Nonlinear system identification under various prior knowledge ⋆

Zygmunt Hasiewicz * **Przemysław Śliwiński** * **Grzegorz Mzyk** *

* *The Institute of Computer Engineering, Control and Robotics, Wrocław University of Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland*

Abstract: In the note the class of block-oriented dynamic nonlinear systems is considered, in particular, Hammerstein and Wiener systems are investigated. Several algorithms for nonlinear system identification are presented. The algorithms exploit various degrees of prior knowledge - from parametric - to nonparametric. Eventually, a semiparametric algorithm, which shares advantages of both approaches is announced.

Keywords: Nonlinear system identification; identification algorithms; parameter identification; non-parametric regression estimation; orthogonal expansions; wavelets; kernel algorithms; semi-parametric approach

1. INTRODUCTION

System identification deals with a problem of establishing a formulae governing a system or phenomenon. Identification algorithms exploit two types of information:

- theoretical, given *a priori* in a form of laws and resulting equations, and
- empirical one, *i.e.* measurements collected during experiments.

Due to an obvious variety of nonlinear dynamic systems there is no *one-fits-all* approach to the problem of their identification and the selection of the proper algorithms is determined by the available prior knowledge about systems and signals. Note also that there is a kind of *Catch 22* here – the prior knowledge is necessary for the experiment to be properly designed and then for the results to be correctly interpreted but – simultaneously – when new problems are explored it is quite obvious that such a knowledge is not available. Therefore, the methods and algorithms capable to work with a small *a priori* knowledge and able to take into account an additional knowledge in a course of experiment are of special interest.

Remark 1. One can point out the determining by Gauss the orbit of the dwarf planet, *Ceres*, accomplished over 200 years age – in 1801 – as the first successful system identification experiment. He used Kepler's laws – as *a priori* knowledge – and a collection of observations gathered by Piazzi – as the measurements; *cf.* Abdulle and Wanner [2002]. Moreover, the mathematical tools proposed then by Gauss constituted the cornerstones of the *modern* system identification: *probability* and *statistics*, accompanied by *linear algebra* (to which one can only add *functional analysis*, supporting recent nonparametric developments).

In the paper we present a class of algorithms working under a common assumption that a system structure, *vis.* its components (blocks) types and interconnections between them are known. The components are simple subsystems being either a static (memoryless) nonlinearity or a linear dynamic element. The approach is called *block-oriented* and the goal of the identification algorithms is to establish the characteristics of these blocks. It can be advocated by the following arguments:

- Since the blocks can be described independently, the algorithms can be tailored appropriately to a different *a priori* knowledge (ranging from a reach – *parametric* to a poor – *nonparametric*) available for either of them separately.
- The resulting algorithms convergence can formally be shown for ample classes of admissible characteristics (*e.g.* for nonlinearities being (dis-)continuous, (non-)invertible, (piecewise-)polynomial and for dynamics with finite or infinite impulse responses).
- It eventually leads to computationally tractable algorithms (as opposed to the algorithms derived within a *black-box* approach, where the system structure is assumed to be unknown, and a where numerically complex, generic Volterra/Wiener kernels methods need to be employed; Billings [1980]).

Within the class of *block-oriented* systems, the two have attracted a significant interest in the literature: the *Hammerstein* system, and the *Wiener* one. Both are quite (yet somehow deceptively) simple as they are merely cascades of either a static nonlinearity followed by a linear dynamics (the former, Fig. 1) or the dynamics followed by the nonlinearity (the latter, Fig. 2). These systems are met in many practical applications to date and they are representative for the broader class of block oriented systems including Uryson systems, parallel-cascade systems, multichannel S_m systems, *etc.*; see Giannakis and Serpedin [2001], *e.g.* in biocybernetics, Hunter and Korenberg [1986], Panescu et al. [1994], Hunt et al. [1998], Jyothi and Chidambaram [2000], Lortie and Kearney [2001], Westwick and Kearney [2001], Dempsey and Westwick [2004], Kukreja et al.

⋆ The paper is supported by the MoSaHE grant N N514 3160 33

[2005], chemistry, Eskinat et al. [1991], control, Lin [1994], Zi-Qiang [1993], Zhu and Seborg [1994], and in economy, Capobianco [2002].

1.1 The systems descriptions

We consider the discrete-time systems (for algorithms for nonparametric continuous-time systems, see *e.g.*, Greblicki [1998, 2000]; for parametric and nonparametric approaches embedded in frequency-domain analysis, see *e.g.* Bai [2003], Pintelon et al. [1994], Schoukens et al. [1998]). In this vein, Hammerstein and Wiener systems can be, in general, described by the following input-output relations (note how the blocks correspond to the basic mathematical notations of a function (nonlinear block, μ) and of a convolution (dynamics, γ_i).

$$y_k = \sum_{i=0}^{\infty} \gamma_i \cdot m\left(u_{k-i}\right) + z_k \tag{1}$$

$$y_k = m\left(\sum_{i=0}^{\infty} \gamma_i \cdot u_{k-i} + z_k\right) \tag{2}$$

We use impulse response description, however, one would prefer difference or state space equations as well. An additive signal, z_k, is a zero-mean noise disturbing the system.

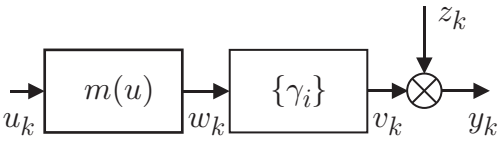

Fig. 1. A Hammerstein system

Remark 2. Neither of interconnecting signals w_k and v_k are available for measurements. Thus, in general, the system characteristics can be recovered up to some, system dependent, multiplicative and additive constants. It will be further shown in more details, however it is noted here to emphasize that this inability is a consequence of the systems assembled structures and hence a result of a lack of measurements of interconnecting signals rather than a shortcoming of any of the proposed algorithms.

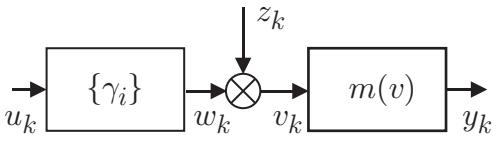

Fig. 2. A Wiener system

2. MATHEMATICAL TOOLS

The main tool the identification algorithms are based upon is a *regression* function, *i.e.* a conditional expectation of the system output given the input.

