Convolutive Blind Source Separation by Efficient Blind Deconvolution and Minimal Filter Distortion

Kun Zhang[a], Lai-Wan Chan[b]

[a]Max Planck Institute for Biological Cybernetics, Spemannstr. 38, 72076 Tübingen, Germany

[b]Department of Computer Science and Engineering, Chinese University of Hong Kong, Hong Kong, China.

***Abstract*— onvolutive blind source separation (BSS) usually encounters two difficulties – the filter indeterminacy in the recovered sources and the relatively high computational load. In this paper we propose an efficient method to convolutive BSS, by dealing with these two issues. It consists of two stages, namely, multichannel blind deconvolution (MBD) and learning the post-filters with the minimum filter distortion (MFD) principle. We present a computationally efficient approach to MBD in the first stage: a vector autoregression (VAR) model is first fitted to the data, admitting a closed-form solution and giving temporally independent errors; traditional independent component analysis (ICA) is then applied to these errors to produce the MBD results. In the second stage, the least linear reconstruction error (LLRE) constraint of the separation system, which was previously used to regularize the solutions to nonlinear ICA, enforces a MFD principle of the estimated mixing system for convolutive BSS. One can then easily learn the post-filters to preserve the temporal structure of the sources. We show that with this principle, each recovered source is approximately the principal component of the contributions of this source to all observations. Experimental results on both synthetic data and real room recordings show the good performance of this method.onvolutive blind source separation (BSS) usually encounters two difficulties – the filter indeterminacy in the recovered sources and the relatively high computational load. In this paper we propose an efficient method to convolutive BSS, by dealing with these two issues. It consists of two stages, namely, multichannel blind deconvolution (MBD) and learning the post-filters with the minimum filter distortion (MFD) principle. We present a computationally efficient approach to MBD in the first stage: a vector autoregression (VAR) model is first fitted to the data, admitting a closed-form solution and giving temporally independent errors; traditional independent component analysis (ICA) is then applied to these errors to produce the MBD results. In the second stage, the least linear reconstruction error (LLRE) constraint of the separation system, which was previously used to regularize the solutions to nonlinear ICA, enforces a MFD principle of the estimated mixing system for convolutive BSS. One can then easily learn the post-filters to preserve the temporal structure of the sources. We show that with this principle, each recovered source is approximately the principal component of the contributions of this source to all observations. Experimental results on both synthetic data and real room recordings show the good performance of this method.C**

***Keywords*— Independent component analysis, Convolutive blind source separation, Least linear reconstruction error, Vector autoregression**

1. INTRODUCTION

Blind source separation (BSS) aims to recover the original sources from their observable mixtures with very little knowledge of the mixing system and the sources. In many scenarios, the original sources are approximately independent; consequently, they can be recovered by the independent component analysis (ICA) technique [1], [2], which transforms the observed data to a set of outputs that are mutually as independent as possible. In the basic ICA model, the mixing system is linear and the number of observed signals is equal to that of the original sources. In this case, under some weak conditions, ICA can recover the sources with trivial scaling and permutation indeterminacies [3].

However, for more complex mixing procedures, the recovered signals by enforcing statistical independence of the

outputs may be different from the original sources. A typical example is nonlinear ICA: it is well-known that solutions to the general nonlinear ICA problem always exist and are highly non-unique [4]. In this paper we are mainly concerned with blind separation of convolutive mixtures, or convolutive BSS (for a recent survey on convolutive BSS, one may see [5]). Since statistical independence amongst a set of signals remains if we apply a filter to each signal, solutions to this problem have the filtering indeterminacy. Many time-domain methods for this problem make the outputs both spatially and temporally as independent as possible. Consequently, if the original sources are not white, their time structures will be lost, causing distortion in the recovered signals. Hence additional information is needed to preserve the temporal information of the sources.

To make ICA result in BSS for the convolutive mixtures, we need to find some additional conditions besides statistical independence. Usually the temporal structure of the sources is approximately preserved in the convolutive mixtures. Therefore, we prefer the independent output signals whose corresponding mixing procedure is as close as possible to a linear instantaneous one, i.e., the mixing procedure is of minimal filter distortion (MFD). In this way the temporal structure in the sources could be recovered. In light of this simple idea, one can separate real room recordings with good performance. Like the minimal nonlinear distortion (MND) constraint for nonlinear ICA [6], MFD for convolutive BSS can be implemented in a simple and convenient way: under the condition that the outputs of the BSS system are independent, we prefer the BSS system that has the least linear reconstruction error (LLRE). Moreover, since convolutive BSS usually involves a large sample size and is computationally expensive, especially when applied on speech signals, we also provide a computationally appealing approach to multichannel blind deconvolution (MBD), which is a major stage in the proposed convolutive BSS method.

