Nonlinear Adaptive Blind Whitening for MIMO Channels

Han-Fu Chen, *Fellow, IEEE*, Xi-Ren Cao, *Fellow, IEEE*, Hai-Tao Fang, *Member, IEEE*, and Jie Zhu, *Senior Member, IEEE*

***Abstract*—A nonlinear adaptive whitening method is proposed for blind deconvolution of MIMO systems by whitening the received signals in both time and space, with a highly nonlinear function of the past output data. The whitened signals are ISI-free and can be viewed as outputs of a memoryless paraunitary mixing system. The convergence of the proposed recursive algorithm is proved. Numerical simulation shows that the whitening method proposed in the paper works well, even if the output signal is corrupted by additive noise.**

***Index Terms*—Blind equalization, blind intersymbol interference (ISI) cancellation, blind source separation (BSS), multiple-input–multiple-output (MIMO) linear systems, nonlinear blind equalization, second-order statistics (SOS), signal uncorrelation, signal whitening.**

NOMENCLATURE

The following notation is the list of mathematical symbols used in the paper.

z	Backward-shift operator.
Ex	Expectation of x.
I	Identity matrix (with a compatible dimension).
$\|A\|$	Norm of a matrix A, defined as the square root of the maximal singular value of A.
$\lambda_{\max}(A)$	Maximum eigenvalue of a non-negative definite matrix A.
$\otimes$	Kronecker product.
trA	Sum of the diagonal elements of matrix A.

I. INTRODUCTION

THE GOAL of blind equalization for multiple-input–multiple-output (MIMO) systems is to recover multiple source signals from the observations of their mixtures without using a reference signal to determine the channels. Such a signal processing problem can be abstracted from many applications in communications, image processing, speech enhancement, and biomedical measurements. When passing through a media in digital communication, a source signal in an MIMO system generally suffers from a convolutive distortion between its symbols and a mixture distortion from other source signals. These two distortions are called intersymbol interference (ISI) and interuser interference (IUI), respectively. Specifically, let s_k be a p-dimensional input signal. An MIMO channel is characterized by a matrix polynomial

$$H(z) = H_0 + H_1 z + \cdots + H_r z^r \tag{1}$$

where $H_i, i = 0, 1, \cdots, r$ are unknown $q \times p$ matrices, $q \geq p$, and z^i denotes the time-delay operator

$$z^i s_k = s_{k-i}.$$

We consider the case without noise first and then extend the results to the noise case. The channel output, thus, can be formulated by

$$x_k = H(z) s_k. \tag{2}$$

Blind equalization of an MIMO system is to recover the channel input s_k by using the output x_k only. For a system having one source signal [i.e., $p = 1$, the single-input–single-output (SISO), or the single-input–multiple-output (SIMO) system], a large amount of algorithms has been developed based on either nonzero higher order statistics of stationary sources or second-order statistics (SOS) of nonstationary (cyclostationary) sources. On the other hand, if the media effect on the signals can be modeled by a memoryless scale, the convolutive distortion vanishes, i.e., $r = 0$ in (1). The problem for this special case is named blind source separation (BSS) [3], [4], [26]. Again, there are many results published in this area. When the mixing matrix is full-column rank, non-Gaussian sources can always be separated by higher order statistics, and nonstationary (cyclostationary) sources can be picked out by SOS, provided that they have distinct power spectra [16]. Due to its potential applications in digital communications, blind equalization of MIMO systems has attracted attention from a growing number of researchers recently. Many equalization solutions to MIMO systems are the extensions of algorithms developed for SISO or SIMO systems based on higherorder statistics (see [20], [21], and references therein). The principle behind it is the Skitovich–Darmois theorem or the maximum entropy formulation. Tugnait [23] developed a blind MIMO equalizer by using a Godard cost function. Chen and Petropulu [5] gave a solution by jointly diagonalizing the polyspectra slices. Touzni *et al.* [22] proposed a set of hierarchical criteria

Manuscript received March 17, 2003; revised September 4, 2004. The work of H.-F. Chen and H.-T. Fang was supported in part by the National Science Foundation of China and in part by the Ministry of Science and Technology of China. The work of X. R. Cao was supported in part by a grant from Hong Kong UGC. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Athina Petropulu.

H.-F. Chen and H.-T. Fang are with Laboratory of Systems and Control, Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, 100080, China (e-mail: hfchen@mail.iss.ac.cn; htfang@mail.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 is with ESS Technology Inc., Fremont, CA 94538 USA (e-mail: jie.zhu@ieee.org).

Digital Object Identifier 10.1109/TSP.2005.850338

from the maximum entropy principle point of view to build an adaptive globally convergent MIMO equalizer. The source signals are required to be sub-Gaussian, i.e., their normalized kurtosises should be less than three.

As it is known, the higher order statistics may not be accurate in comparison with the SOS for a given limit number of signal samples. Also, the higher order statistics require more computational power. Thus, it is usually preferred to use the SOS if it works. Inouye and Liu [11] studied the blind equalization of finite impulse response (FIR) MIMO channel systems on the SOS and concluded that every equalizable channel has an FIR irreducible-paraunitary factorization and can be reduced to a paraunitary FIR system by decorrelation. Although the paraunitary FIR system can be arbitrary, its output signals do not have the ISI distortion, thus reducing the equalization in MIMO systems to the BSS problem. It is worthwhile to mention that because of the resultant paraunitary structure, the BSS can be done by rotating the source signals only with higher order statistics. Consequently, the blind equalization for MIMO systems can be achieved with two steps: ISI cancellation by SOS (decorrelation) and IUI cancellation by higher order statistics or maximum entropy principle. Tugnait and Huang [24] realized this two-step procedure by first whitening the observations up to a unitary mixing matrix and then "unmixing" them with fourth-order statistics. Lopez-Valcarce *et al.* [15] showed that a user channel can be equalized based on only SOS if it is the unique one having the longest memory. Referring to (1), a user channel with longest memory means that the column of matrix H_r corresponding to the user is nonzero. If there is only one such channel, H_r also has only one nonzero column accordingly. As a result, a matrix F satisfying $FH_i = 0$ for $i = 0, \ldots, r-1$ will provide a copy of the user signal up to a scale. To detect a special user signal, the authors suggested filtering the user signal before transmitting it so that the "whole" channel it goes through has the longest memory. By multiplying each user signal with a (known to receiver) complex exponential at a characteristic frequency before transmission so that the whitened signal admits conjugate cyclostationarity, Chevreuil and Loubaton [8] proposed to equalize an MIMO channel up to a paraunitary mixing matrix by SOS and to recover all the user signals on the characteristic frequencies. This method is somehow similar to that in code-division multiple-access (CDMA) systems, where an assigned "code" is used to recognize a particular user. CDMA signal detectors also benefit from the ISI-free whitened signals.

In this paper, we consider the blind ISI cancellation problem for MIMO systems in general cases, i.e., $p > 1$, $r > 0$. The source signals are assumed to be white and mutually independent (i.e., the identification of those MIMO systems driven by colored source signals is, therefore, not involved here.) We propose a self-whitening algorithm to directly output ISI-free estimates for MIMO systems without identifying the channel. The algorithm consists of two steps: whitening in time and whitening in space. The whitened outputs can be seen as all source signals transmitted through a memoryless FIR system. When the received signals are corroded by noises, the whitened outputs give the maximal signal-to-noise (including intersymbol interference) ratio. The algorithm is different from those presented in existing literatures: It is neither a zero-forcing equalizer nor a maximum-likelihood estimator. Its output is not derived via the received signals passing through a linear filter bank but is a direct estimate nonlinearly depending on the past output signals. More realistically, it is a recursive algorithm that is expected to adapt the slow change of the channel, although in the sequel, the time-invariant channel is still assumed.

The paper is organized as follows. In Section II, the extended least-squares (ELS) method combined with the overparameterization technique is applied for whitening in time, and the ELS output is proved to be asymptotic whitening. In Section III, the recursive whitening in space is asymptotically derived in the mean-square sense. In Section IV, performance comparison between the nonlinear whitening (NW) method and the linear prediction deconvolution (LPD) method [11] for FIR channels is presented. Some concluding remarks are given in Section V.

II. Whitening in Time

We consider the case where the observation is free of noise. In this section, by the use of the ELS method, we whiten the observed signal in time.

The following conditions are to be used.

A1) The components of $\{s_k\}$ are mutually independent, $Es_k = 0$, $Es_k s_k^T = \sigma^2 I$, and $\sup_k E\|s_k\|^{2+\delta} < \infty$ for some real number $\delta > 0$, where σ^2 is the power of source signals.

A2) There exists a $p \times q$-matrix L such that $LH(z)$ is stable [i.e., all roots of $\det LH(z)$ are outside the closed unit disk].

Remark 1: Notice that the condition A1) is weaker than common independent and identically distributed (i.i.d.) assumption. In fact, the whitening algorithm developed in the sequel is based on the SOS only; thus, it does not require the source signals to be i.i.d.. Under A1), we have

$$\sum_{i=1}^{\infty} \frac{s_i s_i^T - \sigma^2 I}{i} < \infty \text{ and } \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} s_i s_i^T = \sigma^2 I \text{ a.s.} \quad (3)$$

Remark 2: Without loss of generality, we assume that all source signals have the same power, because if a source signal has a different power level, the power difference may be compensated by changing the gain of the channel it goes through.

Remark 3: The condition A2) does not necessarily imply that the MIMO channel has to be irreducible or equalizable. It ensures the convergence of the recursive whitening procedure developed in the sequel. When the MIMO channel meets A2) and is reducible as well, the whitening procedure still provides white outputs, and the source signal can asymptotically be recovered up to a multiple being an orthogonal matrix. The algorithm cannot whiten the signal when the channel does not meet A2).

Denoting $x_k^L \triangleq Lx_k$, $\varepsilon_k \triangleq C_0 s_k$, $C_0 = LH_0$, and $C_j \triangleq LH_j C_0^{-1}$, $j = 1, \cdots, r$, we then have

$$C(z) = I + C_1 z + \cdots + C_r z^r \quad (4)$$

and

$$x_k^L = C(z)\varepsilon_k. \quad (5)$$

It is worth noting that

$$\varepsilon_k = LH_0 s_k$$

can be viewed as the output of a flat fading channel, and $\{x_k^L\}$ serves as the observed signal. The problem is now reduced to transforming $\{x_k^L\}$ to a sequence that is white in both time and space. In this section, we whiten $\{x_k^L\}$ in time. As a matter of fact, based on $\{x_k^L\}$, we give an estimate $\hat{\varepsilon}_k$ converging to ε_k in the mean-square sense. Then, $\{\hat{\varepsilon}_k\}$ may serve as a sequence whitened in time. Since it depends upon the first tap-matrix H_0 only, the estimate of ε_k may not be acceptable when H_0 is away from full-column rank. We refer to [9] for more information about this issue.

By stability of $C(z)$, we have

$$C^{-1}(z) = \sum_{i=0}^{\infty} D_i z^i, \ \forall |z| \le 1 \tag{6}$$

where $\|D_i\| \le M\lambda^i$, for some $M > 0$ and $\lambda \in (0,1)$.

