2.2.3 Orthogonal expansion

Second kind of nonparametric regression methods are based on orthogonal series expansion. Let us define

$$g(u) \triangleq R(u)f(u) \tag{2.31}$$

and rewrite the regression $R(u)$ as

$$R(u) = \frac{g(u)}{f(u)}. \tag{2.32}$$

Let $\{\varphi_i(u)\}_{i=0}^{\infty}$ be the complete set of orthonormal functions in the input domain. If $g()$ and $f()$ are square integrable, the orthogonal series representations of $g(u)$ and $f(u)$ in the basis $\{\varphi_i(u)\}_{i=0}^{\infty}$ has the form

$$g(u) = \sum_{i=0}^{\infty} \alpha_i \varphi_i(u), \qquad f(u) = \sum_{i=0}^{\infty} \beta_i \varphi_i(u), \tag{2.33}$$

with the infinite number of coefficients ($i = 0, 1, ..., \infty$)

$$\alpha_i = E y_k \varphi_i(u_k), \qquad \beta_i = E\varphi_i(u_k). \tag{2.34}$$

Since the natural estimates of α_i's and β_i's are

$$\widehat{\alpha}_{i,N} = \frac{1}{N}\sum_{k=1}^{N} y_k \varphi_i(u_k), \qquad \widehat{\beta}_{i,N} = \frac{1}{N}\sum_{k=1}^{N} \varphi_i(u_k), \tag{2.35}$$

we obtain the following ratio estimate of $\mu(u)$

$$\widehat{\mu}_N(u) = \widehat{R}_N(u) = \frac{\sum_{i=0}^{q(N)} \widehat{\alpha}_{i,N}\varphi_i(u)}{\sum_{i=0}^{q(N)} \widehat{\beta}_{i,N}\varphi_i(u)}, \tag{2.36}$$

where $q(N)$ is some cut-off (approximation) level [29]. The consistency conditions, with respect to $q(N)$, are given in Remark below. Optimal choice of $q(N)$ with respect to the rate of convergence is considered, e.g., in [29] and [54].

Remark 2.2 *[39] To assure vanishing of the approximation error, the scale $q(N)$ must behave so that $\lim_{N\to\infty} q(N) = \infty$. For the convergence of $\widehat{\mu}_N(u)$ to $\mu(u)$, the rate of $q(N)$-increasing must be appropriately slow, e.g., $\lim_{N\to\infty} q^2(N)/N = 0$ for trigonometric or Legendre series, $\lim_{N\to\infty} q^6(N)/N = 0$ for Laguerre series, $\lim_{N\to\infty} q^{5/3}(N)/N = 0$ for Hermite series.*

2.3 Block-oriented systems

Conception of the block-oriented models (interconnections of static nonlinearities and linear dynamics) in system identification has been introduced in 1980's by Billings and Fakhouri ([10]), as an alternative for Volterra series expansions. It was commonly accepted because of satisfactory approximation capabilities of various real processes and relatively small model complexity. The decentralized approach to the identification of block-oriented complex systems seems to be most natural and desirable, as such an approach corresponds directly to the own nature of systems composed of individual elements distinguished in the structure ([3], [6], [14], [75]) and tries to treat the components 'locally' as independent, autonomous objects. The Hammerstein system, built of a static non-linearity and a linear dynamics connected in a cascade, is the simplest structure in the class and hence for the most part considered in the system identification literature. Unfortunately, the popular, parametric, methods elaborated for Hammerstein system identification do not allow full decentralization of the system identification task, i.e. independent identification of a static nonlinearity and a linear dynamics – first of all, because of inaccessibility for measurements of the inner interconnection signal. They assume that the description of system components, i.e. of a static nonlinearity and a linear dynamics is known up to the parameters (a polynomial model along with a FIR dynamics representation are usually used) and these parameters are "glued" when using standard input-output data of the overall system for identification purposes (e.g. [10], [118], [96], [95]). On the other hand, in a nonparametric setting (the second class of existing identification methods, see, e.g. [37], [38], [80]) no preliminary assumptions concerning the structure of subsystems are used and only the data decide about the obtained characteristics of the system components but then any possible a priori knowledge about the true description of subsystems is not exploited, i.e. inevitably lost.

