1 Introduction

System identification, as a particular process of statistical inference, exploits two types of information. The first is experiment; the other, called a priori, is known before making any measurements. In a wide sense, the a priori information concerns the system itself and signals entering the system. Elements of the information are, for example:

- the nature of the signals, which may be random or nonrandom, white or correlated, stationary or not, their distributions can be known in full or partially (up to some parameters) or completely unknown,
- general information about the system, which can be, for example, continuous or discrete in the time domain, stationary or not,
- the structure of the system, which can be of the Hammerstein or Wiener type, or other,
- the knowledge about subsystems, that is, about nonlinear characteristics and linear dynamics.

In other words, the a priori information is related to the theory of the phenomena taking place in the system (a real physical process) or can be interpreted as a hypothesis (if so, results of the identification should be necessarily validated) or can be abstract in nature.

This book deals with systems consisting of nonlinear memoryless and linear dynamic subsystems, for example, Hammerstein and Wiener systems and other related structures. With respect to them, the a priori information is understood in a narrow sense because it relates to the subsystems only and concerns the a priori knowledge about their descriptions. We refer to such systems as block-oriented.

The characteristic of the nonlinear subsystem is recovered with the help of nonparametric regression estimates. The kernel and orthogonal series methods are used. Ordered statistics are also applied. Both offline and online algorithms are investigated. We examine only these estimation methods and nonlinear models for which we are able to deliver fundamental results in terms of consistency and convergence rates. There are other techniques, for example, neural networks, which may exhibit a promising performance but their statistical accuracy is mostly unknown.

For the theory of nonparametric regression, see Efromovich [78], Györfi, Kohler, Krzyżak, and Walk [140], Härdle [150], Prakasa Rao [241], Simonoff [278], or Wand and Jones [310]. Nonparametric wavelet estimates are discussed in Antoniadis and Oppenheim [6], Härdle, Kerkyacharian, Picard, and Tsybakov [151], Ogden [223], and Walter and Shen [308].

Parametric methods are beyond the scope of this book; nevertheless, we mention Brockwell and Davies [33], Ljung [198], Norton [221], Zhu [332], and Söderström and Stoica [280].

Nonlinear system identification within the parametric framework is studied by Nells [218], Westwick and Kearney [316], Marmarelis and Marmarelis [207], Bendat [16], and Mathews and Sicuranza [208]. These books present identification algorithms based mostly on the theory of Wiener and Volterra expansions of nonlinear systems. A comprehensive list of references concerning nonlinear system identification and applications has been given by Giannakis and Serpendin [102], see also the 2005 special issue on system identification of the IEEE Trans. on Automatic Control [199]. A nonparametric statistical inference for time series is presented in Bosq [26], Fan and Yao [89], and Györfi, Härdle, Sarda, and Vieu [139].

It should be stressed that nonparametric and parametric methods are supposed to be applied in different situations. The first are used when the a priori information is nonparametric, that is, when we wish to recover an infinite-dimensional object with underlying assumptions as weak as possible. Clearly, in such a case, parametric methods can only approximate, but not estimate, the unknown characteristics. When the information is parametric, parametric methods are the natural choice. If, however, the unknown characteristic is a complicated function of parameters convergence analysis becomes difficult. Moreover, serious computational problems can occur. In such circumstances, one can resort to nonparametric algorithms because, from the computational viewpoint, they are not discouraging. On the contrary, they are simple but consume computer memory, because, for example, kernel estimates require all data to be stored. Nevertheless it can be said that the two approaches do not compete with each other since they are designed to be applied in quite different situations. The situations differ from each other by the amount of the a priori information about the identified system. However, a compromise between these two separate worlds can be made by restricting a class of nonparametric models to those that consist of a finite dimensional parameter and nonlinear characteristics, which run through a nonparametric class of univariate functions. Such semiparametric models can be efficiently identified, and the theory of semiparametric identification is examined in this book. The methodology of semiparametric statistical inference is examined in Härdle, Müller, Sperlich, and Werwatz [152], Ruppert, Wand, and Carroll [259], and Yatchev [329].

