Available online at www.sciencedirect.com
SCIENCE DIRECT®
Automatica 41 (2005) 1905–1916
automatica
www.elsevier.com/locate/automatica

Strong consistence of recursive identification for Wiener systems[☆]

Xiao-Li Hu[a, b], Han-Fu Chen[a, *]

[a]Key Laboratory of Systems and Control, Institute of Systems Science, AMSS Chinese Academy of Sciences, Beijing 100080, PR China
[b]Graduate School of the Chinese Academy of Sciences, Beijing 100039, PR China

Received 17 December 2004; received in revised form 18 April 2005; accepted 11 June 2005

Abstract

The paper concerns identification of the Wiener system consisting of a linear subsystem followed by a static nonlinearity $f(\cdot)$ with no invertibility and structure assumption. Recursive estimates are given for coefficients of the linear subsystem and for the value $f(v)$ at any fixed v. The main contribution of the paper consists in establishing convergence with probability one of the proposed algorithms to the true values. This probably is the first strong consistency result for this kind of Wiener systems. A numerical example is given, which justifies the theoretical analysis.

Keywords: Wiener system; Nonparametric nonlinearity; Recursive estimate; Strong consistency; Stochastic approximation

1. Introduction

The Hammerstein and Wiener systems, in particular, their identification issue have attracted a great attention from researchers because of their importance in applications. Since these systems are nonlinear, the identification methods demonstrated by Chen & Guo (1991) and Ljung (1987) are not directly applicable.

A linear system cascaded with a static nonlinearity is called the Wiener (or Hammerstein) system if the nonlinearity follows (or is followed by) the linear subsystem. This paper concerns with identification of the SISO Wiener system presented in Fig. 1 where u_k is the one-dimensional system input to be designed, v_k is the output of the linear subsystem serving as the input of the memoryless nonlinear block, and y_k is the system output which is observed with additive noise ε_k. The coefficients of the linear subsystem and the nonlinear function $f(\cdot)$ are unknown. The problem is how to estimate coefficients contained in the linear subsystem and the static nonlinearity $f(\cdot)$ on the basis of observation $\{z_k\}$ and the adequately designed input $\{u_k\}$, where

$$z_k = y_k + \varepsilon_k. \tag{1}$$

The name Wiener model probably comes from the famous book by Wiener (1958), where the nonlinearity is expanded to the functional series and the correlation analysis is carried out by using the Gaussian input. Based on the method proposed by Wiener (1958) there were many works on analysis and identification of nonlinear systems in 1960s and 1970s. Among early works on identification of Wiener systems, a practical nonparametric algorithm is proposed by Billings & Fakhouri (1978) where no inversion of the nonlinearity is required.

For characterizing the nonlinearity the parametric approach (Bendat, 1999; Hasiewicz, 1987; Hunter & Korenberg, 1986; Nordsjö & Zetterberg, 2001; Pajunen, 1992; Verhaegen & Westwick, 1996; Vörös, 2001; Westwick & Kearney, 1992; Wigren, 1993, 1994) is mostly applied in literature, but the nonparametric approach is also considered (Billings & Fakhouri, 1978; Greblicki, 1997, 2001).

[☆] This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor Johan Schoukens under the direction of Editor Torsten Soederstroem. This work is supported by NSFC (Projects G0221301, 60334040, 60474004.

[*] Corresponding author. Tel.: +86 10 62579540; fax: +86 10 62587343.
E-mail addresses: xlhu@amss.ac.cn (X.-L. Hu), hfchen@iss.ac.cn (H.-F. Chen).

0005-1098/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2005.06.006

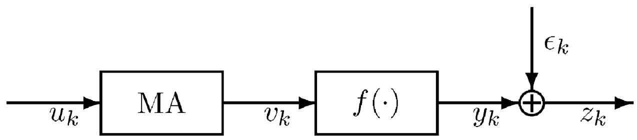

Fig. 1. Wiener system.

When the parametric approach is applied, the nonlinearity is presented either as a linear combination of known functions with unknown coefficients (Hasiewicz, 1987; Hunter & Korenberg, 1986; Nordsjö & Zetterberg, 2001; Westwick & Kearney, 1992) or as a piecewise linear function (Pajunen, 1992; Vörös, 2001; Wigren, 1993). In this case, the parameter estimates may be derived by minimizing some specially designed loss function, and this can be realized by using any optimizing algorithm for data with fixed sample size. Proceeding in this way, the parameters cannot be updated online as can be seen in Vörös (2001). Nevertheless, the estimates may still be made recursive and even with certain kind of convergency, if rather restrictive conditions are imposed as demonstrated by Wigren (1993, 1997, 1998) the nonlinear function is assumed to be known.

When the nonparametric approach is considered, the nonlinear function is usually required to be invertible (Greblicki, 1997, 2001), and the argument v for any given $u = f(v)$ rather than $f(v)$ for any given v is estimated. This may limit applications of corresponding identification methods in practice by the following consideration: saturations are not invertible, but they quite often exist in practical systems and affect the measured outputs; also, inversion of the nonlinearity can lead to severe amplification of possible measurement disturbances as pointed out by Wigren (1993), etc.

The goal of this paper is to recursively estimate the coefficients of the linear subsystem and the value $f(v)$ for any given v without requiring invertibility of $f(\cdot)$. The estimates are required to be strongly consistent, i.e., to converge to the true values with probability one. A similar problem for Hammerstein systems is solved by Chen (2004) by using stochastic approximation (SA) algorithms with expanding truncations (Chen, 2002). There the input is designed to be a sequence of bounded iid random variables, and $f(\cdot)$ is estimated with the help of a kernel function applied to the SA algorithm.

Let us explain why SA is an appropriate tool to deal with the identification problem. When estimating an unknown parameter ϑ on the basis of observation data denoted by $\{g_k\}$, one can always transform this to a SA problem, i.e., to a root-seeking problem for any function $g(\cdot)$ with root ϑ, e.g., $g(x) = -(x - \vartheta)$. This is because g_{k+1} can always be written as $g_{k+1} = g(x_k) + \eta_{k+1}$ with $\eta_{k+1} \triangleq g_{k+1} - g(x_k)$, where x_k denotes the kth estimate for ϑ. In other words, the observation data $\{g_k\}$ can be viewed as a noisy observation on $g(x_k)$ with additive noise η_k.

It is natural to come to the idea: to solve the stated problem for Wiener systems by using SA algorithms with expanding truncations and with kernel functions. However, in doing so, there is an essential difference in analysis for Wiener systems from that for Hammerstein systems. To explain this, we note that the analysis given by Chen (2004) is essentially based on two facts: (1) The correlation function between the input and output of the system has a simple analytic expression connecting parameters to be estimated; (2) All signals passing through the system are bounded when the input is bounded. As shown by Chen (2004), for Hammerstein systems a sequence of bounded iid random variables serving as the system input results in these two properties.

For Wiener systems, though a bounded input still implies the boundedness of all signals in the system, the correlation function between the input and output of the system, in general, does not have a simple analytic expression. This hints us to take an iid Gaussian random variables to serve as the system input. However, the Gaussian random variable is unbounded, and hence the Gaussian input may give rise to the unboundedness of signals in the system. This explains why the analysis method given by Chen (2004) cannot directly be applied to the present case.

The requirement for boundedness of signals passing through the system can also be explained by the following intuitive observation. To estimate $f(v)$ it is important to recover the input v_k of the nonlinear function. The estimate for v_k, denoted by $\hat{v}_k$, is obtained as the output of the estimated linear subsystem, which means the subsystem with coefficients replaced by their estimates. However, $\hat{v}_k$ may not be close to v_k even if the estimates for coefficients of the linear subsystem are sufficiently accurate, when $\{u_k\}$ is unbounded.

To overcome this difficulty, we proceed as follows. While the system input $\{u_k\}$ is taken to be a sequence of iid Gaussian random variables, not all u_k but only such $r + 1$ successive u_k that are bounded by a given constant are used to estimate v_k, where r is the order of the linear subsystem. This selection guarantees that $\{v_k\}$ generated by sets of $r + 1$ successive bounded u_k is bounded. Since the selection depends on sample paths, we have to use the concept of stopping time, which is well developed in probability theory (see, e.g., Chow & Teicher (1978)).

The rest of the paper is organized as follows. The system considered in the paper and conditions imposed on the system are given in Section 2. Also, the basic results of SA used in the paper are described in Section 2. The recursive identification algorithms and their strong consistency for estimating the linear subsystem and the nonlinear block are, respectively, presented in Sections 3 and 4. A numerical example is demonstrated in Section 5 and some concluding remarks are given in Section 6. The mathematical details concerning the properties of stopping times and the behaviors of kernel functions are given in Appendix.

2. Preliminaries

Let us first describe the system more precisely. Assume the linear subsystem is given by

$$v_{k+1} = \sum_{j=0}^{r} d_j u_{k-j}, \quad d_0 = 1 \tag{2}$$

and the output of the nonlinearity is

$$y_k = f(v_k). \tag{3}$$

It is worth noting that d_0 is not necessary equal to 1 but is allowed to be any known constant. The reason to assume d_0 known is technical, because otherwise there is a lack of one equation.

The coefficients $d_1, \ldots, d_r$ and the value $f(v)$ at any fixed $v \in \mathscr{R}$ are to be recursively estimated on the basis of system inputs $\{u_k\}$ and measurements $\{z_k\}$ given by (1).

