Chapter 8
Parametric Versus Nonparametric Approach to Wiener Systems Identification

Grzegorz Mzyk

8.1 Introduction to Wiener System

The problem of nonlinear dynamic systems modelling by means of block-oriented models has been strongly elaborated for the last four decades, due to vast variety of applications. The concept of block-oriented models assumes that the real plant, as a whole, can be treated as a system of interconnected blocks, static nonlinearities (N) and linear dynamics (L), where the interaction signals cannot be measured. The most popular in this class are two-element cascade structures, *i.e.*, Hammerstein-type (N-L), Wiener-type (L-N), and sandwich-type (L-N-L) representations. Particularly, since in the Wiener system (Figure 8.1) the nonlinear block is preceded by the linear dynamics and the nonlinearity input is correlated, its identification is much more difficult in comparison with the Hammerstein system. However the Wiener model allows for better approximation of many real processes. Such difficulties in theoretical analysis forced the authors to consider special cases, and to take somehow restrictive assumptions on the input signal, impulse response of the linear dynamic block and the shape of the nonlinear characteristic. In particular, for Gaussian input the problem of Wiener system identification becomes much easier. Since the internal signal $\{x_k\}$ is then also Gaussian, the linear block can be simply identified by the cross-correlation approach, and the static characteristic can be recovered *e.g.* by the nonparametric inverse regression approach ([14]-[16]). Non-Gaussian random input is very rarely met in the literature. It is allowed *e.g.* in [38], but the algorithm presented there requires prior knowledge of the parametric representation of the linear subsystem. Most of recent methods for Wiener system identification assumes FIR linear dynamics, invertible nonlinearity, or require the use of specially designed input excitations ([2], [12]).

Grzegorz Mzyk,
Institute of Computer Engineering, Control and Robotics,
Wroclaw University of Technology, Poland, e-mail: grzegorz.mzyk@pwr.wroc.pl

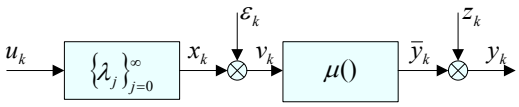

Fig. 8.1: Wiener system

In the chapter we compare and combine two kinds of methods, parametric ([1], [2], [5]-[12], [27]-[32], [39]-[43], [45], [47], [48]) and nonparametric ([14]-[26], [33]-[38], [44]). The method is called 'parametric' if both linear and nonlinear subsystems are described with the use of finite number of unknown parameters, *e.g.* when FIR linear dynamic model and polynomial characteristic with known orders are assumed. The popular parametric methods elaborated for Wiener system identification are not free of the approximation error and do not allow full decentralisation of the identification task of complex system. Moreover the theoretical analysis of identifiability, and convergence of parametric estimates remains relatively difficult. On the other hand, nonparametric approach offers simple algorithms, which are asymptotically free of approximation error, *i.e.* they converge to the true system characteristics. However, the purely nonparametric methods are not commonly exploited in practice for the following reasons: (i) they depend on various tuning parameters and functions; in particular, proper selection of kernel and the bandwidth parameter or orthonormal basis and the scale factor are critical for the obtained results, (ii) the prior knowledge of subsystems is completely neglected; the estimates are based on measurements only, and the resulting model may be not satisfactory when the number of measurements is small, and (iii) bulk number of estimates must be computed when the model complexity grows large.

In Section 8.2 we recollect the traditional parametric least-squares method for Wiener system identification and discuss its weak points. Next, in Section 8.3, we present several purely nonparametric methods, *i.e.*, correlation-based estimate of the linear dynamics, kernel estimate of the inverse regression, and a censored sample mean approach to nonlinearity recovering. Finally, selected parametric and nonparametric methods are combined and the properties of the proposed two-stage procedures are discussed in Section 8.4.

8.2 Nonlinear Least Squares Method

The discrete-time Wiener system, *i.e.* the linear dynamics with the impulse response $\{\lambda_j\}_{j=0}^{\infty}$, connected in the cascade with the static nonlinear block characterised by $\mu()$, is described by the following equation

$$y_k = \mu\left(\sum_{j=0}^{\infty}\lambda_j u_{k-j}\right) + z_k,$$

where u_k, y_k, and z_k are the input, output and the random disturbance, respectively. The goal of identification is to recover both elements, *i.e.* $\left\{\widehat{\lambda}_j\right\}_{j=0}^{\infty}$ and $\widehat{\mu}(x)$ for each $x \in R$, using the set of input-output measurements $\{(u_k, y_k)\}_{k=1}^{N}$. In the traditional (parametric) approach we also assume finite dimensional functionals, *e.g.* the ARMA-type dynamic block

$$\begin{aligned} x_k + a_1^* x_{k-1} + ... + a_r^* x_{k-r} &= b_0^* u_k + b_1^* u_{k-1} + ... + b_s^* u_{k-s}, \\ x_k &= \phi_k^T \theta^*, \\ \phi_k &= (-x_{k-1}, ..., -x_{k-r}, u_k, u_{k-1}, ..., u_{k-s})^T, \\ \theta^* &= (a_1^*, ..., a_r^*, b_0^*, b_1^*, ..., b_s^*)^T, \end{aligned} \tag{8.1}$$

and given formula $\mu(x, c^*) = \mu(x)$ including finite number of unknown true parameters $c^* = (c_1^*, c_2^*, ..., c_m^*)^T$. Respective Wiener model is thus represented by $r + (s+1) + m$ parameters, *i.e.*,

$$\begin{aligned} \overline{x}_k &\triangleq \overline{\phi}_k^T \theta, \text{ and } \overline{x}_k = 0 \text{ for } k \leqslant 0, \\ \text{where } \overline{\phi}_k^T &= (-\overline{x}_{k-1}, ..., -\overline{x}_{k-r}, u_k, u_{k-1}, ..., u_{k-s})^T, \\ \theta &= (a_1, ..., a_r, b_0, b_1, ..., b_s)^T, \\ \text{and } \overline{y}(x, c) &= \mu(x, c), \text{ where } c = (c_1, c_2, ..., c_m)^T. \end{aligned} \tag{8.2}$$