For two number sequences a_n and b_n, $a_n = O(b_n)$ means that a_n/b_n is bounded in absolute value as $n \to \infty$. In particular, $a_n = O(1)$ denotes that a_n is bounded, that is, that $\sup_n |a_n| < \infty$. Writing $a_n \sim b_n$, we mean that a_n/b_n has a nonzero limit as $n \to \infty$.

Throughout the book, "almost everywhere" means "almost everywhere with respect to the Lebesgue measure," whereas "almost everywhere (μ)" means "almost everywhere with respect to the measure μ."

2 Discrete-time Hammerstein systems

In this chapter, we discuss some preliminary aspects of the discrete-time Hammerstein system. In Section 2.1 we form the input–output equations of the system. A fundamental relationship between the system nonlinearity and the nonparametric regression is established in Section 2.2. The use of the correlation theory for recovering the linear subsystem is discussed in Section 2.3.

2.1 The system

A Hammerstein system, shown in Figure 2.1, consists of a nonlinear memoryless subsystem with a characteristic $m(\bullet)$ followed by a linear dynamic one with an impulse response $\{\lambda_n\}$. The output signal W_n of the linear part is disturbed by Z_n and $Y_n = W_n + Z_n$ is the output of the whole system. Neither V_n nor W_n is available to measurement. Our goal is to identify the system, that is, to recover both $m(\bullet)$ and $\{\lambda_n\}$, from observations

$$(U_1, Y_1), (U_2, Y_2), \ldots, (U_n, Y_n), \ldots \tag{2.1}$$

taken at the input and output of the whole system.

Signals coming to the system, that is, the input $\{\ldots, U_{-1}, U_0, U_1, \ldots\}$ and disturbance $\{\ldots, Z_{-1}, Z_0, Z_1, \ldots\}$ are mutually independent stationary white random signals. The disturbance has zero mean and finite variance, that is, $EZ_n = 0$ and $\text{var}[Z_n] = \sigma_Z^2 < \infty$.

Regarding the nonlinear subsystem, we assume that $m(\bullet)$ is a Borel measurable function. Therefore, V_n is a random variable. The dynamic subsystem is described by the state equation

$$\begin{cases} X_{n+1} = AX_n + bV_n \\ W_n = c^T X_n, \end{cases} \tag{2.2}$$

where X_n is a state vector at time n, A is a matrix, b and c are vectors. Thus,

$$\lambda_n = \begin{cases} 0, & \text{for } n = 0, -1, -2, \ldots \\ c^T A^{n-1} b, & \text{for } n = 1, 2, 3, \ldots, \end{cases}$$

and

$$W_n = \sum_{i=-\infty}^{n} \lambda_{n-i} m(U_i). \tag{2.3}$$

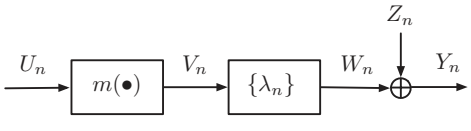

Figure 2.1 The discrete-time Hammerstein system.

Neither b nor c is known. The matrix A and its dimension are also unknown. Nevertheless, the matrix A is stable, all its eigenvalues lie in the unit circle. Therefore, assuming that

$$Em^2(U) < \infty, \tag{2.4}$$

the time index at U is dropped, we conclude that both X_n as well as W_n are random variables. Clearly random processes $\{\ldots, X_{-1}, X_0, X_1, \ldots\}$ and $\{\ldots, W_{-1}, W_0, W_1, \ldots\}$ are stationary. Consequently, the output process $\{\ldots, Y_{-1}, Y_0, Y_1, \ldots\}$ is also a stationary stochastic process. Therefore, the problem is well posed in the sense that all signals are random variables. In the light of this, we estimate both $m(\bullet)$ and $\{\lambda_n\}$ from random observations (2.1).

