HAMMERSTEIN SYSTEM IDENTIFICATION WITH STOCHASTIC APPROXIMATION

W. Greblicki

printed in:
International Journal of Modelling and Simulation,
Vol. 24, No. 3, pp. 131-138, 2004.

Abstract
A recursive algorithm to recover the nonlinear characteristic of the memoryless part of the Hammerstein system is presented. The *a priori* information about the system is nonparametric, no functional form of the nonlinearity is known. The algorithm is derived from the idea of stochastic approximation. Its convergence to the characteristic is shown and the convergence rate is given.

Key Words
System identification, Hammerstein system, recursive identification, nonparametric identification, stochastic approximation.

1. Introduction

A number of nonparametric algorithms have been proposed to recover the nonlinear characteristic in Hammerstein systems, i.e. systems in which a nonlinear memoryless subsystem is followed by a linear dynamic one. In this paper we examine a recursive estimate recovering the characteristic of the nonlinear part from input-output observation of the whole system. In general, recursive algorithms are important since they process measured data sequentially as they become available, see [1,2]. Owing to that, they can be applied in adaptive control.

Parametric methods have been examined in, e.g., [3–5], see also [6]. The nonparametric approach has been proposed in [7–10], and developed in [11-20]. The kernel method has been applied in [7–9,15,16], orthogonal series in [10,11,17], wavelets in [13,14,18], while polynomials in [19,20]. The kernel method applying ordered observations has been studied in [12]. Apart from [9,16] only, however, all cited papers deal with off-line algorithms.

In this paper we present a kernel on-line algorithm based on the stochastic approximation framework, see, e.g., [21]. We show that it converges to the characteristic and examine the convergence rate.

2. Identification Problem

In the Hammerstein system, see fig. 1, a memoryless nonlinear subsystem is followed by a linear dynamic one. First has a characteristic m which means that $W_n = m(U_n)$, where U_n and W_n are input and output of the subsystem. We assume that $\{U_n; n = \cdots, -1, 0, 1, \cdots\}$ is a sequence of independent identically distributed random variables having an unknown density f. The characteristic m is a Borel measurable function. Owing to that, W_n's are also random variables and $\{W_n; n = \cdots, -1, 0, 1, \cdots\}$ is stationary white noise. The linear subsystem is described by the following equation:

$$\begin{cases} X_{n+1} &= AX_n + bW_n \\ V_n &= c^T X_n, \end{cases}$$

with A, b, c unknown. The dimension of the state vector X_n is also unknown. The system is asymptotically stable which means that all eigenvalues of the matrix A lie in the unit circle. To guarantee that X_n's and Y_n's are random variables, we assume that

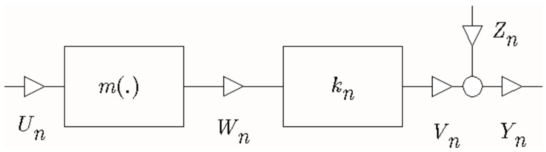

Figure 1: The Hammerstein system.

$EU_n^2 < \infty$ and

$$|m(u)| \le c_1 + c_2 |u| \tag{1}$$

with some unknown c_1 and c_2. Owing to that, $EW_n^2 < \infty$. Because of that and asymptotic stability, X_n's and V_n's are also random variables. Obviously, $\{X_n; n = \cdots, -1, 0, 1, \cdots\}$ and $\{V_n; n = \cdots, -1, 0, 1, \cdots\}$ are stationary white random processes. The output signal of the dynamic subsystem is disturbed by additive stationary white random noise $\{Z_n; n = \cdots, -1, 0, 1, \cdots\}$, i.e.,

$$Y_n = V_n + Z_n.$$

We assume that disturbance is independent of the input signal and that $EZ_n = 0$.

As far as the nonlinear subsystem is concerned, apart form (1), no assumption is imposed on its characteristic m. In particular, its functional form is completely unknown. Observe that the only reason restriction (1) has been imposed is to make the problem well posed, i.e., guarantee Y_n's to be random variables. With respect to the dynamic subsystem, all we assume is its stability. Despite so small amount of the *a priori* information, we recover m from observations $(U_1, Y_1), \cdots, (U_n, Y_n)$ taken at input and output of the whole system. We do it recursively applying an algorithm of the stochastic approximation type.

For simplicity of notation, by U, X, Y we denote random variables distributed like U_n, X_n, Y_n, respectively.

3. The algorithm and its motivation

3.1. The algorithm

We begin with the observation that

$$E\{Y_{n+1}|U_n = u\} = \mu(u), \tag{2}$$

where $\mu(u) = \alpha m(u) + \beta$ with $\alpha = c^T b$, and $\beta = c^T AEX_n$. Thus, estimating the regression $E\{Y_{n+1}|U_n = u\}$, we recover $\mu(u)$, i.e., $m(u)$ up to some multiplicative and additive constants α and β, respectively. Observe that for even distribution of U_n and odd m, we have $EU_n = 0$ and, as a consequence, $EX_n = 0$. In such a case $\beta = 0$ and we recover $\alpha m(u)$.

