[31] S. Ohno, Y. Inouye, and K. Tanebe, "A numerical algorithm for implementing the single-stage maximization criterion for multichannel blind deconvolution," in *Proc. EUSIPCO-98*, Rhodes, Greece, 1998, pp. 1649–1652.
[32] S. Ohno and Y. Inouye, "A least-squares interpretation of the single-stage maximization criterion for multichannel blind deconvolution," in *Proc. ICASSP-98*, Seattle, WA, 1998, pp. 2101–2104.
[33] H. Luo and R. Liu, "Second-order statistics blind MIMO-FIR channel equalization for antenna array-based wireless communications systems," in *Proc. IEEE Sensor Array and Multichannel (SAM) Signal Processing Worksh.*, Cambridge, MA, Mar. 2000.
[34] H. H. Zeng, L. Tong, and C. R. Johnson, Jr., "An analysis of constant modulus receivers," *IEEE Trans. Signal Processing*, vol. 47, pp. 2990–2999, Nov. 1999.
[35] L. Tong, M. Gu, and S. Y. Kung, "A geometrical approach to blind singal estimation," in *Signal Processing Advances in Wireless & Mobile Communications*, G. Giannakis, P. Stoica, Y. Hua, and L. Tong, Eds. Upper Saddle River, NJ: Prentice-Hall, 2001, vol. 1, ch. 8, pp. 255–290.
[36] J. K. Tugnait, "Identification and deconvolution of multichannel linear non-Gaussian processes using higher-order statistics and inverse filter criteria," *IEEE Trans. Signal Processing*, vol. 45, pp. 658–672, Mar. 1997.
[37] ——, "Channel estimatin and equalization using higher-order statistics," in *Signal Processing Advances in Wireless & Mobile Communications*, G. B. Giannikis, P. Stoica, Y. Hua, and L. Tong, Eds. Upper Saddle River, NJ: Prentice-Hall, 2001, vol. 1, ch. 1, pp. 1–40.
[38] ——, "On linear predictors for MIMO channels and related blind identification and equalization," *IEEE Signal Processing Lett.*, vol. 5, no. 11, pp. 289–291, Nov. 1998.
[39] J. K. Tugnait and B. Huang, "On a whitening approach to partial channel estimation and blind equalization of FIR/IIR multiple-input multiple-output channels," *IEEE Trans. Signal Processing*, vol. 48, pp. 832–845, Mar. 2000.
[40] M. Kawamoto, Y. Inouye, and R. Liu, "Blind equalization algorithms for MIMO-FIR channels driven white but higher-order colored source signals," in *Proc. 34th Asilomar Conf. Signal, Systems, and Computers*, vol. 2, Pacific Grove, CA, 2000, pp. 1202–1206.
[41] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 1995.
[42] D. R. Brillinger, *Time Series: Data Analysis and Theory*. San Francisco, CA: Holden-Day, 1981.
[43] M. Rosenblatt, *Stationary Sequences and Random Fields*. Boston, MA: Birkhäuser, 1985.
[44] J. K. Massey and M. K. Sain, "Inverse of linear sequential circuits," *IEEE Trans. Comput.*, vol. C-18, pp. 330–337, 1968.
[45] G. D. Foney, "Minimal bases of rational vector spaces, with applications to multivariable linear systems," *SIAM J. Contr.*, vol. 13, no. 3, pp. 493–520, 1975.
[46] P. P. Vaidyanathan, *Multiple Systems and Filter Banks*. Eaglewood Cliffs, NJ: Prentice-Hall, 1993, p. 724.

Convergence of Stochastic-Approximation-Based Algorithms for Blind Channel Identification

Han-Fu Chen, *Fellow, IEEE*, Xi-Ren Cao, *Fellow, IEEE*, and Jie Zhu, *Member, IEEE*

***Abstract*—In this correspondence, we develop adaptive algorithms for multichannel (single-input–multiple-output, or SIMO) blind identification with both statistic and deterministic models. In these algorithms, the estimates are continuously improved while receiving new signals. Therefore, the algorithms can track the channel continuously and thus are amenable to real applications such as wireless communications. At each step, only a small amount of computation is involved. The algorithms are based on stochastic approximation methods. Convergence properties of these algorithms are proved. Simulation examples are presented to show the performance of the algorithms.**

***Index Terms*—Stochastic approximation, wireless communication.**

I. Introduction

Because of its potential applications in wireless communication and other areas, blind channel identification and equalization have become very active areas of research in recent years (see, e.g., [4], [8]–[12], [14]–[17]). The recent surveys [13], [7], and [6] contain comprehensive overviews in the area.

As characterized in [13], there are two major approaches in blind channel estimation: statistical methods and deterministic methods. In the formal approach, statistic assumptions are made for the input signals; and in the latter, input signals are assumed to be deterministic sequences. Most algorithms developed so far have been "block" algorithms in nature, i.e., a block of data sample is collected first and then processed together to get the estimation of the channel parameters; in the statistic methods, statistic quantities such as moments, etc., are estimated based on the data sample, and in the deterministic methods a set of linear equations usually with large dimensions are established.

In this correspondence, we develop adaptive algorithms for multichannel (single-input–multiple-output, or SIMO) blind identification with both statistic and deterministic models. Compared with the "block" algorithms, adaptive algorithms have their own advantages. Instead of estimating the channel parameters after the entire block of data have been received, the parameter estimates are updated when each single signal is received. The estimates are continuously improved while receiving new signals. Thus, these algorithms are more amenable to real applications such as wireless communications because channels can be tracked continuously (even for time-varying systems). Moreover, instead of a large amount of computation when all data are received, the computation is distributed at every step, and each step involves only a small amount of computation.

Manuscript received October 15, 1999; revised January 10, 2002. The work of H. F. Chen was supported by the National Key Project of China and the National Natural Science Foundation of China. The work of X.-R. Cao was supported by the Hong Kong RGC under Grant HKUST 6203/98E.

H.-F. Chen is with the Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100080, China (e-mail: hfchen@iss03.iss.ac.cn).

X.-R. Cao is with Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong (e-mail: eecao@ust.hk).

J. Zhu was with the Centre for Signal Processing, S2-B4b-05, Nanyang Technological University, Singapore 639798. He is now with Communication Department of ESS Technology Inc., Fremont, CA 94538 USA (e-mail: jie.zhu@esstech.com).

Communicated by T. Fuja, Associate Editor At Large.

Publisher Item Identifier S 0018-9448(02)03092-4.

One distinguishable feature of our algorithms is that at each step of the recursion, the received data are directly used to update the estimation of the channel parameters; statistic quantities, such as moments, are not required. The algorithms are based on the deterministic model developed in [15], and the development of these algorithms was motivated by stochastic approximation methods [1]–[3], [5]. Therefore, our algorithms differ from the previous ones. Algorithms with stochastic models require accurate estimations for stochastic quantities such as moments, which usually need a large data sample, while the deterministic algorithms developed in [15] are "block" in nature and they rely on least squares techniques for noisy channels. However, as it is well known, with fixed size of noisy data least squares estimations may not converge to the true values.

The problem considered in this correspondence is stated in Section II. In Section III, we describe the fundamentals of our two algorithms, one for the noise-free case and the other for noisy channels. The convergence properties of the algorithms are stated in Section IV for different settings. In Section IV-A, we assume that the observed data do not contain noise. Two subcases are considered. In the first subcase, only a finite noise-free data sequence is received; the set of data has to be used repeatedly, i.e., after the recursion finishes at the end of the data sequence, it starts again from the beginning of the sequence, and so on. In the second case, the algorithm is applied to an infinite data sequence obtained from independent input signals. The proofs of the convergence of the algorithms are given in the Appendix. In Section IV-B, the convergence of the algorithm is given for independent input signals with additive channel noise. The proof is presented in the Appendix. A few simulation examples are given in Section V to illustrate the performance of the algorithms. The results indicate that in addition to the desirable recursive nature, the algorithms perform reasonably well compared with the "block" algorithms. One example shows how the algorithms can be used to track a time-varying system. In Section VI, we discuss the pros and cons of the proposed approach.

II. The Problem

Consider a system consisting of p finite impulse response (FIR) channels with L being the maximum order of the channels. Let s_k, $k = 0, 1, 2, \ldots, N$, be the one-dimensional (1-D) input signal, and $x_k = (x_k^{(1)}, x_k^{(2)}, \ldots, x_k^{(p)})^\tau$, $k = L, L+1, \ldots, N$, be the p-dimensional output signal, where N is the number of samples and may not be fixed; the superscript τ denotes transpose, the superscript (i) denotes the ith component, and the subscript k is the time index. Then

$$x_k = \sum_{i=0}^{L} h_i s_{k-i}, \qquad k \geq L \tag{1}$$

where

$$h_i = \left[h_i^{(1)}, \ldots, h_i^{(p)}\right]^\tau.$$

Equation (1) can be written as

$$x_k^{(i)} = h^{(i)}(z) s_k \tag{2}$$

where

$$h^{(i)}(z) \triangleq h_0^{(i)} + h_1^{(i)} z + \cdots + h_L^{(i)} z^L, \qquad i = 1, \ldots, p \tag{3}$$

with z being the shift operator

$$z s_k = s_{k-1}.$$

The observations y_k may be corrupted by noise n_k

$$y_k = x_k + n_k$$

where n_k is a p-dimensional noise vector. The problem is to estimate $h_i, i = 0, \ldots, L$, on the basis of observations $\{y_k\}$. Note that s_k, x_k, n_k, and y_k can be complex numbers.

The channels can be characterized by a $p(L+1)$-dimensional vector h^*. First we define

$$h^{(i)} = \left(h_0^{(i)}, \ldots, h_L^{(i)}\right)^\tau$$

then let

$$h^* = \left[\left(h^{(1)}\right)^\tau, \ldots, \left(h^{(p)}\right)^\tau\right]^\tau. \tag{4}$$

We want to develop an estimate $h(k)$ for h^* at time $k = 1, 2, \ldots$, on the basis of $\{y_i, i \leq k\}$ (or $\{x_i, i \leq k\}$ when there is no noise). Our goal is to give an adaptive algorithm for $h(k)$ so that $h(k)$ is updated on-line on a sample path and $h(k) \underset{k\to\infty}{\to} \alpha h^*$, where α is an arbitrary constant.

