East Asian Journal on Applied Mathematics
doi: 10.4208/eajam.290913.111113a

Vol. **3**, No. 4, pp. 311-332
November 2013

Recursive Identification of Wiener-Hammerstein Systems with Nonparametric Nonlinearity

Xiao-Li Hu[1,*] and Yue-Ping Jiang[2]

[1] *School of Electrical Engineering and Computer Science, University of Newcastle, Newcastle NSW 2308, Australia.*
[2] *School of Natural Sciences, Nanjing University of Posts and Telecommunications, Nanjing 210023, Jiangsu Province, P. R. China.*

Received 29 September 2013; Accepted (in revised version) 11 November 2013

Available online 28 November 2013

Abstract. A recursive scheme is proposed for identifying a single input single output (SISO) Wiener-Hammerstein system, which consists of two linear dynamic subsystems and a sandwiched nonparametric static nonlinearity. The first linear block is assumed to be a finite impulse response (FIR) filter and the second an infinite impulse response (IIR) filter. By letting the input be a sequence of mutually independent Gaussian random variables, the recursive estimates for coefficients of the two linear blocks and the value of the static nonlinear function at any fixed given point are proven to converge to the true values, with probability one as the data size tends to infinity. The static nonlinearity is identified in a nonparametric way and no structural information is directly used. A numerical example is presented that illustrates the theoretical results.

AMS subject classifications: 60G35, 62F12, 93E12

Key words: Wiener-Hammerstein system, nonparametric nonlinearity, recursive estimate, strong consistence.

1. Introduction

The Wiener-Hammerstein (W-H) system comprises two linear dynamic subsystems with a sandwiched static nonlinearity — cf. Fig. 1. A system consisting of the first two blocks is called a Wiener system, whereas a system consisting of the last two blocks is known as a Hammerstein system. Thus W-H systems are a natural extension of Wiener systems and Hammerstein systems, which are all important for modelling many real phenomena. Applications include a distillation column [1], a pH control process [2] for Wiener systems, a cat visual cortex for Hammerstein systems [3], and a light flickering severity-meter for W-H systems [4]. Identification of these systems has therefore been an active research

*Corresponding author. *Email address:* `xlhu@amss.ac.cn` (X.-L. Hu)

area for many years — e.g. see Refs. [5–12] for Wiener systems, [13–16] for Hammerstein systems, and [4, 17–21] for W-H systems.

Since a W-H system is a combination of a Wiener system and a Hammerstein system, a natural starting point is to identify those two simpler components before considering a W-H system. The identification for Wiener systems is much more difficult than for Hammerstein systems, with the essential difference in identification of these two types of nonlinear systems roughly as follows. In the case of a Hammerstein system, if a sequence of independent identically distributed (i.i.d.) random variables is selected as input, it still remains a sequence of i.i.d. random variables after passing through the static nonlinearity of the Hammerstein system. Hence identifying the linear subsystem in a Hammerstein system is a standard Auto-Regressive and Moving Average (ARMA) model issue, where the random variables (outputs of the nonlinearity) in the Moving Average (MA) part may not be zero-mean. However, it is quite different in the case of a Wiener system for i.i.d. input. The intermediate signal (the output of the linear subsystem) is no longer mutually independent, and in general has a complicated distribution unless the input of the linear subsystem is Gaussian. Moreover, the unboundedness of any Gaussian signal may cause additional difficulties in the convergence analysis when using a Gaussian input. This explains why more restrictive conditions are used and less theoretical results are obtained for identifying Wiener systems in comparison with Hammerstein systems. Thus the chief key difficulty in identification for W-H systems is how to identify Wiener systems.

Due to the weak properties of the intermediate signal of a Wiener system, it is no wonder that the static nonlinearity is usually assumed to be invertible or described in a simple parametric form (typically as a low order polynomial). If the order of the polynomial is high, the number of whole terms in the system input-output expansion will be huge. However, even under such strict restrictions, it is still quite difficult to establish convergence results.

Nonparametrisation for the nonlinearity is another way, and some weak convergence results have been established in this context [5]. Several recursive identification algorithms for Wiener systems have also been proposed under different system settings [7–9], and their strong consistences are justified. The first guaranteed consistent recursive identification algorithm for the Wiener system is given in Ref. [7], where it is unnecessary for the nonparametric nonlinear function to be inversive. However, a certain amount of data is excluded for usage there due to a technical reason, and this drawback was cured in Ref. [9]. The Wiener system with a more general linear subsystem is introduced and considered in Ref. [8]. The main idea there is that the coefficients of the linear subsystem are estimated by a system of equations comprising cross-correlation coefficients of input and output, when the nonlinearity is estimated by the estimated intermediate signal and the output with the kernel function technique. It seems possible that nonparametrisation may be used in identifying W-H systems.

Although a number of identification algorithms have been proposed for W-H system (e.g. see [4, 17, 19–21]), relevant theoretical guarantees for the consistency of the corresponding identification algorithms seem rare. In Ref. [17], the best linear approximation method carried out to serve as an initial estimation for W-H systems is actually a parametric way to deal with W-H systems. As in the case of a Wiener system, if the static nonlinearity is in a parametric form the complete expression of the input-output would generally be very long, so in this paper we set the nonlinearity in a nonparametric form. Based on this, we design a recursive identification scheme for W-H systems and prove its strong consistency. The coefficients of the two linear subsystems and the value of the static nonlinear function at any fixed given point are precisely identified recursively. It is worth pointing out that the static nonlinearity is identified in a nonparametric way, and need not be invertible. Meanwhile, the strong consistency (convergence with probability one) of the proposed algorithms is established in Theorems 4.1 and 4.2. A numerical example is tested to justify the theoretical analysis.

The main technical contributions of this article are as follows.

- We avoid using the socalled stochastic approximation with expanding truncations technique as in Refs. [7–9,14], and a direct scheme in a briefer form is proposed and analysed here. This increases the possibility that the algorithms may be extended to the more complicated closed-loop case.

- Some new analytic techniques are introduced to deal with certain stochastic series, when implementing the main idea of Refs. [7–9] mentioned earlier. Thus the coefficients of the linear subsystems are estimated by a system of equations involving the cross-correlation coefficients of input and output, and the nonlinearity is estimated by the estimated first intermediate signal v_k and the estimated secondary intermediate signal x_k with the kernel function technique — cf. in Fig. 1. The non-singularity of the relevant matrix of cross-correlation coefficients is analysed and sufficient conditions are proposed.

It is also worth pointing out that the proposed identification scheme here is a open loop case, but may still shed some light on more complicated closed loop cases.

The statement of the problem and the proposed estimation algorithms are given in Section 2. Some preliminaries to facilitate theoretical analysis are introduced in Section 3. The main results establishing the strong consistency of the estimation algorithms are discussed in Section 4, and then demonstrated by a simple numerical example in Section 5. Finally, our concluding remarks are made in Section 6.

2. The Problem and Estimation Algorithms

The structure and assumptions of the W-H system to be considered are introduced and discussed in Subsection 2.1. Relevant equations involving the input-output correlation coefficients and coefficients of the two linear blocks are established in Subsection 2.2, as theoretical preliminaries for our identification algorithms. Finally, recursive identification algorithms for the two linear blocks and the nonparametric nonlinearity are proposed in Subsection 2.3.

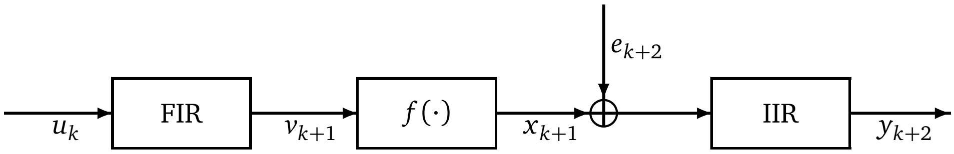

Figure 1: Wiener-Hammerstein system.

