Strong consistence of recursive identification for Wiener systems

Xiao-Li Hu\textsuperscript{a,b}, Han-Fu Chen\textsuperscript{a,∗}

\textsuperscript{a}Key Laboratory of Systems and Control, Institute of Systems Science, AMSS Chinese Academy of Sciences, Beijing 100080, PR China
\textsuperscript{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 $\varepsilon_k$. 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

$$z_k = y_k + \varepsilon_k. \quad (1)$$

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).
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, 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)).

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

\[ v_{k+1} = \sum_{j=0}^{r} d_j u_{k-j}, \quad d_0 = 1 \]  

(2)

and the output of the nonlinearity is

\[ y_k = f(v_k). \]  

(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 \( v \in \mathcal{R} \) 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: \( u_k \in \mathcal{N}(0, 1) \), and \( \{u_k\} \) is independent of the observation noise \( \{z_k\} \).

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 \( v \). The growth rate of \( f(v) \) as \( |v| \to \infty \) is not faster than a polynomial.

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

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 \( \{g_k\} \):

\[ g_{k+1} = g(x_k) + \eta_{k+1}, \]  

(4)

where \( x_k \) is the \( k \)th estimate for \( x^0 \) and \( \eta_{k+1} \) 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:

\[ x_{k+1} = x_k + a_k g_{k+1}, \quad k = 0, 1, \ldots \]  

(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 \( M_k \to \infty \). Let \( x_k \) be generated by the following algorithm:

\[ x_{k+1} = \begin{cases} x_k + a_k g_{k+1} & \text{if } ||x_k + a_k g_{k+1}|| \leq M_{g_k}, \\ x^* & \text{otherwise,} \end{cases} \]  

(6)

\[ \sigma_k = \sum_{j=1}^{k-1} I(||x_j + a_j g_{j+1}|| > M_{g_j}), \quad \sigma_0 = 0. \]  

(7)

From (7) it is seen that \( \sigma_k \leq k - 1 \), and from (6) \( ||x_{k+1}|| \leq ||x^*|| \land M_{g_k} \). 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 \( M_{g_k} \), then we pull \( x_{k+1} \) back to the fixed point \( x^* \) and simultaneously extend the truncation bound from \( M_{g_k} \) to \( M_{g_{k+1}} \). 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. \( a_k > 0, a_k \to 0, k \to \infty \) and \( \sum_{k=1}^{\infty} a_k = \infty, M_k > 0, M_{k+1} > M_k, M_k \to \infty \).

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

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

\[ \sup_{c' \leq ||x - x^0|| \leq c''} g^T(x) v(x) < 0 \]

for any \( c'' > c' > 0 \), where \( v(x) \) 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:

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

(8)

for any \( n_k \) such that \( x_{n_k} \) converges, where

\[ m(k, T) = \max \left\{ m : \sum_{j=k}^{m} a_j \leq T \right\}. \]

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:

\[(5) \quad a_k > 0, \quad a_k \to 0, \quad \text{and} \quad \sum_{k=1}^{\infty} a_k = \infty, \quad \frac{1}{a_{k+1}} - \frac{1}{a_k} \to \gamma \geq 0; \quad M_k > 0, \quad M_{k+1} > M_k, \quad M_k \to \infty.\]

\[(6) \quad \text{For the sample path } \omega \text{ under consideration the observation noise } \{\eta_k\} \text{ in (4) can be decomposed into two parts } \eta_k = \eta_k' + \eta_k'' \text{ such that}\]

\[
\sum_{k=1}^{\infty} \frac{1}{\delta} \eta_k' < \infty, \quad \eta_k'' = O(a_k^\delta)
\]

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

\[(7) \quad g(\cdot) \text{ is measurable and locally bounded, and is differentiable at } x^0 \text{ such that as } x \to x^0 \]

\[
g(x) = F(x - x^0) + A(x), \quad A(x^0) = 0, \quad A(x) = o\left(\|x - x^0\|\right),
\]

and the matrix \(F + \gamma \delta I\) is stable, where \(\gamma\) and \(\delta\) 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 \(\delta\) 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

\[
a_k = \frac{1}{k} \quad \text{and} \quad M_k = M_0 + k^\beta
\]