Hammerstein system. We can rewrite (1) into the equivalent form (see Fig. 1)

$$y_k = \mu\left(u_k\right) + \xi_k + z_k \tag{3}$$

in which the past observations $\{x_{k-i}\}$ induce an *additive* stationary 'system noise'

$$\xi_k = \sum_{i=1}^{\infty} \gamma_i \left[m\left(u_{k-i}\right) - Em\left(u_1\right)\right]$$

correlated because of system dynamics, and disturbing together with the external one, z_k, the output of a *system nonlinearity*

$$\mu\left(u\right) = \gamma_0 m\left(u\right) + \zeta$$

where $\zeta = E\left\{m\left(x_1\right)\sum_{i=1}^{\infty}\gamma_i\right\}$ is a system dependent constant, *cf.* Remark 2. This, since $Ez_k = E\xi_k = 0$, leads eventually to the observation that $Ey_k = E\mu\left(u_k\right)$ and hence that $\mu\left(u\right) = E\left(y_k|u_k = u\right)$, *i.e.* that $\mu\left(u\right)$ is in fact a regression function of y_k on u_k; *cf.* Greblicki and Pawlak [1986]. Thus, to recover the nonlinearity, we merely need to estimate the regression function $\mu\left(u\right)$. As for the linear subsystem, we note that

$$E\left\{y_i u_0\right\} = \gamma_i E\left\{m\left(u_1\right) u_1\right\} \tag{4}$$

and use correlation estimate to recover impulse response coefficients γ_i.

Wiener system. In this case the identification problem is much more intricate. First of all, for Gaussian input u_k and noise z_k we have that (see Fig. 2)

$$E\left(u_k|w_{k+i} = w\right) = \alpha_i w$$

where $\alpha_i = \gamma_i\alpha$ with $\alpha = 1/\left(\sum_{i=0}^{\infty}\gamma_i^2 + \sigma_z^2/\sigma_u^2\right)$, being a system-dependent constant (a usual notation of variance of u_k and z_k by σ_u^2 and σ_z^2, respectively, is applied). Then, for any invertible nonlinearity $m\left(y\right)$ we have

$$E\left(u_k|y_{k+i} = y\right) = \alpha_i m^{-1}\left(y\right) = \mu\left(y\right) \tag{5}$$

that is, a scaled inverse of the nonlinear characteristic is a regression function of u_k on y_k (*i.e.* a conditional expectation holds in with input and output measurements being replaced each other). Hence, estimating the regression is equivalent to estimate the inverse of the nonlinearity (up to the constant α_i, *cf.* Remark 2). For linear system we have

$$E\left\{u_0 y_i\right\} = \beta\gamma_i \tag{6}$$

where $\beta = \alpha E\left\{v_1 m\left(v_1\right)\right\}$ is an another system-dependent constant, *cf.* Greblicki and Pawlak [1992], Greblicki [1994, 1997].

3. IDENTIFICATION ALGORITHMS

Depending on a scope of the prior information at hand, we will present algorithms recovering nonlinearity in Hammerstein and Wiener systems employing either parametric or nonparametric estimates of regression function. In particular, in a former case, a least squares-based estimates will be used, and when only nonparametric knowledge is available, various nonparametric algorithms will exploit a kernel-based Nadaraya-Watson estimate, or an orthogonal series one (with a wavelet series in particular). Also mixed parametric-nonparametric algorithms are presented.

3.1 Least squares identification of Hammerstein system

Prior knowledge. In the classical parameter approach (Narendra and Gallman [1966], Chang and Luus [1971], Billings and Fakhouri [1982], Ljung [1987], Stoica and Söderström [1989], Bai [1998]) a rich a priori knowledge of the system is required, and specific structure on the system description is imposed. Most often the nonlinear system is assumed to be linear in (the finite number of) parameters, i.e.,

$$y_k = \sum_{i=0}^{n} \gamma_i \mu(u_{k-i}) + z_k, \qquad \mu(u) = \sum_{i=1}^{m} c_i f_i(u) \tag{7}$$

where the orders n and m are known and the form of functions $f_1()$,...,$f_m()$ are given a priori (*e.g.* polynomial representation is used, *i.e.*, $f_i(u) = u^{i-1}$), and z_k is immeasurable zero-mean random noise. The purpose is to estimate the parameter vectors

$$c = (c_1, ..., c_m)^T \text{ and } \Gamma = (\gamma_0, ..., \gamma_n)^T \tag{8}$$

representing the static characteristic and the linear dynamics, using the measurements $\{(u_k, y_k)\}_{k=1}^{N}$ of the whole system. Since the internal signal $w_k = \mu(u_k)$ is not accessible for a direct measurement, the systems with the parameter vectors Γ, c and $a\Gamma$, c/a are, for each $a \neq 0$, indistinguishable from the input-output point of view.

Remark 3. To obtain a uniqueness of the solution we assume that $||\Gamma||_2 = 1$, where $||.||_2$ is the euclidean vector norm, and that first nonzero element of Γ is positive.

Let

$$\begin{aligned} \theta &= (\gamma_0 c_1, ..., \gamma_0 c_m, ..., \gamma_n c_1, ..., \gamma_n c_m)^T = \\ &= (\theta_1, \theta_2, ..., \theta_{(n+1)m})^T \end{aligned} \tag{9}$$

be the aggregated parameter vector of the Hammerstein system, and ϕ_k be the generalized input of the form

$$\phi_k = (f_1(u_k), ..., f_m(u_k), ..., f_1(u_{k-n}), ..., f_m(u_{k-n}))^T \tag{10}$$

Since $y_k = \phi_k^T \theta + z_k$, for $k = 1, ..., N$ we obtain the following measurement equation

$$Y_N = \Phi_N \theta + Z_N \tag{11}$$

where $Y_N = (y_1, ..., y_N)^T$, $\Phi_N = (\phi_1, ..., \phi_N)^T$, and $Z_N = (z_1, ..., z_N)^T$.

The identification algorithm The estimation may be performed as follows.

Step 1. Compute the least squares estimate

$$\widehat{\theta}_N^{(LS)} = (\Phi_N^T \Phi_N)^{-1} \Phi_N^T Y_N \tag{12}$$

of θ, and construct the estimate $\widehat{\Theta}_{\Gamma c}^{(LS)}$ of the matrix $\Theta_{\Gamma c} = \Gamma c^T$ using plug in method.

Step 2. Perform the Singular Value Decomposition (*SVD*) of $\widehat{\Theta}_{\Gamma c}^{(LS)}$

$$\widehat{\Theta}_{\Gamma c}^{(LS)} = \sum_{i=1}^{\min(n,m)} \sigma_i \widehat{\mu}_i \widehat{\nu}_i^T \tag{13}$$

and compute the estimates of Γ and c

$$\begin{aligned} \widehat{\Gamma}_N^{(LS)} &= sgn(\widehat{\mu}_1[\kappa_{\mu_1}])\widehat{\mu}_1 \\ \widehat{c}_N^{(LS)} &= sgn(\widehat{\mu}_1[\kappa_{\mu_1}])\sigma_1 \widehat{\nu}_1 \end{aligned}$$

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

It was proved in Bai [1998] that

$$\widehat{\Gamma}_N^{(LS)} \to \Gamma \qquad \text{and} \qquad \widehat{c}_N^{(LS)} \to c \tag{14}$$

with probability 1 as $N \to \infty$.