2.1. Problem setting

We consider a special setting of SISO W-H systems depicted in Fig. 1, since it is difficult to study the identification issue for general W-H systems. With input denoted by u_k and output by y_k, the two linear blocks are

$$v_{k+1} = A(z)u_k \,, \tag{2.1}$$
$$B(z)y_{k+2} = x_{k+1} + e_{k+2} \,, \tag{2.2}$$

where

$$A(z) = 1 + a_1 z + \cdots + a_p z^p \,, \qquad B(z) = 1 + b_1 z + \cdots + b_q z^q \,, \tag{2.3}$$

and the intermediate static nonlinearity is

$$x_{k+1} = f(v_{k+1}) \,. \tag{2.4}$$

For simplicity, the time delays in both linear subsystems are set to be 1. However, identification algorithms and convergence results similar to those established will hold for other time delays.

The first linear block is a finite impulse response (FIR) filter and the second an infinite impulse response (IIR) filter, with all the parameters connected with output y_k. Since the first linear block is a FIR filter, if the input $\{u_k\}$ is a sequence of independent random variables then the subsequence $\{v_{l+(p+1)k} : k = 1,2,\cdots\}$ of the output of this FIR filter is still an independent sequence, for any fixed $l = 0,1,\cdots,p$. This is very convenient to analyse some stochastic series, as shown in Lemma 3.4 below. Given these forms of the two linear blocks, all parameters of the two linear filters are easy to recover by a system of cross correlation equations as in Eqs. (2.8) and (2.9) — although from a practical point of view it may be restrictive for the first linear part due to its FIR structure. Since a Moving Average (MA) linear dynamic model can closely approximate a general model with a real rational transfer function, this restriction is not severe in a real application if the order p is sufficient large. We assume that the noise is introduced before the last linear block, which is more complicated than the case where the noise is included as a mere white observation noise afterward.

The problem is to recursively estimate the coefficients a_i, $i = 1,\cdots,p$ and b_j, $j = 1,\cdots,q$ of the linear subsystems and the value $f(v)$ at any v, given the observations $\{y_k\}$ and inputs $\{u_k\}$. We now list the conditions used, and also provide some brief explanations.

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

H2. The polynomials $A(z)$ and $B(z)$ have no common zeros, and $B(z)$ is stable — i.e. $B(z) \neq 0$ for any $|z| \leq 1$.

H3. The input $\{u_k\}$ of the Wiener-Hammerstein system is a sequence of i.i.d. Gaussian random variables $u_k \in \mathcal{N}(0,1)$, independent of the observation noise $\{e_k\}$. The observation noise $\{e_k\}$ is a sequence of mutually independent random variables with $Ee_k = 0$ and $\sup_k Ee_k^2 < \infty$.

The growth rate of $f(\cdot)$ required in H1 is used to guarantee the existence of moments of v_k. Condition H2 is a standard setting for the identifibility and stability of linear dynamic subsystems, which is clearer for the intermediate nonlinearity $f(x) \equiv x$. Technically, H2 is also needed to avoid pole-zero cancellation in the important formula (2.16) below. When a Gaussian signal is the input, the intermediate signal v_k's distribution is still Gaussian, which is the main reason why we select a Gaussian signal here. For simplicity, we choose the standard Gaussian, but similar theoretical results still hold for any general distribution $\mathcal{N}(0,\sigma^2)$.

2.2. Correlation analysis

The identification idea is to first estimate the coefficients of the two linear subsystems, and then recover the nonlinear block. This subsection serves as a theoretical preparation for identification algorithms of the first step — i.e. for estimations of the coefficients of the two linear blocks. Some cross-correlation coefficients of the input $\{u_k\}$ and output $\{y_k\}$ are introduced and analysed below, and some algebraic equations connected with the cross-correlation coefficients and linear coefficients of the W-H system are then derived. The theoretical analysis is under the assumptions H1-H3.

Let us first point out that the expectation $E[u_{k-i}y_{k+2}]$ is independent of k. From the stability of $B(z)$ under H2, we assume that

$$B^{-1}(z) = \sum_{i=0}^{\infty} \bar{b}_i z^i \,, \tag{2.5}$$

where $\bar{b}_0 = 1$ and $|\bar{b}_k| = O\left(e^{-\tau k}\right)$ for $\tau > 0$ — i.e. $|\bar{b}_k|$ exponentially decays. From Eq. (2.2), we have

$$y_{k+2} = B^{-1}(z)x_{k+1} + B^{-1}(z)e_{k+2} = \sum_{i=0}^{\infty} \bar{b}_i (x_{k+1-i} + e_{k+2-i}) \,;$$

and from Eq. (2.1)(2.4) and the independence in H3, we obtain

$$E[u_{k-i}y_{k+2}] = \sum_{s=0}^{i} \bar{b}_s E[u_{k-i}x_{k+1-s}] \,.$$

Since $\{x_k\}$ is stationary and the coefficients $\{\bar{b}_s\}$ are completely determined by $B(z)$, the expectation $E[u_{k-i}y_{k+2}]$ is independent of k, hence we define

$$\gamma_i \triangleq E[u_{k-i}y_{k+2}] , \quad i = 0, 1, \cdots, p + q . \tag{2.6}$$

It is notable that the input-output correlation coefficient γ_i can be estimated directly from the empirical mean of input-output data as $\frac{1}{n}\sum_{k=1}^{n} u_{k-i}y_{k+2}$ at n-th time. Consequently, if some algebraic equations that are connected with γ_i and the coefficients of the two linear blocks are established, then the coefficients a_i and b_j may be found from these equations.

Another form of the second linear block (2.2) is

$$y_{k+2} = -b_1 y_{k+1} - \cdots - b_q y_{k+2-q} + x_{k+1} + e_{k+2} ;$$

and multiplying this formula by u_{k-i} for $i = 0, 1, \cdots, p+q$ and then taking the expectation, we obtain

$$\gamma_0 = E[u_k x_{k+1}] , \tag{2.7}$$

$$\gamma_i = -\sum_{l=1}^{i\wedge q} b_l \gamma_{i-l} + a_i \gamma_0 , \quad i = 1, \cdots, p , \tag{2.8}$$

$$\gamma_j = -\sum_{l=1}^{j\wedge q} b_l \gamma_{j-l} , \quad j = p + 1, \cdots, p + q , \tag{2.9}$$

where $i \wedge j \triangleq \min\{i, j\}$. The second term $a_i\gamma_0$ on the right-hand side of Eq. (2.8) is derived using Lemma 3.3 — i.e. we have $E[u_{k-i}x_{k+1}] = a_i\gamma_0$. Clearly, if $\gamma_0, \gamma_1, \cdots, \gamma_{p+q}$ are already known and the $p + q$ equations are independent, the parameters a_i and b_j can be calculated from Eqs. (2.8)-(2.9) since there are $p + q$ unknown parameters and $p + q$ linear algebraic equations. As previously mentioned, $\gamma_0, \gamma_1, \cdots, \gamma_{p+q}$ can be estimated empirically in terms of the input-output data. Thus the main idea to identify the two linear blocks is accomplished, as mentioned above.

In order to represent Eq. (2.9) in matrix form, let us define

$$\Upsilon(p, q) \triangleq \begin{pmatrix} \gamma_p & \gamma_{p-1} & \cdots & \gamma_{p-q+1} \\ \gamma_{p+1} & \gamma_p & \cdots & \gamma_{p-q+2} \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_{p+q-1} & \gamma_{p+q-2} & \cdots & \gamma_p \end{pmatrix} , \tag{2.10}$$

where $\gamma_{-i} = 0$ for $i > 0$ by definition. Thus we have the vector equation

$$[\gamma_{p+1}\ \gamma_{p+2}\ \cdots\ \gamma_{p+q}]^T = -\Upsilon(p, q)[b_1\ b_2\ \cdots\ b_q]^T , \tag{2.11}$$

which is similar to the socalled Yule-Walker equation for a linear system. Eq. (2.11) serves as the “breaking point" to identify the linear blocks, and it is of key importance to clarify the conditions that guarantee the invertibility of $\Upsilon(p, q)$ — cf. Lemma 3.1 below.