As explained in Introduction we take u_k to be Gaussian. Let us precisely formulate this as condition H0.

H0. $\{u_k\}$ is a sequence of iid Gaussian random variables: $u_k \in \mathscr{N}(0, 1)$, and $\{u_k\}$ is independent of the observation noise $\{\varepsilon_k\}$.

In addition to H0 the following conditions H1 and H2 are also imposed on the Wiener system under consideration.

H1. $f(\cdot)$ is a measurable function and continuous at v where the value $f(v)$ is estimated. The growth rate of $f(v)$ as $|v| \to \infty$ is not faster than a polynomial.

H2. $\{\varepsilon_k\}$ is a sequence of iid random variables with $E\varepsilon_k=0$ and $E\varepsilon_k^2 < \infty$.

It is noted that no invertibility of $f(\cdot)$ is required.

The recursive estimates are to be generated by SA algorithms, but the classical Robbins–Monro (RM) algorithm does not work here because of its restriction conditions for applicability. In order to ease reading, we cite a general convergence theorem (GCT) for SA algorithms with expanding truncations in the case of single root. For its proof we refer to Chen (2002, Theorem 2.2.1).

The root of a function $g(\cdot)$ is denoted by x^0 if it is single. The problem of SA is to seek the root of $g(\cdot)$ on the basis of its noisy observations $\{g_k\}$:

$$g_{k+1} \triangleq g(x_k) + \eta_{k+1}, \tag{4}$$

where x_k is the kth estimate for x^0 and η_{k+1} is the observation noise.

Let $\{a_k\}$ with $a_k > 0$ be the sequence of stepsizes. By the classical RM algorithm the root x^0 $(g(x^0) = 0)$ is sought by the simple recursion:

$$x_{k+1} = x_k + a_k g_{k+1}, \quad k = 0, 1, \ldots. \tag{5}$$

For convergence of x_k to x^0, a certain growth rate restriction on $g(\cdot)$ or an a priori boundedness assumption on $\{x_k\}$ are required in the classical theory, in addition to conditions on the noise $\{\eta_k\}$.

In order to extend the range of applicability of SA algorithms, the RM algorithm (5) is modified by expanding truncations in Chen (2002). Let $\{M_k\}$ be a sequence of positive real numbers $M_{k+1} > M_k$ and $M_k \underset{k\to\infty}{\longrightarrow} \infty$. Let x_k be generated by the following algorithm:

$$x_{k+1} = \begin{cases} x_k + a_k g_{k+1} & \text{if } \|x_k + a_k g_{k+1}\| \leqslant M_{\sigma_k}, \\ x^* & \text{otherwise}, \end{cases} \tag{6}$$

$$\sigma_k = \sum_{j=1}^{k-1} I_{[\|x_j + a_j g_{j+1}\| > M_{\sigma_j}]}, \quad \sigma_0 = 0. \tag{7}$$

From (7) it is seen that $\sigma_k \leqslant k - 1$, and from (6) $\|x_{k+1}\| \leqslant \|x^*\| \wedge M_{k-1}$. This means that the growth rate of $\|x_{k+1}\|$ is controlled: it should not be faster than M_k. At any time, if $\|x_k + a_k g_{k+1}\|$ exceeds the truncation bound M_{σ_k}, then we pull x_{k+1} back to the fixed point x^* and simultaneously extend the truncation bound from M_{σ_k} to M_{σ_k+1}. Otherwise, it develops as (5). Since the region where x^0 is located is unknown, it is important to allow x_k to grow up in order to have possibility to reach x^0.

General convergence theorem (*GCT for the single root case*). *Assume the following conditions.*

(1) $a_k > 0, a_k \underset{k\to\infty}{\longrightarrow} 0$ *and* $\sum_{k=1}^{\infty} a_k = \infty$, $M_k > 0$, $M_{k+1} > M_k$, $M_k \underset{k\to\infty}{\longrightarrow} \infty$.

(2) $g(\cdot)$ *is measurable and locally bounded.*

(3) *There is a continuously differentiable function* $v(\cdot)$ *such that*

$$\sup_{c' \leqslant \|x - x^0\| \leqslant c''} g^T(x) v_x(x) < 0$$

for any $c'' > c' > 0$, *where* $v_x(x)$ *denotes the gradient of* $v(\cdot)$. *Further,* x^* *used in* (6) *is such that* $\|x^*\| < c_0$ *and* $v(x^*) < \inf_{\|x\|=c_0} v(x)$ *for some* $c_0 > 0$.

Then x_k *defined by* (6) *and* (7) *converges to* x^0 *for any sample paths for which the following condition* (4) *is satisfied*:

(4)

$$\lim_{T\to 0} \limsup_{k\to\infty} \frac{1}{T} \left\| \sum_{i=n_k}^{m(n_k, T_k)} a_i \eta_{i+1} \right\| = 0 \quad \forall T_k \in [0, T] \tag{8}$$

for any n_k *such that* x_{n_k} *converges, where*

$$m(k, T) = \max\left\{ m : \sum_{j=k}^{m} a_j \leqslant T \right\}.$$

It is worth noting that condition (8) is required to verify only along convergent subsequences $\{x_{n_k}\}$ rather than along the whole sequence $\{x_n\}$. As pointed by Chen (2002), in many cases (8) cannot be directly verified along the whole sequence $\{x_n\}$, but it can be done for convergent subsequences. This is the case in Lemma 2 and Theorem 2 of Section 4. The method verifying condition (8) along convergent subsequences has shown a great advantage over verification along the whole sequence, and it is called as trajectory-subsequence (TS) method in Chen (2002).

The convergence analysis to be carried out in Sections 3 and 4 is based on GCT, and also on the convergence rate theorem (CRT), which is presented below.

Convergence rate theorem (*CRT*). *Assume condition* (3) *in GCT holds and the following conditions* (5)–(7) *are satisfied*:

(5) $a_k > 0$, $a_k \underset{k\to\infty}{\longrightarrow} 0$, *and* $\sum_{k=1}^{\infty} a_k = \infty$, $\frac{1}{a_{k+1}} - \frac{1}{a_k} \underset{k\to\infty}{\longrightarrow} \gamma \geqslant 0$; $M_k > 0$, $M_{k+1} > M_k$, $M_k \underset{k\to\infty}{\longrightarrow} \infty$.

(6) *For the sample path ω under consideration the observation noise $\{\eta_k\}$ in* (4) *can be decomposed into two parts $\eta_k = \eta_k' + \eta_k''$ such that*

$$\sum_{k=1}^{\infty} a_k^{1-\delta}\eta_k' < \infty, \quad \eta_k'' = O(a_k^{\delta})$$

for some $\delta \in (0, 1]$.

(7) *$g(\cdot)$ is measurable and locally bounded, and is differentiable at x^0 such that as $x \to x^0$*

$$g(x) = F(x - x^0) + \Delta(x), \quad \Delta(x^0) = 0,$$
$$\Delta(x) = o(\|x - x^0\|),$$

and the matrix $F + \gamma\delta I$ is stable, where γ and δ are given above.

Then x_k given by (4), (6) *and* (7) *converges to x^0 with the following convergence rate*:

$$\|x_k - x^0\| = o(a_k^{\delta}),$$

where δ is given in Condition (5).

For the proof of the theorem we refer to Chen (2002, Theorem 3.1.1).

It may be worth paying attention to the difference in numbering conditions: conditions (1)–(7) are used in GCT and CRT, while conditions H0–H4 are used for the Wiener system.

3. Estimation for linear subsystem

We now use the algorithm (6) and (7) to estimate coefficients d_i, $i = 1, \ldots, r$ in (2). For this we first concretize $\{a_k\}$ and the truncation bounds $\{M_k\}$.

Define

$$a_k = \frac{1}{k} \quad \text{and} \quad M_k = M_0 + k^{\beta} \tag{9}$$

to serve as the stepsizes and truncation bounds to be used in the SA algorithms, where $\beta \in (0, 1/2 - 3\alpha)$, $\alpha \in (0, 1/6)$, and M_0 is a positive number.

Remark. The choice of a_k, M_k, and b_k in (27) has some flexibility: $a_k = 1/k$ is a conventional choice of stepsizes but a_k may be different from $\{1/k\}$ (see Chen (2002)). Here we select $a_k = 1/k$ only for simplicity of description. The selection of $\{a_k\}$ should be coordinated with $\{b_k\}$ given by (27) and $\{M_k\}$, and the rate selection for $\{M_k\}$ is used in the proof of Lemma 1 to upper bound the maximal divergence rate of $|\mu_{\tau_k}(v)|$ in (33).

The algorithms for estimating coefficients of the linear subsystem are the same as those used by Chen (2004), i.e., they are the SA algorithms with expanding truncations

$$\theta_{k+1}(i) = \theta_k(i) - a_k(\theta_k(i) - u_k z_{k+i+1}) \times I_{[|\theta_k(i) - a_k(\theta_k(i) - u_k z_{k+i+1})| \leqslant M_{\sigma_k(i)}]}, \tag{10}$$

$$\sigma_k(i) = \sum_{j=1}^{k-1} I_{[|\theta_j(i) - a_j(\theta_j(i) - u_j z_{j+i+1})| > M_{\sigma_j(i)}]}, \quad \sigma_0(i) = 0 \tag{11}$$

with initial values $\theta_0(i)$, $i = 0, 1, \ldots, r$. Here $\theta_k(0)$ is used to estimate $\rho \triangleq (1/\sigma_v^2)E[v_k f(v_k)]$, and $\theta_k(i)$ for ρd_i, $i = 1, \ldots, r$, where σ_v^2 denotes the variance of v_k.