The restrictions imposed on the signals entering the system and both subsystems apply whenever the Hammerstein system is concerned. They will not be repeated in further considerations, neither lemmas nor theorems.

Input random variables U_ns may have a probability density denoted by $f(\bullet)$ or may be distributed quite arbitrarily. Nevertheless (2.4) holds. It should be emphasized that, apart from few cases, (2.4) is the only restriction in which the nonlinearity is involved.

Assumption (2.4) is irrelevant to identification algorithms and has been imposed for only one reason: to guarantee that both W_n and Y_n are random variables. Nevertheless it certainly has an influence on the restrictions imposed on both $m(\bullet)$ and the distribution of U to meet (2.4). If, for example, U is bounded, (2.4) is satisfied for any $m(\bullet)$. The restriction also holds, if $EU^2 < \infty$ and $|m(u)| \leq \alpha + \beta|u|$ with any α, β. In yet another example, $EU^4 < \infty$ and $|m(u)| \leq \alpha + \beta u^2$. For Gaussian U and $|m(u)| \leq W(u)$, where W is an arbitrary polynomial, (2.4) is also met. Anyway, the a priori information about the characteristic is nonparametric because $m(\bullet)$ cannot be represented in a parametric form. This is because the class of all possible characteristics is very wide.

The family of all stable dynamic subsystems also cannot be parameterized, because its order is unknown. Therefore, the a priori information about the impulse response is nonparametric, too. To form a conclusion we infer about both subsystems under nonparametric a priori information.

In the following chapters, for simplicity, U, W, Y, and Z stand for U_n, W_n, Y_n, and Z_n, respectively.

2.2 Nonlinear subsystem

2.2.1 The problem and the motivation for algorithms

Fix $p \geq 1$ and observe that, since $Y_p = Z_p + \sum_{i=-\infty}^{p} \lambda_{p-i} m(U_i)$ and $\{U_n\}$ is a white process,

$$E\left\{Y_p | U_0 = u\right\} = \mu(u),$$

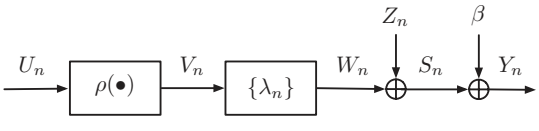

Figure 2.2 The equivalent Hammerstein system.

where

$$\mu(u) = \lambda_p m(u) + \alpha_p$$

with $\alpha_p = Em(U)\sum_{i=1,i\neq p}^{\infty}\lambda_i$. Estimating the regression $E\{Y_p|U_0 = u\}$, we thus recover $m(\bullet)$ up to some unknown constants λ_p and α_p. If $Em(U) = 0$, which is the case, for example, when the distribution of U is symmetrical with respect to zero and $m(\bullet)$ is an even function then $\alpha_p = 0$ and we estimate $m(\bullet)$ only up to the multiplicative constant λ_p.

Since $Y_{p+n} = \mu(U_n) + \xi_{p+n} + Z_{p+n}$ with $\xi_{p+n} = \sum_{i=-\infty,i\neq n}^{p+n}\lambda_{p+n-i}m(U_i)$, it can be said that we estimate $\mu(u)$ from pairs

$$(U_0, Y_p), (U_1, Y_{p+1}), \ldots, (U_n, Y_{p+n}), \ldots,$$

and that the regression $\mu(u)$ is corrupted by the noise $Z_{p+n} + \xi_{p+n}$. The first component of noise is white with zero mean. Because of dynamics the other noise component is correlated. Its mean $E\xi_n = \alpha_p$ is usually nonzero and the variance is equal to $\text{var}[m(U)]\sum_{i=1,i\neq p}^{\infty}\lambda_i^2$. Thus, main difficulties in the analysis of any estimate of $\mu(\bullet)$ are caused by the correlation of $\{\xi_n\}$, that is, the system itself but not by the white disturbance Z_n coming from outside.