To introduce our estimate of $\mu(u)$, we define the following stochastic approximation procedures: for $n = 1, 2, \cdots$, let

$$\hat{g}_n(u) = \hat{g}_{n-1}(u) - \gamma_n \left[\hat{g}_{n-1}(u) - Y_{n+1}\frac{1}{h_n}K\left(\frac{u - U_n}{h_n}\right)\right], \tag{3}$$

$$\hat{f}_n(u) = \hat{f}_{n-1}(u) - \gamma_n \left[\hat{f}_{n-1}(u) - \frac{1}{h_n}K\left(\frac{u - U_n}{h_n}\right)\right], \tag{4}$$

with $\hat{g}_0(u) = 0$, $\hat{f}_0(u) = 0$. In both procedures, $\{\gamma_n\}$ and $\{h_n\}$ are number sequences while K is a kernel function. As an estimate of $m(u)$, we take

$$\hat{\mu}_n(u) = \frac{\hat{g}_n(u)}{\hat{f}_n(u)}. \tag{5}$$

Observe that, for $\gamma_n = n^{-1}$, we get

$$\hat{g}_n(u) = \frac{1}{n}\sum_{i=1}^{n} Y_{i+1}\frac{1}{h_i}K\left(\frac{u - U_i}{h_i}\right),$$

$$\hat{f}_n(u) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_i}K\left(\frac{u - U_i}{h_i}\right)$$

and consequently

$$\hat{\mu}_n(u) = \frac{\sum_{i=1}^{n} Y_{i+1}\frac{1}{h_i}K\left(\frac{u - U_i}{h_i}\right)}{\sum_{i=1}^{n} \frac{1}{h_i}K\left(\frac{u - U_i}{h_i}\right)} \tag{6}$$

which is the estimate examined in [9,16]. Our (5) and $\hat{\mu}_n(u)$ are recursive versions of the following off-line algorithm:

$$\tilde{\mu}_n(u) = \frac{\sum_{i=1}^{n} Y_{i+1} K\left(\frac{u-U_i}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{u-U_i}{h_n}\right)} \tag{7}$$

studied in [7,8]. Other nonparametric regression estimates can be found in [22,23].

Sequences of positive numbers $\{\gamma_n\}$ and $\{h_n\}$ are selected to satisfy the following restrictions:

$$h_n \to 0 \text{ as } n \to \infty, \tag{8}$$

$$\gamma_n \to 0 \text{ as } n \to \infty, \tag{9}$$

$$\sum_{n=1}^{\infty} \gamma_n = \infty, \tag{10}$$

$$\frac{\gamma_n}{h_n} \to 0 \text{ as } n \to \infty. \tag{11}$$

Observe that choosing $h_n \sim n^{-\alpha}$ and $\gamma_n \sim n^{-\gamma}$, we satisfy the above restrictions for $0 < \alpha < \gamma < 1$.

The kernel is a bounded Borel measurable function satisfying the following restrictions:

$$\sup_{-\infty<v<\infty} |K(v)| < \infty, \tag{12}$$

$$\int |K(v)|\, dv < \infty, \tag{13}$$

$$vK(v) \to 0 \text{ as } |v| \to \infty. \tag{14}$$

3.2. Intuitive motivation

Now, we shall give an intuitive explanation that (5) converges to $\mu(u)$. The algorithm is founded on the expectation that

$$\hat{g}_n(u) \to g(u) \int K(v)dv \tag{15}$$

and

$$\hat{f}_n(u) \to f(u) \int K(v)dv \tag{16}$$

as $n \to \infty$, where $g(u) = \mu(u)f(u)$. To verify that, observe that, from (3) it follows that it is reasonable to expect that

$$|\hat{g}_n(u) - g_n(u)| \to 0 \text{ as } n \to \infty \tag{17}$$

in a probabilistic sense, where

$$\begin{aligned} g_n(u) &= \frac{1}{h_n} E\left\{ Y_1 K\left(\frac{u-U_0}{h_n}\right)\right\} \\ &= \frac{1}{h_n} E\left\{ \mu(U_0) K\left(\frac{u-U_0}{h_n}\right)\right\} \\ &= \frac{1}{h_n} \int K\left(\frac{u-v}{h_n}\right) \mu(v) f(v) dv. \end{aligned}$$

Since the kernel K satisfies (12)–(14),

$$\frac{1}{h} K\left(\frac{u-v}{h}\right) \to \delta(u-v) \int K(v)dv \text{ as } h \to 0,$$

where $\delta(.)$ is the Dirac impulse. Thus, recalling (8), we find

$$g_n(u) \to \mu(u) f(u) \int K(v)dv \text{ as } n \to 0.$$

Thus, the convergence in (15) should take place. Since the reasoning applies to (16), too, we finally obtain

$$\hat{\mu}_n(u) \to \mu(u) \text{ as } n \to \infty,$$

in an appropriate probabilistic sense, which has been desired.

Procedures (3) and (4) are just stochastic approximation, see [21]. They are, however, not classic, since

$$Y_{n+1} \frac{1}{h_n} K\left(\frac{u-U_n}{h_n}\right) \text{ and } \frac{1}{h_n} K\left(\frac{u-U_n}{h_n}\right)$$

are only asymptotically unbiased observations of $\alpha m(u) f(u) \int K(v)dv$ and $f(u) \int K(v)dv$, respectively. Unfortunately, under so small *a priori* information no unbiased estimates exist, see [24].

4. Convergence

We shall now show that our algorithm converges to $\mu(u)$.

Lemma 1 *Let $\{h_n\}$ and $\{\gamma_n\}$ be sequences of positive numbers satisfying (8) and (9)–(10), respectively. Let the Borel measurable kernel K satisfy (12)–(14). Then*

$$E\hat{f}_n(u) \to f(u) \text{ as } n \to \infty \tag{18}$$

at every point $u \in R$ at which f is continuous, whereas

$$E\hat{g}_n(u) \to \mu(u)f(u) \text{ as } n \to \infty \tag{19}$$

at every point $u \in R$ at which both m and f are continuous.

