Semiparametric approach to system identification: A meeting at the crossroads

Zygmunt Hasiewicz, Grzegorz Mzyk, and Przemysław Śliwiński

The Institute of Computer Engineering, Control and Robotics
Wrocław University of Technology,
ul. Wybrzeże Wyspiańskiego 27, 50-370 Wrocław
Zygmunt.Hasiewicz,Grzegorz.Mzyk,Przemyslaw.Sliwinski@pwr.wroc.pl

Abstract. The two popular approaches to system identification problems, parametric and nonparametric, are presented and discussed. In particular, the convergence conditions and convergence rates are compared along with their numerical properties. The semiparametric approach, which possesses advantages of these both is proposed and its examplary forms are presented.

1 Introduction

The basic goal of system identification is to discover a formula governing a relation between system (phenomenon) inputs and outputs. Typically, the identification algorithms are derived from statistical framework – in order to cope with a random nature of the measurements.

When dealing with a statistical inference, one distinguishes two types of available information: theoretical and empirical. The former is given *a priori*, and is in a form of laws and resulting equations, and the latter comprises measurements collected during experiment. Depending on the range of the prior knowledge, one can distinguish two approaches: *parametric* and *nonparametric* one. The first should be applied when the knowledge is rich enough to build a parametric model of the system (or its characteristics) – in this case measurements are used to determine model parameters. The other, alternatively, is employed when such a knowledge is poor, and no reliable model can be proposed – there, the only source of information about the system can virtually be gained from measurements.

We present several *state-of-the-art* examples of these approaches, comparing and discussing their properties in a context of recovering a nonlinear characteristic in Hammerstein systems. Then we propose a new approach to system identification – a *semiparametric* one – which combines the former two and has already been investigated in statistics and applied to econometrics, [15].

2 Nonlinearity recovering in Hammerstein systems

The goal of the algorithms is to recover a nonlinearity in Hammerstein systems. We thus *implicitly* assume that the system has blocky structure and both, the types of their components (blocks) and the interconnections between them, are known, and – eventually – that they compose a cascade of static nonlinearity followed by a linear dynamic element, *viz.* the Hammerstein system (see Fig. 1a). This approach, a *block-oriented* one, is very

popular due to the following advantages: *1)* The blocks are described independently, thus the identification algorithms can separately be proposed (accordingly to *e.g.* different *a priori* knowledge available for either of them) and so analyzed. Moreover, such blocks, corresponding to mathematical formulas, facilitate the analysis and interpretation of the systems properties. *2)* Their numerical implementations are computationally encouraging. (Note, for comparison, that in a more general approach, the *black-box* one, where a system structure is assumed to be unknown and where Volterra/Wiener kernels are employed, a numerical complexity of corresponding algorithms makes them intractable already for a few kernels; [2].) *3)* Hammerstein systems have already found many practical applications; see [7], *e.g.* in biocybernetics, [23, 29, 22, 24, 27, 37, 5, 25], chemistry, [6], control, [26], and in economy, [3] and they are authorized example of systems pertaining to a broader class (so called *block-oriented*) including Uryson systems, parallel-cascade systems, multichannel S_m systems, *etc.*; see [13, 30, 19, 18].

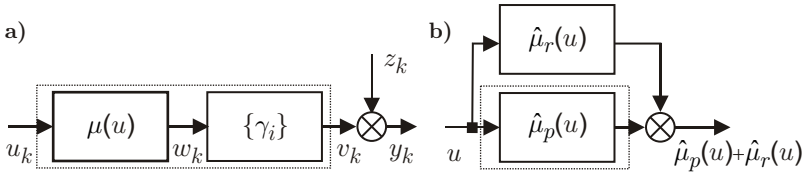

Fig. 1. a) Hammerstein system, **b)** A generic scheme of a semiparametric algorithm

A discrete Hammerstein can, in general, be described by the following input-output relation:

$$y_k = v_k + z_k \text{ where } v_k = \sum_{i=0}^{\infty} \gamma_i \cdot w_{k-i} \text{ and } w_k = \mu\left(u_k\right) \tag{1}$$

that is, by a discrete convolution of an impulse response of a dynamic element $\{\gamma_i\}$ with an output of a static nonlinearity with a characteristic $\mu\left(u\right)$ – being in turn a nonlinearly mapped input $\{u_k\}$. Before splitting into parametric and nonparametric algorithms we will present several assumptions – common for both approaches:

A1 The input signal $\{u_k\}$ is an i.i.d. process, with an arbitrary probability density function $f\left(u\right)$, *e.g.* smooth or discontinuous.