Similar approach can be applied also for the Hammerstein systems with the *IIR* linear subsystems (described by the *ARMA* equation). However for the correlated $\{z_k\}$ the least squares estimate is biased then. One of the techniques which allow to avoid this problem is application of instrumental variables (see e.g. Wong and Polak [1967], Söderström and Stoica [1982]).

3.2 Nonparametric identification of Hammerstein system

Prior knowledge. The following assumptions (typical for nonparametric system identification tasks; *cf.* Greblicki [1989], Hasiewicz et al. [2005], Śliwiński and Hasiewicz [2005], Pawlak and Hasiewicz [1998]) hold: *1)* the input signal, $\{u_k\}$, and the external additive noise, $\{z_k\}$, are zero-mean random processes with finite variances; they are mutually independent and $\{u_k\}$ is an i.i.d. process with a density function $f(u)$, *2)* the density, $f(u)$, and the static nonlinearity, $m(u)$, are bounded and continuous with some Hölder smoothness 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, $\{\gamma_i\}$, is unknown, *4)* only a set of input-output measurements $\{(u_k, y_k)\}$, $k = 1, \ldots, N$, is available.

Kernel algorithm. Assume that input signal is arbitrarily distributed (and in particular, it has no density). The algorithm we employ in such case is based on a kernel function $K(u)$ (satisfying only some additional technical conditions, *cf.* Greblicki and Pawlak [1989]):

$$\hat{\mu}_N(u) = \frac{\sum_{k=1}^{N} y_k K\left(\frac{u-u_k}{h(N)}\right)}{\sum_{k=1}^{N} K\left(\frac{u-u_k}{h(N)}\right)} \tag{15}$$

The convergence of the estimate is governed by the following theorem (*cf.* Greblicki and Pawlak [1989]):

Theorem 1. Let

$$N \to \infty \text{ and } Nh(N) \to 0$$

then

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

for almost all $\tau(u)$ where τ is a probability measure of the input u_k and where $\gamma = \min\{\nu_m, \nu_f, p\}$ with p being a number of vanishing moments of the kernel function $K(u)$.

Orthogonal (wavelet) algorithm. When the input signal u_k has a density, one can apply wavelet based algorithm, *cf.* Hasiewicz et al. [2005], Śliwiński and Hasiewicz [2007] and (3):

$$\hat{\mu}_K(u) = \frac{\hat{g}_K(u)}{\hat{f}_K(u)} \tag{16}$$

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.* [Hasiewicz et al., 2005, Hasiewicz, 2001, Remark 5.1], of a product $g(u) = \mu(u) \cdot f(u)$, and of the input signal density $f(u)$. The estimates presented below in a vector-like form for conciseness and to emphasize similarities of the computations needed by $\hat{g}_K(u)$ and $\hat{f}_K(u)$:

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

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

$$\begin{bmatrix} \hat{\alpha}_{Mn} \\ \hat{a}_{Mn} \end{bmatrix} = \frac{1}{N} \sum_{k=1}^{N} \varphi_{Mn}(u_k) \cdot \begin{bmatrix} y_k \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} \hat{\beta}_{mn} \\ \hat{b}_{mn} \end{bmatrix} = \frac{1}{N} \sum_{k=1}^{N} \psi_{mn}(u_k) \cdot \begin{bmatrix} y_k \\ 1 \end{bmatrix}$$

where $[s_1, s_2]$ and $[t_1, t_2]$ are supports of the *father* and *mother* wavelets, φ and ψ, respectively. For $m = M, M+1, \ldots$, some fixed M, and $n = \ldots, -1, 0, 1, \ldots$, $\{\varphi_{Mn}\}$ and $\{\psi_{mn}\}$ constitute an orthogonal basis of $L^2(R)$ space.

The properties of the reference algorithm is established by the following lemma, being a 'decomposed' version of the Theorem 2 in Hasiewicz et al. [2005] 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 Hasiewicz et al. [2005]).

Theorem 2. Let the wavelet family used in the estimate $\hat{\mu}_K(u)$ in (16) have p vanishing moments. Selecting the estimate scale K according to the rule

$$K = \left\lfloor \tfrac{1}{2\gamma+1} \log_2 N \right\rfloor \text{ where } \gamma = \min\{\nu_m, \nu_f, p\}$$

makes the algorithm $\hat{\mu}_K(u)$ converge to $\mu(u)$ as $N \to \infty$ with the rate

$$\hat{\mu}_K(u) = \mu(u) + \mathcal{O}(N^{-\frac{\gamma}{2\gamma+1}}) \text{ in probability,} \tag{17}$$

for arbitrary u, for which $f(u) > 0$.

Note that the convergence given in (17) is not affected by a structure of a dynamic subsystem $\{\gamma_i\}$; *cf.* discussions in Pawlak and Hasiewicz [1998], Hasiewicz et al. [2005]. Moreover, provided that, for a given number of vanishing moments of the wavelet family, it holds 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 Stone [1980]).

Generalized nonparametric estimates. Recall that the fundamental meaning for the identification routine of Hammerstein system has the following dependence between the regression functions $R_c(u)$ and the genuine system characteristic $m()$

$$\begin{aligned} R_c(u) &= E\{y_k | u_{k-c} = u\} \\ &= E\{\gamma_c m(u_{k-c}) + \sum_{i \neq c} \gamma_i m(u_{k-i}) + z_k | u_{k-c} = u\} \\ &= \gamma_c m(u) + \delta_c \end{aligned} \tag{18}$$

where $\delta_c = Em(u) \sum_{i \neq c}^{\infty} \gamma_i$, and c is any time-lag between input and output. Due to (18), the characteristic $\mu()$ may be estimated only up to some scaling and shifting constants γ_c and δ_c, provided that we can estimate $R_c(u)$. This feature is however independent of the identification method and it is a simple consequence of inaccessibility of the interior signal $\{w_k\}$ for a direct measurement. In the standard nonparametric methods, the Hammerstein system is treated in fact as a nonlinear static element corrupted by a correlated noise. One can namely specify three components of the output

$$y_k = \gamma_c m(u_{k-c}) + \sum_{i \neq c} \gamma_i m(u_{k-i}) + z_k \tag{19}$$

In such a description only the cth term of the sum in (7) is privileged, which means that the most part of the signal y_k is in a sense ignored, although the "system noise"

$$\xi_k \triangleq \sum_{i \neq c} \gamma_i m(u_{k-i}) \tag{20}$$

also depends on the identified function $m()$. To avoid this problem the generalized approach has been proposed in Mzyk [2007]. The following combined regression function

$$\begin{aligned} R_{c_1,c_2}(u) &\triangleq \frac{R_{c_1}(u) + R_{c_2}(u)}{2} \\ &= E\left\{ \frac{y_{k+c_1} + y_{k+c_2}}{2} | u_k = u \right\} \end{aligned}$$

is estimated, which allows to reduce the error by a proper selection of c_1 and c_2.