Below, we present the most popular structures of block-oriented systems [27], i.e., Hammerstein system, Wiener system, Wiener-Hammerstein (sandwich) system, additive NARMAX system, and finally, the system with arbitrary connection structure. In the whole book we assume that the nonlinear characteristics are Borel measurable and the linear blocks are asymptotically stable. Formal description of all systems is given. We also focus on specifics of identification problem of each structure, and give some examples of applications in practice.

2.3.1 Hammerstein system

The Hammerstein system, shown in Fig. 2.6, consists of a static nonlinear block with the unknown characteristic $\mu()$, followed by the linear dynamic object with the unknown impulse response $\{\gamma_i\}_{i=0}^{\infty}$. The internal signal $w_k = \mu(u_k)$ is not accessible, and the output is observed in the presence of the random noise z_k.

$$y_k = \sum_{i=0}^{\infty} \gamma_i \mu(u_{k-i}) + z_k \tag{2.37}$$

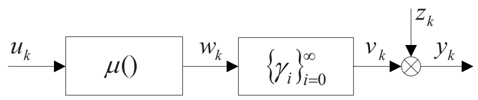

FIGURE 2.6. Hammerstein system

The aim is to estimate both $\mu()$ and $\{\gamma_i\}_{i=0}^{\infty}$ on the basis of N input output measurements $\{(u_k, y_k)\}_{k=1}^{N}$ of the whole Hammerstein system. Obviously, since the signal w_k is hidden, both elements can de identified only up to some scale factor (the system with the characteristic $c\mu()$ and linear dynamics $\{\frac{1}{c}\gamma_i\}_{i=0}^{\infty}$ is not distinguishable with the original one, from the input-output point of view). This problem cannot be avoided without the additional knowledge of the system, and is independent of the identification method. In the literature it is often assumed, without any loss of generality, that $\gamma_0 = 1$.

First attempts to Hammerstein system identification ([75], [14], [9]-[10]), proposed in 1970s and 1980s, assumed polynomial form of $\mu()$ and FIR linear dynamics with known orders p and m. The system was parametrized as follows

$$w_k = c_p u_k^p + c_{p-1} u_k^{p-1} + ... + c_1 u + c_0, \tag{2.38}$$

$$y_k = \sum_{i=0}^{m} b_i w_{k-i} + z_k. \tag{2.39}$$

Since the nonlinear characteristic (2.38) is linear with respect to the parameters, and the second subsystem is linear, also the whole Hammerstein system is linear with respect to parameters and can be presented in the following form

$$y_k = \phi_k \theta + z_k, \tag{2.40}$$

where the regressor ϕ_k includes all combinations of u_{k-i}^l ($i = 0, 1, ..., m$, and $l = 0, 1, ..., p$), and the vector θ is built of all respective mixed products of parameters, i.e., $c_l b_i$ ($i = 0, 1, ..., m$, and $l = 0, 1, ..., p$). Such a description allows for application of the least squares method for estimation of θ in (2.40). Nevertheless, using ordinary polynomials leads to the tasks, which are very badly conditioned numerically. Moreover, if the assumed parametric model is bad, the approximation error appears, which cannot be reduced by increasing number of data.