Proof. We have,

$$E\hat{g}_n(u) = (1-\gamma_n)E\hat{g}_{n-1}(u) + \gamma_n\kappa_n(u), \tag{20}$$

where, by virtue of (2),

$$\kappa_n(u) = \frac{1}{h_n}E\left\{\mu(U)K\left(\frac{u-U}{h_n}\right)\right\}. \tag{21}$$

Using (8) and Lemma 9 in Appendix C, we find

$$\kappa_n(u) \to \mu(u)f(u)\int K(v)dv \text{ as } n \to \infty$$

at every point $u \in R$ at which both m and f are continuous. An application of Lemma 5 in Appendix B yields (19). Since convergence (18) is now obvious, the lemma has been verified. ■

Lemma 2 *Let $\{h_n\}$ and $\{\gamma_n\}$ be sequences of positive numbers satisfying (8) and (9)–(11), respectively. Let the Borel measurable kernel K satisfy (12)–(14). Then*

$$\text{var}[\hat{f}_n(u)] \to 0 \text{ as } n \to \infty$$

at every point $u \in R$ at which f is continuous, whereas

$$\text{var}[\hat{g}_n(u)] \to 0 \text{ as } n \to \infty$$

at every point $u \in R$ at which both m and f are continuous.

Proof. Let u be a point at which both $f$and m are continuous. As (9) holds, with no loss of generality, we assume that $0 < \gamma_n < 1$. Since

$$\hat{g}_n(u) = (1-\gamma_n)\hat{g}_{n-1}(u) + \gamma_n Y_{n+1}\frac{1}{h_n}K\left(\frac{u-U_n}{h_n}\right), \tag{22}$$

we have

$$\begin{aligned}&\text{var}[\hat{g}_n(u)]\\ &\le (1-\gamma_n)\,\text{var}[\hat{g}_{n-1}(u)]\\ &+\gamma_n\left(\frac{\gamma_n}{h_n^2}\,\text{var}\left[Y_{n+1}K\left(\frac{u-U_n}{h_n}\right)\right]\right.\\ &\left.+\,\text{cov}\left[\hat{g}_{n-1}(u), Y_{n+1}\frac{1}{h_n}K\left(\frac{u-U_n}{h_n}\right)\right]\right).\end{aligned}$$

Thus

$$\begin{aligned}\text{var}[\hat{g}_n(u)] \le (1-\gamma_n)\,&\text{var}[\hat{g}_{n-1}(u)]\\ &+\gamma_n\left(\frac{\gamma_n}{h_n}V_n(u) + C_n(u)\right)\end{aligned}$$

with

$$V_n(u) = \frac{1}{h}\,\text{var}\left[Y_{n+1}K\left(\frac{u-U_n}{h}\right)\right]$$

and

$$C_n(u) = \text{cov}\left[\hat{g}_{n-1}(u), Y_{n+1}\frac{1}{h_n}K(\frac{u-U_n}{h_n})\right].$$

By Lemma 3 in Appendix A, $\sup_{0<h<\infty}|V_n(u)| \le \lambda(u)$ with $\lambda(u) < \infty$. To examine $C_n(u)$, observe that

$$\hat{g}_n(u) = \sum_{i=1}^{n}\delta_i(n)Y_{i+1}\frac{1}{h_i}K\left(\frac{u-U_i}{h_i}\right)$$

where

$$\delta_i(n) = \begin{cases}\gamma_i\prod_{j=i+1}^{n}(1-\gamma_j) & \text{for } i = 1,2,...,n-1\\ \gamma_n & \text{for } i = n.\end{cases} \tag{23}$$

Hence

$$\begin{aligned}C_n(u) = \sum_{i=1}^{n-1}\delta_i(n-1)\,\text{cov}\left[Y_i\frac{1}{h_i}K\left(\frac{u-U_{i-1}}{h_i}\right).\right.\\ \left.Y_{n+1}\frac{1}{h_n}K\left(\frac{u-U_n}{h_n}\right)\right].\end{aligned}$$

Applying Lemma 4 in Appendix A, we find the quantity bounded in absolute value by

$\rho(u)\sum_{i=1}^{n-1}\delta_i\,(n-1)\left\|A^{n-i}\right\|$ with finite $\rho(u)$. Since $0<\delta_i(n-1)\le\gamma_i$, $i=1,2,\cdots,n-1$, the quantity is not greater than $\rho(u)d_n$, where

$$d_n=\sum_{i=1}^{n-1}\gamma_i\|A^{n-i}\|. \tag{24}$$

Finally

$$\begin{aligned}\operatorname{var}[\hat{g}_n(u)] &\le (1-\gamma_n)\operatorname{var}[\hat{g}_{n-1}(u)]\\ &\quad+\gamma_n\left(\lambda(u)\frac{\gamma_n}{h_n}+d_n\rho(u)\right).\end{aligned} \tag{25}$$

As (9) holds and all eigenvalues of A lie in the unit circle, applying Lemma 7 in Appendix B, we find $d_n\to 0$ as $n\to\infty$. An application of Lemma 5 in Appendix B completes the proof. ■

From Lemmas 1 and 2, we obtain the following theorem.

Theorem 1 *Let $\{h_n\}$ and $\{\gamma_n\}$ be sequences of positive numbers satisfying (8) and (9)–(11), respectively. Let the Borel measurable kernel K satisfy (12)–(14), Then*

$$\hat{\mu}_n(u)\to\mu(u)\to 0 \text{ as } n\to\infty \text{ in probability}$$

at every point $u\in R$ at which both m and f are continuous and $f(u)>0$.

