Research Article

Blind Deconvolution in Nonminimum Phase Systems Using Cascade Structure

Bin Xia and Liqing Zhang

Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200030, China

We introduce a novel cascade demixing structure for multichannel blind deconvolution in nonminimum phase systems. To simplify the learning process, we decompose the demixing model into a causal finite impulse response (FIR) filter and an anticausal scalar FIR filter. A permutable cascade structure is constructed by two subfilters. After discussing geometrical structure of FIR filter manifold, we develop the natural gradient algorithms for both FIR subfilters. Furthermore, we derive the stability conditions of algorithms using the permutable characteristic of the cascade structure. Finally, computer simulations are provided to show good learning performance of the proposed method.

1. INTRODUCTION

Recently, blind deconvolution has attracted considerable attention in various fields, such as neural network, wireless telecommunication, speech and image enhancement, biomedical signal processing (EEG/MEG signals) [1–4]. Blind deconvolution is to retrieve the independent source signals from sensor outputs using only sensor signals and certain knowledge on statistics of source signals. A number of methods [2, 5–13] have been developed for the blind deconvolution problem.

For blind deconvolution problem in minimum phase systems, causal filters are used as demixing models. Many algorithms work well in learning the coefficients of causal filters, such as the second-order statistical (SOS) approaches [2, 5–11, 13], higher-order statistical (HOS) approaches [2, 5, 9, 10], and the Bussgang algorithms [6–8, 14]. In the real world, the mixing systems are usually nonminimum phase. To deal with the blind deconvolution problem in nonminimum phase systems, Amari et al. [15] used doubly sided infinite impulse response (IIR) filters as demixing model. To our knowledge, it is still a difficult task to develop a practical algorithm for doubly sided IIR filters.

To simplify the problem of blind deconvolution, some researchers introduced the cascade structure for demixing filter. In [16], Douglas discussed a cascade structure for multichannel system. The main idea of cascade structure is to divide the difficult task into several easy subtasks. By introducing this idea in blind deconvolution, we can decompose the demixing filter into subfilters to recover the counterparts in mixing system. Labat et al. [17] presented a cascade structure for single channel blind equalization. Zhang et al. [18] provided a cascade structure to multichannel blind deconvolution. Waheed and Salam [19] discussed several cascade structures for blind deconvolution problem. Theoretically, a nonminimum phase system can be decomposed into a minimum phase subsystem and a corresponding maximum phase subsystem. Therefore, the demixing model can be divided into two subfilters accordingly. Zhang et al. [20] introduced cascade structure which was constructed by a causal FIR filter and an anticausal FIR filter.

In this paper, we introduce a new cascade structure for demixing model by elaborating the structure of mixing model of nonminimum phase systems. The new cascade demixing model is constructed with a causal FIR filter and an anticausal scalar FIR filter. First, we analyze the structure of nonminimum mixing model to obtain a reasonable decomposition of demixing model. Based on this decomposition, we propose a cascade demixing model which is permutable due to the use of an anticausal scalar FIR filter. Then we develop the natural gradient algorithm for both subfilters. The permutable characteristic is also helpful to derive the corresponding stability conditions.

The paper is organized as follows. In Section 2 we formulate the problem of blind deconvolution and discuss the filter decomposition. In Section 3, learning algorithms are

developed for both subfilters. In Section 4, computational complexity and the stability conditions of the proposed algorithms are analyzed. Section 5 presents some simulation results to evaluate the performance of the proposed algorithm. Finally, we devote the conclusions in Section 6.

2. PROBLEM FORMULATION AND FILTER DECOMPOSITION

In this section, the basic problem of blind deconvolution is formulated. By analyzing the geometrical structure of the mixing filter, we divide the demixing model filter into a causal FIR filter and an anticausal scalar FIR filter.

2.1. Basic model

To formulate the problem of blind deconvolution, a linear time-invariant (LTI) system is introduced to describe the mixing model. It is assumed that the measured signals $\mathbf{x}(k)$ are generated from unknown source signals $\mathbf{s}(k)$ by the following convolutive model:

$$\mathbf{x}(k) = \sum_{p=-\infty}^{\infty} \mathbf{H}_p \mathbf{s}(k - p), \qquad (1)$$

where $\mathbf{H}_p$ is an $n \times n$-dimensional matrix of mixing coefficients at time-lag p, which is called the impulse response at time p. In this paper, we assume the number of sensor signals is equal to the number of source signals. $\mathbf{s}(k) = [s_1(k), \ldots, s_n(k)]^T$ is an n-dimensional vector of source signals with mutually independent components and $\mathbf{x}(k) = [x_1(k), \ldots, x_n(k)]^T$ is the vector of sensor signals. We introduce a delay operator z, defined by $z^{-1}\mathbf{x}(k) = \mathbf{x}(k - 1)$. Then the model (1) can be rewritten as

$$\mathbf{x}(k) = \mathbf{H}(z)\mathbf{s}(k), \qquad (2)$$

where $\mathbf{H}(z) = \sum_{p=-\infty}^{\infty} \mathbf{H}_p z^{-p}$.

In blind deconvlution, the source signals $\mathbf{s}(k)$ and coefficients of $\mathbf{H}(z)$ are unknown. The objective is to estimate source signals $\mathbf{s}(k)$ or to identify the channel $\mathbf{H}(z)$ only using observed signals $\mathbf{x}(k)$ and some statistical features of source signals. One solution for blind deconvolution is to estimate the source signals by using an FIR demixing filter as follows:

$$\mathbf{y}(k) = \mathbf{W}(z)\mathbf{x}(k), \qquad (3)$$

where $\mathbf{y}(k) = [y_1(k), \ldots, y_n(k)]^T$ is an n-dimensional vector of the outputs, and $\mathbf{W}(z) = \sum_{p=-N}^{N} \mathbf{W}_p z^{-p}$ is an FIR filter, and $\mathbf{W}_p$ is an $n \times n$-dimensional coefficient matrix at time-lag p.

In independent component analysis (ICA), there exist scaling ambiguity and permutation ambiguity [21] because some prior knowledge of source signals are unknown. Similarly, these indeterminacies remain in the blind deconvolution problem. Therefore the objective of blind deconvolution is to find a demixing model $\mathbf{W}(z)$ which satisfies the following condition:

$$\mathbf{G}(z) = \mathbf{W}(z)\mathbf{H}(z) = \mathbf{P}\boldsymbol{\Lambda}\mathbf{D}(z), \qquad (4)$$

where $\mathbf{G}(z)$ refers to the global transfer function, $\mathbf{P} \in \mathbb{R}^{n \times n}$ is a permutation matrix, $\mathbf{D}(z) = \text{diag}\{z^{-d_1}, \ldots, z^{-d_n}\}$, and $\boldsymbol{\Lambda} \in \mathbb{R}^{n \times n}$ is a nonsingular diagonal scaling matrix.

If the LTI system (1) is minimum phase, the blind deconvolution problem can be solved in a straightforward way [21, 22]. If the LTI system is nonminimum phase, it is still difficult to find a learning algorithm for blind deconvolution. To solve the problem, we introduce a new cascade form for demixing model using filter decomposition method. In the next section, we will discuss the details of filter decomposition.