Identification of the linear part The relation in (4) suggests the following simple estimates of the impulse response coefficients

$$\hat{\gamma}_i = \frac{1}{N-i} \sum_{k=1}^{N-i} y_{k+i} u_k \tag{21}$$

for which holds the mean square convergence holds in a coefficient-wise manner, that is, (*cf.* Greblicki and Pawlak [1989]):

Theorem 3. For any fixed i

$$\lim_{N \to \infty} E(\gamma_i - \hat{\gamma}_i)^2 = 0$$

The same algorithm can be used to estimate coefficients of the impulse response in Wiener systems, as well, *cf.* Greblicki [1997].

3.3 Least squares identification of Wiener system

Since in the Wiener system the nonlinear block precedes the linear dynamics, the identification task is much more difficult. Till now, the sufficient identifiability conditions have been formulated and the convergence of the proposed estimates has been proved only fore some special cases. Many methods require the non-linearity to be known, invertible, differentiable or require special input sequences (see e.g. Billings and Fakhouri [1977], Hasiewicz [1987], Haber and Keviczky [1999]). For example the Wiener system with the polynomial static characteristic and the *FIR* linear dynamics

$$y_k = \sum_{i=0}^{p} c_0 w_k^i + z_k, \qquad w_k = \sum_{i=0}^{m} \gamma_i u_{k-i} \tag{22}$$

can be described similarly to (11)

$$Y_N = \Phi_N \theta + Z_N \tag{23}$$

but now, the meaning and the structure of the matrix $\Phi_N = (\phi_1, ..., \phi_N)^T$ and the vector θ are more sophisticated, i.e.,

$$\phi_k = \left[\frac{|\alpha|!}{\alpha!} \overline{u}_k^{\alpha} \right]_{|\alpha| \leqslant p} \in \mathcal{R}^d \tag{24}$$
$$\theta = \left[c_{|\alpha|} \Gamma^{\alpha} \right]_{|\alpha| \leqslant p} \in \mathcal{R}^d$$

where

$$\alpha = (\alpha_1, \alpha_2, ..., \alpha_{m+1})^T \in \mathcal{N}_0^{m+1}$$

is the multi-index of order $m+1$ (see Lacy and Bernstein [2003]), $|\alpha| = \sum_{i=0}^{m+1} \alpha_i$, $\alpha! = \prod_{i=1}^{m+1} \alpha_i$, $\overline{u}_k = (u_k, u_{k-1}, ..., u_{k-m})^T$, $\overline{u}_k^{\alpha} = \prod_{i=1}^{m+1} u_{k-i-1}^{\alpha_i}$, $\Gamma^{\alpha} = \prod_{i=1}^{m+1} \gamma_{i-1}^{\alpha_i}$, $d = \sum_{i=0}^{p} \frac{(m+i)!}{m!i!}$, and $[f(\alpha)]_{|\alpha| \leqslant p}$ denotes the column vector whose components are evaluated at every multi-index α such that $|\alpha| \leqslant p$ under some established ordering. Step 1 of the identification procedure is the same as for Hammerstein system (see (12)), Step 2 requires application of multi-dimensional *SVD*.

3.4 Nonparametric identification of Wiener system

Assumptions here resemble, to some extent, the assumptions posed for nonparametric Hammerstein system algorithm. They are, however, much more stringent. Namely (*cf.* Greblicki [1992, 1994, 1997]) hold: *1)* the input signal, $\{u_k\}$, and the additive noise, $\{z_k\}$, are zero-mean Gaussian processes, *2)* the static nonlinearity, $m(u)$, is invertible and has $P = 1, 2, \ldots$ continuous derivatives, *3)* the linear dynamic subsystem is asymptotically stable and its impulse response, $\{\gamma_i\}$, is unknown, *4)* only a set of input-output measurements $\{(u_k, y_k)\}$, $k = 1, \ldots, N$, is available.

The kernel algorithms is of the form

$$\hat{\mu}_N(y) = \frac{\sum_{k=1}^{N} u_k K\left(\frac{y-y_k}{h(N)}\right)}{\sum_{k=1}^{N} K\left(\frac{y-y_k}{h(N)}\right)}$$

i.e. it is a version of the algorithm in (15) with switched input and outputs..

The following theorem describes the behavior of the algorithm Greblicki and Pawlak [1992]:

Theorem 4. Let the kernel K satisfies the following restrictions

$$\sup |K(y)| < \infty \text{ and } \int |K(y)| \, dy < \infty$$

and $yK(y) \to 0$ as $y \to \infty$. If, moreover, $K(y)$ has $P+1$ vanishing moments, then for $N \to \infty$ and $Nh(N) \to 0$ it holds that

$$\hat{\mu}_N(y) = \mu(y) + \mathcal{O}\left(N^{-\frac{P}{2P+1}}\right).$$

We would like to point at the end of the Wiener system identification algorithms presentation that the interesting idea of nonparametric identification of Wiener systems with non-invertible characteristics and IIR filters has been proposed in Mzyk [2007b], where the static characteristics is estimated by data pre-selection and local averaging; see also Pawlak et al. [2007] for algorithms working with FIR filters.

3.5 Parameter identification of Hammerstein systems with the help of nonparametric regression methods

Prior knowledge In this section parameter estimation is supported with the nonparametric regression. We admit the *IIR* linear dynamics, correlated output noise and the nonlinear characteristic which is not linear in the parameters.