5. Convergence Rate

In this section, we give convergence rate of our algorithm. In the analysis, $\gamma_n=cn^{-1}$ and $h_n=dn^{-1/5}$. From (20), we get

$$b_n(u)=(1-\gamma_n)b_{n-1}(u)+\gamma_n\eta_n(u),$$

where $b_n(u)=E\hat{g}_n(u)-\mu(u)f(u)$ is the bias and $\eta_n(u)=\kappa_n(u)-\mu(u)f(u)$ with $\kappa_n(u)$ as in (21). For even K,

$$\begin{aligned}\eta_n(u)=\int K(v)\,&[\mu(u+h_nv)f(u+h_nv)\\ &-\mu(u)f(u)]\,dv.\end{aligned} \tag{26}$$

Assuming that at a point u, $u\in R$, both m and f are twice differentiable and expanding μ in a Taylor series, we find

$$\mu(u+h_nv)=\mu(u)+h_nv\mu'(u)+\frac{1}{2}h_n^2v^2\mu''(u+\theta_uh_n)$$

where $0\le\theta_u\le 1$. Since similar property holds for f, assuming that $\int v^2K(v)dv<\infty$, we find, the quantity in (26) equal $h_n^2\rho(u)+o(h_n^2)$, where

$$\begin{aligned}\rho(u)=\Big[&\mu'(u)f'(u)+\frac{1}{2}\mu(u)f''(u+\theta_uh_n)\\ &+\frac{1}{2}f(u)\mu''(u+\theta_uh)\Big]\int v^2K(v)dv.\end{aligned}$$

Assuming now that m'' and f'' are bounded in a neighborhood of u, we find $\rho(u)$ finite, for h_n small enough. Thus, for large n,

$$b_n(u)\le(1-\gamma_n)b_{n-1}(u)+\gamma_nh_n^2\rho(u).$$

For the selected γ_n and h_n, we obtain

$$b_n(u)\le(1-cn^{-1})b_{n-1}(u)+\rho(u)d^2n^{-1}n^{-2/5}.$$

Applying Lemma 6 in Appendix B, we get $b_n(u)=O(n^{-2/5})$, provided that $c>2/5$.

To examine variance, we begin with (25) and observe that d_n defined in (24) is bounded by $d\sum_{i=1}^{n-1}i^{-1}\left\|A^{n-i}\right\|$ which, by virtue of Lemma 8 in Appendix B, is not greater than d_1n^{-1}, with some d_1. Therefore

$$\begin{aligned}&\operatorname{var}[\hat{g}_n(u)]\\ &\le(1-cn^{-1})\operatorname{var}[\hat{g}_{n-1}(u)]+d_1n^{-1}n^{-4/5}\end{aligned}$$

with some d_1. Applying Lemma 6 in Appendix B, we find $\operatorname{var}[\hat{g}_n(u)]=O(n^{-4/5})$, provided that $c>4/5$. Finally $E(\hat{g}_n(u)-\mu(u)f(u))^2=O(n^{-4/5})$, provided that $c>4/5$.

We have thus shown that $\hat{g}_n(u)-\mu(u)f(u)=O(n^{-2/5})$ in probability. For similar reasons $\hat{f}_n(u)-f(u)=O(n^{-2/5})$, in probability, too. Finally

$$\hat{\mu}_n(u)-\mu(u)=O(n^{-2/5}) \text{ in probability},$$

for $\gamma_n=cn^{-1}$ with $c>4/5$ and $h_n=dn^{-1/5}$ with $d>0$, provided that $f(u)>0$. The rate is typical

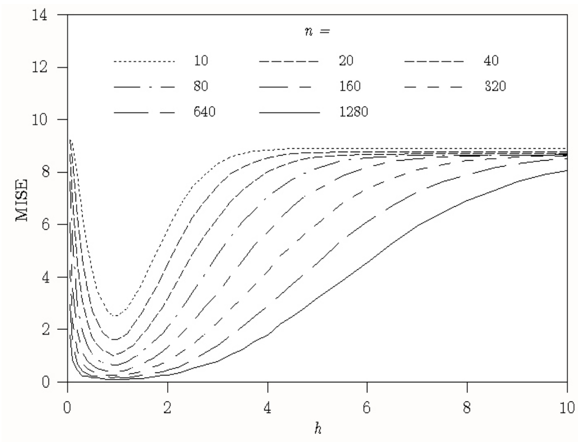

Figure 2: MISE versus h; $h_n = hn^{-0.25}$.

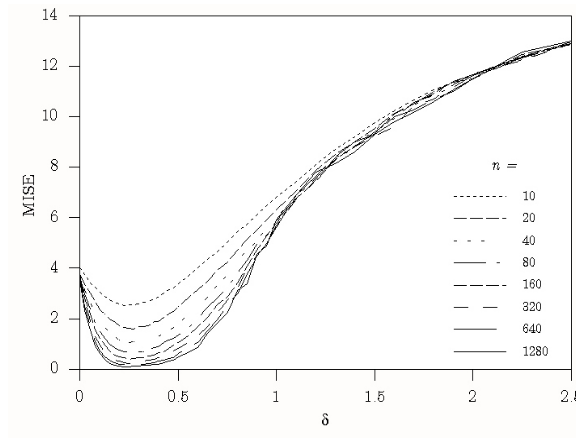

Figure 3: MISE versus δ; $h_n = n^{-\delta}$.

for kernel algorithms, see, e.g., [7–9]. It is not much worse than $n^{-1/2}$, i.e., the rate typical for algorithms designed for parametric *a priori* information.

