Mixed parametric-nonparametric identification of Hammerstein and Wiener systems - a survey

Zygmunt Hasiewicz, Grzegorz Mzyk, Przemyslaw Śliwiński, Paweł Wachel *

* *Institute of Computer Engineering, Control and Robotics, Wrocław University of Technology, 50-370 Wrocław, Poland, phone: +48 71 320 32 77, e-mail: grzegorz.mzyk@pwr.wroc.pl*

Abstract: The paper surveys the ideas of cooperation between parametric and nonparametric (kernel-based) algorithms of nonlinear block-oriented system identification. Various strategies are proposed, dependently on the system structure, number of data and the prior knowledge. The estimates are consistent and their rates of convergence are presented. The aim of the paper is to show some recent results in the field in a systematic, ordered way.

Keywords: Hammerstein system, Wiener system, nonparametric identification, kernel estimate, least squares, instrumental variables, semiparametric methods, recursive algoritms.

1. INTRODUCTION

In the paper we identify the nonlinear static and linear dynamic components in block-oriented (e.g. Hammerstein and Wiener) systems under various kinds of prior knowledge and for various number of measurements. The problem is often met in practise and the selection of appropriate model is of fundamental meaning (Giannakis and Serpedin [2001], Chen [2006], D'Amato *et al.* [2011], Jiang and Fang [2009]). Popular methods (e.g. least squares) are connected with the risk of bad parametrization, which causes systematic approximation error. In the typical approaches, the parameters of both subsystems are aggregated (see e.g. Bai [1998], Giri and Bai [2010]) and jointly estimated. It usually leads to procedures of matrix decomposition and nonlinear optimization, which can be badly conditioned numerically. On the other hand, recent algorithms of nonparametric regression (Greblicki and Pawlak [2008], Hasiewicz *et al.* [2005], Greblicki and Mzyk [2009]), which can be applied without prior assumptions, converge relatively slower, because the signals produced by the dynamic components are treated as the 'system noises'. In the proposed schemes, cooperation between parametric and nonparametric methods allows to solve typical problems like: (i) lack or uncertainty of prior knowledge about one or more components, (ii) long-history dynamics and small number of measurements, (iii) correlated excitations, non-Gaussian inputs, and (iv) nonlinear-in-the-parameters characteristics.

The paper is organized as follows. In Section 1 we identify Hammerstein system by the two step method. First, the pilot nonparametric kernel regression estimate is used to decompose the identification task, and next, the least squares (in FIR case) or instrumental variables method (in IIR case) is applied for identification of particular component. Analogous ideas are presented, in Section 2, for Wiener system, i.e. the correlation-based method and average derivative method for finite memory case (FIR), and two multiple regression estimates for IIR case. Finally, in Section 3, we present more general problems of correlated input, small sample size and the other structures, such as Uryson, MIMO and Doherty models. The algorithms shown in the note have a formal justification and, in particular, the conditions of the algorithms convergence and the convergence rates are all established and presented in the cited literature. Due to the space limitation and survey character of the paper, the standard assumption-theorem-proof convention is therefore omitted.

[1] This research was supported by Polish National Science Centre, Grant No. N N514 700640

2. HAMMERSTEIN SYSTEM

2.1 Dynamics with finite memory

In the simplest case of Hammerstein system (Fig. 1) with the linear-in-the-parameters nonlinearity $w_k = \phi^T(u_k)\, c$, where $\phi(u_k) = (f_1(u_k), f_2(u_k), ..., f_m(u_k))^T$, and MA dynamics $y_k = \sum_{i=0}^{s} \gamma_i w_{k-i} + z_k$ $(s < \infty)$, the identification task can be decomposed by the internal signal estimation, with the help of nonparametric kernel method

$$\widehat{w}_{k,M} = \widehat{R}_M(u_k) - \widehat{R}_M(0), \quad \widehat{R}_M(u) = \frac{\sum_{k=1}^{M} y_k K(\frac{u-u_k}{h(M)})}{\sum_{k=1}^{M} K(\frac{u-u_k}{h(M)})}. \tag{1}$$

where $K()$ is a kernel function and $h(M)$ – a bandwidth parameter. Then, the *least squares* method is used

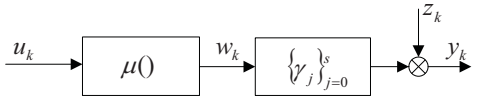

Fig. 1. Hammerstein system

to the independent estimation of both subsystems parameters using, respectively, the pairs $\{(u_k, \widehat{w}_k)\}_{k=1}^{N_0}$ and $\{(\widehat{w}_k, y_k)\}_{k=1}^{N}$, i.e.