III. The Algorithm

A comparison between our results and the approach presented in [15] will help to explain the significance of our approach. Thus, our algorithms are stated in parallel to the approach of [15]. Denote

$$\psi_k^{(i)} = \left[y_k^{(i)} \cdots y_{k-L}^{(i)}\right], \quad \varphi_k^{(i)} = \left[x_k^{(i)} \cdots x_{k-L}^{(i)}\right], \qquad i = 1, \ldots, p,\ k \geq 2L$$

where $y_k^{(i)}$ and $x_k^{(i)}$ are the ith component of y_k and x_k, respectively.

From (2), we have

$$\begin{aligned} h^{(i)}(z) x_k^{(j)} &= h^{(i)}(z) h^{(j)}(z) s_k \\ &= h^{(j)}(z) h^{(i)}(z) s_k = h^{(j)}(z) x_k^{(i)}, \\ &\qquad \forall i, j = 1, \ldots, p,\ k = 2L, 2L+1, \ldots. \end{aligned} \tag{5}$$

Using the observed data (y_k or x_k in the noise-free case), the above set of equations can be written in a matrix form

$$X_L h^* = 0 \tag{6}$$

where X_L is a $(N-2L+1)[p(p-1)/2] \times [(L+1)p]$ matrix, and N is the number of samples (see [15] for the specific form of X_L). Xu *et al.* [15] proposed to solve (6) for h^* in the deterministic case (noise-free), or to solve the following constrained least squares problem when the observations are corrupted by noise

$$h^* = \arg\left\{\min_h \|X_L h\|^2, \quad \|h\| = 1\right\}. \tag{7}$$

This approach describes the essential feature of the problem very well. The only drawback of the approach is that the matrix X_L is usually very large since N is normally a large number, thus, solving (7) is time-consuming. In addition, N is fixed; when new data are available, one has to solve (6) again to obtain new estimates. Besides, as mentioned in the Introduction, applying least squares methods requires a large data sample, which makes the computation in the algorithm very complicated.

In this correspondence, we propose adaptive algorithms in which estimates for h^* are obtained at every step $k = 2L, 2L+1, \ldots, N$ by updating the estimates obtained at the previous step. The estimates converge to h^* when N goes to infinity or the data is repeatedly used. In the noise-free case, this can be viewed as a recursive method for solving (6). For systems with noise, it is a stochastic approximation approach.

Our algorithms can be presented as follows. First, we define two $\frac{p(p-1)}{2} \times p(L+1)$ matrices denoted as Ψ_k and Φ_k

$$\Psi_k = \begin{bmatrix} \psi_k^{(2)} & -\psi_k^{(1)} & 0 & \cdots & \cdots & 0 \\ \psi_k^{(3)} & 0 & -\psi_k^{(1)} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ \psi_k^{(p)} & \psi_k^{(2)} & 0 & \cdots & 0 & -\psi_k^{(1)} \\ 0 & \psi_k^{(3)} & -\psi_k^{(2)} & 0 & \cdots & 0 \\ & \vdots & \vdots & \ddots & \ddots & \\ 0 & \psi_k^{(p)} & 0 & \cdots & 0 & -\psi_k^{(2)} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ & \vdots & & & \vdots & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & 0 & \psi_k^{(p)} & -\psi_k^{(p-1)} \end{bmatrix}. \quad (8)$$

Note that X_L in (6) is $(N-2L+1)$ times as large as $\Psi_k^{(i)}$. We define Φ_k as the matrix that has the same structure as (8) with $\psi_k^{(i)}$ replaced by $\varphi_k^{(i)}$ $\forall\, i = 1, 2, \ldots, p$. Ψ_k (or Φ_k) contains the observation x_k in the noise-free case (or y_k for noisy observations) in a window of size $L+1$ back from time instant k (i.e., $k, k-1, \ldots, k-L$); these are the observations that are related to signal s_{k-L}. It is worth emphasizing that neither Ψ_k nor Φ_k depends on N in contrast to X_L.

Let $\{a_k\}$ be a sequence of step sizes to be specified later. In the noise-free case, we take an initial value $h(2L-1) \neq 0$, and define the adaptive algorithm for $h(k)$ as follows:

$$h(k+1) = h(k) - a_k \Phi_{k+1}^{\dagger} \Phi_{k+1} h(k), \qquad k = 2L, 2L+1, \ldots. \quad (9)$$

In the noisy observation case, let $\|h(2L-1)\| < \kappa$ and define

$$\begin{aligned} h(k+1) &= \left(h(k) - a_k \left(\Psi_{k+1}^{\dagger}\Psi_{k+1} - EN_{k+1}^{\dagger}N_{k+1}\right)h(k)\right) \\ &\quad \cdot I_{[\|h(k)-a_k(\Psi_{k+1}^{\dagger}\Psi_{k+1}-EN_{k+1}^{\dagger}N_{k+1})h(k)\|<\kappa]} \\ &\quad + h(2L-1) I_{[\|h(k)-a_k(\Psi_{k+1}^{\dagger}\Psi_{k+1}-EN_{k+1}^{\dagger}N_{k+1})h(k)\|\geq\kappa]}, \\ &\qquad k = 2L, 2L+1, \ldots \quad (10) \end{aligned}$$

where $I_{[\bullet]}$ is an indicator function, the superscript † denotes transpose with complex conjugate, and $N_k = \Psi_k - \Phi_k$, $k = 2L, 2L+1, \ldots$. Without loss of generality, we can choose $\kappa = 1$. In both cases, we will prove that

$$h(k) \underset{k\to\infty}{\to} \alpha h^*$$

where α is a scalar multiple.

In fact, (10) is a truncated algorithm, which does not allow $\|h(k)\|$ to be greater than or equal to $\kappa = 1$. Once $\|h(k)\|$ reaches the truncation bound 1, the estimate is pulled back to the initial $h(2L-1)$. This does not mean that it repeats over the estimate from the beginning, because the step size has been reduced. This truncation is needed only when the initial step size is not small enough to avoid the "blow-up" of the estimate. Theoretically, we will prove that the number of such truncations is finite; in practice, truncation does not happen in our simulation presented in Section VI. The necessity for truncation also depends on the noise. When $N_k = 0$ for all k, truncation will not occur and (10) becomes (9) for the deterministic case. In the following two sections, we will study the convergence properties of the above two algorithms.

From the input sequence $\{s_0, s_1, \ldots, s_N, N \geq 2L+1\}$, we form the $(N-2L+1) \times (2L+1)$ Hankel matrix $S_N(2L+1)$

$$S_N(2L+1) \triangleq \begin{bmatrix} s_0 & s_1 & \cdots & s_{2L} \\ s_1 & s_2 & \cdots & s_{2L+1} \\ \cdots & & & \\ s_{N-2L} & s_{N-2L+1} & \cdots & s_N \end{bmatrix}. \quad (11)$$

It is clear that the maximal rank of $S_N(2L+1)$ is $2L+1$. If $S_N(2L+1)$ is of full rank, then the finite sequence $\{s_i, i = 1, \ldots, N\}$ in [15] is said to have linear complexity greater than or equal to $2L+1$.

We need the following lemma.

Lemma 1: Assume that L is known and the following conditions hold.

A1 $h^{(i)}(z), i = 1, \ldots, p$, given by (3) have no common factor.

A2 The Hankel matrix $S_N(2L+1)$ given by (11) is of full rank $(=2L+1)$.

Then h^* is the unique (up to a scalar multiple) nonzero vector simultaneously satisfying

$$\Phi_k h^* = 0, \qquad k = 2L, 2L+1, \ldots, N. \quad (12)$$

This lemma can be deduced from [15, Theorem 1]. For completeness, we provide a direct proof in the Appendix.

IV. Main Results

A. Noise-Free Observations

We first consider the case where a finite noise-free data sequence $\{x_L, x_{L+1}, \ldots, x_N\}$ is observed. In this case, $\varphi_k^{(i)}$, $i = 1, \ldots, p$, $2L \leq k \leq N$, are available. Thus, we can construct Φ_k, $k = 2L, \ldots, N$. Since the algorithm will not converge to the true value in a finite number of steps, we need to use these data repeatedly to form a sequence of infinitely many samples. To this end, we denote

$$\Phi_{k(N-2L+1)+i} = \Phi_i, \qquad i = 2L, \ldots, N,\ k = 0, 1, \ldots. \quad (13)$$

The sequence can be divided into periods each with a length of $N-2L+1$ and consisting of the same data as the first one.

For step sizes $\{a_k\}$, we need the following condition.

A3 $a_k > 0$, $a_{k+1} \leq a_k$, $\forall\, k = 1, 2, \ldots$, $a_k \underset{k\to\infty}{\to} 0$ and $\sum_{k=1}^{\infty} a_k = \infty$.

Theorem 1: Assume A1–A3 hold, and L is known. Then, for $h(k)$ given by (9) and (13) with initial value $h(2L-1)$

$$h(k) \underset{k\to\infty}{\to} \alpha h^*.$$

The proof of this theorem is given in the Appendix. As shown in the proof, we have

$$\alpha = \frac{(h^*)^{\dagger} h(2L-1)}{\|h^*\|^2}.$$

However, because h^* is unknown, this equation does not help us in determining the absolute value of h^*. It is worth noting that $(h^*)^{\dagger}h(2L-1) = 0$ is a noninteresting case, because the fact that $h(k)$ tends to zero gives no information about h^*. So, the initial value $h(2L-1)$ should have a nonzero projection on h^*.

Theorem 1 shows that our algorithm converges to the true value (with a scaling factor α) if we repeatedly apply the sequence of finite samples. Starting with any initial value $h(2L-1)$, after applying the adaptive algorithm to the sequence of samples, we obtain an estimate $h(N)$. When N is small, $h(N)$ may not be very accurate; in this case, we can

use $h(N)$ as the initial value and apply the algorithm again on the same set of data (see (13)). When N is large, $h(N)$ is accurate enough even after applying the adaptive algorithm to the sequence only once; repeatedly using the data is not needed.