If the assumed model (8.2) agrees with the true system description (8.1), then the results of identification can be significantly improved in comparison with the nonparametric approach, particularly, if the number of measurements is small. On the other hand, the risk of bad parametrisation and existence of systematic approximation error must be taken into account together with the warranty of the parameter estimates converge. If x_k had been accessible for measurements then the true system parameters could have been estimated by the following minimisation

$$\widehat{\theta} = \arg\min_{\theta} \sum_{k=1}^{N} (x_k - \overline{x}_k(\theta))^2, \qquad \widehat{c} = \arg\min_{c} \sum_{k=1}^{N} (y_k - \overline{y}(x_k, c))^2. \tag{8.3}$$

Here we assume that only the input-output measurements (u_k, y_k) of the whole Wiener system are accessible, and the internal signal x_k is hidden. This observation leads to the following nonlinear least squares problem

$$\widehat{\theta}, \widehat{c} = \arg\min_{\theta, c} \sum_{k=1}^{N} [y_k - \overline{y}(\overline{x}_k(\theta), c)]^2, \tag{8.4}$$

which is usually very complicated numerically. Moreover, uniqueness of the solution in (8.4) cannot be guaranteed in general, as it depends on both input distribution, types of models, and values of parameters. Recent publications concerning applica-

tion of neural networks or soft computing methods to Wiener system identification problem do not include any theoretical analysis of convergence.

8.3 Nonparametric Identification Tools

The nonparametric approach to block-oriented system identification was introduced in eighties by Greblicki and Pawlak ([20], [37]). For the system of Hammerstein structure, the reverse connection described by the equation $y_k^H = \sum_{j=0}^{\infty} \gamma_j m(u_{k-j}) + z_k$, it was noticed that the input-output regression function is equivalent to nonlinear static characteristic, up to some scale and offset factors

$$R(u) \triangleq E\left(y_{k+l}^H | u_k = u\right) = \gamma_l m(u) + \sum_{j \neq l} \gamma_j E m(u_1).$$

Since then, two kinds of nonparametric methods have been examined – first, based on the kernel regression estimation, and second, employing the orthogonal series expansion of the nonlinearity. Also the cross-correlation method was proposed for estimation of the impulse response of linear dynamic block in Hammerstein system. The analogous ideas was also applied by Greblicki for a class of Wiener systems, with Gaussian input and locally invertible characteristics [16]. Respective algorithms are shortly reminded in Sections 8.3.1 and 8.3.2. In Section 8.3.3 we introduce and analyse a new kind of nonlinearity estimate in Wiener system, which works under the least possible prior knowledge, *i.e.* under non-Gaussian input, IIR linear dynamics and any continuous, but not necessary invertible, static characteristic.

8.3.1 Inverse Regression Approach

Assume that the input u_k and the noise ε_k are white Gaussian, mutually independent processes with finite variances $\sigma_u^2, \sigma_\varepsilon^2 < \infty$, the noise ε_k is zero-mean, *i.e.* $E\varepsilon_k = 0$, and the output measurement noise z_k is not present, *i.e.* $y_k = \overline{y}_k$. The nonparametric estimation of the inverted regression relies on the following lemma.

Lemma 8.1. *[21] If $\mu()$ is invertible then for any $y \in \mu(R)$ it holds that*

$$E\left(u_k | y_{k+p} = y\right) = \alpha_p \mu^{-1}(y) \tag{8.5}$$

where $\alpha_p = \lambda_p \frac{\sigma_u^2}{\sigma_v^2}$.

Since for any time lag p, the $\mu^{-1}(y)$ can be identified only up to some multiplicative constant α_p, let us denote, for convenience, $v(y) = \alpha_p \mu^{-1}(y)$. The nonparametric of $v(y)$ has the form

$$\widehat{v}(y) = \frac{\sum_{k=1}^{N} u_k K(\frac{y-y_{k+p}}{h(N)})}{\sum_{k=1}^{N} K(\frac{y-y_{k+p}}{h(N)})}, \tag{8.6}$$

where $K()$ and $h(N)$ is a kernel function and bandwidth parameter, respectively. The following theorem holds.

Theorem 8.1. *[21] If $\mu()$ is invertible, $K()$ is Lipschitz and such that $c_1 H(|y|) \leqslant K(y) \leqslant c_2 H(|y|)$ for some c_1 and c_2, where $H()$ is nonnegative and non-increasing function, defined of $[0,\infty)$, continuous and positive at $t = 0$, and such that $tH(t) \to 0$ as $t \to \infty$, then for $h(N) \to 0$ and $Nh^2(N) \to \infty$ as $N \to \infty$ it holds that*

$$\widehat{v}(y) \to v(y) \text{ in probability as } N \to \infty, \tag{8.7}$$

at every point y, in which the probability density $f(y)$ is positive and continuous.

The rate of convergence in (8.7) depends on the smoothness of the identified characteristic and is provided by the following lemma.

Lemma 8.2. *[21] Let us define $g(y) \triangleq v(y)f(y)$ and denote $v(y) = \frac{g(y)}{f(y)}$. If $\mu^{-1}()$, $f()$ and $g()$ have q bounded derivatives in a neighbourhood of y, then*

$$|\widehat{v}(y) - v(y)| = \mathscr{O}\left(N^{-\frac{1}{2}+\frac{1}{2q+2}}\right) \text{ in probability.}$$

e.g. $\mathscr{O}(N^{-1/4})$ for $q = 1$, $\mathscr{O}(N^{-1/3})$ for $q = 2$, and $\mathscr{O}(N^{-1/2})$ for q large.