$$\widehat{c}_{N_0,M} = \left(\Phi_{N_0}^T \Phi_{N_0}\right)^{-1} \Phi_{N_0}^T \widehat{W}_{N_0,M}, \tag{2}$$

$$\widehat{\gamma}_{N,M} = (\widehat{\Theta}_{N,M}^T \widehat{\Theta}_{N,M})^{-1} \widehat{\Theta}_{N,M}^T Y_N, \tag{3}$$

where

$$\Phi_{N_0} = (\phi(u_1), \phi(u_2), ..., \phi(u_{N_0}))^T,$$
$$\widehat{W}_{N_0,M} = (\widehat{w}_{1,M}, \widehat{w}_{2,M}, ..., \widehat{w}_{N_0,M})^T,$$
$$\widehat{\Theta}_{N,M} = (\widehat{\vartheta}_{1+s,M}, \widehat{\vartheta}_{2+s,M}, ..., \widehat{\vartheta}_{N+s,M})^T,$$
$$\widehat{\vartheta}_{t,M} = (\widehat{w}_{t,M}, \widehat{w}_{t-1,M}, ..., \widehat{w}_{t-s,M})^T,$$

and Y_N is a noisy output vector. The proofs of consistency and convergence rate analysis can be found in Hasiewicz and Mzyk [2004]. In particular, if in the nonparametric stage $\left|\widehat{R}_M(u) - R(u)\right| = O(M^{-\tau})$ in probability as $M \to \infty$ for each $u \in \{0, u_n;\ n = 1, 2, ..., N_0\}$ then also $\|\widehat{c}_{N_0,M} - c\| = O(M^{-\tau})$ in probability as $M \to \infty$. Moreover, if for $u \in \{0, u_{t-r}; t = n+s; n = 1, 2, ..., N; r = 0, 1, ..., s\}$ the estimate $\widehat{R}_M(u)$ is bounded and the asymptotic nonparametric estimation error behaves like $\left|\widehat{R}_M(u) - R(u)\right| = O(M^{-\tau})$ in probability, then $\widehat{\gamma}_{N,M} \to \gamma$ in probability provided that $N, M \to \infty$ and $NM^{-\tau} \to 0$. For $M \sim N^{(1+\alpha)/\tau}$, equivalently $N \sim M^{\tau/(1+\alpha)}$, $\alpha > 0$, the asymptotic convergence rate in the parametric stage is $\left\|\widehat{\gamma}_{N,M} - \gamma\right\| = O(N^{-\min(1/2,\alpha)})$ in probability.

2.2 Dynamics with infinite memory

For the infinite impulse response ARMA linear dynamics

$$v_k = b_0 w_k + ... + b_s w_{k-s} + a_1 v_{k-1} + ... + a_p v_{k-p}$$

with unknown parameters $\theta = (b_0, b_1, ..., b_s, a_1, a_2, ..., a_p)^T$, the combined least squares estimate (2) can be generalized to instrumental variables one

$$\widehat{\theta}_{N,M}^{(IV)} = (\widehat{\Psi}_{N,M}^T \widehat{\Theta}_{N,M})^{-1} \widehat{\Psi}_{N,M}^T Y_N, \tag{4}$$

where

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

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 $\left|\widehat{R}_M(u) - R(u)\right| = O(M^{-\tau})$ in probability then $Plim_{M,N\to\infty} \frac{1}{N} \widehat{\Psi}_{N,M}^T Z_N = 0$, where $Z_N = (z_1, ..., z_N)^T$, and $Plim_{M,N\to\infty} \frac{1}{N} \widehat{\Psi}_{N,M}^T \widehat{\Theta}_{N,M}$ does exist and is not singular, provided that $NM^{-\tau} \to 0$. Consequently, it holds that $\widehat{\theta}_{N,M}^{(IV)} \to \theta$ in probability 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 $\left\|\widehat{\theta}_{N,M}^{(IV)} - \theta\right\| = O(N^{-\min(\frac{1}{2},\alpha)})$ in probability.

In Mzyk [2011] the idea of instrumental variables approach was generalized in the sense that it allows to eliminate the bias caused by correlation of both the noise and input processes of a nonlinear system.

3. WIENER SYSTEM

3.1 Dynamics with finite memory

In Pawlak *et al.* [2007] the Wiener system (Fig. 2)

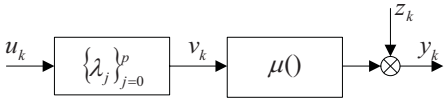

Fig. 2. Wiener system

$$v_n = \sum_{l=0}^{p} \lambda_l u_{n-l}, \tag{5}$$

$$y_n = \mu(v_n) + z_n. \tag{6}$$

with a finite memory (of length p) is identified by a two-step parametric-nonparametric approach. The method permits recovery of a wide-class of nonlinearities which need not be invertible. Furthermore, the algorithm allows non-Gaussian input signals and the presence of additive output measurement noise. We begin with splitting the measurement set into two disjoint subsets. Let I_1, I_2 denote sets of measurements indexes in these subsets (n_1, n_2 are the numbers of elements in I_1 and I_2, respectively). The impulse response function of the linear part is recovered via minimization ($\hat{\lambda} = \arg\min_{\boldsymbol{\lambda}} \hat{Q}(\boldsymbol{\lambda})$) of the following nonlinear least squares criterion