Let us next point out an important formula to be used in Lemma 3.1, which connects $\{\gamma_i\}$ and the linear coefficients $\{a_i, i = 1, \cdots, p\}$ and $\{b_j, j = 1, \cdots, q\}$. Noticing that $\gamma_{-i} = 0$ for $i > 0$, from Eqs. (2.8) and (2.9) we have

$$B(z)\gamma_i = a_i\gamma_0 , \quad 0 \le i \le p , \tag{2.12}$$

$$B(z)\gamma_j = 0 , \quad j > p . \tag{2.13}$$

The cross-correlation generating function of $\{\gamma_i\}$ is

$$\Gamma(z) \triangleq \sum_{k=-\infty}^{\infty} \gamma_k z^k , \tag{2.14}$$

which satisfies

$$\begin{aligned} \Gamma(z)B(z) &= \sum_{k=-\infty}^{\infty} B(z)\gamma_k z^k = \sum_{k=0}^{\infty} B(z)\gamma_k z^k \\ &= \sum_{k=0}^{p} B(z)\gamma_k z^k = \gamma_0 \sum_{k=0}^{p} a_k z^k \\ &= \gamma_0 A(z) \end{aligned} \tag{2.15}$$

from Eqs. (2.12) and (2.13), hence we obtain the formula

$$\Gamma(z) = \frac{\gamma_0 A(z)}{B(z)} . \tag{2.16}$$

Since $\gamma_i = E[u_{k-i}y_{k+2}]$ is included and the output y_{k+2} can be expressed in terms of $f(\cdot)$, in an indirect way the formula (2.16) does contain some information on $f(\cdot)$.

2.3. Identification algorithms

The identification problem may be viewed as two steps — the first to estimate the coefficients of the two linear blocks, and the second to identify the nonparametric function $f(v)$ at a given point v. In practice, we can simultaneously estimate $f(\cdot)$ at (say) 100 uniformly distributed points, to find its graph. On recalling the definition of the correlation coefficient (2.6), it is natural to calculate the k-th estimation of γ_i from the available input-output data as

$$\hat{\gamma}_i(k) = \frac{1}{k} \sum_{l=1}^{k} u_{l-i} y_{l+2} . \tag{2.17}$$

Equivalently, γ_i can be estimated recursively as

$$\hat{\gamma}_i(k) = \hat{\gamma}_i(k-1) - \frac{1}{k} \left(\hat{\gamma}_i(k-1) - u_{k-i} y_{k+2} \right) , \tag{2.18}$$

with the initial real values $\hat{\gamma}_i(0) = 0$, $i = 0, 1, \cdots, p+q$. After $\hat{\gamma}_i(k)$, $i = 0, 1, \cdots, p+q$, are obtained, the k-th estimates for $b_1, \cdots, b_q$ (denoted by $\hat{b}_{1k}, \cdots, \hat{b}_{qk}$) can be calculated from

Eq. (2.11) by replacing Υ with its k-th estimation $\hat{\Upsilon}_k$, derived by replacing each element of Υ with the corresponding $\hat{\gamma}_i(k)$ given by

$$[\hat{b}_{1k}, \cdots, \hat{b}_{qk}]^T = -\hat{\Upsilon}_k^{-1}[\hat{\gamma}_{p+1}(k), \cdots, \hat{\gamma}_{p+q}(k)]^T \,, \tag{2.19}$$

if $\hat{\Upsilon}_k$ is invertible.

From Theorem 4.1, under the assumptions H1-H3 we have $\hat{\Upsilon}_k \to \Upsilon$ as $k \to \infty$, with the same convergence rate of $\hat{\gamma}_i(k)$ to γ_i — i.e.

$$\|\hat{\Upsilon}_k - \Upsilon\| = o(k^{-\delta}) \,,$$

where $\delta \in (0, 1/2)$ and $\|\cdot\|$ denotes the matrix Euclidean norm. Further, if $\gamma_0 \neq 0$ then Υ is nonsingular from Lemma 3.1, hence $\hat{\Upsilon}_k$ is clearly invertible when k is sufficiently large. To avoid initial singularities, $\hat{\Upsilon}_k$ can be replaced by $(\frac{1}{k^\nu} I + \hat{\Upsilon}_k)$ where I is the corresponding identity matrix and $\nu \geq 1$, so estimates of b_j can be constructed using Eq. (2.19). Then from Eq. (2.8) the k-th estimates for the parameters a_i, $i = 1, \cdots, p$ (denoted by $\hat{a}_{ik}$, $i = 1, \cdots, p$) are calculated as

$$\hat{a}_{ik} = \frac{1}{\hat{\gamma}_0(k)} \left(\hat{\gamma}_i(k) + \sum_{l=1}^{i \wedge q} \hat{b}_{lk} \hat{\gamma}_{i-l}(k) \right), \quad i = 1, \cdots, p \,, \tag{2.20}$$

where $\hat{\gamma}_0(k)$ can also be replaced by $\hat{\gamma}_0(k) + 1/k^{\upsilon}$ with $\upsilon \geq 1$ to overcome any initial singularity. Thus all of the parameters of the two linear blocks are estimated at step k by the recursive algorithms (2.18)-(2.20), and the strong consistency of these estimation algorithms is ensured from Theorem 4.1 given below.

Now let us turn to the identification issue for the nonlinear part. After the estimates for the parameters a_i, $i = 1, \cdots, p$ and b_j, $j = 1, \cdots, q$ are obtained, the output v_{k+1} and the deviation of the linear subsystem are reconstructed by

$$\hat{v}_{k+1} \triangleq u_k + \hat{a}_{1k} u_{k-1} + \cdots + \hat{a}_{pk} u_{k-p} \,, \tag{2.21}$$

$$\hat{\sigma}_k \triangleq \left(\sum_{i=0}^{p} \hat{a}_{ik}^2 \right)^{1/2} , \tag{2.22}$$

where $\hat{a}_{0k} = 1$. In order to estimate $f(v)$ at a given v, we introduce the kernel (or weight) function

$$w_k \triangleq \sqrt{2}\, \sigma k^{\alpha} \exp\left\{ -(v_k - v)^2 k^{2\alpha} + \frac{v_k^2}{2\sigma^2} \right\} \tag{2.23}$$

where $\alpha \in (0, 1/6)$, and the standard deviation of v_{k+1} (denoted by σ) is given by

$$\sigma \triangleq \left(\sum_{i=0}^{p} a_i^2 \right)^{1/2} \tag{2.24}$$

where $a_0 = 1$. The weight w_k measures the importance of the v_k, the input of the nonlinear block, for estimation of the value $f(v)$ at a fixed point v. In practice, σ is estimated by Eq. (2.22) and the step-size parameter α is chosen to be large for a fast convergence rate, such that the weight is actually replaced by

$$\hat{w}_k \triangleq \sqrt{2}\hat{\sigma}_k k^{\alpha} \exp\left\{-(\hat{v}_k - v)^2 k^{2\alpha} + \frac{\hat{v}_k^2}{2\hat{\sigma}_k^2}\right\} , \tag{2.25}$$

where $\hat{v}_k$ is defined by Eq. (2.21) and $\hat{\sigma}_k$ by Eq. (2.22) respectively. On ignoring the noise component in Eq. (2.2), we estimate the output x_{k+1} of the nonlinear block as

$$\hat{x}_{k+1} = y_{k+2} + \hat{b}_{1k} y_{k+1} + \cdots + \hat{b}_{qk} y_{k+2-q} , \tag{2.26}$$

since eventually the noise can be averaged out empirically because $Ee_{k+2} = 0$.

Inspired by the formula

$$\lim_{k\to\infty} Ew_k x_k = f(v)$$

and the fact that Ew_k^2 is proportional to k^{α} (as shown in Lemma 3.2 below), the value of $f(v)$ at any fixed v is calculated recursively as

$$\hat{\mu}_k(v) = \hat{\mu}_{k-1}(v) - \frac{1}{k}\hat{w}_k \left(\hat{\mu}_{k-1}(v) - \hat{x}_k\right) , \tag{2.27}$$

with the initial real value $\hat{\mu}_0(v) = 0$. The $\hat{w}_k$ involved here are defined by Eq. (2.25), and then $\hat{x}_k$ follows from Eq. (2.26).

In brief, the recursive procedure for identifying $f(\cdot)$ is defined by Eqs. (2.21),(2.22) and (2.25)-(2.27), on the basis of recursive estimations for the linear coefficients from Eqs. (2.18)-(2.20). As noted previously, the strong consistency of this recursive algorithm is established in Theorem 4.2 below.