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 stepsize, 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 \(|u_{\eta_k}(v)|\) 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

\[
\theta_k(i) = \theta_k(i) - a_k(\theta_k(i) - u_k z_{k+i+1})
\]

\[
\times I(\|\theta_k(i) - a_k(\theta_k(i) - u_k z_{k+i+1})\| < M_k(i)),
\]

\[
\sigma_k(i) = \sum_{j=1}^{k-1} I(\|\theta_j(i) - a_j(\theta_j(i) - u_j z_{j+i+1})\| < M_j(i))\]

\[
\sigma_0(i) = 0
\]

with initial values \(\theta_0(i), i = 0, 1, \ldots, r.\) Here \(\theta_k(0)\) is used to estimate \(\rho \overset{\text{d}}{=} (1/\sigma_v^2) E[v_k f(v_k)]\), and \(\theta_k(i)\) for \(\rho d_i, i = 1, \ldots, r,\) where \(\sigma_v^2\) denotes the variance of \(v_k\).

It is clear that \((\theta_k(i)/\theta_k(0)) \overset{\text{d}}{=} d_i\) may serve as the estimate for \(d_i\) at time \(k\) whenever \(\theta_k(0) \neq 0.\)

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

\[
\theta_k(0) \overset{\text{d}}{\to} \rho \frac{E(v_k f(v_k))}{\sigma_v^2} \quad \text{a.s.}
\]

and

\[
\theta_k(i) \overset{\text{d}}{\to} \rho d_i \quad \text{a.s.,} \quad i = 1, \ldots, r
\]

with rates of convergence

\[
|\theta_k(0) - \rho| = o(k^{-\gamma}) \quad \text{a.s.}, \quad |\theta_k(i) - \rho d_i| = o(k^{-\gamma}) \quad \text{a.s.}, \quad \gamma \in (0, 1/2), \quad i = 1, \ldots, r.
\]

**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

\[
E\left[u_k - \frac{d_i v_{k+i+1}}{\sigma_v^2}\right] v_{k+i+1}\]

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

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...
variables independence is equivalent to uncorrelatedness. From independence it follows that
\[ E\left( u_k - \frac{d_i v_{k+i+1}}{\sigma^2_V} \right) = 0, \]
which implies
\[ E(u_k|v_{k+i+1}) = \frac{d_i v_{k+i+1}}{\sigma^2_V}. \]
Consequently, we have
\[ E(u_k v_{k+i+1}) = \frac{d_i v_{k+i+1}}{\sigma^2_V} = d_i \rho. \]
(15)
The recursion (10) can be rewritten as
\[ \theta_{k+1}(i) = \theta_k(i) - a_k(\theta_k(i) - d_i \rho) - a_k \tilde{\theta}_{k+1}(i) \times I(\theta_k(i) - a_k(\theta_k(i) - d_i \rho) - a_k \tilde{\theta}_{k+1}(i) \leq M_d(i)). \]
(16)
where
\[ \tilde{\theta}_{k+1}(i) = -u_k z_{k+i+1} + d_i \rho, \quad i = 0, 1, \ldots, r. \]
(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 \( g^{(i)}(x) \equiv -(x - d_i \rho), \) \( i = 0, 1, \ldots, r. \) 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
\[ \lim_{T \to \infty} \lim_{n \to \infty} \frac{1}{T} \sum_{t=0}^{T} \frac{m(n,t)}{m} \leq 1 \]
for all \( t \in [0, T], \quad i = 0, 1, \ldots, r, \)
(18)
where
\[ m(n,t) = \max \left\{ m : \sum_{j=n}^{m} \frac{1}{j} \leq t \right\}. \]
(19)
Notice that
\[ \tilde{\theta}_{k+1}(i) = -(u_k y_{k+i+1} - d_i \rho) - u_k \tilde{\theta}_{k+1}(i). \]
(20)
By (15), \( \{u_k y_{k+i+1} + d_i \rho, k = 0, 1, \ldots\} \) 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
\[ \sum_{j=1}^{\infty} a_j (u_j y_{j+i+1} - d_i \rho) = \sum_{j=1}^{r+1} \sum_{k \in A_j} a_k (u_k y_{k+i+1} - d_i \rho) < \infty \quad \text{a.s.}, \]
(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 \( \{u_i\} \) and \( \{v_k\} \) we conclude
\[ \sum_{j=1}^{\infty} a_j \tilde{\theta}_{j+i+1}(i) < \infty \quad \text{a.s.} \]
(22)
which implies (18).

