Journal of Global Optimization **27:** 249–271, 2003.

Blind Channel Identification Based on Noisy Observation by Stochastic Approximation Method ⋆

HAI-TAO FANG and HAN-FU CHEN
Laboratory of Systems and Control, Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, 100080, P.R. China
(E-mail: hfchen@isso3.iss.ac.cn)

Abstract. A stochastic approximation algorithm for estimating multichannel coefficients is proposed, and the estimate is proved to converge to the true parameters a.s. up-to a constant scaling factor. The estimate is updated after receiving each new observation, so the output data need not be collected in advance. The input signal is allowed to be dependent and the observation is allowed to be corrupted by noise, but no noise statistics are used in the estimation algorithm.

Key words: Stochastic approximation, blind identification, on-line update, noisy observation, strong consistency.

1. Introduction

For recent years the blind channel identification and blind equalization have attracted great research interest in the area of signal processing and communication ([9,10,13]), and many estimation algorithms have been proposed (see e.g. [6,7,11,12,14-16]). Most results published so far are concerned with "block" algorithms, i.e., the estimation for channel coefficients and for input signal is carried out after having entire data been collected. In contrast to this, in the recent papers [5,18] the on-line recursive channel estimation algorithms have been proposed, where the sample size N of the output data is not fixed and the estimate is updated by use of each observation of the channel output. It is proved in [5], that the estimate for channel coefficients converges a.s. to the true ones up-to a constant scaling factor where the channel input may be random or deterministic and the observations may be free of or corrupted by noise. However, in the case where the observation is with additive noise, the noise variance is used in the algorithm proposed in [5]. This greatly limits the potential application of the algorithm. Further, the input signal is required to be mutually independent in [5]. The aim of this paper is to remove using the noise variance in the algorithm, to extend the input signal

⋆ Supported by the National Key Project of China and the National Natural Science foundation of China.

from independent to dependent, and to prove the a.s. convergence of the proposed algorithm.

Stochastic approximation ([1,3,4,8]) is a tool to deal with root-seeking problems for an unknown regression function which can be observed but the observations are corrupted by errors which may contain both the random noise and the structural error where the vector $\mathbf{h}^0$ composed of channel coefficients is the unique root of the unknown regression function. In the present case the noise variance is unknown, and this causes additional error in the observations. As a result, the root set of the corresponding regression function no longer consists of a singleton but a set of isolated points including the sought-for $\mathbf{h}^0$. After establishing the convergence of the applied stochastic approximation algorithm the key difficulty is to clarify of the limit is $\mathbf{h}^0$ or not.

We overcome this difficulty by using a property of stochastic approximation consisting in that the algorithm cannot converge to an unstable equilibrium of the associated homogeneous difference equation if the noise added to the difference equation effects in all directions. As a matter of fact, it will be shown that the algorithm using noisy data converges to an eigenvector of the matrix C to be defined later on, and under some reasonable conditions on the observation noise the limit of the algorithm must be the eigenvector corresponding to the minimum eigenvalue ($= 0$) of C. On the other hand, it turns out that the vector composed of the channel coefficients coincides with this eigenvector up-to a constant scaling factor. This will be demonstrated in the subsequent sections.

In Section 2, the recursive algorithm for estimating channel coefficients is defined. In Section 3, conditions used for convergence of the algorithm are listed and some auxiliary lemmas are proved. The main convergence theorem and its proof are given in Section 4, but the proof for a technical point is placed in the Appendix. A brief conclusion is contained in Section 5.

2. Recursive Algorithm for Blind Identification

Let s_k be one-dimensional input, $\mathbf{x}_k = (x_k^{(1)}, \cdots, x_k^{(p)})^\tau$ be the output of p sensors at time k, and let $\mathbf{x}_k$ be related with s_k as follows:

$$\mathbf{x}_k = \sum_{i=0}^{L} \mathbf{h}_i s_{k-i},$$

where $\mathbf{h}_i = (h_i^{(1)}, \cdots, h_i^{(p)})^\tau$, $i = 0, \cdots, L$, are unknown channel coefficients.

Denote by $\mathbf{h}^{(j)} = (h_0^{(j)}, \cdots, h_L^{(j)})^\tau$, the coefficients of the jth channel $j = 1, \cdots, p$, and by a long vector

$$\mathbf{h}^0 = (\mathbf{h}_0^{(\tau)}, \cdots, \mathbf{h}_L^{(\tau)})^\tau \tag{1}$$

the coeffficients of the whole system.

Assume the observation of the output is corrupted by noise, and the observation at time k is

$$\mathbf{y}_k = \mathbf{x}_k + \xi_k = \sum_{i=0}^{L} \mathbf{h}_i s_{k-i} + \xi_k,$$

where $\xi_k = (\xi_k^{(1)}, \cdots, \xi_k^{(p)})^\tau$ is the observation noise.

The problem of blind channel identification based on noisy observations is to estimate $\mathbf{h}^0$ by using the data $\{\mathbf{y}_i, i = 1, \cdots, k\}$. Further, we want to recursively estimate $\mathbf{h}^0$ updating the estimate $\mathbf{h}_k$ for $\mathbf{h}^0$ at time k by using the new observation $\mathbf{y}_{k+1}$, $k = 1, 2, \cdots$.

For defining estimation algorithm we introduce the matrices X_k and N_k as follows:

$$X_k = \begin{pmatrix} x_k^{(2)} & -x_k^{(1)} & 0 & 0 & 0 \\ x_k^{(3)} & 0 & -x_k^{(1)} & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_k^{(p)} & 0 & 0 & 0 & -x_k^{(1)} \\ 0 & x_k^{(3)} & -x_k^{(2)} & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & x_k^{(p)} & 0 & 0 & -x_k^{(2)} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & x_k^{(p)} & -x_k^{(p-1)} \end{pmatrix}.$$

Define matrix N_k with the same structure as X_k but with $x_k^{(i)}$ replaced by $\xi_k^{(i)}$.

Further define

$$\Phi_k = (X_k, \cdots, X_{k-L}), \ \ \Xi_k = (N_k, \cdots, N_{k-L}) \tag{2}$$

and

$$\Psi_k = \Phi_k + \Xi_k. \tag{3}$$

It is clear that Φ_k, Ξ_k and Ψ_k are $p(p-1)/2 \times p(L+1)$-matrices.

The estimate for the channel coefficients $\mathbf{h}^0$ is given by the following normalized stochastic approximation algorithm:

$$\tilde{\mathbf{h}}_{k+1} = \mathbf{h}_k - a_k \Psi_{k+1}^\tau \Psi_{k+1} \mathbf{h}_k, \tag{4}$$

$$\mathbf{h}_{k+1} = \tilde{\mathbf{h}}_{k+1} / \left\| \tilde{\mathbf{h}}_{k+1} \right\|, \tag{5}$$

where a_k is the stepsize.

From (4)(5) it is seen that Ψ_k is changed to Ψ_{k+1} after receiving a new observation $\mathbf{y}_{k+1}$, and this yields the update of the estimate from $\mathbf{h}_k$ to $\mathbf{h}_{k+1}$. In contrast to the algorithm used in [5], here the variance of $\{\xi_k\}$ is not used and the algorithm in normalized and hence is nonlinear. The aim of the paper is to show that $\mathbf{h}_k$ converges to the channel coefficient $\mathbf{h}^0$ up-to scaling factor.

3. Auxiliary Lemmas

We first list conditions to be used in the paper.

A1) The input $\{s_k\}$ is a ϕ-mixing sequence, i.e. there exist a constant $M \geqslant 0$ and a function $\phi(m) \underset{m\to\infty}{\longrightarrow} 0$ such that for any $n \geqslant 1$

$$\sup_{V\in\mathcal{F}_1^n, U\in\mathcal{F}_{n+m}^\infty} |P(U|V) - P(U)| \leqslant \phi(m),\ \forall m \geqslant M,$$

where $\mathcal{F}_i^j = \sigma\{s_k, i \leqslant k \leqslant j\}$;

A2) There exists a distribution function $F_0(\cdot)$ over $\mathbb{R}^{2L+1}$ such that

$$\left|\sup_{S\in\mathcal{B}^{2L+1}} P\{(s_{k-2L}, \cdots s_k) \in S\} - \int_S dF_0(\mathbf{w})\right| \underset{k\to\infty}{\longrightarrow} 0,$$

where $\mathcal{B}^{2L+1}$ denotes the Borel σ-algebra in $\mathbb{R}^{2L+1}$ and $\mathbf{w} = (w_1, \cdots, w_{2L+1})^\tau$;

A3) The $(2L+1)\times(2L+1)$-matrix $Q = (q_{ij})$ with $q_{ij} = \int_{\mathbb{R}^{2L+1}} w_i w_j dF_0(\mathbf{w})$ is nondegenerate;

A4) The signal $\{s_k\}$ is independent of $\{\xi_k\}$ and $\sup_k |s_k(\omega)| \leqslant \zeta(\omega) < \infty$, where $\zeta(\omega)$ is a random variable with

$$E\zeta^{2+\gamma} < +\infty$$

for some $\gamma > 0$;

A5) All components $\{\xi_k^{(i)}, i = 1, \cdots, p, k = 1, 2, \cdots\}$ of $\{\xi_k\}$ are mutually independent with $E\{\xi_k\} = 0$, $E\{(\xi_k^{(i)})^3\} = 0$, $E\{(\xi_k^{(i)})^2\} = c > 0$, and $E\{((\xi_k^{(i)})^2 - c)^2\} > 0$, $\forall i, k$, and $\{\xi_k^{(i)}\}$ is bounded by a constant, i.e. $\sup_k \|\xi_k(\omega)\| < \xi < \infty$, where ξ is a constant;

A6) The polynomials $\{h^{(i)}(z)\}$ characterizing subchannels do not share common zeros, where

$$h^{(i)}(z) = h_0^{(i)} + h_1^{(i)} z + \cdots + h_L^{(i)} z^L,\ i = 1, \cdots, p; \tag{6}$$

A7) $a_k > 0$, $\sum_k a_k = +\infty$, $\sum_k a_k^2 < \infty$ and $a_{k+1}/a_k = 1 + O(a_k)$.