In a sense, the algorithm discussed in this section can be viewed as a recursive approach to solving (6). The recursive nature makes the implementation more flexible: channel estimates can be updated when new samples are received. One nice feature of the recursive algorithm is that N is not required to be fixed. That is, the algorithm can be applied to an infinitely long sequence. As the length of the sequence increases, the estimate improves and eventually may converge to the true value under some very mild assumptions. Next, we will prove that this is indeed true if the samples are chosen from a system with independent random input signals.

We assume that the input signal $\{s_i\}$ is a sequence of infinitely many mutually independent random variables and that the observations do not involve noise; i.e., $n_k \equiv 0$. Similarly to Lemma 1 for deterministic sequences, we have the following.

Lemma 2: Assume A1 holds, L is known, and s_i is a sequence of mutually independent random variables with $E|s_i|^2 \neq 0$. Then $\overline{h}^* \triangleq h^*/\|h^*\|$ is the unique unit eigenvector corresponding to zero eigenvalue for the matrices

$$B_{j,k} \triangleq \sum_{i=j+k(2L+1)}^{j+(k+1)(2L+1)-1} E\Phi_i^{\dagger}\Phi_i, \qquad \forall\, j \geq 0,\ \forall\, k \geq 0 \tag{14}$$

and the rank of $B_{j,k}$ is $p(L+1)-1$.

The proof is given in the Appendix.

Since we have an infinitely long sequence of $\Phi_k^{\dagger}\Phi_k$, we need some assumptions on the minimal nonzero eigenvalues of $B_{j,k}$. Let $\lambda_{\min}(k)$ denote the minimal nonzero eigenvalue of $B_{0,k}$. On $\{s_i\}$ we need the following condition.

A2′ $\{s_k\}$ is a sequence of mutually independent random variables with $E|s_k|^2 \neq 0$, $\sup_k E|s_k^{2+\gamma}| < \infty$ for some $\gamma > 0$, and

$$\sum_{j=1}^{\infty} a_{(j+1)(2L+1)-1}\lambda_{\min}(j) = \infty. \tag{15}$$

A3 will be strengthened to the following.

A3′ A3 holds and

$$\sum_{i=1}^{\infty} a_i^{1+\gamma/2} < \infty$$

where γ is given in A2′.

It is obvious that if $\{s_i\}$ is an independent and identically distributed (i.i.d.) sequence, then $\lambda_{\min}(j)$ is a positive constant, and hence (15) is automatically satisfied because of A3.

Theorem 3: Assume L is known, A1, A2′, and A3′ hold, and $h(k)$ is given by (9) with initial value $h(2L-1)$. Then

$$h(k) \to \alpha h^* \qquad \text{a.s.}$$

where $\alpha = \frac{h^{*\dagger}h(2L-1)}{\|h^*\|^2}$.

Again, the proof is given in the Appendix.

B. Noisy Observations

We now consider the general case where the observations are corrupted by noise: $y_k = x_k + n_k$, and $\{s_k\}$ is a sequence of mutually independent random variables.

We will use the following conditions.

A3″ A3′ holds and $\frac{a_k}{a_{k+1}} < c$, $\forall\, k$, where c is a constant.

A4 $\{s_k\}$ and $\{n_k\}$ are mutually independent and each of them is a sequence of mutually independent random variables such that $E|s_k|^2 \neq 0$, and

$$\sup_k \{|s_k| + |n_k|\} \leq \eta < \infty, \quad E\eta^{2+\gamma} < \infty.$$

A5 $\lambda_{\min}(j,k) \geq \lambda > 0, \forall\, j \geq 0, \forall\, k \geq 0$, where $\lambda_{\min}(j,k)$ is the minimal nonzero eigenvalue of $B_{j,k}$ defined in (14).

Theorem 3: Assume L is known, A1, A3″, A4, and A5 hold, and $h(k)$ is given by (10) with initial value $h(2L-1)$. Then after a finite number of steps there is no truncation in (10) and

$$h(k) \underset{k\to\infty}{\to} \alpha h^*, \qquad \text{a.s.}$$

where α is a random variable specified later by (55).

The proof is given in the Appendix.

V. Simulation Examples

In this section, we present three computer simulation examples. In the first example, we illustrate the convergence of the algorithm in a noisy time-invariant wireless environment. In particular, we investigate the impact of the noise covariance matrix on the channel estimation. In the second one, we compare the performance of the recursive algorithm with that of the "block" algorithms proposed in [15]. In the third one, we illustrate the capability of the algorithm in tracing a time-variant channel. In all the examples, the modulation of the input signal is quaternary phase shift keying (QPSK) and the channel order is assumed to be known *a priori*. We repeat the simulation for N different random environments and compute the root-mean-square error (RMSE) as the measure of the accuracy of the channel estimates; the RMSE is defined as

$$\text{RMSE} = \frac{1}{\|h^*\|}\sqrt{\frac{1}{N}\sum_{i=1}^{N}\|\beta_i\hat{h}_i - h^*\|^2} \tag{16}$$

where $\hat{h}_i$ is the channel estimate in the ith Monte Carlo run and β a scalar that minimizes the value of $\|\beta_i\hat{h}_i - h^*\|$; i.e.,

$$\beta_i = \frac{\hat{h}_i' * h^*}{\hat{h}_i' * \hat{h}_i}. \tag{17}$$

Example 1: In this example, we select a transmitter with raised-cosine pulse $p(t)$ whose rolloff factor is 0.1. The raised-cosine pulse $p(t)$ is truncated to $4T$ where T is the baud period. A two-ray wireless radio channel with a long delay multipath is used as the channel model. The overall channel impulse response is

$$h(t) = p(t) - 0.7(1+j)p(t-1.3T). \tag{18}$$

The channel is corrupted by a Gaussian noise of 25 dB. The received signal is oversampled by a factor of 3 ($p = 3$) and thus each input symbol corresponds to a three-dimensional (3-D) output signal (a multichannel system with one input and three outputs). The number of input symbols is fixed to 100 thus the received signal consists of 3×100 samples. All components of the initial value $h(2L-1)$ are set to one. In each Monte Carlo run, the set of samples is repeatedly used 100 times in the same manner as Theorem 1. Thus, the channel estimates are updated 9700 times. The step size is initially chosen to be 0.1 and is reduced by 2% each time the received sequence of signals is repeatedly used.

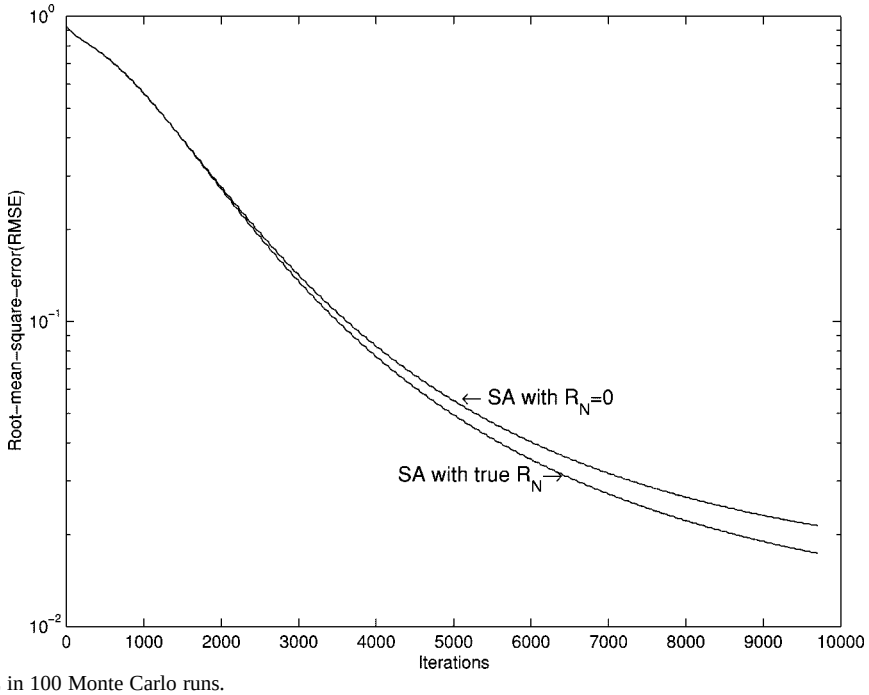

Fig. 1. The Average RMSE in 100 Monte Carlo runs.

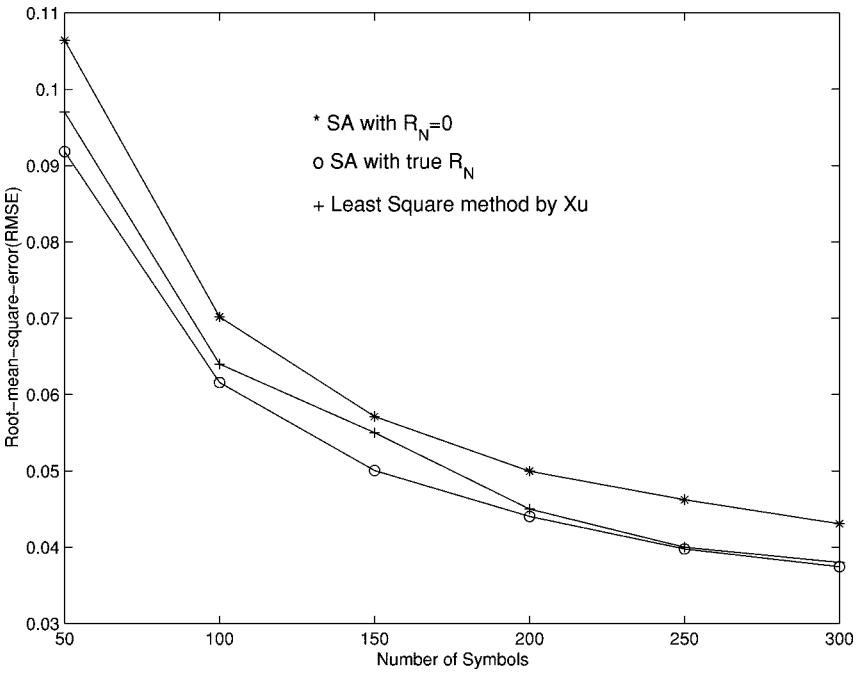

Fig. 2. Comparison of Xu's least squares and the adaptive algorithm for varying symbol size 50–300.