Let m be a sufficiently large integer such that

$$m > \left[\frac{\left(\log\left[\|C(z)\|_\infty^{-1} M^{-1}(1-\lambda)\right]\right)}{\log \lambda} \right] - 1 \tag{7}$$

where $\|C(z)\|_\infty = \max_{|z|=1} \lambda_{\max}[C(z)C^T(z^{-1})]$, and denote

$$D(z) \triangleq \sum_{i=0}^{m} D_i z^i, \quad D_0 = I. \tag{8}$$

Lemma 1: If $C(z)$ is stable and $D(z)$ is defined by (6)–(8), then

$$[D(z)C(z)]^{-1} - \frac{1}{2}I \tag{9}$$

is strictly positive real (SPR), i.e.,

$$\left[D(e^{j\lambda})C(e^{j\lambda})\right]^{-1} + \left(\left[D(e^{-j\lambda})C(e^{-j\lambda})\right]^T\right)^{-1} - I > 0$$

$\forall \lambda \in [0, 2\pi]$, and $D(z)$ is stable.

Proof: The SPR property is proved in [7, p. 139].

SPR of $[D(z)C(z)]^{-1} - (1/2)I$ implies SPR of $[D(z)C(z)]^{-1}$. Then, $D(z)C(z)$ is also SPR (see, e.g., [7, Lemma 4.1]), and hence, $D(z)$ is stable. ■

Remark 4: If $C(z)$ is SPR, then we may take $D(z) = I$.

Denote

$$F(z) \triangleq D(z)C(z). \tag{10}$$

Then, from (5), we obtain the following ARMA system:

$$D(z)x_k^L = F(z)\varepsilon_k. \tag{11}$$

It is clear that $F(0) = I$. Let $F(z) = I + F_1 z + \cdots + F_{mr} z^{mr}$, and

$$\theta^T = [-D_1, \cdots, -D_m, F_1, \cdots, F_{mr}]. \tag{12}$$

The recursive algorithm whitening $\{x_k^L\}$ in time is defined as follows:

$$\hat{\varepsilon}_{k+1} = x_{k+1}^L - \theta_{k+1}^T \varphi_k \tag{13}$$

$$\theta_{k+1} = \theta_k + a_k P_k \varphi_k \left(x_{k+1}^{LT} - \varphi_k^T \theta_k\right) \tag{14}$$

$$P_{k+1} = P_k - a_k P_k \varphi_k \varphi_k^T P_k^T$$

$$a_k = \left(1 + \varphi_k^T P_k \varphi_k\right)^{-1} \tag{15}$$

with $P_0 = \alpha I$ and arbitrary θ_0, where

$$\varphi_k^T = \left[x_k^{LT}, \cdots, x_{k-m+1}^{LT}, \hat{\varepsilon}_k^T, \hat{\varepsilon}_{k-1}^T, \cdots, \hat{\varepsilon}_{k-mr+1}^T\right]. \tag{16}$$

Theorem 1: Under Conditions A1) and A2), $\{\hat{\varepsilon}_k\}$ is a sequence asymptotically whitened in time with

$$\frac{1}{n}\sum_{k=1}^{n} \|\hat{\varepsilon}_k - \varepsilon_k\|^2 = O\left(\frac{\log n}{n}\right) \tag{17}$$

and

$$\frac{1}{n}\sum_{k=1}^{n} \hat{\varepsilon}_k \hat{\varepsilon}_{k+j}^T \underset{n\to\infty}{\longrightarrow} \begin{cases} \sigma^2 C_0 C_0^T > 0, & j = 0 \\ 0, & j > 0 \end{cases}. \tag{18}$$

Proof: Let $\xi_k = \hat{\varepsilon}_k - \varepsilon_k$, and denote by $\lambda_{\max}(n)$ the maximum eigenvalue of $(P_{n+1})^{-1} = \sum_{i=0}^{n} \varphi_i \varphi_i^T + (1/\alpha)I$, i.e., $\lambda_{\max}(n) = \lambda_{\max}(P_{n+1}^{-1})$. By Lemma 1, we see that [7, Theorem 4.1] is applicable to (11) by taking $\beta = 2 + \delta$ and $u_k \equiv 0$ in that theorem.

Then, it is proved in [7] (see (4.62) of [7]) that

$$\sum_{i=0}^{n} \|\xi_{i+1}\|^2 = O\left(\log\left(\lambda_{\max}(n)\right)\right). \tag{19}$$

Let us define

$$\psi_k^T = \left[\hat{\varepsilon}_k^T, \cdots, \hat{\varepsilon}_{k-r+1}^T\right] \text{ and } Q_{k+1} = \left(\sum_{i=0}^{k} \psi_i \psi_i^T + \frac{1}{\alpha}I\right)^{-1} \tag{20}$$

and denote by $\mu_{\max}(n)$ the maximum eigenvalue of Q_n^{-1}, i.e., $\mu_{\max}(n) = \lambda_{\max}(Q_n^{-1})$.

Further, denote

$$\psi_k^0 = [\varepsilon_k, \cdots, \varepsilon_{k-r+1}]^T \text{ and } \psi_k^\xi = \psi_k - \psi_k^0 \tag{21}$$

and notice that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \varepsilon_k \varepsilon_{k+j}^T = \begin{cases} \sigma^2 C_0 C_0^T, & j = 0 \\ 0, & j > 0. \end{cases} \tag{22}$$

From (22) it follows that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n} \psi_k^0 \psi_k^{0T} = I \otimes \sigma^2 C_0 C_0^T \tag{23}$$

where $\otimes$ is the Kronecker product, and I is an r-dimensional identity matrix.

From (5) and (22), it follows that

$$\sum_{k=1}^{n} \left\|x_k^L\right\|^2 = O(n).$$

Therefore, we have

$$\lambda_{\max}(n) = O\left(\sum_{k=1}^{n} \left\|x_k^L\right\|^2 + \sum_{k=1}^{n} \|\hat{\varepsilon}_k\|^2\right) = O\left(n + \mu_{\max}(n)\right). \tag{24}$$

Let x be a unit vector with the same dimension as ψ_n. Then, by the Schwartz inequality, it follows that

$$\begin{aligned}\sum_{i=0}^{n} \left(x^T \psi_i\right)^2 &= \sum_{i=0}^{n} \left(x^T \psi_i^\xi + x^T \psi_i^0\right)^2 \\ &\le 2\sum_{i=0}^{n} \left(x^T \psi_i^0\right)^2 + 2\sum_{i=0}^{n} \left\|\psi_i^\xi\right\|^2. \end{aligned} \tag{25}$$

From (23), we see $2\sum_{i=0}^{n} \left(x^T \psi_i^0\right)^2 = O(n)$, while from (19) and (24), it follows that

$$2\sum_{i=0}^{n} \left\|\psi_i^\xi\right\|^2 = O\left(\log\left(n + \mu_{\max}(n)\right)\right).$$

By noticing $\mu_{\max}(n) = \max_{\|x\|=1} \sum_{i=0}^{n} \left(x^T \psi_i\right)^2 + (1/\alpha)$, from (25), we conclude that

$$\mu_{\max}(n) = O(n) + O\left(\log\left(n + \mu_{\max}(n)\right)\right)$$

which implies

$$\mu_{\max}(n) = O(n). \tag{26}$$

Incorporating (26) with (24), from (19), we derive

$$\sum_{i=0}^{n} \|\xi_{i+1}\|^2 = O(\log n)$$

which proves (17).

By (17) and (22), we see that all terms on the right-hand side of (27) stated below tend to zero for $j > 1$ and to $\sigma^2 C_0 C_0^T$ for $j = 0$ as $n \to \infty$

$$\begin{aligned}\frac{1}{n}\sum_{k=1}^{n} \hat{\varepsilon}_k \hat{\varepsilon}_{k+j}^T &= \frac{1}{n}\sum_{k=1}^{n} (\hat{\varepsilon}_k - \varepsilon_k)\hat{\varepsilon}_{k+j}^T \\ &+ \frac{1}{n}\sum_{k=1}^{n} \varepsilon_k (\hat{\varepsilon}_{k+j} - \varepsilon_{k+j})^T + \frac{1}{n}\sum_{k=1}^{n} \varepsilon_k \varepsilon_{k+j}^T. \end{aligned} \tag{27}$$

This proves (18). ■

This means that $\{\hat{\varepsilon}_k\}$ has asymptotically been whitened in time. We note that (13)–(16) is the extended least-squares estimate for θ. Under the conditions of Theorem 1, θ_k actually is strongly consistent for θ. If $C(z)$ is stable, then for sufficiently large m, we have (18). So, for A2), we may try different L and kick off those L for which $(1/n)\sum_{k=1}^{n} \hat{\varepsilon}_k \hat{\varepsilon}_{k+j}^T$ diverges or becomes nondegenerate.

Since $\sigma^2 C_0 C_0^T$ may not be diagonal, we have to further whiten $\hat{\varepsilon}_k$ in space.

III. Whitening in Space

Assume L has been selected.

We have

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \hat{\varepsilon}_k \hat{\varepsilon}_k^T = \sigma^2 C_0 C_0^T \triangleq R_\varepsilon. \tag{28}$$

Using the data $\{\hat{\varepsilon}_k \hat{\varepsilon}_k^T\}$, we proceed to recursively diagonalize R_ε. In fact, we present a recursive method for principal component analysis.

Recursively define

$$\tilde{u}_{k+1}^{(1)} = u_k^{(1)} + \frac{1}{k}\hat{\varepsilon}_{k+1}\hat{\varepsilon}_{k+1}^T u_k^{(1)} \tag{29}$$

$$u_{k+1}^{(1)} = \frac{\tilde{u}_{k+1}^{(1)}}{\left\|\tilde{u}_{k+1}^{(1)}\right\|}, \text{ if } \left\|\tilde{u}_{k+1}^{(1)}\right\| \neq 0. \tag{30}$$

If $\|\tilde{u}_{k+1}^{(1)}\| = 0$, $u_k^{(1)}$ is reset to be a vector with norm 1. Define

$$P_k^{(i)} \triangleq I - V_k^{(i)} V_k^{(i)T}, \quad i = 1, \cdots, j-1 \tag{31}$$

$$V_k^{(j)} = \left[u_k^{(1)} P_k^{(1)} u_k^{(2)} \cdots P_k^{(j-1)} u_k^{(j)}\right] \tag{32}$$

$$\tilde{u}_{k+1}^{(j+1)} = P_k^{(j)} u_k^{(j+1)} + \frac{1}{k} P_k^{(j)} \hat{\varepsilon}_{k+1}\hat{\varepsilon}_{k+1}^T P_k^{(j)} u_k^{(j+1)} \tag{33}$$

$$u_{k+1}^{(j+1)} = \frac{\tilde{u}_{k+1}^{(j+1)}}{\left\|\tilde{u}_{k+1}^{(j+1)}\right\|}, \text{ if } \left\|\tilde{u}_{k+1}^{(j+1)}\right\| \ge \varepsilon,\ 0 < \varepsilon < \frac{1}{4}. \tag{34}$$

If $\|\tilde{u}_{k+1}^{(j+1)}\| < \varepsilon$, define a $u_k^{(j+1)}$ with $\|u_k^{(j+1)}\| = 1$ such that $\|P_k^{(j)} u_k^{(j+1)}\| = 1$.