Assumption 1. The form of a static nonlinearity is known up to the parameters, i.e. we are given the function $\mu(u, c)$ such that $\mu(u, c^*) = \mu(u)$, where $c^* = (c_1^*, c_2^*, ..., c_m^*)^T$ is a vector of the unknown true parameters of the nonlinearity. The function $\mu(u, c)$ is by assumption differentiable with respect to c, and the gradient $\bigtriangledown_c \mu(u, c)$ is bounded in some convex neighborhood $\mathcal{O}(c^*)$ of c^*:

$$\| \bigtriangledown_c \mu(u, c) \| \leqslant G_{\max} < \infty, \qquad c \in \mathcal{O}(c^*)$$

Assumption 2. The linear element is of the $ARMA(s, p)$ type, i.e. it can be described by the following difference equation

$$v_k = \alpha_0 w_k + ... + \alpha_s w_{k-s} + \beta_1 v_{k-1} + ... + \beta_p v_{k-p} \tag{25}$$

$p \geqslant s$, with unknown parameters $\alpha_0, \alpha_1, ..., \alpha_s$ and $\beta_1, \beta_2, ..., \beta_p$, or equivalently as

$$\beta(q^{-1}) v_k = \alpha(q^{-1}) w_k$$

where

$$\alpha(q^{-1}) = \alpha_0 + \alpha_1 q^{-1} + ... + \alpha_s q^{-s}$$
$$\beta(q^{-1}) = 1 - \beta_1 q^{-1} - ... - \beta_p q^{-p}$$

and q^{-1} is a backward shift operator

As it will be seen, when identifying dynamic subsystem Assumption 1 may be omitted, and conversely Assumption 2 may be omitted during identification process of the static part. The aim is to discover the true parameters of subsystems, respectively $c^* = (c_1^*, c_2^*, ..., c_m^*)^T$ and $\theta =$

$(\alpha_0, \alpha_1, ..., \alpha_s, \beta_1, \beta_2, ..., \beta_p)^T$, using a set of input-output data $\{(u_k, y_k)\}$ collected from the whole system in an identification experiment.

Estimation of the nonlinearity parameters First, exploiting a nonparametric regression estimation technique, the immeasurable inner signal $\{w_k\}$ is estimated from the measurement data (u_k, y_k). Then, the least squares/ instrumental variables method is used to the independent estimation of the two subsystems parameters using, respectively, the pairs $(u_k, \widehat{w}_k)$ and $(\widehat{w}_k, y_k)$ where $\{\widehat{w}_k\}$ is the estimate of the interaction sequence obtained by a nonparametric method.

Stage 1 (nonparametric): On the basis of M input-output measurement data $\{(u_k, y_k)\}_{k=1}^{M}$, 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

$$\widehat{w}_{n,M} = \widehat{R}_M(\overline{u}_n) - \widehat{R}_M(0), \tag{26}$$

where $\widehat{R}_M(u)$ is a nonparametric estimate of the regression function $R(u) = E[y_k | u_k = u]$.

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

$$\widehat{Q}_{N_0,M}(c) = \sum_{n=1}^{N_0} \left[\widehat{w}_{n,M} - \mu(\overline{u}_n, c)\right]^2 \tag{27}$$

and minimize them, getting the solution $\widehat{c}_{N_0,M}$. Take the computed $\widehat{c}_{N_0,M}$ as the estimate of c^*.

Theorem 5. (Hasiewicz and Mzyk [2004b]) Assume that the computed $\widehat{c}_{N_0,M}$ is unique and for each M the minimizers $c_{N_0,M}$, $c^* \in C$, where C is a bounded convex set in R^m. If in stage 1 it holds that

$$\widehat{R}_M(\overline{u}_n) = R(\overline{u}_n) + O(M^{-\tau}) \text{ in probability} \tag{28}$$

as $M \to \infty$ for $n = 1, 2, ..., N_0$ and for $\overline{u}_n = 0$ then

$$\widehat{c}_{N_0,M} = c^* + O(M^{-\tau}) \text{ in probability} \tag{29}$$

as $M \to \infty$.

3.6 Identification of ARMA *dynamics by nonparametric instrumental variables*

Since $v_k = y_k - z_k$ thus

$$y_k = \vartheta_k^T \theta + \overline{z}_k \tag{30}$$

where $\theta = (\alpha_0, \alpha_1, ..., \alpha_s, \beta_1, \beta_2, ..., \beta_p)^T$ is a vector of unknown true parameters of the linear dynamics ($p \geqslant s$), $\vartheta_k = (w_k, w_{k-1}, ..., w_{k-s}, y_{k-1}, y_{k-2}, ..., y_{k-p})^T$ is a generalized input vector and

$$\overline{z}_k = z_k - \beta_1 z_{k-1} - ... - \beta_p z_{k-p}$$

is a proper, zero-mean and stationary resultant disturbance. For a set of N input-output data $\{(\vartheta_k, y_k)\}$ we can write concisely

$$Y_N = \Theta_N \theta + Z_N$$

where $Y_N = (y_1, y_2, ..., y_N)^T$, $\Theta_N = (\vartheta_1, \vartheta_2, ..., \vartheta_N)^T$ and $Z_N = (\overline{z}_1, \overline{z}_2, ..., \overline{z}_N)^T$.

Since the matrix Θ_N contains among others the regressors $y_{k-1}, y_{k-2}, ..., y_{k-p}$ and the noise $\{\overline{z}_k\}$ is not white, the least squares estimate of θ, of the form

$$\widehat{\theta}_N^{(LS)} = (\Theta_N^T \Theta_N)^{-1} \Theta_N^T Y_N \tag{31}$$

is obviously biased. As is well known, we can overcome this weakness by using instrumental variables approach, yielding the estimate (see Stoica and Söderström [1989])

$$\widehat{\theta}_N^{(IV)} = (\Psi_N^T \Theta_N)^{-1} \Psi_N^T Y_N \tag{32}$$

where Ψ_N is a matrix of properly selected instruments $\Psi_N = (\psi_1, \psi_2, ..., \psi_N)^T$, $\psi_k = (\psi_{k,1}, \psi_{k,2}, ..., \psi_{k,s+p+1})^T$ such that the following two properties hold

(a) $Plim_{N\to\infty} \left(\frac{1}{N} \Psi_N^T \Theta_N\right)$ exists and is not singular

(b) $Plim_{N\to\infty} \left(\frac{1}{N} \Psi_N^T Z_N\right) = 0$

Under such conditions the estimation error

$$\Delta_N^{(IV)} = \widehat{\theta}_N^{(IV)} - \theta = \left(\frac{1}{N} \mathbf{\Psi}_N^T \Theta_N\right)^{-1} \left(\frac{1}{N} \mathbf{\Psi}_N^T Z_N\right) \tag{33}$$

tends to zero (in probability) as $N \to \infty$, i.e. $\widehat{\theta}_N^{(IV)} \to \theta$ in probability as N grows large.

The conditions *(a)* and *(b)* require in fact the elements of Ψ_N be correlated with inputs and simultaneously not correlated with the noise $\{\overline{z}_k\}$. The simplest Ψ_N-generation techniques exploit directly former inputs of linear dynamics (see e.g. Söderström and Stoica [1982]), i.e. we take

$$\psi_{k,i} = w_{k-i+1} \tag{34}$$

which yields

$$\psi_k = (w_k, ..., w_{k-s}, w_{k-s-1}, ..., w_{k-s-p})^T \tag{35}$$

or, more generally, are based on a linear filtering of the input process

$$\psi_{k,i} = F_i(q^{-1}) w_k \tag{36}$$

where $F_i(q^{-1})$ is a polynomial in q^{-1} (a backward shift operator).