Remark 2.1. An alternative recursive form replacing (2.27) for estimating $f(v)$ is

$$\hat{\mu}_k(v) = \hat{\mu}_{k-1}(v) - \frac{1}{k}\left(\hat{\mu}_{k-1}(v) - \hat{w}_k \hat{x}_k\right) , \tag{2.28}$$

which directly produces

$$\hat{\mu}_k(v) = \frac{1}{k}\sum_{i=1}^{k} \hat{w}_i \hat{x}_i .$$

The proof of convergence is simpler, and essentially contained in the proof of Theorem 4.2. However, the algorithm (2.28) generally performs worse than (2.27), perhaps because the kernel function w_i has a different weight at different i and Ew_k^2 is proportional to the k^{α} — cf. Lemma 3.2. The algorithm (2.27) also seems to more effectively eliminate the noise included from the estimate of the weight $\hat{w}_k$. For these reasons, we have preferred to use the algorithm (2.27) in this article.

3. Auxiliary Lemmas

We now prove some preliminaries to facilitate our subsequent theoretical analysis. The following lemma specifies conditions to guarantee the invertibility of $\Upsilon(p,q)$ defined by (2.10). It is of key importance to recover the two linear blocks, especially the last. The proof here is a simplification of counterparts in Refs. [22,23].

Lemma 3.1. *Under the conditions* H1-H3 *and for* $\gamma_0 \neq 0$ *defined by Eq.* (2.7)*,* $\Upsilon(p,q)$ *given by Eq.* (2.10) *is nonsingular.*

Proof. By standard matrix theory, it is sufficient to show that it is impossible to find q constants $h_1, h_2, \cdots, h_q$ such that

$$H(z)\gamma_k = \sum_{j=1}^{q} h_j \gamma_{k-(j-1)} = 0 \tag{3.1}$$

for $k = p, p+1, \cdots, p+q-1$, where $H(z) \triangleq \sum_{j=1}^{q} h_j z^{j-1}$. Obviously, the degree of H is $q-1$.

Let us rewrite the formula (2.16) connecting the cross-correlation coefficients $\{\gamma_i\}$ and polynomials $A(z)$ and $B(z)$:

$$\Gamma(z) = \frac{\gamma_0 A(z)}{B(z)} . \tag{3.2}$$

From the stability of $B(z)$, $\Gamma(z)$ is analytic in the disk $\{z : |z| < r\}$ for some $r > 1$. Since $A(z)$ and $B(z)$ are co-prime and $\gamma_0 \neq 0$, $\Gamma(z)$ has just q (including multiple) nonzero poles. Then given Eq. (2.9), it is easy to prove that Eq. (3.1) holds for any $k \geq p$. Multiplying Eq. (3.2) by $H(z)$ and noticing $\gamma_j = 0$ for $j < 0$, from Eqs. (2.14) and (3.1) we have

$$\Gamma(z)H(z) = \sum_{k=-\infty}^{\infty} H(z)\gamma_k z^k = \sum_{k=0}^{\infty} H(z)\gamma_k z^k = \sum_{k=0}^{p-1} H(z)\gamma_k z^k . \tag{3.3}$$

Clearly, the right-hand side of Eq. (3.3) is a polynomial and thus analytic in the whole complex plane, while the left-hand side has at least one pole in the region $\{z : |z| \geq r\}$ since the order of $H(z)$ is $q-1$ while that of $B(z)$ is q. This contradiction leads to the assertion. $\square$

The following lemma gives some conclusions for the first and second moments involving the weight w_k.

Lemma 3.2. *Under the conditions* H1-H3*, for* w_k *defined by Eq.* (2.23) *we have the following results:*

$$Ew_k = 1 , \qquad E[k^{-\alpha/2} w_k]^2 = \sigma \exp\left\{\frac{v^2}{2\sigma^2}\right\} ,$$

$$\lim_{k\to\infty} E[w_k x_k] = f(v) , \quad \lim_{k\to\infty} E[k^{-\alpha/2} w_k x_k]^2 = \sigma f^2(v) \exp\left\{\frac{v^2}{2\sigma^2}\right\} .$$

Proof. Let us prove the last two conclusions only, as the other two may be proven in a similar manner. Thus we see that

$$
\begin{aligned}
&\lim_{k\to\infty} E[w_k x_k] \\
&= \lim_{k\to\infty} \sqrt{2}\sigma k^{\alpha} \int_R \exp\left(-(x-v)^2 k^{2\alpha} + \frac{x^2}{2\sigma^2}\right) f(x) \frac{e^{-\frac{x^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}\, dx \\
&= \lim_{k\to\infty} \frac{1}{\sqrt{\pi}} \int_R \exp[-t^2 f(v+k^{-\alpha}t)]\, dt \\
&= f(v)\,;
\end{aligned}
$$

and similarly,

$$
\begin{aligned}
&\lim_{k\to\infty} E[k^{-\alpha/2} w_k x_k]^2 \\
&= \lim_{k\to\infty} 2\sigma^2 k^{\alpha} \int_R \exp\left(-2(x-v)^2 k^{2\alpha} + \frac{x^2}{\sigma^2}\right) f^2(x) \frac{e^{-\frac{x^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}\, dx \\
&= \lim_{k\to\infty} \sqrt{\frac{2}{\pi}}\sigma \int_R e^{-2t^2} f^2(v+k^{-\alpha}t) \exp\left(\frac{(v+t/k^{\alpha})^2}{2\sigma^2}\right) dt \\
&= \sigma f^2(v) \exp\left(\frac{v^2}{2\sigma^2}\right) . \qquad \square
\end{aligned}
$$

The following lemma is fundamental for the derivation of Eq. (2.8), which is used to recover the first linear block. (Similar results can be found in Refs. [5,7–9].)

Lemma 3.3. *Under the conditions* H1-H3, *we have that*

$$
E[u_{k-j} x_{k+1}] = \gamma_0 a_j\,, \quad j = 0, 1, \cdots, p\,, \tag{3.4}
$$

where γ_0 is given by Eq. (2.7).

Proof. By noticing $v_{k+1} = \sum_{j=0}^{p} a_j u_{k-j}$, we see that v_{k+1} and u_j are jointly Gaussian. Consequently, from

$$
E\left[\left(u_{k-j} - \frac{a_j}{\sigma^2} v_{k+1}\right) v_{k+1}\right] = 0
$$

it follows that $u_{k-j} - (a_j/\sigma^2) v_{k+1}$ and v_{k+1} are independent. Thus we have

$$
E\left[u_{k-j} | v_{k+1}\right] = \frac{a_j}{\sigma^2} v_{k+1}\,,
$$

and hence

$$
\begin{aligned}
E\left[u_{k-j} f(v_{k+1})\right] &= E\left(E\left[u_{k-j} f(v_{k+1}) | v_{k+1}\right]\right) \\
&= E\left(f(v_{k+1}) E\left[u_{k-j} | v_{k+1}\right]\right) \\
&= \gamma_0 a_j\,, \quad j = 0, 1, \cdots, p\,. \qquad \square
\end{aligned}
$$

Remark 3.1. Based on Lemma 3.3, let us analyze γ_0 in more detail to get some sense of the conditions to derive $\gamma_0 \neq 0$. From Eq. (2.7), clearly

$$\begin{aligned}\gamma_0 &= E[u_k x_{k+1}] = E[u_k f(v_{k+1})] \\ &= E[v_{k+1} f(v_{k+1})] - a_1 E[u_{k-1} f(v_{k+1})] - \cdots - a_p E[u_{k-p} f(v_{k+1})] ,\end{aligned}$$

on substituting $u_k = v_{k+1} - a_1 u_{k-1} - \cdots - a_p u_{k-p}$ in the third step. Recalling the fact that $v_{k+1} \sim \mathcal{N}(0, \sigma^2)$ and Lemma 3.3, we obtain

$$\gamma_0 = \frac{1}{\sqrt{2\pi}\sigma^3} \int_R t f(t) e^{-\frac{t^2}{2\sigma^2}} \, dt , \tag{3.5}$$

where σ is given by Eq. (2.24). The existence of γ_0 is guaranteed by condition H1, since the growth rate of $f(\cdot)$ is at most polynomial. A typical case satisfying the restriction $\gamma_0 \neq 0$ is where $f(\cdot)$ is an continuous odd function with a nonzero value at a certain point.