$$\hat{Q}(\boldsymbol{\lambda}) = \frac{1}{n_2} \sum_{i \in I_2} \{y_i - \hat{\mu}(v_i(\boldsymbol{\lambda}); \boldsymbol{\lambda})\}^2,$$

where the system non-linearity $\mu(\cdot)$ is pre–estimated by a pilot non-parametric kernel regression estimate of the form

$$\hat{\mu}(v; \boldsymbol{\lambda}) = \frac{\sum_{j \in I_1} y_j K(\frac{v - v_j(\boldsymbol{\lambda})}{h})}{\sum_{j \in I_1} K(\frac{v - v_j(\boldsymbol{\lambda})}{h})}.$$

and where $h = h(n_1)$ is the bandwidth parameter. The estimate $\hat{\lambda}$ is then used to recover interconnection v_n by the estimate $v_n(\hat{\lambda})$. In consequence, using the sequence of pairs $(v_n(\hat{\lambda}), y_n)$, the non-parametric kernel estimate of the nonlinearity $\mu(\cdot)$ is built in a following way

$$\hat{\mu}(v) = \frac{\sum_{j \in I_1} y_j K(\frac{v - v_j(\hat{\lambda})}{h})}{\sum_{j \in I_1} K(\frac{v - v_j(\hat{\lambda})}{h})}.$$

The consistency of the estimates $\hat{\lambda}$ and $\hat{\mu}(v)$ is established in Pawlak *et al.* [2007] for a large class of input signals and non-invertible nonlinearities.

Average derivative estimation Identification of the Wiener system (5)–(6) by the average derivative approach (originated from Powell *et al.* [1989]) was independently introduced in Wachel [2010] and Greblicki and Pawlak [2008]. Similarly to the algorithm described above, the method allows to recover unknown nonlinearity directly (*i.e.* not by the inverse function) and in the systems having non-Gaussian input. The algorithm requires the following alternative system description

$$y_n = G\left(\underline{u}_n\right) + z_n, \tag{7}$$

where $\underline{u}_n = (u_n, u_{n-1}, \ldots, u_{n-p})$ is generalized vector input, and $G(\cdot) : \mathbf{R}^{p+1} \to \mathbf{R}$ is of the form

$$G\left(\underline{u}\right) = \mu\left(\underline{\lambda}^T \underline{u}\right), \qquad \underline{u} \in \mathbf{R}^{p+1}. \tag{8}$$

In representation (7)–(8) the impulse response of the dynamic subsystem is given by the vector $\underline{\lambda} = (\lambda_0, \lambda_1, \ldots, \lambda_p)^T$. Let $D_G\left(\underline{u}\right)$ be the gradient of $G(\cdot)$ in $\underline{u}$. Observe that

$$D_G\left(\underline{u}_1\right) = \mu'\left(\underline{\lambda}^T \underline{u}_1\right)\underline{\lambda}. \tag{9}$$

To further elucidate the main idea of the method, let $f_{\underline{u}}\left(\underline{u}\right)$ be the density function of the generalized vector input $\underline{u}_n$. Clearly, for the *i.i.d.* scalar input u_n having density $f(\cdot)$, the density $f_{\underline{u}}\left(\underline{u}\right)$ takes the form $f_{\underline{u}}\left(\underline{u}\right) = \prod_{j=1}^{p+1} f\left(\underline{u}^{(j)}\right)$, $\underline{u} \in \mathbf{R}^{p+1}$,where $\underline{u}^{(j)}$ denotes j–th entry of $\underline{u}$. Due to (9), the following equality holds

$$f_{\underline{u}}\left(\underline{u}_1\right) D_G\left(\underline{u}_1\right) = f_{\underline{u}}\left(\underline{u}_1\right)\mu'\left(\underline{\lambda}^T \underline{u}_1\right)\underline{\lambda},$$

and therefore

$$E\left\{f_{\underline{u}}\left(\underline{u}_1\right) D_G\left(\underline{u}_1\right)\right\} = E\left\{f_{\underline{u}}\left(\underline{u}_1\right)\mu'\left(\underline{\lambda}^T \underline{u}_1\right)\right\}\underline{\lambda}. \tag{10}$$

Applying integration by parts we note, in turn, that $E\left\{f_{\underline{u}}\left(\underline{u}_1\right) D_G\left(\underline{u}_1\right)\right\} = -2E\left\{y_1 D_f\left(\underline{u}_1\right)\right\}$ where $D_f\left(\underline{u}_1\right)$ is a gradient of $f_{\underline{u}}(\cdot)$. This and (10) gives

$$\underline{\lambda} = \kappa E\left\{y_1 D_f\left(\underline{u}_1\right)\right\}, \tag{11}$$

where κ is constant equal to $-2\left(E\left\{f_{\underline{u}}\left(\underline{u}_1\right)\mu'\left(\underline{\lambda}^T \underline{u}_1\right)\right\}\right)^{-1}$. The formula (11) allows now to construct the estimate $\hat{\underline{\lambda}}$ of the scaled impulse response $\underline{\lambda}$ by estimating expectation in (11). To this end observe, that (in explicit form) gradient $D_f\left(\underline{u}_1\right)$ equals $\left[d_{f,1}, d_{f,2}, \ldots, d_{f,p+1}\right]^T$ where

$$d_{f,i}\left(\underline{u}\right) = f'\left(\underline{u}^{(i)}\right)\prod_{j=1, j\neq i}^{p+1} f\left(\underline{u}^{(j)}\right). \tag{12}$$