This paper is organized as follows. Section 2 discusses how to enforce the LLRE constraint of a given BSS system. This constraint is used to implement the MFD principle for convolutive BSS in Section 3; related work is also discussed there. Section 4 presents a convenient two-stage method, which consists of an efficient MBD approach and learning the post-filters with MFD, to achieve convolutive BSS with MFD. Experimental results on both synthetic data and real room recordings are given in Section 5.

2. BSS System with the Least Linear Reconstruction Error

Here we assume that the sources to be recovered are mutually independent, and consider a ICA-based BSS system. Denote by $\mathbf{x} = (x_1, ..., x_M)^T$ the vector of observed signals and by $\mathbf{y} = (y_1, ..., y_N)^T$ the vector of output signals of the BSS system. In addition to the independence condition, sometimes we expect the BSS system to have the least mean squared deviation from its best-fitting linear approximation, such that certain structure in the observations is approximately preserved in the separation results. Denote by R_{MSE} the mean square error (MSE) of the best-fitting linear approximation, or the LLRE, of the separation system. Let $\breve{\mathbf{A}}$ be the affine mapping which fits the transformation from $\mathbf{y}$ to $\mathbf{x}$ best and let $\breve{\mathbf{x}} = (\breve{x}_1, ..., \breve{x}_M)^T$ be its output. Let $\tilde{\mathbf{y}} = [\mathbf{y}; 1]$. $R_{MSE}(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ denotes the parameter set of the BSS system, can then be written as the MSE between x_i and $\breve{x}_i$:

$$R_{MSE}(\boldsymbol{\theta}) = E\{\| \mathbf{x} - \breve{\mathbf{x}} \|^2\} \text{ , where} \qquad (1)$$
$$\breve{\mathbf{x}} = \breve{\mathbf{A}}\tilde{\mathbf{y}}, \text{ and } \breve{\mathbf{A}} = \arg_{\mathbf{A}} \min E\{\| \mathbf{x} - \mathbf{A}\mathbf{y} \|^2\}$$

Here $\breve{\mathbf{A}}$ is an $M \times (N+1)$ matrix. If all components of $\mathbf{x}$ and $\mathbf{y}$ are zero-mean, which is usually assumed in what follows, $\breve{\mathbf{x}}$ can be obtained as $\breve{\mathbf{x}} = \breve{\mathbf{A}}\mathbf{y}$ instead, and here $\breve{\mathbf{A}}$ is an $M \times N$ matrix. The generating procedure of R_{MSE} is depicted in Figure 1.

The derivative of R_{MSE} w.r.t. $\breve{\mathbf{A}}$ is

$$\frac{\partial R_{MSE}}{\partial \breve{\mathbf{A}}} = -2E\{(\mathbf{x} - \breve{\mathbf{A}}\tilde{\mathbf{y}})\tilde{\mathbf{y}}^T\}.$$

Setting this derivative to $\mathbf{0}$ gives $\breve{\mathbf{A}}$:

$$E\{(\mathbf{x} - \breve{\mathbf{A}}\tilde{\mathbf{y}})\tilde{\mathbf{y}}^T\} = \mathbf{0} \Longrightarrow \breve{\mathbf{A}} = E\{\mathbf{x}\tilde{\mathbf{y}}^T\}[E\{\tilde{\mathbf{y}}\tilde{\mathbf{y}}^T\}]^{-1}.$$

We can see that $\breve{\mathbf{A}}$ is obtained in closed form, which greatly simplifies the expression for the LLRE R_{MSE}.

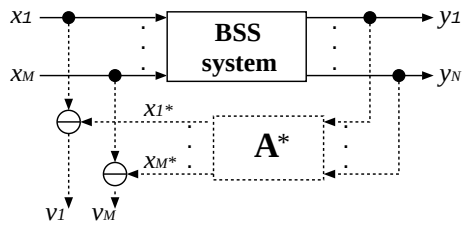

Fig. 1. Generating procedure of R_{MSE} (dashed line). $R_{MSE} = \sum_{i=1}^{M} E(v_i^2)$, where $v_i = x_i - \breve{x}_i$. Here it is assumed that $\mathbf{x}$ and $\mathbf{y}$ are zero-mean; consequently $\breve{\mathbf{x}} = \breve{\mathbf{A}}\mathbf{y}$ and $\breve{\mathbf{A}}$ is $M \times N$.