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

4. Estimation for \( f(v) \)

As explained in Introduction, the input of the nonlinear part is estimated by
\[ \hat{u}_k \triangleq u_{k-1} + d_1 u_{k-2} + \cdots + d_k u_{k-r-1}. \]
(23)
It is conceivable to estimate \( f(v) \) on the basis of \( \{\hat{u}_k\} \) and \( \{v_k\} \) by using SA algorithm with kernel functions as done in Chen (2004). However, \( \hat{u}_k \) is, in general, unbounded because of the unboundedness of \( \{u_i\} \), 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
\[ \tau_1 = \inf\{k > r + 1 : |u_{k-j}| \leq c, \quad j = 1, \ldots, r+1\} \]
(24)
and
\[ \tau_k = \inf\{i > \tau_{k-1} : |u_i-j| \leq c, \quad j = 1, \ldots, r+1, \}
\]
(25)
\[ k = 2, 3, \ldots. \]
From (24) and (25) we see that \( \tau_k \geq \tau_{k-1} \), and starting from \( u_{\tau_k-r-1} \) there are \( r+1 \) successive \( u_i \) with \( |u_i| \leq c, \quad i = \tau_k - r - 1, \ldots, \tau_k - 1. \) Thus, among \( \hat{u}_i \) 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 \( u_1, \ldots, u_{\tau_k-1}. \) In other words, \( \{\tau_k = s\} \in \mathcal{F}_s \triangleq \sigma\{u_1, \ldots, u_{s-1}\}, \) the \( \sigma \)-algebra generated by \( \{u_1, \ldots, u_{s-1}\}. \) We recall that the nonnegative random variable \( \tau \) is called the Markov time with respect to the family of nondecreasing \( \sigma \)-algebras \( \mathcal{F}_s \) if \( \tau = \tau_k \in \mathcal{F}_s, \) \( s = 1, 2, \ldots. \) Further, if a Markov time \( \tau < \infty \) a.s., then \( \tau \) is called the stopping time with respect to \( \{\mathcal{F}_s\} \). Consequently, \( \tau_{k+1} = 1, 2, \ldots \) are the Markov times.

It is worth noting that
\[ |v_{\tau_k}| = |u_{\tau_k-1} + d_1 u_{\tau_k-2} + \cdots + d_r u_{\tau_k-r-1}| \leq c \sum_{j=0}^{r} |d_j|, \quad k = 1, 2, \ldots. \]
(26)
Let
\[ b_k = \frac{1}{k^2}, \]
(27)
where \( \varepsilon \) is figured in the definition of \( \beta \) appearing in (9).
For a fixed \( v \in \mathcal{R} \), define the kernel function

\[
w_{\tau_k} = \frac{1}{b_k} e^{-(v_{\tau_k} - v)^2/(2b_k^2)},
\]

and its estimate

\[
\hat{w}_{\tau_k} = \frac{1}{b_k} e^{-(\hat{v}_{\tau_k} - v)^2/(2b_k^2)},
\]

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

\[
\mu_k(v) = \mu_k(v) - a_k \hat{w}_{\tau_k} (\mu_k(v) - z_{\tau_k})
\]

\[
\lambda_k(v) = \sum_{j=1}^{k-1} I(\mu_j(v) - a_j \hat{w}_{\tau_j} (\mu_j(v) - z_{\tau_j})) \leq M_{\lambda_j(v)}.
\]

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 \( \tau_k \) and the analysis carried out in this section would be much simpler. As a matter of fact, in this case all \( \tau_k \) 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

\[
\sum_{k=1}^{\infty} a_k |(\hat{w}_{\tau_k} - w_{\tau_k})(\mu_k(v) - z_{\tau_k})| < \infty \text{ a.s.,}
\]