We note that Conditions A1)-A4) are imposed on the input signal. By these conditions the input is allowed to be a $\phi-$mixing sequence of not equally distributed random variables having a $(2L+1)$-dimensional joint limit distribution with nondegenerate covariance matrix. The input is also assumed to be bounded by a random variable. Condition A5) is on the observation noise, requiring it be bounded by a constant among other requirements. Conditions A6), A7) are quite standard, but a rate for a_k is required when it tends to zero.

In the sequel, $I_{n\times n}$ denotes the n-dimensional identity matrix.

LEMMA 1. *If A2), A3) hold, then*

$$E\{\Phi_k^\tau \Phi_k\} \underset{k\to\infty}{\longrightarrow} H^\tau (I_{\frac{p(p-1)}{2}\times\frac{p(p-1)}{2}}) \otimes Q)H \triangleq C,$$

where C is a $p(L+1)\times p(L+1)$-matrix, Q is given in A3), $\otimes$ denotes the Kronecker product and $H = (H_0, \cdots, H_L)$ with

$$H = \begin{pmatrix} H_t^{(2)} & -H_t^{(1)} & 0 & 0 & 0 \\ H_t^{(3)} & 0 & -H_t^{(1)} & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ H_t^{(p)} & 0 & 0 & 0 & -H_t^{(1)} \\ 0 & H_t^{(3)} & -H_t^{(2)} & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & H_t^{(p)} & 0 & 0 & -H_t^{(2)} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & H_t^{(p)} & -H_t^{(p-1)} \end{pmatrix}$$

and $H_t^{(i)} = (\underbrace{0, \cdots, 0}_{t}, h_0^{(i)}, \cdots, h_L^{(i)}, \underbrace{0, \cdots, 0}_{L-t})^\tau$.

Proof. By the definition of Φ_k, we have

$$\Phi_k = (I_{\frac{p(p-1)}{2}\times\frac{p(p-1)}{2}}) \otimes (s_k, \cdots, s_{k-2L}))H. \tag{7}$$

Since

$$\begin{aligned}&((I_{\frac{p(p-1)}{2}\times\frac{p(p-1)}{2}}) \otimes (s_k, \cdots, s_{k-2L}))^\tau ((I_{\frac{p(p-1)}{2}\times\frac{p(p-1)}{2}}) \otimes (s_k, \cdots, s_{k-2L}))\\ &= (I_{\frac{p(p-1)}{2}\times\frac{p(p-1)}{2}}) \otimes ((s_k, \cdots, s_{k-2L})^\tau (s_k, \cdots, s_{k-2L})),\end{aligned}$$

and $E\{(s_k, \cdots, s_{k-2L})^\tau (s_k, \cdots, s_{k-2L})\} \underset{k\to\infty}{\longrightarrow} Q$ by A2), the lemma immediately follows. □

We now show that $\mathbf{h}^0$ is the unique eigenvector of C corresponding to the zero eigenvalue.

LEMMA 2. *Under A2)–A3) and A6),* $\mathbf{h}^0$ *is the unique (up-to a constant multiple) non-zero vector satisfying the following equations:*

$$\Phi_k \mathbf{h}^0 = 0,\ \forall k = 2L+1, \cdots ,$$

and

$$C\mathbf{h}^0 = 0.$$

Proof. By the definition of Φ_k the equations listed in the lemma are satisfied by $\mathbf{h}^0$. The only thing remains to prove is the uniqueness. Let $\hat{\mathbf{h}}$ be another solution of these linear equations. Then from that $C\hat{\mathbf{h}} = 0$ and $C = [H^\tau(I_{\frac{p(p-1)}{2}\times\frac{p(p-1)}{2}}) \otimes Q^{1/2}][H^\tau(I_{\frac{p(p-1)}{2}\times\frac{p(p-1)}{2}}) \otimes Q^{1/2}]^\tau$ it follows that

$$(I_{\frac{p(p-1)}{2}\times\frac{p(p-1)}{2}}) \otimes Q^{1/2})H\hat{\mathbf{h}} = 0.$$

Since $(I_{\frac{p(p-1)}{2}\times\frac{p(p-1)}{2}}) \otimes Q^{1/2}$ is nondegenerate, we have $H\hat{\mathbf{h}} = 0$. Then by A6) along with lines of [17] (See from (14) of [17] to the end of the proof for Theorem 1 of [17]) it is shown that $\hat{\mathbf{h}}$ is identical to $\mathbf{h}^0$ up-to a constant multiple. □

LEMMA 3. *Under Condition A4)*

$$D \triangleq E\{\Xi_k^\tau \Xi_k\} = (p-1)cI_{p(L+1)\times p(L+1)}.$$

Proof. By the definition of Ξ_k, it is seen that

$$\Xi_k^\tau \Xi_k = \begin{pmatrix} N_k^\tau N_k & \cdots & N_k^\tau N_{k-L} \\ \vdots & \ddots & \vdots \\ N_{k-L}^\tau N_k & \cdots & N_{k-L}^\tau N_{k-L} \end{pmatrix}.$$

By A4) it follows that $E\{N_n^\tau N_n\} = (p-1)cI_{p\times p}$, and $E\{N_m^\tau N_n\} = 0$ for $m \neq n$. Then the lemma follows immediately. □

We need a fact from stochastic approximation and formulate it as a lemma. For its proof we refer to [2].

LEMMA 4. *Let* $\{\mathcal{F}_k\}$ *be a family of nondecreasing* σ*-algebras and* $\{\varepsilon_k, \mathcal{F}_k\}$ *be martingale difference sequence with*

$$E\{\|\varepsilon_{k+1}\|^2|\mathcal{F}_k\} < \infty,\ E\{\varepsilon_{k+1}|\mathcal{F}_k\} = 0.$$

Let $\{\Theta_k, \mathcal{F}_k\}$ be an adapted random sequence and $\{c_k\}$ be a real sequence with $c_k > 0$, $\sum_k c_k = +\infty$ and $\sum |c_k|^2 < \infty$. Suppose that on $\Gamma \subset \Omega$, the following conditions 1,2 and 3 hold.

1. $$\limsup_{k\to\infty} E\{\|\varepsilon_{k+1}\|^2 \,|\mathcal{F}_k\} < \infty \quad \liminf_{k\to\infty} E\{\|\varepsilon_{k+1}\| \,|\mathcal{F}_k\} > 0; \tag{8}$$

2. *Θ_k can be decomposed into two adapted sequences $\{r_k, \mathcal{F}_k\}$ and $\{R_k, \mathcal{F}_k\}$ such that $\Theta_k = r_k + R_k$ and*

$$\sum_k \|r_k\|^2 < \infty \; and \; E\{I_\Gamma \sum_{k=n}^{\infty} \|c_k R_k\|\} = o\left(\sum_{k=n}^{\infty} |c_k|^2\right)^{1/2} \; as \; n \to \infty. \tag{9}$$

3. *$\sum_{k=n}^{\infty} c_k(\Theta_k + \varepsilon_k)$ coincides with an $\mathcal{F}_n$-measurable random variable for some n.*

Then $P\{\Gamma\} = 0$.

The following lemma shows a general property for a $\phi-$mixing sequence.

LEMMA 5. *Let $g(\cdot)$ be a measurable function such that $\|g(\mathbf{s}_k)\| \leqslant a\,\|\mathbf{s}_k\|^2$ where a is a constant. If Conditions A1), A2), A3) and the condition on $\mathbf{s}_k$ in A5) are satisfied, then*

$$\left\| E\{g(\mathbf{s}_k)|\mathcal{F}_1^{k-j}\} - E\{g(\mathbf{s}_k)\}\right\| \leqslant \chi(\omega)\,(\phi(k-L-j))^{\frac{\gamma}{2+\gamma}},$$

where $\chi(\omega) < \infty$.