To study how sensitive the channel estimate is to the knowledge of R_N, we implemented two sets of simulations. In the first set, we use the true value for R_N, i.e., 25 dB; and in the second set, we simply set $R_N = 0$ in the algorithm (in (10)). Each set contains 100 Monte Carlo runs. The RMSEs are computed after each update. Fig. 1 presents the results after 100 Monte Carlo runs. It is apparent that for both cases (i.e., when R_N is known and R_N is simply set to zero) the adaptive procedure is convergent. When the number of updates increases, the effect of ignoring R_N also increases. However, the difference of the RMSEs for both cases always remains reasonably small. This shows that the algorithm is relatively insensitive to errors in estimating the variance matrix R_N.

Example 2: In the second example, we compare our algorithm with Xu's least squares algorithm in [15]. To this end, we use the same channel responses as presented in [15, Table II]. In addition, as in [15], we choose SNR to be 20 dB and let the number of symbols vary from 50 to 300. The step size for our algorithm is chosen as 0.04 initially and is reduced by 0.08% after each channel estimate update. For each case, 500 Monte Carlo runs are conducted. In each Monte Carlo run the channel estimate is updated for 6000 times. Again, two cases using a true R_N and $R_N = 0$ are simulated. Fig. 2 shows the RMSEs of the channel estimate from Xu's paper [15] and our simulation results. It is not surprising that the adaptive algorithm (10) performs a channel estimate quite close to Xu's.

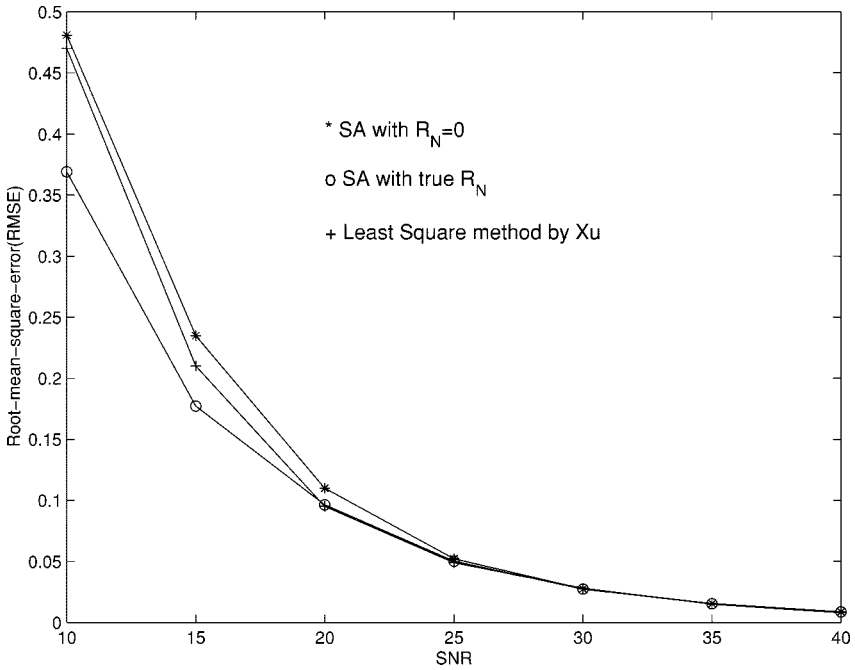

Fig. 3. Comparison of Xu's least squares and the adaptive algorithm for varying SNR 10–40 dB.

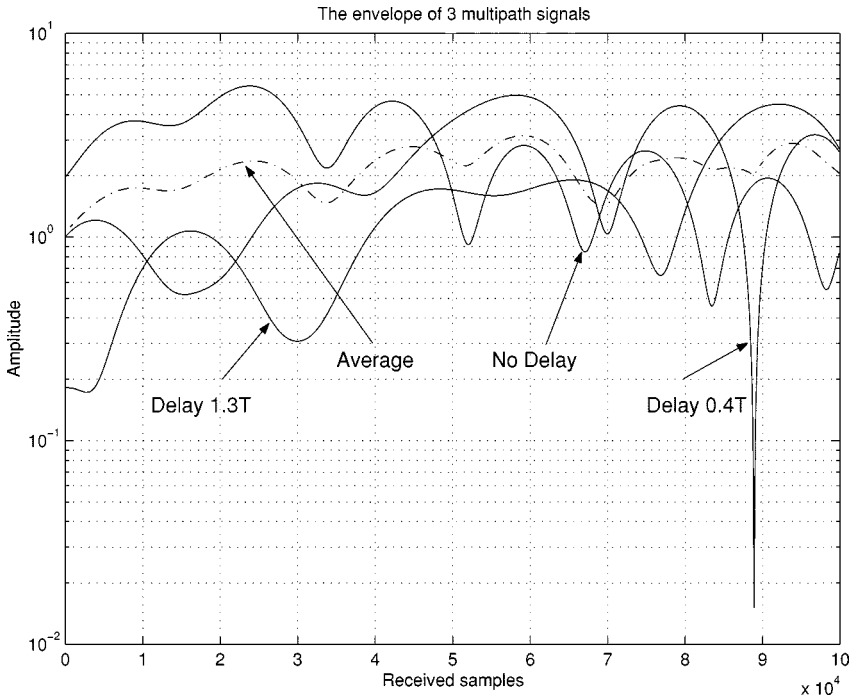

Fig. 4. Envelopes of 3 uncorrelated radio rays, Doppler frequency = 10 Hz.

We repeated the comparison under the above conditions. This time the number of symbols is kept at 100 but the signal-to-noise ratio (SNR) changes from 10 to 40 dB. Fig. 3 shows the results after 500 Monte Carlo runs. Again, the results are quite similar. The channel estimate with a known R_N is even better than Xu's result when SNR is poorer than 20 dB.

Example 3: The channel in the above examples is assumed to be time-invariant, which is a common requirement by most blind channel identification methods. As the algorithm developed is inherently adaptive, in this example we apply it to a time-varying three-ray wireless radio channel to evaluate its adaptability. It is worthwhile to mention that the convergence of the algorithm is still an open problem in such time-varying environment.

The time-varying simulation channel used in this example is defined as

$$h(t) = E_1(t)p(t) + E_2(t)p(t - 0.4T) + E_3(t)p(t - 1.3T). \quad (19)$$

where $E_1(t)$, $E_2(t)$, and $E_3(t)$ are the envelopes of three radio rays at time t, generated by Jakes model with Doppler frequency 10 Hz. Fig. 4

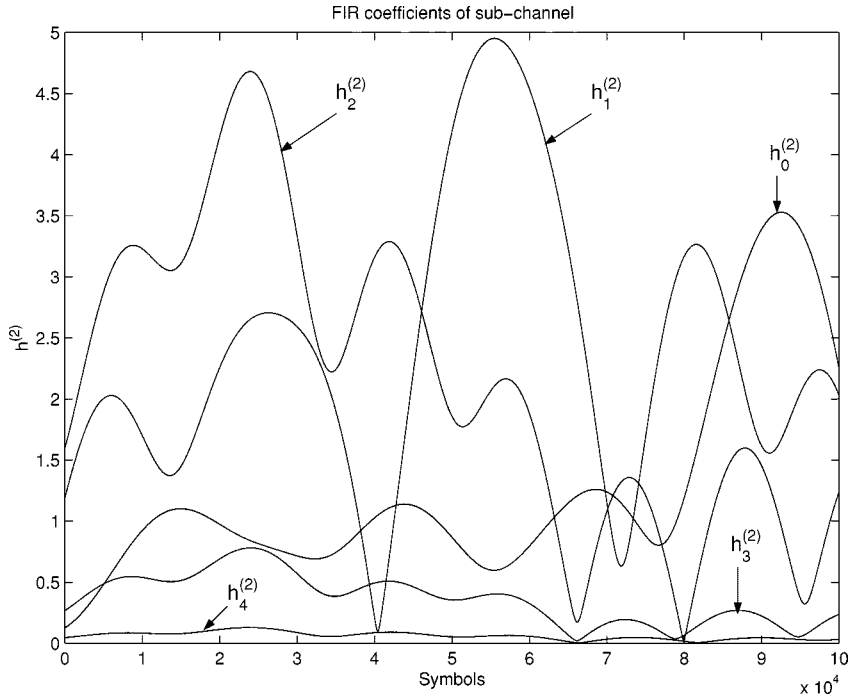

Fig. 5. Channel finite impulse response, $h^{(2)}$, $\mathrm{SNR} = 35$ dB.