2.2. Model decomposition

To split the difficult task into some easy subtasks, filter decomposition method was introduced in blind deconvolution problems [17, 19, 20, 23]. In [20], the demixing filter $\mathbf{W}(z)$ was decomposed into a causal filter and an anticausal filter with cascade form. The filter decomposition is helpful to keep the demixing filter stable during training and to develop the natural gradient algorithm for training one-sided FIR filters. The learning algorithms [20] for both subfilters are dependent. Since error feedback propagation exists in the training process, the algorithm performance will be affected.

In this paper, we study the structure of nonminimum phase mixing model and filter decomposition method. The purpose is to find an efficient algorithm for blind deconvolution. Generally, the demixing model can be regarded as the inverse of mixing model. According to the matrix theory, the inverse of $\mathbf{H}(z)$ can be calculated by

$$\mathbf{H}^{-1}(z) = \mathbf{H}^{\sharp}(z) \det\left(\mathbf{H}(z)\right)^{-1}, \qquad (5)$$

where $\mathbf{H}^{\sharp}(z)$ is the adjoint matrix of $\mathbf{H}(z)$. If the mixing model $\mathbf{H}(z)$ is nonminimum phase system, the $\det(\mathbf{H}(z))^{-1}$ can be described as follows:

$$
\begin{aligned}
&\det\left(\mathbf{H}(z)\right)^{-1}\\
&= \left(cz^{-L_0} \prod_{p=1}^{L_1} \left(1 - b_p z^{-1}\right) \prod_{p=1}^{L_2} \left(1 - d_p z^{-1}\right) \right)^{-1}\\
&= c^{-1} z^{L_0} \prod_{p=1}^{L_1} \left(1 - b_p z^{-1}\right)^{-1} \prod_{p=1}^{L_2} \left(1 - d_p z^{-1}\right)^{-1}\\
&= c^{-1} z^{L_0+L_2} \prod_{p=1}^{L_1} \left(1 - b_p z^{-1}\right)^{-1} \prod_{p=1}^{L_2} \left(-d_p\right)^{-1} \sum_{q=0}^{\infty} d_p^{-q} z^q,
\end{aligned}
$$
$$(6)$$

where c is a nonzero constant, L_0, L_1, and L_2 are certain natural numbers, $0 < \|b_p\| < 1$, for $p = 1, \ldots, L_1$ and $\|d_p\| > 1$ for $p = 1, \ldots, L_2$. The b_p, d_p refer to the zeros of the FIR filter $\mathbf{H}(z)$. In nonminimum phase system, the zeros locate in the interior and outer of the unit circle. If all zeros of a system are in the interior of the unit circle of complex plane, the system is minimum phase. Submitting (6) in (5), we obtain

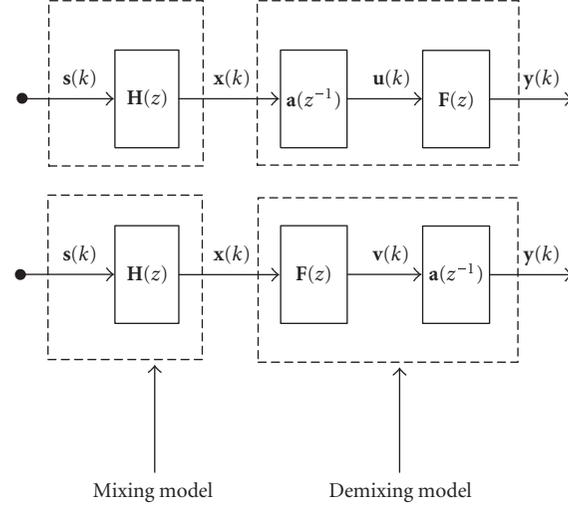$$\mathbf{H}^{-1}(z) = c^{-1} z^{L_0+L_2} \prod_{p=1}^{L_2} \left(-d_p\right)^{-1} \mathbf{F}(z)\mathbf{a}(z^{-1}), \qquad (7)$$

FIGURE 1: Illustration of filter decomposition for blind deconvolution.

where

$$\mathbf{F}(z) = \sum_{r=0}^{\infty} \mathbf{F}_r z^{-r} = \mathbf{H}^{\sharp}(z) \prod_{p=1}^{L_1} \left(1 - b_p z^{-1}\right)^{-1},$$

$$\mathbf{a}(z^{-1}) = \sum_{r=0}^{\infty} \mathbf{a}_r z^r = \prod_{p=1}^{L_2} \sum_{q=0}^{\infty} d_p^{-q} z^q. \tag{8}$$

From the above analysis, we know that the demixing model can be constructed by two parts: a causal filter $\mathbf{F}(z)$ and an anticausal scalar filter $\mathbf{a}(z^{-1})$. The two subfilters can exchange their positions because the filter $\mathbf{a}(z^{-1})$ is a scalar. As shown in Figure 1, we can obtain two decomposition forms as follows:

$$\mathbf{W}(z) = \mathbf{a}(z^{-1})\mathbf{F}(z) \quad \text{or} \quad \mathbf{W}(z) = \mathbf{F}(z)\mathbf{a}(z^{-1}). \tag{9}$$

In (8), $\|\mathbf{F}_r\|$ and $\|\mathbf{a}_r\|$ decay exponentially to zero as r tends to infinity. Hence, the decomposition of demixing filter is reasonable. After being decomposed, we can use two one-sided FIR filters to approximate filters $\mathbf{F}(z)$ and $\mathbf{a}(z^{-1})$ due to the decay properties of the coefficient of the inverse filter:

$$\mathbf{F}(z) = \sum_{p=0}^{N} \mathbf{F}_p z^{-p},$$

$$\mathbf{a}(z^{-1}) = \sum_{p=0}^{N} \mathbf{a}_p z^p, \tag{10}$$

where $\mathbf{F}_p$ is an $n \times n$-dimensional coefficient matrix at time-lag p, $\mathbf{a}_p$ is a scalar at time-lag p, and N is a given positive integer. This approximation will cause a model error in blind decovolution. If we choose an appropriate filter length N, the model error will become negligible and will not increase computational cost.

3. LEARNING ALGORITHM

In the previous section, we decomposed the demixing filter and introduced a new permutable cascade structure. To obtain self-closed multiplication and inverse operations in the manifold of FIR filters, we introduce some Lie Group's properties. Based on the geometrical structure of the FIR filter manifold, the natural gradient algorithms are developed for both subfilters.

3.1. Lie group

To discuss the geometrical property of FIR filter, we denote the set of all one-sided FIR filters of length N as $\mathcal{M}(N)$:

$$\mathcal{M}(N) = \left\{ \mathbf{A}(z) \mid \mathbf{A}(z) = \sum_{p=0}^{N} \mathbf{A}_p z^{-p} \right\}. \tag{11}$$

In $\mathcal{M}(N)$, the operations of *multiplication* $*$ and *inverse* † are defined as

$$\mathbf{A}(z) * \mathbf{B}(z) = \left[\mathbf{A}(z)\mathbf{B}(z)\right]_N, \tag{12}$$

where $[\cdot]_N$ is the truncating operator that any term with order higher than N is omitted.

$$\mathbf{B}^{\dagger}(z) = \sum_{p=0}^{N} \mathbf{B}_p^{\dagger} z^{-p}, \tag{13}$$

where $\mathbf{B}_p^{\dagger}$ are recurrently defined by $\mathbf{B}_0^{\dagger} = \mathbf{B}_0^{-1}$, $\mathbf{B}_1^{\dagger} = -\mathbf{B}_0^{\dagger}\mathbf{B}_1\mathbf{B}_0^{\dagger}$, $\mathbf{B}_p^{\dagger} = -\sum_{q=1}^{p} \mathbf{B}_{p-q}^{\dagger}\mathbf{B}_q\mathbf{B}_0^{\dagger}$, $p = 1, \dots, N$.