Let M_k be a sequence of positive real numbers such that $M_{k+1} > M_k$, $M_k > 0$, $M_k \underset{k\to\infty}{\to} \infty$. Define

$$\begin{aligned}\lambda_{k+1}^{(j)} &= \left[\lambda_k^{(j)} - \frac{1}{k}\left(\lambda_k^{(j)} - u_k^{(j)T}\hat{\varepsilon}_{k+1}\hat{\varepsilon}_{k+1}^T u_k^{(j)}\right)\right] \\ &\quad \times I_{\left[\left|\lambda_k^{(j)} - \frac{1}{k}\left(\lambda_k^{(j)} - u_k^{(j)T}\hat{\varepsilon}_{k+1}\hat{\varepsilon}_{k+1}^T u_k^{(j)}\right)\right| \le M_{\sigma_k}\right]} \qquad (35) \\ \sigma_k &= \sum_{i=1}^{k-1} I_{\left[\left|\lambda_i^{(j)} - \frac{1}{i}\left(\lambda_i^{(j)} - u_i^{(j)T}\hat{\varepsilon}_{i+1}\hat{\varepsilon}_{i+1}^T u_i^{(j)}\right)\right| > M_{\sigma_i}\right]} \\ \sigma_0 &= 0, \quad j = 1, \cdots, p \\ \Lambda_k &\triangleq \begin{bmatrix} \lambda_k^{(1)} & & 0 \\ & \ddots & \\ 0 & & \lambda_k^{(p)} \end{bmatrix} \\ U_k &\triangleq \left[u_k^{(1)}, \cdots, u_k^{(p)}\right] \qquad (36)\end{aligned}$$

where $I_{[\bullet]}$ is an indicator function, which equals 1 if the relation in the bracket is true and 0 otherwise. Set

$$\hat{s}_k \triangleq \Lambda_k^{-\frac{1}{2}} U_k^T \hat{\varepsilon}_k \left(= \Lambda_k^{-\frac{1}{2}} U_k^T \left(Lx_k - \theta_k^T \varphi_{k-1}\right)\right). \tag{37}$$

The following theorem shows that $\hat{s}_k$ is the desired estimate whitened in both time and space.

Theorem 2: Assume that Conditions A1) and A2) hold and $(\|s_k\|^3/k) \underset{k\to\infty}{\to} 0$ a.s. Then, $\hat{s}_k$ is asymptotically whitened in space

$$\frac{1}{n}\sum_{k=1}^{n} \hat{s}_k\hat{s}_k^T \underset{n\to\infty}{\to} I \tag{38}$$

$$\frac{1}{n}\sum_{k=1}^{n} \hat{s}_k\hat{s}_{k+j}^T \underset{n\to\infty}{\to} 0,\ \forall j \geq 1 \tag{39}$$

and

$$U_k\Lambda_k U_k^T \to R_\varepsilon. \tag{40}$$

Proof: Applying the Taylor's expansion leads to

$$\begin{aligned} u_{k+1}^{(1)} &= \left(u_k^{(1)} + \frac{1}{k}\hat{\varepsilon}_{k+1}\hat{\varepsilon}_{k+1}^T u_k^{(1)}\right) \\ &\quad\times\left(1 + \frac{2}{k}u_k^{(1)T}\hat{\varepsilon}_{k+1}\hat{\varepsilon}_{k+1}^T u_k^{(1)}\right. \\ &\qquad\left. + \frac{1}{k^2}u_k^{(1)T}\left(\hat{\varepsilon}_{k+1}\hat{\varepsilon}_{k+1}^T\right)^2 u_k^{(1)}\right)^{-\frac{1}{2}} \\ &= \left(u_k^{(1)} + \frac{1}{k}\hat{\varepsilon}_{k+1}\hat{\varepsilon}_{k+1}^T u_k^{(1)}\right) \\ &\quad\times\left\{1 - \frac{1}{k}u_k^{(1)T}\hat{\varepsilon}_{k+1}\hat{\varepsilon}_{k+1}^T u_k^{(1)}\right. \\ &\qquad - \frac{1}{2k^2}u_k^{(1)}\left(\hat{\varepsilon}_{k+1}\hat{\varepsilon}_{k+1}^T\right)^2 u_k^{(1)} \\ &\qquad + \frac{3}{8}\left[\frac{4}{k^2}\left(u_k^{(1)T}\hat{\varepsilon}_{k+1}\hat{\varepsilon}_{k+1}^T u_k^{(1)}\right)^2\right. \\ &\qquad\qquad + \frac{4}{k^3}u_k^{(1)T}\hat{\varepsilon}_{k+1}\hat{\varepsilon}_{k+1}^T u_k^{(1)}u_k^{(1)T} \\ &\qquad\qquad \left.\times\left(\hat{\varepsilon}_{k+1}\hat{\varepsilon}_{k+1}^T\right)^2 u_k^{(1)}\right] \\ &\qquad \left. - \frac{5}{2k^3}\left(u_k^{(1)T}\hat{\varepsilon}_{k+1}\hat{\varepsilon}_{k+1}^T u_k^{(1)}\right)^3 + r_k\right\}. \end{aligned} \tag{41}$$

Since $(\|s_k\|^3/k) \underset{k\to\infty}{\to} 0$, we have $(\|\varepsilon_k\|^3/k) \underset{k\to\infty}{\to} 0$ a.s. It is clear that (17) implies that $\|\hat{\varepsilon}_k - \varepsilon_k\| = O(\log k)$. Therefore

$$\frac{1}{k}\|\hat{\varepsilon}_k\|^3 \leq \frac{4}{k}\left(\|\varepsilon_k\|^3 + \|\hat{\varepsilon}_k - \varepsilon_k\|^3\right) \underset{k\to\infty}{\to} 0 \text{ a.s.} \tag{42}$$

Therefore, in (41)

$$\|r_k\| \leq c\left\|\left(\frac{1}{k}\hat{\varepsilon}_{k+1}\hat{\varepsilon}_{k+1}^T\right)^4\right\| \to 0$$

where c may depend on a sample path.

We then rewrite (41) as

$$\begin{aligned} u_{k+1}^{(1)} = u_k^{(1)} &+ \frac{1}{k}\hat{\varepsilon}_{k+1}\hat{\varepsilon}_{k+1}^T u_k^{(1)} \\ &- \frac{1}{k}\left(u_k^{(1)T}\hat{\varepsilon}_{k+1}\hat{\varepsilon}_{k+1}^T u_k^{(1)}\right)u_k^{(1)} + \frac{1}{k}\nu_{k+1}^{(1)} \end{aligned} \tag{43}$$

where

$$\begin{aligned} \nu_{k+1}^{(1)} \triangleq \frac{1}{k}&\left[-\frac{1}{2}u_k^{(1)T}\left(\hat{\varepsilon}_{k+1}\hat{\varepsilon}_{k+1}^T\right)^2 u_k^{(1)}\right. \\ &\quad + \frac{3}{2}\left(u_k^{(1)T}\hat{\varepsilon}_{k+1}\hat{\varepsilon}_{k+1}^T u_k^{(1)}\right)^2 u_k^{(1)} \\ &\quad \left. - \left(u_k^{(1)T}\hat{\varepsilon}_{k+1}\hat{\varepsilon}_{k+1}^T u_k^{(1)}\right)\hat{\varepsilon}_{k+1}\hat{\varepsilon}_{k+1}^T u_k^{(1)}\right] \\ + \frac{1}{k^2}&\left[\frac{3}{2}\left(u_k^{(1)T}\hat{\varepsilon}_{k+1}\hat{\varepsilon}_{k+1}^T u_k^{(1)}\right)\right. \\ &\quad\times\left(u_k^{(1)T}\left(\hat{\varepsilon}_{k+1}\hat{\varepsilon}_{k+1}^T\right)^2 u_k^{(1)}\right)u_k^{(1)} \\ &\quad - \frac{5}{2}\left(u_k^{(1)}\hat{\varepsilon}_{k+1}\hat{\varepsilon}_{k+1}^T u_k^{(1)}\right)^3 u_k^{(1)} \\ &\quad - \frac{1}{2}\left(u_k^{(1)T}\left(\hat{\varepsilon}_{k+1}\hat{\varepsilon}_{k+1}^T\right)^2 u_k^{(1)}\right)\hat{\varepsilon}_{k+1}\hat{\varepsilon}_{k+1}^T u_k^{(1)} \\ &\quad + \frac{3}{2}\left(u_k^{(1)T}\hat{\varepsilon}_{k+1}\hat{\varepsilon}_{k+1}^T u_k^{(1)}\right)^2 \\ &\quad \left.\times\ \hat{\varepsilon}_{k+1}\hat{\varepsilon}_{k+1}^T u_k^{(1)}\right] + \delta_k \end{aligned} \tag{44}$$

where $\|\delta_k\| \leq c_1(1/k^3)\|\hat{\varepsilon}_{k+1}\|^8$ with c_1 possibly varying in a different sample path.

By (42)

$$\nu_{k+1}^{(1)} \underset{k\to\infty}{\to} 0 \text{ a.s.} \tag{45}$$

Rewrite (43) as follows:

$$\begin{aligned} u_{k+1}^{(1)} = u_k^{(1)} &+ \frac{1}{k}\left(R_\varepsilon u_k^{(1)} - \left(u_k^{(1)T}R_\varepsilon u_k^{(1)}\right)u_k^{(1)}\right) \\ &+ \frac{1}{k}\left(\nu_{k+1}^{(1)} + \mu_{k+1}^{(1)} + \gamma_{k+1}^{(1)}\right) \end{aligned} \tag{46}$$

where

$$\begin{aligned} \gamma_{k+1}^{(1)} &= \left(\hat{\varepsilon}_{k+1}\hat{\varepsilon}_{k+1}^T - \varepsilon_{k+1}\varepsilon_{k+1}^T\right)u_k^{(1)} \\ &\quad + \left(u_k^{(1)T}\left(\varepsilon_{k+1}\varepsilon_{k+1}^T - \hat{\varepsilon}_{k+1}\hat{\varepsilon}_{k+1}^T\right)u_k^{(1)}\right)u_k^{(1)} \end{aligned} \tag{47}$$

$$\begin{aligned} \mu_{k+1}^{(1)} &= \left(\varepsilon_{k+1}\varepsilon_{k+1}^T - R_\varepsilon\right)u_k^{(1)} \\ &\quad + \left(u_k^{(1)T}\left(R_\varepsilon - \varepsilon_{k+1}\varepsilon_{k+1}^T\right)u_k^{(1)}\right)u_k^{(1)}. \end{aligned} \tag{48}$$

We may consider the truncated version of (46), but it will coincide with the untruncated version for large k since $\{\|u_k^{(1)}\|\}$ is known to be bounded by 1. Therefore, the algorithm (46) is a special form of stochastic approximation algorithm (66) given in Appendix A, where $u_k^{(1)}$, $R_\varepsilon u_k^{(1)} - (u_k^{(1)T}R_\varepsilon u_k^{(1)})u_k^{(1)}$, and $\nu_{k+1}^{(1)} + \mu_{k+1}^{(1)} + \gamma_{k+1}^{(1)}$ correspond to ϑ_k, $f(\vartheta_k)$, and ν_{k+1} in (66), respectively. Therefore, C3) is clearly satisfied.