Let us now state a direct corollary of Theorem 2.2 in Ref. [24] (also a slight extension of Corollary 2.2.1), to establish some strong laws of large numbers.

Proposition 3.1. *Let $\{X_k\}$ be a sequence of random variables satisfying $EX_k = 0$ and*

$$E[X_i X_j] \le \rho_{j-i} \sqrt{EX_i^2 EX_j^2}$$

for $i \le j$, where $\rho_k \ge 0$ for $\forall k$. If

$$\sum_{k=1}^{\infty} \rho_k < \infty \quad \text{and} \quad \sum_{k=1}^{\infty} \ln^2 k \cdot EX_k^2 < \infty ,$$

then $\sum_{k=1}^{\infty} X_k$ converges almost surely.

Lemma 3.4. *Under the conditions* H1-H3*, the following strong laws of large numbers hold:*

$$\frac{1}{n^{1-\delta}} \sum_{k=1}^{n} (u_{k-j} y_{k+2} - E[u_{k-j} y_{k+2}]) \xrightarrow[n\to\infty]{a.s.} 0 , \tag{3.6}$$

$$\frac{1}{n} \sum_{k=1}^{n} (w_k x_k - E[w_k x_k]) \xrightarrow[n\to\infty]{a.s.} 0 , \tag{3.7}$$

$$\frac{1}{n} \sum_{k=1}^{n} w_k e_k \xrightarrow[n\to\infty]{a.s.} 0 , \tag{3.8}$$

where $\delta \in (0, 1/2)$ and $j = 0, 1, \cdots, p+q$ in Eq. (3.6) .

Proof. We first prove Eq. (3.6) from Proposition 3.1. From Eq. (2.5),

$$\begin{aligned} y_{k+2} &= B^{-1}(z) x_{k+1} + B^{-1}(z) e_{k+2} \\ &= \sum_{i=0}^{k+1} \bar{b}_i x_{k+1-i} + \sum_{i=0}^{k+2} \bar{b}_i e_{k+2-i} . \end{aligned} \tag{3.9}$$

Denoting $X_k = (u_{k-j}y_{k+2} - E[u_{k-j}y_{k+2}])/k^{1-\delta}$, from Eq. (3.9) and H3 we have

$$\begin{aligned}
&[k(k+l)]^{1-\delta}E[X_kX_{k+l}] \\
&= E[u_{k-j}y_{k+2}u_{k+l-j}y_{k+l+2}] - E[u_{k-j}y_{k+2}]\cdot E[u_{k+l-j}y_{k+l+2}] \\
&= \sum_{s=0}^{k+1}\sum_{t=0}^{k+l+1}\bar{b}_s\bar{b}_tE[u_{k-j}u_{k+l-j}(x_{k+1-s}x_{k+l+1-t} + x_{k+1-s}e_{k+l+2-t} \\
&\quad + x_{k+l+1-t}e_{k+2-s} + e_{k+2-s}e_{k+l+2-t})] - \left(\sum_{i=0}^{j}\bar{b}_iE[u_{k-j}x_{k+1-i}]\right)^2 \\
&= \sum_{s=0}^{k+1}\sum_{t=0}^{k+l+1}\bar{b}_s\bar{b}_tE[u_{k-j}u_{k+l-j}x_{k+1-s}x_{k+l+1-t}] - \left(\sum_{i=0}^{j}\bar{b}_iE[u_{k-j}x_{k+1-i}]\right)^2 \\
&= \left(\sum_{i=0}^{j}\bar{b}_iE[u_{k-j}x_{k+1-i}]\right)^2 - \left(\sum_{i=0}^{j}\bar{b}_iE[u_{k-j}x_{k+1-i}]\right)^2 \\
&= 0
\end{aligned}$$

for $l \geq 2p+q+1$, on noticing that $\{x_k\}$ is a $p+1$-dependent sequence and $x_{k+1} \in \sigma\{u_k, u_{k-1}, \cdots, u_{k-p}\}$. Hence all of the conditions of Proposition 3.1 are satisfied trivially — i.e. $\rho_k = 0$ for $k \geq 2p+q+1$ and

$$\ln^2 k \cdot EX_k^2 = O\left(\frac{\ln^2 k}{k^{2(1-\delta)}}\right) = O\left(\frac{1}{k^{\delta_1}}\right)$$

with $\delta_1 \in (1, 2(1-\delta))$. Thus from Proposition 3.1 and Kronecker's lemma, Eq. (3.6) follows.

We consequently prove Eq. (3.7). Thus on noticing that the first linear subsystem is a moving averaging form with finite order $p+1$, clearly $\{w_kx_k,\ k \in A_j\}$ is an i.i.d. sequence, where $A_j = \{j + i(p+1):\ i = 0, 1, \cdots\}$. Then from Lemma 3.2,

$$\begin{aligned}
&\sum_{k=1}^{\infty}\frac{1}{k^{1-\alpha}}[k^{-\alpha}w_kx_k - k^{-\alpha}E(w_kx_k)] \\
=&\sum_{j=0}^{p}\sum_{k\in A_j}\frac{1}{k^{1-\alpha}}[k^{-\alpha}w_kx_k - k^{-\alpha}E(w_kx_k)]
\end{aligned}$$

converges almost surely. From Kronecker's lemma, the desired assertion therefore follows directly. The proof for Eq. (3.8) is similar to that for Eq. (3.7). □

Finally, we introduce a basic fact for later reference.

Lemma 3.5. *If a sequence of random variables $\{\xi_k\}$ satisfies $\sup E|\xi_k|^r < \infty$ for any $r > 0$, then for any $\delta > 0$ we have the following limit:*

$$\frac{\xi_k}{k^{\delta}} \xrightarrow[k\to\infty]{a.s.} 0\,.$$

Proof. Since

$$P\left[\frac{|\xi_k|}{k^\delta} > \varepsilon\right] = P\left[\frac{|\xi_k|^{2/\delta}}{k^2} > \varepsilon^{2/\delta}\right] < \frac{1}{\varepsilon^{2/\delta}k^2}E|\xi_k|^{2/\delta}$$

for any given $\varepsilon > 0$, it follows that

$$\sum_{k=1}^{\infty} P\left[\frac{|\xi_k|}{k^\delta} > \varepsilon\right] < \infty.$$

From the Borel-Cantelli lemma, we therefore have $\xi_k/k^\delta \xrightarrow[k\to\infty]{a.s.} 0$. □

4. Main Results

We now proceed to establish the strong consistency of the recursive estimates given by Eqs. (2.18)-(2.20) for the linear coefficients, and Eqs. (2.21),(2.22) and (2.25)-(2.27) for nonlinear $f(\cdot)$ under reasonable conditions.

The convergence of algorithms (2.18)-(2.20) for linear coefficients is assured by the following theorem. Hereafter the abbreviation “a.s.”' means “almost surely convergent" — i.e. we have convergence with probability 1.

Theorem 4.1. *Assume conditions* H1-H3 *hold for the W-H system given by Eqs.* (2.1), (2.2) and (2.4). *Then*

$$\hat{\gamma}_i(k) \xrightarrow[k\to\infty]{a.s.} \gamma_i, \quad i = 0, 1, \cdots, p+q, \tag{4.1}$$

where $\hat{\gamma}_i(k)$ is generated by algorithm (2.18) *and with the convergence rate:*

$$|\hat{\gamma}_i(k) - \gamma_i| = o\left(k^{-\delta}\right) \ a.s. \tag{4.2}$$

for $\forall\delta \in (0, 1/2)$, $i = 0, 1, \cdots, p+q$. If further $\gamma_0 = E[u_k y_{k+2}] \neq 0$, where γ_0 is given by Eq. (3.5)*, then the estimates for a_i, $i = 1, \cdots, p$ and b_j, $j = 1, \cdots, q$ from Eqs.* (2.19) and (2.20) *converge to the true values almost surely and have the same convergence rate as above.*

Proof. As already mentioned, the recursive algorithm (2.18) until step n is equivalent to

$$\hat{\gamma}_i(n) = \frac{1}{n}\sum_{k=1}^{n} u_{k-i} y_{k+2}\,. \tag{4.3}$$

It is sufficient to show that

$$\frac{1}{n^{1-\delta}}\sum_{k=1}^{n}[u_{k-i}y_{k+2} - E(u_{k-i}y_{k+2})] \tag{4.4}$$

converges to zero almost surely for $\delta \in (0,1/2)$, which is just (3.6) in Lemma 3.4. Recalling $\gamma_i = E(u_{k-i}y_{k+2})$, the assertions for γ_i follow directly. Consequently, from Eq. (4.2) we have

$$\|\hat{\Upsilon}_k - \Upsilon\| = o(k^{-\delta}),$$

where $\delta \in (0,1/2)$ and $\|\cdot\|$ denotes the matrix Euclidean norm, which immediately implies the subsequent assertions regarding a_i and b_j. $\square$

Finally, the recursive algorithms given by Eqs. (2.21), (2.22) and (2.25)-(2.27) to recover the value $f(v)$ at a fixed v are analysed in the following theorem.