For the sake of simplicity, we only give some properties of Lie Group here. More detailed information can be found in [20].

Property 1.

$$\mathbf{A}(z) * \left(\mathbf{B}(z) * \mathbf{C}(z)\right) = \left(\mathbf{A}(z) * \mathbf{B}(z)\right) * \mathbf{C}(z). \tag{14}$$

Property 2.

$$\mathbf{B}(z) * \mathbf{B}^{\dagger}(z) = \mathbf{B}^{\dagger}(z) * \mathbf{B}(z) = \mathbf{I}. \quad (15)$$

Within the Lie group framework, the inverse $\mathbf{F}^{\dagger}(z)$ of the causal filter $\mathbf{F}(z)$ still lies in the manifold $\mathcal{M}(N)$, while the inverse $\mathbf{a}^{\dagger}(z^{-1})$ is in the same manifold with anticausal filter $\mathbf{a}(z^{-1})$.

3.2. Learning algorithm

The purpose of blind deconvolution is to find an FIR demixing filter $\mathbf{W}(z)$ such that the output of the demixing model is maximally mutually independent and temporally i.i.d. The Kullback-Leibler Divergence has been used as a criterion for blind deconvolution [20, 24, 25] to measure the mutual independence of the output signals. In [20], the authors introduced the following simple cost function for blind deconvolution:

$$l(\mathbf{y}, \mathbf{W}(z)) = -\log |\det(\mathbf{F}_0)| - \sum_{i=1}^{n} \log p_i(y_i), \quad (16)$$

where the output signals $y_i = \{y_i(k), k = 1, 2, \dots\}$, $i = 1, \dots, n$ is a stochastic process, $p_i(y_i(k))$ is the marginal probability density function of $y_i(k)$ for $i = 1, \dots, n$ and $k = 1, \dots, T$, and $\mathbf{F}_0$ is an $n \times n$-dimensional coefficient matrix at time-lag 0 of filter $\mathbf{F}(z)$. The first term in the cost function is introduced to prevent the matrix $\mathbf{F}_0$ from being singular.

Using the cascade form in (9), we will develop the algorithms for both $\mathbf{F}(z)$ and $\mathbf{a}(z^{-1})$. Here we introduce an intermediate variable $\mathbf{u}$, defined as

$$\mathbf{u}(k) = [\mathbf{a}(z^{-1})]\mathbf{x}(k),$$
$$\mathbf{y}(k) = [\mathbf{F}(z)]\mathbf{u}(k). \quad (17)$$

To calculate the natural gradient of the cost function, we consider the differential of the cost function:

$$dl(\mathbf{y}, \mathbf{W}(z)) = -d \log |\det(\mathbf{F}_0)| - \sum_{i=1}^{n} d \log p_i(y_i). \quad (18)$$

Using the relation $d \log |\det(\mathbf{F}_0)| = \mathrm{tr}(d\mathbf{F}_0\mathbf{F}_0^{-1})$, we have

$$dl(\mathbf{y}, \mathbf{W}(z)) = -\mathrm{tr}(d\mathbf{F}_0\mathbf{F}_0^{-1}) + \boldsymbol{\varphi}(\mathbf{y})^T d\mathbf{y}, \quad (19)$$

where $\varphi(\mathbf{y}) = (\varphi_1(y_1), \dots, \varphi_n(y_n))^T$ is the vector of nonlinear activation functions, defined by

$$\varphi_i(y_i) = -\frac{d}{dy} \log(p_i(y_i)) = -\frac{p_i'(y_i)}{p_i(y_i)}, \quad \text{for } i = 1, \dots, n. \quad (20)$$

In order to develop the natural gradient algorithms for both filters, we introduce nonholonomic transforms here:

$$d\mathbf{X}(z) = d\mathbf{F}(z) * \mathbf{F}^{\dagger}(z),$$
$$d\mathbf{b}(z^{-1}) = d\mathbf{a}(z^{-1}) * \mathbf{a}^{\dagger}(z^{-1}). \quad (21)$$

In particular,

$$d\mathbf{X}_0 = d\mathbf{F}_0\mathbf{F}_0^{-1}, \quad (22)$$
$$d\mathbf{a}_0 = 0, \qquad d\mathbf{b}_0 = 0. \quad (23)$$

Using the nonholonomic transforms, we can easily calculate

$$
\begin{aligned}
d\mathbf{y}(k) &= d[\mathbf{W}(z)]\mathbf{x}(k) \\
&= [d\mathbf{F}(z)\mathbf{a}(z^{-1})]\mathbf{x}(k) + [\mathbf{F}(z)]d\mathbf{a}(z^{-1})\mathbf{x}(k) \\
&= [d\mathbf{F}(z) * \mathbf{F}^{\dagger}(z) * \mathbf{F}(z)]\mathbf{u}(k) \\
&\quad + [\mathbf{F}(z)d\mathbf{a}(z^{-1}) * \mathbf{a}^{\dagger}(z^{-1}) * \mathbf{a}(z^{-1})]\mathbf{x}(k) \\
&= [d\mathbf{X}(z)]\mathbf{y}(k) + [d\mathbf{b}(z^{-1})]\mathbf{y}(k).
\end{aligned}
\quad (24)
$$

Substituting (22) and (24) into (19), we have

$$
\begin{aligned}
dl(\mathbf{y}, \mathbf{W}(z)) = &-\mathrm{tr}(d\mathbf{X}_0) + \boldsymbol{\varphi}^T(\mathbf{y})[d\mathbf{X}(z)]\mathbf{y}(k) \\
&+ \boldsymbol{\varphi}^T(\mathbf{y})[d\mathbf{b}(z^{-1})]\mathbf{y}(k).
\end{aligned}
\quad (25)
$$

Therefore, we obtain the derivatives of the cost function with respect to $\mathbf{X}(z)$ and $\mathbf{b}(z^{-1})$

$$
\begin{aligned}
\frac{\partial l(\mathbf{y}, \mathbf{W}(z))}{\partial \mathbf{X}_p} &= -\delta_{0,p}\mathbf{I} + \boldsymbol{\varphi}(\mathbf{y}(k))\mathbf{y}^T(k - p), \\
\frac{\partial l(\mathbf{y}, \mathbf{W}(z))}{\partial \mathbf{b}_q} &= \boldsymbol{\varphi}^T(\mathbf{y}(k))\mathbf{y}(k + q),
\end{aligned}
\quad (26)
$$

for $p = 0, 1, \dots, N$; $q = 1, \dots, N$. The gradient descent algorithms for $\mathbf{X}(z)$ and $\mathbf{b}(z^{-1})$ are given by

$$
\begin{aligned}
\triangle \mathbf{X}_p &= -\eta \frac{\partial l(\mathbf{y}, \mathbf{W}(z))}{\partial \mathbf{X}_p} = \eta(\delta_{0,p}\mathbf{I} - \boldsymbol{\varphi}(\mathbf{y}(k))\mathbf{y}^T(k - p)), \\
\triangle \mathbf{b}_q &= -\eta \frac{\partial l(\mathbf{y}, \mathbf{W}(z))}{\partial \mathbf{b}_p} = -\eta \boldsymbol{\varphi}^T(\mathbf{y}(k))\mathbf{y}(k + q),
\end{aligned}
\quad (27)
$$

for $p = 0, 1, \dots, N$; $q = 1, \dots, N$. Using the differential relations (21), we derive learning algorithms for updating the filters $\mathbf{F}(z)$ and $\mathbf{a}(z^{-1})$ as follows:

$$
\begin{aligned}
\triangle \mathbf{F}(z) &= \triangle \mathbf{X}(z) * \mathbf{F}(z), \\
\triangle \mathbf{a}(z^{-1}) &= \triangle \mathbf{b}(z^{-1}) * \mathbf{a}(z^{-1}).
\end{aligned}
\quad (28)
$$