As regards the nonparametric methods ([74], [29]-[38]), strongly elaborated in 1980s, they are based on the observation that, for i.i.d. excitation $\{u_k\}$, the input-output regression in Hammerstein system is equivalent to the nonlinear characteristic of the static element, up to some constant δ, i.e.,

$$R(u) = E\{y_k|u_k = u\} = E\left\{\gamma_0\mu(u_k) + \sum_{i=1}^{\infty}\gamma_i\mu(u_{k-i})|u_k = u\right\} =$$
$$= \gamma_0\mu(u_k) + \delta,$$

where the offset $\delta = \sum_{i=1}^{\infty}\gamma_i E\mu(u_1)$ can be simply avoided if the characteristic is known in at least one point, e.g., we know that $\mu(0) = 0$. As it was shown in e.g. [37] and [39], standard nonparametric regression estimates $\widehat{R}_N(u)$ (kernel or orthogonal) can be successfully applied for correlated data $\{(u_k, y_k)\}_{k=1}^{N}$. Under some general conditions, they recover the shape of true nonlinear characteristic, and are asymptotically free of approximation error. The cost paid for neglecting of the prior knowledge is large variance of the model for small number of measurements. The reason of the variance is that the Hammerstein system is in fact treated as the static element, and the historical term $\sum_{i=1}^{\infty}\gamma_i\mu(u_{k-i})$ is treated as the 'system' noise. Hence, nonparametric methods are better choice as regards the asymptotic properties, when the number of measurements is large enough (required number obviously depends on specifics of the system).

2.3.2 Wiener system

Since in the Wiener system the nonlinear block precedes the linear dynamics, the identification task is much more difficult. Till now, sufficient identifiability conditions have been formulated have 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. [9], [48], [42]).The discrete-time Wiener system (see Fig. 2.7), i.e. the linear dynamics with the impulse response $\{\lambda_j\}_{j=0}^{\infty}$, connected in the cascade with the static nonlinear block characterized by $\mu()$, is described by the following equation

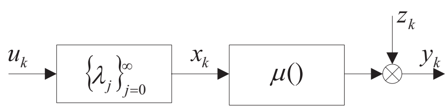

FIGURE 2.7. Wiener system

$$y_k = \mu\left(\sum_{j=0}^{\infty}\lambda_j u_{k-j}\right) + z_k, \tag{2.41}$$

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 functional, e.g. the ARMA-type dynamic block

$$\begin{aligned} x_k + a_1^* x_{k-1} + ... + a_r^* x_{k-r} &= b_0^* u_k + b_1^* u_{k-1} + ... + b_s^* u_{k-s}, \\ x_k &= \phi_k^T \theta^*, \\ \phi_k &= (-x_{k-1}, ..., -x_{k-r}, u_k, u_{k-1}, ..., u_{k-s})^T, \\ \theta^* &= (a_1^*, ..., a_r^*, b_0^*, b_1^*, ..., b_s^*)^T, \end{aligned} \tag{2.42}$$

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.,

$$\begin{aligned} \overline{x}_k &\triangleq \overline{\phi}_k^T \theta, \text{ and } \overline{x}_k = 0 \text{ for } k \leqslant 0, \\ \text{where } \overline{\phi}_k^T &= (-\overline{x}_{k-1}, ..., -\overline{x}_{k-r}, u_k, u_{k-1}, ..., u_{k-s})^T, \\ \theta &= (a_1, ..., a_r, b_0, b_1, ..., b_s)^T, \\ \text{and } \overline{y}(x, c) &= \mu(x, c), \text{ where } c = (c_1, c_2, ..., c_m)^T. \end{aligned} \tag{2.43}$$

If x_k had been accessible for measurements then the true system parameters could have been estimated by the following minimizations

$$\widehat{\theta} = \arg\min_{\theta} \sum_{k=1}^{N} (x_k - \overline{x}_k(\theta))^2, \qquad \widehat{c} = \arg\min_{c} \sum_{k=1}^{N} (y_k - \overline{y}(x_k, c))^2. \tag{2.44}$$

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

$$\widehat{\theta}, \widehat{c} = \arg\min_{\theta, c} \sum_{k=1}^{N} [y_k - \overline{y}(\overline{x}_k(\theta), c)]^2, \tag{2.45}$$

which is usually very complicated. Moreover, uniqueness of the solution in (7.10) cannot be guaranteed in general, as it depends on both input distribution, types of models, and values of parameters.