Theorem 4.2. *Assume the conditions* H1-H3 *hold for the W-H system given by Eqs.* (2.1), (2.2) and (2.4) *with* $\gamma_0 \neq 0$. *Then*

$$\hat{\mu}_k(v) \xrightarrow[k\to\infty]{a.s.} f(v), \tag{4.5}$$

where $\hat{\mu}_k(v)$ *is given by Eq.* (2.27).

Proof. For convenience, we divide the proof into five steps. Roughly speaking, the steps 1), 2) and 4) are preliminaries, and the steps 3) and 5) are the mainstream.

1). We first specify the estimation errors between the kernel function w_k and its estimation $\hat{w}_k$, and between x_k and $\hat{x}_k$ at step k, respectively. From Theorem 4.1 we have $|\hat{v}_k - v_k| = o\left(k^{-\delta}\right)$, where and hereafter a possible subset with zero probability in the whole sample path set Ω is ignored and $\delta \in (0,1/2)$. Since the derivative of w_k with respective to v_k is $\mathrm{O}(k^{3\alpha})$ as $k \to \infty$, the standard mean value theorem implies

$$|\hat{w}_k - w_k| = o\left(k^{-\delta_1}\right),$$

where $\delta_1 = \delta - 3\alpha > 0$ by suitable selection of α. This is the reason why we require $\alpha \in (0,1/6)$ in Section 2. Thus,

$$\hat{w}_k = w_k + o\left(k^{-\delta_1}\right). \tag{4.6}$$

Similarly, by Theorem 4.1, for reconstruction of intermediate signal x_k from Eq. (2.26) we have

$$|\hat{x}_k - (x_k + e_{k+1})| = o\left(k^{-\delta}\right),$$

so that

$$\hat{x}_k = x_k + e_{k+1} + o\left(k^{-\delta}\right). \tag{4.7}$$

2). We develop a strong law of large numbers for $\{\hat{w}_k\hat{x}_k\}$ in this step. From Eqs. (4.6) and (4.7),

$$\begin{aligned}\hat{w}_k\hat{x}_k &= \left[w_k + o\left(k^{-\delta_1}\right)\right]\cdot\left[x_k + e_{k+1} + o\left(k^{-\delta}\right)\right]\\ &= w_k(x_k+e_{k+1}) + o\left(w_k k^{-\delta} + (x_k+e_{k+1})k^{-\delta_1} + k^{-\delta_1-\delta}\right).\end{aligned} \tag{4.8}$$

From Lemma 3.2, and similar to the proof for Eq. (3.7), we have

$$\sum_{k=1}^{\infty}\frac{1}{k}w_k k^{-\delta}=\sum_{k=1}^{\infty}\frac{w_k}{k^{1+\delta}}=\sum_{k=1}^{\infty}\frac{w_k-Ew_k}{k^{1+\delta}}+\sum_{k=1}^{\infty}\frac{Ew_k}{k^{1+\delta}}<\infty\,.$$

Thus by Kronecker's lemma,

$$\frac{1}{n}\sum_{k=1}^{n}w_k k^{-\delta}\xrightarrow[n\to\infty]{}0\,.$$

From Lemmas 3.2, 3.4 and 3.5, all other terms in Eq. (4.8) can be analysed similarly. A strong law of large numbers therefore holds — viz.

$$\frac{1}{n}\sum_{k=1}^{n}[\hat{w}_k\hat{x}_k-f(v)]\xrightarrow[n\to\infty]{}0\,,$$

which is equivalent to

$$\frac{1}{n}\sum_{k=1}^{n}\hat{w}_k\hat{x}_k=f(v)+o(1). \tag{4.9}$$

Similarly,

$$\frac{1}{n}\sum_{k=1}^{n}|\hat{w}_k\hat{x}_k|=|f(v)|+o(1)\,. \tag{4.10}$$

3). We now analyse the asymptotic performance of the recursive formula Eq. (4.5) by introducing a subset — i.e. $\Omega(n_0)$ given in Eq. (4.13), for the set of whole sample paths Ω. We rewrite Eq. (2.27) as

$$\mu_n(v)=\Phi_{n,1}\mu_0(v)+\sum_{k=1}^{n}\Phi_{n,k+1}\frac{1}{k}\hat{w}_k\hat{x}_k\,, \tag{4.11}$$

where

$$\Phi_{i,j}\triangleq\left(1-\frac{\hat{w}_i}{i}\right)\left(1-\frac{\hat{w}_{i-1}}{i-1}\right)\cdots\left(1-\frac{\hat{w}_j}{j}\right),\quad i\geq j,\quad \Phi_{j\,,j+1}=1. \tag{4.12}$$

For convenience, we define a subset of Ω as

$$\Omega(n_0)\triangleq\left\{\omega\in\Omega:\ \left|\frac{\hat{w}_k(\omega)}{k}\right|<\frac{1}{2}\,,\ k=n_0,\,n_0+1\,,\cdots\right\} \tag{4.13}$$

and first point out that

$$P[\Omega(n_0)]\to 1\,,\quad n_0\to\infty\,. \tag{4.14}$$

Actually, from Chebyshev's inequality, Eq. (4.6) and Lemma 3.2 we have

$$1 - P[\Omega(n_0)] \le \sum_{k=n_0}^{\infty} P\left[\left|\frac{\hat{w}_k(\omega)}{k}\right| \ge \frac{1}{2}\right] \le \sum_{k=n_0}^{\infty} \frac{4E\hat{w}_k^2}{k^2} \xrightarrow[n_0\to\infty]{} 0\,.$$

4). To prepare for the final step, we analyse the order of $\Phi_{n,k}$ defined by Eq. (4.12), for $n > k \ge n_0$ as n tends to infinity at a fixed sample path in $\Omega(n_0)$. We specify the distance between $\Phi_{n,k+1}$ and k/n, as described in Eq. (4.19) below. From the definition (4.12),

$$\Phi_{n,k+1} = \exp\left\{\sum_{s=k+1}^{n} \log(1 - \frac{\hat{w}_s}{s})\right\}\,,$$

hence from the inequalities $\frac{x}{1+x} \le \log(1+x) \le x$ for $x > -1$ we obtain

$$\exp\left\{-\sum_{s=k+1}^{n} \frac{\hat{w}_s}{s-\hat{w}_s}\right\} \le \Phi_{n,k+1} \le \exp\left\{-\sum_{s=k+1}^{n} \frac{\hat{w}_s}{s}\right\}\,. \tag{4.15}$$