In [16] and [19], the estimate (8.6) was generalised for the larger class of Wiener systems, admitting the "locally invertible" nonlinear static blocks and correlated excitation. The strongest limitation of the inverse regression approach is thus assumption about the Gaussianity of the input signal.

8.3.2 Cross-correlation Analysis

The nonparametric identification of the linear dynamic block is based on the following property.

Lemma 8.3. *[21] If $E\,|v_k \mu(v_k)| < \infty$ then*

$$E\left\{u_k y_{k+p}\right\} = \beta \lambda_p,$$

where $\beta = \frac{\sigma_u^2}{\sigma_v^2} E\left\{v_k \mu(v_k)\right\}$.

Since λ_p can be identified only up to some multiplicative constant β, let us denote, for convenience, $\kappa_p \triangleq \beta \lambda_p$, and consider its natural estimate of the form

$$\widehat{\kappa}_p = \frac{1}{N}\sum_{k=1}^{N} u_k y_{k+p}. \tag{8.8}$$

Theorem 8.2. *[21] If $\mu()$ is the Lipschitz function, then*

$$\lim_{N\to\infty} E\left(\widehat{\kappa}_p - \kappa_p\right)^2 = \mathscr{O}\left(\frac{1}{N}\right). \tag{8.9}$$

Consequently, when the stable IIR linear subsystem is modelled by the filter with the impulse response $\widehat{\kappa}_0, \widehat{\kappa}_1, ..., \widehat{\kappa}_{n(N)}$, then it is free of the asymptotic approximation error if $n(N) \to \infty$ and $n(N)/N \to 0$ as $N \to \infty$.

8.3.3 A Censored Sample Mean Approach

In this section we assume that the input $\{u_k\}$ is an i.i.d., bounded ($|u_k| < u_{\max}$; unknown $u_{\max} < \infty$), but not necessary Gaussian random process. There exists a probability density of the input, $\vartheta_u(u_k)$ say, which is a continuous and strictly positive function around the estimation point x, *i.e.*, $\vartheta_u(x) \geqslant \varepsilon > 0$. The unknown impulse response $\{\lambda_j\}_{j=0}^{\infty}$ of the linear IIR filter is exponentially upper bounded, that is

$$\left|\lambda_j\right| \leqslant c_1 \lambda^j, \text{ some unknown } 0 < c_1 < \infty, \tag{8.10}$$

where $0 < \lambda < 1$ is an *a priori* known constant. The nonlinearity $\mu(x)$ is an arbitrary function, continuous almost everywhere on $x \in (-u_{\max}, u_{\max})$ (in the sense of Lebesgue measure). The output noise $\{z_k\}$ is a zero-mean stationary and ergodic process, which is independent of the input $\{u_k\}$. For simplicity of presentation we also let $L \triangleq \sum_{j=0}^{\infty} \lambda_j = 1$ and $u_{\max} = \frac{1}{2}$. We note that the members of the family of Wiener systems composed by series connection of linear filters with the impulse responses $\{\overline{\lambda}_j\} = \{\frac{\lambda_j}{c_2}\}_{j=0}^{\infty}$ and the nonlinearities $\overline{\mu}(x) = \mu(c_2 x)$ are, for $c_2 \neq 0$, indistinguishable from the input-output point of view. In consequence, the characteristic $\mu()$ can be recovered in general only up to some domain scaling factor c_2, independently of the applied identification method. Observe that, in particular, for the FIR linear dynamics, the condition (8.10) is fulfilled for arbitrarily small constant $\lambda > 0$. Moreover, it holds that $|x_k| < x_{\max} < \infty$, where $x_{\max} \triangleq u_{\max} \sum_{j=0}^{\infty} \left|\lambda_j\right|$. Since $\sum_{j=0}^{\infty} \left|\lambda_j\right| \geqslant L$ and $L = 1$, thus the support of the random variables x_k, *i.e.* $(-x_{\max}, x_{\max})$, is generally wider than the estimation interval $x \in (-u_{\max}, u_{\max})$. We introduce and analyse the nonparametric estimate of the part of characteristic $\mu(x)$, for $x \in (-u_{\max}, u_{\max})$, and next we expand the obtained results for $x \in (-x_{\max}, x_{\max})$, when the parametric knowledge of $\mu()$ is provided.

Let x be a chosen estimation point of $\mu(\cdot)$. For a given x let us define a "weighted distance" between the measurements $u_k, u_{k-1}, u_{k-2}, ..., u_1$ and x as

$$\delta_k(x) \triangleq \sum_{j=0}^{k-1} \left|u_{k-j} - x\right| \lambda^j = |u_k - x| \lambda^0 + |u_{k-1} - x| \lambda^1 + ... + |u_1 - x| \lambda^{k-1}, \quad (8.11)$$

i.e. $\delta_1(x) = |u_1 - x|$, $\delta_2(x) = |u_2 - x| + |u_1 - x| \lambda$, $\delta_3(x) = |u_3 - x| + |u_2 - x| \lambda + |u_1 - x| \lambda^2$, *etc.*, which can be computed recursively as follows

$$\delta_k(x) = \lambda \delta_{k-1}(x) + |u_k - x|. \quad (8.12)$$

Under above assumptions we obtain

$$\begin{aligned} |x_k - x| &= \left| \sum_{j=0}^{\infty} \lambda_j u_{k-j} - \sum_{j=0}^{\infty} \lambda_j x \right| = \left| \sum_{j=0}^{\infty} \lambda_j \left(u_{k-j} - x\right) \right| = \\ &= \left| \sum_{j=0}^{k-1} \lambda_j \left(u_{k-j} - x\right) + \sum_{j=k}^{\infty} \lambda_j \left(u_{k-j} - x\right) \right| \leqslant \\ &\leqslant \sum_{j=0}^{k-1} \left|\lambda_j\right| \left|u_{k-j} - x\right| + 2u_{\max} \sum_{j=k}^{\infty} \left|\lambda_j\right| \leqslant \delta_k(x) + \frac{\lambda^k}{1-\lambda} \triangleq \Delta_k(x). \end{aligned} \quad (8.13)$$