Every algorithm estimating the nonlinearity in Hammerstein systems studied in this book, the estimate is denoted here as $\hat{\mu}(U_0, \ldots, U_n; Y_p, \ldots, Y_{p+n})$, is linear with respect to output observations, which means that

$$\begin{aligned}&\hat{\mu}(U_0, \ldots, U_n; \theta_p + \eta_p, \ldots, \theta_{p+n} + \eta_{p+n})\\ &\quad = \hat{\mu}(U_0, \ldots, U_n; \theta_p, \ldots, \theta_{p+n}) + \hat{\mu}(U_0, \ldots, U_n; \eta_p, \ldots, \eta_{p+n})\end{aligned} \tag{2.5}$$

and has a natural property that, for any number θ,

$$\hat{\mu}(U_0, \ldots, U_n; \theta, \ldots, \theta) \to \theta \text{ as } n \to \infty \tag{2.6}$$

in an appropriate stochastic sense. This property, or rather its consequence, is exploited when proving consistency. To explain this, observe that with respect to U_n and Y_n, the identified system shown in Figure 2.1 is equivalent to that in Figure 2.2 with nonlinearity $\rho(u) = m(u) - Em(U)$ and an additional disturbance $\beta = Em(U)\sum_{i=1}^{\infty}\lambda_i$. In the equivalent system, $E\rho(U) = 0$ and $E\{Y_p|U_0 = u\} = \mu(u)$. From (2.5) and (2.6), it follows that

$$\begin{aligned}\hat{\mu}(U_0, \ldots, U_n; Y_p, \ldots, Y_{p+n}) &= \hat{\mu}(U_0, \ldots, U_n; S_p + \beta, \ldots, S_{p+n} + \beta)\\ &= \hat{\mu}(U_0, \ldots, U_n; S_p, \ldots, S_{p+n})\\ &\quad + \hat{\mu}(U_0, \ldots, U_n; \beta, \ldots, \beta)\end{aligned}$$

with $\hat{\mu}(U_0, \ldots, U_n; \beta, \ldots, \beta) \to \beta$ as $n \to \infty$. Hence, if

$$\hat{\mu}(U_0, \ldots, U_n; S_p, \ldots, S_{p+n}) \to E\{S_p|U_0 = u\}, \text{ as } n \to \infty,$$

we have

$$\hat{\mu}(U_0, \ldots, U_n; Y_p, \ldots, Y_{p+n}) \to E\{Y_p|U_0 = u\}, \text{ as } n \to \infty,$$

where convergence is understood in the same sense as that in (2.6).

Thus, if the estimate recovers the regression $E\{S_p|U_0 = u\}$ from observations

$$(U_0, S_p), (U_1, S_{1+p}), (U_2, S_{2+p}), \ldots,$$

it also recovers $E\{Y_p|U_0 = u\}$ from

$$(U_0, Y_p), (U_1, Y_{1+p}), (U_2, Y_{2+p}), \ldots.$$

We can say that if the estimate works properly when applied to the system with input U_n and output S_n (in which $E\rho(U) = 0$), it behaves properly also when applied to the system with input U_n and output Y_n (in which $Em(U)$ may be nonzero).

The result of the reasoning is given in the following remark:

REMARK 2.1 *Let an estimate have properties (2.5) and (2.6). If the estimate is consistent for $Em(U) = 0$, then it is consistent for $Em(U) \neq 0$, too.*

Owing to the remark, with no loss of generality, in all proofs of consistency of algorithms recovering the nonlinearity, we assume that $Em(U) = 0$.

In parametric problems the nonlinearity is usually a polynomial $m(u) = \alpha_0 + \alpha_1 u + \cdots + \alpha_q u^q$ of a fixed degree with unknown true values of parameters $\alpha_0, \ldots, \alpha_q$. Therefore, to apply parametric methods, we must have a great deal more a priori information about the subsystem. It seems that in many applications, it is impossible to represent $m(\bullet)$ in a parametric form.

Since the system with the following $ARMA$ type difference equation:

$$w_n + a_{k-1}w_{n-1} + \cdots + a_0 w_{n-k} = b_{k-1}m(u_{n-1}) + \cdots + b_0 m(u_{n-k})$$

can be described by (2.2), all presented methods can be used to recover the nonlinearity $m(\bullet)$ in the previous $ARMA$ system.