Set

$$m(k,t) \triangleq \max\left\{m, \sum_{i=k}^{m}\frac{1}{i} < t\right\}. \tag{49}$$

We now show

$$\left\|\sum_{l=k}^{m(k,T)}\frac{1}{l}\gamma_{l+1}^{(1)}\right\| \underset{k\to\infty}{\to} 0,\ \forall T > 0. \tag{50}$$

Denote

$$S_n^{(1)} \triangleq \sum_{l=1}^{n} \|\hat{\varepsilon}_{l+1} - \varepsilon_{l+1}\|,$$
$$S_n^{(2)} \triangleq \sum_{l=1}^{n} \|\hat{\varepsilon}_{l+1}\|^2$$
$$S_n^{(3)} \triangleq \sum_{l=1}^{n} \|\varepsilon_{l+1}\|^2.$$

By (3), (17), and (18), we see that

$$S_n^{(1)} = O(\log n) \text{ a.s.} \tag{51}$$
$$\frac{S_n^{(2)}}{n} \to \mathrm{tr} R_\varepsilon \text{ a.s.} \tag{52}$$
$$\frac{S_n^{(3)}}{n} \to \mathrm{tr} R_\varepsilon \text{ a.s.} \tag{53}$$

and

$$\begin{aligned}
&\left\| \sum_{l=k}^{m(k,T)} \frac{1}{l} \gamma_{l+1}^{(1)} \right\| \\
&\leq 2 \sum_{l=k}^{m(k,T)} \frac{1}{l} \left\| \hat{\varepsilon}_{l+1}\hat{\varepsilon}_{l+1}^T - \varepsilon_{l+1}\varepsilon_{l+1}^T \right\| \\
&= \sum_{l=k}^{m(k,T)} \frac{1}{l} \left\| (\hat{\varepsilon}_{l+1} - \varepsilon_{l+1})\hat{\varepsilon}_{l+1}^T + \varepsilon_{l+1}(\hat{\varepsilon}_{l+1} - \varepsilon_{l+1})^T \right\| \\
&\leq \left(\sum_{l=k}^{m(k,T)} \frac{1}{l} \|\hat{\varepsilon}_{l+1} - \varepsilon_{l+1}\|^2 \right)^{\frac{1}{2}} \\
&\quad \times \left[\left(\sum_{l=k}^{m(k,T)} \frac{1}{l} \|\hat{\varepsilon}_{l+1}\|^2 \right)^{\frac{1}{2}} + \left(\sum_{l=k}^{m(k,T)} \frac{1}{l} \|\varepsilon_{l+1}\|^2 \right)^{\frac{1}{2}} \right].
\end{aligned} \tag{54}$$

Notice that by (31), we have

$$\begin{aligned}
\sum_{l=k}^{m(k,T)} \frac{1}{l} \|\hat{\varepsilon}_{l+1} - \varepsilon_{l+1}\|^2 &= \sum_{l=k}^{m(k,T)} \frac{1}{l} \left(S_l^{(1)} - S_{l-1}^{(1)} \right) \\
&= \frac{S_{m(k,T)}^{(1)}}{m(k,T)} - \frac{S_{k-1}^{(1)}}{k} \\
&\quad + \sum_{l=k}^{m(k,T)} S_l^{(1)} \left(\frac{1}{l-1} - \frac{1}{l} \right) \underset{k\to\infty}{\longrightarrow} 0
\end{aligned}$$

by (53)

$$\begin{aligned}
\sum_{l=k}^{m(k,T)} \frac{1}{l} \|\varepsilon_{l+1}\|^2 &= \sum_{l=k}^{m(k,T)} \frac{1}{l} \left(S_l^{(3)} - S_{l-1}^{(3)} \right) \\
&= \frac{S_{m(k,T)}^{(3)}}{m(k,T)} - \frac{S_{k-1}^{(3)}}{k} \\
&\quad + \sum_{l=k}^{m(k,T)} S_l^{(3)} \left(\frac{1}{l-1} - \frac{1}{l} \right) \underset{k\to\infty}{\longrightarrow} T \mathrm{tr} R_\varepsilon
\end{aligned}$$

and by (52), we have

$$\sum_{l=k}^{m(k,T)} \frac{1}{l} \|\hat{\varepsilon}_{l+1}\|^2 \underset{k\to\infty}{\longrightarrow} T \mathrm{tr} R_\varepsilon.$$

From these, we see that the right-hand side of (54) tends to zero as $k \to \infty$. This proves (50).

Notice that $(\mu_k^{(1)}, \mathcal{F}_k)$ is a martingale difference sequence, where $\mathcal{F}_k$ is the σ-algebra generated by $\{u_l^{(j)}, j = 1, \cdots, p, l < k\}$, and

$$\sup_k E \left(\left\| \mu_{k+1}^{(1)} \right\|^{1+\frac{\delta}{2}} |\mathcal{F}_k \right) < \infty.$$

Therefore

$$\sum_{l=1}^{\infty} \frac{1}{l} \mu_{l+1}^{(1)} < \infty \text{ a.s.} \tag{55}$$

Combining (35), (50), and (55) leads to

$$\sum_{l=k}^{m(k,T)} \frac{1}{l} \left(\nu_{l+1}^{(1)} + \mu_{l+1}^{(1)} + \gamma_{l+1}^{(1)} \right) \underset{k\to\infty}{\longrightarrow} 0 \text{ a.s.} \tag{56}$$

This verifies C2) in Appendix A.

Denote by S the unit sphere in $\mathbb{R}^p$. Then, $u_k^{(1)}$ evolves on S. Define

$$f(u) = R_\varepsilon u - (u^T R_\varepsilon u)u, \quad u \in S.$$

The root set of $f(\cdot)$ on S is

$$J \triangleq \{f_i, i = 1, \cdots, p\} \tag{57}$$

where f_i are unit eigenvectors of R_ε.

Defining

$$v(u) \triangleq -\frac{1}{2} u^T R_\varepsilon u$$

for $u \in S$, we have

$$\begin{aligned}
v_u(u) f(u) &= -u^T R_\varepsilon \left[R_\varepsilon u - (u^T R_\varepsilon u)u \right] \\
&= -u^T R_\varepsilon^2 u + (u^T R_\varepsilon u)^2 \\
&= \begin{cases} < \|R_\varepsilon u\|^2 \|u\|^2 - u^T R_\varepsilon^2 u = 0, & \text{if } u \notin J \\ 0, & \text{if } u \in J. \end{cases}
\end{aligned} \tag{58}$$

This verifies C1') in Appendix A. Thus, the convergence theorem of stochastic approximation given in Appendix A is applicable, and we conclude that $u_k^{(1)}$ converges to one of f_i, say, f_1.

It is shown in Appendix B that by induction $u_k^{(j)}$ given by (29)–(34) converge to different unit eigenvectors of R_ε.

Rewrite (35) as

$$\begin{aligned}
\lambda_{k+1}^{(j)} = &\left\{ \lambda_k^{(j)} + \frac{1}{k} \left[\lambda^{(j)} - \lambda_k^{(j)} + u_k^{(j)T} R_\varepsilon u_k^{(j)} - \lambda^{(j)} \right. \right. \\
&\left. \left. + u_k^{(j)T} \left(\hat{\varepsilon}_{k+1}\hat{\varepsilon}_{k+1}^T - R_\varepsilon \right) u_k^{(j)} \right] \right\} \\
&\times I_{\left[\left| \lambda_k^{(j)} - \frac{1}{k} \left(\lambda_k^{(j)} - u_k^{(j)T} \hat{\varepsilon}_{k+1}\hat{\varepsilon}_{k+1}^T u_k^{(j)} \right) \right| \leq M_{\sigma_k} \right]}.
\end{aligned} \tag{59}$$

We see that this is in the form of (66) with $\vartheta^* = 0$ and $f(\vartheta) = \lambda^{(j)} - \vartheta$.

Since $u_k^{(j)}$ converges and $u_k^{(j)T} R_\varepsilon u_k^{(j)} - \lambda^{(j)} \underset{k\to\infty}{\rightarrow} 0$, by the treatment similar to that used for (47)–(55), we have

$$\sum_{l=k}^{m(k,T)} \frac{1}{l} \left[u_l^{(j)T} R_\varepsilon u_l^{(j)} - \lambda^{(j)} + u_l^{(j)T} \left(\hat{\varepsilon}_{l+1} \hat{\varepsilon}_{l+1}^T - R_\varepsilon \right) u_l^{(j)} \right] \underset{k\to\infty}{\rightarrow} 0.$$

Then, by the convergence theorem of stochastic approximation given in Appendix A for linear regression functions, we have

$$\lambda_k^{(j)} \to \lambda^{(j)}, \quad j = 1, \cdots, p.$$

Then, Λ_k and U_k given by (36) have limits

$$\Lambda_k \to \begin{bmatrix} \lambda^{(1)} & & 0 \\ & \ddots & \\ 0 & & \lambda^{(p)} \end{bmatrix}, \; U_k \to [f_1, \cdots, f_p].$$

From (18), it follows that

$$\begin{aligned} \frac{1}{n} \sum_{l=1}^{n} \hat{s}_l \hat{s}_l^T &= \frac{1}{n} \sum_{l=1}^{n} \Lambda_l^{-\frac{1}{2}} U_l^T \hat{\varepsilon}_l \hat{\varepsilon}_l^T U_l \Lambda_l^{\frac{1}{2}} \\ &\to \Lambda^{-\frac{1}{2}} U^T R_\varepsilon U \Lambda^{-\frac{1}{2}} = I. \\ \frac{1}{n} \sum_{l=1}^{n} \hat{s}_l \hat{s}_{l+j}^T &= \frac{1}{n} \sum_{l=1}^{n} \Lambda_l^{-\frac{1}{2}} U_l^T \hat{\varepsilon}_l \hat{\varepsilon}_{l+j}^T U_{l+j} \Lambda_{l+j}^{-\frac{1}{2}} \\ &\to 0, \; \forall j > 0. \end{aligned}$$

Then, (38) and (39) have been proved. ■

By Theorem 1, we see that when the original source sequence $\{s_k\}$ is Gaussian, the estimated sequence $\{\hat{\varepsilon}_k\}$ is asymptotically Gaussian. Since $\{\hat{s}_k\}$ is transformed linearly from $\{\hat{\varepsilon}_k\}$, by (37) and (40), $\{\hat{s}_k\}$ also is asymptotically Gaussian.

IV. Simulation Results

In this section, we consider a numerical example to illustrate our approach. For generating the signal $\{s_k\}$, we proceed as follows: Take a random sequence consisting of 0 and 1, and encode the sequence by a turbo encoder. The coded bits $\{b_k\}$ are interleaved and passed through a serial to parallel (S/P) converter. Then, they are fed into p transmission paths corresponding to p transmitter antennas. At each path, the modulator maps each of its inputs into one point of a quadrature phase-shift keying (QPSK) constellation. The output of the modulator serves as the channel input used in our example, i.e., $\{s_k\}$. Let the matrix polynomial $H(z)$ characterizing the channel be given by

$$H(z) = \begin{bmatrix} 1 & \frac{2}{3} & -1 \\ -\frac{1}{2} & 1 & \frac{7}{4} \\ 1 & -\frac{4}{5} & 0 \\ 0 & 1 & -\frac{2}{3} \end{bmatrix} + \begin{bmatrix} \frac{1}{5} & -\frac{2}{5} & \frac{1}{2} \\ \frac{2}{5} & -\frac{1}{5} & \frac{3}{5} \\ -\frac{3}{5} & \frac{8}{25} & \frac{3}{10} \\ -\frac{3}{10} & \frac{3}{5} & \frac{2}{7} \end{bmatrix} z + \begin{bmatrix} \frac{1}{8} & -\frac{2}{7} & \frac{1}{5} \\ -\frac{2}{9} & \frac{1}{6} & -\frac{3}{7} \\ -\frac{3}{8} & -\frac{8}{25} & -\frac{3}{11} \\ \frac{3}{11} & \frac{3}{8} & \frac{2}{13} \end{bmatrix} z^2 \quad (60)$$

with $p = 3$, $q = 4$, $r = 2$, and the observations by

$$y_k = x_k + n_k = H(z)s_k + n_k \quad (61)$$

where the noise $\{n_k\}$ is an i.i.d. Gaussian sequence with $En_k = 0$, $En_k n_k^T = \sigma^2 I$.