A2 The noise $\{z_k\}$ is a second order zero-mean process (white or correlated) of arbitrary distribution.

A3 Only a set of input-output measurements $\{(u_k, y_k)\}$, $k = 1, \ldots, N$, is available.

2.1 Parametric prior knowledge

The word "parametric" means that the system can be represented with the use of a finite number of unknown parameters, and the input-output equations describing the system are known a priori. In particular, it is assumed that:

A4 The form of a static nonlinearity is known, i.e., there exists a function $\mu(u, c)$ such that $\mu(u, c) = \mu(u, c^*)$, where $c^* = (c_1^*, c_2^*, ..., c_m^*)^T$ is a vector of the unknown true parameters of the nonlinearity.

A5 The linear element is of the $ARMA(s,p)$ type, i.e. it can be described by the following difference equation $v_k = b_0^* w_k + ... + b_s^* w_{k-s} + a_1^* v_{k-1} + ... + a_p^* v_{k-p}$ with an unknown vector of true parameters $\theta = (b_0^*, b_1^*, ..., b_s^*, a_1^*, a_2^*, ..., a_p^*)^T$, but with the known orders s, p.

This kind of knowledge is often met in practical applications and is strictly connected with laws of physics or provided by experts, and – usually – it lets to speed up the convergence. However, if that knowledge is uncertain, an application of the parametric algorithms can cause a systematic (approximation) error resulting in the lack of convergence of the underlying models to system characteristics (and such 'gap' between model and system cannot be reduced further even if the number of measurements grows).

2.2 Nonparametric prior knowledge

Here, the term "nonparametric" denotes the situation where the model with a *finite* and *known* number of parameters is not available prior to the experiment. The assumptions that encompass this situation are typical for nonparametric system identification task, *cf.* [8, 18, 30], and are:

A6 The input density $f(u)$ and the static nonlinearity, $\mu(u)$, are bounded and continuous with arbitrary Hölder exponents ν_f and ν_μ (*i.e.* both, the density and the nonlinearity, can have an arbitrary number of continuous derivatives; that is we only assume that $\nu_f, \nu_\mu > 0$).

A7 The linear dynamic subsystem, $\{\gamma_i\}$, is asymptotically stable.

3 A parametric approach: polynomial model with FIR dynamics

In the early papers dealing with Hammerstein system identification, the static characteristic was assumed to be *linear-in-parameters*, i.e., $\mu(u, c^*) = \sum_{i=1}^m c_i^* f_i(u)$ and the linear dynamics was assumed to have a *finite impulse response* ($p = 0$), i.e., $v_k = \sum_{i=0}^s b_i^* w_{k-i}$, where the orders m and s and the form of functions $f_1()$,...,$f_m()$ were known a priori (usually, a polynomial representation was used, i.e., $f_i(u) = u^{i-1}$). Within these setting the purpose of the identification is to estimate the parameter vector $c^* = (c_1^*, ..., c_m^*)^T$ representing the static characteristic, using the measurements $\{(u_k, y_k)\}_{k=1}^N$ of the whole Hammerstein system. Since the internal signal $w_k = \mu(u_k)$ is not accessible for a direct measurement, the systems with the parameter vectors and $\theta = (b_0^*, ..., b_s^*)^T$, c^* and $\alpha\theta$, c^*/α are, for each $\alpha \neq 0$, indistinguishable from the input-output point of view. To obtain uniqueness of the solution we assume that $||\theta||_2 = 1$, where $||.||_2$ is the Euclidean vector norm, and that first nonzero element of θ is positive. Let

$$P^* = (b_0^* c_1^*, ..., b_0^* c_m^*, ..., b_s^* c_1^*, ..., b_s^* c_m^*)^T = (p_1^*, p_2^*, ..., p_{(s+1)m}^*)^T$$