Proof. According to the notation introduced in A1) $\mathcal{F}_1^n = \sigma\{s_k,\ k = 1, \cdots, n\}$. Denote by $F_k(z, \mathcal{F}_1^{k-j})$ the conditional distribution function of $\mathbf{s}_k$ given $\mathcal{F}_1^{k-j}$, where $k - L > j$, and by $F_k(z)$ the distribution function of $\mathbf{s}_k$. Then by the Jordan-Hahn decomposition for the signed measure

$$dG_{k,j}(z,\omega) \triangleq dF_k(z, \mathcal{F}_1^{k-j}) - dF_k(z),$$

there is a Borel set $U \in \mathbb{R}^{2L+1}$ such that for any Borel set $V \in \mathbb{R}^{2L+1}$

$$\int_V dG_{k,j}^+(z,w) = \int_{V\cap U^c} dG_{k,j}(z,w) \leqslant \phi(k-L-j),$$
$$\int_V dG_{k,j}^-(z,w) = \int_{V\cap U} dG_{k,j}(z,w) \leqslant \phi(k-L-j)$$

and

$$dG_{k,j}(z,\omega) = dG_{k,j}^+(z,\omega) - dG_{k,j}^-(z,\omega).$$

Therefore, by the Hölder inequality

$$\begin{aligned}
&\left\| E\{g(\mathbf{s}_k)|\mathcal{F}_1^{k-j}\} - E\{g(\mathbf{s}_k)\}\right\| = \left\| \int_{-\infty}^{\infty} g(z) dF_k(z, \mathcal{F}_1^{k-j}) - \int_{-\infty}^{\infty} g(z) dF_k(z) \right\| \\
&\leqslant \left\| \int_{-\infty}^{\infty} \|g(z)\| \, (dG_{k,j}^{+}(z,\omega) + dG_{k,j}^{-}(z,\omega)) \right\| \\
&\leqslant \left(\left(\int_{-\infty}^{\infty} \|g(z)\|^{1+\frac{\gamma}{2}} \, dG_{k,j}^{+}(z,\omega) \right)^{\frac{2}{2+\gamma}} \right. \\
&\left. + \left(\int_{-\infty}^{\infty} \|g(z)\|^{1+\frac{\gamma}{2}} \, dG_{k,j}^{-}(z,\omega) \right)^{\frac{2}{2+\gamma}} \right) \phi^{\frac{\gamma}{2+\gamma}}(k-L-j).
\end{aligned}$$

Since $\sup_k |s_k(\omega)| \leqslant \zeta(\omega) < \infty$ and $E\zeta^{2+\gamma} < \infty$, the integrals in the last expression are finite a.s. Denoting the sum of two integrals by $\chi(\omega)$ leads to the desired result. □

4. Main Results

We now in a position to formulate and prove the main results for the algorithm defined by (4) (5).

THEOREM 1. *Under A1)–A7), for any given initial* $\mathbf{h}_0$ *the distance between* $\mathbf{h}_k$ *and* J *converges to zero, i.e.*

$$d(\mathbf{h}_k, J) \underset{k\to\infty}{\longrightarrow} 0,$$

where J *is the set of unit eigenvectors of the matrix* C *defined in Lemma 1.*

Proof. To prove the theorem, by Theorem 2 in [19] or Theorem 5.2.1 in [20], we need only to prove that for any $t \in [0, T]$

$$\lim_{T\to 0} \limsup_{n\to\infty} \frac{1}{T} \left\| \sum_{k=n}^{m(n,t)} a_k(\Psi_{k+1}^{\tau}\Psi_{k+1} - B) \right\| = 0 \ a.s., \tag{10}$$

where $B = C + D$ and $m(n,t) = \max\{k : \sum_{i=n}^{k} a_i \leqslant t\}$. This is because any eigenvector of B is also an eigenvector of C.

By (3) we have

$$\begin{aligned}&\left\|\sum_{k=n}^{m(n,t)} a_k(\Psi_{k+1}^\tau \Psi_{k+1} - B)\right\| \\ &\leqslant \left\|\sum_{k=n}^{m(n,t)} a_k(\Phi_{k+1}^\tau \Phi_{k+1} - C)\right\| + \left\|\sum_{k=n}^{m(n,t)} a_k(\Xi_{k+1}^\tau \Xi_{k+1} - E\{\Xi_{k+1}^\tau \Xi_{k+1}\})\right\| \\ &+ \left\|\sum_{k=n}^{m(n,t)} a_k(\Phi_{k+1}^\tau \Xi_{k+1} + \Xi_{k+1}^\tau \Phi_{k+1})\right\|. \qquad (11)\end{aligned}$$

By A4), A7) and the convergence theorem of martingale difference sequence it is seen that the last two terms in (11) are of $o(T)$ as $T \to 0$. Therefore, it remains to prove that

$$\limsup_{n\to\infty} \left\|\sum_{k=n}^{m(n,t)} a_k(\Phi_{k+1}^\tau \Phi_{k+1} - C)\right\| = o(T).$$

Note that

$$\begin{aligned}&\left\|\sum_{k=n}^{m(n,t)} a_k(\Phi_{k+1}^\tau \Phi_{k+1} - C)\right\| \\ &\leqslant \left\|\sum_{k=n}^{m(n,t)} a_k(\Phi_{k+1}^\tau \Phi_{k+1} - E\{\Phi_{k+1}^\tau \Phi_{k+1} | \mathcal{F}_1^{k-j}\})\right\| \\ &+ \left\|\sum_{k=n}^{m(n,t)} a_k(E\{\Phi_{k+1}^\tau \Phi_{k+1} | \mathcal{F}_1^{k-j}\} - C)\right\|, \qquad (12)\end{aligned}$$

and for any $j > 0$, $\{\Phi_{k+1}^\tau \Phi_{k+1} - E\{\Phi_{k+1}^\tau \Phi_{k+1} | \mathcal{F}_1^{k-j}\}\}$ is a sum of j martingale difference sequences. By the convergent theorem for martingale difference sequence, from A5) and A7), it follows that for any $j > 0$,

$$\sum_{k=L}^{\infty} a_k(\Phi_{k+1}^\tau \Phi_{k+1} - E\{\Phi_{k+1}^\tau \Phi_{k+1} | \mathcal{F}_1^{k-j}\}) < \infty. \text{ a.s.}$$

The second term of the right hand side of (12) is less than the sum of the following two terms

$$\begin{aligned}&\left\|\sum_{k=n}^{m(n,t)} a_k(E\{\Phi_{k+1}^\tau \Phi_{k+1} | \mathcal{F}_1^{k-j}\} - E\{\Phi_{k+1}^\tau \Phi_{k+1}\})\right\| \\ &+ \left\|\sum_{k=n}^{m(n,t)} a_k(E\{\Phi_{k+1}^\tau \Phi_{k+1}\} - C)\right\|.\end{aligned}$$

By Lemma 5 the first term is less than $\tilde{\chi}(\omega)T\phi^{\frac{\gamma}{2+\gamma}}(n-L-j)$, where $\tilde{\chi}(\omega)<\infty$, while the second term is $o(T)$ by Lemma 1. Combining all of these leads to the desired result. □

By Lemma 2 zero is an eigenvalue of C with multiplicity one and the corresponding eigenvector is $\mathbf{h}^0$, $\mathbf{h}^0/\left\|\mathbf{h}^0\right\|\in J$. Theorem 1 guarantees that estimate $\mathbf{h}_k$ approaches to J, but it is not clear if $\mathbf{h}_k$ tends to the direction of $\mathbf{h}^0$. Let $0=\lambda_1<\lambda_2<\cdots<\lambda_m$, $m\leqslant p(L+1)$ be all different eigenvalues of C. J is composed of disconnected sets $J_s=\{\mathbf{h}\in\mathbb{R}^p,\ \|\mathbf{h}\|=1 \text{ and } C\mathbf{h}=\lambda_s\mathbf{h}\}$, $s=1,\cdots,m$, where $J_1=\{\mathbf{h}^0/\|\mathbf{h}^0\|,-\mathbf{h}^0/\|\mathbf{h}^0\|\}$. Note that the limit points of $\mathbf{h}_k$ are in a connected set, so h_k converges to a J_s for some s. Let $\Gamma_s=\{\omega, d(\mathbf{h}_k(\omega),J_s)\underset{k\to\infty}{\longrightarrow}0\}$. We want to prove that $d(\mathbf{h}_k,J_1)\underset{k\to\infty}{\longrightarrow}0$ a.s. or $P\{\Gamma_1\}=1$.

THEOREM 2. *Assume A1)-A7) hold. Then* $\mathbf{h}_k$ *defined by (4) (5) a.s. converges to* $\mathbf{h}^0$ *up-to a constant multiple:*

$$\mathbf{h}_k\to\alpha\mathbf{h}^0,$$

where α *equals either* $\left\|\mathbf{h}^0\right\|^{-1}$ *or* $-\left\|\mathbf{h}^0\right\|^{-1}$.

Proof. Assume the contrary, that $P\{\Gamma_s\}>0$ for some $s>1$, $\lambda_s>0$. Since C is a symmetric matrix, $\mathbf{h}^{0\tau}\mathbf{h}_k\to0$ for $\omega\in\Gamma_s$, where and hereafter a possible set with zero probability in Γ_s is ignored.