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

**Proof.** Let \( \delta \in (0, 1/3 - 3\beta - \beta) \), then \( 1/3 - 3\beta - \beta - \delta > 1 \). By the boundedness of \( \{v_{\tau_k}\}, \) (14), and the fact \( (\tau_k/k) \rightarrow E \tau \) (see Lemma A in Appendix), it follows that

\[
\left( \frac{v_{\tau_k} - v}{b_k} \right)^2 - \left( \frac{\hat{v}_{\tau_k} - v}{b_k} \right)^2 = \frac{1}{b_k^2} (v_{\tau_k} \hat{v}_{\tau_k} - 2v^2)(v_{\tau_k} \hat{v}_{\tau_k}) = o \left( k^{-1/2 - 2\delta \beta} \right) \rightarrow 0 \text{ a.s.}
\]

From this we have

\[
\hat{w}_{\tau_k} - w_{\tau_k} = \frac{1}{b_k} e^{-(\hat{v}_{\tau_k} - v)^2/(2b_k^2)} - e^{-(v_{\tau_k} - v)^2/(2b_k^2)}
\]

\[
\times \left[ e^{((v_{\tau_k} - v)/b_k)^2} - (v_{\tau_k} - v)^2 \right] = o(k^{-1/2 - 2\delta \beta}) \text{ a.s.,}
\]

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

\[
|\hat{w}_{\tau_k} - w_{\tau_k}| \leq M \frac{1}{k^{1/2 - 3\beta \delta}} \text{ a.s. } \forall k.
\]

Therefore, taking notice of \( M_k = M_{\lambda} + k^\delta \) and \( M_{\lambda_{k-1}(v)} \leq M_{\lambda} + (k-1)^\delta \), we have

\[
\sum_{k=1}^{\infty} a_k |(\hat{w}_{\tau_k} - w_{\tau_k})\mu_k(v)| \leq \sum_{k=1}^{\infty} a_k |(\hat{w}_{\tau_k} - w_{\tau_k})| \cdot |\mu_k(v)|
\]

\[
\leq M(M_0 + 1) \sum_{k=1}^{\infty} \frac{1}{(k^{1/2 - 3\beta \delta} M_{\lambda_{k-1}(v)})}
\]

\[
\leq M(M_0 + 1) \sum_{k=1}^{\infty} \frac{1}{(k^{1/2 - 3\beta \delta} M_{\lambda_{k-1}(v)})} < \infty \text{ a.s.}
\]

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

\[
\sum_{j=1}^{\infty} a_k |(\hat{w}_{\tau_k} - w_{\tau_k})\tau_k| < \infty \text{ a.s.,}
\]

where \( 0 < M' < \infty \).

Further, by (A18) in Appendix it follows that

\[
\sum_{j=1}^{\infty} a_k |(\hat{w}_{\tau_k} - w_{\tau_k})e_{\tau_k}| < \infty \text{ a.s.,}
\]

\[
\sum_{j=1}^{\infty} a_k |(\hat{w}_{\tau_k} - w_{\tau_k})| \cdot |e_{\tau_k}| \leq \sum_{j=1}^{\infty} M \frac{1}{k^{3/2 - 3\beta \delta}} |e_{\tau_k}|
\]

\[
= \sum_{j=1}^{\infty} M |e_{\tau_k}| \left( \frac{1}{(k^{3/2 - 3\beta \delta})^2} \right) + \sum_{j=1}^{\infty} \frac{M E |e_{\tau_k}|}{k^{3/2 - 3\beta \delta}} < \infty \text{ a.s.}
\]

Combining (33), (34) and (35) implies (32).  \( \square \)

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

\[
\mu_{k+1}(v) = \mu_k(v) - a_k \hat{w}_{\tau_k} (\mu_k(v) - z_{\tau_k}),
\]

and

\[
\|\mu_{k+1}(v) - \mu_k(v)\| \leq cT, \text{ s } s = k, k + 1, \ldots, m(k_i, T)
\]

for all sufficiently large \( i \) and small enough \( T > 0 \), where \( c > 0, c \) may depend on sample path \( \omega \) but is independent of \( k_i \).
Proof. Before proving the lemma we first note that here a convergent subsequence \( \{ \mu_k(v) \} \) of \( \{ \mu_k(v) \} \) is considered, and from the subsequent proof it can be seen that replacing \( \mu_k(v) \) 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 \( \mu_k(v) \).

Set
\[
\Phi_{i,j} := (1 - a_i w_{r_1}) \cdots (1 - a_j w_{r_j}),
\]
\( i > j \), \( \Phi_{j,j} = 1 \).