Observe that if in turn

$$\Delta_k(x) \leqslant h(N), \quad (8.14)$$

then the true (but unknown) interaction input x_k is located close to x, provided that $h(N)$ (further, a calibration parameter) is small. The distance given in (8.13) may be easily computed as the point x and the data $u_k, u_{k-1}, u_{k-2}, ..., u_1$ are each time at ones disposal. In turn, the condition (8.14) selects k's for which the input sequences $\{u_k, u_{k-1}, u_{k-2}, ..., u_1\}$ are such that the true nonlinearity inputs $\{x_k\}$ surely belong to the neighbourhood of the estimation point x with the radius $h(N)$. Let us also notice that asymptotically, as $k \to \infty$, it holds that

$$\delta_k(x) = \Delta_k(x), \quad (8.15)$$

with probability 1.

Proposition 8.1. *If, for each $j = 0, 1, ..., \infty$ and some $d > 0$, it holds that*

$$\left|u_{k-j} - x\right| \leqslant \frac{d}{\lambda^j}, \quad (8.16)$$

then

$$|x_k - x| \leqslant d \log_\lambda d + d \frac{\lambda}{1-\lambda}. \quad (8.17)$$

Proof. The condition (8.16) is fulfilled with probability 1 for each $j > j_0$, where $j_0 = \lfloor \log_\lambda d \rfloor$ is the solution of the following inequality

$$\frac{d}{\lambda^j} \geqslant 2u_{\max} = 1.$$

Analogously as in (8.13), we obtain

$$|x_k - x| \leqslant \sum_{j=0}^{j_0} \lambda^j \frac{d}{\lambda^j} + \frac{\lambda^{j_0+1}}{1-\lambda},$$

which yields (8.17).

We propose the following nonparametric kernel-like estimate of the nonlinear characteristic $\mu()$ at the given point x, exploiting the distance $\delta_k(x)$ between x_k and x, and having the form

$$\widehat{\mu}_N(x) = \frac{\sum_{k=1}^N y_k \cdot K\left(\frac{\delta_k(x)}{h(N)}\right)}{\sum_{k=1}^N K\left(\frac{\delta_k(x)}{h(N)}\right)}, \tag{8.18}$$

where $K()$ is the window kernel function of the form

$$K(v) = \begin{cases} 1, \text{ as } |v| \leqslant 1 \\ 0, \text{ elsewhere} \end{cases}. \tag{8.19}$$

Since the estimate (8.18) is of the ratio form we treat the case $0/0$ as 0.

Theorem 8.3. *If $h(N) = d(N)\log_\lambda d(N)$, where $d(N) = N^{-\gamma(N)}$, and $\gamma(N) = \left(\log_{1/\lambda} N\right)^{-w}$, then for each $w \in \left(\frac{1}{2}, 1\right)$ the estimate (8.18) is consistent in the mean square sense, i.e., it holds that*

$$\lim_{N\to\infty} E\left(\widehat{\mu}_N(x) - \mu(x)\right)^2 = 0. \tag{8.20}$$

Proof. Let us denote the probability of selection as $p(N) \triangleq P\left(\Delta_k(x) \leqslant h(N)\right)$. To prove (8.20) it suffices to show that (see (19) and (22) in [33])

$$h(N) \to 0, \tag{8.21}$$

$$Np(N) \to \infty, \tag{8.22}$$

as $N \to \infty$. They assure vanishing of the bias and variance of $\widehat{\mu}_N(x)$, respectively. Since under assumptions of Theorem 8.3

$$d(N) \to 0 \Rightarrow h(N) \to 0, \tag{8.23}$$

in view of (8.17), the bias-condition (8.21) is obvious. For the variance-condition (8.22) we have

$$p(N) \geqslant P\left\{\bigcap_{j=0}^{\min(k,j_0)} \left(\left|u_{k-j} - x\right| < \frac{d(N)}{\lambda^j}\right)\right\} \geqslant$$

$$\geqslant P\left\{\bigcap_{j=0}^{\min(k,j_0)}\left(\left|u_{k-j}-x\right|<\frac{d(N)}{\lambda^j}\right)\right\}=\prod_{j=0}^{j_0}P\left(\left|u_{k-j}-x\right|<\frac{d(N)}{\lambda^j}\right)\geqslant$$

$$\geqslant \varepsilon\frac{d(N)}{\lambda^0}\cdot\varepsilon\frac{d(N)}{\lambda^1}\cdot\ldots\cdot\varepsilon\frac{d(N)}{\lambda^{j_0}}=\frac{(\varepsilon d(N))^{j_0+1}}{\lambda^{\frac{j_0(j_0+1)}{2}}}=$$

$$=\left(\frac{\varepsilon d(N)}{\lambda^{\frac{j_0}{2}}}\right)^{j_0+1}=\left(\varepsilon\sqrt{d(N)}\right)^{j_0+1}=\varepsilon\cdot d(N)^{\frac{1}{2}\log_\lambda d(N)+\log_\lambda\varepsilon+\frac{1}{2}}. \quad (8.24)$$

By inserting $d(N)=N^{-\gamma(N)}=(1/\lambda)^{-\gamma(N)\log_{1/\lambda}N}$ to (8.24) we obtain

$$N\cdot p(N)=\varepsilon\cdot N^{1-\gamma(N)\left(\frac{1}{2}\gamma(N)\log_{1/\lambda}N+\log_\lambda\varepsilon+\frac{1}{2}\right)}. \quad (8.25)$$

For $\gamma(N)=\left(\log_{1/\lambda}N\right)^{-w}$ and $w\in\left(\frac{1}{2},1\right)$ from (8.25) we simply conclude (8.22) and consequently (8.20).