The NW method developed in this paper is compared with the LPD method proposed in [11], by which a matrix polynomial $W(z)$ is first derived such that $W(z)H(z) = H_0$. In this example, we set the order of $W(z)$ to 8. Then, the estimate for ε_k is defined as

$$\hat{\varepsilon}_k = W(z)y_k = H_0 s_k + W(z)n_k. \quad (62)$$

To evaluate a whitening or deconvolution method, we use the following performance indices: the mean-squared error (MSE) of the estimate for $H_0 s_k$, the ISI, which characterizes the whiteness of $\hat{\varepsilon}_k$ in time, and the component correlatedness (CC) of $\hat{s}_k$ in space. They are defined, respectively, by

$$\text{MSE} = \frac{\sum_{k=1}^{n} \|\hat{\varepsilon}_k - H_0 s_k\|^2}{\sum_{k=1}^{n} \|H_0 s_k\|^2} \quad (63)$$

$$\text{ISI} = \max_{\rho, j=1,2} \left| \rho \left(\frac{n \sum_{k=1}^{n-j} \hat{\varepsilon}_k \hat{\varepsilon}_{k+j}^T}{(n-j) \sum_{k=1}^{n} \hat{\varepsilon}_k^T \hat{\varepsilon}_k} \right) \right| \quad (64)$$

and

$$\text{CC} = \max_{\rho} \left| \rho \left(\frac{1}{n} \sum_{k=1}^{n} \hat{s}_k \hat{s}_k^T - I \right) \right| \quad (65)$$

where $\rho(A)$ denotes the eigenvalue of a matrix A. The efficiency of the whitening method proposed in the paper is measured by the bit-error rate (BER).

When applying NW, in order to keep ε_k as $H_0 s_k$, we take

$$L^{(1)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \end{bmatrix} \text{ and } L^{(2)} = \begin{bmatrix} 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The data to be used are $\eta_k^{(i)}$, $i = 1, 2$

$$\eta_k^{(i)} \triangleq L^{(i)} y_k = L^{(i)}(x_k + n_k), \quad i = 1, 2.$$

In Appendix C, it is shown that $\eta_k^{(i)}$ has the innovation representation [13]

$$\eta_{k+1}^{(i)} = w_{k+1}^{(i)} + G_1^{(i)} w_k^{(i)} + \cdots + G_r^{(i)} w_{k-r+1}^{(i)}.$$

Let $D(z) \equiv I + D_1 z$ in (11), and treat $\eta_k^{(i)}$ and $w_k^{(i)}$ as x_k^L and ε_k in (11), respectively. By (13)–(16), we derive $\hat{w}_k^{(i)}$. The initial values for (14) and (15) are $\theta_0 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ and $P_0 = 0.2I$, respectively. In order to reduce the effect caused by inaccuracy of initial values, each frame with length of 1200 bits is used twice to go through (13)–(16): The estimates θ_k and P_k obtained at the end of the first round of computation serve as the initial values of the second round of computation, and the estimates $\hat{w}_k^{(i)}$, $i = 1, 2$ derived from the second round of computation are used for comparison. Define $\hat{w}_k = L^{(1)T} \hat{w}_k^{(1)} + L^{(2)T} \hat{w}_k^{(2)}$ to serve as the estimate $\hat{\varepsilon}_k$ for ε_k. Putting the computed $\hat{\varepsilon}_k$ into

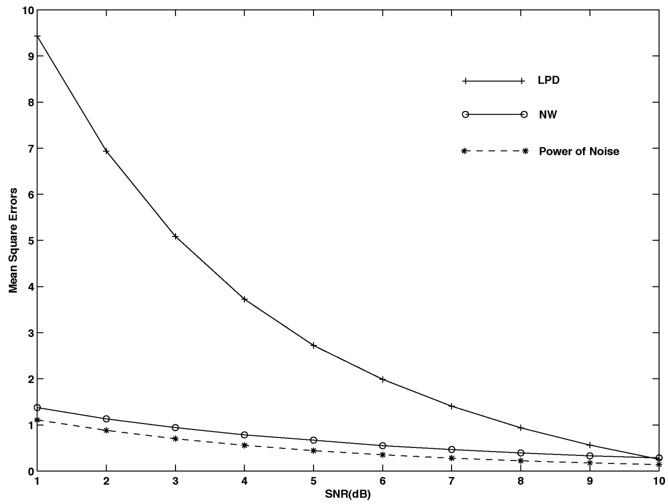


Fig. 1. MSEs.

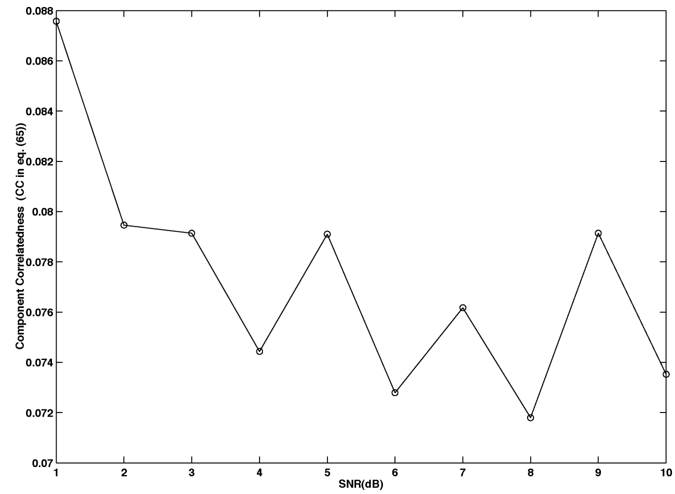


Fig. 3. Components correlatedness.

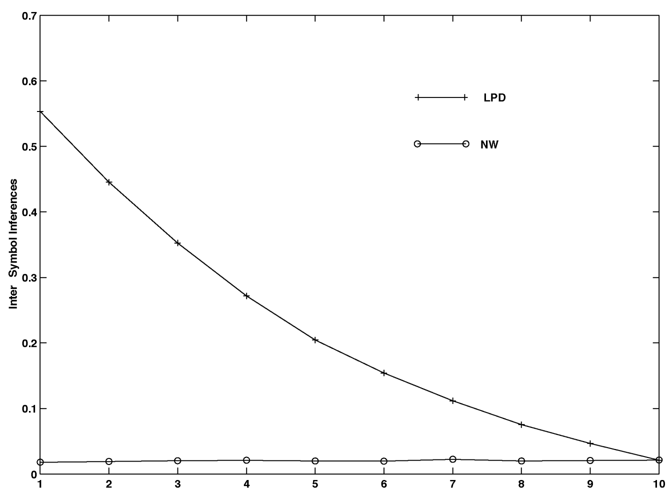


Fig. 2. ISIs.

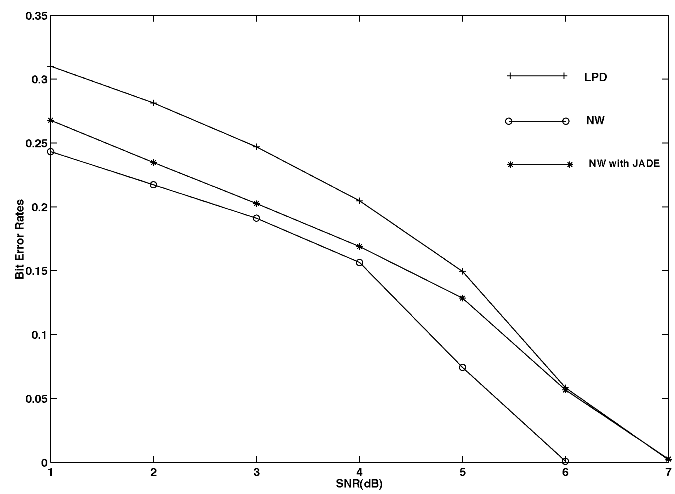


Fig. 4. BERs.

(63) and (64) immediately yields MSE and ISI. In order to derive CC, we replace $\hat{\varepsilon}_{k+1}$ in (29)–(37) by $w_{k+1}^{(1)}$ and derive a whitened sequence $\{\hat{s}_k\}$ in space, which gives $CC^{(1)}$ according to (65). Similarly, by $\hat{w}_{k+1}^{(2)}$, we obtain $CC^{(2)}$. To demonstrate the efficiency of the NW method, we elect CC in the worse case, i.e., $CC \triangleq \max(CC^{(1)}, CC^{(2)})$, and illustrate it in Fig. 3.

In Figs. 1–3, the performance indices are plotted as functions of signal-to-noise ratio (SNR), where the performance indices are obtained by 100 *Monte Carlo* runs, while SNR is defined as

$$\mathrm{SNR} = 10\log_{10}\left\{\frac{\sum_{i=1}^{q} E\left\{\left|x_k^i\right|^2\right\}}{\sum_{i=1}^{q} E\left\{\left|n_k^i\right|^2\right\}}\right\}.$$

In Figs. 1–3, the lines with cycles are given by NW, while the lines with plus signs are given by LPD; the dashed line with asterisks in Fig. 1 is the relative power of noise, i.e., $E\{\|n_k\|^2\}/E\{\|H_0 s_k\|^2\}$. Figs. 1 and 2 show that the method NW proposed in the paper gives better results over the LPD method, especially for low SNR. Fig. 3 demonstrates that the NW provides a good performance of recursive whitening in space.

The overall efficiency of our whitening method is measured by BER. The BERs of the recovered signal are computed when LPD and NW are used for deconvoluting the system. The lines with cycles and plus signs in Fig. 4 are computed on the "best case scenario," i.e., for the case where the channel matrix H_0 is assumed to be known. From Fig. 4, it is seen that NW gives better results than LPD. For the case where H_0 is unknown, we apply the signal rotation part of the JADE method [2], [3] to the signal whitened in space according to NW. The JADE algorithm is taken here because it is a simple algorithm and is widely applied in BSS. The resulting signal differs from the true signal s_k by a multiple being a diagonal matrix with diagonal elements $e^{i\theta_j}$, $j = 1, \cdots, q$, where q is the dimension of s_k. By using the property that each component of s_k is of the form $\pm(\sqrt{2}/2) \pm i(\sqrt{2}/2)$, the estimates for θ_j, $j = 1, \cdots, q$, are then obtained by the least-squares method. For this case, BER is shown by the line with asterisks in Fig. 4. Therefore, even in the case where H_0 is unknown, the proposed method NW still works well.

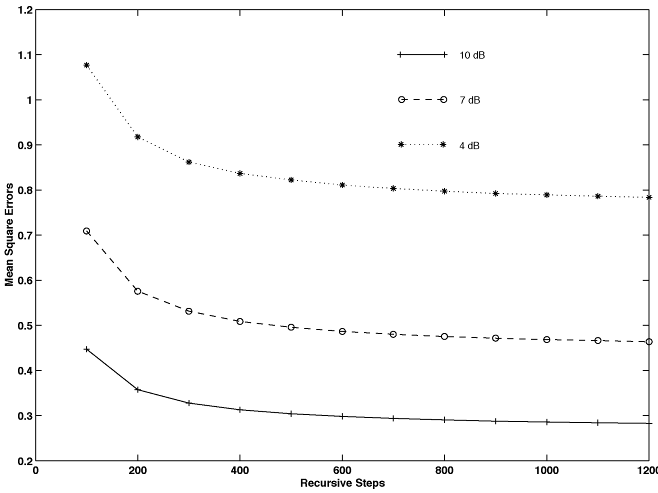


Fig. 5. MSEs.