be the aggregated parameter vector of the Hammerstein system, and ϕ_k be the generalized input of the form $\phi_k = (f_1(u_k), ..., f_m(u_k), ..., f_1(u_{k-s}), ..., f_m(u_{k-s}))^T$. Since $y_k = \phi_k^T P^* + z_k$, for $k = 1, ..., N$ we obtain the following measurement equation $Y_N = \Phi_N P^* + Z_N$, where $Y_N = (y_1, ..., y_N)^T$, $\Phi_N = (\phi_1, ..., \phi_N)^T$, and

$Z_N = (z_1, ..., z_N)^T$. The estimation may be performed in two steps: *Step 1.* Compute the least squares estimate of the aggregated vector P^*

$$\widehat{P}_N^{(LS)} = (\Phi_N^T \Phi_N)^{-1} \Phi_N^T Y_N \tag{2}$$

and construct the estimate $\widehat{A}_{\theta c^*}^{(LS)}$ of the matrix $A_{\theta c^*} = \theta c^{*T}$ using plug in method. *Step 2.* Perform the Singular Value Decomposition (*SVD*) of $\widehat{A}_{\theta c^*}^{(LS)}$, i.e., $\widehat{A}_{\theta c^*}^{(LS)} = \sum_{i=1}^{\min(n,m)} \sigma_i \widehat{\mu}_i \widehat{\nu}_i^T$ and compute the estimate of c^*

$$\widehat{c}_N^{(LS)} = sgn(\widehat{\mu}_1[\kappa_{\mu_1}])\sigma_1 \widehat{\nu}_1 \tag{3}$$

where $\kappa_{\mathbf{x}} = \min\{k : x[k] \neq 0\}$.

It was proved in [1] that $\widehat{c}_N^{(LS)} \to c^*$ with probability 1 as $N \to \infty$. Nevertheless, the approach has the following disadvantages: *1)* the class of admissible nonlinearities is narrow, *2)* the convergence takes place only if the output disturbance is a white noise, and *3)* the recovering of $\mu()$ is dependent on the correctness of parametric knowledge about the linear dynamics.

4 A nonparametric approach

A nonparametric approach, originated in [9], is based on the observation for Hammerstein systems it holds that

$$E\left(y_k | \, u_k = u\right) = \mu\left(u\right) \tag{4}$$

i.e. that the system nonlinearity $\mu\left(u\right)$ is a regression function of the output y_k on the input signal u_k. To illustrate the approach, we will present two different algorithms, one based on the regression estimates employing employing Nadaraya-Watson (kernel) estimate (*cf.* [9]), and the other, using wavelet orthogonal expansion and ordered statistics (*cf.* [10]). Their asymptotic properties (convergencies and their rates are jointly presented in section 4.3).

4.1 Kernel estimate

Kernel regression estimate (studied in [9]) has the form

$$\hat{\mu}_K(u) = \sum_{k=1}^{N} K(\frac{u - u_k}{h(N)}) \cdot y_k \Big/ \sum_{k=1}^{N} K(\frac{u - u_k}{h(N)}) \tag{5}$$

where $K(u)$ is a kernel (weighting) function and $h(N)$ is a bandwidth parameter controlling the range of data used for estimating a regression function $\mu(u)$ at a given point u.

4.2 Orthogonal estimate

Assume that the input signals come from the unit interval, *i.e.* $u_k \in [0, 1]$. Thus, the nonlinearity – being bounded – is a square integrable function and can be recovered by a following wavelet estimate (*cf.* [11, 12]):

$$\hat{\mu}_\psi\left(u\right) = \sum_{n=0}^{2^M - 1} \hat{\alpha}_{Mn} \varphi_{Mn}\left(u\right) + \sum_{m=M}^{K-1} \sum_{n=0}^{2^m - 1} \hat{\beta}_{mn} \psi_{mn}\left(u\right) \tag{6}$$