Meanwhile, from Eq. (4.6) and Lemma 3.2 we have

$$\begin{aligned}\sum_{s=k+1}^{n} \frac{\hat{w}_s}{s} &= \sum_{s=k+1}^{n}\left[\frac{w_s}{s} + o\left(\frac{1}{s^{1+\delta_1}}\right)\right]\\ &= \sum_{s=k+1}^{n} \frac{w_s - Ew_s}{s} + \sum_{s=k+1}^{n} \frac{Ew_s}{s} + o\left(\sum_{s=k+1}^{n} \frac{1}{s^{1+\delta_1}}\right)\\ &= k^{-\beta}\sum_{s=k+1}^{n} \frac{k^{\beta}(w_s - Ew_s)}{s} + \sum_{s=k+1}^{n} \frac{1}{s} + o\left(\frac{1}{k^{\delta_1}}\right)\\ &= o\left(\frac{1}{k^{\beta}}\right) + \sum_{s=k+1}^{n} \frac{1}{s} + o\left(\frac{1}{k^{\delta_1}}\right) = \log\frac{n}{k} + o\left(\frac{1}{k^{\delta_1}}\right),\end{aligned} \tag{4.16}$$

where $0 < \delta_1 < \beta < (1/2) - (\alpha/2)$. Further,

$$\begin{aligned}\left|\sum_{s=k+1}^{n} \frac{\hat{w}_s}{s-\hat{w}_s} - \sum_{s=k+1}^{n} \frac{\hat{w}_s}{s}\right| &\le \sum_{s=k+1}^{n}\left|\frac{\hat{w}_s}{s}\right|^2 \cdot \left|\frac{1}{1-\frac{\hat{w}_s}{s}}\right| \le 2\sum_{s=k+1}^{n}\left|\frac{\hat{w}_s}{s}\right|^2\\ &= 2\sum_{s=k+1}^{n}\left[\frac{\hat{w}_s^2 - w_s^2}{s^2} + \frac{w_s^2 - Ew_s^2}{s^2} + \frac{Ew_s^2}{s^2}\right]\\ &= O\left(k^{-1}\right)\,,\end{aligned}$$

and consequently

$$\begin{aligned}-\sum_{s=k+1}^{n} \frac{\hat{w}_s}{s-\hat{w}_s} &\ge -\sum_{s=k+1}^{n} \frac{\hat{w}_s}{s} - \left|\sum_{s=k+1}^{n} \frac{\hat{w}_s}{s-\hat{w}_s} - \sum_{s=k+1}^{n} \frac{\hat{w}_s}{s}\right|\\ &\ge -\sum_{s=k+1}^{n} \frac{\hat{w}_s}{s} - O\left(k^{-1}\right)\,.\end{aligned} \tag{4.17}$$

From Eqs. (4.16) and (4.17), Eq. (4.15) turns to be

$$\frac{k}{n}\exp\left(o(k^{-\delta_1})-O\left(k^{-1}\right)\right)\le\Phi_{n,k+1}\le\frac{k}{n}\exp(o(k^{-\delta_1}))\,. \tag{4.18}$$

Let us assume that n_0 is selected big enough such that $n_0^{-\delta_1}<1$. From the assumption $n>k\ge n_0$ at the beginning of this step, and the inequality $1+x<e^x<1+e|x|$ for $|x|<1$, Eq. (4.18) becomes

$$\frac{k}{n}\left(1+o(k^{-\delta_1})-O\left(k^{-1}\right)\right)\le\Phi_{n,k+1}\le\frac{k}{n}(1+o(k^{-\delta_1}))\,. \tag{4.19}$$

5). Now we are ready to establish the convergence result Eq. (4.5), by first considering the convergence issue at a fixed sample path in $\Omega(n_0)$, and then deducing the desired formula Eq. (4.5). From Eqs. (4.19) and (4.10),

$$\sum_{k=n_0}^{n}\left|\Phi_{n,k+1}-\frac{k}{n}\right|\frac{\left|\hat{w}_k\hat{x}_k\right|}{k}\le\sum_{k=n_0}^{n}\left[O\left(k^{-1}\right)+o(k^{-\delta_1})\right]\frac{\left|\hat{w}_k\hat{x}_k\right|}{n}=o(n_0^{-\delta_1})\,,$$

so

$$\begin{aligned}\sum_{k=n_0}^{n}\Phi_{n,k+1}\frac{1}{k}\hat{w}_k\hat{x}_k&=\sum_{k=n_0}^{n}\frac{1}{n}\hat{w}_k\hat{x}_k+\sum_{k=n_0}^{n}(\Phi_{n,k+1}-\frac{k}{n})\frac{1}{k}\hat{w}_k\hat{x}_k\\&=\sum_{k=n_0}^{n}\frac{1}{n}\hat{w}_k\hat{x}_k+O\left(\sum_{k=n_0}^{n}\left|\Phi_{n,k+1}-\frac{k}{n}\right|\frac{\left|\hat{w}_k\hat{x}_k\right|}{k}\right)\\&=f(v)+o(n_0^{-\delta_1})\xrightarrow[n_0\to\infty]{}f(v)\,,\end{aligned} \tag{4.20}$$

and hence from Eqs. (4.19) and (4.20) we have

$$\begin{aligned}\hat{\mu}_n(v)&=\Phi_{n,n_0+1}\hat{\mu}_{n_0}(v)+\sum_{k=n_0+2}^{n}\Phi_{n,k+1}\frac{1}{k}\hat{w}_k\hat{x}_k\\&=\frac{n_0}{n}\left(1+o(n_0^{-\delta_1})\right)\hat{\mu}_{n_0}(v)+f(v)+o(n_0^{-\delta_1})\\&\xrightarrow[n\to\infty]{}f(v)+o(n_0^{-\delta_1})\xrightarrow[n_0\to\infty]{}f(v)\,.\end{aligned} \tag{4.21}$$

Now the sample path set $\Omega(n_0)$ tends to include nearly all sample paths when $n_0\to\infty$ since its probability tends to 1, so Eq. (4.21) holds with probability 1 — i.e.

$$\hat{\mu}_n(v)\xrightarrow[n\to\infty]{a.s.}f(v)\,,$$

which completes the proof. □

Theorems 4.1 and 4.2 declare the strong consistency of the recursive identification algorithms — viz. Eqs. (2.18)-(2.20) for linear coefficients and (2.21)(2.22)(2.25)-(2.27) for the nonlinear block, as proposed in Section 2.3. (Strong consistency means asymptotic convergence to the true values with probability 1 as the data size tends to infinity.) The restriction $\gamma_0\neq 0$ appears in both theorems, as we mentioned in Remark 3.1. A typical case is that $f(\cdot)$ is a continuous odd function, satisfying $f(v_0)\neq 0$ for certain v_0.

5. Numerical Example

We now illustrate our results with a numerical example, where the first and second linear subsystems of a W-H system are

$$v_{k+1} = u_k + 0.27u_{k-1} - 0.5u_{k-2} + 0.79u_{k-3}$$
$$y_{k+2} - 0.3y_{k+1} + 0.6y_k = x_{k+1} + e_{k+2} \,.$$

Clearly, the parameters are

$$a_1 = 0.27\,, a_2 = -0.5\,, a_3 = 0.79\,,\ b_1 = -0.3\,, b_2 = 0.6\,.$$

Let the nonlinear static block be

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

and the observation noise be Gaussian $e_k \in \mathcal{N}(0, 0.1^2)$. Clearly, all conditions in our Theorems 4.1 and 4.2 are satisfied. The signal to noise ratio (SNR) between input and noise is 100 (or 20dB) – and although the SNR is large, convincing performances of the estimate algorithms will also hold for lower SNR provided the data size is large enough. We use Eqs. (2.18)-(2.20) and (2.27) to estimate a_i, $i = 1, \cdots, p$; b_j , $j = 1, \cdots, q$ and $f(v)$ respectively.

The estimates for $a_i, i = 1, \cdots, p$ and $b_j, j = 1, \cdots, q$ appear in Fig. 2, where the dotted lines denote the true values and the solid lines the estimates of the coefficients. In Fig. 3 it is shown how the true $f(v)$ denoted by the dotted line is approximated by its estimate

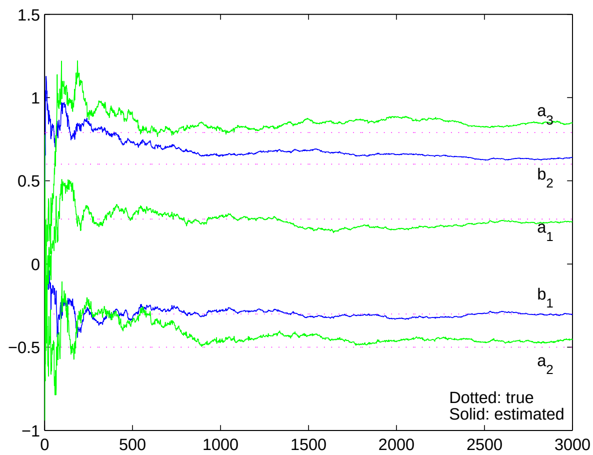

Figure 2: Estimates for linear coefficients as $k = 1, \cdots, 3000$.