It will be convenient to denote

$$\phi(u) = E\left\{W_p^2|U_0 = u\right\}. \tag{2.7}$$

Since $W_p = \sum_{i=-\infty}^{p-1} \lambda_i m(U_i)$, denoting $c_0 = Em^2(U)\sum_{i=1, i\neq p}^{\infty} \lambda_i^2 + E^2 m(U)$ $(\sum_{i=1, i\neq p}^{\infty} \lambda_i)^2$, $c_1 = 2\lambda_p Em(U)\sum_{i=1, i\neq p}^{\infty} \lambda_i$, and $c_2 = \lambda_p^2$, we find

$$\phi(u) = c_0 + c_1 m(u) + c_2 m^2(u).$$

To avoid complicated notation, we do not denote explicitly the dependence of the estimated regression and other functions on p and simply write $\mu(\bullet)$ and $\phi(\bullet)$.

Results presented in further chapters can be easily generalized on the system shown in Figure 2.3, where $\{\ldots, \xi_0, \xi_1, \xi_2, \ldots\}$ is another zero mean noise. Moreover, $\{Z_n\}$ can

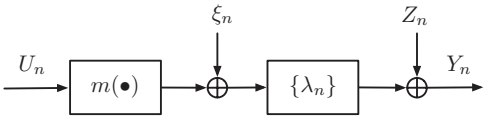

Figure 2.3 Possible generalization of the system shown in Figure 2.1.

be correlated, that is, it can be the output of a stable linear dynamic system stimulated by white random noise. So can $\{\xi_n\}$.

It is worth noting that a class of stochastic processes generated by the output process $\{Y_n\}$ of the Hammerstein system is different from the class of strong mixing processes considered extensively in the statistical literature concerning the nonparametric inference from dependent data, see, for example, [26] and [89]. Indeed, the ARMA process $\{X_n\}$ in which $X_{n+1} = aX_n + V_n$, where $0 < a \leq 1/2$, and where V_ns are Bernoulli random variables is not strong mixing, see [4] and [5]. Such a process can be easily generated by the Hammerstein system if the input of the whole system has a normal density and a nonlinear characteristic takes two different values. In the light of that, the strong mixing approach developed in the statistical literature does not apply in general to the identification problem of nonlinear systems.

2.2.2 Simulation example

In the chapters devoted to the Hammerstein system, the behavior of the identification algorithms presented in this book is illustrated with results of simulation examples. In all examples, the system is described by the following scalar equation:

$$X_{n+1} = aX_n + m(U_n),$$

where

$$m(u) = (1 - e^{-|u|})\,\mathrm{sign}(u),$$

(see Figure 2.4). The input signal has a normal density with zero mean and variance 1. In all algorithms, $p = 1$, which means that $\mu(u) = m(u)$. For $a = 0.5$, an example of a

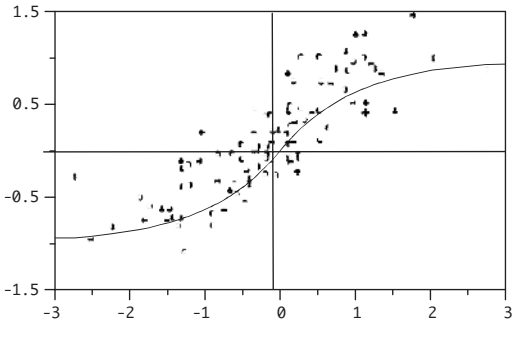

Figure 2.4 The characteristic m and 200 pairs of input–output observations; $a = 0.5$.