It is clear that $(\theta_k(i)/\theta_k(0)) \triangleq d_{ik}$ may serve as the estimate for d_i at time k whenever $\theta_k(0) \neq 0$.

Theorem 1. *Assume* H0–H2 *hold. Then*

$$\theta_k(0) \underset{k\to\infty}{\longrightarrow} \rho \triangleq \frac{E(v_j f(v_j))}{\sigma_v^2} \quad a.s. \tag{12}$$

and

$$\theta_k(i) \underset{k\to\infty}{\longrightarrow} \rho d_i \quad a.s., \quad i = 1, \ldots, r \tag{13}$$

with rates of convergence

$$|\theta_k(0) - \rho| = o(k^{-\varsigma}) \text{ a.s.}, \quad |\theta_k(i) - \rho d_i| = o(k^{-\varsigma}) \ a.s., \quad \varsigma \in \left(0, \tfrac{1}{2}\right), \quad i = 1, \ldots, r. \tag{14}$$

Proof. The proof is essentially based on the convergence theorems GCT and CRT of SA given in Section 2.

Noticing that $\{u_k\}$ is Gaussian with zero mean and iid, from

$$E(u_k v_{k+i+1}) = E\left(u_k \sum_{j=0}^{r} d_j u_{k+i-j}\right) = d_i,$$

we conclude that

$$E\left[\left(u_k - \frac{d_i v_{k+i+1}}{\sigma_v^2}\right) v_{k+i+1}\right] = E(u_k v_{k+i+1}) - \frac{d_i}{\sigma_v^2} E(v_{k+i+1}^2) = 0,$$

which implies $u_k - (d_i v_{k+i+1}/\sigma_v^2)$ is uncorrelated with and hence is independent of v_{k+i+1}, since for Gaussian random variables independence is equivalent to uncorrelatedness. From independence it follows that

$$E\left(u_k - \frac{d_i v_{k+i+1}}{\sigma_v^2}\middle| v_{k+i+1}\right) = 0,$$

which implies

$$E(u_k|v_{k+i+1}) = \frac{d_i v_{k+i+1}}{\sigma_v^2}.$$

Consequently, we have

$$E(u_k y_{k+i+1}) = E\left[E(u_k y_{k+i+1}|v_{k+i+1})\right] = E\left[y_{k+i+1}\frac{d_i}{\sigma_v^2}v_{k+i+1}\right] = d_i\rho. \tag{15}$$

The recursion (10) can be rewritten as

$$\theta_{k+1}(i) = \theta_k(i) - a_k(\theta_k(i) - d_i\rho) - a_k\bar{\varepsilon}_{k+1}(i) \times I_{[|\theta_k(i)-a_k(\theta_k(i)-d_i\rho)-a_k\bar{\varepsilon}_{k+1}(i)|\leqslant M_{\sigma_k}(i)]}, \tag{16}$$

where

$$\bar{\varepsilon}_{k+1}(i) = -u_k z_{k+i+1} + d_i\rho, \quad i = 0, 1, \ldots, r. \tag{17}$$

Comparing (16) with (4) and (6), we find that $g(x)$ in (4) corresponds to the linear function $-(x - d_i\rho)$ in (16). In other words, the algorithm (16) and (11) seeks for the root of the linear function $g^{(i)}(x) \triangleq -(x - d_i\rho)$, $i = 0, 1, \ldots, r$. It is also noticed that x^* in (6) corresponds to 0 in (16).

It is clear that for the linear function $-(x - d_i\rho)$, $v(x) = (x - d_i\rho)^2$ satisfies condition (3) in GCT, while (1) and (2) in GCT obviously hold in the present case. Thus, by GCT given in Section 2, for (12) and (13) it suffices to show

$$\lim_{T\to 0}\limsup_{n\to\infty}\frac{1}{T}\left|\sum_{j=n}^{m(n,t)} a_j\bar{\varepsilon}_{j+1}(i)\right| = 0 \quad \forall t \in [0, T], \quad i = 0, 1, \ldots, r, \tag{18}$$

where

$$m(n, t) = \max\left\{m : \sum_{j=n}^{m}\frac{1}{j} \leqslant t\right\}. \tag{19}$$

Notice that

$$\bar{\varepsilon}_{k+1}(i) = -(u_k y_{k+i+1} - d_i\rho) - u_k\varepsilon_{k+i+1}. \tag{20}$$

By (15), $\{u_{j+k(r+1)}y_{j+k(r+1)+i+1} - d_i\rho,\ k = 0, 1, \ldots\}$ is a sequence of iid random variables for any fixed $i : i = 1, \ldots, r$, and by the convergence theorem for independent random variables (Chow & Teicher, 1978) we find

$$\sum_{j=1}^{\infty} a_j(u_j y_{j+i+1} - d_i\rho) = \sum_{j=1}^{r+1}\sum_{k\in A_j} a_k(u_k y_{k+i+1} - d_i\rho) < \infty \quad \text{a.s.}, \tag{21}$$

where $A_j = \{j + k(r+1) : k = 0, 1, \ldots\}$, $j = 1, 2, \ldots, r+1$. From (20) and (21) and the independence of $\{u_k\}$ and $\{\varepsilon_k\}$ we conclude

$$\sum_{j=1}^{\infty} a_j\bar{\varepsilon}_{j+i+1}(i) < \infty \quad \text{a.s.} \tag{22}$$

which implies (18).

The rates of convergence are derived from CRT given in Section 2 for the case where $\gamma = 1$, $F = -1$, $\varepsilon''_k = 0$. $\square$

4. Estimation for $f(v)$

As explained in Introduction, the input of the nonlinear part is estimated by

$$\hat{v}_k \triangleq u_{k-1} + d_{1k}u_{k-2} + \cdots + d_{rk}u_{k-r-1}. \tag{23}$$

It is conceivable to estimate $f(v)$ on the basis of $\{\hat{v}_k\}$ and $\{z_k\}$ by using SA algorithm with kernel functions as done in Chen (2004). However, $\hat{v}_k$ is, in general, unbounded because of the unboundedness of $\{u_k\}$, and hence we will use only such $r+1$ successive u_k that are bounded by a given constant c to estimate $f(v)$. To be precise, we introduce a sequence of Markov times as follows. Fix a positive constant $c > |v|$ and define

$$\tau_1 = \inf\{k > r + 1 : \quad |u_{k-j}| \leqslant c, \quad j = 1, \ldots, r + 1\} \tag{24}$$

and

$$\tau_k = \inf\{i > \tau_{k-1} : \quad |u_{i-j}| \leqslant c, \quad j = 1, \ldots, r + 1\}, \quad k = 2, 3, \ldots . \tag{25}$$

From (24) and (25) we see that $\tau_k > \tau_{k-1}$, and starting from u_{τ_k-r-1} there are $r + 1$ successive u_i with $|u_i| \leqslant c$, $i = \tau_k - r - 1, \ldots, \tau_k - 1$. Thus, among $\hat{v}_i$ given by (23) only those with $i = \tau_k$ are used to estimate $f(v)$. It is clear that the set $[\tau_k = s]$ is completely determined by random variables $u_1, \ldots, u_{s-1}$. In other words, $[\tau_k = s] \in \mathscr{F}_s \triangleq \sigma\{u_1, \ldots, u_{s-1}\}$, the σ-algebra generated by $\{u_1, \ldots, u_{s-1}\}$. We recall that the nonnegative random variable τ is called the Markov time with respect to the family of nondecreasing σ-algebras $\mathscr{F}_s$ if $[\tau_k = s] \in \mathscr{F}_s$, $\forall s = 1, 2, \ldots$. Further, if a Markov time $\tau < \infty$ a.s., then τ is called the stopping time with respect to $\{\mathscr{F}_s\}$. Consequently, τ_k, $k = 1, 2, \ldots$ are the Markov times.

It is worth noting that

$$|v_{\tau_k}| = |u_{\tau_k-1} + d_1u_{\tau_k-2} + \cdots + d_r u_{\tau_k-r-1}| \leqslant c\sum_{i=0}^{r}|d_i|, \quad k = 1, 2, \ldots . \tag{26}$$

Let

$$b_k = \frac{1}{k^{\alpha}}, \tag{27}$$

where α is figured in the definition of β appearing in (9).

For a fixed $v \in \mathscr{R}$, define the kernel function

$$w_{\tau_k} \triangleq \frac{1}{b_k} e^{-((v_{\tau_k}-v)/b_k)^2+\frac{1}{2}\sum_{j=1}^{r+1} u^2_{\tau_k-j}}, \tag{28}$$

and its estimate

$$\hat{w}_{\tau_k} \triangleq \frac{1}{b_k} e^{-((\hat{v}_{\tau_k}-v)/b_k)^2+\frac{1}{2}\sum_{j=1}^{r+1} u^2_{\tau_k-j}}, \tag{29}$$

which is used in the SA algorithm for estimating $f(v)$:

$$\mu_{k+1}(v) = \mu_k(v) - a_k \hat{w}_{\tau_k}(\mu_k(v) - z_{\tau_k}) \times I_{[|\mu_k(v)-a_k\hat{w}_{\tau_k}(\mu_k(v)-z_{\tau_k})|\leqslant M_{\lambda_k(v)}]}, \tag{30}$$

$$\lambda_k(v) = \sum_{j=1}^{k-1} I_{[|\mu_j(v)-a_j\hat{w}_{\tau_j}(\mu_j(v)-y_{\tau_j})|>M_{\lambda_j(v)}]},$$
$$\lambda_0(v) = 0 \tag{31}$$

with an initial value $\mu_0(v)$.