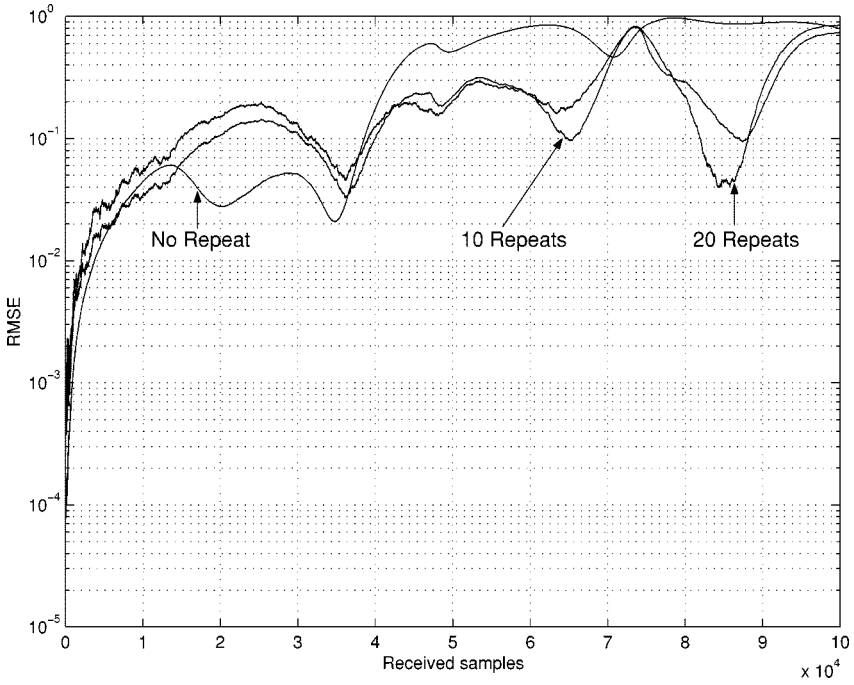

Fig. 6. RMSE for the time-varying system with SNR $= 35$ dB.

shows the change of these envelopes where $E_1(t)$, $E_2(t)$, and $E_3(t)$ are indicated by "No Delay," "Delay 0.4 T," and "Delay 1.3 T," respectively. When the carrier frequency is 900 MHz in the group special mobile (GSM) context, 10-Hz Doppler frequency is equivalent to the receiver moving toward the transmitter with velocity 12 km/h. Again, the shaping filter $p(t)$ is chosen with raised-cosine pulse truncated to $4T$. The rolloff factor is 0.5. $p = 3$ and SNR $= 35$ dB. The channel order used is fixed to 4. Consequently, there are 15 channel coefficients to be identified. As an example, the channel coefficients of subchannels $h^{(2)}$ are drawn in Fig. 5. The symbols are unit-variant white QPSK with symbol rate 300 kHz. The noise effect on the update is neglected, i.e., $R_N = 0$.

To study the tracking capability of our adaptive algorithm, we choose the initial channel parameters be their true values. The estimated channel parameters are updated once at the arrival of each symbol signal with previous estimated parameters as their current initial values. The step size is chosen to be 0.0015. The estimated RMSE is presented by the curve in Fig. 6. From the figure, the algorithm has a good tracing capability for the 35 000 symbols where the RMSE is always less than 0.06.

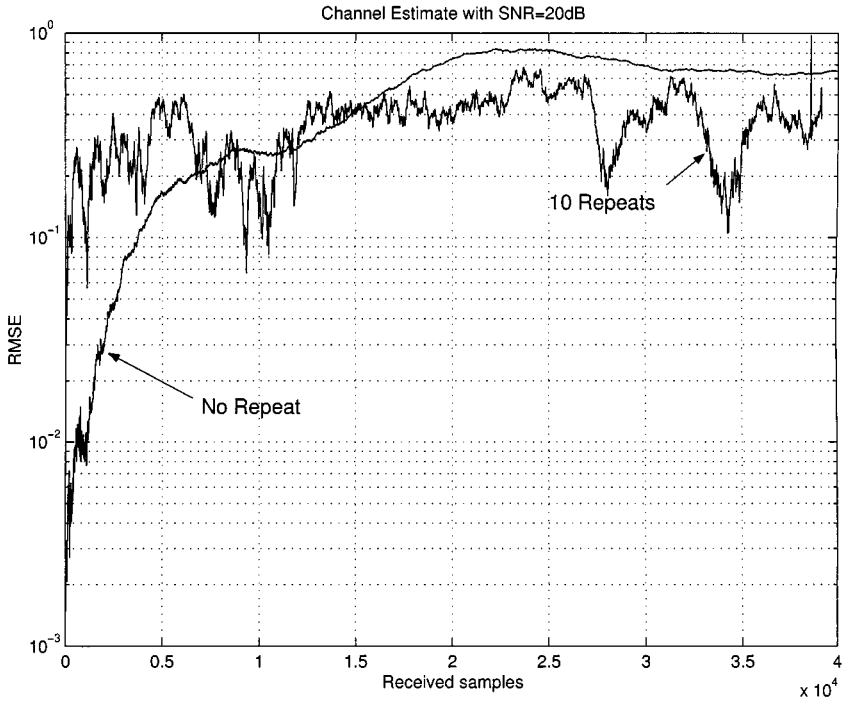

Fig. 7. RMSE for the time-varying system with SNR $=$ 20 dB.

Fig. 7 shows the same results for SNR $=$ 20 dB. One can notice that the performance of tracking deteriorates as the noise increases.

This example shows that even for the time-varying system shown in Fig. 5, the algorithm can trace the channel reasonably well with one update each symbol. The initial parameters in this simulation are chosen to be the true value. Thus, a possible application is to combine this algorithm together with some existing identification techniques such as the training signal approach. That is, to use training signals to identify the channel at the beginning of each interval, and then apply our adaptive algorithm to follow the channel in the internal. Only recursive algorithms can achieve such a goal. Further research is certainly needed.

VI. Conclusion

In this correspondence, we proposed some adaptive algorithms for blind channel identification by exploiting channel diversity. In these algorithms, the channel estimates are updated every time a new sample is received. The convergence of the algorithms is proved. Recursive algorithms are easyly implemented in real systems such as wireless channels. The algorithms converge to the true channel parameters even when the number of samples is finite if the samples are repeatedly used. In this case, the algorithm becomes a recursive version of the approach proposed in [15] for deterministic inputs.

The algorithm for noisy observations requires knowledge of the covariance R_N of the noise. In principle, R_N can be estimated based on the sequence of the observed data; the implementation details remain a future research topic. However, from the simulation results, we may expect that the estimate does not depend so much on R_N. As shown in Fig. 1, one can simply set $R_N \equiv 0$ and the additional RMSE induced is reasonably small.

The channel order L is, in general, unknown in advance. However, if Condition A4 holds and if each of the sequences are i.i.d., then the covariance matrix of observations

$$R_y(k) = E[(y_i - Ey_i)(y_{i+k} - Ey_{i+k})]$$

is independent of i, and it does not equal zero for $k = 0, 1, 2, \ldots, L$ and is 0 for $k = L + 1$. Using this property and replacing $R_y(k)$ by its sampling approximation, we can obtain consistent estimates for L as the data size increases to infinity. In practice, one may not estimate L in advance, but simply take it sufficiently large. This kind of overparametrization technique works quite well in modeling for systems like (1).

The adaptive nature of the algorithms allow them to be applied to time-varying systems. We did one simulation example to explore this property. The example shows that the tracking ability is reasonably good with 35 000 symbols. Further research is needed to study the applicability to time-varying systems.

Appendix

Proof of Lemma 1: From (5), it is readily seen that h^* satisfies (12). It remains to prove the uniqueness.

Assume that $\overline{h} \triangleq [(\overline{h}^{(1)})^\tau, \ldots, (\overline{h}^{(p)})^\tau]^\tau$ is also a solution to (12) but $\overline{h} \neq h^*$, where $\overline{h}^{(i)} = [\overline{h}_0^{(i)}, \ldots, \overline{h}_L^{(i)}]^\tau$ is $(L+1)$-dimensional, $i = 1, \ldots, p$. Define

$$\overline{h}^{(i)}(z) \triangleq \overline{h}_0^{(i)} + \overline{h}_1^{(i)} z + \cdots + \overline{h}_L^{(i)} z^L.$$

Since $\overline{h}$ is a solution to (12), we have

$$\overline{h}^{(i)}(z) x_k^{(j)} - \overline{h}^{(j)}(z) x_k^{(i)} = 0, \quad \forall\, i, j = 1, \ldots, p, \quad k = 2L, \ldots, N - (2L+1),$$

and by (2)

$$\left(\overline{h}^{(i)}(z) h^{(j)}(z) - \overline{h}^{(j)}(z) h^{(i)}(z)\right) s_k = 0, \quad \forall\, i, j = 1, \ldots, p, \quad k = 2L, \ldots, N - (2L+1). \tag{20}$$

The above set of equations implies

$$h^\dagger(i, j) S_N(2L+1) = 0, \qquad \forall\, i, j = 1, \ldots, p$$

where $h(i, j)$ denotes the $(2L+1)$-dimensional vector composed of coefficients of $\overline{h}^{(i)}(z) h^{(j)}(z) - \overline{h}^{(j)}(z) h^{(i)}(z)$ written in an increasing order of z. By A2, $h(i, j) = 0$. In other words

$$\overline{h}^{(i)}(z) h^{(j)}(z) - \overline{h}^{(j)}(z) h^{(i)}(z) = 0, \qquad \forall\, i, j = 1, \ldots, p. \tag{21}$$

For a fixed j, (20) is valid for all $i = 1, \ldots, p$, $i \neq j$. Therefore, all roots of $h^{(j)}(z)$ should be roots of $\overline{h}^{(j)}(z)h^{(i)}(z)$ for all $i \neq j$. By A1, it follows that all roots of $h^{(j)}(z)$ coincide with roots of $\overline{h}^{(j)}(z)$. This means that there is a constant α_j such that

$$\overline{h}^{(j)}(z) = \alpha_j h^{(j)}(z), \qquad \forall j = 1, \ldots, p.$$

Substituting this into (20) leads to

$$\alpha_i h^{(i)}(z)h^{(j)}(z) - \alpha_j h^{(j)}(z)h^{(i)}(z) = 0$$

and hence $\alpha_i = \alpha_j \triangleq \alpha$, $\forall i, j = 1, \ldots, p$. Thus, we conclude that $\overline{h} = \alpha h^*$. □

Proof of Theorem 1: First, we decompose $h(2L-1)$ and $h(k)$, respectively, into orthogonal vectors

$$h(2L-1) = \alpha h^* + h'(2L-1), \quad h(k) = \frac{h^\dagger(k)h^*}{\|h^*\|^2}h^* + h'(k),$$
$$h'^\dagger(k)h^* = 0,\ k = 2L-1, 2L, \ldots. \quad (22)$$

If αh^* serves as the initial value for (9), then by (12) $h(k) \equiv \alpha h^*$. Again, by (12)

$$h^{*\dagger}\left(I - a_k \Phi_{k+1}^\dagger \Phi_{k+1}\right) h'(k) = 0, \qquad \forall k \geq 2L-1.$$

We then have

$$h'(k+1) = h'(k) - a_k \Phi_{k+1}^\dagger \Phi_{k+1} h'(k) \quad (23)$$
$$h(k) = \alpha h^* + h'(k). \quad (24)$$

Therefore, for proving the theorem it suffices to show $h'(k) \longrightarrow 0$ as $k \longrightarrow \infty$.