However, as it was already pointed out, in the Hammerstein system the inputs $w_k, ..., w_{k-s-p}$ of linear dynamics, $w_k = \mu(u_k)$, are not accessible for measurements thus precluding the direct use of such an instrumental variables estimate. According to our leading idea to overcome this drawback we propose instead implementation (in stage 2) of the following plug in estimate

$$\widehat{\theta}_{N,M}^{(IV)} = (\widehat{\Psi}_{N,M}^T \widehat{\Theta}_{N,M})^{-1} \widehat{\Psi}_{N,M}^T Y_N \tag{37}$$

where

$$\widehat{\Theta}_{N,M} = (\widehat{\vartheta}_{1,M}, ..., \widehat{\vartheta}_{N,M})^T \tag{38}$$

$$\widehat{\vartheta}_{k,M} = (\widehat{w}_{k,M}, ..., \widehat{w}_{k-s,M}, y_{k-1}, ..., y_{k-p})^T$$

$$\widehat{\Psi}_{N,M} = (\widehat{\psi}_{1,M}, ..., \widehat{\psi}_{N,M})^T \tag{39}$$

$$\widehat{\psi}_{k,M} = (\widehat{w}_{k,M}, ..., \widehat{w}_{k-s,M}, \widehat{w}_{k-s-1,M}, ..., \widehat{w}_{k-s-p,M})^T$$

and where appropriate $\widehat{w}_{k-r,M}$ are computed (in stage 1) by a nonparametric technique (see (26)). If the nonparametric estimate $\widehat{R}_M(u)$ is bounded, converges pointwise to

the regression function $R(u)$ and at the estimation points $u \in \{0, u_{k-r}$; for $k = 1, 2, ..., N$ and $r = 0, 1, ..., s + p\}$ the error behaves like

$$\left|\widehat{R}_M(u) - R(u)\right| = O(M^{-\tau}) \text{ in probability} \tag{40}$$

then
(a') $Plim_{M,N\to\infty}\left(\frac{1}{N}\widehat{\Psi}^T_{N,M}\widehat{\Theta}_{N,M}\right)$ exists and is not singular
(b') $Plim_{M,N\to\infty}\left(\frac{1}{N}\widehat{\Psi}^T_{N,M}Z_N\right) = 0$
provided that $NM^{-\tau} \to 0$.

Theorem 6. (Hasiewicz and Mzyk [2004b]) For the estimate (37) with $\widehat{\Theta}_{N,M}$ and the instruments $\widehat{\Psi}_{N,M}$ as in (38) and (39) it holds that

$$\widehat{\theta}^{(IV)}_{N,M} \to \theta \text{ in probability} \tag{41}$$

as $N, M \to \infty$, provided that $NM^{-\tau} \to 0$. Particularly, for $M \sim N^{(1+\alpha)/\tau}$, $\alpha > 0$, the asymptotic rate of convergence is

$$\left\|\widehat{\theta}^{(IV)}_{N,M} - \theta\right\| = O(N^{-\min(\frac{1}{2},\alpha)}) \text{ in probability} \tag{42}$$

Theorem 7. (Hasiewicz and Mzyk [2004b]) For Hammerstein systems, the index $Q(\Psi_N)$ is asymptotically optimal for the instrumental matrix

$$\Psi^*_N = (\psi^*_1, \psi^*_2, ..., \psi^*_N)^T \text{ with} \tag{43}$$
$$\psi^*_k = (w_k, w_{k-1}, ..., w_{k-s}, v_{k-1}, v_{k-2}, ..., v_{k-p})^T$$

where $w_k, w_{k-1}, ..., w_{k-s}$ are interactions and v_{k-1}, v_{k-2}, $..., v_{k-p}$ are noise-free outputs of the system, i.e. for Ψ^*_N as in (43) and all other admissible choices of Ψ_N it holds that

$$\lim_{N\to\infty} Q(\Psi^*_N) \leqslant \lim_{N\to\infty} Q(\Psi_N) \quad \text{with probability 1} \tag{44}$$

As compared to the parametric identification techniques developed to date, the potential advantages of the approach are that: 1) we get simple estimates of both subsystems, given by the explicit formulas, 2) the routine is not using any type of alternate updating, 3) the method works with systems having non-polynomial static characteristics, 4) the algorithm operates efficiently for both white and colored noise, without the need of recovering the noise model, 5) each part of the system is identified separately making the estimates robust against lack or falsity of a priori information about the other part, and 6) convergence properties are established and rates of convergence are given.

4. FINAL REMARKS

A priori knowledge incorporation As it was noticed in the introduction, a desirable property of identification algorithms is their ability to assimilate (preferable in a systematic and elegant way) various kinds of prior information. Let as shortly retrieve the presented algorithm in this context. In parametric case the selection of basis function spanning the model space can be adjusted to the information about a special form of nonlinearity (*e.g.* polynomial of known order can be used, or a model with piecewise-smooth function can be applied, Vörös [1999, 2003]). In a nonparametric case, for known range of the input signal (in Hammerstein system) or known bound of the inverse of the nonlinearity (in Wiener system) allows to employ an appropriate orthogonal series, *e.g.* for unbounded input one can use Hermite polynomial or wavelet series, and trigonometric or Legendre, or Chebyshev polynomials otherwise. Also, if the smoothness of the nonlinearity (or its inverse) is known, the optimal bandwidth of kernel algorithms together with optimal kernel can both be applied – resulting in the fastest available convergence rate of the algorithms.

Remark 4. In case of decoupled algorithms recovering independently a nonlinear and linear parts, an incorrect information about the other subsystem does not affect the quality of the identification (in particular, both convergencies and their rates remain unchanged).

A note on computational complexity While both classes of parametric and nonparametric algorithms complete each other rather than compete (hence, in particular, can not directly comparable), one can make a safe comparison between them using a computational complexity criterion (*i.e.* a number of operations necessary to evaluate a final estimate). It seems to be intuitively clear and (can somehow be supported by the proper interpretation of the convergence rates of the presented algorithms), that the smaller *a priori* knowledge, the more measurements need to be collected to get similar quality of the identification algorithm. Therefore, it would be favorable to have a complexity of the computational routines corresponding to algorithms decreasing with the smaller *a priori* information since (intuitively) the smaller knowledge, the smaller number of (usually simpler) formulae (relations binding the measurement data) hold. These natural expectations are sheerly fulfilled in presented algorithms, as:

- the parametric algorithms, based on least squares principle, share a polynomial complexity $\mathcal{O}(N^p)$, $p \geq 3$, while
- the complexity of their nonparametric versions reduces to a linear complexity $\mathcal{O}(N)$ in case of wavelet and some kernel algorithms (for other orthogonal series algorithms it grows to $\mathcal{O}(N^p)$, $p \leq 2$, however).