Hence, assuming that input density $f(\cdot)$ is known and differentiable, the estimate of i–th entry of the impulse response $\underline{\lambda}$ has the following simple form

$$\hat{\lambda}_i = \frac{1}{N}\sum_{k\in I} Y_k d_{f,i+1}\left(\underline{u}\right), \tag{13}$$

and as an estimate of vector $\underline{\lambda}$ we take $\hat{\lambda} = \left[\hat{\lambda}_0, \hat{\lambda}_1, \ldots, \hat{\lambda}_p\right]^T$. Having $\hat{\lambda}$ in turn, the estimate $v_n(\hat{\lambda})$ of the interconnection v_n is introduced and the unknown nonlinearity $\mu(\cdot)$ is recovered by

$$\hat{\mu}\left(v\right) = \frac{\sum_{j\in I} y_j K(\frac{v - v_j(\hat{\lambda})}{h})}{\sum_{j\in I} K(\frac{v - v_j(\hat{\lambda})}{h})}.$$

If the input density $f(\cdot)$ is unknown, partial derivatives $d_{f,i+1}$ in (13) can not be directly evaluated. In such cases however, one can utilize plug–in estimates of $f(\cdot)$ and its derivative $f'(\cdot)$ in (12).

Due to the fundamental role of differentiation in the considered approach (see (9)), certain requirements regarding smoothness of nonlinearity and density of input signal have to be imposed. These and other assumptions are discussed in Wachel [2010] and Greblicki and Pawlak [2008] where the formal analysis of the estimates can also be found.

3.2 Dynamics with infinite memory

For Wiener systems with IIR linear dynamics we propose one of the following nonlinearity estimates (see Mzyk [2007], Mzyk [2010*a*] and Greblicki [2010])

$$\widehat{\mu}_N^{(1)}(u) = \frac{\sum_{k=1}^N y_k \cdot K\left(\frac{\sum_{j=0}^k |u_{k-j} - u|\lambda^j}{h(N)}\right)}{\sum_{k=1}^N K\left(\frac{\sum_{j=0}^k |u_{k-j} - u|\lambda^j}{h(N)}\right)}, \tag{14}$$

$$\widehat{\mu}_N^{(2)}(u) = \frac{\sum_{k=1}^N y_k \prod_{i=0}^p K\left(\frac{u - u_{k-i}}{h(N)}\right)}{\sum_{k=1}^N \prod_{i=0}^p K\left(\frac{u - u_{k-i}}{h(N)}\right)} \tag{15}$$

Both of them are based on multiple regression and admit arbitrary input density function. The symbols λ and p denote the upper bound of impulse response decaying, and the order of approximation, respectively. The estimates $\widehat{\mu}_N^{(1)}(u)$ and $\widehat{\mu}_N^{(2)}(u)$ can be also successfully applied (Mzyk [2010*b*]) for the Hammerstein system and sandwich (Wiener-Hammerstein) system, shown in Fig. 3.

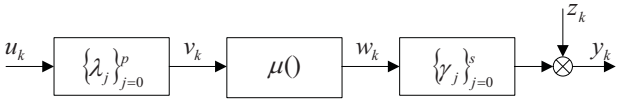

Fig. 3. Wiener-Hammerstein (sandwich) system

4. GENERALIZATIONS

4.1 Long-range dynamics and hard nonlinearities in Hammerstein system

It was shown in Mzyk [2009] that the error of nonparametric kernel estimate of the regression function in Hammerstein system can be significantly reduced by the inverse linear pre-filtering of the output process, which leads to the estimate

$$\widehat{\mu}_N^f(u) = \frac{\sum_{k=1}^N y_k^f K\left(\frac{u_k - u}{h_N}\right)}{\sum_{k=1}^N K\left(\frac{u_k - u}{h_N}\right)}, \tag{16}$$

where

$$y_k^f = y_k - \widehat{a}_{1,N} y_{k-1} - \widehat{a}_{2,N} y_{k-2} - \ldots - \widehat{a}_{p,N} y_{k-p}, \tag{17}$$

and the coefficients $\widehat{a}_{1,N}$, $\widehat{a}_{2,N}$,...,$\widehat{a}_{p,N}$ are obtained by any deconvolution method. Moreover, the idea of mixed, parametric-nonparametric approach allows for significant generalization of the class of admitted nonlinear characteristics (Hasiewicz and Mzyk [2009]). If the parametric

model $\mu(u,c)$ of the true nonlinearity $\mu(u)$ is not linear with respect to c, application of the nonlinear least squares procedure is still possible, i.e., we perform minimization of the following criterion

$$\widehat{Q}_{N_0,M}(c)=\sum_{n=1}^{N_0}\left[\widehat{w}_{n,M}-\mu(\overline{u}_n,c)\right]^2, \text{ i.e.,} \tag{18}$$