It is worth noting that if all signals in the system were bounded, then it would be unnecessary to introduce τ_k and the analysis carried out in this section would be much simpler. As a matter of fact, in this case all τ_k in what follows could be replaced by k and Lemmas A and B in Appendix would no longer be needed, while Lemmas C and D could be proved much simpler.

We intend to show $\mu_k(v)$ a.s. converges to $f(v)$. For this we first prove lemmas.

Lemma 1. *Assume* H0–H2 *hold and* $\rho \neq 0$. *Then*

$$\sum_{k=1}^{\infty} a_k \left|(\hat{w}_{\tau_k} - w_{\tau_k})(\mu_k(v) - z_{\tau_k})\right| < \infty \quad \text{a.s.}, \tag{32}$$

where w_{τ_k}, $\hat{w}_{\tau_k}$ *and* $\mu_k(v)$ *are given by* (28), (29) *and* (30), *respectively*.

Proof. Let $\delta \in (0, \frac{1}{2} - 3\alpha - \beta)$, then $\frac{3}{2} - 3\alpha - \beta - \delta > 1$. By the boundedness of $\{\hat{v}_{\tau_k}\}$, (14), and the fact $(\tau_k/k) \underset{k\to\infty}{\longrightarrow} E\tau$ (see Lemma A in Appendix), it follows that

$$\begin{aligned}
&\left(\frac{v_{\tau_k}-v}{b_k}\right)^2 - \left(\frac{\hat{v}_{\tau_k}-v}{b_k}\right)^2 \\
&\quad = \frac{1}{b_k^2}(v_{\tau_k} + \hat{v}_{\tau_k} - 2v)(v_{\tau_k} - \hat{v}_{\tau_k}) \\
&\quad = o\left(k^{-(1/2-2\alpha-\delta)}\right) \underset{k\to\infty}{\longrightarrow} 0 \quad \text{a.s.}
\end{aligned}$$

From this we have

$$\begin{aligned}
\hat{w}_{\tau_k} - w_{\tau_k} &= \frac{1}{b_k}[e^{-((\hat{v}_{\tau_k}-v)/b_k)^2} - e^{-((v_{\tau_k}-v)/b_k)^2}] \\
&= \frac{1}{b_k} e^{-((v_{\tau_k}-v)/b_k)^2} \\
&\quad \times [e^{((v_{\tau_k}-v)/b_k)^2 - ((\hat{v}_{\tau_k}-v)/b_k)^2} - 1] \\
&= o(k^{-(1/2-3\alpha-\delta)}) \quad \text{a.s.},
\end{aligned}$$

or there is $0 < M < \infty$, which may depend on samples, such that

$$|\hat{w}_{\tau_k} - w_{\tau_k}| \leqslant M \frac{1}{k^{1/2-3\alpha-\delta}} \quad \text{a.s.} \quad \forall k.$$

Therefore, taking notice of $M_k = M_0 + k^\delta$ and $M_{\lambda_{k-1}(v)} \leqslant M_0 + \left(\lambda_{k-1}(v)\right)^\beta \leqslant M_0 + (k-2)^\beta \leqslant (M_0+1)k^\beta$, we have

$$\begin{aligned}
&\sum_{k=1}^{\infty} a_k |(\hat{w}_{\tau_k} - w_{\tau_k})\mu_k(v)| \leqslant \sum_{k=1}^{\infty} a_k |(\hat{w}_{\tau_k} - w_{\tau_k})| \cdot |\mu_k(v)| \\
&\quad \leqslant M(M_0+1) \sum_{k=1}^{\infty} \frac{1}{(k)^{3/2-3\alpha-\delta}} M_{\lambda_{k-1}(v)} \\
&\quad \leqslant M(M_0+1) \sum_{k=1}^{\infty} \frac{1}{(k)^{3/2-3\alpha-\beta-\delta}} < \infty \quad \text{a.s.}
\end{aligned} \tag{33}$$

Since $\{y_{\tau_k}\}$ is bounded by H1 and (26), we have

$$\begin{aligned}
&\sum_{l=1}^{\infty} a_k |(\hat{w}_{\tau_k} - w_{\tau_k}) y_{\tau_k}| \\
&\quad \leqslant \sum_{l=1}^{\infty} a_k |(\hat{w}_{\tau_k} - w_{\tau_k})| \cdot |y_{\tau_k}| \\
&\quad \leqslant M' \sum_{l=1}^{\infty} \frac{1}{k^{3/2-3\alpha-\delta}} < \infty \quad \text{a.s.},
\end{aligned} \tag{34}$$

where $0 < M' < \infty$.

Further, by (A18) in Appendix it follows that

$$\begin{aligned}
&\sum_{l=1}^{\infty} a_k |(\hat{w}_{\tau_k} - w_{\tau_k})\varepsilon_{\tau_k}| \\
&\quad \leqslant \sum_{l=1}^{\infty} a_k |(\hat{w}_{\tau_k} - w_{\tau_k})| \cdot |\varepsilon_{\tau_k}| \leqslant \sum_{l=1}^{\infty} \frac{M}{k^{3/2-3\alpha-\delta}} |\varepsilon_{\tau_k}| \\
&\quad = \sum_{l=1}^{\infty} \frac{M(|\varepsilon_{\tau_k}| - E|\varepsilon_{\tau_k}|)}{(k)^{3/2-3\alpha-\delta}} + \sum_{l=1}^{\infty} \frac{ME|\varepsilon_{\tau_k}|}{k^{3/2-3\alpha-\delta}} < \infty \quad \text{a.s.}
\end{aligned} \tag{35}$$

Combining (33), (34) and (35) implies (32). □

Lemma 2. *Assume* H0–H2 *hold and* $\rho \neq 0$. *Then there is an* Ω_0 *with* $P\Omega_0 = 1$ *such that for any fixed sample path* $\omega \in \Omega_0$ *if* $\mu_{k_i}(v)$ *is a convergent subsequence of* $\{\mu_k(v)\}$, *then*

$$\mu_{s+1}(v) = \mu_s(v) - a_s \hat{w}_{\tau_s}(\mu_s(v) - z_{\tau_s}), \tag{36}$$

and

$$\|\mu_{s+1}(v) - \mu_{k_i}(v)\| \leqslant cT, \quad s = k_i, k_i+1, \ldots, m(k_i, T) \tag{37}$$

for all sufficiently large i *and small enough* $T > 0$, *where* $c > 0$, c *may depend on sample path* ω *but is independent of* k_i.

Proof. Before proving the lemma we first note that here a convergent subsequence $\{\mu_{k_i}(v)\}$ of $\{\mu_k(v)\}$ is considered, and from the subsequent proof it can be seen that replacing $\mu_{k_i}(v)$ with $\mu_k(v)$ does not work. As mentioned in Section 2, here the TS method is applied.

Temporarily ignoring (30) and (31), consider recursion (36) with initial value $\mu_{k_i}(v)$.

Set

$$\Phi_{i,j} \triangleq (1-a_i w_{\tau_i})\cdots(1-a_j w_{\tau_j}), \quad i \geqslant j, \quad \Phi_{j,j+1}=1.$$

By (A14) in Appendix and the fact that $Ew_{\tau_k} \underset{k\to\infty}{\longrightarrow} w_0$ proved in Lemma C it follows that

$$\sum_{j=k_i}^{s} a_j w_{\tau_j} = O(T) \quad \forall s \in [k_i, \dots, m(k_i, T)] \tag{38}$$

and hence

$$\log \Phi_{s,k_i} = O\left(\sum_{j=k_i}^{s} a_j w_{\tau_j}\right),$$

$$\Phi_{s,k_i} = 1 + O(T) \quad \forall s \in [k_i, \dots, m(k_i, T)] \tag{39}$$

as $i \to \infty$ and $T \to 0$.

By (A19) we have

$$\left|\sum_{j=k_i}^{s} \Phi_{s,j+1} a_j w_{\tau_j} \varepsilon_{\tau_j}\right| = O(T) \tag{40}$$

and by (38) and the boundedness of $\{y_{\tau_j}\}$

$$\sum_{j=k_i}^{s} \Phi_{s,j+1} a_j w_{\tau_j} y_{\tau_j} = O\left(\sum_{j=k_i}^{s} a_j w_{\tau_j}\right) = O(T) \quad \forall s \in [k_i, \dots, m(k_i, T)] \tag{41}$$

as $i \to \infty$ and $T \to 0$.