cloud of 200 input–output observations, we infer from is presented in Figure 2.4. The quality of each estimate, denoted here by $\hat{m}(u)$, is measured with

$$\text{MISE} = \int_{-3}^{3} (\hat{m}(u) - m(u))^2 du.$$

2.3 Dynamic subsystem identification

Passing to the dynamic subsystem, we use (2.3) and recall $EZ_n = 0$ to notice $E\{Y_i U_0\} = \sum_{j=-\infty}^{i} \lambda_{i-j} E\{m(U_i)U_0\} = \lambda_i E\{m(U)U\}$. Denoting $\kappa_i = \lambda_i E\{Um(U)\}$, we obtain

$$\kappa_i = E\{Y_i U_0\},$$

which can be estimated in the following way:

$$\hat{\kappa}_i = \frac{1}{n} \sum_{j=1}^{n-i} Y_{i+j} U_j.$$

THEOREM 2.1 *For any i,*

$$\lim_{n\to\infty} E(\hat{\kappa}_i - \kappa_i)^2 = 0.$$

Proof. The estimate is unbiased, that is, $E\hat{\kappa}_i = E\{Y_i U_0\} = \kappa_i$. Moreover, $\text{var}[\hat{\kappa}_i] = P_n + Q_n + R_n$ with

$$P_n = \frac{1}{n^2} \text{var}\left[\sum_{j=1}^{n} Z_{i+j} U_j\right] = \frac{1}{n^2} \sum_{j=1}^{n} \text{var}\left[Z_{i+j} U_j\right] = \frac{1}{n} \sigma_Z^2 EU^2,$$

$$Q_n = \frac{1}{n} \text{var}\left[W_i U_0\right],$$

and

$$R_n = \frac{1}{n^2} \sum_{j=1}^{n} \sum_{j=1, j\neq i}^{n} \text{cov}\left[W_{i+j} U_j, W_{i+m} U_m\right]$$
$$= \frac{1}{n^2} \sum_{j=1}^{n} (n-j) \,\text{cov}\left[W_{i+j} U_j, W_i U_0\right].$$

Since $W_i = \sum_{j=-\infty}^{i} \lambda_{i-j} m(U_j)$, $Q_n = n^{-1} \lambda_i^2 \,\text{var}\,[m(U)U]$. For the same reason, for $j > 0$,

$$\text{cov}\left[W_{i+j} U_j, W_i U_0\right] = \sum_{p=-\infty}^{i+j} \sum_{q=-\infty}^{i} \lambda_{i+j-p} \lambda_{i-q} \,\text{cov}\left[m(U_p) U_j, m(U_q) U_0\right]$$
$$= E^2\{Um(U)\} \lambda_{i+j} \lambda_{i-j}$$

See Lemma C.3 in Appendix C, which leads to

$$|R_n| \le \frac{1}{n^2} E^2\{Um(U)\} \sum_{j=1}^{n} (n-j)|\lambda_{i+j}\lambda_{i-j}| \le \frac{1}{n} E^2\{Um(U)\} \max_s |\lambda_s| \sum_{j=1}^{\infty} |\lambda_j|.$$

Thus,

$$E(\hat{\kappa}_i - \kappa_i)^2 = \text{var}\,[\hat{\kappa}_i] = O\left(\frac{1}{n}\right) \tag{2.8}$$

which completes the proof. ■

The theorem establishes convergence of the local error $E(\hat{\kappa}_i - \kappa_i)^2$ to zero as $n \to \infty$. As an estimate of the whole impulse response $\{\kappa_1, \kappa_2, \kappa_3, \ldots\}$, we take a sequence $\{\hat{\kappa}_1, \hat{\kappa}_2, \hat{\kappa}_3, \ldots, \hat{\kappa}_{N(n)}, 0, 0, \ldots\}$ and find the mean summed square error (MSSE) is equal to

$$\text{MSSE}(\hat{\kappa}) = \sum_{i=1}^{N(n)} E(\hat{\kappa}_i - \kappa_i)^2 + \sum_{i=N(n)+1}^{\infty} \kappa_i^2.$$