To establish the asymptotic rate of convergence we additionally assume that the nonlinear characteristic $\mu(x)$ is a Lipschitz function, *i.e.*, it exists a positive constant $l<\infty$, such that for each $x_a,x_b\in R$ it holds that $|\mu(x_a)-\mu(x_b)|\leqslant l\,|x_a-x_b|$.

For a window kernel (8.19) we can rewrite (8.18) as $\widehat{\mu}_N(x)=\frac{1}{S_0}\sum_{i=1}^{S_0}y_{[i]}$, where $[i]$'s are indexes, for which $K\left(\frac{\delta_{[i]}(x)}{h(N)}\right)=1$, and S_0 is a random number of selected output measurements. For each $y_{[i]}$, $i=1,2,\ldots,S_0$, respective $x_{[i]}$ is such that $\left|x_{[i]}-x\right|\leqslant h(N)$ and consequently

$$\left|\mu(x_{[i]})-\mu(x)\right|\leqslant lh(N),$$

which for $Ez_k=0$ leads to

$$|\text{bias}\widehat{\mu}_N(x)|=\left|Ey_{[i]}-\mu(x)\right|=\left|E\mu(x_{[i]})-\mu(x)\right|\leqslant lh(N),$$
$$\text{bias}^2\widehat{\mu}_N(x)=\mathscr{O}\left(h^2(N)\right). \quad (8.26)$$

For the variance we have

$$\text{var}\widehat{\mu}_N(x)=\sum_{n=0}^{N}P(S_0=n)\cdot\text{var}\left(\widehat{\mu}_N(x)|S_0=n\right)=\sum_{n=1}^{N}P(S_0=n)\cdot\text{var}\left(\frac{1}{n}\sum_{i=1}^{n}y_{[i]}\right).$$

Since, under strong law of large numbers and Chebychev inequality, it holds that $\lim_{N\to\infty}P\left(S_0>\alpha ES_0\right)=1$ for each $0<\alpha<1$ (see [33]), we obtain asymptotically

$$\text{var}\widehat{\mu}_N(x)=\sum_{n>\alpha ES_0}P(S_0=n)\cdot\text{var}\left(\frac{1}{n}\sum_{i=1}^{n}y_{[i]}\right) \quad (8.27)$$

with probability 1. Taking into account that $y_{[i]} = \overline{y}_{[i]} + z_{[i]}$, where $\overline{y}_{[i]}$ and $z_{[i]}$ are independent random variables we obtain

$$\mathrm{var}\left(\frac{1}{n}\sum_{i=1}^{n} y_{[i]}\right) = \mathrm{var}\left(\frac{1}{n}\sum_{i=1}^{n} \overline{y}_{[i]}\right) + \mathrm{var}\left(\frac{1}{n}\sum_{i=1}^{n} z_{[i]}\right). \tag{8.28}$$

Since the process $\left\{z_{[i]}\right\}$ is ergodic, under strong law of large numbers, it holds that

$$\mathrm{var}\left(\frac{1}{n}\sum_{i=1}^{n} z_{[i]}\right) = \mathscr{O}\left(\frac{1}{Np(N)}\right) = \mathscr{O}\left(\frac{1}{N}\right). \tag{8.29}$$

The process $\{\overline{y}_{[i]}\}$ is in general not ergodic, but in consequence of (8.14) it has compact support $[\mu(x) - lh(N), \mu(x) + lh(N)]$ and the following inequality holds

$$\mathrm{var}\left(\frac{1}{n}\sum_{i=1}^{n} \overline{y}_{[i]}\right) \leqslant \mathrm{var}\overline{y}_{[i]} \leqslant (2lh(N))^2. \tag{8.30}$$

From (8.27), (8.28), (8.29) and (8.30) we conclude that

$$\mathrm{var}\widehat{\mu}_N(x) = \mathscr{O}(h^2(N)), \tag{8.31}$$

which in view of (8.26) leads to

$$|\widehat{\mu}_N(x) - \mu(x)| = \mathscr{O}(h^2(N)) \tag{8.32}$$

in the mean square sense. A relatively slow rate of convergence, guaranteed in a general case, for $h(N)$ as in Theorem 8.3, is a consequence of small amount of *a priori* information. We emphasise that for, *e.g.*, often met in applications piecewise constant functions $\mu(x)$, it exists $\overline{N} < \infty$, such that $\mathrm{bias}^2\widehat{\mu}_N(x) = 0$ and $\mathrm{var}\left(\frac{1}{n}\sum_{i=1}^{n}\overline{y}_{[i]}\right) = 0$ for $N > \overline{N}$, and consequently $|\widehat{\mu}_N(x) - \mu(x)| = \mathscr{O}(\frac{1}{N})$ as $N \to \infty$ (see (8.29)).

8.4 Combined Parametric-Nonparametric Approach

The idea of the combined parametric-nonparametric approach to system identification was introduced by Hasiewicz and Mzyk in [24], and continued in [25], [33], [35], and [38]. The algorithms decompose the complex system identification task on independent identification problems for each components. The decomposition is based on the estimation of interaction inputs x_k. Next, using the resulting pairs $(u_k, \widehat{x}_k)$ and $(\widehat{x}_k, y_k)$, both linear dynamic and static nonlinear subsystems are identified separately. In the contrary to Hammerstein system, where $x_k^H = m(u_k)$ may be estimated directly by any regression estimation method, for Wiener system the

situation is more complicated as $x_k = \sum_{j=0}^{\infty} \lambda_j u_{k-j}$, and the impulse response of the linear dynamics must be estimated first, to provide indirect estimates of x_k.

