathrm{var}\hat{\beta}_{mn}}_{\mathrm{var}\hat{\mu}_r}. \end{aligned} \quad (16)$$

The first term, $\mathrm{approx}^2\mu_r^K$, is an integrated (over the interval [0,1]) squared error of approximation of the residual nonlinearity, $\mu_r(x)$, by its wavelet series (4) truncated at the scale $K-1$. For this error it holds naturally (due to a completeness of a wavelet basis) that

$$\mathrm{approx}^2\mu_r^K = \sum_{m=K}^{\infty}\sum_{n=0}^{2^m-1}\beta_{mn}^2 \to 0 \text{ as } K\to\infty. \quad (17)$$

The remaining terms, $\mathrm{bias}^2\hat{\mu}_r$, and $\mathrm{var}\hat{\mu}_r$ are integrated squared bias and variance errors of $\hat{\mu}_r$, composed of squared bias and variance errors of the wavelet expansion coefficients estimates $\hat{\alpha}_{Mn}$ and $\hat{\beta}_{mn}$ in (8), respectively. From the following simple decomposition of β_{mn}, cf. (5) and (11):

$$\beta_{mn} = \sum_{l=1}^{k}\int_{x_{l-1}}^{x_l}\mu(x)\psi_{mn}(x)dx + \int_{x_k}^{1}\mu(x)\psi_{mn}(x)dx$$

we easily get that (cf. eqs. (14)–(18) in [14, p. 1477])

$$\mathrm{bias}\hat{\beta}_{mn} = E\int_{x_k}^{1}\mu(x)\psi_{mn}(x)dx$$

which in turn (by virtue of (B.4) in [14, p. 1488]) results in $\mathrm{bias}\hat{\beta}_{mn} = \mathcal{O}(k^{-1})$ and, after application of the same arguments to $\hat{\alpha}_{Mn}$, in $\mathrm{bias}\hat{\alpha}_{Mn} = \mathcal{O}(k^{-1})$. Thus [cf. (16)]

$$\mathrm{bias}^2\hat{\mu}_r = \mathcal{O}\left(\frac{2^M}{k^2}+\sum_{m=M}^{K-1}\frac{2^m}{k^2}\right) = \mathcal{O}\left(\frac{2^K}{k^2}\right). \quad (18)$$

Using now Lemma 1 from [14, p. 1475] we get that

$$\mathrm{var}\hat{\alpha}_{Mn} = \mathcal{O}\left(\frac{2^M}{k}\right) \text{ and } \mathrm{var}\hat{\beta}_{mn} = \mathcal{O}\left(\frac{2^m}{k}\right)$$

and hence [cf. (16)]

$$\mathrm{var}\hat{\mu} = \mathcal{O}\left(\frac{2^M}{k}+\sum_{m=M}^{K-1}\frac{2^m}{k}\right) = \mathcal{O}\left(\frac{2^K}{k}\right). \quad (19)$$

Taking together (14), and (16)–(19) concludes the proof. ■

Appendix B
Proof of Theorem 2

Recalling that if the nonlinearity $\mu(x)$ has a Hölder exponent ν (and so does $\mu_r(x)$), then

$$\beta_{mn} = \mathcal{O}\left(2^{-\frac{2\gamma+1}{2}m}\right), \text{ where } \gamma = \min\{\nu, p\} \quad (20)$$

(see e.g. [18]), and combining it with the fact that the number of wavelet coefficients β_{mn} at each scale $m = K, K+1, \ldots$, equals 2^m [see e.g. [16]–[18] and cf. (4)], yields

$$\mathrm{approx}^2\mu_r^K = \mathcal{O}(2^{-2\gamma K}). \quad (21)$$

For piecewise-smooth nonlinearities, i.e., those having separate jump-type discontinuities and a Hölder exponent ν between them, the smooth parts remain characterized by wavelet coefficients with the bound as in (20), however, at each scale m there is also a finite number of coefficients of order $\mathcal{O}(2^{-m/2})$ corresponding to wavelets with supports located in the *"cones of influence"* of nonlinearity jumps, [18, Fig. 6.1]. Their presence deteriorates the convergence rate of the approximation error, giving in result a bound as in (21), yet with $\gamma = \min\{\nu, 1/2\}$; cf. [24, S. V].