By (A14) in Appendix and the fact that \( E w_{r_k} \to w_0 \) proved in Lemma C it follows that
\[
\sum_{j=k_i}^s a_j w_{r_j} = O(T) \quad \forall s \in [k_i, \ldots, m(k_i, T)]
\]
and hence
\[
\log \Phi_{s,k_i} = O \left( \sum_{j=k_i}^s a_j w_{r_j} \right), \quad \Phi_{s,k_i} = 1 + O(T) \quad \forall s \in [k_i, \ldots, m(k_i, T)]
\]
as \( i \to \infty \) and \( T \to 0 \).

By (A19) we have
\[
\left| \sum_{j=k_i}^s \Phi_{s,j+1} a_j w_{r_j} z_{r_j} \right| = O(T)
\]
and by (38) and the boundedness of \( \{ y_{r_j} \} \)
\[
\sum_{j=k_i}^s \Phi_{s,j+1} a_j w_{r_j} y_{r_j} = O \left( \sum_{j=k_i}^s a_j w_{r_j} \right) = O(T)
\]
as \( i \to \infty \) and \( T \to 0 \).

Combining (39), (40) and (41) yields
\[
\Phi_{s,k_i} \mu_k(v) + \sum_{j=k_i}^s \Phi_{s,j+1} a_j w_{r_j} z_{r_j} = \mu_k(v) + O(T)
\]
\( \forall s \in [k_i, \ldots, m(k_i, T)] \).

From (36) we have
\[
\mu_{s+1}(v) = \mu_s(v) - a_s w_{r_s} (\mu_s(v) - z_{r_s})
+ a_s (w_{r_s} - \hat{w}_{r_s}) (\mu_s(v) - z_{r_s})
= \Phi_{s,k_i} \mu_k(v) + \sum_{j=k_i}^s \Phi_{s,j+1} a_j w_{r_j} z_{r_j}
+ \sum_{j=k_i}^s \Phi_{s,j+1} a_j (w_{r_j} - \hat{w}_{r_j}) (\mu_j(v) - z_{r_j}),
\]
which incorporating with (42) by Lemma 1 implies
\[
\mu_{s+1}(v) = \mu_k(v) + O(T) \quad \forall s \in [k_i, \ldots, m(k_i, T)].
\]

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

The set \( \Theta_0 \) 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
\[
\mu_k(v) \to f(v) \quad a.s.,
\]
where \( \mu_k(v) \) is given by (30) and (31).

Proof. Write (30) as
\[
\begin{align*}
\mu_{k+1}(v) &= \mu_k(v) - a_k w_0 (\mu_k(v) - f(v)) + a_k e_k(v) \\
&\times I(|\mu_k(v) - a_k w_0 (\mu_k(v) - f(v)) + a_k e_k(v)| \leq M_k(v)),
\end{align*}
\]
where \( w_0 \) is given in Lemma C, and
\[
e_k(v) = w_0 (\mu_k(v) - f(v)) - \hat{w}_{r_k} (\mu_k(v) - z_{r_k}).
\]

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

for any convergent \( \mu_k(v) \to \mu(v) \), where
\[
e_k^{(1)}(v) + e_k^{(2)}(v) + e_k^{(3)}(v) + e_k^{(4)}(v) = e_k(v),
\]
and
\[
e_k^{(1)}(v) \triangleq (w_{r_k} - \hat{w}_{r_k}) (\mu_k(v) - y_{r_k}),
\]
\[
e_k^{(2)}(v) \triangleq (w_0 - w_{r_k}) \mu_k(v),
\]
\[
e_k^{(3)}(v) \triangleq \hat{w}_{r_k} y_{r_k} - w_0 f(v),
\]

and \( e_k^{(4)}(v) \) is free of \( v \) and equals \( e_k^{(4)} \) with
\[
e_k^{(4)} \triangleq w_{r_k} e_{r_k}.
\]