Combining (39), (40) and (41) yields

$$\Phi_{s,k_i}\mu_{k_i}(v) + \sum_{j=k_i}^{s} \Phi_{s,j+1} a_j w_j z_{\tau_j} = \mu_{k_i}(v) + O(T) \quad \forall s \in [k_i, \dots, m(k_i, T)]. \tag{42}$$

From (36) we have

$$\begin{aligned}\mu_{s+1}(v) &= \mu_s(v) - a_s w_{\tau_s}(\mu_s(v) - z_{\tau_s}) \\ &\quad + a_s(w_{\tau_s} - \hat{w}_{\tau_s})(\mu_s(v) - z_{\tau_s}) \\ &= \Phi_{s,k_i}\mu_{k_i}(v) + \sum_{j=k_i}^{s} \Phi_{s,j+1} a_j w_{\tau_j} z_{\tau_j} \\ &\quad + \sum_{j=k_i}^{s} \Phi_{s,j+1} a_j (w_{\tau_j} - \hat{w}_{\tau_j})(\mu_j(v) - z_{\tau_j}),\end{aligned}$$

which incorporating with (42) by Lemma 1 implies

$$\mu_{s+1}(v) = \mu_{k_i}(v) + O(T) \quad \forall s \in [k_i, \dots, m(k_i, T)]. \tag{43}$$

This means that $\mu_s(v)$ generated by the recursion (36) is close to $\mu_{k_i}(v)$, which has a finite limit as $i \to \infty$, for $s \in [k_i, \dots, m(k_i, T)]$. In other words, (36) can be considered as a part of the algorithm (30) and (31), and there is no truncation for $s \in [k_i, \dots, m(k_i, T)]$ for large i and small $T > 0$, while (37) is an alternative writing of (43).

The set Ω_0 may be taken as the one where all (A14) (A19) and (32) are simultaneously satisfied. □

Theorem 2. *Assume* H0–H2 *hold and* $\rho \neq 0$. *Then*

$$\mu_k(v) \underset{k\to\infty}{\longrightarrow} f(v) \quad a.s., \tag{44}$$

where $\mu_k(v)$ *is given by* (30) and (31).

Proof. Write (30) as

$$\begin{aligned}\mu_{k+1}(v) &= \mu_k(v) - a_k w_0(\mu_k(v) - f(v)) + a_k e_k(v) \\ &\quad \times I_{[|\mu_k(v) - a_k w_0(\mu_k(v)-f(v)) + a_k e_k(v)| \leqslant M_{\lambda_k(v)}]},\end{aligned} \tag{45}$$

where w_0 is given in Lemma C, and

$$e_k(v) = w_0(\mu_k(v) - f(v)) - \hat{w}_{\tau_k}(\mu_k(v) - z_{\tau_k}). \tag{46}$$

Since $f(v)$ is the root of the linear function $w_0(x - f(v))$, by GCT given in Section 2 for (44) it suffices to show

$$\lim_{T\to 0}\limsup_{i\to\infty}\frac{1}{T}\left|\sum_{j=k_i}^{m(k_i,T_i)} a_j \mathrm{e}_j^{(s)}(v)\right| = 0, \quad s = 1, 2, 3, 4 \quad \forall T_i \in [0, T] \tag{47}$$

for any convergent $\mu_{k_i}(v) \underset{i\to\infty}{\longrightarrow} \mu(v)$, where

$$e_k^{(1)}(v) + e_k^{(2)}(v) + e_k^{(3)}(v) + e_k^{(4)}(v) = e_k(v),$$

and

$$\begin{aligned} e_k^{(1)}(v) &\triangleq (w_{\tau_k} - \hat{w}_{\tau_k})(\mu_k(v) - y_{\tau_k}), \\ e_k^{(2)}(v) &\triangleq (w_0 - w_{\tau_k})\mu_k(u), \\ e_k^{(3)}(v) &\triangleq w_{\tau_k} y_{\tau_k} - w_0 f(v), \end{aligned}$$

and $e_k^{(4)}(v)$ is free of v and equals $e_k^{(4)}$ with

$$e_k^{(4)} \triangleq w_{\tau_k}\varepsilon_{\tau_k}.$$

It is clear that (47) holds for $s = 1$ by Lemma 1, while for $s = 4$ by (A17). We now show (47) for $s = 2$. From the following inequality

$$\begin{aligned}
&\lim_{T\to 0}\limsup_{i\to\infty}\frac{1}{T}\left|\sum_{j=k_i}^{m(k_i,T_i)} a_j e_j^{(2)}(v)\right| \\
&\leqslant \lim_{T\to 0}\limsup_{i\to\infty}\frac{1}{T}\left|\sum_{j=k_i}^{m(k_i,T_i)} a_j(\mu_j(v)-\mu(v))(w_{\tau_j}-Ew_{\tau_j})\right| \\
&\quad + \lim_{T\to 0}\limsup_{i\to\infty}\frac{1}{T}|\mu(v)|\left|\sum_{j=k_i}^{m(k_i,T_i)} a_j(w_{\tau_j}-Ew_{\tau_j})\right| \\
&\quad + \lim_{T\to 0}\limsup_{i\to\infty}\frac{1}{T}\left|\sum_{j=k_i}^{m(k_i,T_i)} a_j\mu_j(v)(Ew_{\tau_j}-w_0)\right|,
\end{aligned}$$

we see that at its right-hand side the first term is zero by (37) and (A15), the second term is zero because of (A14), while the last term is zero by the first part of (A8). Thus (47) also holds for $s=2$. Finally, by (A16) and (A17) and the second part of (A8) we conclude that (47) holds for $s=3$ too. This completes the proof of the theorem. □

5. Numerical example

Let the linear subsystem be a second order MA process

$$v_{k+1}=u_k+0.65u_{k-1}+0.28u_{k-2}, \quad u_k\in\mathcal{N}(0,1),$$

and let the nonlinear function be

$$f(v)=\begin{cases}\sin(\pi v/2) & \text{if } |v|\leqslant 1,\\ 1 & \text{if } v>1,\\ -1 & \text{if } v<-1,\end{cases}$$

and the observation noise be Gaussian $\varepsilon_k\in\mathcal{N}(0,0.1)$.

Let $a_k=k^{-1}$, $b_k=k^{-(1/7)}$ and $M_k=50+k^{(1/15)}$, and let $c=3$ in the definition of τ_k.

In Fig. 2 the estimates (dotted lines) for coefficients (solid lines) of the linear subsystem are demonstrated, while in Fig. 3 it is shown how the true $f(v)$ (solid line) is approximated by its estimate (dotted line). The estimate for $f(v)$ is derived in the following way: the interval $[-1.5, 1.5]$ where $f(v)$ is defined on is equally divided into 100 subintervals, and at each endpoint v of subintervals, $f(v)$ is estimated by (30) and (31). The dotted line consists of the estimates given at $k=3000$. In Fig. 4 the estimates (dotted lines) for $f(v)$ (solid lines) are plotted vs time for $v=-1, -0.5, 0, 0.5, 1$ to demonstrate the procedure of convergence of the estimates.

From Figs. 2, 3, and 4 it is seen that the numerical simulation justifies Theorems 1 and 2.

6. Concluding remarks

This work concerns the nonparametric approach to identification of Wiener systems. The invertibility of $f(\cdot)$ often used in literature is not assumed in the paper. All estimates for coefficients of the linear subsystem as well as for the values $f(v)$ of the nonlinear block are given recursively with the help of SA algorithms with expanding truncations, and are proved strongly consistent. To the authors' knowledge this probably is the first piece of work on strong consistency of estimates for identifying Wiener systems with nonparametric nonlinearity.

Although the SA algorithms with expanding truncations were applied to identifying Hammerstein system in Chen

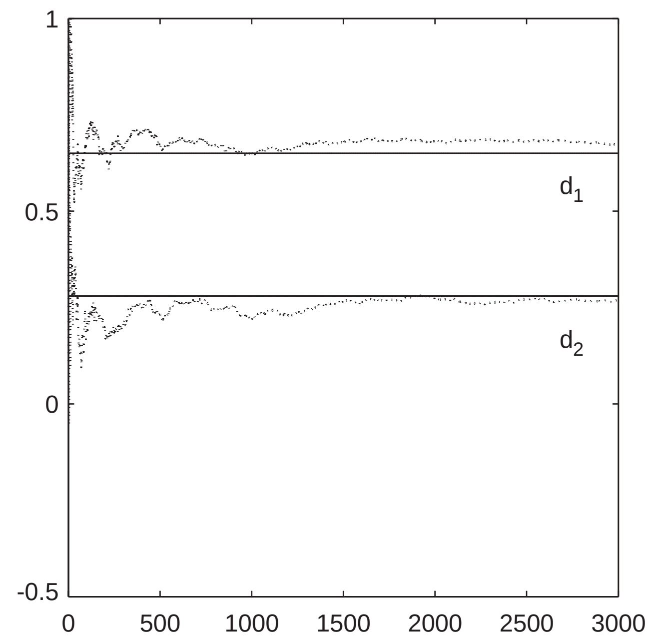

Fig. 2. Estimates for linear subsystem.