where $\varphi_{Mn}(u)$ and $\psi_{mn}(u)$ are scaling functions and wavelets forming an orthogonal basis in the identification interval $[0,1]$, see [4], such that wavelets have $p = 1, 2, \ldots$ vanishing moments; $M \geq 1 + \log_2 p$ is a fixed initial scale of the estimate, and K is the scale of the estimate (which increases with a growing measurement number k; *cf.* [19, 18]). The expansion coefficients, $\hat{\alpha}_{Mn}$ and $\hat{\beta}_{mn}$, are estimated with the help of ordered measurements (*i.e.* the pairs (u_k, y_k) sorted according to increasing values of u_k):

$$\hat{\alpha}_{Mn} = \sum_{l=1}^{k} y_l \int_{u_{l-1}}^{u_l} \varphi_{Mn}(u)\, du \text{ and } \hat{\beta}_{mn} = \sum_{l=1}^{k} y_l \cdot \int_{u_{l-1}}^{u_l} \psi_{mn}(u)\, du \tag{7}$$

4.3 Convergence and convergence rates

The following theorem describes jointly the asymptotic properties of the nonparametric algorithms.

Theorem 1. *If in the estimate $\hat{\mu}_K(u)$, the bandwidth parameter $h(N)$ (and the scale K in case of $\hat{\mu}_\psi(u)$) are governed by the following rules, respectively*

$$h(N) = N^{-1/(2\nu_K+1)} \text{ and } K(N) = \lceil 1/(2\nu_\psi + 1) \cdot \log_2 N \rceil \tag{8}$$

with $\nu_K = \min\{\nu_\mu, \nu_f, p\}$ and $\nu_\psi = \min\{\nu_\mu, p\}$, then both estimates $\hat{\mu}_K(u)$ and $\hat{\mu}_\psi(u)$ converge the nonlinearity $\mu(u)$, in probability, with the rates $\mathcal{O}(N^{-\nu_K/(2\nu_K+1)})$ and $\mathcal{O}(N^{-\nu_\psi/(2\nu_\psi+1)})$, respectively.

For the proof, see [34, 9]. Note that the optimal convergence rates (amongst all nonparametric estimates for a given ν_μ; see *e.g.* [14, 35]) are achieved by the estimates. Moreover, the rates approach the optimal *parametric* rate $\mathcal{O}\left(k^{-1}\right)$, for large ν_K or ν_ψ, *i.e.* for smooth nonlinearities (and input densities), and for appropriately selected number of vanishing moments (*i.e.* such that $p \geq \min\{\nu_\mu, \nu_K\}$) of wavelet functions in (6), and kernel in (5). Note also, that the convergence rate of the first algorithm does not depend on the input density.

4.4 Generalized methods with filtered output

Let us notice that the Hammerstein system, connected in the cascade with the additional (programmable) linear filter, belongs also to the Hammerstein class, and has the same static characteristic as the original one. In some cases it is worth to apply the output linear filtering to reduce the variance of the estimate. In particular, the following averaged regression function may be estimated with only an obvious slight modification of the formulas in (7) and in (5)), *cf.* (4):

$$\mu(u)_{d_1,d_2}(u) \triangleq E\left\{\frac{y_{k+d_1}+y_{k+d_2}}{2} | u_k = u\right\}$$

instead of $\mu(u)$. For discussion concerning optimal selection of d_1, d_2 see [28].

5 Combined parametric-nonparametric approach

The idea is to join the results obtained for the nonparametric identification of nonlinear characteristics in Hammerstein systems with parametric knowledge of subsystems and standard results concerning least squares and instrumental variables, taking advantages of both. In [16], the combined parametric-nonparametric algorithm has been proposed for the identification of parameters appearing linearly in the static nonlinear element followed by the *FIR* linear dynamics. The mixed approach can be generalized (see [17]) for: 1) static characteristic $\mu()$ not linear in the parameters, 2) linear dynamics with the infinite impulse response (*IIR*), 3) correlated output noise. The method is two stage. Stage 1 of the algorithm (nonparametric) consists in nonparametric estimation of interaction inputs $w_n = \mu(u_n)$ in some N_0 points $\overline{u}_1, \overline{u}_2, ..., \overline{u}_{N_0}$, to cope with their inaccessibility for direct measurements. In Stage 2 (parametric), using the obtained estimates $\widehat{w}_k$ of w_k, we identify parameters of the nonlinear block using the pairs $\{u_n, \widehat{w}_n\}_{n=1}^{N_0}$. When identifying $\mu()$, no a priori knowledge (model) of the linear dynamics is required. Due to the general form of subsystems and correlation of the output noise, the linear least squares approach fails, and the nonlinear least squares technique is applied instead of. The algorithm has the following form.