Concerning the complexity of LPD and NW, they are not comparable, in general, since LPD is a batch algorithm, while NW is an adaptive one. The computation of NW is distributed to each step when a new received sample comes. The LPD collects a certain number of received samples before whitening. The computation of NW is proportional to the length of the observed signal sequence. For a given signal sequence that is long enough to have an acceptable estimate, the overall computation of NW is larger than that of LPD. Fig. 5 shows the behaviors of the MSE as the number of steps in our recursive algorithm grows up for different levels of the noise.

V. Conclusion

The paper proposes a recursive direct method for MIMO channels to whiten the output signal in both time and space. Unlike the conventional deconvolution methods that normally construct a weighting matrix $W(z)$ to form a linear filter acting on the past output data to produce the estimate for input signal, the estimate for input signal proposed in the paper is a highly nonlinear function of the past output data. From (89), it can be seen that the nonlinear method helps us in suppressing noise in an optimal way, while for all linear methods, the influence of noise is neglected during designing $W(z)$, and as a result, the noise term $W(z)n_k$ additively appears in the estimate $\hat{\varepsilon}_k$ for $H_0 s_k$ [see (62)]. This explains why our nonlinear method is better than linear methods.

The convergence of the algorithm is proved in the paper. Numerical simulation demonstrates that the proposed method works very well when the observation is corrupted by noise. As a matter of fact, in this case, the noisy output can still be expressed as the output of an MIMO channel without noise by using the innovation representation. The simulation results show that the observation noise does not affect too much on the innovation representation, especially for cases of large SNR, and that ignoring such an effect leads to BERs much less than those when simply ignoring the existence of noise in the LPD method.

Although the proof of the method assumes that the channel is time invariant, it is expected to work well for a slowly changing MIMO channel. Unlike the batch methods, the algorithm proposed in this paper is recursive and updates at each sample. It is an inherent property of recursive algorithms that they may adapt to the change in system parameters if the change is slower than the convergence speed of the algorithms. Intuitively, the recursive algorithms spread the whitening process into each time when the signal sample received, thus, is expected to give a more accurate estimate than the batch method when the channel varies slowly.

The key step when applying the NW method given in the paper is to adequately select a matrix L satisfying A2) and a sufficiently large m satisfying (7). One may first assume that there exists a stable submatrix of $H(z)$. If the stable submatrix is available, then L can be taken such that each its column has only one nonzero element that equals 1 and corresponds to the row of $H(z)$ that should be selected. If only the existence of a stable submatrix is known but the submatrix itself is unknown, then we may try different L with columns having only one nonzero (equal to 1) element. The total number of such matrices is $C_p^q = [p\cdot(p-1)\cdots(p-q)]/[q\cdot(q-1)\cdots 1]$, where p is the number of rows of $H(z)$, and q is the number of columns of $H(z)$. So, after at most C_p^q trials, the desired L will be obtained. However, in general, it may happen that $LH(z)$ is stable for some matrix L, even if there is no stable submatrix of $H(z)$. In this case, we do not have a general way to define L. At present, we need work by "trial and error." The development of a practical algorithm for selecting L is a future research topic. Besides, to quantitatively analyze the effect of the observation noise is also of interest for further research.

Appendix A
Convergence Theorem of Stochastic Approximation

For the stochastic approximation algorithm with expanding truncations

$$\begin{aligned}\vartheta_{k+1} = &\left(\vartheta_k + \frac{1}{k}\left(f(\vartheta_k)+\nu_{k+1}\right)\right)\\ &\times I_{\left[\left\|\vartheta_k+\frac{1}{k}(f(\vartheta_k)+\nu_{k+1})\right\|\le M_{\sigma_k}\right]}\\ &+\vartheta^* I_{\left[\left\|\vartheta_k+\frac{1}{k}(f(\vartheta_k)+\nu_{k+1})\right\|> M_{\sigma_k}\right]} \end{aligned}\tag{66}$$

$$\begin{aligned}\sigma_k &= \sum_{i=1}^{k-1} I_{\left[\left\|\vartheta_i+\frac{1}{i}(f(\vartheta_k)+\nu_{k+1})\right\|> M_{\sigma_i}\right]}\\ \sigma_0 &= 0\end{aligned}\tag{67}$$

assume that the following conditions hold.

C1) There is a continuously differentiable function $v(\cdot): \mathbb{R}^l \to \mathbb{R}$ such that

$$\sup_{\delta\le d(\vartheta,J)\le\Delta} f^T(\vartheta)v_\vartheta(\vartheta) < 0$$

for any $\Delta > \delta > 0$, where J is the zero set of $f(\cdot)$ consisting of isolated points. $d(\vartheta, J) = \inf_\phi\{\|\vartheta-\phi\| : \phi\in J\}$ and $v_\vartheta(\cdot)$ denote the gradient of $v(\cdot)$. Further, ϑ^* used in (66) is such that $v(\vartheta^*) < \inf_{\|\vartheta\|=c_0} v(\vartheta)$ for some $c_0 > 0$ and $\|\vartheta^*\| < c_0$.

C2) For the sample path under consideration

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

for any $\{n_k\}$ such that ϑ_{n_k} converges, where $m(k,T)$ is given by (49).

C3) $f(\cdot)$ is measurable and locally bounded.

Then, $d(\vartheta_k, J) \to 0$ for any given initial value ϑ_0 for the sample path for which C2) holds (see [6, Th. 2.2.1]).

Remark 5: If it is known that $\{\vartheta_k\}$ evolves in a subspace S of $\mathbb{R}^l$, then C1) can be weakened to C1'). There is a continuously differentiable function $v(\cdot): \mathbb{R}^l \to \mathbb{R}$ such that

$$\sup_{\substack{\delta \le d(\vartheta, J\cap S)\le \Delta\\ \vartheta\in S}} f^T(\vartheta) v_\vartheta(\vartheta) < 0$$

for any $\Delta > \delta > 0$ (see Remark 2.2.6 in [6]).

Appendix B

Inductively assume

$$P_k^{(i-1)} u_k^{(i)} \to f_i, \quad i = 1, \cdots, j-1 \tag{68}$$

and denote

$$V^{(i)} \triangleq [f_1 \cdots f_i],\ P^{(i)} \triangleq I - V^{(i)} V^{(i)T},\ P^{(0)} = I. \tag{69}$$

By (68), we have

$$V_k^{(i)} \underset{k\to\infty}{\to} V^{(i)},\ P_k^{(i)} \to P^{(i)},\ i = 1, \cdots, j-1 \tag{70}$$

and

$$P^{(i-1)} f_i = f_i, \quad i = 1, \cdots, j-1. \tag{71}$$

By (42) and (68), it follows that

$$\tilde{u}_k^{(i)} \to f_i, \quad i = 1, \cdots, j-1 \tag{72}$$

and by (34)

$$u_k^{(i)} \to f_i, \quad i = 1, \cdots, j-1. \tag{73}$$

Since

$$\begin{aligned} V_{k+1}^{(1)} V_{k+1}^{(1)T} - V_k^{(1)} V_k^{(1)T} &= \left(V_{k+1}^{(1)} - V_k^{(1)}\right) V_{k+1}^{(1)T} + V_k^{(1)} \left(V_{k+1}^{(1)T} - V_k^{(1)T}\right) \\ &= \left(u_{k+1}^{(1)} - u_k^{(1)}\right) u_{k+1}^{(1)T} + u_k^{(1)} \left(u_{k+1}^{(1)T} - u_k^{(1)T}\right) \end{aligned}$$

and from (46)

$$u_{k+1}^{(1)} - u_k^{(1)} = o\left(\frac{1}{k}\right) + \frac{1}{k}\left(\nu_{k+1}^{(1)} + \mu_{k+1}^{(1)} + \gamma_{k+1}^{(1)}\right)$$

we see

$$V_{k+1}^{(1)} V_{k+1}^{(1)T} - V_k^{(1)} V_k^{(1)T} = o\left(\frac{1}{k}\right) + \frac{1}{k}\delta_{k+1}^{(1)} \tag{74}$$

where

$$\delta_{k+1}^{(1)} = \left(\nu_{k+1}^{(1)} + \mu_{k+1}^{(1)} + \gamma_{k+1}^{(1)}\right) u_{k+1}^{(1)T} + u_k^{(1)} \left(\nu_{k+1}^{(1)} + \mu_{k+1}^{(1)} + \gamma_{k+1}^{(1)}\right)^T.$$

Since $u_k^{(1)}$ converges, by (56), it follows that

$$\sum_{l=k}^{m(k,T)} \frac{1}{l}\delta_{l+1}^{(1)} \underset{k\to\infty}{\to} 0 \text{ a.s.} \tag{75}$$

Together with (68), inductively we also assume

$$V_{k+1}^{(i)} V_{k+1}^{(i)T} - V_k^{(i)} V_k^{(i)T} = o\left(\frac{1}{k}\right) + \frac{1}{k}\delta_{k+1}^{(i)},\ i = 1, \cdots, j-1 \tag{76}$$

and

$$\sum_{l=k}^{m(k,T)} \frac{1}{l}\delta_{l+1}^{(i)} \underset{k\to\infty}{\to} 0 \text{ a.s.} \quad i = 1, \cdots, j-1. \tag{77}$$

We now show that $u_k^{(i)}$ converges to one of the unit eigenvectors contained in

$$J \setminus \{f_1, \cdots, f_{j-1}\}$$

and that (76) and (77) hold also for $i = j$.

By definition

$$\tilde{u}_{k+1}^{(j)} = P_k^{(j)} u_k^{(j)} + \frac{1}{k} P_k^{(j-1)} \hat{\varepsilon}_{k+1} \hat{\varepsilon}_{k+1}^T P_k^{(j-1)} u_k^{(j)} \tag{78}$$

$$u_{k+1}^{(j)} = \frac{\tilde{u}_{k+1}^{(j)}}{\left\|\tilde{u}_{k+1}^{(j)}\right\|}, \text{ if } \left\|\tilde{u}_{k+1}^{(j+1)}\right\| \ge \varepsilon. \tag{79}$$

Since the last term in (78) tends to zero and $P_{k+1}^{j-1} \to P^{(j-1)}$, we need only to reset $u_k^{(j)}$ to a new $\bar{u}_k^{(j)}$ with $\|\bar{u}_k^{(j)}\| = 1$ and $\|P_k^{(j-1)} \bar{u}_k^{(j)}\| = 1$ at most for a finite number of times.