Selecting $\gamma_n = cn^{-(1-5\varepsilon/2)}$ with $\varepsilon > 0$, and $h_n = dn^{-1/5}$, $d > 0$, we can drop the restriction $c > 4/5$ at very small loss of convergence rate. One can easily verify that now both $b_n^2(u)$ and var $[\hat{g}_n(u)]$ are of order $O(n^{-1+2\varepsilon})$. Thus,

$$\hat{\mu}_n(u) - \mu(u) = O(n^{-2/5+\varepsilon}) \text{ in probability,}$$

provided that ε, c and d are all positive. The ε can be selected arbitrarily small.

6. Simulation Example

To illustrate the behavior of our algorithm, we present results of numerical simulation in which the state vector has only one dimension and $X_{n+1} = aX_n + m(U_n)$, $Y_n = X_n + Z_n$ with m equal $0.25 + u$ or $-0.25 + u$ for $u > 0$ or $u \leq 0$, respectively. The input signal has a normal density with variance 1. The noise is also Gaussian with variance 0.1. In the estimate, $h_n = hn^{-\delta}$, $\gamma_n = cn^{-\gamma}$, while $K(u)$ equals 1 for $|u| \leq 1$ and 0 otherwise. The quality of the estimate has been measured with MISE $= \int_{-2}^{2}(\hat{\mu}_n(u) - m(u))^2 du$. In figs 2–5, $a = 0.75$. In fig. 6, a varies.

In figs. 2 and 3, $\gamma_n = n^{-0.9}$. In first $h_n = hn^{-0.25}$, in the other $h_n = n^{-\delta}$. Observe that too small h_n, i.e., too small h and δ are not recommended. Nevertheless, all curves have very wide valleys. Moreover, the larger n, the less critical problem of the proper selection of both h and δ. In turn, in figs. 4 and 5, $h_n = n^{-0.25}$. In first $\gamma_n = cn^{-0.9}$, in the other $\gamma_n = n^{-\gamma}$. The MISE is minimal for γ_n close too 1. Observe that c and γ should not be too small. For large n, the problem of the proper selection of both c and γ is less critical.

The MISE versus n, for $h_n = n^{-0.25}$ and $\gamma_n = n^{-0.9}$, is shown in fig. 6. The greater a, the greater error. Notice, however, that for a smaller than 0.75, MISE is almost independent of a, but increases rapidly for $a > 0.9$. Our algorithm has been compared with (6) and off-line (7). Results are shown in fig. 7. This time $a = 0.75$ while $h_n = n^{-0.25}$ and $\gamma_n = n^{-0.9}$. Loss incurred by recursiveness appears for large n but is very small.

7. Conclusion

The idea of the class of algorithms presented in the paper is derived from the stochastic approximation framework. As we have mentioned, a particular version (6) of our (5) has been already examined in [9,16]. Our simulation suggests that (5), in

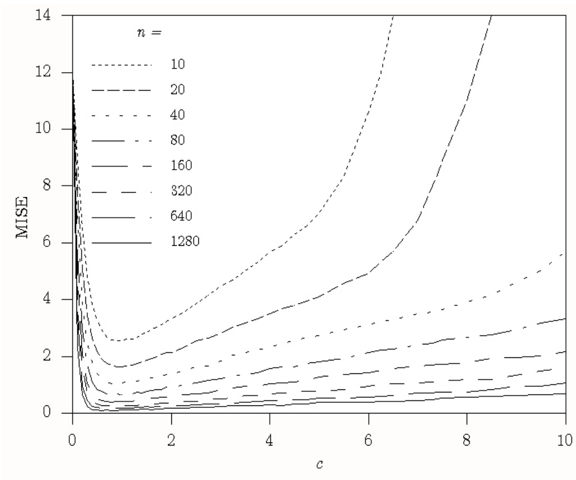

Figure 4: MISE versus c; $\gamma_n = cn^{-0.9}$.

particular (6), behaves very much like off-line (7) which means that the accuracy loss always incurred by recursiveness is small this time.

In parametric problems, nonlinear characteristics are usually polynomials of known degree and, therefore, parametric algorithms are successful only if the unknown characteristic belongs to such a narrow class of functions. Otherwise, their behavior is unclear. In this paper, the *a priori* information about the nonlinear characteristic is nonparametric, i.e., extremely small, since (1) is the only restriction. Thus, the family of all admissible nonlinearities is very wide and includes, e.g., functions which are not polynomials nor continuous. The advantage of our algorithm over parametric ones is that it recovers the nonlinear characteristic in such adverse circumstances, i.e., under such small *a priori* information.

Appendix A

Hammerstein System

In this Appendix, we deal with the Hammerstein system described in Section 2 and denote $Q = EXX^T$.

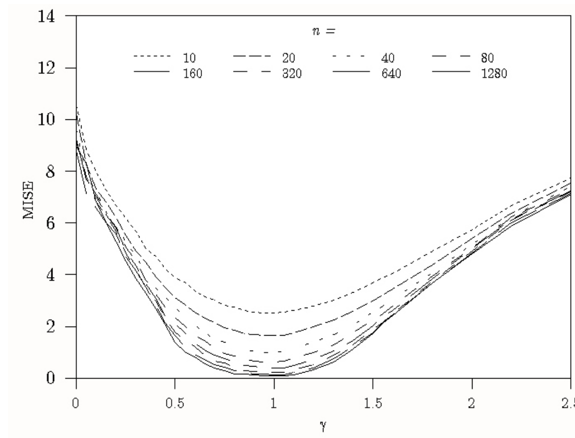

Figure 5: MISE versus γ; $\gamma_n = n^{-\gamma}$.

Lemma 3 *Let (12)–(14) hold. Then*

$$\sup_{h>0} \frac{1}{h} \operatorname{var}\left[Y_1 K\left(\frac{u-U_0}{h}\right)\right] \le \lambda(u),$$

where $\lambda(u)$ is finite at every point $u \in R$ at which both f and m are continuous.