Select $\overline{u}_1, \overline{u}_2, ..., \overline{u}_{N_0}$ such that the parameter vector c^* is identifiable.

Stage 1 (nonparametric): On the basis of N input-output measurement data $\{(u_k, y_k)\}_{k=1}^{N}$, for the selected N_0 input points $\{\overline{u}_n;\ n = 1, 2, ..., N_0\}$ estimate the corresponding interactions $\{w_n = \mu(\overline{u}_n, c^*);\ n = 1, 2, ..., N_0\}$ as

$$\widehat{w}_{n,N} = \hat{\mu}_K(\overline{u}_n) - \hat{\mu}_K(0), \tag{9}$$

where $\hat{\mu}_K(u)$ is a nonparametric estimate of the regression function $\mu(u)$.

Stage 2 (parametric): Plug in the estimates $\widehat{w}_{n,N}$ obtained in Stage 1 to the loss function

$$\widehat{Q}_{N_0,N}(c) = \sum_{n=1}^{N_0} \left[\widehat{w}_{n,N} - \mu(\overline{u}_n, c)\right]^2 \tag{10}$$

and minimize it, getting the solution $\widehat{c}_{N_0,N}$. Take the computed $\widehat{c}_{N_0,N}$ as the estimate of c^*. The following theorem holds.

Theorem 2. *Assume that the computed $\widehat{c}_{N_0,N}$ is unique and for each N the minimizers $c_{N_0,N}$, $c^* \in C$, where C is a bounded convex set in R^m. If in stage 1, for some $\tau > 0$, it holds that $\hat{\mu}_K(\overline{u}_n) = \mu(\overline{u}_n) + O(N^{-\tau})$ in probability as $N \to \infty$ for $n = 1, 2, ..., N_0$ and for $\overline{u}_n = 0$ then $\widehat{c}_{N_0,N} = c^* + O(N^{-\tau})$ in probability as $N \to \infty$.*

For the proof see [17]. The theorem says that under our assumptions both convergence and the rate of convergence of the nonparametric estimate $\hat{\mu}_K(u)$ are conveyed on the estimate $\widehat{c}_{N_0,N}$ of the parameter vector c^*, i.e., the estimate $\widehat{c}_{N_0,N}$ converges to c^* in the same sense and with the same speed as $\hat{\mu}_K(u)$ to $\mu(u)$.

6 Computational aspects

6.1 Linear and nonlinear least squares procedures

For the linear in parameters representation $Y_N = \Phi_N P^* + Z_N$ one can distinguish two kinds of numerical procedures. First of them use the normal equation $\Phi_N^T \Phi_N P = \Phi_N^T Y_N$, obtained by analytical minimization of $\|Y_N - \Phi_N P\|_2$ with respect to P. It has a unique solution if the matrix $\Phi_N^T \Phi_N$ is a full rank one. The necessary condition is that basis functions $f_1(), ..., f_m()$ must be linearly independent. The sufficient condition requires input to be reach enough (see e.g. [16]).

Second group of algorithms is based on performing orthogonal linear transformations directly on the undetermined equation $Y_N = \Phi_N P$. The most popular is singular value decomposition Φ_N. It allows for save and effective computing of pseudo-inverse Φ_N^+ of Φ_N (see [21]), providing the best possible result $\widehat{P} = \Phi_N^+ Y_N$ even if Φ_N is rank-deficient. The computational complexity of the least squares algorithms is $O\left((\dim P)^3\right) + O(N)$; [31].