Denote

$$A_{k+1} \triangleq \left(I - a_{(k+1)(N-2L+1)-1}\Phi^\dagger_{(k+1)(N-2L+1)} \cdot \Phi_{(k+1)(N-2L+1)}\right) \cdot \left(I - a_{(k+1)(N-2L+1)-2}\Phi^\dagger_{(k+1)(N-2L+1)-1} \cdot \Phi_{(k+1)(N-2L+1)-1}\right) \cdots \cdot \left(I - a_{k(N-2L+1)}\Phi^\dagger_{k(N-2L+1)+1}\Phi_{k(N-2L+1)+1}\right)$$

and

$$g_k \triangleq h'(k(N-2L+1)).$$

From (23) it follows that

$$g_{k+1} = A_{k+1} g_k. \quad (25)$$

Since $a_k \longrightarrow 0$ and Φ_k is uniformly bounded with respect to k, there is a large k_0 such that

$$A_{k+1} \leq I - \frac{1}{2}\sum_{i=k(N-2L+1)}^{(k+1)(N-2L+1)-1} a_i \Phi_{i+1}^\dagger \Phi_{i+1}, \quad A_k^2 \leq A_k, \quad \forall k \geq k_0. \quad (26)$$

By (13)

$$\sum_{i=k(N-2L+1)}^{(k+1)(N-2L+1)-1} \Phi_{i+1}^\dagger \Phi_{i+1} = \sum_{i=0}^{N-2L} \Phi_{i+1}^\dagger \Phi_{i+1} \quad (27)$$

which, by Lemma 1, is of rank $p(L+1)-1$ and has the unique (up to a constant) multiple eigenvector h^* corresponding to the zero eigenvalue.

Denote by $\lambda_{\min}$ the minimal nonzero eigenvalue of

$$\sum_{i=0}^{N-2L} \Phi_{i+1}^\dagger \Phi_{i+1}.$$

Let h' be an arbitrary $p(L+1)$-dimensional vector such that $h'^\dagger h^* = 0$. Then h' can be expressed by

$$h' = \sum_{i=1}^{p(L+1)-1} \alpha_i u_i$$

where u_i, $i = 1, \ldots, p(L+1)-1$, are unit eigenvectors of

$$\sum_{i=0}^{N-2L} \Phi_{i+1}^\dagger \Phi_{i+1}$$

corresponding to its nonzero eigenvalues.

Noticing

$$h'^\dagger \sum_{i=0}^{N-2L} \Phi_{i+1}^\dagger \Phi_{i+1} h' \geq \lambda_{\min}\|h'\|^2$$

and from (25) and (26), we have for $k \geq h_0$

$$\|g_{k+1}\|^2 \leq g_k^\dagger A_{k+1} g_k \leq \|g_k\|^2 - \frac{\lambda_{\min} a_{(k+1)(N-2L+1)-1}}{2}\|g_k\|^2$$

and

$$\|g_{k+1}\|^2 \leq \prod_{i=k_0}^{k+1}\left(1 - \frac{\lambda_{\min}}{2} a_{i(N-2L+1)-1}\right)\|g_{k_0}\|^2. \quad (28)$$

Since

$$\sum_{i=1}^{\infty} a_{i(N-2L+1)} \geq \frac{1}{N-2L+1}\sum_{i=N-2L+1}^{\infty} a_i = \infty$$

it follows that the right-hand side of (28) tends to zero as $k \longrightarrow \infty$. From (23), it is seen that $\|h'(k)\|$ is nonincreasing for $k \geq k_0$. Hence, the convergence $g_k \xrightarrow[k\longrightarrow\infty]{} 0$ implies $h'(k) \xrightarrow[k\longrightarrow\infty]{} 0$. □

Proof of Lemma 2: Since $\{s_i\}$ is a sequence of mutually independent random variables and $E|s_i|^2 \neq 0$, it follows that

$$ES_{4L}^{(k)}(2L+1)S_{4L}^{(k)\dagger}(2L+1) > 0, \qquad \forall k \quad (29)$$

where

$$S_{4L}^{(k)} \triangleq \begin{bmatrix} s_k & s_{k+1} & \cdots & s_{k+2L} \\ s_{k+1} & s_{k+2} & \cdots & s_{k+2L+1} \\ \cdots & & & \\ s_{k+2L} & s_{k+2L+1} & \cdots & s_{k+4L} \end{bmatrix}. \quad (30)$$

Proceeding along the same lines as the proof of Lemma 1, we obtain (20), $\forall i, j = 1, \ldots, p$, $\forall k \geq 2L$. This implies that

$$h^\dagger(i,j)S_{4L}^{(k)}(2L+1) = 0$$

and

$$h^\dagger(i,j)ES_{4L}^{(k)}(2L+1)s_{4L}^{(k)\dagger}(2L+1)h(i,j) = 0, \quad \forall i, j = 1, \ldots, p, \ \forall k \geq 2L. \quad (31)$$

From (29) and (31), it follows that $h(i,j) = 0$. Following the proof of Lemma 1, we conclude that $\overline{h}^*$ is the unique unit vector satisfying (12). Consequently, $\overline{h}^*$ is the unique unit vector satisfying

$$E\Phi_i^\dagger \Phi_i \overline{h}^* = 0, \quad \forall i\colon j + k(2L+1) \leq i \leq j + (k+1)(2L+1) - 1, \ ; \ \forall j \geq 0.$$

This shows that $B_{j,k}$ is with rank $p(L+1)-1$, $\forall j \geq 0$, $\forall k \geq 0$, and $\overline{h}^*$ is its unique unit eigenvector corresponding to zero eigenvalue. □

Proof of Theorem 2: We have shown (12) in the proof of Lemma 2. So, we still have (23) and (24), and, again, it suffices to show that $h'(k) \xrightarrow[k\longrightarrow\infty]{} 0$.

With N replaced by $4L$ in the definitions for A_k and g_k, we again obtain (25). Since $a_k \xrightarrow[k\longrightarrow\infty]{} 0$, $\{E\Phi_{k+1}^{\dagger}\Phi_{k+1}\}$ is bounded and

$$\sum_{i=1}^{\infty} a_i(\Phi_{i+1}^{\dagger}\Phi_{i+1} - E\Phi_{i+1}^{\dagger}\Phi_{i+1})$$

converges a.s. by A2′, there is a large k_0 such that

$$\begin{aligned} A_{k+1} &\le I - \sum_{i=k(2L+1)}^{(k+1)(2L+1)-1} a_i E\Phi_{i+1}^{\dagger}\Phi_{i+1} \\ &\quad - \sum_{i=k(2L+1)}^{(k+1)(2L+1)-1} a_i\left(\Phi_{i+1}^{\dagger}\Phi_{i+1} - E\Phi_{i+1}^{\dagger}\Phi_{i+1}\right) \\ &\quad + o\left(a_{k(2L+1)}\right) \\ &\le I - \frac{1}{2}\sum_{i=k(2L+1)}^{(k+1)(2L+1)-1} a_i E\Phi_{i+1}^{\dagger}\Phi_{i+1}. \end{aligned}$$

Let h' be an arbitrary $p(L+1)$-dimensional vector such that $h'^{\dagger}h^* = 0$. Then, by Lemma 2

$$h'^{\dagger}B_{0,k}h' \ge \lambda_{\min}(k)\|h'\|^2$$

and hence

$$\begin{aligned} \|g_{k+1}\|^2 &\le g_k^{\dagger}A_{k+1}g_k \\ &\le \|g_k\|^2 - \frac{a_{(k+1)(2L+1)-1}\lambda_{\min}(k)}{2}\|g_k\|^2, \qquad k \ge k_0. \end{aligned}$$

Therefore,

$$\|g_{k+1}\|^2 \le \prod_{i=k_0}^{k+1}\left(1 - \frac{\lambda_{\min}(i)}{2}\,a_{(i+1)(2L+1)-1}\right)\|g_{k_0}\|^2$$

which tends to zero since

$$\sum_{i=1}^{\infty}\lambda_{\min}(i)a_{(i+1)(2L+1)-1} = \infty$$

by A2′. This implies $h'(k) \xrightarrow[k\longrightarrow\infty]{} 0$. □

Proof of Theorem 3: Set $\|h(2L-1)\| = \delta$. Without loss and generality and for convenience in the proof, we take $\delta = \frac{1}{4}$. Define

$$R_i = \Phi_i^{\dagger}\Phi_i - E\Phi_i^{\dagger}\Phi_i + \Phi_i^{\dagger}N_i + N_i^{\dagger}\Phi_i + N_i^{\dagger}N_i - EN_i^{\dagger}N_i. \quad (32)$$

Then

$$\Psi_i^{\dagger}\Psi_i - EN_i^{\dagger}N_i = E\Phi_i^{\dagger}\Phi_i + R_i. \quad (33)$$

Denote by τ_k, $k = 1, 2, \ldots$, the truncation times, i.e., $h(\tau_k) = h(2L-1)$.