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

\[
\lim_{T \to 0} \limsup_{i \to \infty} \frac{1}{T} \left| \sum_{j=k_i}^{m(k_i,T_i)} a_j e_j^{(2)}(v) \right| \\
\leq \lim_{T \to 0} \limsup_{i \to \infty} \frac{1}{T} \left| \sum_{j=k_i}^{m(k_i,T_i)} a_j (\mu_j(v) - \mu(v)) (w_{\tau_j} - E w_{\tau_j}) \right| \\
+ \lim_{T \to 0} \limsup_{i \to \infty} \frac{1}{T} \left| \mu(v) \sum_{j=k_i}^{m(k_i,T_i)} a_j (w_{\tau_j} - E w_{\tau_j}) \right| \\
+ \lim_{T \to 0} \limsup_{i \to \infty} \frac{1}{T} \left| \sum_{j=k_i}^{m(k_i,T_i)} a_j \mu_j(v) (E w_{\tau_j} - w_0) \right|
\]

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. \( \square \)

5. Numerical example

Let the linear subsystem be a second order MA process

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

and let the nonlinear function be

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

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 \( \tau_k \).

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...
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 \( \tau_k \), \( k = 1, 2, \ldots \) 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 \( \tau \) be a stopping time with respect to a sequence \( \{ \mathcal{F}_k \} \) of nondecreasing \( \sigma \)-algebras. If \( \mathcal{F}_k \) is independent of \( \sigma(\xi_j, \ j > k) \), \( k \geq 1 \), then \( \mathcal{F}_\tau \) is independent of \( \sigma(\xi_\tau+1, \xi_{\tau+2}, \ldots) \) and \( \{ \xi_\tau+k, \ k \geq 1 \} \) is a sequence of iid random variables with the same distribution as that for \( \xi_1 \).

Define
\[
\tau \triangleq \tau_1 - (r + 1) = \inf\{k \geq 1 : |u_{k+r+1-j}| \leq c, \ j = 1, \ldots, r + 1\}. \quad (A1)
\]

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

**Proof.** We first show that for \( E\tau_k < \infty \) and \( E\tau^2_k < \infty \) it suffices to prove
\[
E\tau < \infty \quad \text{and} \quad E\tau^2 < \infty. \quad (A2)
\]

It is clear that \( \tau < \infty \) a.s. and \( \tau \) is a stopping time with respect to \( \mathcal{G}_k \triangleq \sigma(u_j, j = 1, \ldots, k + r) \). Then for any integer \( k \geq 0 \) there is a Borel set \( B_{j+r} \) such that \( \{ \tau = j \} = B_{j+r} \in B_{j+r+1} \). By Proposition 1 it follows that
\[
P[\tau = j] = P[(u_1, \ldots, u_{j+r}) \in B_{j+r}] = P[(u_{j+1}, \ldots, u_{j+r}) \in B_{j+r}] = P[\tau_j - \tau_{j-1} = j] \quad \text{with} \quad \tau_0 \triangleq r + 1, \quad (A3)
\]
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 \( \mathcal{G}_{\tau_k-1} \) and \( \sigma(u_{k-1}, u_{k-2}, \ldots) \), respectively. Consequently, we have
\[
E\tau_k = E\tau_0 + \sum_{j=1}^{r} E(\tau_j - \tau_{j-1}) = (r + 1) + kE\tau
\]
and
\[
E\tau^2_k = E\left( \tau_0 + \sum_{j=1}^{r} (\tau_j - \tau_{j-1}) \right)^2 \leq (r + 1)^2 + 2(r + 1)kE\tau + k^2E\tau^2.
\]

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

We now prove (A2).

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

Set \( p \triangleq P[|u_1| \leq c] \) 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 \( n \geq 2k \) and let \( P_n(k) \) be the probability that in \( n \) trials there are no \( k \) successive successes.

It is clear that
\[
P_n(k) = qP_{n-1}(k) + pqP_{n-2}(k) + \cdots + p^{k-1}qP_{n-k}(k), \quad (A4)
\]
or

\[ A(z) P_n(k) = 0, \]

where

\[ A(z) = 1 - qz - pqz^2 - \cdots - p^{k-1}qz^k \quad (A5) \]

with \( z \) being the backward-shift operator.