As regards the parametric optimization task (unconstrained nonlinear least squares) to be solved in Stage 2 of the combined methods (see (10)) we can use to this end for instance the standard Levenberg-Marquardt method. Minimization of $\widehat{Q}_{N_0,N}(c)$ is then performed with the use of the following iterative routine:

$$\widehat{c}_{N_0,N}^{(i+1)} = \widehat{c}_{N_0,N}^{(i)} - \left[J^T(\widehat{c}_{N_0,N}^{(i)}) J(\widehat{c}_{N_0,N}^{(i)}) + \lambda_i I\right]^{-1} J^T(\widehat{c}_{N_0,N}^{(i)}) r(\widehat{c}_{N_0,N}^{(i)})$$

where $J(c)$ is the Jacobian matrix of the form $J(c)[n,j] = \frac{\partial r_n(c)}{\partial c_j}$ ($n = 1, 2, ..., N_0$; $j = 1, 2, ..., m$), where $r_n(c) = \mu(\overline{u}_n, c) - \widehat{w}_{n,N}$ and $r(c) = (r_1(c), r_2(c), ..., r_{N_0}(c))^T$. The procedure is of on-line type and the dimensions of the inverted matrix are $\dim c \times \dim c$.

6.2 Procedures for orthogonal and kernel algorithms

The wavelet algorithm demonstrated here is a remarkable example of orthogonal algorithm not only because of its good asymptotic behavior but also due to its minimal numerical complexity – the number of operations necessary to calculate the coefficients values in (7) from N measurements is of order $O(N \log N)$ and, in case of sorted measurements reduces even further, to $O(N)$ – thanks to the application of the *FWT – Fast Wavelet Transform*: *cf.* [33] and [20]. Note that such complexity can be achieved in spite of the randomness of the measurement data, and that it is independent of the estimate scale K (and *a fortiori* a number of coefficients to be determined). Furthermore, after computing the coefficients, the measurement set can be removed – the estimate values can be computed for an arbitrary u like in the parametric case (here with a complexity $O(\log N)$; *cf.* again [20]. Kernel algorithms are also of moderate complexity – the main difference between them and their orthogonal counterparts lies in the fact that all the measurements need to be now stored to allow computing estimate values for arbitrary u. Hence, for N measurements one needs $O(N)$ operations for each u (in case of compactly supported kernels and sorted data, the complexity reduces only slightly and the cost $O(N \log N)$ of the initial sorting can only be compensated if the kernel estimate is evaluated for a large number of different u's); *cf.* [20].

7 Semiparametric approach

Let us start with a short comparison of the two demonstrated approaches. The models in *parametric* algorithms have usually an intuitive or formal interpretation (being *e.g.* either a physical or chemical, or economical law), but the identification algorithm often involves algebraic operations and thus is more computationally intricate (particularly, when the model comprises a number of parameters). They are often global in their nature and therefore appropriate for prediction/extrapolation tasks. Finally, they converge faster, given the model is correct.

On the other hand-side, the estimates in *nonparametric* algorithms are of quantitative rather than qualitative nature as they usually lack of 'natural' or 'established' interpretations. They need only simple arithmetic operations when computed from measurements. They are also locally dependent on the measurements and thus predisposed to filtering problems. Their convergence is universal (yet slower), *i.e.* they recover virtually all characteristics (at the expense of greater measurements data – necessary for compensation of a poorer prior knowledge).

Even from this sketchy comparison one can immediately conclude that the parametric and nonparametric algorithms complete each other rather than compete. It is therefore reasonable, to consider a mixed approach, which 'picks up' from them advantageous properties. This idea results in a semiparametric approach which has already been considered in statistic literature, see *e.g.* a concept of additive regression [36, 14, 15] and the Fig. 1b. The first algorithm from this class (proposed in [34]) is based on the polynomial parametric model from Section 3 and the nonparametric wavelet estimate (6)-(7), and the other exploits the nonlinear-in-parameters model from Section 5 and the kernel estimate (5). Both estimate the residuum $\mu_r(x) = \mu(u) - \mu_p(u)$ where $\mu_p(u)$ is an appropriate parametric model from section 3 or 5.