The learning algorithm can be written in the matrix form:

$$
\begin{aligned}
\triangle \mathbf{F}_p &= -\eta \sum_{q=0}^{p} \frac{\partial l(\mathbf{y}, \mathbf{W}(z))}{\partial \mathbf{X}_p} \mathbf{F}_{p-q} \\
&= \eta \sum_{q=0}^{p} (\delta_{0,q}\mathbf{I} - \boldsymbol{\varphi}(\mathbf{y}(k))\mathbf{y}^T(k - q))\mathbf{F}_{p-q}, \\
\triangle \mathbf{a}_p &= -\eta \sum_{q=0}^{p} \frac{\partial l(\mathbf{y}, \mathbf{W}(z))}{\partial \mathbf{b}_p} \mathbf{a}_{p-q} \\
&= -\eta \sum_{q=0}^{p} (\boldsymbol{\varphi}^T(\mathbf{y}(k))\mathbf{y}(k + q))\mathbf{a}_{p-q}
\end{aligned}
\quad (29)
$$

for $p = 1, \ldots, N$. In (29), there exists an unknown parameter $\boldsymbol{\varphi}$, that is, the nonlinear activation function, which depends on the probability density functions of the unknown sources. According to the semiparameter theory, $\boldsymbol{\varphi}$ can be regarded as a nuisance parameter, therefore it is not necessary to estimate it precisely. However, if we choose a better $\boldsymbol{\varphi}$, it is helpful for improving performance of the algorithm. For example, a suitable activation function can greatly improve the stability of the learning algorithm [20, 26].

4. COMPUTATIONAL COMPLEXITY AND STABILITY CONDITIONS

As mentioned above, we use an anticausal scalar filter in new cascade structure of demixing filter. It is not only to make the structure permutable, but also to halve the computation requirements. In [20], the demixing FIR filter was decomposed into two one-sided FIR filters. If the order of the FIR filters is N and the number of sensors is n, so we must compute $2 * n^2 * N$ parameters for each iteration. In the proposed algorithm, we only need to compute $n^2 * N$ parameters for causal FIR filter and to compute N parameters for the scalar anticausal FIR filter at each iteration. So the computation cost is lower than that in [20].

Amari et al. [26] derived the stability conditions for instantaneous blind source separation. In [27], authors analyzed the stability of blind deconvolution and presented the stability conditions. The proposed algorithms, developed by using filter decomposition, are different from the algorithms in [27]. So the stability conditions in [27] cannot be applied directly to the algorithm developed for noncausal demixing filters.

From (29) we know that the learning algorithms for updating $\mathbf{F}_p$ and $\mathbf{a}_p$, $p = 0, 1, \ldots, N$, are linear combination of $\mathbf{X}_p$ and $\mathbf{b}_p$, respectively. It is easy to see that the stability of $\mathbf{X}_p$ and $\mathbf{b}_p$ implies the stability of the learning algorithm. Here we suppose that the estimated signals $\mathbf{y} = (y_1, \ldots, y_n)^T$ are not only spatially mutually independent but also temporally i.i.d.

The learning algorithms of $\mathbf{X}_p$ and $\mathbf{b}_p$ can be written as follows:

$$\frac{d\mathbf{X}_p}{dt} = \eta(\delta_{0,p}\mathbf{I} - \boldsymbol{\varphi}(\mathbf{y}(k))\mathbf{y}^T(k-p)),$$
$$\frac{d\mathbf{b}_p}{dt} = -\eta(\boldsymbol{\varphi}^T(\mathbf{y}(k))\mathbf{y}(k+p)), \tag{30}$$

where $p = 0, 1, \ldots, N$. To analyze those asymptotic properties of the learning algorithms, we take expectation on the above equations:

$$\frac{d\mathbf{X}_p}{dt} = \eta(\delta_{0,p}\mathbf{I} - E[\boldsymbol{\varphi}(\mathbf{y}(k))\mathbf{y}^T(k-p)]),$$
$$\frac{d\mathbf{b}_p}{dt} = -\eta(E[\boldsymbol{\varphi}^T(\mathbf{y}(k))\mathbf{y}(k+p)]). \tag{31}$$
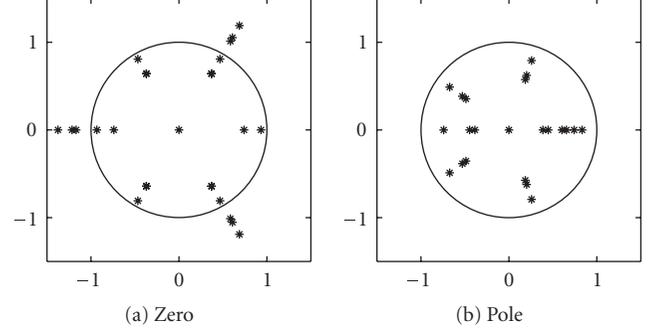

(a) Zero   (b) Pole

FIGURE 2: (a) Zero distributions of mixing model; (b) pole distributions of mixing model.

The stability conditions for (31) are obtained as follows:

$$k_i > 0, \quad \text{for } i = 1, \ldots, n,$$
$$k_i k_j \sigma_i^2 \sigma_j^2 > 1, \quad \text{for } i, j = 1, \ldots, n,$$
$$m_i + 1 > 0, \quad \text{for } i = 1, \ldots, n,$$
$$\sum_i k_i \sigma_i^2 > \sum_i (k_i \sigma_i^2)^{-1}, \tag{32}$$

where $m_i = E[\varphi'(y_i)y_i^2]$, $k_i = E[\varphi_i'(y_i)]$, $\sigma_i^2 = E[|y_i|^2]$, $i = 1, \ldots, n$. Detailed derivation is left in the appendix.

5. SIMULATIONS

We now present several examples for simulating to illustrate the performance of the proposed blind deconvolution algorithm. The proposed algorithm is named as permutable filter decomposed method (PFD) and its performance is compared with the decomposition method (FD) in [20] and the natural gradient algorithm (NG) [5]. In this section, we provide three simulation examples.