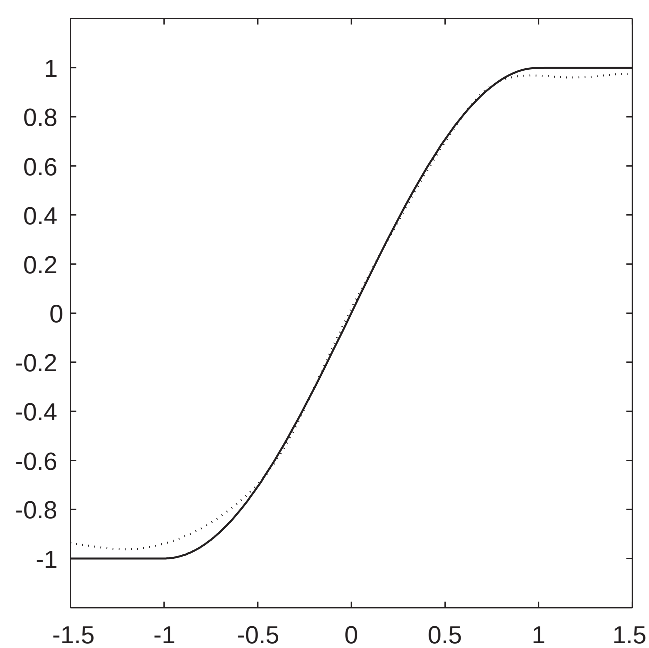

Fig. 3. Estimate for $f(\cdot)$.

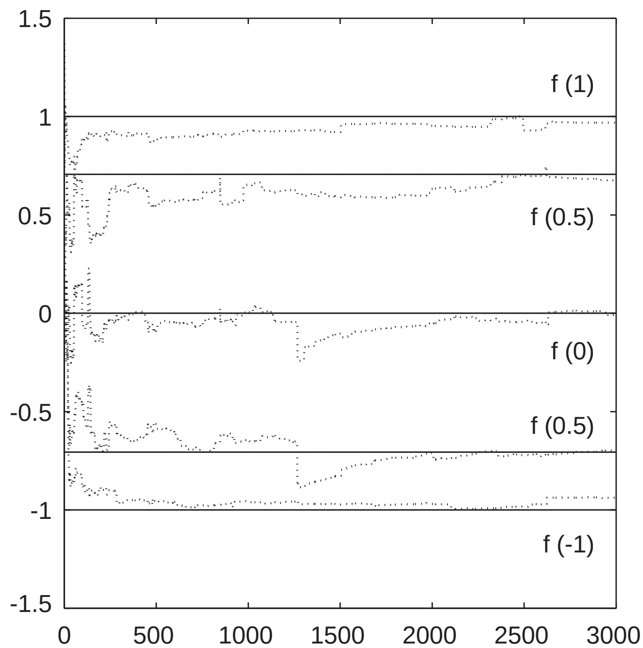

Fig. 4. Estimates for $f(v)$ with $v=0$, ± 0.5, ± 1.

(2004), extension of the method to identifying Wiener systems is not straightforward. This is because we have to use Gaussian rather than bounded input in order to appropriately estimate the linear subsystem based on the observation data which are made not directly on the output of the linear subsystem but on the output of the nonlinear block. The unboundedness of signals gives rise to additional difficulty in estimating $f(v)$. If the signals were bounded, the stopping times introduced in the paper would be unnecessary and the proof of Theorem 2 could be greatly simplified.

The identification problems for multidimensional systems, for Hammerstein–Wiener systems, and for systems with more complicated linear systems are under consideration.

Acknowledgements

The authors would like to thank the Associate Editor and the reviewers for their comments and suggestions which are helpful for improving the paper.

Appendix

In the Appendix we demonstrate some properties of τ_k, $k=1,2,\ldots$ and kernel functions defined by (28). For this we need the following well-known fact for whose proof we refer to Chow & Teicher (1978).

Proposition 1. *Let $\{\xi_k\}$ be a sequence of iid random variables and let τ be a stopping time with respect to a sequence $\{\mathscr{F}_k\}$ of nondecreasing σ-algebras. If $\mathscr{F}_k$ is independent of $\sigma\{\xi_j,\ j>k\}$, $k\geqslant 1$, then $\mathscr{F}_\tau$ is independent of $\sigma\{\xi_{\tau+1},\xi_{\tau+2},\ldots\}$ and $\{\xi_{\tau+k},\ k\geqslant 1\}$ is a sequence of iid random variables with the same distribution as that for ξ_1.*

Define

$$\tau \triangleq \tau_1-(r+1) \\ =\inf\{k\geqslant 1: |u_{k+r+1-j}|\leqslant c,\ j=1,\ldots,r+1\}. \quad \text{(A1)}$$

Lemma A. *If $\{u_k\}$ is iid and Gaussian, then $E\tau_k<\infty$, $E\tau_k^2<\infty$, and $\frac{\tau_k}{k}\underset{k\to\infty}{\longrightarrow} E\tau$ a.s., where τ_k is defined by* (24) *and* (25).

Proof. We first show that for $E\tau_k<\infty$ and $E\tau_k^2<\infty$ it suffices to prove

$$E\tau<\infty \quad \text{and} \quad E\tau^2<\infty. \quad \text{(A2)}$$

It is clear that $\tau<\infty$ a.s. and τ is a stopping time with respect to $\mathscr{G}_k\triangleq\sigma\{u_j,\ \ j=1,\ldots,k+r\}$. Then for any integer $j\geqslant 0$ there is a Borel set B_{j+r} such that $[\tau=j]=[(u_1,\ldots,u_{j+r})\in B_{j+r}]$. By Proposition 1 it follows that

$$P[\tau=j]=P[(u_1,\ldots,u_{j+r})\in B_{j+r}] \\ =P[(u_{\tau_{k-1}+1},\ldots,u_{\tau_{k-1}+j+r})\in B_{j+r}] \\ =P[\tau_k-\tau_{k-1}=j] \quad \text{with } \tau_0\triangleq r+1, \quad \text{(A3)}$$

and that $\{\tau_j-\tau_{j-1}\}$ is a sequence of mutually independent random variables by taking notice of measurability of $\{\tau_j-\tau_{j-1}\}$, $j=1,\ldots,k-1$ and $\tau_k-\tau_{k-1}$ with respect to $\mathscr{G}_{\tau_{k-1}}$ and $\sigma\{u_{\tau_{k-1}+1},u_{\tau_{k-1}+2},\ldots\}$, respectively. Consequently, we have

$$E\tau_k=E\tau_0+\sum_{j=1}^{k}E(\tau_j-\tau_{j-1})=(r+1)+kE\tau$$

and

$$E\tau_k^2=E\left(\tau_0+\sum_{j=1}^{k}(\tau_j-\tau_{j-1})\right)^2 \\ \leqslant (r+1)^2+2(r+1)kE\tau+k^2E\tau^2.$$

This proves sufficiency of (A2) for Lemma A. Therefore, (A2) implies $E\tau_k<\infty$ and $E\tau_k^2<\infty$, $\forall\ k\geqslant 1$.

We now prove (A2).

From (A1) it is seen that τ is the first time that the immediately past $r+1$ (with the τth random variable included) successive random variables in $\{u_k\}$ are bounded by c.

Set $p\triangleq P[|u_1|\leqslant c]$ and $q\triangleq 1-p$. Consider a sequence of independent trials. Each trial with probability p has outcome "success" and with probability q "failure". Assume $n\geqslant 2k$ and let $P_n(k)$ be the probability that in n trials there are no k successive successes.

It is clear that

$$P_n(k)=qP_{n-1}(k)+pqP_{n-2}(k) \\ +\cdots+p^{k-1}qP_{n-k}(k), \quad \text{(A4)}$$

or

$A(z)P_n(k)=0,$

where

$$A(z)=1-qz-pqz^2-\cdots-p^{k-1}qz^k \tag{A5}$$

with z being the backward-shift operator.

Since

$$|qz+pqz+\cdots+p^{k-1}qz^k|\leqslant\frac{q(1-p^k)}{1-p}=1-p^k<1$$

on the unit circle $|z|=1$, by the Rouché theorem from the theory of complex variable functions it is concluded that $A(z)$ and 1 have the same number of roots inside the unit circle. In other words, all roots of $A(z)$ are outside the closed unit disk. Therefore, $P_n(k)$ as a solution of the stable difference equation exponentially decays, i.e., there are constants $C(k)$ and $0<\rho(k)<1$ possibly depending on k such that

$$P_n(k)\leqslant C(k)\rho^n(k). \tag{A6}$$

For $n\geqslant 2(r+1)+2$, from the definition (A1) we see that $[\tau=n]$ means that $|u_{n-1}|>c$, there are no $r+1$ successive random variables bounded by c in the first $n-2$ random variables of $\{u_k\}$, and $|u_{n+i}|\leqslant c$ for all $i=0,1,\ldots,r$. Thus, by (A6) we have

$$\begin{aligned}P[\tau=n]&=P_{n-2}(r+1)qp^{r+1}\\&\leqslant C(r+1)\rho^{n-2}(r+1)qp^{r+1},\end{aligned}$$

and hence

$$\begin{aligned}E\tau&=\sum_{n=1}^{\infty}nP[\tau=n]\\&=\sum_{n=1}^{2r+3}nP[\tau=n]+\sum_{n=2r+4}^{\infty}nP[\tau=n]\\&\leqslant\sum_{n=1}^{2r+3}nP[\tau=n]\\&\quad+C(r+1)qp^{r+1}\sum_{n=2r+4}^{\infty}n\rho^{n-2}(r+1)<\infty,\end{aligned}$$

since $0<\rho(r+1)<1$.

The proof of $E\tau^2<\infty$ is completed in a similar way.