Expanding $\mathbf{h}_{k+1}$ defined by (5) to the Taylor's series with respect to a_k, we derive

$$\mathbf{h}_{k+1}=\mathbf{h}_k-a_k(B\mathbf{h}_k-(\mathbf{h}_k^\tau\Psi_{k+1}^\tau\Psi_{k+1}\mathbf{h}_k)\mathbf{h}_k+\mu_{k+1}+\beta_{k+1}), \tag{13}$$

where

$$\mu_{k+1}=(\Psi_{k+1}^\tau\Psi_{k+1}-B)\mathbf{h}_k, \tag{14}$$

$$\beta_{k+1}=O(a_k). \tag{15}$$

Defining $\theta_k=\mathbf{h}^{0\tau}\mathbf{h}_k$ and noting $\mathbf{h}^{0\tau}C=0$ and $\mathbf{h}^{0\tau}\Phi_{k+1}=0$, we derive

$$\theta_{k+1}=\theta_k+a_k((\mathbf{h}_k^\tau\Psi_{k+1}^\tau\Psi_{k+1}\mathbf{h}_k-(p-1)c)\theta_k-\mathbf{h}^{0\tau}\mu_{k+1}-\mathbf{h}^{0\tau}\beta_{k+1}), \tag{16}$$

and

$$\begin{aligned}
\mathbf{h}^{0\tau}\mu_{k+1}&=\mathbf{h}^{0\tau}(\Psi_{k+1}^\tau\Psi_{k+1}-B)\mathbf{h}_k\\
&=\mathbf{h}^{0\tau}(\Xi_{k+1}^\tau\Xi_{k+1}+\Xi_{k+1}^\tau\Phi_{k+1}^\tau)\mathbf{h}_k-(p-1)c\theta_k\\
&=\left(\sum_{i=0}^{L}h_i^\tau N_{k+1-i}^\tau\right)\left(\sum_{i=0}^{L}N_{k+1-i}h_{k,i}\right)\\
&\quad+\left(\sum_{i=0}^{L}h_i^\tau N_{k+1-i}^\tau\right)\left(\sum_{i=0}^{L}X_{k+1-i}h_{k,i}\right)-(p-1)c\theta_k.
\end{aligned}$$

By A4) and A5), there exists $\alpha(\omega) < \infty$ a.s. such that $\|\Psi_{k+1}^{\tau}\Psi_{k+1} - D\| < \alpha(\omega)$ a.s. For any integers m and n define $\Gamma_m = \{\omega, \alpha(\omega) < m\} \cap \Gamma_s$ and

$$B_n = \prod_{k=n_0}^{n} \{1 + a_k(\mathbf{h}_k^{\tau}(\Psi_{k+1}^{\tau}\Psi_{k+1} - D)\mathbf{h}_k)\}. \tag{17}$$

Note that for $\omega \in \Gamma_m$,

$$\mathbf{h}_k^{\tau} C \mathbf{h}_k \to \lambda_s > 0,$$

and by the convergence of $\mathbf{h}_k$ from (13) it follows that $\|\mathbf{h}_j - \mathbf{h}_k\| < c_0 T,\ \forall j : k \leqslant j \leqslant m(k, T)$ where c_0 is a constant for all ω in Γ_m. By (10) we then have

$$\begin{aligned}
&\left| \sum_{k=j}^{m(j,T)} a_k(\mathbf{h}_k^{\tau}(\Psi_{k+1}^{\tau}\Psi_{k+1} - B)\mathbf{h}_k) \right| \\
&\leqslant \left\| \sum_{k=j}^{m(j,T)} a_k(\Psi_{k+1}^{\tau}\Psi_{k+1} - B) \right\| + 2c_0 T m \sum_{k=j}^{m(j,T)} a_k = o(T).
\end{aligned}$$

Choose large enough n_0 and sufficiently small T such that $o(T)/T < \lambda_s/4, \forall j \geqslant n_0$. Let $k_0 = n_0$, $k_1 = m(n_0, T) + 1$, $k_2 = m(k_1, T) + 1, \cdots, k_{j+1} = m(k_j, T) + 1, \cdots$, and $m(k_l, T) \leqslant n \leqslant m(k_{l+1}, T)$. It then follows that for $\omega \in \Gamma_m$

$$\begin{aligned}
\ln\ B_n &= \ln \left\{ \prod_{k=n_0}^{n} \{1 + a_k(\mathbf{h}_k^{\tau}(\Psi_{k+1}^{\tau}\Psi_{k+1} - D)\mathbf{h}_k\} \right\} \\
&= \sum_{k=n_0}^{n} a_k(\mathbf{h}_k^{\tau}(\Psi_{k+1}^{\tau}\Psi_{k+1} - D)\mathbf{h}_k) + O\left(\sum_{k=n_0}^{n} a_k^2\right) \\
&= \sum_{k=n_0}^{n} \mathbf{h}_k^{\tau} C \mathbf{h}_k a_k + \sum_{k=n_0}^{n} a_k(\mathbf{h}_k^{\tau}(\Psi_{k+1}^{\tau}\Psi_{k+1} - B)\mathbf{h}_k) + O\left(\sum_{k=n_0}^{n} a_k^2\right) \\
&\geqslant \sum_{j=0}^{l} \sum_{k=k_j}^{m(k_j,T)} \frac{\lambda_s}{2} a_k > \frac{\lambda_s}{3} \sum_{k=n_0}^{n} a_k
\end{aligned} \tag{18}$$

for n_0 sufficiently large.

Consequently, for $\omega \in \Gamma_m$ with fixed m

$$B_n \geqslant e^{\frac{\lambda_s}{3}\sum_{k=n_0}^{n} a_k} \tag{19}$$

and hence

$$B_n / \left(\sum_{k=n_0}^{n} a_k \right)^2 \to \infty. \tag{20}$$

Define

$$\Lambda_l = \left\{ \omega, B_n > \left(\sum_{k=1}^{n} a_k \right)^2, \forall n \geqslant l \right\}.$$

From (16) it follows that

$$\theta_n = B_{n-1}(\theta_{n_0} - \sum_{j=n_0}^{n-1} B_j^{-1} a_j (\mathbf{h}^{0\tau} \mu_{j+1} + \mathbf{h}^{0\tau} \beta_{j+1})). \tag{21}$$

Tending $n \to \infty$ in (21) and replacing n_0 by n in the resulting equality, by (19) we have

$$\theta_n = \sum_{j=n}^{\infty} B_j^{-1} a_j (\mathbf{h}^{0\tau} \mu_{j+1} + \mathbf{h}^{0\tau} \beta_{j+1}), \quad \forall n, \ \omega \in \Gamma_m \cap \Lambda_l. \tag{22}$$

Let $\mathcal{F}_k = \sigma\{\xi_l, l = 0, \cdots, k, s_l, l = 0, \cdots, k + 2L + 1\}$. We intend to show that θ_n given by (22) can be expressed in the form of condition 3 in Lemma 4. If this can be done, then noticing that by (21) θ_n is $\mathcal{F}_n$-measurable, by Lemma 4 it follows that $P\{\bigcup_{m,l}(\Gamma_m \cap \Lambda_l)\} = 0$ or $P\{\Gamma_s\} = 0, \ \forall s > 1$ and the theorem will be proved.

We first show that the series

$$S_n \triangleq \sum_{k=n}^{\infty} a_k (\mathbf{h}^{0\tau} \mu_{k+1} + \mathbf{h}^{0\tau} \beta_{k+1}) \tag{23}$$

is convergent on Γ_s. By (15) and A7) it suffices to show $\sum_{k=n}^{\infty} a_k / \mu_{k+1}$ is convergent on Γ_s.

Define

$$\varepsilon_{k+1}^{(1)} = \sum_{i=0}^{L} (h_i^\tau N_{k+1}^\tau)(N_{k+1} h_{k,i}) - (p-1) c \theta_k, \tag{24}$$

$$\varepsilon_{k+1}^{(2)} = \sum_{i=0}^{L-1} \left[(h_i^\tau N_{k+1}^\tau) \left(\sum_{l=i+1}^{L} N_{k+i+1-l} h_{k,l} \right) + \left(\sum_{l=i+1}^{L-1} h_l^\tau N_{k+i+1-l}^\tau \right) (N_{k+1} h_{k,i}) \right], \tag{25}$$

$$\varepsilon_{k+1}^{(3)} = \sum_{i=0}^{L-1} \left[(h_i^\tau N_{k+1}^\tau) \left(\sum_{l=0}^{L} N_{k+i+1-l} h_{k,l} \right) \right], \tag{26}$$

and

$$\delta_{k+1} = \sum_{i=1}^{3} \varepsilon_{k+1}^{(i)}. \tag{27}$$

Clearly, δ_k, is measurable with respect to $\mathcal{F}_k$ and $E\{\delta_{k+1}|\mathcal{F}_k\} = 0$. Then by the convergence theorem for martingale difference sequences

$$\sum_{k=m}^{\infty} a_k \delta_{k+1} < \infty. \tag{28}$$

By (2), (3) and (14) it follows that

$$\begin{aligned}
&\sum_{k=n}^{\infty} a_k[\mathbf{h}^{0\tau}\mu_{k+1} + (p-1)c\theta_k] \\
&= \sum_{k=n}^{\infty} a_k \left[\sum_{i=0}^{L} (h_i^\tau N_{k+1-i}^\tau) \left(\sum_{s=0}^{L} N_{k+1-s} h_{k,s} \right) \right. \\
&\quad \left. + \left(\sum_{i=0}^{L} h_i^\tau N_{k+1-i}^\tau \right) \left(\sum_{s=0}^{L} X_{k+1-s} h_{k,s} \right) \right] \\
&= \sum_{i=0}^{L} \sum_{k=n}^{\infty} \left[a_k h_i^\tau N_{k+1-i}^\tau \left(\sum_{s=0}^{L} N_{k+1-s} h_{k,s} \right) \right. \\
&\quad \left. + a_k h_i^\tau N_{k+1-i}^\tau \left(\sum_{s=0}^{L} X_{k+1-s} h_{k,s} \right) \right] \\
&= \sum_{i=0}^{L} \sum_{l=n-i}^{\infty} \left[a_{l+i} h_i^\tau N_{l+1}^\tau \left(\sum_{s=0}^{L} N_{l+i+1-s} h_{l+i,s} \right) \right. \\
&\quad \left. + a_{l+i} h_i^\tau N_{l+1}^\tau \left(\sum_{s=0}^{L} X_{l+i+1-s} h_{l+i,s} \right) \right].
\end{aligned} \tag{29}$$