Replacing $u_k^{(1)}$ by $P_k^{(j-1)} u_k^{(j)}$ in (41)–(48), we arrive at the following recursions corresponding to (46)–(48):

$$\begin{aligned} u_{k+1}^{(j)} = u_k^{(j)} &+ \frac{1}{k}\Big[P_k^{(j-1)} R_\varepsilon P_k^{(j-1)} u_k^{(j)} \\ &\quad - \left(u_k^{(j)T} P_k^{(j-1)} R_\varepsilon P_k^{(j-1)} u_k^{(j)}\right) P_k^{(j-1)} u_k^{(j)}\Big] \\ &+ \frac{1}{k}\left(\nu_{k+1}^{(j)} + \mu_{k+1}^{(j)} + \gamma_{k+1}^{(j)}\right) \end{aligned} \tag{80}$$

where $\nu_{k+1}^{(j)} \underset{k\to\infty}{\to} 0$

$$\begin{aligned} \gamma_{k+1}^{(j)} &= \left(\hat{\varepsilon}_{k+1}\hat{\varepsilon}_{k+1}^T - \varepsilon_{k+1}\varepsilon_{k+1}^T\right) P_k^{(j-1)} u_k^{(j)} \\ &\quad + \left[u_k^{(j)T} P_k^{(j-1)} \left(\varepsilon_{k+1}\varepsilon_{k+1}^T - \hat{\varepsilon}_{k+1}\hat{\varepsilon}_{k+1}^T\right) P_k^{(j-1)} u_k^{(j)}\right] \\ &\quad \times P_k^{(j-1)} u_k^{(j)} \\ \mu_{k+1}^{(j)} &= \left(\hat{\varepsilon}_{k+1}\hat{\varepsilon}_{k+1}^T - R_\varepsilon\right) P_k^{(j-1)} u_k^{(j)} \\ &\quad + \left[u_k^{(j)T} P_k^{(j-1)} \left(R_\varepsilon - \hat{\varepsilon}_{k+1}\hat{\varepsilon}_{k+1}^T\right) P_k^{(j-1)} u_k^{(j)}\right] \\ &\quad \times P_k^{(j-1)} u_k^{(j)}. \end{aligned}$$

Similar to (50) and (55), we have

$$\sum_{i=k}^{m(k,T)} \frac{1}{i}\gamma_{i+1}^{(j)} \underset{k\to\infty}{\to} 0 \text{ a.s.}$$

and

$$\sum_{i=1}^{\infty} \frac{1}{i}\mu_{i+1}^{(j)} < \infty \text{ a.s.}$$

Noticing $P_k^{(j-1)}P_k^{(j-1)} = P_k^{(j-1)}$ and using (70), we can rewrite (80)

$$\begin{aligned} P_k^{(j-1)}u_{k+1}^{(j)} &= P_k^{(j-1)}u_k^{(j)} - V_{k+1}^{(j-1)}V_{k+1}^{(j-1)T}P_k^{(j-1)}u_k^{(j)} + \frac{1}{k}P^{(j-1)} \\ &\times \Big[P^{(j-1)}R_\varepsilon P^{(j-1)}P_k^{(j-1)}u_k^{(j)} \\ &\quad - \left(u_k^{(j)T}P_k^{(j-1)}P^{(j-1)}R_\varepsilon P^{(j-1)}P_k^{(j-1)}u_k^{(j)}\right) \\ &\quad \times P_k^{(j-1)}u_k^{(j)}\Big] + \frac{1}{k}\left(o(1) + \mu_{k+1}^{(j)} + \gamma_{k+1}^{(j)}\right) \end{aligned} \tag{81}$$

where we have used the assumption

$$P_k^{(j-1)} \to P^{(j-1)}.$$

Denote the second term on the right-hand side of (81) by

$$\begin{aligned} \frac{1}{k}\beta_{k+1}^{(j)} &\triangleq -V_{k+1}^{(j-1)}V_{k+1}^{(j-1)T}P_k^{(j-1)}u_k^{(j)} \\ &= -V_{k+1}^{(j-1)}V_{k+1}^{(j-1)T} \\ &\quad \times \left(I - V_{k+1}^{(j-1)}V_{k+1}^{(j-1)T} + V_{k+1}^{(j-1)}V_{k+1}^{(j-1)T} \right. \\ &\quad \left. - V_k^{(j-1)}V_k^{(j-1)T}\right)u_k^{(j)} \\ &= -V_{k+1}^{(j-1)}V_{k+1}^{(j-1)T} \\ &\quad \times \left(V_{k+1}^{(j-1)}V_{k+1}^{(j-1)T} - V_k^{(j-1)}V_k^{(j-1)T}\right). \end{aligned}$$

By (70), (76), and (77), we have

$$\frac{1}{k}\beta_{k+1}^{(j)} = o\left(\frac{1}{k}\right) + \frac{1}{k}\delta_{k+1}^{(j)} \tag{82}$$

and

$$\sum_{i=k}^{m(k,T)} \frac{1}{i}\beta_{i+1}^{(j)} \underset{k\to\infty}{\to} 0 \text{ a.s.} \tag{83}$$

Setting $z_k^{(j)} = P_k^{(j-1)}u_k^{(j)}$, from (81)–(83), we see

$$\begin{aligned} z_{k+1}^{(j)} &= z_k^{(j)} + \frac{1}{k}P^{(j-1)} \\ &\quad \times \Big[P^{(j-1)}R_\varepsilon P^{(j-1)}z_k^{(j)} \\ &\quad\quad - \left(z_k^{(j)T}P^{(j-1)}R_\varepsilon P^{(j-1)}z_k^{(j)}\right)z_k^{(j)}\Big] \\ &\quad + \frac{1}{k}\left(\mu_{k+1}^{(j)} + \gamma_{k+1}^{(j)} + \beta_{k+1}^{(j)} + o(1)\right). \end{aligned} \tag{84}$$

Again, applying the convergence theorem of stochastic approximation given in Appendix A, we can prove the convergence of $z_k^{(i)}$ to an unit eigenvector of $P^{(j-1)}R_\varepsilon P^{(j-1)}$.

By (32) and (33), $u_k^{(j)}$ and, hence, $u_k^{(j)}$ converges: $u_k^{(j)} \to u^{(j)}$. Thus

$$z_k^{(j)} \to P^{(j-1)}u^{(j)}.$$

From (33), we have

$$\begin{aligned} V^{(j-1)}V^{(j-1)T}u_{k+1}^{(j)} &= V^{(j-1)}V^{(j-1)T}P_k^{(j-1)}u_k^{(j-1)} \\ &\quad + \frac{1}{k}V^{(j-1)}V^{(j-1)T}P_k^{(j-1)}\hat{\varepsilon}_{k+1}\hat{\varepsilon}_{k+1}^T \\ &\quad \times P_k^{(j-1)}u_k^{(j)} \underset{k\to\infty}{\to} 0 \end{aligned}$$

and by (34)

$$V^{(j-1)}V^{(j-1)T}u_{k+1}^{(j)} \underset{k\to\infty}{\to} 0.$$

This incorporated with $u_k^{(j)} \to u^{(j)}$ leads to

$$V^{(j-1)}V^{(j-1)T}u^{(j)} = 0 \text{ or } P^{(j-1)}u^{(j)} = u^{(j)}. \tag{85}$$

Since the limit of $z_k^{(j)}$, $P^{(j-1)}u^{(j)}$, is an unit eigenvector of $P^{(j-1)}R_\varepsilon P^{(j-1)}$, we have

$$\left[P^{(j-1)}R_\varepsilon P^{(j-1)} - \left(u^{(j)}R_\varepsilon u^{(j)}\right)\right]u^{(j)} = 0$$

or

$$P^{(j-1)}R_\varepsilon u^{(j)} - \left(u^{(j)}R_\varepsilon u^{(j)}\right)u^{(j)} = 0. \tag{86}$$

From (85), it follows that $u^{(j)}$ can be expressed by a linear combination of eigenvectors $f_j, \cdots, f_p$. Consequently

$$P^{(j-1)}R_\varepsilon u^{(j)} = R_\varepsilon u^{(j)}$$

which, combined with (86), implies

$$R_\varepsilon u^{(j)} = \left(u^{(j)}R_\varepsilon u^{(j)}\right)u^{(j)}.$$

This means that $u^{(j)}$ is an eigenvector of R_ε, and $u^{(j)}$ is different from $f_1, \cdots, f_{j-1}$ by (85). Thus, we have shown (68) for $i = j$.

Since

$$P^{(j-1)}R_\varepsilon P^{(j-1)}z_k^{(j)} - \left(z_k^{(j)T}P^{(j-1)}R_\varepsilon P^{(j-1)}z_k^{(j)}\right)z_k^{(j)} \underset{k\to\infty}{\to} 0$$

from (84), we have

$$z_{k+1}^{(i)} - z_k^{(i)} = o\left(\frac{1}{k}\right) + \frac{1}{k}\alpha_{k+1}^{(i)}, \quad i = 1, \cdots, j \tag{87}$$

where

$$\alpha_{k+1}^{(i)} = \mu_{k+1}^{(i)} + \gamma_{k+1}^{(i)} + \beta_{k+1}^{(i)} + o(1) \tag{88}$$

and

$$\sum_{l=k}^{m(k,T)} \frac{1}{l}\alpha_{l+1}^{(i)} \underset{k\to\infty}{\to} 0,\ \forall T > 0.$$

Elementary manipulation leads to

$$\begin{aligned}
&V_{k+1}^{(j)} V_{k+1}^{(j)T} - V_k^{(j)} V_k^{(j)T} \\
&= \left(V_{k+1}^{(j)} - V_k^{(j)}\right) V_{k+1}^{(j)T} + V_k^{(j)} \left(V_{k+1}^{(j)T} - V_k^{(j)T}\right) \\
&= \left(V_{k+1}^{(j)} - V_k^{(j)}\right) V_{k+1}^{(j)T} + V_k^{(j)} \left(V_{k+1}^{(j)T} V_{k+1}^{(j)}\right)^{-1} V_{k+1}^{(j)T} \\
&\quad - V_k^{(j)} \left(V_k^{(j)T} V_k^{(j)}\right)^{-1} V_k^{(j)T} \\
&= \left(V_{k+1}^{(j)} - V_k^{(j)}\right) V_{k+1}^{(j)T} + V_k^{(j)} \left(V_{k+1}^{(j)T} V_{k+1}^{(j)}\right)^{-1} \\
&\quad \times \left(V_k^{(j)T} V_k^{(j)} - V_{k+1}^{(j)T} V_{k+1}^{(j)}\right) V_k^{(j)T} \\
&= \left(V_{k+1}^{(j)} - V_k^{(j)}\right) V_{k+1}^{(j)T} + V_k^{(j)} \left(V_{k+1}^{(j)T} V_{k+1}^{(j)}\right)^{-1} \\
&\quad \times \left[\left(V_k^{(j)T} - V_{k+1}^{(j)T}\right) V_k^{(j)} + V_{k+1}^{(j)T} \left(V_k^{(j)} - V_{k+1}^{(j)}\right)\right] V_k^{(j)T}.
\end{aligned}$$

This equation, together with (87) and (88), proves (76) and (77) for $i = j$.

Appendix C

Consider the case where the received signal is corrupted by an additive noise n_k, which is uncorrelated with $\{s_k\}$ with $En_k = 0$, $En_k n_{k+j}^T = 0, \forall j > 0$, $En_k n_k^T = R \,\forall k$.

Denote

$$\eta_k = L(x_k + n_k)$$

i.e.,

$$\eta_k = LH_0 s_k + LH_1 s_{k-1} + \cdots + LH_r s_{k-r} + Ln_k.$$

As in Section II [see (4) and (5)], setting $\varepsilon_k = LH_0 s_k$

$$C(z) = I + C_1 z + \cdots + C_r z^r,\ C_i = LH_i (LH_0)^{-1}$$

we have

$$\eta_k = \varepsilon_k + C_1 \varepsilon_{k-1} + \cdots + C_r \varepsilon_{k-r} + Ln_k.$$

Comparing with (5), we find that when the signal is received without noise the signal x_k^L to be whitened is an MA process, while here we want to whiten η_k, which consists of not only the MA part but also an exogenous input $L\eta_k$. In what follows, we show that by using the innovation representation, η_k can still be expressed as an MA process. Let

$$\zeta_k = \begin{bmatrix} \varepsilon_k \\ Ln_k \end{bmatrix}$$

$$A = \begin{bmatrix} 0 & I & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & & \cdots & I \\ 0 & & \cdots & 0 \end{bmatrix}$$

$$B = \begin{bmatrix} I & 0 \\ C_1 & 0 \\ \vdots & \vdots \\ C_r & 0 \end{bmatrix}$$

$$C = \underbrace{[I, 0, \cdots, 0]}_{(r+1)p}.$$

Then, we have the state-space representation of $\{\eta_k\}$

$$\begin{aligned} \xi_{k+1} &= A\xi_k + B\zeta_{k+1} \\ \eta_k &= C\xi_k + [0,\ I]\zeta_k. \end{aligned}$$

If (A, B, C) is controllable and observable, then for $\{\zeta_k\}$ being any uncorrelated sequence with $E\zeta_k = 0$ and $E\zeta_k \zeta_k^T = R$, it is well known [14] that the Kalman filter gain K_k converges to a limit: $K_k \underset{k\to\infty}{\rightarrow} K < \infty$. Further, if $\{\zeta_k\}$ is Gaussian, then innovation of $\{\eta_k\}$

$$w_k \triangleq \eta_k - E\left(\eta_k | \mathcal{F}_{k-1}^\eta\right) \tag{89}$$

is an i.i.d. sequence with $Ew_k = 0$ and $\mathcal{F}_k^w = \mathcal{F}_k^\eta$, where $\mathcal{F}_k^w$ denotes $\sigma\{w_1, \cdots, w_k\}$, and $\mathcal{F}_k^\eta$ is defined in a similar way.