Proof. Let u be a point at which both f and m are continuous. Since X_0 and U_0 are independent and $Y_1 = c^T AX_0 + c^T bm(U_0)$, we have

$$\begin{aligned}
&\frac{1}{h} \operatorname{var}\left[Y_1 K\left(\frac{u-U_0}{h}\right)\right] \\
&\le \frac{1}{h} E\left\{Y_1^2 K^2\left(\frac{u-U_0}{h}\right)\right\} \\
&= c^T AQA^T c\lambda_0(u;h) \\
&+ 2c^T AE\{X\}c^T b\lambda_1(u;h) + (c^T b)^2 \lambda_2(u;h)
\end{aligned}$$

with $\lambda_i(u;h) = h^{-1}E\{m^i(U)K^2(h^{-1}(u-U))\}$. As, by virtue of Lemma 10 in Appendix C, $\sup_{h>0}|\lambda_i(u;h)|$ is finite for $i = 1,2,3$, we have completed the proof. ■

Lemma 4 *Let $i \neq j$. Let (12)–(14) hold. Then*

$$\sup_{\substack{h>0\\H>0}} \left|\operatorname{cov}\left[Y_{i+1}\frac{1}{h}K\left(\frac{u-U_i}{h}\right),\right.\right.$$
$$\left.\left. Y_{j+1}\frac{1}{H}K\left(\frac{u-U_j}{H}\right)\right]\right| \le \left\|A^{|j-i|}\right\| \rho(u),$$

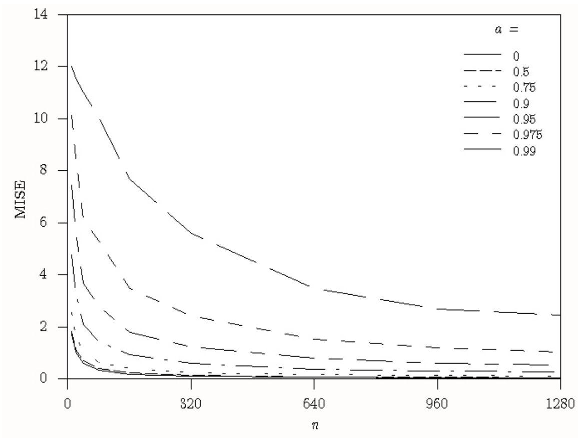

Figure 6: MISE versus n.

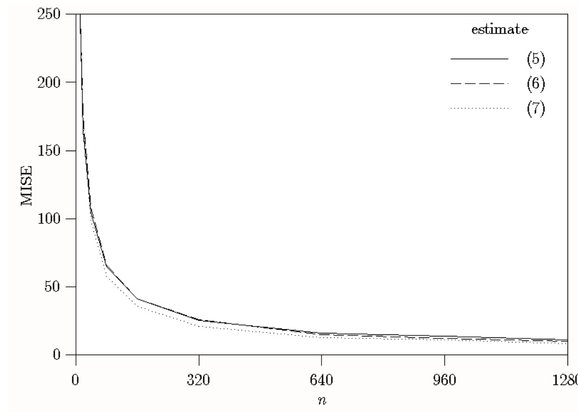

Figure 7: MISE versus n.

where ρ is finite at every point $u \in R$ at which both f and m are continuous.

Proof. Let u be a point at which both f and m are continuous. First of all, we show that for $i \neq j$, and for any Borel functions φ and ψ,

$$\begin{aligned}&\operatorname{cov}\left[X_{j+1}\varphi(U_j), X_{i+1}\psi(U_i)\right] = A^{|j-i|}E\{\psi(U)\}\\&\times\left[AQA^TE\{\varphi(U)\} + 2AE\{X\}b^TE\{m(U)\varphi(U)\}\right.\\&\left.+\, bb^TE\{m^2(U)\varphi(U)\}\right].\end{aligned} \tag{27}$$

For the sake of the simplicity, suppose that $i < j$. Since $X_{j+1} = A^{j-i}X_{i+1} + \xi_{i+1,j}$ with $\xi_{i+1,j} = \sum_{i=i+1}^{j} A^{j-i}bm(U_i)$, the examined covariance equals

$$\begin{aligned}&\operatorname{cov}\left[\xi_{i+1,j}\varphi(U_j), X_{i+1}\psi(U_i)\right]\\&+ A^{j-i}\operatorname{cov}\left[X_{i+1}\varphi(U_j), X_{i+1}\psi(U_i)\right].\end{aligned}$$

Since $\{U_j; j = \cdots, -1, 0, 1, \cdots\}$ is white noise, $\xi_{i+1,j}$ is independent of both X_{i+1} and U_i. As so is U_j, we find the first term in the above sum equal zero. The other equals $A^{j-i}E\{X_{i+1}X_{i+1}^T\varphi(U_i)\}E\{\psi(U)\}$. Because $X_{i+1} = AX_i + bm(U_i)$, we obtain

$$\begin{aligned}&E\{X_{i+1}X_{i+1}^T\varphi(U_i)\}\\&= AQA^TE\varphi(U) + 2AE\{X\}b^TE\{m(U)\varphi(U)\}\\&+ bb^TE\{m^2(U)\varphi(U)\}\end{aligned}$$

and find (27) true.