8.4.1 Kernel Method with the Correlation-Based Internal Signal Estimation

Here we assume that the input u_k is white and Gaussian, the nonlinear characteristic $\mu()$ is bounded by polynomial of any finite order, $cov(u_1, y_1) \neq 0$, and the linear dynamics is FIR with known order s, *i.e.* $x_k = \sum_{j=0}^{s} \lambda_j u_{k-j}$.

Observe that $E\{y_k | x_k = x\} = \mu(x)$. Since the internal signal x_k cannot be measured, the following kernel regression estimate is proposed (see [44])

$$\widehat{\mu}(x) = \frac{\sum_{k=1}^{N} y_k K\left(\frac{x-\widehat{x}_k}{h(N)}\right)}{\sum_{k=1}^{N} K\left(\frac{x-\widehat{x}_k}{h(N)}\right)}, \tag{8.33}$$

where $\widehat{x}_k$ is indirect estimate of βx_k (*i.e.* scaled x_k)

$$\widehat{x}_k = \sum_{j=0}^{s} \widehat{\kappa}_j u_{k-j}$$

based on the input-output sample correlation (see (8.8)). The following theorem holds.

Theorem 8.4. *[44] If $K()$ is Lipschitz then $\widehat{\mu}(x) \to \mu(x/\beta)$ in probability as $N \to \infty$. Moreover, if both $\mu()$ and $K()$ are twice differentiable, then it holds that*

$$|\widehat{\mu}(x) - \mu(x/\beta)| = \mathscr{O}\left(N^{-\frac{2}{5}+\varepsilon}\right) \tag{8.34}$$

for any small $\varepsilon > 0$, provided that $h(N) \sim N^{-1/5}$.

In practise, due to assumed Gaussianity of excitations, the algorithm (8.34) is rather recommended for the tasks, in which the input process can be freely generated.

8.4.2 Identification of IIR Wiener Systems with Non-Gaussian Input

The presented kernel-type algorithm (8.18) is applied in this section to support estimation of parameters, when our prior knowledge about the system is large, and in particular, the parametric model of the characteristic is known.

Assume that we are given the class $\mu(x,c)$, such that $\mu(x) \subset \mu(x,c)$, where $c = (c_1, c_2, ..., c_m)^T$, *i.e.* for the vector of true parameters $c^* = (c_1^*, c_2^*, ..., c_m^*)^T$ it holds that $\mu(x,c^*) = \mu(x)$. Let moreover the function $\mu(x,c)$ be by assumption differentiable with respect to c, and the gradient $\bigtriangledown_c \mu(x,c)$ be bounded in some convex neighbourhood of c^* for each x. We assume that c^* is identifiable, *i.e.*, there exists such a sequence $x^{(1)}, x^{(2)}, ..., x^{(N_0)}$ of estimation points, that

$$\mu(x^{(i)},c) = \mu(x^{(i)}),\ i = 1,2,...,N_0 \Longrightarrow c = c^*.$$

The proposed estimate has two steps.

Step 1. For the sequence $x^{(1)}, x^{(2)}, ..., x^{(N_0)}$ compute N_0 pairs

$$\left\{ \left(x^{(i)}, \widehat{\mu}_N(x^{(i)}) \right) \right\}_{i=1}^{N_0},$$

using the estimate (8.18).

Step 2. Perform the minimisation of the cost-function

$$Q_{N_0,N}(c) = \sum_{i=1}^{N_0} \left(\widehat{\mu}_N(x^{(i)}) - \mu(x^{(i)},c) \right)^2,$$

with respect to the variable vector c, and take

$$\widehat{c}_{N_0,N} = \arg\min_c Q_{N_0,N}(c) \tag{8.35}$$

as the estimate of c^*.

Theorem 8.5. *Since in* Step 1 *(nonparametric) for the estimate (8.18) it holds that* $\widehat{\mu}_N(x^{(i)}) \to \mu(x^{(i)})$ *in probability as* $N \to \infty$ *for each* $i = 1,2,...,N_0$*, thus*

$$\widehat{c}_{N_0,N} \to c^*$$

in probability, as $N \to \infty$*.*

Proof. The proof is analogous to that of Theorem 1 in [25].

8.4.3 Recent Ideas

The interesting new attempt to the impulse response estimation of the linear block in Wiener system is presented in [44]. It is assumed that the input probability density $f(u)$ has compact support, both $\mu()$ and $f()$ have continuous derivatives, and the linear dynamics is FIR with known order s. We emphasise that similarly as in the correlation-based algorithm (see Section 8.3.2) the characteristic $\mu()$ need not to be invertible and moreover the input density is not assumed to be Gaussian. The idea follows from the chain rule. Introducing the vectors

$$\underline{u}_k = (u_k, u_{k-1}, ..., u_{k-s})^T \text{ and } \underline{\lambda} = (\lambda_0, \lambda_1, ..., \lambda_s)^T$$

one can describe the Wiener system by the following formula

$$y_k = F(\underline{u}_k) + z_k, \text{ where } F(\underline{u}_k) = \mu(\underline{\lambda}^T \underline{u}_k).$$

Let $D_F(\underline{u})$ be the gradient of $F()$. It holds that $D_F(\underline{u}_k) = \mu'(\underline{\lambda}^T \underline{u}_k)\underline{\lambda}$ and consequently

$$E\{D_F(\underline{u}_k)\} = c_0\underline{\lambda}, \text{ where } c_0 = E\left\{\mu'(\underline{\lambda}^T \underline{u}_k)\right\}. \tag{8.36}$$