The first term on the right-hand side of the last equality of (29) can be expressed in the following form:

$$\begin{aligned}
&\sum_{i=0}^{L} \sum_{l=n-i}^{\infty} a_{l+i} (h_i^\tau N_{l+1}^\tau)(N_{l+1} h_{l+i,i}) \\
&\quad + \sum_{i=0}^{L-1} \sum_{l=n-i}^{\infty} a_{l+i} (h_i^\tau N_{l+1}^\tau) \left(\sum_{s=i+1}^{L} N_{l+i+1-s} h_{l+i,s} \right) \\
&\quad + \sum_{i=1}^{L} \sum_{l=n-i}^{\infty} a_{l+i} (h_i^\tau N_{l+1}^\tau) \left(\sum_{s=0}^{i-1} N_{l+i+1-s} h_{l+i,s} \right),
\end{aligned} \tag{30}$$

where the last term equals

$$\begin{aligned}
&\sum_{s=0}^{L-1} \sum_{i=s+1}^{L-1} \sum_{l=n-i}^{\infty} a_{l+i} (h_i^\tau N_{l+1}^\tau)(N_{l+1+i-s} h_{l+i,s}) \\
&= \sum_{s=0}^{L-1} \sum_{i=s+1}^{L-1} \sum_{m=n-s}^{\infty} a_{m+s} (h_i^\tau N_{m-i+s+1}^\tau)(N_{m+1} h_{m+s,s}).
\end{aligned} \tag{31}$$

Combining (30) and (31) we derive that the first term on the right-hand side of the last equality of (29) is

$$\begin{aligned}&\sum_{i=0}^{L}\sum_{l=n-i}^{\infty} a_{l+i}(h_i^{\tau}N_{l+1}^{\tau})(N_{l+1}h_{l+i,i})\\&+\sum_{i=0}^{L-1}\sum_{l=n-i}^{\infty}\sum_{s=i+1}^{L-1} a_{l+i}[(h_i^{\tau}N_{l+1}^{\tau})(N_{l+i+1-s}h_{l+i,s})+(h_s^{\tau}N_{l+i+1-s}^{\tau})(N_{l+1}h_{l+i,i})].\end{aligned} \tag{32}$$

By A4), A5) and A7) it is clear that $\|\mathbf{h}_{k+l}-\mathbf{h}_k\| = O(a_k)$, $\forall l, 0 \leqslant l \leqslant L$. Hence replacing h_{l+i} by h_l in (29) results in producing an additional term of magnitude $O(a_j)$. Thus, by (24)–(26) we can rewrite (29) as

$$\sum_{k=n}^{\infty} a_k \mathbf{h}^{0\tau}\mu_{k+1} = \sum_{k=n}^{\infty} a_k\left(\sum_{i=1}^{3}\varepsilon_{k+1}^{(i)} + \nu_{k+1}\right) = \sum_{k=n}^{\infty} a_k(\delta_{k+1}+\nu_{k+1}), \tag{33}$$

where $\nu_{k+1} = O(a_{k+1})$ and is $\mathcal{F}_{k+1}$-measurable. By (28) and A7) the series (33) is convergent, and hence S_n given by (23) is a convergent series.

Let $B_{n-1} = I$. We now have

$$\begin{aligned}\theta_n &= \sum_{k=n}^{\infty} B_k^{-1}(S_k - S_{k+1}) = \sum_{k=n}^{\infty}(B_k^{-1}-B_{k-1}^{-1})S_k + S_{n_0}\\&= \sum_{k=n}^{\infty}\left[(B_k^{-1}-B_{k-1}^{-1})S_k + a_k(\mathbf{h}^{0\tau}\mu_{k+1}+\mathbf{h}^{0\tau}\beta_{k+1})\right]\\&= \sum_{j=0}^{\infty}\left[\sum_{l=j(L+1)+n}^{(j+1)(L+1)+n-1} R_l^1 + a_{j(L+1)+n}\sum_{l=j(L+1)+n}^{(j+1)(L+1)+n-1}(\delta_{l+1}+\tilde{\nu}_{l+1}+\mathbf{h}^{0\tau}\beta_{l+1})\right],\end{aligned}$$

where $R_j^1 = (B_j^{-1} - B_{j-1}^{-1})S_j$,

$$\begin{aligned}\tilde{\nu}_{l+1} &= \left(\frac{a_l}{a_{j(L+1)+n}} - 1\right)(\delta_{l+1}+\nu_{l+1}+\mathbf{h}^{0\tau}\beta_{l+1}) + \nu_{l+1}\\&= O(a_{j(L+1)+n}), \quad \forall l : j(L+1)+n \leqslant l < (j+1)(L+1)+n.\end{aligned}$$

Denote

$$R_j = \frac{1}{c_j}\sum_{l=j(L+1)+n}^{(j+1)(L+1)+n-1} R_l^1, \quad r_j = \sum_{l=j(L+1)+n}^{(j+1)(L+1)+n-1}(\tilde{\nu}_{l+1}+\mathbf{h}^{0\tau}\beta_{l+1}),$$

$$\varepsilon_j = \sum_{l=j(L+1)+n}^{(j+1)(L+1)+n-1}\delta_{l+1}, \quad c_j = a_{j(L+1)+n} \text{ and } \mathcal{F}_j' \triangleq \mathcal{F}_{(j+1)(L+1)+n}.$$

Then $\{R_j, \mathcal{F}'_j\}$, $\{r_j, \mathcal{F}'_j\}$ are adapted sequences and $\{\varepsilon_j, \mathcal{F}'_j\}$ is a martingale difference sequence, and θ_n is written in the form of Lemma 4: $\theta_n = \sum_{j=n}^{\infty} c_j(R_j + r_j + \varepsilon_j)$.

It remains to verify (8) and (9).

From (23) and (33) it follows that there is a constant $\eta > 0$ such that $E\{|S_n|^2\} \leqslant \eta \sum_{k=n}^{\infty} a_k^2$. Then for $n > l$ noticing

$$|B_j^{-1} - B_{j-1}^{-1}| \leqslant B_j^{-1} a_j |(\mathbf{h}_j^{\tau} \Psi_{j+1}^{\tau} \Psi_{j+1} \mathbf{h}_j) - (p-1)c|,$$

and

$$\begin{aligned}\sum_{j=n}^{\infty} \left(E\left\{ I_{\Gamma_m \cap \Lambda_l} |B_j^{-1} - B_{j-1}^{-1}|^2 \right\} \right)^{1/2} &\leqslant \sum_{j=n}^{\infty} \left(E\left\{ 1_{\Gamma_m \cap \Lambda_l} B_j^{-2} a_j^2 (m)^2 \right\} \right)^{1/2} \\ &\leqslant m \sum_{j=n}^{\infty} \frac{a_j}{\left(\sum_{k=1}^{j} a_k \right)^2} \leqslant \int_{\sum_{k=1}^{n-1} a_k}^{\infty} \frac{1}{x^2} dx < \infty,\end{aligned}$$

we have

$$\begin{aligned}E\left\{ I_{\Gamma_m \cap \Lambda_l} \sum_{j=n}^{\infty} |c_j R_j| \right\} &\leqslant E\left\{ I_{\Gamma_m \cap \Lambda_l} \sum_{k=n}^{\infty} |R_k^1| \right\} \\ &= E\left\{ I_{\Gamma_m \cap \Lambda_l} \sum_{k=n}^{\infty} |B_k^{-1} - B_{k-1}^{-1}||S_k| \right\} \\ &\leqslant \sum_{k=n}^{\infty} \left(E\{I_{\Gamma_m \cap \Lambda_l} |B_k^{-1} - B_{k-1}^{-1}|^2\} E\{I_{\Gamma_m \cap \Lambda_l} |S_k|^2\} \right)^{1/2} \\ &\leqslant o\left(\sum_{k=n}^{\infty} a_k^2 \right)^{1/2} = o\left(\sum_{j=n}^{\infty} c_j^2 \right)^{1/2} \quad \text{as } n \to \infty.\end{aligned}$$

By A4) and A5) it follows that

$$\limsup_{k \to \infty} E\left\{ \delta_{k+1}^{2+\gamma} | \mathcal{F}_k \right\} < \infty \text{ for some } \gamma > 0. \tag{34}$$