5.1. Separation experiment in nonminimum phase system

In this simulation, we verify separation performance of the proposed algorithm for nonminimum phase system. Here we employ a mixing model generated by an ARMA model, described as follows:

$$\mathbf{x}(k) + \sum_{i=1}^{N} \mathbf{A}_i \mathbf{x}(k-i) = \sum_{i=0}^{N} \mathbf{B}_i \mathbf{s}(k-i) + \mathbf{v}(k), \tag{33}$$

where $\mathbf{x}(k)$ is the vector of mixing signals, $\mathbf{s}(k)$ is the vector of source signals, and $\mathbf{v}(k)$ is the Gaussian noise with zero mean and a covariance matrix $0.1\mathbf{I}$. Using this ARMA model, we can generate minimum phase or nonminimum phase mixing model by choosing different $\mathbf{A}_i$ and $\mathbf{B}_i$. From the distributions of zeros and poles shown in Figure 2, the mixing system is stable and of nonminimum phase. The source signals are three independent i.i.d. signals uniformly distributed in range $(-1, 1)$. The nonlinear activation function
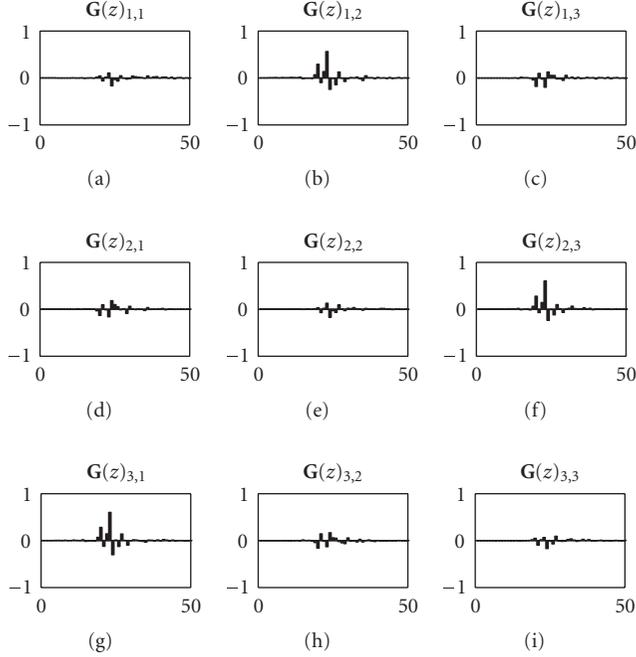
FIGURE 3: The coefficients of the global function at initiation.

FIGURE 4: The coefficients of the global function after convergence.

is $\boldsymbol{\varphi}(y) = y^3$. We use batch method in this example to implement the proposed algorithm and the batch window is set as 6000. In the proposed algorithm, we use FIR filter to approximate IIR filter, which will cause a model error. We should choose an optimal filter length to minimize this model error. In general, the MDL criterion is used to choose filter length [20]. In this simulation, we set the filter length N as 20. The initial learning rate η is set to 0.01, and update learning rate by $\eta = \max\{0.9\eta, 10^{-4}\}$ for every 10 iterations. As we defined before, the $G(z)$ is the global function whose coefficients initial are shown in Figure 3. Generally, if the global function is close to an identity filter, the source signals can be estimated well. Figure 4 shows the coefficients of $\mathbf{G}(z)$ after convergence. It is obvious that the $\mathbf{G}(z)$ is very close to an identity filter. That means the proposed algorithm achieves good separation performance. Figures 5 and 6 show the coefficients of the causal filter $\mathbf{F}(z)$ and the anticausal filter $\mathbf{a}(z^{-1})$, respectively. The coefficients of both filters decay while the delay number p increases.

5.2. Comparison of PFD, FD, and NG in minimum phase system

The key point of filter decomposition method [20] is to divide the nonminimum phase system into a minimum phase part and a maximum part, and then use a causal filter and an anticausal filter to demix the counterparts, respectively. As shown in [20] and simulation 1, both PFD and FD work well in nonminimum phase system. How about the performance in minimum phase system? We compare the PFD, FD, and NG [24] algorithms in minimum phase system here and
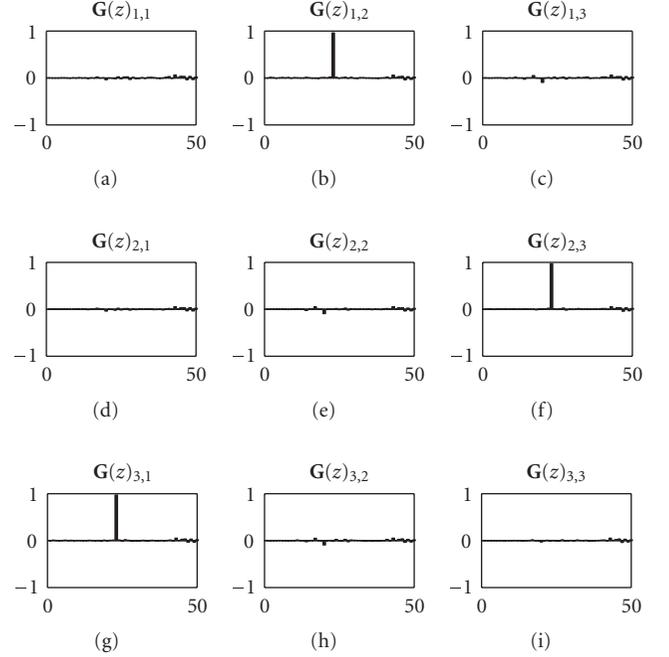
analyze the different performances of them.

We introduce the intersymbol interference as a performance criterion. In [12, 28], the M_{ISI} is defined as

$$
M_{\text{ISI}} = \sum_{i=1}^{n} \frac{\left| \sum_{j=1}^{n} \sum_{p=-N}^{N} |g_{p,ij}| - \max_{p,j} |g_{p,ij}| \right|}{\max_{p,j} |g_{p,ij}|}
$$
$$
+ \sum_{j=1}^{n} \frac{\left| \sum_{i=1}^{n} \sum_{p=-N}^{N} |g_{p,ij}| - \max_{p,i} |g_{p,ij}| \right|}{\max_{p,i} |g_{p,ij}|}.
$$

(34)

In this simulation, we choose different $\mathbf{A}_i$ and $\mathbf{B}_i$ in (33) to obtain a minimum phase mixing model. The source signals in simulation 1 is used in this simulation. We set the filter length of demixing filters and the learning rate update rule as those in simulation 1 for three algorithms.