From (27), it follows that the covariance in the assertion equals

$$\begin{aligned}&c^TA^{|j-i|}\rho_1(u;h)\left[AQA^T\rho_0(u;H)\right.\\&\left.+\, 2AE\{X\}b^T\rho_1(u;H) + bb^T\rho_2(u;h)\right]c\end{aligned}$$

with $\rho_i(u;h) = E\{m^i(U)h^{-1}K(h^{-1}(u-U))\}$. Finding $\sup_{h>0}|\rho_i(u;h)| < \infty$, $i = 1, 2, 3$, see Lemma 10 in Appendix C, we complete the proof.

Appendix B

Number Sequences

Lemma 5 *Let $\{\gamma_n\}$ be a positive number sequence satisfying (9) and (10). Let, for $n = 1, 2, \cdots$,*

$$\xi_n = (1-\gamma_n)\xi_{n-1} + \gamma_n c_n$$

where $\{c_n\}$ is a sequence such that $c_n \to c$ as $n \to \infty$. Then, for any ξ_0,

$$\xi_n \to c \text{ as } n \to \infty.$$

Proof. Since (9) holds, for the sake of simplicity we assume that $0 < \gamma_n < 1$, $n = 0, 1, 2, \cdots$. We begin with the following observations:

$$\prod_{i=1}^{n}(1-\gamma_i) \to 0 \text{ as } n \to \infty. \tag{28}$$

and

$$\sum_{i=1}^{n} \delta_i(n) = 1, \tag{29}$$

where $\delta_i(n)$'s are defined by (23). Observing that $\prod_{i=1}^{n}(1-\gamma_i) \le \exp(-\sum_{i=1}^{n}\gamma_i)$ and using (10), we get (28). To verify (29) suppose that $n \ge 2$. From (23), it follows that, for $i = 1, 2, \cdots, n-1$, $\delta_i(n) = (1-\gamma_n)\,\delta_i(n-1)$. Thus $\sum_{i=1}^{n-1}\delta_i(n) = (1-\gamma_n)\sum_{i=1}^{n-1}\delta_i(n-1)$. Therefore, denoting $s_n = \sum_{i=1}^{n}\delta_i(n)$, we get $s_n = \gamma_n + (1-\gamma_n)s_{n-1}$, for $n = 2, 3, \cdots$. Noticing that $s_1 = 1$, we obtain (29).

The main part of the proof we begin with the following observation: $\xi_n = \eta_n + \omega_n$, where $\eta_n = \xi_0\prod_{i=1}^{n}(1-\gamma_i)$ and $\omega_n = \sum_{i=1}^{n}\delta_i(n)c_i$. From (28), it follows that $\eta_n \to 0$ as $n \to \infty$, any ξ_0. To examine ω_n, denote

$$\rho_i = \begin{cases} \gamma_i & \text{for } i = 1 \\ \gamma_i \Big/ \prod_{j=2}^{i}(1-\gamma_j) & \text{for } i = 2, 3, \cdots. \end{cases}$$

Owing to $0 < \gamma_n < 1$, ρ_i exists and is nonzero. As $\delta_i(n) = \rho_i\prod_{j=2}^{n}(1-\gamma_j)$, applying (29), we get $\sum_{i=1}^{n}\rho_i\prod_{j=2}^{n}(1-\gamma_j) = 1$. Hence $\delta_i(n) = \rho_i/\sum_{i=1}^{n}\rho_i$. Thus $\omega_n = \sum_{i=1}^{n}\rho_i c_i/\sum_{i=1}^{n}\rho_i$. Since $\gamma_i \le \rho_i$, $i = 1, 2, \cdots$, (10) implies $\sum_{n=1}^{\infty}\rho_n = \infty$. Therefore $\omega_n \to c$ as $n \to \infty$ which completes the proof. ∎

The lemma given below can be found in [25].

Lemma 6 *Let*

$$\xi_n = \left(1 - \frac{A}{n^\alpha}\right)\xi_{n-1} + \frac{B}{n^{\alpha+\beta}}, \tag{30}$$

where $0 < \alpha \le 1$, $0 < \beta$. *Let*

$$C = \begin{cases} A, & \text{for } 0 < \alpha < 1 \\ A - \beta, & \text{for } \alpha = 1. \end{cases}$$

If $C > 0$, *then, for any* ξ_0,

$$n^\beta \xi_n \to \frac{B}{C} \text{ as } n \to \infty.$$

Proof. Observe that $(1+n^{-1})^\beta = 1 + \beta_n n^{-1}$ with $\beta_n \to \beta$ as $n \to \infty$. Thus

$$\begin{aligned}&(1+n^{-1})^\beta(1-An^{-\alpha}) \\ &= 1 - An^{-\alpha} + \beta_n n^{-1} + A\beta_n n^{-(1+\alpha)} = 1 - C_n n^{-\alpha}\end{aligned}$$

with $C_n \to C$ as $n \to \infty$. Multiplying both sides of (30) by $(n+1)^\beta = n^\beta(1+n^{-1})^\beta$ and denoting $\eta_n = (n+1)^\beta\xi_n$, we obtain

$$\eta_n = (1-\gamma_n)\eta_{n-1} + \gamma_n B_n/C_n$$

with $B_n = B(1+n^{-1})^\beta$ and $\gamma_n = C_n/n^\alpha$. Since $B_n/C_n \to B/C$ as $n \to \infty$, an application of Lemma 5 completes the proof. ∎

Lemma 7 *Let* $|q| < 1$ *and let* $\alpha_n \to 0$ *as* $n \to \infty$. *Then*

$$\sum_{i=1}^{n}\alpha_i q^{n-i} \to 0 \text{ as } n \to \infty.$$

Lemma 8 *Let* $0 < \alpha$, *and* $0 < q < 1$. *Then*

$$\sum_{i=1}^{n}\frac{1}{i^\alpha}q^{n-i+1} \le \frac{c}{n^\alpha}$$

with some c.