For example the Wiener system with the polynomial static characteristic and the *FIR* linear dynamics

$$y_k = \sum_{i=0}^{p} c_0 w_k^i + z_k, \qquad w_k = \sum_{i=0}^{m} \gamma_i u_{k-i}, \tag{2.46}$$

can be described as

$$Y_N = \Phi_N \theta + Z_N, \tag{2.47}$$

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.,

$$\begin{aligned} \phi_k &= \left[\frac{|\alpha|!}{\alpha!} \overline{u}_k^{\alpha} \right]_{|\alpha| \leqslant p} \in \mathcal{R}^d, \\ \theta &= \left[c_{|\alpha|} \Gamma^{\alpha} \right]_{|\alpha| \leqslant p} \in \mathcal{R}^d, \end{aligned} \tag{2.48}$$

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 [61]), $|\alpha| = \sum_{i=0}^{m+1} \alpha_i$, $\alpha! = \prod_{i=1}^{m+1} \alpha_i$, $\overline{u}_k = (u_k, u_{k-1}, ..., u_{k-m})^T$, $\overline{u}_k^{\alpha} = \prod_{i=1}^{m+1} u_{k-i-1}^{\alpha_i}$, $\Gamma^{\alpha} = \prod_{i=1}^{m+1} \gamma_{i-1}^{\alpha_i}$, $d = \sum_{i=0}^{p} \frac{(m+i)!}{m!i!}$, 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. Estimate of θ can be computed by the standard least squares, analogously as for Hammerstein system, but extraction of parameters of components requires application of multi-dimensional singular value decomposition (SVD) procedure.

As regards nonparametric methods, proposed in 1980s, for Wiener system identification, they were usually based on the restrictive assumptions that the nonlinear characteristic is invertible or locally invertible and the input process is Gaussian. Moreover, existence of output random noise were excluded. It was shown, that under above conditions the reverse regression function

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

is equivalent to the inversion of the identified characteristic up to some (impossible to identify) scale α_p. Recently, several new ideas was proposed in the literature (see e.g. [81], [69], [33]), admitting arbitrary input density, nonivertible characteristic and IIR linear block. Nevertheless, the rate of convergence of estimates is still not satisfying.

2.3.3 Sandwich (Wiener-Hammerstein) system

In the Wiener-Hammerstein (sandwich) system, presented in Fig. 2.8, the nonlinear block $\mu()$ is surrounded by two linear dynamics $\{\lambda_j\}_{j=0}^{\infty}$ and $\{\gamma_i\}_{i=0}^{\infty}$. It can be described as follows

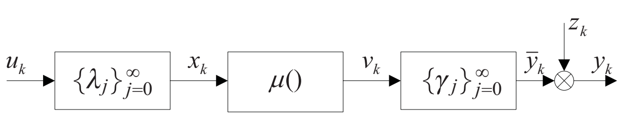

FIGURE 2.8. Wiener-Hammerstein (sandwich) system

$$y_k = \sum_{i=0}^{\infty} \gamma_i v_{k-i} + z_k, \text{ where } v_k = \mu\left(\sum_{j=0}^{\infty} \lambda_j u_{k-j}\right). \tag{2.49}$$

Since both x_k and v_k cannot be measured, the system as a whole cannot be distinguished with the system composed with the elements $\{\frac{1}{c_1}\lambda_j\}$, $c_2\mu(c_1 x)$, $\{\frac{1}{c_2}\gamma_i\}$, and the nonlinear characteristic can be identified only up to $c_1$and c_2. Analogously as for Hammerstein and Wiener systems, without any loss of generality, we assume that $\gamma_0 = 1$ and $\lambda_0 = 1$. The Hammerstein system and Wiener systems are special cases of (2.49), obtained for $\lambda_j = 1, 0, 0, ...$ or $\gamma_i = 1, 0, 0, ...$, respectively.