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

Let the computed $\widehat{c}_{N_0,M}$ be unique for each M. If in the nonparametric stage, for some $\tau>0$, it holds that

$$\widehat{R}_M(\overline{u}_n)=R(\overline{u}_n)+O(M^{-\tau}) \text{ in probability} \tag{19}$$

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

$$\widehat{c}_{N_0,M}=c^*+O(M^{-\tau}) \text{ in probability.} \tag{20}$$

4.2 Semiparametric approach

The interesting idea of combining parametric methods with nonparametric ones was introduced in Śliwiński *et al.* [2009] and Greblicki and Mzyk [2009]. Since the nonparametric algorithms have usually large error for small number of data, the traditional (parametric) model $\mu(u,\widehat{c})$ can be elastically substituted by the nonparametric regression estimate $\widehat{\mu}_N(u)$, when the number of observations grows large. It leads to the following semiparametric routines $\hat{\mu}_1(u)$ and $\hat{\mu}_2(u)$

$$\hat{\mu}_1(u)=\lambda_N\mu(u,\widehat{c})+(1-\lambda_N)\,\widehat{\mu}_N(u), \tag{21}$$

with $\lambda_N\to 0$ as $N\to\infty$, or

$$\hat{\mu}_2(u)=\mu(u,\widehat{c})+\widehat{r}_N(u), \tag{22}$$

where $\widehat{r}_N(u)$ is the nonparametric estimate of the remaining part $r(u)=\mu(u,c^*)-\mu(u)$, and $\mu(u,c^*)$ is the best approximation of the true characteristic $\mu(u)$ in the parametric class $\mu(u,c)$. The convergence analysis of (21) and (22) are straightforward.

4.3 Examples of Hammerstein-type structures

From the regression estimation-based algorithms point of view, the Hammerstein systems are seen as the static (memoryless) elements disturbed by the additive and correlated *system noise* (*i.e.* properly reinterpreted system dynamics). There exist a bunch of other structures, for which such an interpretation remains valid and, therefore, the identification algorithms designed for the (canonical) Hammerstein systems can be applied; *cf. e.g.* Gallman [1975], Pawlak and Hasiewicz [1998], Hasiewicz *et al.* [2005], [Greblicki and Pawlak 2008, Ch. 12]. Furthermore, the Hammerstein structures are basic components the Uryson and MISO (*i.e. multiple-input single-output*) dynamic systems with input nonlinearities; see Fig. 4 and 5. These nonlinear multibranched structures are composed of Hammerstein systems connected in parallel, where, in the former, the input signal, u_k, is common for all subsystems, while in the latter, each uth branch, $u=1,\ldots,U$, is driven independently. In the first part of this section we consider the identification limits for Uryson and MISO systems and then we present three Haar wavelet-based algorithms which recover the system nonlinearities in a recursive fashion.

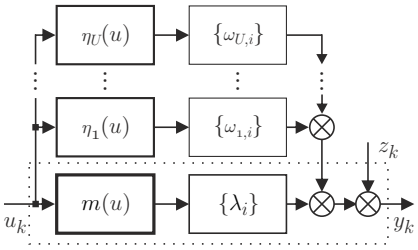

Fig. 4. Uryson system

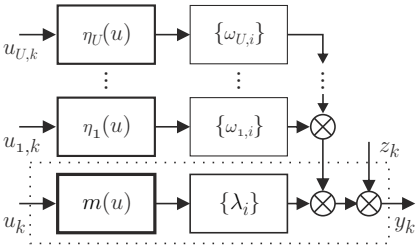

Fig. 5. MISO system

Uryson system The input-output description equation of the Uryson system has the form:

$$y_k=\sum_{i=0}^{\infty}\lambda_i m\left(u_{k-i}\right)+\sum_{u=1}^{U}\sum_{i=0}^{\infty}\omega_{u,i}\eta_u\left(u_{k-i}\right)+z_k.$$

The regression function of the system output on the system input in the case of Uryson system is a weighted sum of scaled nonlinearities from all the system branches (shifted by some (unknown) constant factor β):

$$\begin{aligned}&E\left\{y_k\,|u_k=u\right\}=\lambda_0 E\left\{m\left(u_k\right)|u_k=u\right\}\\&+\sum_{u=1}^{U}\omega_{u,0}E\left\{\eta_u\left(u_k\right)|u_k=u\right\}\\&+\underbrace{\sum_{i=1}^{\infty}\lambda_{u,i}Em\left(u_{k-i}\right)+\sum_{u=1}^{U}\sum_{i=1}^{\infty}\omega_{u,i}E\eta_u\left(u_{k-i}\right)}_{=\beta}\\&=\underbrace{\lambda_0 m\left(u\right)+\beta}_{=\mu(u)}+\underbrace{\sum_{u=1}^{U}\omega_{u,0}\eta_u\left(u\right)}_{=\mu_u(u)}=\mu_U\left(u\right),\end{aligned} \tag{23}$$