To remove the effect of a single numerical trial, we use the ensemble average of 100 trails. Figure 7 illustrates the comparison results of the three algorithms. It shows that the performances of PFD and NG are similar, and both of them are better than FD algorithm. That is because the FD algorithm uses an error back propagation method to develop algorithms for both subfilters. In the minimum phase system, the anticausal filter should be an identity filter. But in FD algorithm, the coefficients of anticausal filter did not achieve the identity filter due to the error back propagation which degenerates the convergence performance. In PFD algorithm, there is not an error back propagation process. Therefore PFD algorithm can obtain the same performance with natural gradient algorithm.
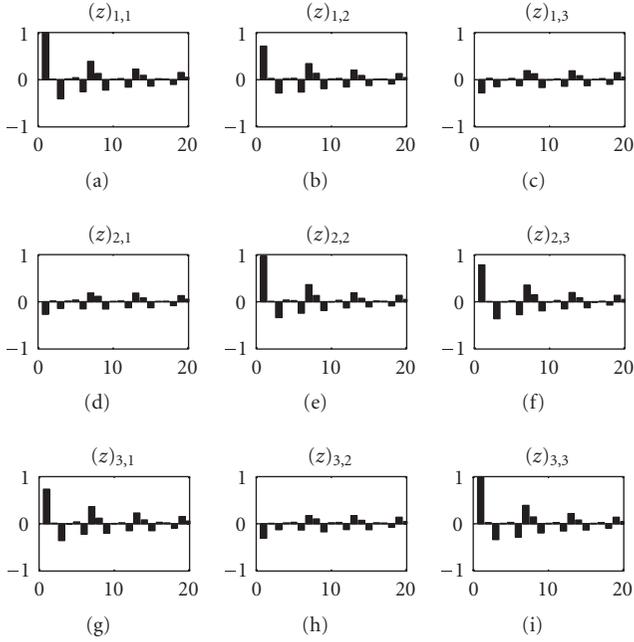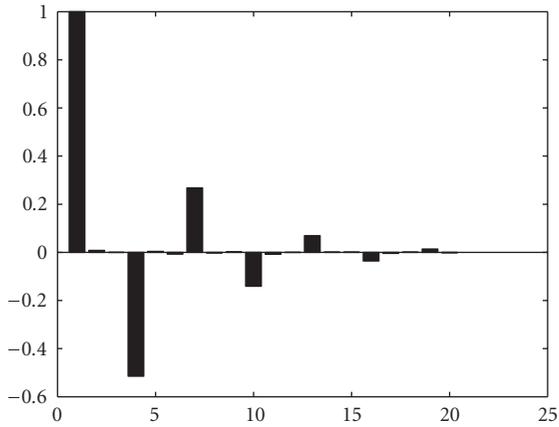
FIGURE 5: Coefficients of $\mathbf{F}(z)$.

FIGURE 6: Coefficients of $\mathbf{a}(z^{-1})$.

FIGURE 7: Comparison results of M_{ISI} in minimum phase system.
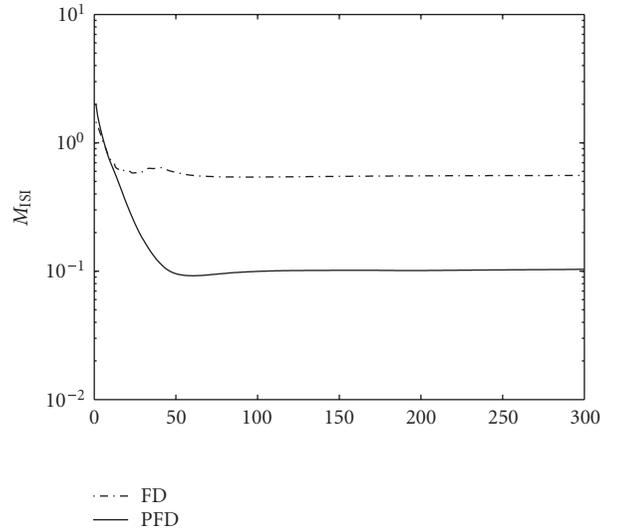
FIGURE 8: Comparison results of M_{ISI} in nonminimum phase system.

5.3. Comparison of PFD and FD in the nonminimum phase system

We intend to compare the proposed algorithm with other algorithms in nonminimum phase system. But some algorithms cannot work well in the situation of simulation 1, such as NG algorithm and Bussgang algorithm. In this simulation, we only compared PFD and FD algorithms in nonminimum phase system because both algorithms can separate mixing signals. The coefficients of mixing filter $\mathbf{H}(z)$ are set the same as experiment 1. We set the filter length N to 20 at both sides. Figure 8 shows the 100 trails ensemble average comparison result. The PFD algorithm converges faster than FD. Because the computational cost is lower in PFD at each
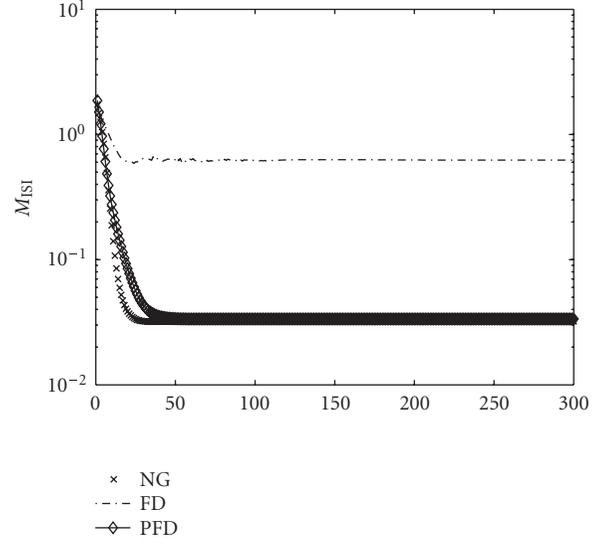
iteration than in FD. During the computing, we find the M_{ISI} fluctuates at the initiation in FD algorithm due to the error back propagation. In PFD algorithm, we use scalar anticausal filter in PFD and then avoid the error back propagation. So the convergence processing is smooth.

6. CONCLUSION

In this paper we present a permutable cascade form for multichannel blind deconvolution in nonminimum phase system. By decomposing the demixing anticausal FIR filter into two sub-FIR filters, the difficult problem is divided into several easy subtasks. The structure of demixing model is permutable because an anticausal scalar FIR filter is used.

Natural gradient-based algorithms can be easily developed for two one-sided filters. Using the permutable characteristic of this cascade structure, we derive the stability conditions for the proposed algorithm. Finally, the simulation results show the efficiency and performance of the proposed algorithm.

APPENDIX

In this appendix, we provide the detailed derivation for the stability conditions. The learning algorithms for updating $\mathbf{F}_k$ and $\mathbf{a}_k$, $k = 0, 1, \ldots, N$, are linear combination of $\mathbf{X}_k$ and $\mathbf{b}_k$, respectively. The stability of $\mathbf{X}_k$ and $\mathbf{b}_k$ implies the stability of the learning algorithm. Here we suppose that the separating signals $\mathbf{y} = (y_1, \ldots, y_n)^T$ are not only spatially mutually independent but also temporally i.i.d.

Consider (31), if the variational matrix at equilibrium point is negative definite, then the system is stable in the vicinity of the equilibrium point. Taking a variation $\delta\mathbf{X}_p$ on $\mathbf{X}_p$ and a variation $\delta\mathbf{b}_p$ on $\mathbf{b}_p$, respectively, we have

$$
\begin{aligned}
\frac{d\delta\mathbf{X}_p}{dt} &= -\eta E\big[\boldsymbol{\varphi}'(\mathbf{y}(k))\delta\mathbf{y}\mathbf{y}^T(k-p) + \boldsymbol{\varphi}(\mathbf{y}(k))\delta\mathbf{y}^T(k-p)\big], \\
\frac{d\delta\mathbf{b}_p}{dt} &= -\eta E\big[\big[\boldsymbol{\varphi}'(\mathbf{y}(k))\big]^T\delta\mathbf{y}(k)\mathbf{y}(k+p) \\
&\qquad + \boldsymbol{\varphi}^T(\mathbf{y}(k))\delta\mathbf{y}(k+p)\big].
\end{aligned}
\tag{A.1}
$$

Furthermore, we write the differential expression of $\delta\mathbf{y}(k)$

$$
\begin{aligned}
\delta\mathbf{y}(k) &= \big[\mathbf{a}(z)\delta\mathbf{F}(z) + \delta\mathbf{a}(z)\mathbf{F}(z)\big]\mathbf{x}(k) \\
&= \big[\delta\mathbf{X}(z) + \mathbf{I}\delta\mathbf{b}(z)\big]\mathbf{y}(k).
\end{aligned}
\tag{A.2}
$$