Example 1. Assume that the parametric model is a polynomial of known order. Exploiting the vanishing moments property of the wavelets $\psi_{mn}(u)$, one can use directly the estimate in (6) to recover $\mu_r(u)$ (and hence a genuine $\mu(u)$) if the coefficients estimates in (7) are initiated with $\hat{\alpha}_{Mn} := \langle -\mu_p, \varphi_{Mn} \rangle$ and $\hat{\beta}_{mn} := 0$.

Example 2. In a more general case, when a parametric model is unknown (or the class of admissible models is broader than polynomials) one can simply reuse the kernel algorithm (5) with arbitrary kernel function $K(u)$, and modified measurements set $\{u_k, y'_k\}$ where $y'_k = y_k - \mu_p(u_k)$ to estimate the residuum $\mu_r(u)$.

8 Summary

The two approaches to the problem of system identification, the *parametric* and *nonparametric* ones, have already established position in the literature, they are found in applications, and their properties are well known. The presented here *semiparametric* approach is an attempt to get 'the best' from these two and reveals several potential advantages: *1)* Nonparametric parts are able to recover irregular nonlinearities better than *e.g.* polynomials due to their good localization properties, *2)* There are numerous implementations of wavelet and kernel algorithms available for different platforms

(`Matlab`, `Mathematica`, `C/C++/C#/Java`, *etc.* see *e.g.* [32]), and *3)* The existing models remain unaltered which can be important when the model is implemented in a form of software library or service, or eventually in hardware. The proposed concept can further be used for better understanding of an investigated phenomena (*e.g.* to furnish more accurate simulation results) or to get more accurate nonlinearity compensation (*e.g.* to enable application of linear optimization algorithms to the resulting system).

References

1. E.W. Bai, An optimal two-stage identification algorithm for Hammerstein-Wiener nonlinear systems, *Automatica*, vol.34, No.3, pp. 333-338, 1998.
2. S. A. Billings. Identification of non-linear systems—a survey. *Proceedings of IEE*, 127(6):272–285, 1980.
3. E. Capobianco. Hammerstein system representation of financial volatility processes. *The European Physical Journal B - Condensed Matter*, 27:201–211, 2002.
4. A. Cohen, I. Daubechies, and P. Vial. Wavelet bases on the interval and fast algorithms. *Journal of Applied and Computational Harmonic Analysis*, 1(1):54–81, 1993.
5. E. Dempsey and D. Westwick. Identification of Hammerstein models with cubic spline nonlinearities. *IEEE Transactions on Biomedical Engineering*, 51(2):237–245, 2004.
6. E. Eskinat, S. H. Johnson, and W. L. Luyben. Use of Hammerstein models in identification of non-linear systems. *American Institute of Chemical Engineers Journal*, 37:255–268, 1991.
7. G. B. Giannakis and E. Serpedin. A bibliography on nonlinear system identification. *Signal Processing*, 81:533–580, 2001.
8. W. Greblicki. Nonparametric orthogonal series identification of Hammerstein systems. *International Journal of Systems Science*, 20:2355–2367, 1989.
9. W. Greblicki and M. Pawlak. Identification of discrete Hammerstein system using kernel regression estimates. *IEEE Transactions on Automatic Control*, 31:74–77, 1986.
10. W. Greblicki and M. Pawlak. Dynamic system identification with order statistics. *IEEE Transactions on Information Theory*, 40:1474–1489, 1994.
11. W. Greblicki and P. Śliwiński. Non-linearity estimation in Hammerstein system based on ordered observations and wavelets. In *Proceedings 8th IEEE International Conference on Methods and Models in Automation and Robotics — MMAR 2002*, pages 451–456, Szczecin, September 2002. Institute of Control Engineering, Technical University of Szczecin.
12. W. Greblicki and P. Śliwiński. Non-parametric identification of non-linearity in Hammerstein system. In *3th IFAC Symposium on System Identification SYSID-2003*, pages 60–65, Rotterdam, Netherlands, August 26-28 2003. IFAC.
13. R. Haber and L. Keviczky. *Nonlinear System Parameter Identification*. Kluwer Academic Publishers, Dordrecht-Boston-London, 1999.
14. W. Härdle. *Applied Nonparametric Regression*. Cambridge University Press, Cambridge, 1990.
15. W. Härdle, M. Müller, S. Sperlich, and A. Werwatz. *Nonparametric and Semiparametric Models*. Springer-Verlag, Berlin Heidelberg, 2004.
16. Z. Hasiewicz, G. Mzyk, Combined parametric-nonparametric identification of Hammerstein systems, *IEEE Transactions on Automatic Control,* vol. 49, No. 8, pp. 1370-1375, August 2004.
17. Z. Hasiewicz, G. Mzyk, Hammerstein system identification by nonparametric instrumental variables, submitted for publication in the *International Journal of Control*.