In general, the branch nonlinearities cannot be recovered separately. Nevertheless, there are several specific situations of practical significance, when, for instance, the nonlinearity $\mu\left(u\right)$ can still be estimated. For instance:

- When it holds that $\omega_{u,0}=0$ for all $u=1,\ldots,U$, *i.e.* when all other dynamic subsystems, $\{\omega_{u,i}\}$, have a non-zero delay (its transmittance has a pole at $z=0$); then clearly $E\left\{y_k\,|u_k=u\right\}=\mu_U\left(u\right)=\mu\left(u\right)$.
- When it holds that $\operatorname{supp}\mu\left(u\right)\cap\operatorname{supp}\eta_u\left(u\right)=\emptyset$ for all $u=1,\ldots,U$, *i.e.* when all the branch nonlinearities are active in the input signal ranges non-overlapping with the active input range of $\mu\left(u\right)$; then, in the activity region of $\mu\left(u\right)$, we obtain $E\left\{y_k\,|u_k=u\right\}=\mu_U\left(u\right)=\mu\left(u\right)$.

Example 1. The structure of the well known *Doherty power amplifier* (see Fig. 6), which is commonly used in

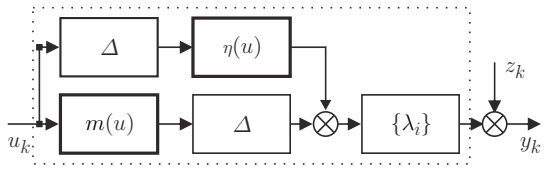

Fig. 6. Doherty amplifier model (Δ denotes a unit delay system, *e.g.* $\Delta[u_k] = u_{k-1}$)

RF systems; *cf. e.g.* Cripps [2006]) can be described with the help of the following input-output equation

$$y_k = \sum_{i=0}^{\infty} \lambda_i \{\Delta[m(u_{k-i})] + \eta(\Delta[u_{k-i}])\} + z_k$$
$$= \sum_{i=1}^{\infty} \lambda_{i-1}[m(u_{k-i}) + \eta(u_{k-i})] + z_k,$$

where $m(u)$, $\eta(u)$ are the so called *carrier* and *peak* nonlinear amplifiers, respectively. The whole system nonlinearity can be recovered using any Hammerstein system identification algorithm since for the regression function in this system we have that

$$E\{y_k | u_{k-1} = u\} = \lambda_0[m(u) + \eta(u)]$$
$$+\underbrace{\sum_{i=2} \lambda_{i-1} E[m(u_{k-i}) + \eta(u_{k-i})]}_{=\beta} = \mu(u).$$

MISO system The identification conditions for the MISO one in Fig. 5 are much less stringent: if the input signals u_k and $u_{u,k}$, $u = 1, \ldots, U$, are stochastically independent, then it holds that, *cf.* (23):

$$E\{y_k | u_k = u\} = \lambda_0 m(u)$$
$$+\underbrace{\sum_{i=1}^{\infty} \lambda_{u,i} E m(u_{k-i}) + \sum_{u=1}^{U} \sum_{i=0}^{\infty} \omega_{u,i} E \eta_u(u_{u,k-i})}_{=\beta} = \mu(u),$$

and, by estimating the regression function from the measurement pairs (u_k, y_k), the nonlinearity $\mu(u)$ can be recovered regardless of the other nonlinearities and properties of the component dynamic subsystems.

4.4 Recursive recovery of the system nonlinearity

The problem of nonparametric recursive recovery of the system nonlinearity in Hammerstein-type systems has been addressed by several authors, but mostly parametric recursive algorithms (aimed at estimation of the model unknown parameters) were proposed; see *e.g.* Boutayeb and Darouach [1995]-Dempsey and Westwick [2004]). A few nonparametric recursive algorithms, recovering the unknown characteristics point-by-point, were proposed in Greblicki and Pawlak [1989], Chen [2004]; *cf.* the summary in [Greblicki and Pawlak 2008, Chs. 4 and 5]. In the remainder, we present three simple nonparametric algorithms, derived from the *off-line* quotient-form estimates and *off-line* order statistic-based estimates, and eventually based on the stochastic approximation algorithms. For simplicity, the algorithms employ Haar wavelets, which basic properties are now shortly recollected. Let $I(x)$ be the indicator function of the unit interval $[0, 1]$. The reproducing kernel of the Haar series has the form $\phi_{K(k)}(u, v) = 2^{K(k)} \phi\left(2^{K(k)} u, 2^{K(k)} v\right)$ (denoted further by $\phi_k(u)$ for shortness), where $\phi(u, v) = I(u - \lfloor v \rfloor)$. The scale factors sequence $K(k)$ for $k = 1, 2, \ldots$, forms an increasing sequence dependent on the running index k of processed data pairs (u_k, y_k) (and is a wavelet counterpart of the bandwidth parameter used in kernel algorithms).