Since

\[ |qz + pqz + \cdots + p^{k-1}qz^k| \leq \frac{q(1 - p^k)}{1 - p} = 1 - p^k < 1 \]

on the unit circle \(|z|=1\), by the Rouche 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

\[ P_n(k) \leq C(k) \rho^n(k). \quad (A6) \]

For \( n \geq 2(r + 1) + 2 \), 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 \( |u_{n+i}| \leq c \) for all \( i = 0, 1, \ldots, r \). Thus, by (A6) we have

\[ P[\tau = n] = P_{n-2}(r + 1)q^r \rho^{r+1} \leq C(r + 1) \rho^{n-r}(r + 1)q^r \rho^{r+1}, \]

and hence

\[ E\tau = \sum_{n=1}^{\infty} n P[\tau = n] \leq \sum_{n=2r+4}^{\infty} \sum_{n=1}^{\infty} n P[\tau = n] \leq \sum_{n=1}^{\infty} n P[\tau = n] + C(r + 1)q^r \rho^{r+1} \sum_{n=2r+4}^{\infty} \rho^{n-r}(r + 1) < \infty, \]

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

\[ \frac{\tau_k}{k} \xrightarrow{k \to \infty} E\tau \quad \text{a.s.} \]

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

---

**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 \( \{\mathcal{F}_j\} \) and \( \tau_1 < \tau_2 < \cdots < \tau_k \), we see that \( \{\xi_{\tau_1}, \ldots, \xi_{\tau_k}\} \) are \( \mathcal{F}_{\tau_k} \)-measurable. On the other hand, by the definition of \( \tau_{k+1} \), \( \xi_{\tau_{k+1}} \) is measurable with respect to \( \sigma(\xi_{\tau_{k+1}}, \xi_{\tau_{k+2}}, \ldots) \) which is independent of \( \mathcal{F}_{\tau_k} \) 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

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

Introduce the following notations:

\[ s^+ \triangleq c + \sum_{j=1}^{r} d_j s_j - v, \quad s^- \triangleq -c + \sum_{j=1}^{r} d_j s_j - v, \quad (A7) \]

\[ B^+(\varepsilon) \triangleq \{ (s_1, s_2, \ldots, s_r) \in [-c, c]^r : s^+ > \varepsilon, s^- < \varepsilon, \varepsilon \geq 0 \}, \]

\[ B^0(\varepsilon) \triangleq \{ (s_1, s_2, \ldots, s_r) \in [-c, c]^r : s^+ > \varepsilon, s^- < \varepsilon, s^+ < \varepsilon, \varepsilon \geq 0 \}, \]

\[ B^-(-\varepsilon) \triangleq \{ (s_1, s_2, \ldots, s_r) \in [-c, c]^r : s^+ < -\varepsilon, \varepsilon \geq 0 \}. \]

Clearly, \( B^+(\varepsilon) \), \( B^0(\varepsilon) \) and \( B^-(-\varepsilon) \) are bounded sets. Denote their Lebesgue measures by \( V(B^+(\varepsilon)) \), \( V(B^0(\varepsilon)) \) and \( V(B^-(-\varepsilon)) \), respectively.

**Lemma C.** Assume H0–H2 hold. Then

\[ Ew_{\tau_k} \xrightarrow{k \to \infty} w_0, \quad E[w_{\tau_k} f(v_{\tau_k})] \xrightarrow{k \to \infty} w_0 f(v) \quad (A8) \]

and

\[ \sup_k E(\sqrt{b_k w_{\tau_k}})^2 < \infty, \quad (A9) \]

where \( w_0 = \frac{V(B^+(0))}{(2\pi)^r p^{r+1}} \) and \( p = P[|u_k| \leq c] \).
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

$$E[w_{\tau_k} f(v_{\tau_k}) I_{\{\tau_k = n\}}] = \frac{1}{b_k} E\left[ e^{-\frac{(u_n - v_i)/b_k}{2} + \frac{1}{2} \sum_{j=0}^{r+1} s_j^2} f(v_n) \right]$$

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

$$= P(u_1, \ldots, u_{n-r-2}) \in B_{n-r-2}(k)] \int_{-c}^{c} \cdots \int_{-c}^{c} \exp \left\{ -\left( \sum_{j=0}^{r} d_j x_j - v \right)^2 \right\}$$

$$\times f \left( \sum_{j=0}^{r} d_j x_j \right) dx_0 \cdots dx_r$$

$$= \frac{1}{(\sqrt{\pi})^{r+1} b_k} \int_{-c}^{c} \cdots \int_{-c}^{c} \exp \left\{ -\left( \sum_{j=0}^{r} d_j x_j - v \right)^2 \right\}$$

$$\times f \left( \sum_{j=0}^{r} d_j x_j \right) dx_0 \cdots dx_r$$

$$= \int_{-c}^{c} \cdots \int_{-c}^{c} ds_1 \cdots ds_r$$

$$\times \left( \int_{-\infty}^{c} \frac{e^{-s^2}}{\sqrt{2\pi}} f(v + b_k s_0) ds_0 \right).$$