It is proved in Lemma 6 in the Appendix that

$$\liminf_{k \to \infty} E\left\{ \left| \sum_{l=k}^{k+L} \delta_{l+1} \right|^2 | \mathcal{F}_k \right\} > 0,$$

which implies that

$$\liminf_{k \to \infty} E\left\{ |\varepsilon_{k+1}|^2 | \mathcal{F}'_k \right\} = \liminf_{k \to \infty} E\left\{ \left| \sum_{l=(k+1)(L+1)+n}^{(k+2)(L+1)+n} \delta_{l+1} \right|^2 | \mathcal{F}_{(k+1)(L+1)+n} \right\} > 0. \tag{35}$$

Then from the following inequality

$$E\{|\varepsilon_{k+1}|^2|\mathcal{F}'_k\} < E\{|\varepsilon_{k+1}|^{2+\gamma}|\mathcal{F}'_k\}^{\frac{1}{1+\gamma}} E\{|\varepsilon_{k+1}||\mathcal{F}'_k\}^{\frac{\gamma}{1+\gamma}}$$

by (34)(35) it follows that

$$\liminf_{k\to\infty} E|\varepsilon_{k+1}||\mathcal{F}'_k\} > 0.$$

Therefore all conditions required in Lemma 4 are met, and we conclude $P\{\Gamma_m \cap \Lambda_l\} = 0$. Since $\Gamma_s = \bigcup_{m,l} \Gamma_m \cap \Lambda_l$, it follows that $P\{\Gamma_s\} = 0$, and $\mathbf{h}_k$ must converge to $\alpha\mathbf{h}^0$, a.s. □

5. Conclusion Remarks

In this paper we have presented a recursive algorithm for channel identification. The algorithm is featured by the following points: 1) The algorithm is on-line updated needless to collect entire data in advance; 2) No noise statistics are used in the algorithm; 3) The input signal is allowed to be dependent; 4) The estimate is proved to converge a.s. to the true channel coefficients up-to a constant scaling factor.

For further study it is of interest to consider the case where the input signal is also multidimensional. It is also of interest to weaken conditions imposed on the input and on the observation noise.

Appendix

LEMMA 6. *Under Conditions A4), A5)*

$$\liminf_{n\to\infty} E\left\{\left|\sum_{l=k}^{k+L} \delta_{l+1}\right|^2 |\mathcal{F}_k\right\} > 0,$$

where δ_k *is given by (27) and* $\mathcal{F}_k = \sigma\{\xi_l, l = 0, \cdots, k, s_l, l = 0, \cdots, k+2L+1\}$.

Proof. If the lemma were not true, then there would exist a subsequence $\{k_n\}$ such that

$$E\left\{\left|\sum_{l=k_n}^{k_n+L} \delta_{l+1}\right|^2 |\mathcal{F}_{k_n}\right\} \underset{n\to\infty}{\longrightarrow} 0. \tag{36}$$

For notational simplicity, let us denote the subsequence k_n still by k.

Since by A5) $E\{\varepsilon^{(s)}_{j+1}\varepsilon^{(t)}_{i+1}|\mathcal{F}_k\} = 0$ for $j, i \geqslant k$ if $s \neq t$, and for any $j, i \geqslant k$ but $j \neq i$ if $s = t$, we then have

$$E\left\{\left|\sum_{l=k}^{k+L}\delta_{l+1}\right|^2 |\mathcal{F}_k\right\} = \sum_{l=k}^{k+L}\left[E\{(\varepsilon^{(1)}_{l+1})^2|\mathcal{F}_k\} + E\{(\varepsilon^{(2)}_{l+1})^2|\mathcal{F}_k\} + E\{(\varepsilon^{(3)}_{l+1})^2|\mathcal{F}_k\}\right]$$

which incorporating with (36) implies that

$$E\left\{\left(\varepsilon^{(1)}_{k+1}\right)^2 |\mathcal{F}_k\right\} \underset{k\to\infty}{\to} 0, \tag{37}$$

and

$$E\left\{\left(\varepsilon^{(2)}_{k+L+1}\right)^2 |\mathcal{F}_k\right\} \underset{k\to\infty}{\to} 0. \tag{38}$$

Noticing that $\theta_k \underset{k\to\infty}{\to} 0$ and $|h_{j-L,i} - h_{j,i}| = O(a_j)$, from (37) and (24) it follows that

$$E\left\{\left(\sum_{i=0}^{L}(h_i^\tau N_{k+1}^\tau)(N_{k+1}h_{k,i})\right)^2 |\mathcal{F}_k\right\} \underset{k\to\infty}{\to} 0.$$

On the other hand, we have

$$\begin{aligned}\sum_{i=0}^{L}(h_i^\tau N_{k+1}^\tau)(N_{k+1}h_{k,i}) &= \sum_{i=0}^{L}\sum_{n=1}^{p}\sum_{m=n+1}^{p}\left(h_i^{(n)}\xi^{(m)}_{k+1} - h_i^{(m)}\xi^{(n)}_{k+1}\right)\\ &\quad\times\left(h_{k,i}^{(n)}\xi^{(m)}_{k+1} - h_{k,i}^{(m)}\xi^{(n)}_{k+1}\right)\\ &= \sum_{i=0}^{L}\sum_{n=1}^{p}\sum_{\substack{m=1\\ m\neq n}}^{p}\left(h_i^{(n)}h_{k,i}^{(n)}\xi^{(m)}_{k+1}\xi^{(m)}_{k+1}\right.\\ &\quad\left.-h_i^{(m)}h_{k,i}^{(n)}\xi^{(n)}_{k+1}\xi^{(m)}_{k+1}\right),\end{aligned}$$

and hence,

$$E\left\{\left[\sum_{i=0}^{L}\sum_{m=1}^{p}\sum_{\substack{n=1\\ n\neq m}}^{p}\left(h_i^{(n)}h_{k,i}^{(n)}\xi^{(m)}_{k+1}\xi^{(m)}_{k+1} - h_i^{(m)}h_{k,i}^{(n)}\xi^{(n)}_{k+1}\xi^{(m)}_{k+1}\right)\right]^2 |\mathcal{F}_k\right\} \underset{k\to\infty}{\to} 0. \tag{39}$$

Since for any $s \neq t$,

$$E\left\{\left(h_i^{(n)}h_{j,i}^{(n)}\xi^{(m)}_{j+1}\xi^{(m)}_{j+1}\right)\left(h_i^{(t)}h_{j,i}^{(s)}\xi^{(s)}_{j+1}\xi^{(t)}_{j+1}\right)|\mathcal{F}_j\right\} = 0,$$

we have

$$E\left\{\left[\sum_{m=1}^{p}\sum_{\substack{n=1\\n\neq m}}^{p}\left(\left(\sum_{i=0}^{L}h_i^{(n)}h_{k,i}^{(n)}\right)\xi_{k+1}^{(m)}\xi_{k+1}^{(m)}-\left(\sum_{i=0}^{L}h_i^{(m)}h_{k,i}^{(n)}\right)\xi_{k+1}^{(n)}\xi_{k+1}^{(m)}\right)\right]^2|\mathcal{F}_k\right\}$$
$$=E\left\{\left[\sum_{m=1}^{p}\sum_{\substack{n=1\\n\neq m}}^{p}\left(\sum_{i=0}^{L}h_i^{(n)}h_{k,i}^{(n)}\right)\xi_{k+1}^{(m)}\xi_{k+1}^{(m)}\right]^2|\mathcal{F}_k\right\}$$
$$+E\left\{\left[\sum_{m=1}^{p}\sum_{\substack{n=1\\n\neq m}}^{p}\left(\sum_{i=0}^{L}h_i^{(m)}h_{k,i}^{(n)}\right)\xi_{k+1}^{(n)}\xi_{k+1}^{(m)}\right]^2|\mathcal{F}_k\right\}.$$

Hence (39) implies that

$$E\left\{\left[\sum_{m=1}^{p}\sum_{\substack{n=1\\n\neq m}}^{p}\left(\sum_{i=0}^{L}h_i^{(n)}h_{k,i}^{(n)}\right)\left(\xi_{k+1}^{(m)}\right)^2\right]^2|\mathcal{F}_k\right\},\underset{k\to\infty}{\longrightarrow}0. \tag{40}$$

and

$$E\left\{\left[\sum_{m=1}^{p}\sum_{\substack{n=1\\n\neq m}}^{p}\left(\sum_{i=0}^{L}h_i^{(m)}h_{k,i}^{(n)}\right)\left(\xi_{k+1}^{(n)}\xi_{k+1}^{(m)}\right)\right]^2|\mathcal{F}_k\right\}\underset{k\to\infty}{\longrightarrow}0. \tag{41}$$