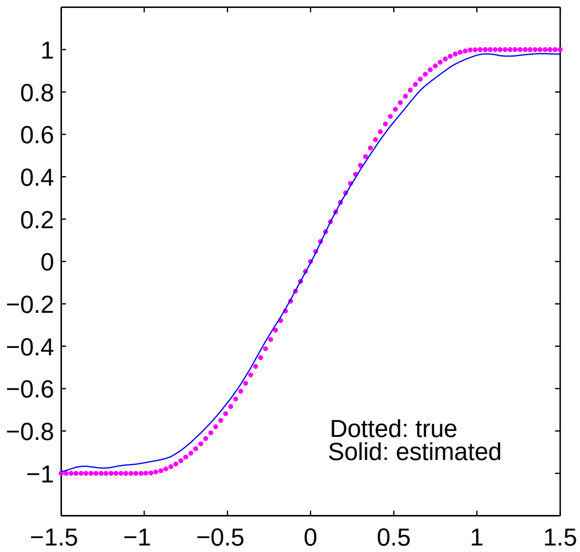

Figure 3: Estimates of $f(\cdot)$ at $k=3000$.

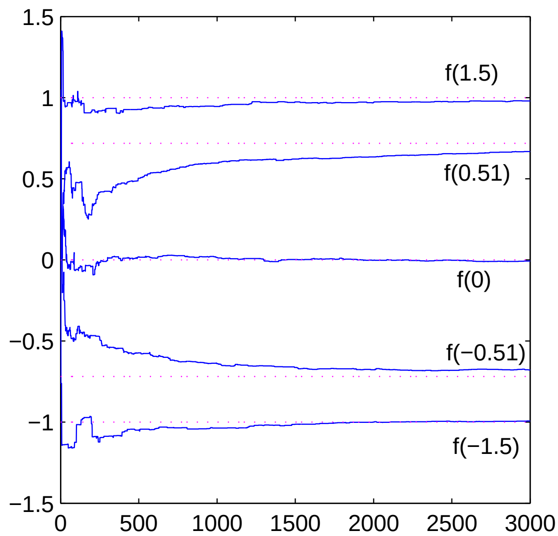

Figure 4: Estimates of $f(v)$ as $k=1,\cdots,3000$ for $v=-1.5,\ -0.51,\ 0,\ 0.51,\ 1.5$.

denoted by the solid line. To derive the estimate curve (solid line) for $f(v)$, the interval $[-1.5, 1.5]$ on which $f(v)$ is defined is divided into 100 equal subintervals, and at the endpoints v of subintervals $f(v)$ is estimated by Eqs. (2.21),(2.22) and (2.25)-(2.27). The solid line is the result from the estimates given at $k=3000$.

In Fig. 2, the estimates of the linear parts evidently tend to the true values, which justifies the theoretical assertion in Theorem 4.1. The nonparametric point-wise estimates of $f(v)$ at $k=3000$ in Fig. 3, and the tendencies of the estimates for $v=-1.5$, -0.51, 0, 0.51, 1.5 as $k=1,\cdots,3000$ in Fig. 4, are consistent with the assertion in Theorem 4.2.

6. Concluding Remarks

A recursive identification scheme for SISO Wiener-Hammerstein systems with non-parametrised nonlinearity is discussed in this article, and the strong consistency of the estimation algorithms for the linear blocks and nonlinearity is established in Theorems 4.1 and 4.2, respectively. Technically, these algorithms relate to those in Refs. [7–9,14], but are simpler by avoiding the socalled stochastic approximation (SA) with expanding truncations and the theoretical analysis is more straightforward. A numerical experiment illustrates our theoretical results. The first linear block we considered is a Moving Averaging (MA) type and the second one an Auto-Regressive (AR) type. More general Auto-Regressive and Moving Average (ARMA) types for both subsystems could be considered, where similar algorithms and strong convergence results may be sought.

References

[1] Y. Zhu, *Distillation column identification for control using Wiener model*, Proceedings of the American Control Conference, vol. 5, pp. 3462–3466, San Diego, California, June 1999.

[2] A. Kalafatis, L. Wang, and W.R. Cluett, *Identification of time-varying pH processes using sinusoidal signals*, Automatica **41**, 685–691 (2005).

[3] R.C. Emerson, M.J. Korenberg and M.C. Citron, *Identification of complex-cell intensive nonlinearities in a cascade model of cat visual cortex*, Biol. Cybern. **66**, 291–300 (1992).

[4] M. Szyper and A. Bien, *Application of Wiener-Hammerstein models for modeling of light flickering severity-meter*, Syst. Anal. Model. Sim. **40**, 245–255 (2001).

[5] W. Greblicki, *Nonparametric approach to Wiener system identification*, IEEE T Circuits-I: Fundamental Theory and Applications **44**, 538–545 (1997).

[6] W. Greblicki and M. Pawlak, *Nonparametric System Identification*, Cambridge University Press, 2008.

[7] X.L. Hu and H.F. Chen, *Strong consistency of recursive identification for Wiener systems*, Automatica **41**, 1905–1916 (2005).

[8] X.L. Hu and H.F. Chen,*Identification for Wiener Systems with RTF subsystems*, Eur. J. Control **6**, 581–594 (2006).

[9] X.L. Hu and H.F. Chen, *Recursive identification for Wiener systems using Gaussian signals*, Asian J. Control **10**, 341–350 (2008).

[10] A. Kalafatis, L. Wang and W.R. Cluett, *Identification of Wiener-type nonlinear systems in a noisy environment*, Int. J. Control **66**, 923–941 (1997).

[11] M. Verhaegen and D. Westwick, *Identifying MIMO Wiener systems in the context of subspace model identification methods*, Int. J. Control **63**, 331–349 (1996).

[12] Y.L. Zhao, L.Y. Wang, G.G. Yin and J.F. Zhang, *Identification of Wiener systems with binary-valued output observations*, Automatica **43**, 1752–1765 (2007).

[13] E.W. Bai, *Frequency domain identification of Hammerstein models*, IEEE Trans. Automat. Control **48**, 530–542 (2003).

[14] H.F. Chen, *Pathwise convergence of recursive identification algorithms for Hammerstein systems*, IEEE Trans. Automat. Control **49**, 1641–1649 (2004).

[15] W. Greblicki, *Stochastic approximation in nonparametric identification of Hammerstein systems*, IEEE Trans. Automat. Control **47**, 1800–1810 (2002).

[16] J. Vörös, *Recursive identification of Hammerstein systems with discontinuous nonlinearities containing dead-zones*, IEEE Trans. Automat. Control **48**, 2203–2206 (2003).

[17] J. Schoukens, R. Pintelon and M. Enqvist, *Study of the LTI relations between the outputs of two coupled Wiener systems and its application to the generation of initial estimates for Wiener-Hammerstein systems*, Automatica **44**, 1654–1665 (2008).

[18] A.H. Tan, *Wiener-Hammerstein modeling of nonlinear effects in bilinear systems*, IEEE Trans. Automat. Control **51**, 648–652 (2006).

[19] A.H. Tan and K.R. Godfrey, *Identification of Wiener-Hammerstein models using linear interpolation in the frequency domain (LIFRED)*, IEEE T. Instrum. Meas. **51**, 509–521 (2002).

[20] N.J. Bershad, P. Celka, and S. Mclaughlin, *Analysis of stochastic gradient identification of Wiener-Hammerstein systems for nonlinearities with Hermite polynomial expansions*, IEEE T. Signal Process. **49**, 1060–1072 (2001).

[21] J. Vörös, *An iterative method for Wiener-Hammerstein systems parameter identification*, J. Elec. Eng. **58**, 114–117 (2007).

[22] W.X. Zhao and H.F. Chen, *Recursive identification for Hammerstein system with ARX subsystem*, IEEE T. Automat. Control **51**, 1966–1974 2006.

[23] A.M. Walker, *Large-sample estimation of parameters for autoregressive process with moving-average residuals*, Biometrika **49**, 117–131 (1962).

[24] R.J. Serfling, *Convergence properties of S_n under moment restrictions*, Ann. Math. Stat. **41**, 1235–1248 (1970).