(A10)

where

$$s_0 = x_0 + d_1 x_1 + \cdots + d_r x_r - v$$

$$s_j = x_j, \ j = 1, 2, \ldots, r,$$

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

Noticing

$$s^+/b_k \rightarrow \infty \quad \text{and} \quad s^-/b_k \rightarrow -\infty$$

in $B_+^\pm(c)$ with $c > 0$, we find that

$$\int_{B_+^\pm(c)} ds_1 \cdots ds_r \left( \int_{-\infty}^{c} e^{-s^2} f(v + b_k s_0) ds_0 \right)$$

$$\rightarrow \sqrt{2\pi} V(B_+^\pm(c)) f(v),$$

and by tending $c \rightarrow 0$

$$\int_{B_+^\pm(0)} ds_1 \cdots ds_r \left( \int_{-\infty}^{c} e^{-s^2} f(v + b_k s_0) ds_0 \right)$$

$$\rightarrow \sqrt{2\pi} V(B_+^\pm(0)) f(v).$$

(A11)

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

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

$$E[w_{\tau_k} f(v_{\tau_k})] = \sum_{n=1}^{\infty} E[w_{\tau_k} f(v_{\tau_k}) I_{\{\tau_k = n\}}]$$

$$= \frac{1}{(\sqrt{2\pi})^{r+1} b_k} \int_{-c}^{c} \cdots \int_{-c}^{c} ds_1 \cdots ds_r$$

$$\times \left( \int_{-\infty}^{c} \frac{e^{-s^2}}{\sqrt{2\pi}} f(v + b_k s_0) ds_0 \right)$$

$$\rightarrow \frac{1}{k \rightarrow \infty} (\sqrt{2\pi})^{r+1} V(B_+^\pm(0)) f(v) = w_0 f(v).$$

(A12)

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

By a similar approach we find that

$$E(\sqrt{b_k} w_{\tau_k})^2$$

$$= \frac{1}{(\sqrt{2\pi})^{r+1} b_k} \int_{-c}^{c} \cdots \int_{-c}^{c} ds_1 \cdots ds_r$$

$$\times \left( \int_{-\infty}^{c} e^{-s^2} f(v + b_k s_0) ds_0 \right)$$

$$\rightarrow \frac{1}{k \rightarrow \infty} (\sqrt{2\pi})^{r+1} V(B_+^\pm(0)) f(v) = w_0 f(v).$$

(A13)

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

Lemma D. Assume H0–H2 hold. Then

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

(A14)

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

(A15)

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

(A16)

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

(A17)

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

(A18)

and

$$\sum_{j=1}^{\infty} a_j (w_{\tau_j} |\varepsilon_{\tau_j}| - E w_{\tau_j} |\varepsilon_{\tau_j}|) < \infty \ a.s.$$
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, 
\[
\left\langle \sqrt{b_j} w_{1k} - E \sqrt{b_j} w_{1k}, \right\rangle < \infty \quad \text{a.s.,}
\]
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 \( \{v_k\} \) we find that
\[
E[w_{1k}^2 v_{1k}^2 I[\tau_k=n]] = E \left[ \exp \left\{ -2 \left( \frac{v_n - v}{b_k} \right)^2 + \sum_{j=1}^{r+1} u_{n-j}^2 \right\} \right] \times \sum_{n=1}^{\infty} P[\tau_k=n] \sup_k E(\sqrt{b_k} w_{1k})^2
\]
and hence by H2 and Lemma C
\[
E[b_k w_{1k}^2 v_{1k}^2] = E \sup_k E(\sqrt{b_k} w_{1k})^2 < \infty.
\]

By the similar treatment and using the independence of \( \{u_k\} \) and \( \{v_k\} \) and Lemma B we find that
\[
E[w_{1k} v_{1k}] = 0.
\]

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

References


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.