As mentioned above, the matrix $\mathbf{F}_0$ is nonsingular. This means that the learning algorithms keep the filters $\mathbf{F}(z)$ and $\mathbf{a}(z)$ on the same manifold with the initial filter. This property implies that the equilibrium point of the learning algorithm satisfies the following equations:

$$
E\big[\mathbf{I} - \boldsymbol{\varphi}(\mathbf{y}(k))\mathbf{y}^T(k)\big] = 0.
\tag{A.3}
$$

Using the mutual independence and i.i.d. properties of the output signals y_i, $i = 1, \ldots, n$ and the normalized condition (A.3), we deduce (A.1) to

$$
\begin{aligned}
\frac{d\delta\mathbf{X}_p}{dt} &= -\eta E\big\{\boldsymbol{\varphi}'(\mathbf{y}(k))\big[\big([\delta\mathbf{X}(z)] + \mathbf{I}\delta\mathbf{b}(z)\mathbf{y}(k)\big)\big]\mathbf{y}^T(k-p) \\
&\qquad + \boldsymbol{\varphi}(\mathbf{y}(k))\mathbf{y}^T(k-p)\big[(\delta\mathbf{X}(z) + \delta\mathbf{b}(z))^T\big]\big\}, \\
\frac{d\delta\mathbf{b}_p}{dt} &= -\eta E\big\{\big[\boldsymbol{\varphi}'(\mathbf{y}(k))\big]^T(\delta\mathbf{X}(z) + \delta\mathbf{b}(z) * \mathbf{I})\mathbf{y}(k)\mathbf{y}(k+p) \\
&\qquad + \boldsymbol{\varphi}^T(\mathbf{y}(k))\big[\delta\mathbf{X}(z) + \delta\mathbf{b}(z) * \mathbf{I}\big]\mathbf{y}(k+p)\big\}.
\end{aligned}
\tag{A.4}
$$

When $p = 0$,

$$
\frac{d\delta\mathbf{X}_0}{dt} = -\eta E\big[\boldsymbol{\varphi}'(\mathbf{y}(k))\delta\mathbf{X}_0\mathbf{y}(k)\mathbf{y}^T(k) + \boldsymbol{\varphi}(\mathbf{y}(k))\mathbf{y}^T(k)\delta\mathbf{X}_0^T\big],
\tag{A.5}
$$

$$
\frac{d\delta\mathbf{b}_0}{dt} = 0.
\tag{A.6}
$$

Rewrite (A.5) in component form

$$
\begin{aligned}
\frac{d\delta\mathbf{X}_{0,ij}}{dt} &= -\eta\big[k_i\sigma_j^2\delta\mathbf{X}_{0,ij} + \delta\mathbf{X}_{0,ji}\big], \\
\frac{d\delta\mathbf{X}_{0,ji}}{dt} &= -\eta\big[k_j\sigma_i^2\delta\mathbf{X}_{0,ji} + \delta\mathbf{X}_{0,ij}\big],
\end{aligned}
\tag{A.7}
$$

for $i \neq j$, and

$$
\frac{d\delta\mathbf{X}_{0,ii}}{dt} = -\eta(m_i + 1)\delta\mathbf{X}_{0,ii},
\tag{A.8}
$$

for $p = 1, \ldots, N$, and $i, j = 1, \ldots, n$, where $m_i = E[\varphi'(y_i)y_i^2]$, $k_i = E[\varphi_i'(y_i)]$, $\sigma_i^2 = E[|y_i|^2]$, $i = 1, \ldots, n$. The stability conditions of (A.7) and (A.8) are given by

$$
\begin{aligned}
k_i &> 0, \quad \text{for } i = 1, \ldots, n, \\
k_i k_j \sigma_i^2 \sigma_j^2 &> 1, \quad \text{for } i, j = 1, \ldots, n, \\
m_i + 1 &> 0, \quad \text{for } i = 1, \ldots, n.
\end{aligned}
\tag{A.9}
$$

When $p \neq 0$

$$
\begin{aligned}
\frac{d\delta\mathbf{X}_p}{dt} &= -\eta E\big[\boldsymbol{\varphi}'(\mathbf{y}(k))\delta\mathbf{X}_p\mathbf{y}(k-p)\mathbf{y}^T(k-p) \\
&\qquad + \boldsymbol{\varphi}(\mathbf{y}(k))\delta\mathbf{b}_p\mathbf{y}^T(k)\big] \\
&= -\eta\big[k\mathbf{I}\delta\mathbf{X}_p\sigma^2 + \delta\mathbf{b}_p\mathbf{I}\big],
\end{aligned}
\tag{A.10}
$$

$$
\begin{aligned}
\frac{d\delta\mathbf{b}_p}{dt} &= -\eta E\big[\big[\boldsymbol{\varphi}'(\mathbf{y}(k))\big]^T\delta\mathbf{b}_p\mathbf{y}^T(k+q)\mathbf{y}(k+q) \\
&\qquad + \boldsymbol{\varphi}^T(\mathbf{y}(k))\delta\mathbf{X}_p\mathbf{y}(k)\big] \\
&= -\eta\Big[\sum_i k_i\sigma_i^2\delta\mathbf{b}_p + \sum_i \delta\mathbf{X}_{p,ii}\Big].
\end{aligned}
\tag{A.11}
$$

For $i \neq j$, the components form of (A.10) can be rewritten as follows:

$$
\frac{d\delta\mathbf{X}_p}{dt} = -\eta(k_i\sigma_j^2\delta\mathbf{X}_{p,ij}).
\tag{A.12}
$$

The stability condition for (A.12) is as follows:

$$
k_i > 0, \quad \text{for } i = 1, \ldots, n.
\tag{A.13}
$$

From (A.11), we know that only the diagonal entries of $\delta\mathbf{X}_p$ are relative with $\delta\mathbf{b}_p$, for $i = j$. The diagonal component form of $\delta\mathbf{X}_p$ can be written as

$$
\frac{d\delta\mathbf{X}_{p,ii}}{dt} = -\eta\big[k_i\sigma_i^2\delta\mathbf{X}_{p,ii} + \delta\mathbf{b}_p\big].
\tag{A.14}
$$

Combining (A.11) and (A.14), we get

$$\frac{d}{dt}\begin{bmatrix} \delta\mathbf{X}_{p,11} \\ \vdots \\ \delta\mathbf{X}_{p,nn} \\ \delta\mathbf{b}_p \end{bmatrix} = -\eta \begin{bmatrix} k_1\sigma_1^2 & 0 & \cdots & 1 \\ \vdots & \ddots & \cdots & \vdots \\ 0 & \cdots & k_n\sigma_n^2 & 1 \\ 1 & \cdots & 1 & \sum_i k_i\sigma_i^2 \end{bmatrix} \begin{bmatrix} \delta\mathbf{X}_{p,11} \\ \vdots \\ \delta\mathbf{X}_{p,nn} \\ \delta\mathbf{b}_p \end{bmatrix}. \tag{A.15}$$

If we want to make the variation matrix be negative, we should let

$$\sum_i k_i\sigma_i^2 - \sum_i \left(k_i\sigma_i^2\right)^{-1} > 0. \tag{A.16}$$

So we obtain the stability condition for $p \neq 0$

$$\sum_i k_i\sigma_i^2 > \sum_i \left(k_i\sigma_i^2\right)^{-1}, \quad \text{for } i = 1,\ldots,n. \tag{A.17}$$

In summary, we have the total stability conditions for the natural gradient algorithm of the blind deconvolution as follows:

$$\begin{aligned} k_i &> 0, \quad \text{for } i = 1,\ldots,n, \\ k_i k_j \sigma_i^2 \sigma_j^2 &> 1, \quad \text{for } i,j = 1,\ldots,n, \\ m_i + 1 &> 0, \quad \text{for } i = 1,\ldots,n, \\ \sum_i k_i\sigma_i^2 &> \sum_i \left(k_i\sigma_i^2\right)^{-1}. \end{aligned} \tag{A.18}$$

ACKNOWLEDGMENTS

REFERENCES

[1] S. Amari, "Natural gradient works efficiently in learning," *Neural Computation*, vol. 10, no. 2, pp. 251–276, 1998.