Proof. We have

$$\begin{aligned}\sum_{i=1}^{n}\frac{1}{i^\alpha}q^{n-i+1} &= \frac{1}{n^\alpha}\sum_{i=1}^{n}\frac{n^\alpha}{(n+1-i)^\alpha}q^i \\ &\le \frac{1}{n^\alpha}\sum_{i=1}^{n}i^\alpha q^i \le \frac{c}{n^\alpha},\end{aligned}$$

where $c = \sum_{i=1}^{\infty}i^\alpha q^i$. The proof has been completed. ∎

Appendix C

General Results

Our next lemma is due to [26, Theorem 9.8].

Lemma 9 *Let* X *be a random variable with a density function* f *and let* φ *be a Borel measurable function such that* $E|\varphi(X)| < \infty$. *If (12)–(14) hold, then*

$$\lim_{h\to 0}\frac{1}{h}E\left\{\varphi(X)K\left(\frac{x-X}{h}\right)\right\} = \varphi(x)f(x)\int K(y)dy$$

at every point $x \in R$ *at which both* φ *and* f *are continuous.*

As a simple consequence of the lemma, we get

Lemma 10 *Let X be a random variable with a density function f and let φ be a Borel measurable function such that $E\left|\varphi(X)\right| < \infty$. If (12)–(14) are satisfied, then*

$$\sup_{h>0} \frac{1}{h} E \left| \varphi(X) K \left(\frac{x-X}{h} \right) \right| < \infty$$

at every point $x \in R$ at which both φ and f are continuous.

References

[1] L. Ljung, System Identification: Theory for the User (Prentice-Hall, Englewood Cliffs, 1987).

[2] T. Söderström & P. Stoica, System Identification (Prentice-Hall, New York, 1989).

[3] K. S. Narendra & P. G. Gallman, An iterative method for identification of nonlinear systems using a Hammerstein model, IEEE Transactions on Automatic Control, AC-11, 1966, 546–550.

[4] S. A. Billings & S. Y. Fakhouri, Theory of separable processes with applications to the identification of nonlinear systems, IEE Proceedings, 125(9), 1978, 1051–1058.

[5] J. S. Bendat, Nonlinear System Analysis and Identification from Random Data (Wiley, New York, 1990).

[6] G. B. Giannakis & E. Serpendin, A bibliography on nonlinear system identification, Signal Processing, 81, 2001, 533–580.

[7] W. Greblicki & M. Pawlak, Identification of discrete Hammerstein systems using kernel regression estimates, IEEE Transactions on Automatic Control, AC-31, 1986, 74–77.

[8] W. Greblicki & M. Pawlak, Nonparametric identification of Hammerstein systems, IEEE Transactions on Information Theory, IT-35, (2), 1989, 409–418.

[9] W. Greblicki & M. Pawlak, Recursive identification of Hammerstein systems, Journal of the Franklin Institute, 326(4), 1989, 461–481.

[10] W. Greblicki & M. Pawlak, Nonparametric identification of a cascade nonlinear time series system, Signal Processing, 22, 1991, 61–75.

[11] W. Greblicki, Nonparametric orthogonal series identification of Hammerstein systems, International Journal of Systems Science, 20(12), 1989, 2355–2367.

[12] W. Greblicki, Nonlinearity estimation in Hammerstein systems based on ordered observations, IEEE Transactions on Signal Processing, 44(5), 1996, 1224–1233.

[13] Z. Hasiewicz, Hammerstein system identification by the Haar multiresolution approximation, International Journal of Adaptive Control and Signal Processing, 13, 1999, 691–717.

[14] Z. Hasiewicz, Non-parametric estimation of nonlinearity in a cascade time-series system by multiscale approximation, Signal Processing, 81, 2001, 791–807.

[15] A. Krzyżak. An estimation of a class of nonlinear systems by the kernel regression estimate, IEEE Transactions on Information Theory, 36(1), 1990, 141–152.

[16] A. Krzyżak, Global convergence of the recursive kernel regression estimates with applications in classification & nonlinear system estimation, IEEE Transactions on Information Theory, 38(4), 1992, 1323-1338.

[17] M. Pawlak, On the series expansion approach to the identification of Hammerstein system, IEEE Transactions on Automatic Control, AC-36, 1991, 763–767.

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

[19] Z. Q. Lang, Controller design oriented model identification method for Hammerstein system, Automatica, 39(3), 1993, 767–771.

[20] Z. Q. Lang, A nonparametric polynomial identification algorithm for the Hammerstein system, IEEE Transactions on Automatic Control, 42(10), 1997, 1435–1441.

[21] M. T. Wasan, Stochastic Approximation (Cambridge University Press, Cambridge, 1969).

[22] M. P. Wand & M. C. Jones, Kernel Smoothing (Chapman & Hall, London, 1995).

printed in: International Journal of Modelling and Simulation, Vol. 24, No. 3, pp. 131-138, 2004.

[23] L. Györfi, M. Kohler, A. Krzyżak, & H. Walk, A Distribution-free Theory of Nonparametric Regression (Springer, New York, 2002).
[24] M. Rosenblatt, Remarks on some nonparametric estimates of a density function, Annals of Mathematical Statistics, 27, 1956, 832–837.
[25] V. Fabian, Stochastic approximation of minima with improved asymptotic speed, Annals of Mathematical Statistics, 38, 1967, 191–200.
[26] R. L. Wheeden & A. Zygmund, Measure and Integral (Dekker, New York, 1977).

printed in: International Journal of Modelling and Simulation, Vol. 24, No. 3, pp. 131-138, 2004.