4.1 Gauss discovery revisited

Let us consider the Gauss discovery once again, now a bit more from a "various prior knowledge" perspective. First, Ceres was spotted by Piazzi as a result of an exhaustive search (in an attempt to verify Titius-Body rule governing the distance of the Solar system objects from Sun). The observation of its position were recorded yet no orbit parameters had been established. Therefore, the dwarf-planet was lost after traversing behind Sun. Several astronomers (Body, von Zach, Olbers) tried to determine the orbit. They however used a wrong model (inappropriate *a priori* knowledge) assuming circular shape of the orbit (which result in *a biased model with systematic error*), and did also not correctly deal with error in measurements. It was Gauss who ingeniously not only took into account these errors (proposing his *least squares* algorithm to cope with *random errors*) but also used a better model admitting elliptical orbits (*e.g.* the one based on Kepler's laws); Abdulle and Wanner [2002]. That the Kepler's laws

were not an ultimate model for celestial bodies motion was discovered and explained another 100 years later by another genius, Albert Einstein, whose *general relativity theory* finally explained Mercury's orbit anomalies.

This history has been recalled to illustrate a well known fact that possessing a proper *a priori* information and measurements are both *conditia sine qua non* of the successful identification and that an improper prior knowledge cannot automatically be compensated by the measurements.

4.2 A proposition of identification based on semiparametric approach

At the end of this short summary we would like to propose an approach which would gain from both parametric and nonparametric algorithms and moreover would furnish a possibility of verification of the prior knowledge (represented by a parametric model). The approach is known in statistics and econometric literature as a *semiparametric* approach and therefore we would like to 'advertize' in a context of system identification.

Assume the underlying system is a Hammerstein one and we are interested in recovering its nonlinear element characteristic. The initial model is a polynomial one

$$\hat{\mu}_p(u) = \sum_{i=0}^{P} \hat{\alpha}_i p_i(u)$$

and has been selected for its convenience – as a kind of 'quick'n'dirty' solution. To establish parameters α_i least-squares algorithms is applied. There are three possibilities now:

- The model is correct and its parameters are correctly estimated (by happenstance).
- The model remains correct but its parameters are wrongly estimated (*e.g.* because of correlation of the external noise z_k).
- Neither model nor its parameters are correct (*e.g.* the genuine nonlinearity has jumps).

The proposed algorithm exploits wavelet, and particularly their property of having $P+1$ vanishing moments (*vis.* of being orthogonal to all polynomials of order P, where P can be set arbitrarily). Moreover, for simplicity, we assume that the system input is bounded and of uniform distribution. Hence, assuming that the nonlinearity is square integrable (which is a non issue in practice) we have (both scaling indices m, M and translation factors n are assumed to run through the proper indices ranges in order to represent $\mu(u)$ in a unit interval $[0,1]$; see Cohen et al. [1993])

$$\mu(u) = \sum_n \alpha_{Mn}\varphi_{Mn}(u) + \sum_m \sum_n \beta_{mn}\psi_{mn}(u)$$

Given a model $\hat{\mu}_p(u)$ we are interested in studying the residue function

$$\mu_r(u) = \mu(u) - \hat{\mu}_p(u)$$

Clearly, due to the aforementioned property of vanishing moment of wavelet functions, it holds that

$$\alpha_{Mn} = \langle \mu(u) - \hat{\mu}_p(u), \varphi_{Mn}(u) \rangle$$
$$\beta_{mn} = \langle \mu(u), \psi_{mn}(u) \rangle$$

and therefore the semiparametric algorithm $\hat{\mu}_p(u) + \hat{\mu}_r(u)$ converges to the actual nonlinearity $\mu(u)$ regardless of the initial model, with the wavelet expansion coefficients computed as

$$\hat{\alpha}_{Mn} = \frac{1}{N}\sum_{k=1}^{N} y_k \varphi_{Mn}(u_k) - \alpha_{Mn}$$

$$\hat{\beta}_{mn} = \frac{1}{N}\sum_{k=1}^{N} y_k \psi_{mn}(u_k).$$

Examining the coefficients $\hat{\alpha}_{Mn}$ and $\hat{\beta}_{mn}$ we can moreover draw the following conclusions about the initial model:

- if all $\hat{\beta}_{mn} \sim 0$ (*i.e.* they are statistically insignificant), and if also all $\hat{\alpha}_{Mn} \sim 0$, then $\hat{\mu}_p(u)$ is a proper model,
- if all $\hat{\beta}_{mn} \sim 0$ but any $\hat{\alpha}_{Mn}$ is significant, then the model is correct yet has wrongly estimated parameters,
- if any of $\hat{\beta}_{mn}$ is significant, then the model is incorrect.

Semiparametric algorithms seem to be a valuable proposal for nonlinear system identification since they are statistically and numerically advantageous for a wide range of sizes of measurement sets: from small (when parametric estimate works) to large (where nonparametric estimate plays the main role).

ACKNOWLEDGEMENTS

We would like to thank Prof. Fouad Giri for invitation to this session.

REFERENCES

A. Abdulle and G. Wanner. 200 years of least squares method. *Elemente der Mathematik*, 57:45–60, 2002.

E.W. Bai, An optimal two-stage identification algorithm for Hammerstein-Wiener nonlinear systems, *Automatica*, vol.34, No.3, pp. 333–338, 1998.

E.W. Bai, Frequency domain identification of Hammerstein models, IEEE Transactions on Automatic Control, Vol. 48, No. 4, pp. 530–542, 2003.

S. A. Billings. Identification of non-linear systems—a survey. *Proceedings of IEE*, 127(6):272–285, 1980.

S.A. Billings, S.Y. Fakhouri, Identification of nonlinear systems using the Wiener model, *Electronics Letters*, vol. 13, No. 17, pp. 502-504, 1977.

S.A. Billings, S.Y. Fakhouri, Identification of systems containing linear dynamic and static nonlinear elements, *Automatica*, vol. 18, No. 1, pp. 15-26, 1982.

E. Capobianco. Hammerstein system representation of financial volatility processes. *The European Physical Journal B - Condensed Matter*, 27:201–211, 2002.

F.H.I. Chang, R. Luus, A non-iterative method for identification using Hammerstein model, *IEEE Transactions on Automatic Control*, vol. AC-16, pp. 464-468, 1971.

Albert Cohen, Ingrid Daubechies, and Pierre Vial. Wavelet bases on the interval and fast algorithms. *Journal of*

Applied and Computational Harmonic Analysis, 1(1): 54–81, 1993.