2.3.4 NARMAX system

We also consider in this book a special case of NARMAX systems, consisting of two branches of Hammerstein structure, from which one plays the role of nonlinear feedback (see Fig. 2.9). The system is described by the equation

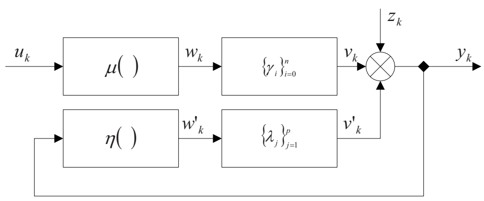

FIGURE 2.9. The additive NARMAX system

$$y_k = \sum_{j=1}^{\infty} \lambda_j \eta(y_{k-j}) + \sum_{i=0}^{\infty} \gamma_i \mu(u_{k-i}) + z_k, \tag{2.50}$$

where $\eta()$ is nonlinear static characteristic in feedback. The system is assumed to be stable as a whole. As it was shown in the Appendix A.3, the Hammerstein system is a special case of (2.50) when the characteristic $\eta()$ is linear.

2.3.5 Interconnected MIMO system

Finally, we introduce the system with arbitrary structure of connections (Fig. 2.10). It consists of n blocks described by unknown functionals $F_i(\{u_i\}, \{x_i\})$, $i = 1, 2, ..., n$. Only external inputs u_i and outputs y_i of the system can be measured. The interactions x_i are hidden, but the structure of connections

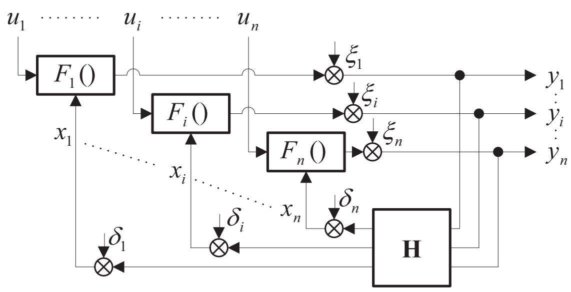

FIGURE 2.10. The system with arbitrary structure

is known and coded in the zero-one matrix H, i.e.,

$$x_i = H_i \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} + \delta_i,$$

where H_i denotes ith row of H, and δ_i is a random disturbance.

In the simplest case of static linear system, the single block can be described as follows

$$y_i = [a_i, b_i] \begin{bmatrix} x_i \\ u_i \end{bmatrix} + \xi_i \qquad (i = 1, 2, ..., n),$$

where a_i and b_i are unknown parameters and ξ_i is a random output noise. In more general case of nonlinear and dynamic system, the single block can be represented by, e.g., two channels of Hammerstein models (see Fig. 2.11).

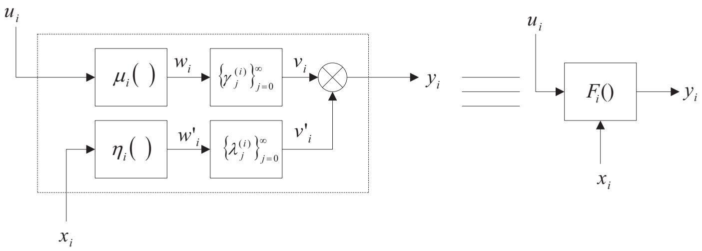

FIGURE 2.11. Example of the single nonlinear dynamic block in complex system

2.3.6 Applications in science and technology

Block-oriented models are often met in various domains of science and technology. For details, we refer the reader to the paper [26], including a huge number of publications on this topic. Examples of the most popular applications are given below:

- signal processing (time series modeling and forecasting, noise reduction, data compression, EEG analysis),
- automation and robotics (fault detection, adaptive control),
- telecommunication and electronics (channel equalization, image compression, diode modeling, testing of AM and FM decoders),
- acoustics (noise and echo cancellation, loudspeaker linearization),
- biocybernetics (artificial eye, artificial muscle, model of neuron),
- geology (anti-flood systems, water level modeling, deconvolution),
- chemistry (modeling of distillation and fermentation processes, pH neutralization),
- physics (adaptive optics, heat exchange processes, modeling of the Diesel motor).