The first, *recursive quotient identification algorithm*, has the following simple formula

$$\hat{\mu}_k(u) = \hat{\mu}_{k-1}(u) + \gamma_k(u)\left[y_k - \hat{\mu}_{k-1}(u)\right], \tag{24}$$

where $\gamma_k(u) = \vartheta_k(u) / \hat{\kappa}_k(u)$ is a weighting factor with $\vartheta_k(u) = 2^{-K(k)} \phi_{K(k)}(u, u_k) = \phi\left(2^{K(k)} u, 2^{K(k)} v\right)$ being the non-normalized version of the reproducing kernel $\phi_{K(k)}(u, v)$ and with a denominator $\hat{\kappa}_k(x) = \hat{\kappa}_{k-1}(x) + \vartheta_k(x)$ acting as a counter of those measurement pairs, which input values are in a vicinity of the estimation point u. The algorithm's initial conditions are clearly $\hat{\mu}_0(u) = \hat{\kappa}_0(u) = 0$. Observe that the weighting factor $\gamma_k(u)$ has a random (yet for the examined Haar wavelet instance) nonnegative denominator.

The second, *stochastic approximation identification algorithm*, has a similar form, *i.e.*

$$\hat{\mu}_k(u) = \hat{\mu}_{k-1}(u) + \gamma_k(u)\left[y_k - \hat{\mu}_{k-1}(u)\right], \tag{25}$$

where the weighting factor $\gamma_k(u) = \vartheta_k(u) / \kappa_k$, with $\kappa_k = k^{-1+\varepsilon}$, for arbitrarily small $\varepsilon > 0$, remains random, but its denominator is now deterministic. The algorithm's initial conditions is clearly $\hat{\mu}_0(u) = 0$.

In order to present the last, *order statistics recursive identification algorithm*, we need some additional notation. Let $\Phi_{\mathfrak{m}}(u, v)$ be the *indefinite integral*, $\Phi_{\mathfrak{m}}(u, v) = \int \phi_{\mathfrak{m}}(u, v)\, dv$, of the kernel function $\phi_{\mathfrak{m}}(u, v)$. Let the measurements are kept sorted w.r.t. the input values u_k. The algorithm takes the form

$$\hat{\mu}_k(u) = \hat{\mu}_{k-1}(u) + \gamma_k(u)(y_k - y_k^R) \tag{26}$$

where $\gamma_k(u) = \Phi_{\mathfrak{m}}(u_k, u) - \Phi_{\mathfrak{m}}(u_k^L, u)$ plays the role of the random gain (weighting) factor, and where (u_k^L, y_k^L) and (u_k^R, y_k^R) are earlier measurements being the left- and the right-closest to the new incoming pair (u_k, y_k) with respect to the input values. The algorithm starts with $\hat{\mu}_1(u) = 0$ and, by definition, with two initial pairs $\{(u_{-1}, y_{-1}), (u_0, y_0)\} = \{(0, 0), (1, 0)\}$.

All three algorithms reveal only minor differences in their asymptotic properties. The convergence properties of the algorithms (24) and (26) are independent of the smoothness of the input probability density function $f(u)$. In particular, in any identification point, where the input density does not vanish: (i) they converge to the system nonlinearity $\mu(u)$ in each point of its continuity (thus, the convergence holds in almost everywhere and the nonlinearity can be *e.g.* only piecewise-continuous), (ii) the convergence is not affected by either the structure of the overall system or the distribution of the external noise z_k, (iii) the rates of their convergence grow with growing smoothness of the nonlinearity $\mu(u)$. For instance, they converge with respective rates $\mathcal{O}\left(k^{-1/3}\right)$ and $\mathcal{O}\left(k^{-1/4}\right)$ in regions, where the nonlinearity is Lipschitz.

If $f(u)$ is Lipschitz, then the stochastic approximation algorithm (25) converges with the rate $\mathcal{O}\left(k^{\varepsilon-1/3}\right)$, that is, somewhat slower than the recursive quotient one (24), but faster than the order statistics algorithm (26).

5. CONCLUSION

The nonparametric algorithms are the only choice when the prior knowledge of the systems is small. Since they converge to the true system characteristics (are free of the approximation error), they can successfully support the traditional methods if the prior knowledge is partial or uncertain. Moreover, application of mixed parametric-nonparametric strategy allows to improve numerical properties of the algorithm, by proper selection of estimation points in te first stage.

REFERENCES

Bai, E.W. (1998). An optimal two-stage identification algorithm for Hammerstein-Wiener nonlinear systems. *Automatica* **34**(3), 333–338.