It leads to the idea of estimation of the scaled vector $\underline{\lambda}$, including the true elements of the impulse response, by the gradient averaging. Since for a given $\underline{u}_k$, $\mu'(\underline{\lambda}^T \underline{u}_k)$ is unknown, $D_F(\underline{u}_k)$ cannot be computed directly. Introducing $f_{\underline{u}}()$ – the joint probability density of $\underline{u}_k$, the property (8.36) can be transformed to the more applicable form ([44])

$$E\left\{y_k D_f(\underline{u}_k)\right\} = c_1\underline{\lambda}, \text{ where } c_1 = \frac{1}{2}E\left\{f(u_k)\mu'(\underline{\lambda}^T \underline{u}_k)\right\},$$

and $D_f(u)$ is a gradient of $f_{\underline{u}}()$. Since for white u_k we have $f_{\underline{u}}(\underline{u}_k) = \prod_{j=0}^{s} f(u_{k-j})$, it leads to the following scalar estimates of the impulse response

$$\widehat{\lambda}_j = \frac{1}{N}\sum_{k=1}^{N} y_k d_{f,j}(\underline{u}_k), \text{ where } d_{f,j}(u_k) = D_f(\underline{u}_k)[j] = f'(u_{k-j}) \prod_{i=0, i\neq j}^{s} f(u_{k-i}). \tag{8.37}$$

If the input probability density function $f(u)$ is unknown, it can be simply estimated, *e.g.*, by the kernel method. The open question is generalisation of the approach for IIR linear subsystems and correlated input cases.

8.5 Conclusion

The principal question in Wiener system identification problem is selection of adequate method. The scope of application of each estimates is limited by specific set of associated assumptions. Most of them requires *a priori* known parametric type of model, Gaussian input, FIR dynamics or invertible characteristic. In fact, the authors address particular cases, and the problems they solve are quite different (see references below). Since the general Wiener system identification problem includes many difficult aspects, existence of one universal algorithm cannot be expected. In the light of this, the nonparametric approach seems to be good tool, which allows for combining selected methods, depending on specificity of the particular task. Moreover, pure nonparametric estimates are the only possible choice, when the prior knowledge of the system is poor.

References

1. Bai, E. W. (2002) "A blind approach to Hammerstein–Wiener model identification", *Automatica*, vol. 38, pp. 969-979
2. Bai, E. W. (2003) "Frequency domain identification of Wiener models", *Automatica*, vol. 39, pp. 1521-1530
3. Bai, E. W. and J. Reyland (2008) "Towards identification of Wiener systems with the least amount of *a priori* information on the nonlinearity", *Automatica*, vol. 44, pp. 910-919
4. Bai, E. W. and J. Reyland (2009) "Towards identification of Wiener systems with the least amount of *a priori* information: IIR cases", *Automatica*, vol. 45, No. 4, pp. 956-964
5. Bershad, N. J., S. Bouchired, F. Castanie (2000) "Stochastic analysis of adaptive gradient identification of Wiener–Hammerstein systems for Gaussian inputs", *IEEE Transactions on Signal Processing*, vol. 48, No. 2, pp. 557-560
6. Bershad, N. J., P. Celka, and J. M. Vesin (2000) "Analysis of stochastic gradient tracking of time-varying polynomial Wiener systems", *IEEE Transactions on Signal Processing*, vol. 48, No. 6, pp. 1676-1686
7. Billings, S.A. and S.Y. Fakhouri (1977) "Identification of nonlinear systems using the Wiener model", *Automatica*, vol. 13, No. 17, pp. 502-504
8. Billings, S.A. and S.Y. Fakhouri (1982) "Identification of systems containing linear dynamic and static nonlinear elements", *Automatica*, vol. 18, No. 1, pp. 15-26
9. Celka, P., N.J. Bershad, and J.M. Vesin (2000) "Fluctuation analysis of stochastic gradient identification of polynomial Wiener systems", *IEEE Transactions on Signal Processing*, vol. 48, No. 6, pp. 1820-1825
10. Celka, P., N.J. Bershad, and J.M. Vesin (2001) "Stochastic gradient identification of polynomial Wiener systems: analysis and application", *IEEE Transactions on Signal Processing*, vol. 49, No. 2, pp. 301-313
11. Chen, H. F. (2006) "Recursive identification for Wiener model with discontinuous piece-wise linear function", *IEEE Transactions on Automatic Control*, vol. 51, No. 3, pp. 390-400
12. Giri, F., Y. Rochdi, and F. Z. Chaoui (2009) "An analytic geometry approach to Wiener system frequency identification", *IEEE Transactions on Automatic Control*, vol. 54, No. 4, pp. 683-696
13. Giannakis, G. B., and E. Serpedin (2001) "A bibliography on nonlinear system identification", *Signal Processing*, vol. 81, pp. 533-580
14. Greblicki, W. (1992) "Nonparametric identification of Wiener systems", *IEEE Transactions on Information Theory*, vol. 38, pp. 1487-1493
15. Greblicki, W. (1994) "Nonparametric identification of Wiener systems by orthogonal series", *IEEE Transactions on Automatic Control*, vol. 39, No. 10, pp. 2077-2086
16. Greblicki, W. (1997) "Nonparametric approach to Wiener system identification", *IEEE Transactions on Circuits and Systems – I: Fundamental Theory and Applications* , vol. 44, No. 6, pp. 538-545
17. Greblicki, W. (1998) "Continuous-time Wiener system identification", *IEEE Transactions on Automatic Control*, vol. 43, No. 10, pp. 1488-1493
18. Greblicki, W. (2001) "Recursive identification of Wiener systems", *International Journal of Applied Mathematics and Computer Science*, vol. 11, No. 4, pp. 977–991
19. Greblicki, W. (2004) "Nonlinearity recovering in Wiener system driven with correlated signal", *IEEE Transactions on Automatic Control*, vol. 49, No. 10, pp. 1805-1810
20. Greblicki, W. and M. Pawlak Identification of discrete (1986) Hammerstein systems using kernel regression estimates, *IEEE Transactions on Automatic Control*, vol. 31, pp. 74–77
21. Greblicki, W., M. Pawlak (2008) *Nonparametric System Identification*, Cambridge University Press
22. Greblicki, W. and G. Mzyk (2009) "Semiparametric approach to Hammerstein system identification", *Proceedings of the 15th IFAC Symposium on System Identification*, pp. 1680-1685, Saint-Malo, France, July 6-8