From (2.8), it follows that the error is not greater than

$$O\left(\frac{N(n)}{n}\right) + \sum_{i=N(n)+1}^{\infty} \kappa_i^2.$$

Therefore, if $N(n) \to \infty$ as $n \to \infty$ and $N(n)/n \to 0$ as $n \to \infty$,

$$\lim_{n\to\infty} \text{MSSE}(\hat{\kappa}) = 0.$$

The identity $\lambda_s \tau = E\{Y_s U_0\}$, where $\tau = E\{Um(U)\}$, allows us to form a nonparametric estimate of the linear subsystem in the frequency domain. Indeed, formation of the Fourier transform of the identity yields

$$\Lambda(\omega)\tau = S_{YU}(\omega), \quad |\omega| \le \pi, \tag{2.9}$$

where $S_{YU}(\omega) = \sum_{s=-\infty}^{\infty} \kappa_s e^{-is\omega}$ is the cross-spectral density function of the processes $\{Y_n\}$ and $\{U_n\}$. Moreover,

$$\Lambda(\omega) = \sum_{s=0}^{\infty} \lambda_s e^{-is\omega}$$

is the transfer function of the linear subsystem. Note also that if $\lambda_0 = 1$, then $\tau = \kappa_0$. See Chapter 12 for further discussion on the frequency domain identification of linear systems.

2.4 Bibliographic notes

Various aspects of parametric identification algorithms of discrete-time Hammerstein systems have been studied by Narendra and Gallman [216]; Haist, Chang, and Luus

[142], Thatchachar and Ramaswamy [289], Kaminskas [175], Gallman [92], Billings [19], Billings and Fakhouri [20,24], Shih and Kung [276], Kung and Shih [190], Liao and Sethares [195], Verhaegen and Westwick [301], Giri, Chaoui, and Rochidi [103], Ninness and Gibson [220], Bai [11,12], and Vörös [305]. The analysis of block–oriented systems and, in particular, Hammerstein ones, useful for various aspects of identification and its applications can be found in Bendat [16], Chen [45], Marmarelis and Marmarelis [207], Mathews and Sicuranza [208], Nells [218], and Westwick and Kearney [316].

Sometimes results concerning Hammerstein systems are given, however not explicitly, in works devoted to more complicated Hammerstein–Wiener or Wiener–Hammerstein structures, see, for example, Gardiner [94], Billings and Fakhouri [22, 23], Fakhouri, Billlings, and Wormald [86], Hunter and Korenberg [168], Korenberg and Hunter [177], Emara-ShaBaik, Moustafa, and Talaq [79], Boutayeb and Darouach [27], Vandersteen, Rolain, and Schoukens [296], Bai [10], Bershad, Celka, and McLaughlin [18], and Zhu [333].

The nonparametric approach offers a number of algorithms to recover the characteristics of the nonlinear subsystem. The most popular kernel estimate can be used in the offline version, see Chapter 3. For semirecursive and fully recursive forms, see Chapter 4 and Chapter 5, respectively. Nonparametric orthogonal series identification algorithms, see Chapter 6, utilize trigonometric, Legendre, Laguerre, Hermite functions or wavelets. Both classes of estimates can be modified to use ordered input observations (see Chapter 7), which makes them insensitive to the roughness of the input density.

The Hammerstein model has been used in various and diverse areas. Eskinat, Johnson, and Luyben [82] applied it to describe processes in distillation columns and heat exchangers. The hysteresis phenomenon in ferrites was analyzed by Hsu and Ngo [166], pH processes were analyzed by Patwardhan, Lakshminarayanan, and Shah [227], biological systems were studied by Hunter and Korenberg [168], and Emerson, Korenberg, and Citron [80] described some neuronal processes. The use of the Hammerstein model for modeling aspects of financial volatility processes is presented in Capobianco [38]. In Giannakis and Serpendin [102] a comprehensive bibliography on nonlinear system identification is given, see also the 2005 special issue on system identification of the IEEE Trans. on Automatic Control [199].

It is also worth noting that the concept of the Hammerstein model originates from the theory of nonlinear integral equations developed by Hammerstein in 1930 [148], see also Tricomi [292].