By Proposition 1 and (A3) it follows that $\{\tau_j-\tau_{j-1}\}$ with $\tau_0=r+1$ is iid. By the strong law of large numbers (Chow & Teicher, 1978) we have

$$\frac{\tau_k}{k}=\frac{r+1}{k}+\frac{1}{k}\sum_{j=1}^{k}(\tau_j-\tau_{j-1})\underset{k\to\infty}{\longrightarrow}E\tau\quad\text{a.s.}\qquad\square$$

Corollary 1. *Since $E\tau_k<\infty$, we have $\tau_k<\infty$ a.s. Hence τ_k is a stopping time with respect to $\{\mathscr{F}_k\}$ and $\tau_k-\tau_{k-1}$ is a stopping time with respect to $\{\mathscr{H}_j\}$, where $\mathscr{H}_j\triangleq\sigma\{u_{\tau_{k-1}+i}, i=1,\ldots,j-1\}$.*

Lemma B. *Let $\{\xi_k\}$ be a sequence of iid random variables and let $\{\xi_j, j=k,k+1,\ldots\}$ be independent of $\{u_j, j=1,\ldots,k-1\}$. Then $\{\xi_{\tau_k}\}$ is iid where $\{\tau_k\}$ is given by* (24) *and* (25).

Proof. Since $\{\tau_j\}$ is a sequence of stopping times with respect to $\{\mathscr{F}_j\}$ and $\tau_1<\tau_2<\cdots<\tau_k$, we see that $\{\xi_{\tau_1},\ldots,\xi_{\tau_k}\}$ are $\mathscr{F}_{\tau_k}$-measurable. On the other hand, by the definition of τ_{k+1}, $\xi_{\tau_{k+1}}$ is measurable with respect to $\sigma\{\xi_{\tau_k+1},\xi_{\tau_k+2},\ldots\}$ which is independent of $\mathscr{F}_{\tau_k}$ by Proposition 1. This implies that$\xi_{\tau_{k+1}}$ is independent of $(\xi_{\tau_1},\ldots,\xi_{\tau_k})$ and hence proves the mutually independence of $\{\xi_{\tau_k}\}$ by induction.

We now show that they are identically distributed. Similar to (A3) we have

$$\begin{aligned}&P[\xi_{\tau_k}<\lambda]\\&=\sum_{j=1}^{\infty}P[\tau_k-\tau_{k-1}=j,\ \xi_{\tau_{k-1}+j}<\lambda]\\&=\sum_{j=1}^{\infty}P[(u_{\tau_{k-1}+1},u_{\tau_{k-1}+2},\ldots)\in B_{j+r},\ \xi_{\tau_{k-1}+j}<\lambda]\\&=\sum_{j=1}^{\infty}P[(u_1,u_2,\ldots)\in B_{j+r},\ \xi_j<\lambda]\\&=\sum_{j=1}^{\infty}P[\tau=j,\ \xi_\tau<\lambda]=P[\xi_\tau<\lambda]=P[\xi_1<\lambda].\qquad\square\end{aligned}$$

Introduce the following notations:

$$s^+\triangleq c+\sum_{j=1}^{r}d_js_j-v,\quad s^-\triangleq -c+\sum_{j=1}^{r}d_js_j-v, \tag{A7}$$

$$\begin{aligned}&B_-^+(\varepsilon)\triangleq\{(s_1,s_2,\ldots,s_r)\in[-c,c]^r:\\&\quad s^+>\varepsilon,\ s^-<-\varepsilon,\ \varepsilon\geqslant 0\},\\&B_+^+(\varepsilon)\triangleq\{(s_1,s_2,\ldots,s_r)\in[-c,c]^r:\ s^->\varepsilon,\ \varepsilon\geqslant 0\},\\&B_-^-(\varepsilon)\triangleq\{(s_1,s_2,\ldots,s_r)\in[-c,c]^r:\ s^+<-\varepsilon,\ \varepsilon\geqslant 0\}.\end{aligned}$$

Clearly, $B_-^+(\varepsilon)$, $B_+^+(\varepsilon)$ and $B_-^-(\varepsilon)$ are bounded sets. Denote their Lebesgue measures by $V(B_-^+(\varepsilon))$, $V(B_+^+(\varepsilon))$ and $V(B_-^-(\varepsilon))$, respectively.

Lemma C. *Assume* H0–H2 *hold. Then*

$$Ew_{\tau_k}\underset{k\to\infty}{\longrightarrow}w_0,\quad E[w_{\tau_k}f(v_{\tau_k})]\underset{k\to\infty}{\longrightarrow}w_0f(v) \tag{A8}$$

and

$$\sup_k E(\sqrt{b_k}w_{\tau_k})^2<\infty, \tag{A9}$$

where $w_0=\dfrac{V(B_-^+(0))}{(\sqrt{2\pi})^r p^{r+1}}$ *and* $p=P[|u_k|\leqslant c]$.

Proof. For any integers k and n there is a Borel set $B_{n-r-2}(k)$ such that

$$[\tau_k = n] = \{(u_1, \ldots, u_{n-r-2}) \in B_{n-r-2}(k),\ u_i \in [-c, c],\ i = n-r-1, \ldots, n-1\}.$$

Then, by the independence of $\{u_j\}$ we have

$$\begin{aligned}
&E[w_{\tau_k} f(v_{\tau_k}) I_{[\tau_k=n]}] \\
&\quad = \frac{1}{b_k} E[\mathrm{e}^{-((v_n - v)/b_k)^2 + \frac{1}{2}\sum_{j=1}^{r+1} u_{n-j}^2} f(v_n) \\
&\qquad \times I_{[(u_1,\ldots,u_{n-r-2}) \in B_{n-r-2}(k),\ u_i \in [-c,c],\ i=n-r-1,\ldots,n-1]}] \\
&\quad = \frac{P[(u_1, \ldots, u_{n-r-2}) \in B_{n-r-2}(k)]}{(\sqrt{2\pi})^{r+1} b_k} \int_{-c}^{c} \cdots \int_{-c}^{c} \\
&\qquad \times \exp\left\{-\left(\frac{\sum_{j=0}^{r} d_j x_j - v}{b_k}\right)^2\right\} \\
&\qquad \times f\left(\sum_{j=0}^{r} d_j x_j\right) \mathrm{d}x_0 \cdots \mathrm{d}x_r \\
&\quad = \frac{P[\tau_k = n]}{(\sqrt{2\pi} p)^{r+1} b_k} \int_{-c}^{c} \cdots \int_{-c}^{c} \\
&\qquad \times \exp\left\{-\left(\frac{\sum_{j=0}^{r} d_j x_j - v}{b_k}\right)^2\right\} \\
&\qquad \times f\left(\sum_{j=0}^{r} d_j x_j\right) \mathrm{d}x_0 \cdots \mathrm{d}x_r \\
&\quad = \frac{P[\tau_k = n]}{(\sqrt{2\pi} p)^{r+1}} \int_{-c}^{c} \cdots \int_{-c}^{c} \mathrm{d}s_1 \cdots \mathrm{d}s_r \\
&\qquad \times \left(\int_{s^-/b_k}^{s^+/b_k} \mathrm{e}^{-s_0^2} f(v + b_k s_0)\, \mathrm{d}s_0\right),
\end{aligned} \tag{A10}$$

where

$$s_0 = \frac{x_0 + d_1 x_1 + \cdots + d_r x_r - v}{b_k},$$
$$s_j = x_j, \quad j = 1, 2, \ldots, r,$$

and s^+ and s^- are given by (A7).

Noticing

$$\frac{s^+}{b_k} \xrightarrow[k\to\infty]{} \infty \quad \text{and} \quad \frac{s^-}{b_k} \xrightarrow[k\to\infty]{} -\infty$$

in $B_-^+(\varepsilon)$ with $\varepsilon > 0$, we find that

$$\int_{B_-^+(\varepsilon)} \mathrm{d}s_1 \cdots \mathrm{d}s_r \left(\int_{s^-/b_k}^{s^+/b_k} \mathrm{e}^{-s_0^2} f(v + b_k s_0)\, \mathrm{d}s_0\right) \xrightarrow[k\to\infty]{} \sqrt{2\pi} V(B_-^+(\varepsilon)) f(v),$$

and by tending $\varepsilon \to 0$

$$\int_{B_-^+(0)} \mathrm{d}s_1 \cdots \mathrm{d}s_r \left(\int_{s^-/b_k}^{s^+/b_k} \mathrm{e}^{-s_0^2} f(v + b_k s_0)\, \mathrm{d}s_0\right) \xrightarrow[k\to\infty]{} \sqrt{2\pi} V(B_-^+(0)) f(v). \tag{A11}$$

Similarly, it is proved that the limit of integral (A11) is zero if the integral is taken over $B_+^+(0)$ or $B_-^-(0)$.

Thus, by (A10) and (A11) it follows that

$$\begin{aligned}
&E[w_{\tau_k} f(v_{\tau_k})] \\
&\quad = \sum_{n=1}^{\infty} E[w_{\tau_k} f(v_{\tau_k}) I_{[\tau_k=n]}] \\
&\quad = \frac{1}{(\sqrt{2\pi} p)^{r+1}} \int_{-c}^{c} \cdots \int_{-c}^{c} \mathrm{d}s_1 \cdots \mathrm{d}s_r \\
&\qquad \times \left(\int_{s^-/b_k}^{s^+/b_k} \mathrm{e}^{-s_0^2} f(v + b_k s_0) \mathrm{d}s_0\right) \\
&\quad \xrightarrow[k\to\infty]{} \frac{1}{(\sqrt{2\pi})^r p^{r+1}} V(B_-^+(0)) f(v) = w_0 f(v).
\end{aligned} \tag{A12}$$

As a special case of (A12), we have $E w_{\tau_k} \xrightarrow[k\to\infty]{} w_0$.