By A4) the left hand side of (40) equals

$$E\left\{\left[\sum_{m=1}^{p}\left(\sum_{n=1}^{p}\sum_{i=0}^{L}h_i^{(n)}h_{k,i}^{(n)}-\sum_{i=0}^{L}h_i^{(m)}h_{k,i}^{(m)}\right)\left(\xi_{k+1}^{(m)}\right)^2\right]^2|\mathcal{F}_k\right\}$$
$$=E\left\{\left[\sum_{m=1}^{p}\left(\theta_k-\sum_{i=0}^{L}h_i^{(m)}h_{k,i}^{(m)}\right)\left(\xi_{k+1}^{(m)}\right)^2\right]^2|\mathcal{F}_k\right\}$$
$$=\sum_{m=1}^{p}\left(\theta_k-\sum_{i=0}^{L}h_i^{(m)}h_{k,i}^{(m)}\right)^2E\left\{\left(\left(\xi_{k+1}^{(m)}\right)^2-c\right)^2\right\}$$
$$+c^2\left(\sum_{m=1}^{p}\left(\theta_k-\sum_{i=0}^{L}h_i^{(m)}h_{k,i}^{(m)}\right)\right)^2$$
$$=\sum_{m=1}^{p}\left(\theta_k-\sum_{i=0}^{L}h_i^{(m)}h_{k,i}^{(m)}\right)^2E\left\{\left(\left(\xi_{k+1}^{(m)}\right)^2-c\right)^2\right\}+(p-1)^2c^2\theta_k^2.$$

Since $\theta_k \to 0$, it follows that for any m,

$$\sum_{i=0}^{L} h_i^{(m)} h_{k,i}^{(m)} \underset{k\to\infty}{\to} 0. \tag{42}$$

The left side of (41) equals

$$E\left\{\sum_{m=1}^{p}\sum_{\substack{n=1\\n\neq m}}^{p}\left(\sum_{i=0}^{L}\left(h_i^{(m)}h_{k,i}^{(n)}+h_i^{(n)}h_{k,i}^{(m)}\right)\right)^2\left(\xi_{k+1}^{(n)}\xi_{k+1}^{(m)}\right)^2|\mathcal{F}_k\right\}$$
$$= c^2\sum_{m=1}^{p}\sum_{n=1}^{m-1}\left(\sum_{i=0}^{L}h_i^{(m)}h_{k,i}^{(n)}+h_i^{(n)}h_{k,i}^{(m)}\right)^2.$$

Thus (41) implies that for any $m \neq n$,

$$\sum_{i=0}^{L}\left(h_i^{(m)}h_{k,i}^{(n)}+h_i^{(n)}h_{k,i}^{(m)}\right)\underset{k\to\infty}{\to} 0. \tag{43}$$

Noticing $|\mathbf{h}_{k+i}-\mathbf{h}_k| = O(a_k)\ \forall i = 1,\cdots,L$ from (25) we have

$$\begin{aligned}\varepsilon_{k+L+1}^{(2)} &= \sum_{i=0}^{L-1}\sum_{l=0}^{L-i-1}\left[\left(h_i^\tau N_{k+L+1}^\tau\right)\left(N_{k+L-l}h_{k,l+i+1}\right)\right.\\ &\quad\left.+\left(h_{l+i+1}^\tau N_{k+L-l}^\tau\right)\left(N_{k+1}h_{k,i}\right)\right]\\ &= \sum_{l=0}^{L-1}\sum_{i=0}^{L-l-1}\left[\left(h_i^\tau N_{k+L+1}^\tau\right)\left(N_{k+L-l}h_{k,l+i+1}\right)\right.\\ &\quad\left.+\left(h_{l+i+1}^\tau N_{k+L-l}^\tau\right)\left(N_{k+L+1}h_{k,i}\right)\right]+O(a_k).\end{aligned}$$

Then by A5) (38) implies that for any l

$$\begin{aligned}E\left\{\left[\sum_{i=0}^{L-l-1}\left[\left(h_i^\tau N_{k+L+1}^\tau\right)\left(N_{k+L-l}h_{k,l+i+1}\right)\right.\right.\right.\\ \left.\left.\left.+\left(h_{l+i+1}^\tau N_{k+L-l}^\tau\right)\left(N_{k+L+1}h_{k,i}\right)\right]\right]|\mathcal{F}_k\right\}\underset{k\to\infty}{\to}0.\end{aligned} \tag{44}$$

Notice that

$$\begin{aligned}&\left(h_i^\tau N_{k+L+1}^\tau\right)\left(N_{k+L-l}h_{k,l+i+1}\right)\\ &=\sum_{m=1}^{p}\sum_{n=1}^{m-1}\left(h_i^{(n)}\xi_{k+L+1}^{(m)}-h_i^{(m)}\xi_{k+L+1}^{(n)}\right)\left(h_{k,l+i+1}^{(n)}\xi_{k+L-l}^{(m)}-h_{k,l+i+1}^{(m)}\xi_{k+L-l}^{(n)}\right)\\ &=\sum_{m=1}^{p}\sum_{\substack{n=1\\n\neq m}}^{p}\left(h_i^{(n)}h_{k,l+i+1}^{(n)}\xi_{k+L+1}^{(m)}\xi_{k+L-l}^{(m)}-h_i^{(n)}h_{k,l+i-1}^{(m)}\xi_{k+L+1}^{(m)}\xi_{k+L-l}^{(n)}\right),\end{aligned} \tag{45}$$

and

$$\left(h_{l+i+1}^{\tau} N_{k+L-l}^{\tau}\right)\left(N_{k+L+1} h_{k,i}\right) = \sum_{m=1}^{p} \sum_{\substack{n=1 \\ n \neq m}}^{p} \Big(h_{l+i+1}^{(n)} h_{k,i}^{(n)} \xi_{k+1}^{(m)} \xi_{k+L-l}^{(m)} - h_{l+i+1}^{(m)} h_{k,i}^{(n)} \xi_{k+L+1}^{(m)} \xi_{k+L-l}^{(n)} \Big). \tag{46}$$

Then by A5), from (44)–(46) it follows that

$$E\left\{\left[\sum_{i=0}^{L-l-1}\left[\left(h_i^{\tau} N_{k+L+1}^{\tau}\right)\left(N_{k+L-l} h_{k,l}\right)+\left(h_{l+i+1}^{\tau} N_{k+L-l}^{\tau}\right)\left(N_{k+L+1} h_{k,i}\right)\right]\right]^2\right\}$$
$$= c^2 \sum_{m=1}^{p}\left[\left(\sum_{i=0}^{L-l-1} \sum_{\substack{n=1 \\ n \neq m}}^{p}\left(h_i^{(n)} h_{k,l+i+1}^{(n)}+h_{l+i+1}^{(n)} h_{k,i}^{(n)}\right)\right)^2 + \sum_{\substack{n=1 \\ n \neq m}}^{p}\left(\sum_{i=0}^{L-l-1}\left(h_i^{(m)} h_{k,l+i+1}^{(n)}+h_{l+i+1}^{(n)} h_{k,i}^{(m)}\right)\right)^2\right] \underset{k\to\infty}{\longrightarrow} 0,$$

and hence for any $l = 0, \cdots, L-1$

$$\sum_{i=0}^{L-l-1}\left(h_i^{(m)} h_{k,l+i+1}^{(n)}+h_{l+i+1}^{(n)} h_{k,i}^{(m)}\right) \underset{k\to\infty}{\longrightarrow} 0 \tag{47}$$

and

$$\sum_{i=0}^{L-l-1} \sum_{\substack{n=1 \\ n \neq m}}^{p}\left(h_i^{(n)} h_{k,l+i+1}^{(n)}+h_{l+i+1}^{(n)} h_{k,i}^{(n)}\right) \underset{k\to\infty}{\longrightarrow} 0. \tag{48}$$

Notice that (48) means that

$$p \sum_{n=1}^{p} \sum_{i=0}^{L-l-1}\left(h_i^{(n)} h_{k,l+i+1}^{(n)}+h_{l+i+1}^{(n)} h_{k,i}^{(n)}\right) - \sum_{m=1}^{p} \sum_{i=0}^{L-l-1}\left(h_i^{(m)} h_{k,l+i+1}^{(m)}+h_{l+i+1}^{(m)} h_{k,i}^{(m)}\right) \underset{k\to\infty}{\longrightarrow} 0.$$

However, the above expression equals

$$(p-1) \sum_{n=1}^{p} \sum_{i=0}^{L-l-1}\left(h_i^{(n)} h_{k,l+i+1}^{(n)}+h_{l+i+1}^{(n)} h_{k,i}^{(n)}\right).$$

Therefore

$$\sum_{i=0}^{L-l-1} \left(h_i^{(n)} h_{k,l+i+1}^{(n)} + h_{l+i+1}^{(n)} h_{k,i}^{(n)} \right) \underset{k\to\infty}{\to} 0. \tag{49}$$

In the sequel, it will be shown that (42), (43), (47) and (49) imply that $\mathbf{h}_k \underset{k\to\infty}{\to} 0$, which contracts with $\|h_k\| = 1$. This means that the converse assumption (36) is not true.