Using the innovation property of residuals in the steady-state Kalman filter, we derive the innovation representation [13]

$$\eta_{k+1} = w_{k+1} + G_1 w_k + \cdots + G_r w_{k-r+1} \tag{90}$$

where $G_i = CA^i K$. Instead of A2), we now assume

$$G(z) = I + G_1 z + \cdots + G_r z^r \tag{91}$$

is stable. Corresponding to (5), we now have

$$\eta_k = G(z) w_k. \tag{92}$$

Since $\{w_i\}$ is Gaussian, we have $E\|w_k\|^l < \infty$ for any $l > 0$. Then, $(1/n)\sum_{k=1}^n \|w_k\|^l \to E\|w_n\|^l$. Consequently

$$\frac{1}{n}\|w_n\|^l \underset{n\to\infty}{\rightarrow} 0, \text{ for any } l > 0.$$

Therefore, if in addition to A1), $\{\zeta_k\}$ is an i.i.d. Gaussian sequence and $G(z)$ defined by (91) is stable, then the same method as that used in Theorems 1 and 2 can still be applied to whiten $\{\eta_k\}$, which corresponds to $\{x_k^L\}$ in (5).

References

[1] K. Abed-Meraim, J. Cardoso, A. Y. Gorokhov, P. Loubaton, and E. Moulines, "On subspace methods for blind identification of single-input multiple-output FIR systems," *IEEE Trans. Signal Process.*, vol. 45, no. 1, pp. 42–55, Jan. 1997.

[2] J. Cardoso and A. Souloumiac, "Blind beamforming for non Gaussian signals," *Proc. Inst. Elect. Eng. F, Radar Signal Process.*, vol. 140, no. 6, pp. 362–370, Dec. 1993.

[3] J. Cardoso, "High-order contrasts for independent components analysis," *Neural Computat.*, vol. 11, pp. 157–192, 1999.

[4] X. R. Cao and R. Liu, "General approach to blind signal separation," *IEEE Trans. Signal Process.*, vol. 44, no. 3, pp. 562–571, Mar. 1996.

[5] B. Chen and A. P. Petropulu, "Frequency domain blind MIMO system identification based on second- and higher order statistics," *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1677–1688, Aug. 2001.

[6] H. F. Chen, *Stochastic Approximation and Its Application*. Dordrecht, The Netherlands: Kluwer, 2002.

[7] H. F. Chen and L. Guo, *Identification and Stochastic Adaptive Control*. Cambridge, MA: Birkhäuser, 1991.

[8] A. Chevreuil and P. Loubaton, "MIMO blind second-order equalization method and conjugate cyclostationarity," *IEEE Trans. Signal Process.*, vol. 47, no. 2, pp. 572–578, Feb. 1999.

[9] D. Gesbert and P. Duhamel, "Robust blind identification and equalization based on multi-step predictions," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 5, 1997, pp. 3621–3624.

[10] A. Gorokhov and P. Loubaton, "Subspace-based techniques for blind separation of convolutive mixtures with temporally correlated sources," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 44, no. 9, pp. 813–820, Sep. 1997.

[11] Y. Inouye and R. W. Liu, "A system-theoretic foundation for blind equalization for an FIR MIMO channel system," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 49, no. 4, pp. 425–436, Apr. 2002.

[12] Y. Inouye and T. Sato, "Iterative algorithms based on multistage criteria for multichannel blind deconvolution," *IEEE Trans. Signal Process.*, vol. 47, no. 6, pp. 1759–1764, Jun. 1999.

[13] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Englewood Cliffs, NJ: Prentice-Hall, 2000.

[14] R. S. Liptser and A. N. Shiryayev, *Statistics of Random Processes*. New York: Springer-Verlag, 1977.

[15] R. Lopez-Valcarece, Z. Ding, and S. Dasgupta, "Equalization and interference cancellation in linear multiuser systems based on second-order statistics," *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 2042–2049, Sep. 2001.

[16] C. T. Ma, Z. Ding, and S. F. Yau, "A two stage algorithm for MIMO blind deconvolution of nonstationary colored signals," *IEEE Trans. Signal Process,*, vol. 48, no. 4, pp. 1187–1192, Apr. 2000.

[17] L. Parra and C. Spence, "Convolutive blind source separation of nonstationary sources," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 3, pp. 320–327, May 2000.

[18] J. D. Shen and Z. Ding, "Zero-forcing blind equalization based on subspace estimation for multiuser systems," *IEEE Trans. Commun.*, vol. 49, no. 2, pp. 262–271, Feb. 2001.

[19] D. W. E. Schobben and P. C. W. Sommen, "A frequency domain blind signal separation method based on decorrelation," *IEEE Trans. Signal Process.*, vol. 50, no. 8, pp. 1855–1865, Aug. 2002.

[20] A. Swami, G. B. Giannakis, and S. Shamsunder, "Multichannel ARMA processes," *IEEE Trans. Signal Process.*, vol. 42, no. 4, pp. 898–913, Apr. 1994.

[21] L. Tong, Y. Inouye, and R. Liu, "A finite-step global convergence algorithm for the parameter estimation of multichannel MA processes," *IEEE Trans. Signal Process.*, vol. 40, no. 10, pp. 2547–2558, Oct. 1992.

[22] A. Touzni, I. Fijalkow, M. G. Larimore, and J. R. Treichler, "A globally convergent approach for blind MIMO adaptive deconvolution," *IEEE Trans. Signal Process.*, vol. 49, no. 6, pp. 1166–1178, Jun. 2001.

[23] J. K. Tugnait, "Blind spatio-temporal equalization and impulse response estimation for MIMO channels using a Godard cost function," *IEEE Trans. Signal Process.*, vol. 45, no. 1, pp. 268–271, Jan. 1997.

[24] 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 Process.*, vol. 48, no. 3, pp. 832–845, Mar. 2000.

[25] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Trans. Signal Process.*, vol. 43, no. 12, pp. 2982–2993, Dec. 1995.

[26] J. Zhu, X.-R. Cao, and Z. Ding, "An algebraic principle for blind separation of white non-Gaussian sources," *Signal Process.*, vol. 76, pp. 105–115, 1999.

[27] J. Zhu, Z. Ding, and X. R. Cao, "Column anchored zeroforcing blind equalization for multiuser wireless FIR channels," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 3, pp. 411–423, Mar. 1998.

Han-Fu Chen (SM'94–F'97) graduated from Leningrad (St. Petersburg) University, Leningrad (St. Petersburg), Russia, in 1961.

He joined the Institute of Mathematics, Chinese Academy of Sciences (CAS), Beijing, China, in 1961. Since 1979, he has been with the Institute of Systems Science, which now is a part of the Academy of Mathematics and Systems Science, CAS. He is a Professor with the Laboratory of Systems and Control of the Institute. His research interests are mainly in stochastic systems, including system identification, adaptive control, and stochastic approximation and its applications to systems, control, and signal processing. He authored and coauthored more than 160 journal papers and seven books. He was elected as a Member of CAS in 1993. He now serves as a Council Member of the International Federation of Automatic Control (IFAC) and as the Editor of both *Systems Science and Mathematical Sciences* and *Control Theory and Applications*. He is also involved with the editorial boards of several international and domestic journals.

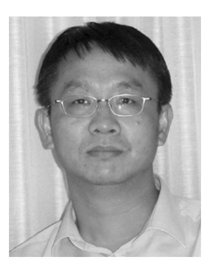
Xi-Ren Cao (SM'89–F'96) received the M.S. and Ph.D. degrees from Harvard University, Cambridge, MA, in 1981 and 1984, respectively.

He was a research fellow at Harvard University from 1984 to 1986. He then worked as a Principal and Consultant Engineer/Engineering Manager at Digital Equipment Corporation, Marlboro, MA, until October 1993. Since then, he was been a Professor with the Hong Kong University of Science and Technology (HKUST), Kowloon, Hong Kong, where he is the Director of the Center for Networking. He held visiting positions at Tsinghua University, Beijing, China, Harvard University, and many other universities. His current research areas include discrete event dynamic systems, communication systems, signal processing, stochastic processes, and system optimization.

Dr. Cao owns three patents in data communications and telecommunications and published two books: *Realization Probabilities—the Dynamics of Queuing Systems* (New York: Springer-Verlag, 1994) and *Perturbation Analysis of Discrete-Event Dynamic Systems* (Norwell, MA: Kluwer, 1991, coauthored with Y. C. Ho). He received the Outstanding Transactions Paper Award from the IEEE Control System Society in 1987 and the Outstanding Publication Award from the Institution of Management Science in 1990. He is Associate Editor at Large of IEEE TRANSACTIONS OF AUTOMATIC CONTROL, is on the Board of Governors of the IEEE Control Systems Society, is a member of the standing committee of the Chinese Association of Automation, is an associate editor of a number of international journals, and is chairman of a few technical committees of international professional societies.

Hai-Tao Fang (M'00) received the B.S. degree in probability and statistics in 1990 from Peking University, Beijing, China, the M.S. degree in applied mathematics in 1993 from Tsinghua University, Beijing, and the Ph.D. degree in 1996 from Peking University.

He now is with the Laboratory of Systems and Control, Institute of Systems Science, Chinese Academy of Sciences, Beijing, as an Associate Professor. From 1996 to 1998, he was a postdoctoral at the Institute of Systems Science and joined the Institute as an Assistant Professor in 1998. During 1998, 1999, and 2001, he was with Hong Kong University of Science and Technology as a Research Associate. His current research interests include stochastic optimization and systems control, communication systems, and signal processing.



Jie Zhu (S'95–M'98–SM'03) received the B.Eng. degree from Southwest Jiaotong University, Chengdu, China, in 1985 and the M.Eng. degree from Northwestern Polytechnic University, Chengdu, in 1988, both in computer engineering. He received the Ph.D. degree in electrical and electronic engineering from the Hong Kong University of Science and Technology (HKUST), Kowloon, Hong Kong, in 1997.

From 1988 to 1992, he was an Engineer with the Aeronautical Computing Technique Institute of China, Xi'an. From 1998 to 2001, he was a Research Fellow at the Center for Signal Processing, Nanyang Technological University, Singapore. He has been a Senior Engineer at ESS Technology Inc., Fremont, CA, since 2001, working on the R&D of communication and DVD severo products. He holds one patent in blind equalization and has published about twenty journal and conference papers. His research interests include blind adaptive equalization, blind signal separation, higher order statistics, and signal detection.