By a similar approach we find that

$$\begin{aligned}
&E(\sqrt{b_k} w_{\tau_k})^2 \\
&\quad = \frac{1}{(\sqrt{2\pi} p)^{r+1} b_k} \int_{-c}^{c} \cdots \int_{-c}^{c} \mathrm{d}x_0 \cdots \mathrm{d}x_r \\
&\qquad \times \exp\left\{-2\left(\frac{\sum_{j=0}^{r} d_j x_j - v}{b_k}\right)^2 + \frac{1}{2} \sum_{j=0}^{r} x_j^2\right\} \\
&\quad \leqslant \frac{\mathrm{e}^{((r+1)c^2)/2}}{(\sqrt{2\pi} p)^{r+1} b_k} \int_{-c}^{c} \cdots \int_{-c}^{c} \\
&\qquad \times \exp\left\{-2\left(\frac{\sum_{j=0}^{r} d_j x_j - v}{b_k}\right)^2\right\} \mathrm{d}x_0 \cdots \mathrm{d}x_r
\end{aligned} \tag{A13}$$

and the right side of (A13) is uniformly bounded with respect to k. This proves (A9). □

Lemma D. *Assume* H0–H2 *hold. Then*

$$\sum_{j=1}^{\infty} a_j (w_{\tau_j} - E w_{\tau_j}) < \infty \quad a.s., \tag{A14}$$

$$\sum_{j=1}^{\infty} a_j (|w_{\tau_j} - E w_{\tau_j}| - E|w_{\tau_j} - E w_{\tau_j}|) < \infty \quad a.s., \tag{A15}$$

$$\sum_{j=1}^{\infty} a_j (w_{\tau_j} y_{\tau_j} - E w_{\tau_j} y_{\tau_j}) < \infty \quad a.s., \tag{A16}$$

$$\sum_{j=1}^{\infty} a_j w_{\tau_j} \varepsilon_{\tau_j} < \infty \quad a.s., \tag{A17}$$

$$\sum_{j=1}^{\infty} a_j \left(|\varepsilon_{\tau_j}| - E|\varepsilon_{\tau_j}|\right) < \infty \quad a.s., \tag{A18}$$

and

$$\sum_{j=1}^{\infty} a_j \left(w_{\tau_j} |\varepsilon_{\tau_j}| - E w_{\tau_j} |\varepsilon_{\tau_j}|\right) < \infty \quad a.s. \tag{A19}$$

Proof. Since for any fixed $j=0,1,\ldots,r$, the sequence $\{v_{j+k(r+1)},\ k=0,1,\ldots\}$ is iid, by Lemmas B and C, $\{\sqrt{b_{j+k(r+1)}}w_{\tau_{j+k(r+1)}} - E\sqrt{b_{j+k(r+1)}}w_{\tau_{j+k(r+1)}},\}$ is a sequence of mutually independent random variables with bounded variances, $j=1,2,\ldots,r+1$. Therefore, by the convergence theorem for sum of independent random variables (Chow & Teicher, 1978) we derive (A14) by writing

$$\sum_{j=1}^{\infty} a_j(w_{\tau_j} - Ew_{\tau_j}) = \sum_{j=1}^{r}\sum_{k\in A_j}\frac{a_k}{\sqrt{b_k}}(\sqrt{b_k}w_{\tau_k} - E\sqrt{b_k}w_{\tau_k}) < \infty \quad \text{a.s.},$$

where $A_j=\{j+k(r+1) : k=0,1,2,\ldots\}$, $j=1,2,\ldots,r+1$.

Similarly, we can show (A15) and (A16) by noticing that $\{y_k\}$ is bounded by H1 and (26).

We now show (A17). Proceeding as (A10) by the independence of $\{u_k\}$ and $\{\varepsilon_k\}$ we find that

$$\begin{aligned} &E[w_{\tau_k}^2\varepsilon_{\tau_k}^2 I_{[\tau_k=n]}] \\ &= E\left[\exp\left\{-2\left(\frac{v_n - v}{b_k}\right)^2 + \sum_{j=1}^{r+1}u_{n-j}^2\right\}\varepsilon_n^2 \right. \\ &\qquad \left. \times\ I_{[(u_1,\ldots,u_{n-r-2})\in B_{n-r-2}(k),\ u_i\in[-c,c],\ i=n-r-1,\ldots,n-1]}\right] \\ &\leqslant P[\tau_k=n]E\varepsilon_n^2\sup_k E(\sqrt{b_k}w_{\tau_k})^2 \end{aligned}$$

and hence by H2 and Lemma C

$$\begin{aligned} E[b_k w_{\tau_k}^2\varepsilon_{\tau_k}^2] &= \sum_{n=1}^{\infty}P[\tau_k=n]E\varepsilon_n^2\sup_k E(\sqrt{b_k}w_{\tau_k})^2 \\ &= E\varepsilon_n^2\sup_k E(\sqrt{b_k}w_{\tau_k})^2 < \infty. \end{aligned} \quad \text{(A20)}$$

By the similar treatment and using the independence of $\{u_k\}$ and $\{\varepsilon_k\}$ and Lemma B we find that

$$E[w_{\tau_k}\varepsilon_{\tau_k}] = 0. \quad \text{(A21)}$$

Combining (A20) and (A21) leads to (A17) and (A19), while (A18) follows from Lemma B. □

References

Bendat, J. S. (1999). *Nonlinear system analysis and identification from random data*. New York: Wiley.
Billings, S. A., & Fakhouri, F. Y. (1978). Identification of nonlinear systems using the Wiener model. *Electronics Letters*, *13*, 502–504.
Chen, H. F. (2002). *Stochastic approximation and its applications*. Dordrecht: Kluwer.
Chen, H. F. (2004). Pathwise convergence of recursive identification algorithms for Hammerstein systems. *IEEE Transactions on Automatic Control*, *49*(10), 1641–1649.
Chen, H. F., & Guo, L. (1991). *Identification and stochastic adaptive control*. Boston: Birkhäuser.
Chow, Y. S., & Teicher, H. (1978). *Probability theory*. New York: Springer.
Greblicki, W. (1997). Nonparametric approach to Wiener system identification. *IEEE Transactions on Circuits and Systems-I: Fundamental Theory and Applications*, *44*(6), 538–545.
Greblicki, W. (2001). Recursive identification of Wiener system. *Applied Mathematics and Computer Science*, *11*(4), 977–991.
Hasiewicz, Z. (1987). Identification of a linear system observed though zero-memory nonlinearity. *International Journal of Systems Science*, *18*, 1595–1607.
Hunter, I. W., & Korenberg, M. J. (1986). The identification of nonlinear biological systems: Wiener and Hammerstein cascade models. *Biological Cybernetics*, *55*, 135–144.
Ljung, L. (1987). *System identification*. Englewood Cliffs, NJ: Prentice-Hall.
Nordsjö, A. E., & Zetterberg, L. H. (2001). Identification of certain time-varying nonlinear Wiener and Hammerstein systems. *IEEE Transactions on Signal Processings*, *49*, 577–592.
Pajunen, G. A. (1992). Adaptive control of Wiener type nonlinear systems. *Automatica*, *28*, 781–785.
Verhaegen, M., & Westwick, D. (1996). Identifying MIMO Wiener systems in the context of subspace model identification methods. *International Journal of Control*, *63*(2), 331–349.
Vörös, J. (2001). Parameter identification of Wiener systems with discontinuous nonlinearities. *Systems and Control Letters*, *44*, 363–372.
Westwick, D. T., & Kearney, R. E. (1992). A new algorithm for the identification of multiple input Wiener systems. *Biological Cybernetics*, *68*, 75–85.
Wiener, N. (1958). *Nonlinear problems in random theory*. NewYork: Wiley.
Wigren, T. (1994). Convergence analysis of recursive algorithms based on the nonlinear Wiener model. *IEEE Transactions on Automatic Control*, *39*, 2191–2206.
Wigren, T. (1993). Recursive prediction error identification using the nonlinear Wiener mode. *Automatica*, *29*, 1011–1025.
Wigren, T. (1997). Circle criteria in recursive identification. *IEEE Transactions on Automatic Control*, *42*, 975–979.
Wigren, T. (1998). Output error convergence of adaptive filters with compensation for output nonlinearities. *IEEE Transactions on Automatic Control*, *43*, 975–978.

Xiao-Li Hu was born in Hunan, China, in 1975. He received the B.S. degree in Mathematics from Hunan Normal University in 1997, and the MSC degree in Applied Mathematics from Kunming University of Science and Technology in 2003. Now he is pursuing his Ph.D. degree in the Key Laboratory of Systems and Control, Chinese Academy of Sciences. His current research interests are system identification and stochastic approximation and its applications.

Han-Fu Chen after graduation from the Leningrad (St. Petersburg) University in 1961 joined the Institute of Mathematics, Chinese Academy of Sciences (CAS). Since 1979 he has been with the Institute of Systems Science, which is now a part of the Academy of Mathematics and Systems Sciences, CAS. His research interests are stochastic systems, including system identification, adaptive control, and stochastic approximation and its applications to systems, control, and signal processing. He authored and coauthored more than 160 journal papers and seven books.