E.J. Dempsey and D.T. Westwick. Identification of Hammerstein models with cubic spline nonlinearities. *IEEE Transactions on Biomedical Engineering*, 51(2): 237–245, 2004.

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.

G. B. Giannakis and E. Serpedin. A bibliography on nonlinear system identification. *Signal Processing*, 81: 533–580, 2001.

W. Greblicki. Nonparametric orthogonal series identification of Hammerstein systems. *International Journal of Systems Science*, 20:2355–2367, 1989.

W. Greblicki. Nonparametric identification of Wiener systems. *IEEE Transactions on Information Theory*, 38(5):1487–1493, 1992.

W. Greblicki. Nonparametric identification of Wiener systems by orthogonal series. *IEEE Transactions on Automatic Control*, 39(10):2077–2086, 1994.

W. Greblicki. Nonparametric approach to Wiener system identification. *IEEE Transactions on Circuits and Systems - I: Fundamental Theory and Applications*, 44:538–545, 1997.

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

W. Greblicki and M. Pawlak. Nonparametric identification of Hammerstein systems. *IEEE Transactions on Information Theory*, 35:409–418, 1989.

W. Greblicki and M. Pawlak. Nonparametric identification of a particular nonlinear time series system. *IEEE Transactions on Signal Processing*, 40:985–989, 1992.

W. Greblicki. Continuous-time Hammerstein system identification. *IEEE Transactions on Automatic Control*, 45: 1232–1236, 2000.

W. Greblicki. Continuous-time Wiener system identification. *IEEE Transactions on Automatic Control*, 43: 1488–1493, 1998.

R. Haber, L. Keviczky, *Nonlinear System Identification – Input-Output Modeling Approach*, vol. 1 & 2, Kluwer Academic Publishers, Dordrecht, 1999.

Z. Hasiewicz, Identification of a linear system observed through zero-memory non-linearity, *International Journal of Systems Science*, vol. 18, pp. 1595-1607, 1987.

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

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.

Z. Hasiewicz, G. Mzyk, Hammerstein system identification by nonparametric instrumental variables, submitted for publication in the *International Journal of Control*.

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*, 52(1):427–442, 2005.

K.J. Hunt, M. Munih, N.de.N. Donaldson, and F.M.D. Barr. Investigation of the Hammerstein hypothesis in the modeling of electrically stimulated muscle. *IEEE Transactions on Biomedical Engineering*, 45(8):998–1009, 1998.

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.

S. N. Jyothi and M. Chidambaram. Identification of Hammerstein model for bioreactors with input multiplicities. *Bioprocess Engineering*, 23(4):323–326, 2000.

S.L. Kukreja, R.E. Kearney, and H.L. 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.

S. L. Lacy, D. S. Bernstein, Identification of FIR Wiener systems with unknown, non-invertible, polynomial nonlinearities, *International Journal of Control*, vol. 76, No. 15, pp. 1500-1507, 2003.

L. Ljung, *System identification: theory for user*, Prentice Hall, Englewood Cliffs, 1987.

Shir-Kuan Lin. Identification of a class of nonlinear deterministic systems with application to manipulators. *IEEE Transactions on Automatic Control*, 39(9):1886–1893, 1994.

M. Lortie and R. E. Kearney. Identification of time-varying Hammerstein systems from ensemble data. *Annals of Biomedical Engineering*, 29(2):619–635, 2001.

G. Mzyk, Generalized kernel regression estimate for the identification of Hammerstein systems, *International Journal of Applied Mathematics and Computer Science*, vol. 17, No. 2, pp. 189-198, 2007.

G. Mzyk, A censored sample mean approach to nonparametric identification of nonlinearities in Wiener systems, *IEEE Transacions on Circuits and Systems II*, vol. CAS-54, No. 10, pp. 897-901, 2007.

K.S. Narendra, P.G. Gallman, An iterative method for the identification of nonlinear systems using the Hammerstein model, *IEEE Transactions on Automatic Control*, vol. AC-11, pp. 546-550, 1966.

D. Panescu, J.G. Webster, and R.A. Stratbucker. A nonlinear electrical-thermal model of the skin. *IEEE Transactions on Biomedical Engineering*, 41(7):672–680, 1994.

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.

M. Pawlak, Z. Hasiewicz, and P. Wachel. On nonparametric identification of Wiener systems. *IEEE Transactions on Signal Processing*, 55(5):482–492, 2007.

R. Pintelon, P. Guillaume, Y. Rolain, J. Schoukens, and H. Van Hamme. Parametric identification of transfer functions in the frequency domain-a survey. *IEEE Transactions on Automatic Control*, 39(11):2245–2260, 1994.

J. Schoukens, T. Dobrowiecki, and R. Pintelon. Parametric and nonparametric identification of linear systems inthe presence of nonlinear distortions-a frequency domain approach. *IEEE Transactions on Automatic Control*,

43(2):176–190, 1998.

T. Söderström, P. Stoica, Instrumental-variable methods for identification of Hammerstein system, *International Journal of Control*, vol. 35, No. 3, pp. 459-476, 1982.

P. Stoica, T. Söderström, *System Identification*, Englewood Cliffs, NJ: Prentice Hall, 1989.

C. J. Stone. Optimal rates of convergence for nonparametric regression. *Annals of Statistics*, 8(6):1348–1360, 1980.

P. Śliwiński and Z. Hasiewicz. Computational algorithms for multiscale identification of nonlinearities in Hammerstein systems with random inputs. *IEEE Transactions on Signal Processing*, 53(1):360–364, 2005.

P. Śliwiński and Z. Hasiewicz. Computational algorithms for wavelet identification of nonlinearities in Hammerstein systems with random inputs. *IEEE Transactions on Signal Processing*, vol. SP-56, No. 2, pp. 845-851, 2008.

J. Vörös. Recursive identification of Hammerstein systems with discontinuous nonlinearities containing dead-zones. *IEEE Transactions on Automatic Control*, 48(12):2203–2206, 2003.

J. Vörös. Iterative algorithm for parameter identification of Hammerstein systems with two-segment nonlinearities. *IEEE Transactions on Automatic Control*, 44(12): 2145–2149, 1999.

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.

K. Wong, E. Polak, Identification of linear discrete time systems using the instrumental variable method, *IEEE Transactions on Automatic Control*, vol. AC-12, No. 6, pp. 707-718, 1967.

X. Zhu and D. E. Seborg. Nonlinear predictive control based on Hammerstein system. *Control Theory Applications*, 11:564–575, 1994.

L. Zi-Qiang. Controller design oriented model identification method for Hammerstein system. *Automatica*, 29: 767–771, 1993.