18. 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: Regular Papers*, 52(1):427–442, 2005.
19. Z. Hasiewicz and P. Śliwiński. Identification of non-linear characteristics of a class of block-oriented non-linear systems via Daubechies wavelet-based models. *International Journal of Systems Science*, 33(14):1121–1144, 2002.
20. Z. Hasiewicz and P. Śliwiński. *Falki ortogonalne o zwartym nośniku. Zastosowanie do nieparametrycznej identyfikacji systemów*. Akademicka Oficyna Wydawnicza EXIT, Warszawa, 2005.
21. D. Kincaid, W. Cheney, *Numerical Analysis*, 3rd Edition, Published by Brooks/Cole, 2002.
22. K. Hunt, M. Munih, N. Donaldson, and F. Barr. Investigation of the Hammerstein hypothesis in the modeling of electrically stimulated muscle. *IEEE Transactions on Biomedical Engineering*, 45(8):998–1009, 1998.
23. I. W. Hunter and M. J. Korenberg. The identification of non-linear biological systems: Wiener and Hammerstein cascade models. *Biological Cybernetics*, 55(2–3):135–144, 1986.
24. S. N. Jyothi and M. Chidambaram. Identification of Hammerstein model for bioreactors with input multiplicities. *Bioprocess Engineering*, 23(4):323–326, 2000.
25. S. Kukreja, R. Kearney, and H. Galiana. A least-squares parameter estimation algorithm for switched Hammerstein systems with applications to the VOR. *IEEE Transactions on Biomedical Engineering*, 52(3):431–444, 2005.
26. S.-K. Lin. Identification of a class of nonlinear deterministic systems with application to manipulators. *IEEE Transactions on Automatic Control*, 39(9):1886–1893, 1994.
27. M. Lortie and R. E. Kearney. Identification of time-varying Hammerstein systems from ensemble data. *Annals of Biomedical Engineering*, 29(2):619–635, 2001.
28. G. Mzyk, "Generalised kernel regression estimate for the identification of Hammerstein systems", *International Journal of Applied Mathematics and Computer Science*, vol. 17, No. 2, pp. 189-197, 2007.
29. D. Panescu, J. Webster, and R. Stratbucker. A nonlinear electrical-thermal model of the skin. *IEEE Transactions on Biomedical Engineering*, 41(7):672–680, 1994.
30. M. Pawlak and Z. Hasiewicz. Nonlinear system identification by the Haar multiresolution analysis. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 45(9):945–961, 1998.
31. A. Pazman. *Nonlinear statistical models*. Kluwer Academic Publishers, Dordrecht, 1993.
32. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, Cambridge, 1993.
33. P. Śliwiński and Z. Hasiewicz. Computational algorithms for wavelet identification of nonlinearities in Hammerstein systems with random inputs. *IEEE Transactions on Signal Processing*, 56(2), 2008.
34. P. Śliwiński, J. Rozenblit, M. W. Marcellin, and R. Klempous. Wavelet amendment of polynomial models in nonlinear system identification. *IEEE Transactions on Automatic Control*, 2008. Submitted for review.
35. C. J. Stone. Optimal global rates of convergence for nonparametric regression. *Annals of Statistics*, 10(4):1040–1053, 1982.
36. C. J. Stone. Additive regression and other nonparametric models. *Annals of Statistics*, 13(2):689–705, 1985.
37. D. T. Westwick and R. E. Kearney. Separable least squares identification of nonlinear Hammerstein models: Application to stretch reflex dynamics. *Annals of Biomedical Engineering*, 29(8):707–718, 2001.