Boutayeb, M. and M. Darouach (1995). Recursive identification method for MISO Wiener-Hammerstein model. *IEEE Transactions on Automatic Control* **40**(2), 281–291.

Chen, H.F. (2004). Pathwise convergence of recursive identification algorithms for Hammerstein systems. *IEEE Transactions on Automatic Control* **49**(10), 1641–1649.

Chen, H.F. (2006). Recursive identification for Wiener model with discontinuous piece-wise linear function. *Automatic Control, IEEE Transactions on* **51**(3), 390–400.

Cripps, S.C. (2006). *RF Power Amplifiers for Wireless Communications*. Artech House.

D'Amato, A.M., B.O.S. Teixeira and D.S. Bernstein (2011). Semi-parametric identification of Wiener systems using a single harmonic input and retrospective cost optimisation. *Control Theory & Applications, IET* **5**(4), 594–605.

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

Gallman, P. (1975). An iterative method for the identification of nonlinear systems using a Uryson model. *Automatic Control, IEEE Transactions on* **20**(6), 771–775.

Giannakis, G.B. and E. Serpedin (2001). A bibliography on nonlinear system identification. *Signal Processing* **81**(3), 533–580.

Giri, F. and E. W. Bai (2010). *Block-Oriented Nonlinear System Identification*. Springer.

Greblicki, W. (2010). Nonparametric input density-free estimation of the nonlinearity in Wiener systems. *Information Theory, IEEE Transactions on* **56**(7), 3575–3580.

Greblicki, W. and G. Mzyk (2009). Semiparametric approach to Hammerstein system identification. In: *Proceedings of the 15th IFAC Symposium on System Identification, Saint-Malo, France*. pp. 1680–1685.

Greblicki, W. and M. Pawlak (1989). Recursive nonparametric identification of Hammerstein systems. *Journal of the Franklin Institute* **326**(4), 461–481.

Greblicki, W. and M. Pawlak (2008). *Nonparametric System Identification*. Cambridge University Press.

Haber, R. and L. Keviczky (1999). *Nonlinear System Parameter Identification*. Kluwer Academic Publishers. Dordrecht-Boston-London.

Hasiewicz, Z. and G. Mzyk (2004). Combined parametric-nonparametric identification of Hammerstein systems. *IEEE Transactions on Automatic Control* **48**(8), 1370–1376.

Hasiewicz, Z. and G. Mzyk (2009). Hammerstein system identification by non-parametric instrumental variables. *International Journal of Control* **82**(3), 440–455.

Hasiewicz, Z., M. Pawlak and P. Śliwiński (2005). Nonparametric identification of non-linearities in block-oriented complex systems by orthogonal wavelets with compact support. *IEEE Transactions on Circuits and Systems I: Regular Papers* **52**(1), 427–442.

Jiang, Y.P. and H.T. Fang (2009). Recursive identification of MIMO Wiener systems with general inputs. In: *Decision and Control, CDC/CCC 2009. Proceedings of the 48th IEEE Conference on*. IEEE. pp. 3626–3631.

Mzyk, G. (2007). A censored sample mean approach to nonparametric identification of nonlinearities in Wiener systems. *IEEE Transactions on Circuits and Systems – II: Express Briefs* **54**(10), 897–901.

Mzyk, G. (2009). Nonlinearity recovering in Hammerstein system from short measurement sequence. *IEEE Signal Processing Letters* **16**(9), 762–765.

Mzyk, G. (2010*a*). *Parametric versus nonparametric approach to Wiener systems identification*. Chap. 8. Vol. 404 of *Lecture Notes in Control and Information Sciences*. Springer.

Mzyk, G. (2010*b*). Wiener-Hammerstein system identification with non-gaussian input. In: *IFAC International Workshop on Adaptation and Learning in Control and Signal Processing, Antalya, Turkey*.

Mzyk, G. (2011). Instrumental variables for nonlinearity recovering in block-oriented systems driven by correlated signal. *submitted to the International Journal of Systems Science*.

Pawlak, M. and Z. Hasiewicz (1998). Nonlinear system identification by the Haar multiresolution analysis. *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on* **45**(9), 945–961.

Pawlak, M., Z. Hasiewicz and P. Wachel (2007). On nonparametric identification of Wiener systems. *Signal Processing, IEEE Transactions on* **55**(2), 482–492.

Powell, J.L., J.H. Stock and T.M. Stoker (1989). Semiparametric estimation of index coefficients. *Econometrica: Journal of the Econometric Society* pp. 1403–1430.

Śliwiński, P., J. Rozenblit, M.W. Marcellin and R. Klempous (2009). Wavelet amendment of polynomial models in Hammerstein systems identification. *Automatic Control, IEEE Transactions on* **54**(4), 820–825.

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

Wachel, P. (2010). Parametric-Nonparametric Identification of Wiener Systems (in Polish). PhD thesis. Wroclaw University of Technology, Poland.