Putting (18), (19) and (21) into (16) yields

$$\begin{aligned}\mathrm{MISE}\hat{\mu}_r &= \mathcal{O}\left(2^{-2\gamma K}+\left(1+\frac{1}{k}\right)\cdot\frac{2^K}{k}\right)\\ &= \mathcal{O}\left(2^{-2\gamma K}+\frac{2^K}{k}\right).\end{aligned}$$

Application of the rule (15), completes the proof. ■

References

[1] S. A. Billings, "Identification of non-linear systems—A survey," *Proc. IEE*, vol. 127, no. 6, pp. 272–285, 1980.
[2] S. A. Billings and S. Y. Fakhouri, "Identification of systems containing linear dynamic and static non-linear elements," *Automatica*, vol. 18, pp. 15–26, 1982.
[3] R. Haber and L. Keviczky, *Nonlinear System Parameter Identification*. London, U.K.: Kluwer Academic Publishers, 1999.

[4] L. Ljung, *System Identification—Theory for the User*. Englewood Cliffs, NJ: Prentice-Hall, 1987.
[5] T. Söderström and P. Stoica, *System Identification*. Englewood Cliffs, NJ: Prentice Hall, 1989.
[6] E. W. Bai and D. Li, "Convergence of the iterative Hammerstein system identification algorithm," *IEEE Trans. Automat. Control*, vol. 49, no. 11, pp. 1929–1940, Nov. 2004.
[7] Z. Hasiewicz and G. Mzyk, "Combined parametric-nonparametric identification of Hammerstein systems," *IEEE Trans. Automat. Control*, vol. 49, no. 8, pp. 1370–1375, Aug. 2004.
[8] J. Vörös, "Identification of Hammerstein systems with time-varying piecewise-linear characteristics," *IEEE Trans. Circuits Syst. II: Express Briefs*, vol. 52, no. 12, pp. 865–869, Dec. 2005.
[9] *"System Identification Toolbox 7, User's Guide"* 2008 [Online]. Available: http://www.mathworks.com/access/helpdesk/help/pdf_doc/ident/ident.pdf
[10] C. J. Stone, "Additive regression and other nonparametric models," *Annals Stat.*, vol. 13, no. 2, pp. 689–705, 1985.
[11] W. Härdle, *Applied Nonparametric Regression*. Cambridge, U.K.: Cambridge University Press, 1990 [Online]. Available: http://www.quantlet.com/mdstat/scripts/anr/pdf/anrpdf.pdf
[12] W. Härdle, M. Müller, S. Sperlich, and A. Werwatz, *Nonparametric and Semiparametric Models*. Berlin, Germany: Springer-Verlag, 2004.
[13] D. Ruppert, M. P. Wand, and R. Carroll, *Semiparametric Regression*, ser. Cambridge Series in Statistical and Probabilistic Mathematics. New York: Cambridge University Press, 200.
[14] W. Greblicki and M. Pawlak, "Dynamic system identification with order statistics," *IEEE Trans. Inform. Theory*, vol. 40, no. 5, pp. 1474–1489, Sep. 1994.
[15] W. Greblicki, "Nonlinearity estimation in Hammerstein systems based on ordered observations," *IEEE Trans. Signal Processing*, vol. 44, no. 5, pp. 1224–1233, May 1996.
[16] L. Andersson, N. Hall, B. Jawerth, and G. Peters, "Wavelets on closed subsets of the real line," in *Topics in the Theory and Applications of Wavelets*, L. L. Schumaker and G. Webb, Eds. Boston, MA: Academic Press, 1993, p. 60 [Online]. Available: http://citeseer.ist.psu.edu/3753.html
[17] A. Cohen, I. Daubechies, and P. Vial, "Wavelet bases on the interval and fast algorithms," *J. Appl. Computat. Harmonic Anal.*, vol. 1, no. 1, pp. 54–81, 1993.
[18] S. G. Mallat, *A Wavelet Tour of Signal Processing*. San Diego, CA: Academic Press, 1998.
[19] T. C. M. Lee and H.-S. Oh, "Automatic polynomial wavelet regression," *Stat. Comput.*, vol. 14, pp. 337–341, 2004.
[20] N. D. Haist, F. H. L. Chang, and R. Luus, "Non-linear identification in the presence of the correlated noise using Hammerstein model," *IEEE Trans. Automat. Control*, vol. 18, no. 5, pp. 552–555, Oct. 1973.