From (10) we have

$$h(\tau_k + j) = h(2L-1) - \sum_{i=\tau_k}^{\tau_k+j-1} a_i E(\Phi_{i+1}^{\dagger}\Phi_{i+1}h(i) + R_{i+1})h(i)$$

and

$$\|h(\tau_k + j)\| < 1, \qquad j = 1, \ldots, \tau_{k+1} - \tau_k - 1, \quad \forall k. \quad (34)$$

Denote $\overline{h}^* = h^*/\|h^*\|$. Let

$$U = \begin{bmatrix} V & \overline{h}^* \end{bmatrix}$$

be a Hermite matrix. Denoting

$$h'(k) \triangleq VV^{\dagger}h(k), \qquad \overline{h}(k) \triangleq \overline{h}^{*\dagger}h(k) \quad (35)$$

we have

$$h(k) = h'(k) + \overline{h}(k)\overline{h}^*. \quad (36)$$

Noticing $\Phi_i\overline{h}^* = 0$, $\forall i$ and

$$VV^{\dagger}E\Phi_{i+1}^{\dagger}\Phi_{i+1} = E\Phi_{i+1}^{\dagger}\Phi_{i+1}$$

by (34), (36) we find that

$$\begin{aligned} h'(\tau_k + j) &= h'(2L-1) - \sum_{i=\tau_k}^{\tau_k+j-1} a_i E\Phi_{i+1}^{\dagger}\Phi_{i+1}h'(i) \\ &\quad - \sum_{i=\tau_k}^{\tau_k+j-1} a_i VV^{\dagger}R_{i+1}h(i) \end{aligned} \quad (37)$$

and

$$\begin{aligned} \overline{h}(\tau_k + j) &= \overline{h}(2L-1) - \sum_{i=\tau_k}^{\tau_k+j-1} a_i\overline{h}^{*\dagger}R_{i+1}h(i), \\ &\qquad j = 1, \ldots, \tau_{k+1} - \tau_k - 1. \end{aligned} \quad (38)$$

We complete the proof in five steps.

Step 1: First, we show that

$$\sum_{i=1}^{\infty} a_i R_{i+1}h(i) < \infty \qquad \text{a.s.} \quad (39)$$

By A4, $D_i \triangleq \Phi_i N_i + N_i^{\dagger}\Phi_i + N_i^{\dagger}N_i - EN_i^{\dagger}N_i$ is a martingale difference sequence with $\sup_i E\|D_i\|^{1+\frac{\gamma}{2}} < \infty$. Since

$$\sum_{i=1}^{\infty} a_i^{1+\frac{\gamma}{2}} < \infty \quad \text{and} \quad \|h(i)\| < 1$$

it follows that

$$\sum_{i=1}^{\infty} a_i D_{i+1}h(i) < \infty \qquad \text{a.s.} \quad (40)$$

by the convergence theorem for martingale difference sequences.

Since $\Phi_i^{\dagger}\Phi_i - E\Phi_i^{\dagger}\Phi_i$ is independent of $\Phi_{i+2L+1}^{\dagger}\Phi_{i+2L+1} - E\Phi_{i+2L+1}^{\dagger}\Phi_{i+2L+1}$ and

$$\sup_i E\|\Phi_i^{\dagger}\Phi_i - E\Phi_i^{\dagger}\Phi_i\|^{1+\frac{\gamma}{2}} < \infty$$

we also have

$$\sum_{i=1}^{\infty} a_i\left(\Phi_{i+1}^{\dagger}\Phi_{i+1} - E\Phi_{i+1}^{\dagger}\Phi_{i+1}\right)h(i) < \infty \qquad \text{a.s.}$$

Incorporating the above inequality with (40) yields (39).

Step 2: Next, we show that for any $j \ge \tau_k$ and $T > 0$ with $m(j,T) \le \tau_{k+1} - \tau_k - 1$, there is a $c_0 > 0$, which possibly depends on the sample path but is independent of k, j, and T, such that

$$\|h(i+1) - h(j)\| \le c_0 t, \qquad \forall\, i\colon j \le i \le m(j,t), \ \ \forall t \in [0,T] \quad (41)$$

where

$$m(j,T) = \max\left\{k\colon \sum_{i=j}^{k} a_i \le T\right\}.$$

By A4, there is a $c_0 > 0$, possibly depending on the sample path, such that

$$\left\|E\Phi_{i+1}^{\dagger}\Phi_{i+1} + R_{i+1}\right\| < c_0, \qquad \forall i.$$

Then, by noticing $\|h(i)\| < 1$, (41) immediately follows from (34).

Step 3: Third, we show that for any $h' \triangleq VV^{\dagger}h \neq 0$, there exists an N which is large enough, such that

$$-h'^{\dagger}\sum_{i=j}^{m(j,t)} a_i E\Phi_{i+1}^{\dagger}\Phi_{i+1}h' \leq -\epsilon t\|h'\|^2, \quad \forall j \geq N,\ \forall t \in (0,T] \tag{42}$$

where $\epsilon = \frac{\lambda}{2(2L+1)c^{2L+1}} > 0$ with c and λ given by A3$''$ and A5, respectively.

Let $[u_1^{(j,k)}, \ldots, u_{p(L+1)-1}^{(j,k)}, \overline{h}^*]$ be the Hermite matrix composed of eigenvectors of $B_{j,k}$. By Lemma 2, $\overline{h}^*$ is its unique eigenvector corresponding to the zero eigenvalue.

Since $h'^{\dagger}\overline{h}^* = 0$, h' can be expressed as

$$h' = \sum_{i=1}^{p(L+1)-1} \alpha_i u_i^{(j,k)}.$$

Then

$$-h'^{\dagger}B_{j,k}h' \leq -\lambda_{\min}(j,k)\sum_{i=1}^{p(L+1)-1}|\alpha_i|^2 \leq -\lambda\|h'\|^2, \quad \forall j, \forall k. \tag{43}$$

Noticing that by A4 and (43), and the fact that $\|E\Phi_{i+1}^{\dagger}\Phi_{i+1}\|$ is bounded with respect to i and $\{a_i\}$ is nonincreasing, we have

$$\begin{aligned}
&-h'^{\dagger}\sum_{i=j}^{m(j,t)} a_i E\Phi_{i+1}^{\dagger}\Phi_{i+1}h' \\
&\quad \leq -h'^{\dagger}\sum_{k=0}^{[\frac{m(j,k)-j+1}{2L+1}]}\ \sum_{i=j+k(2L+1)}^{j+(k+1)(2L+1)-1} a_i E\Phi_{i+1}^{\dagger}\Phi_{i+1}h' \\
&\quad \leq -\lambda\|h'\|^2\sum_{k=0}^{[\frac{m(j,k)-j+1}{2L+1}]} a_{j+(k+1)(2L+1)-1},
\end{aligned} \tag{44}$$

where $[x]$ denotes the integer part of the real number x.

Since $\frac{a_{k+1}}{a_k} \geq \frac{1}{c}$ and $c \geq 1$, from (44) it follows that for a large enough j

$$\begin{aligned}
&-h'^{\dagger}\sum_{i=j}^{m(j,t)} a_i E\Phi_{i+1}^{\dagger}\Phi_{i+1}h' \\
&\quad \leq -\lambda\|h'\|^2\sum_{k=0}^{[\frac{m(j,t)-j+1}{2L+1}]}\frac{1}{2L+1}\sum_{i=0}^{2L} a_{j+(k+1)(2L+1)-1} \\
&\quad \leq -\frac{\lambda\|h'\|^2}{(2L+1)c^{2L+1}}\sum_{k=0}^{[\frac{m(j,t)-j+1}{2L+1}]}\sum_{i=0}^{2L} a_{j+k(2L+1)+i} \\
&\quad = -\frac{\lambda\|h'\|^2}{(2L+1)c^{2L+1}}\sum_{i=0}^{[\frac{m(j,t)-j+1}{2L+1}](2L+1)+2L} a_{j+i} \\
&\quad = -\frac{\lambda\|h'\|^2}{(2L+1)c^{2L+1}}\sum_{i=0}^{m(j,t)} a_i + o(1) \leq -\epsilon t\|h'\|^2
\end{aligned}$$

where $o(1) \longrightarrow 0$ as $j \longrightarrow \infty$.

Step 4: We now show that the number of truncations in (10) is finite. To this end, let us first assume that the converse is true: $\tau_k \xrightarrow[k\longrightarrow\infty]{} \infty$. Then, by (39) and $\|h(2L-1)\| = \frac{1}{4}$, there exists a large K such that

$$\left\|\sum_{i=\tau_k}^{\tau_k+j} a_i R_{i+1}h(i)\right\| \leq \frac{1}{8}$$

and

$$\left\|h^{\dagger}(2L-1)\overline{h}^*\overline{h}^* - \sum_{i=\tau_k}^{\tau_k+j} a_i R_{i+1}h(i)\right\| < \frac{3}{8}, \quad \forall j \geq 1,\ \forall k \geq K. \tag{45}$$

By the definition of τ_{k+1}, we have

$$\left\|h(2L-1) - \sum_{i=\tau_k}^{\tau_{k+1}-1} a_i\left(E\Phi_{i+1}^{\dagger}\Phi_{i+1} + R_{i+1}\right)h(i)\right\| \geq 1.$$

Incorporating the above with (45) implies

$$\left\|h'(2L-1) - \sum_{i=\tau_k}^{\tau_{k+1}-1} a_i E\Phi_{i+1}^{\dagger}\Phi_{i+1}h'(i)\right\| \geq \frac{5}{8},$$

and

$$\left\|h'(2L-1) - \sum_{i=\tau_k}^{\tau_{k+1}-1} a_i E\Phi_{i+1}^{\dagger}\Phi_{i+1}h'(i) - \sum_{i=\tau_k}^{\tau_{k+1}-1} a_i VV^{\dagger}R_{i+1}h(i)\right\| > \frac{1}{2}. \tag{46}$$

Define

$$\begin{aligned}
j(k) \triangleq \Bigg\{ &\min\ j\colon j < \tau_{k+1} - \tau_k, \\
&\left\|h'(2L-1) - \sum_{i=\tau_k}^{\tau_{k+1}-1} a_i E\Phi_{i+1}^{\dagger}\Phi_{i+1}h'(i) - \sum_{i=\tau_k}^{\tau_{k+1}-1} a_i VV^{+}R_{i+1}h(i)\right\| > \frac{1}{2}\Bigg\}.
\end{aligned} \tag{47}$$