[2] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, no. 6, pp. 1129–1159, 1995.

[3] J.-F. Cardoso and B. Laheld, "Equivariant adaptive source separation," *IEEE Transactions on Signal Processing*, vol. 44, no. 12, pp. 3017–3030, 1996.

[4] P. Comon, "Independent component analysis: a new concept?" *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.

[5] S. Amari, S. Douglas, A. Cichocki, and H. Yang, "Novel on-line algorithms for blind deconvolution using natural gradient approach," in *Proceedings of the 11th IFAC Symposium on System Identification (SYSID '97)*, pp. 1057–1062, Kitakyushu, Japan, July 1997.

[6] S. Bellini, "Bussgang techniques for blind equalization," in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM '86)*, pp. 1634–1640, Houston, Tex, USA, December 1986.

[7] A. Benveniste, M. Goursat, and G. Ruget, "Robust identification of a nonminimum phase system: blind adjustment of a linear equalizer in data communications," *IEEE Transactions on Automatic Control*, vol. 25, no. 3, pp. 385–399, 1980.

[8] D. N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Transactions on Communications Systems*, vol. 28, no. 11, pp. 1867–1875, 1980.

[9] Y. Hua, "Fast maximum likelihood for blind identification of multiple FIR channels," *IEEE Transactions on Signal Processing*, vol. 44, no. 3, pp. 661–672, 1996.

[10] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)," *IEEE Transactions on Information Theory*, vol. 36, no. 2, pp. 312–321, 1990.

[11] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: a time domain approach," *IEEE Transactions on Information Theory*, vol. 40, no. 2, pp. 340–349, 1994.

[12] J. K. Tugnait, "Channel estimation and equalization using high-order statistics," in *Signal Processing Advances in Wireless and Mobile Communications*, G. B. Giannakis, Y. Hua, P. Stoica, and L. Tong, Eds., vol. 1, pp. 1–40, Prentice-Hall, Upper Saddle River, NJ, USA, 2000.

[13] J. K. Tugnait and B. Huang, "Multistep linear predictors-based blind identification and equalization of multiple-input multiple-output channels," *IEEE Transactions on Signal Processing*, vol. 48, no. 1, pp. 26–38, 2000.

[14] Y. Li, A. Cichocki, and L. Zhang, "Blind source estimation of FIR channels for binary sources: a grouping decision approach," *Signal Processing*, vol. 84, no. 12, pp. 2245–2263, 2004.

[15] S. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," in *Advances in Neural Information Processing Systems. Vol. 8 (NIPS '95)*, D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, Eds., pp. 757–763, MIT Press, Cambridge, Mass, USA, 1996.

[16] S. C. Douglas, "Simplified plant estimation for multichannel active noise control," in *Proceedings of 18th International Congress on Acoustics (ICA '04)*, Kyoto, Japan, April 2004.

[17] J. Labat, O. Macchi, and C. Laot, "Adaptive decision feedback equalization: can you skip the training period?" *IEEE Transactions on Communications*, vol. 46, no. 7, pp. 921–930, 1998.

[18] L.-Q. Zhang, A. Cichocki, and S. Amari, "Multichannel blind deconvolution of nonminimum phase systems using information backpropagation," in *Proceedings of the 6th International Conference on Neural Information Processing (ICONIP '99)*, pp. 210–216, Perth, Australia, November 1999.

[19] K. Waheed and F. M. Salam, "Cascaded structures for blind source recovery," in *Proceedings of the 45th IEEE International Midwest Symposium on Circuits and Systems (MSCAS '02)*, vol. 3, pp. 656–659, Tulsa, Okla, USA, August 2002.

[20] L.-Q. Zhang, A. Cichocki, and S. Amari, "Multichannel blind deconvolution of nonminimum-phase systems using filter decomposition," *IEEE Transactions on Signal Processing*, vol. 52, no. 5, pp. 1430–1442, 2004.

[21] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, New York, NY, USA, 2001.

[22] S. Haykin, *Unsupervised Adaptive Filtering, Volume 2: Blind Deconvolution*, John Wiley & Sons, New York, NY, USA, 2000.

[23] A. K. Nandi and S. N. Anfinsen, "Blind equalization with recursive filter structures," *Signal Processing*, vol. 80, no. 10, pp. 2151–2167, 2000.

[24] S. Amari, S. C. Douglas, A. Cichocki, and H. H. Yang, "Multichannel blind deconvolution and equalization using the natural gradient," in *Proceedings of the 1st IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications (SPAWC '97)*, pp. 101–104, Paris, France, April 1997.

[25] D. T. Pham, "Mutual information approach to blind separation of stationary sources," *IEEE Transactions on Information Theory*, vol. 48, no. 7, pp. 1935–1946, 2002.

[26] S. Amari, T.-P. Chen, and A. Cichocki, "Stability analysis of learning algorithms for blind source separation," *Neural Networks*, vol. 10, no. 8, pp. 1345–1351, 1997.

[27] L.-Q. Zhang, A. Cichocki, and S. Amari, "Geometrical structures of FIR manifold and multichannel blind deconvolution," *The Journal of VLSI Signal Processing*, vol. 31, no. 1, pp. 31–44, 2002.

[28] Y. Inouye and S. Ohno, "Adaptive algorithms for implementing the single-stage criterion for multichannel blind deconvolution," in *Proceedings of the 5th International Conference on Neural Information Processing (ICONIP '98)*, pp. 733–736, Kitakyushu, Japan, October 1998.

Bin Xia received his B.S. degree in mechanical engineering from Luoyang Institute of Technology, in 1997, and M.S. degree in mechanical engineering from Guizhou University, China, in 2001. He is currently a Ph.D. candidate of Department of Computer Sciences and Engineering, Shanghai Jiao Tong University, China. His research interests include statistical signal processing, blind signal processing, and machine learning.

Liqing Zhang received his B.S. degree in mathematics from Hangzhou University, in 1983, and the Ph.D. degree in computer sciences from Zhongshan University, China, in 1988. He became an Associate Professor in 1990 and then a Full Professor in 1995 at the Department of Automation, South China University of Technology. He joined Laboratory for Advanced Brain Signal Processing, RIKEN Brain Science Institute, Japan, in 1997 as a Research Scientist. Since 2002, he has been working in Department of Computer Sciences and Engineering, Shanghai Jiao Tong University, China. His research interests include neuroinformatics, perception computing, adaptive systems, and statistical learning. He has published more than 110 papers.