[21] D. T. Westwick and R. E. Kearney, "Separable least squares identification of nonlinear Hammerstein models: Application to stretch reflex dynamics," *Annal. Biomed. Eng.*, vol. 29, no. 8, pp. 707–718, 2001.
[22] W. Greblicki and M. Pawlak, "Identification of discrete Hammerstein system using kernel regression estimates," *IEEE Trans. Automat. Control*, vol. AC-31, no. 1, pp. 74–77, Jan. 1986.
[23] M. Pawlak and Z. Hasiewicz, "Nonlinear system identification by the Haar multiresolution analysis," *IEEE Trans. Circuits Syst. I*, vol. CAS-45, no. 9, pp. 945–961, Sep. 1998.
[24] Z. Hasiewicz and P. Śliwiński, "Identification of non-linear characteristics of a class of block-oriented non-linear systems via Daubechies wavelet-based models," *Int. J. Syst. Sci.*, vol. 33, no. 14, pp. 1121–1144, 2002.
[25] 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 Trans. Circuits Syst. I*, vol. 52, no. 1, pp. 427–442, Jan. 2005.
[26] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992.
[27] W. Greblicki and P. Śliwiński, "Non-parametric identification of nonlinearity in Hammerstein system," in *Proc. 3th IFAC Symp. Syst. Identification SYSID-2003*, Rotterdam, The Netherlands, Aug. 26–28, 2003, pp. 60–65.
[28] P. Śliwiński and Z. Hasiewicz, "Computational algorithms for wavelet identification of nonlinearities in Hammerstein systems with random inputs," *IEEE Trans. Signal Processing*, vol. 56, no. 2, pp. 846–851, Feb. 2008.
[29] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge, U.K.: Cambridge University Press, 1993.
[30] C. J. Stone, "Optimal global rates of convergence for nonparametric regression," *Annals Stat.*, vol. 10, no. 4, pp. 1040–1053, 1982.
[31] C. R. Rao, *Linear Statistical Inference and Its Applications*, 2nd ed. New York: Wiley, 1973.
[32] R. A. DeVore, "Nonlinear approximation," *Acta Numerica* vol. 7, pp. 51–150, 1998 [Online]. Available: http://citeseer.ist.psu.edu/429884.html
[33] Z. Hasiewicz and G. Mzyk, "Hammerstein system identification by nonparametric instrumental variables," *Int. J. Control*, vol. 82, no. 3, pp. 440–455, Mar. 2009.
[34] G. Szego, *Orthogonal Polynomials*, 3rd ed. Providence, R.I.: American Mathematical Society, 1974.
[35] E. Capobianco, "Hammerstein system representation of financial volatility processes," *Eur. Phys. J. B—Condensed Matter*, vol. 27, pp. 201–211, 2002.
[36] W. Jang and G. Kim, "Identification of loudspeaker non-linearities using the NARMAX modelling technique," *J. Audio Eng. Soc.*, vol. 42, pp. 50–59, 1994.
[37] G. B. Giannakis and E. Serpedin, "A bibliography on nonlinear system identification," *Signal Processing*, vol. 81, pp. 533–580, 2001.
[38] K. Hunt, M. Munih, N. Donaldson, and F. Barr, "Investigation of the Hammerstein hypothesis in the modeling of electrically stimulated muscle," *IEEE Trans. Biomed. Eng.*, vol. 45, no. 8, pp. 998–1009, 1998.
[39] D. Panescu, J. Webster, and R. Stratbucker, "A nonlinear electrical-thermal model of the skin," *IEEE Trans. Biomed. Eng.*, vol. 41, no. 7, pp. 672–680, Jul. 1994.
[40] E. Eskinat, S. H. Johnson, and W. L. Luyben, "Use of Hammerstein models in identification of non-linear systems," *Amer. Inst. Chem. Eng. J.*, vol. 37, pp. 255–268, 1991.
[41] L. Rutkowski, "Generalized regression neural networks in time-varying environment," *IEEE Trans. Neural Networks*, vol. 15, no. 3, pp. 576–596, May 2004.
[42] M. Niedźwiecki and P. Kaczmarek, "Identification of quasi-periodically varying systems using the combined nonparametric/parametric approach," *IEEE Trans. Signal Processing*, vol. 53, no. 12, pp. 4588–4598, Dec. 2005.