23. Hasiewicz, Z. (1987) "Identification of a linear system observed through zero-memory nonlinearity", *International Journal of Systems Science*, vol. 18, pp. 1595-1607
24. Hasiewicz, Z., G. Mzyk (2004) "Combined parametric-nonparametric identification of Hammerstein systems", *IEEE Transactions on Automatic Control*, vol. 49, pp. 1370-1376
25. Hasiewicz, Z., G. Mzyk (2009) "Hammerstein system identification by non-parametric instrumental variables", *International Journal of Control,* vol. 82, No. 3, pp. 440-455
26. Hasiewicz, Z., P. Śliwiński, G. Mzyk (2008) "Nonlinear system identification under various prior knowledge", *Proceedings of the 17th World Congress The IFAC*, pp. 7849-7858, Seoul, Korea
27. Hagenblad, A., L. Ljung, A. Wills (2008) "Maximum likelihood identification of Wiener models", *Automatica*, vol. 44, pp. 2697-2705
28. Hu, X. L. and H. F. Chen (2005) "Strong consistence of recursive identification of Wiener systems", *Automatica*, vol. 41, pp. 1905-1916
29. Hughes, M. C. and D. T. Westwick (2005) "Identification of IIR Wiener systems with spline nonlinearities that have variable knots", *IEEE Transactions on Automatic Control*, vol. 50, No. 10, pp. 1617-1622
30. Hunter, I. W. and M. J. Korenberg (1986) "The identification of nonlinear biological systems: Wiener and Hammerstein cascade models", *Biological Cybernetics*, vol. 55, pp. 135-144
31. Korenberg, M. J. and I. W. Hunter (1986) "The identification of nonlinear biological systems: LNL cascade models", *Biological Cybernetics*, vol. 55, pp. 125-134
32. Lacy, S. L. and D. S. Bernstein (2003) "Identification of FIR Wiener systems with unknown, non-invertible, polynomial non-linearities", *International Journal of Control*, vol. 76, No. 15, pp. 1500-1507
33. 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*, vol. 54, No. 10, pp. 897-901
34. Mzyk, G. (2007) "Generalized kernel regression estimate for the identification of Hammerstein systems", *International Journal of Applied Mathematics and Computer Science*, vol. 17, No. 2, pp. 101-109
35. Mzyk, G. (2009) "Nonlinearity recovering in Hammerstein system from short measurement sequence", *IEEE Signal Processing Letters*, vol. 16, No. 9, pp. 762-765
36. Mzyk, G. (2010) "Kernel-type identification of IIR Wiener systems with non-gaussian input", submitted to the *IEEE Transactions on Signal Processing*.
37. Nadaraya, E. A. (1964) On estimating regression, *Theory of Probability and its Applications*, vol. 9, pp. 141-142
38. Pawlak, M., Z. Hasiewicz, P. Wachel (2007) "On nonparametric identification of Wiener systems", *IEEE Transactions on Signal Processing*, vol. 55, No. 2, pp. 482-492
39. Pupeikis, R. (2005) "On the identification of Wiener systems having saturation-like functions with positive slopes", *Informatica*, vol. 16, No. 1, pp. 131-144
40. Rafajłowicz, E. (1994) "Non-parametric identification with errors in independent variables", *International Journal of Systems Science*, vol. 25, No. 9, pp. 1513-1520
41. Vanbeylen, L., R. Pintelon, J. Schoukens (2009) "Blind maximum-likelihood identification of Wiener systems", *IEEE Transactions on Signal Processing*, vol. 57, No. 8, pp. 3017-3029
42. Vandersteen, G., J. Schoukens (1999) "Measurement and identification of nonlinear systems consisting of linear dynamic blocks and one static nonlinearity", *IEEE Transactions on Automatic Control*, vol. 44, No. 6, pp. 1266-1271
43. Vörös, J. (2007) "Parameter identification of Wiener systems with multisegment piecewise-linear nonlinearities", *Systems and Control Letters*, vol. 56, pp. 99-105
44. Wachel, P. (2008) *Parametric-Nonparametric Identification of Wiener Systems*, PhD Thesis (in Polish), `http://diuna.iiar.pwr.wroc.pl/wachel/rozprawa.pdf`, Wrocław University of Technology, Poland
45. Westwick, D., M. Verhaegen (1996) "Identifying MIMO Wiener systems using subspace model identification methods", *Signal Processing*, vol. 52, pp. 235-258
46. Wiener, N. (1958) *Nonlinear Problems in Random Theory*, Wiley, New York

47. Wigren, T. (1994) “Convergence analysis of recursive identification algorithms based on the nonlinear Wiener model”, *IEEE Transactions on Automatic Control*, vol. 39, pp. 2191-2206
48. Zhao, Y., L. Y. Wang, G. G. Yin, J. F. Zhang (2007) “Identification of Wiener systems with binary-valued output observations”, *Automatica*, vol. 43, pp. 1752-1765