For any $m \neq n$, since $h^{(n)}(z)$, $h^{(m)}(z)$ are coprime, where $h^{(n)}(z)$ is given in (6), there exist polynomials $d_1(z)$, $d_2(z)$ such that

$$d_1(z)h^{(n)}(z) + d_2(z)h^{(m)}(z) = 1. \tag{50}$$

Let r_1 and r_2 be the degrees of $d_1(z)$ and $d_2(z)$, respectively. Set $q = 4(r_1 + r_2) + 5L + 1$. Introduce the q-dimensional vector g_k^s and $q \times q$ square matrices T and A as follows

$$g_k^s = (\underbrace{0, \cdots, 0}_{2(r_1+r_2+L)}, h_{k,0}^{(s)}, \cdots, h_{k,L}^{(s)}, \underbrace{0, \cdots, 0}_{2(r_1+r_2+L)})^\tau,$$

$$T = \begin{pmatrix} 0 & \cdots & 0 & 1 \\ \vdots & & \diagup & 0 \\ 0 & \diagup & & \vdots \\ 1 & 0 & \cdots & 0 \end{pmatrix}, \quad A = \begin{pmatrix} 0 & \cdots & \cdots & 0 \\ 1 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \vdots \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Notes that $Tg = (g_q, \cdots, g_1)^\tau$, where $g = (g_1, \cdots, g_q)^\tau$ and $Ag = (0, g_1, \cdots, g_{q-1})^\tau$, $A^\tau g = (g_2, \cdots, g_q, 0)^\tau$. Then (42), (43), (47) and (49) can be written in the following compact form:

$$h^{(s)}(A)Tg_k^t + h^{(t)}(A^\tau)A^L g_k^t \underset{k\to\infty}{\to} 0, \quad \forall s, t = 1, \cdots p. \tag{51}$$

To see this, note that for any fixed s and t, on the left hand sides of (47) and (49) there are $2L$ different sums when l varies from 0 to $L-1$ and s, t replace roles each other. These together with (42) and (43) give us $2L + 1$ sums, and each of them tends to zero. Explicitly expressing (51), we find that there are $2L + 1$ nonzero rows and each row corresponds to one of the relationships in (42), (43), (47) and (49).

Since we have put enough zeros in the definition of g_k^s, multiplying the left hand side of (51) by A^i, $\forall i \leqslant r_1 + r_2$, $A^i(h^{(s)}(A)Tg_k^t + h^{(t)}(A^\tau)A^L g_k^s)$ has only shifted nonzero elements in $h^{(s)}(A)Tg_k^t + h^{(t)}(A^\tau)A^L g_k^s$.

From (51) it follows that for any $l : l = 1, \cdots, p$, and m, n in (50)

$$\begin{aligned} &(d_1(A)(h^{(n)}(A)Tg_k^l + h^{(l)}(A^\tau)A^L g_k^n) + d_2(A)(h^{(m)}(A)Tg_k^l + h^{(l)}(A^\tau)A^L g_k^m) \\ &= d_1(A)h^{(n)}(A) + d_2(A)h^{(m)}(A))Tg_k^l + h^{(l)}(A^\tau)A^L(d_1(A)g_k^n + d_2(A)g_k^m) \\ &= Tg_k^l + h^{(l)}(A^\tau)A^L(d_1(A)g_k^n + d_2(A)g_k^m) \underset{k\to\infty}{\to} 0. \end{aligned} \tag{52}$$

From (52) it follows that

$$\begin{aligned}&d_1(A^\tau)[Tg_k^n + h^{(n)}(A^\tau)A^L(d_1(A)g_k^n + d_2(A)g_k^m)]\\&+d_2(A^\tau)[Tg_k^m + h^{(m)}(A^\tau)A^L(d_1(A)g_k^n + d_2(A)g_k^m)]\underset{k\to\infty}{\to} 0.\end{aligned} \tag{53}$$

Note that for any polynomial $g(z)$ of degree r, $d(A^\tau)Ty = Td(A)y$, if the last r elements of y are zeros. From (53) it follows that

$$\begin{aligned}&d_1(A^\tau)Tg_k^n + d_2(A^\tau)Tg_k^m + (d_1(A^\tau)h^{(n)}(A^\tau)\\&+d_2(A^\tau)h^{(m)}(A))z^L(d_1(A)g_k^n + d_2(A)g_k^m)\\&= T(d_1(A)g_k^n + d_2(A)g_k^m) + A^L(d_1(A)g_k^n + d_2(A)g_k^m)\underset{k\to\infty}{\to} 0.\end{aligned} \tag{54}$$

Denoting

$$g_k = (g_{k,1}, \cdots, g_{k,q})^\tau \triangleq d_1(A)g_k^n + d_2(A)g_k^m,$$

from (54) we find that

$$Tg_k + A^L g_k \underset{k\to\infty}{\to} 0. \tag{55}$$

By the definition of g_k^n, the first $2(r_1 + r_2 + L)$ elements of g_k are zeros, i.e., $g_{k,i} = 0, i = 1, \cdots, 2(r_1 + r_2 + L)$. This means that the last $2(r_1 + r_2 + L)$ elements of Tg_k are zeros, i.e.,

$$Tg_k = (\underbrace{g_{k,q}, g_{k,q-1}, \cdots,}_{2(r_1+r_2+L)+L} g_{k,2(r_1+r_2+L)+1}, \underbrace{0, \cdots, 0}_{2(r_1+r_2+L)})^\tau. \tag{56}$$

On the other hand,

$$A^L g_k = (\underbrace{0, \cdots, 0,}_{2(r_1+r_2+L)+L} g_{k,2(r_1+r_2+L)+1}, \underbrace{\cdots, g_{k,q-L}}_{2(r_1+r_2+L)}). \tag{57}$$

By (55), from (56)(57) it is seen that $g_k \underset{k\to\infty}{\to} 0$, i.e.,

$$d_1(A)g_k^n + d_2(A)g_k^m \underset{k\to\infty}{\to} 0.$$

From (52) it then follows that

$$g_k^l \underset{k\to\infty}{\to} 0, \quad \forall l = 1, \cdots, p,$$

i.e., $h_k^{(l)} \underset{k\to\infty}{\to} 0$, $\forall l = 1, \cdots p$. But this is impossible, because $\mathbf{h}_k$ are unit vectors. Consequently, (36) is impossible and this proves the lemma. □

References

1. Benveniste, A., Metivier, M. and Priouret, P. (1990), *Adaptive Algorithms and Stochastic Approximations*, Springer, Berlin.

2. Brandière, O. and Duflo, M. (1996), Les algorithmes stochastiques contournents-ils les pièges? *Ann. Inst. Henri Poincaré*, 32, 395–427.
3. Chen, H.F. (1994), *Stochastic appoximation and its new applications*, Proc. Hong Kong International Workshop on New Directions of Control and Manufacturing, pp. 2–12.
4. Chen, H.F. (1998), Stochastic approximation with non-additive measurement noise, *J. Appl. Probab.*, 35, 407–417.
5. Chen, H.F., Cao, X.R. and Zhu, J., Recursive algorithms for multichannel blind identification, IEEE Trans. Information Theory, Vol. 48, No. 5, 2002, 1214–1225.
6. Ding, Z. and Li, Y. (1994), On channel identification based on second order cyclic spectra, *IEEE Trans. Signal Process.*, 42, 1260–1264.
7. Hua, Y. and Wax, M. (1996), Strict identifiability of multiple FIR channels driven by an unknown arbitrary sequence, *IEEE Trans. Signal Processing*, 44, 756–759.
8. Kushner H.J. and Yin, G. (1997), *Stochastic Approximation Algorithms and Applications*, Springer, Berlin.
9. Liu, R. (1996), Blind signal processing: An introduction, *Proc. 1996 Intl. Symp. Circuits and Systems*, 2, 81–83.
10. Liu, H., Xu, G. Tong, L. and Kailath, T. (1996), Recent developments in blind channel equalization: From cyclostationarity to subspace, *Signal Processing*, 50, 516–525.
11. Moulines, E., Duhamel, P. Cardoso, J.-F. and Mayrargue, S. (1995), Subspace methods for the blind identification of multichannel FIR filters, *IEEE Trans. on Signal Process.*, 43, 516–525.
12. Sato, D. (1975), A method of self-recovering equalization for multilevel amplitude-modulation, *IEEE Trans. on Commun.*, 23, 679–682.
13. Tong L. and Perreau, S. (1998), Multichannel blind identification: From subspace to maximum likelihood methods, *IEEE Proceeding*, 86, 1951–1968.
14. Tong, L., Xu, G. and Kailath, T. (1994), Blind identification and equalization based on second-order statistics: A time domain approach, *IEEE Trans. Inform. Theory*, 40, 340–349.
15. Tong, L., Xu, G. and Kailath, T. (1995), Blind channel identification based on second-order statistics: A frequency-domain approach, *IEEE Trans. Inform. Theory*, 41, 329–334.
16. van der Veen, A.J., Talwar, S. and Paulraj, A. (1995), Blind estimation of multiple digital signals transmitted over FIR channels, *IEEE Signal Process. Lett.*, 2, 99–102.
17. Xu, G.H., Liu, H., Tong, L. and Kailath, T. (1995), A least-square approach to blind channel identification, *IEEE Trans. Signal Process.*, 43, 2982–2993.
18. Zhao, Q. and Tong, L. (1999), Adaptive blind channel estimation by least squares smoothing, *IEEE Trans. Signal Process.*, 47, 3000–3012.
19. Zhang, J.H. and Chen, H.F. (1997), Convergence of algorithms used for principal component analysis, *Sci. in China (Series E)*, 40, 597–604.
20. Chen, H.F. (2002), Stochastic Approximation and its Applications, Kluwer Academic Publishers, Dordrecht.