By (39), $a_i \longrightarrow 0$, and the boundedness of $E\Phi_{i+1}^{\dagger}\Phi_{i+1}h'(i)$, we have

$$a_i E\Phi_{i+1}^{\dagger}\Phi_{i+1}h'(i) - a_i VV^{\dagger}R_{i+1}h(i) \xrightarrow[i\longrightarrow\infty]{} 0.$$

Thus, $j(k)$ is well-defined. Consequently

$$\begin{aligned}
h(\tau_k + j) &= h(2L-1) - \sum_{i=\tau_k}^{\tau_{k+1}+j-1} a_i(E\Phi_{i+1}^{\dagger}\Phi_{i+1} + R_{i+1})h(i) \qquad (48)\\
h'(\tau_k + j) &= h'(2L-1) - \sum_{i=\tau_k}^{\tau_k+j-1} a_i E\Phi_{i+1}^{\dagger}\Phi_{i+1}h'(i) \\
&\quad - \sum_{i=\tau_k}^{\tau_k+j-1} a_i VV^{\dagger}R_{i+1}h(i), \\
&\qquad \forall j\colon 0 \leq j \leq j(k),\ \forall k \geq K. \qquad (49)
\end{aligned}$$

Next, we say that the sequence

$$\{\|h'(\tau_k + i)\|^2,\ i = l_k, l_k+1, \ldots, m_k\}$$

crosses an interval $[a, b]$ with $0 < a < b < \infty$, if $\|h'(\tau_k + l_k)\|^2 \leq a$, $\|h'(\tau_k + m_k)\|^2 \geq b$, $a < \|h'(\tau_k + i)\|^2 < b, \forall i\colon l_k < i < m_k$, and if there is no truncation for any $i\colon l_k \leq i \leq m_k$.

From (47)–(49) and $\|h(2L-1)\| = \frac{1}{4}$ it is seen that the sequence $\{\|h'(\tau_k + i)\|^2,\ i = 0, 1, \ldots, j(k)\}$ crosses the interval $[\frac{1}{16}, \frac{1}{4}]$ for each $k \geq K$.

Without loss of generality, we may assume $h'(\tau_k + l_k)$ converges: $h'(\tau_k + l_k) \longrightarrow h'$ as $k \longrightarrow \infty$. It is clear that $\|h'\| = \frac{1}{4}$ and $h'^{\dagger}h^* = 0$. By (41), there is no truncation for $h(i)$

$$i = \tau_k + l_k,\ \tau_k + l_k + 1, \ldots, m(\tau_k + l_k, T) + 1$$

if T is small enough. Then by Taylor's expansion and (39) and (41), it follows that

$$\begin{aligned}
&\|h'(m(\tau_k+l_k,T)+1)\|^2-\|h'(\tau_k+l_k)\|^2\\
&\quad=-\sum_{i=\tau_k+l_k}^{m(\tau_k+l_k,T)} h'^{\dagger}a_i\\
&\qquad\cdot\Big(E\Phi_{i+1}^{\dagger}\Phi_{i+1}h'(i)+VV^{\dagger}R_{i+1}h(i)\Big)+o(T)\\
&\quad=-\sum_{i=\tau_k+l_k}^{m(\tau_k+l_k,T)} h'^{\dagger}a_iE\Phi_{i+1}^{\dagger}\Phi_{i+1}(h'+h'(i)-h')\\
&\qquad-\sum_{i=\tau_k+l_k}^{m(\tau_k+l_k,T)} h'^{\dagger}a_iVV^{\dagger}R_{i+1}h(i)+o(T)\\
&\quad=-\sum_{i=\tau_k+l_k}^{m(\tau_k+l_k,T)} h'^{\dagger}a_iE\Phi_{i+1}^{\dagger}\Phi_{i+1}h'(i)+o(1)+o(T) \qquad (50)
\end{aligned}$$

where $o(1)\longrightarrow 0$ as $k\longrightarrow\infty$ and $o(T)\longrightarrow 0$ as $T\longrightarrow 0$.

By (42), for a large k and a small T, we have

$$\|h'(m(\tau_k+l_k,T)+1)\|^2-\|h'(\tau_k+l_k)\|^2\le-\frac{\epsilon T}{2}\|h'\|^2. \quad (51)$$

By (41),

$$h'(m(\tau_k+l_k,T)+1)\xrightarrow[T\longrightarrow 0]{} h'(\tau_k+l_k).$$

Noticing that $\|h'(\tau_k+l_k)\|^2\le\frac{1}{16}$ and $\|h'(\tau_k+m_k)\|^2\ge\frac{1}{4}$, by the definition of crossing we see that $m(\tau_k+l_k,T)+1<\tau_k+m_k$ for small enough T. In order words

$$\|h'(m(\tau_k+l_k,T)+1)\|^2>\frac{1}{16}. \quad (52)$$

Letting $k\longrightarrow\infty$ in (51) leads to a contradictory inequality

$$\limsup_{k\longrightarrow\infty}\|h'(m(\tau_k+l_k,T)+1)\|^2<\frac{1}{16}-\frac{\epsilon T}{2}\|h'\|^2 \quad (53)$$

since the left-hand side of (53) should be greater than or equal to $\frac{1}{16}$ by (52). The obtained contradiction shows that $\tau_k\longrightarrow\tau<\infty$ (i.e., there are only a finite number of truncations) a.s.

Step 5: Now we are ready to complete the proof of the theorem.

We have shown that starting from a random τ, the algorithm (10) suffers from no more truncation. Now, let us first show that $\|h'(k)\|^2$ converges. If this were not true, then

$$\liminf_{k\longrightarrow\infty}\|h'(k)\|^2<\limsup_{k\longrightarrow\infty}\|h'(k)\|^2$$

and $\|h'(k)\|^2$ would cross a nonempty interval $[a,b]$ for infinitely many times. As shown above, this is impossible. Therefore, $\|h'(k)\|^2$ converges. If the limit of $\|h'(k)\|$ were not zero, then there would exist a convergent subsequence $h'(k_j)\longrightarrow h'\neq 0$. Replacing τ_k+l_k in (50) by k_j, from (51) it follows that

$$\|h'(m(k_j,\tau)+1)\|^2-\|h'(k_j)\|^2\le-\frac{\epsilon T}{2}\|h'\|. \quad (54)$$

Since $\|h'(k)\|^2$ converges, the left-hand side of (54) tends to zero, which makes (54) a contradictory inequality. Thus, we have proved $h'(k)\xrightarrow[k\longrightarrow\infty]{}0$ a.s.

Since $\lim_{k\longrightarrow\infty}\tau_k=\tau<\infty$, from (38) it follows that

$$\lim_{k\longrightarrow\infty}\overline{h}(k)=h^{\dagger}(2L-1)\overline{h}^{*}-\sum_{i=\tau}^{\infty}a_i\overline{h}^{*\dagger}R_{i+1}h(i)\triangleq\alpha. \quad (55)$$

Finally, by the fact $h'(k)\xrightarrow[k\longrightarrow\infty]{}0$, from (36) and (55) it follows that

$$\lim_{k\longrightarrow\infty}h(k)=\alpha\overline{h}^{*}. \qquad \square$$

REFERENCES

[1] A. Benveniste, M. Metivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximations*. Berlin, Germany: Springer, 1990.

[2] H. F. Chen, "Stochastic approximation and its new applications," in *Proc. 1994 Hong Kong Int. Worksh. New Directions of Control and Manufacturing*, 1994, pp. 2–12.

[3] H. F. Chen and Y. M. Zhu, *Stochastic Approximation* (in Chinese). Shanghai, China: Shanghai Scientific and Technical Publishers, 1996.

[4] Z. Ding and Y. Li, "On channel identification based on second order cyclic spectra," *IEEE Trans. Signal Processing*, vol. 42, pp. 1260–1264, May 1994.

[5] H. J. Kushner and G. Yin, *Stochastic Approximation Algorithms and Applications*. Berlin, Germany: Springer, 1997.

[6] R. Liu, "Blind signal processing: An introduction," in *Proc. 1996 Int. Symp. Circuits and Systems*, vol. 2, May 1996, pp. 81–83.

[7] H. Liu, G. Xu, L. Tong, and T. Kailath, "Recent developments in blind channel equalization: From cyclostationarity to subspacs," *Signal Processing*, vol. 50, pp. 83–99, 1996.

[8] E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Processing*, vol. 43, pp. 516–525, Feb. 1995.

[9] Y. Sato, "A method of self-recovering equalization for multilevel amplitude-modulation," *IEEE Trans. Commun.*, vol. COM-23, pp. 679–682, June 1975.

[10] D. Slock, "Blind fractionally-spaced equalization, perfect-reconstruction filter banks and multichannel linear prediction," in *Proc. IEEE ICASSP*, Adelaide, Australia, May 1994, pp. 585–588.

[11] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: A time domain approach," *IEEE Trans. Inform. Theory*, vol. 40, pp. 340–349, Mar. 1994.

[12] ——, "Blind channel identification based on second-order statistics: A frequency-domain approach," *IEEE Trans. Inform. Theory*, vol. 41, pp. 329–334, Mar. 1995.

[13] L. Tong and S. Perreau, "Multichannel blind identification: From subspace to maximum likelihood methods," *Proc. IEEE* , to be published.

[14] A. J. van der Veen, S. Talwar, and A. Paulraj, "Blind estimation of multiple digital signals transmitted over FIR channels," *IEEE Signal Processing Lett.*, vol. 2, no. 5, pp. 99–102, May 1995.

[15] G. Xu, L. Tong, and T. Kailath, "A least-squares approach to blind identification," *IEEE Trans. Signal Processing*, vol. 43, pp. 2982–2993, Dec. 1995.

[16] J. Zhu, X.-R. Cao, and R.-W. Liu, "A blind fractionally-spaced equalizer using higher order statistics," *IEEE Trans. Circuits Syst., II*, vol. 46, pp. 755–764, June 1999.

[17] J. Zhu, Z. Ding, and X.-R. Cao, "Column anchoring zeroforcing blind equalization for wireless FIR channels," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 411–423, Mar. 1999.