R_{MSE} can then be written as

$$\begin{aligned} R_{MSE} &= \mathrm{Tr}\big(E\{(\mathbf{x} - \breve{\mathbf{A}}\tilde{\mathbf{y}})(\mathbf{x} - \breve{\mathbf{A}}\tilde{\mathbf{y}})^T\}\big) \\ &= -\mathrm{Tr}\big(E\{\breve{\mathbf{A}}\tilde{\mathbf{y}}\mathbf{x}^T\}\big) + \text{const} \\ &= -\mathrm{Tr}\big(E\{\mathbf{x}\tilde{\mathbf{y}}^T\}[E\{\tilde{\mathbf{y}}\tilde{\mathbf{y}}^T\}]^{-1}E\{\tilde{\mathbf{y}}\mathbf{x}^T\}\big) + \text{const} \quad (2) \end{aligned}$$

Since ICA makes y_i independent from each other, y_i are uncorrelated. Moreover, we can easily make y_i zero-mean. Consequently, $E\{\tilde{\mathbf{y}}\tilde{\mathbf{y}}^T\} = \mathrm{diag}\{E(y_1^2), E(y_2^2), ..., E(y_N^2), 1\}$, and R_{MSE} becomes

$$R_{MSE} = -\sum_{j=1}^{M}\sum_{i=1}^{N} \frac{E^2(x_j y_i)}{E(y_i^2)} + \text{const} \quad (3)$$

In the update of the parameters, the gradient of R_{MSE} w.r.t. $\boldsymbol{\theta}$ is involved. Define $\mathbf{K} = (K_1, ..., K_N)^T$, with K_i given by

$$K_i = 2\sum_{j=1}^{M}\Big[\frac{E^2(x_j y_i)}{E^2(y_i^2)} y_i - \frac{E(x_j y_i)}{E(y_i^2)} x_j\Big] \quad (4)$$

One can check that the gradient of R_{MSE} w.r.t. the parameter θ_i would be $\frac{\partial R_{MSE}}{\partial \theta_i} = E\big(\mathbf{K}^T \cdot \frac{\partial \mathbf{y}}{\partial \theta_i}\big)$, where $\frac{\partial \mathbf{y}}{\partial \theta_i}$ depends on the separation system.

Recently, to alleviate the ill-posedness of nonlinear ICA, nonlinear ICA with MND, implemented by regularizing the nonlinear ICA system with the LLRE (Figure 1), was proposed; for details, see [6], [7]. Here we are interested in the use of R_{MSE} for constraining the solutions of convolutive BSS.

3. Convolutive BSS with Minimal Filter Distortion

3.1. Convolutive BSS

In convolutive BSS, the observed data $\mathbf{x}(t) = (x_1(t), \cdots, x_M(t))^T$ are assumed to be convolutive mixtures of spatially independent stochastic sequences $s_i(t)$, $i = 1, \cdots, N$. In matrix form, this generating procedure of $\mathbf{x}$ is described as $\mathbf{x}(t) = \sum_\tau \mathbf{B}_\tau \mathbf{s}(t-\tau)$, where $\mathbf{s}(t) = (s_1(t), \cdots, s_N(t))^T$. Or in the z-domain, it can be written as

$$\mathcal{X}(z) = \mathcal{B}(z)\mathcal{S}(z),$$

where $\mathcal{B}(z) = \sum \mathbf{B}_\tau z^{-\tau}$. Convolutive BSS aims to recover the source signals $s_i(t)$ from the observed signals $x_i(t)$. Denote by $\mathcal{W}(z)$ the separation system. Its output is $\mathbf{y}(t) = \sum_\tau \mathbf{W}_\tau \mathbf{x}(t-\tau)$, or

$$\mathcal{Y}(z) = \mathcal{W}(z)\mathcal{X}(z).$$

Here we assume that $N \leq M$ and that both $\mathcal{B}(z)$ and $\mathcal{W}(z)$ are stable. Previous work shows that under certain weak conditions, when the spatial independence between the output sequences $y_i(t)$ is achieved, the sources $s_i(t)$ could be recovered up to the filter and permutation indeterminacies. In other words, the learned $\mathcal{W}(z)$ satisfies

$$\mathcal{W}(z)\mathcal{B}(z) = \mathbf{P}\boldsymbol{\Lambda}(z), \quad (5)$$

where $\mathbf{P}$ is an $N \times N$ permutation matrix and $\boldsymbol{\Lambda}(z)$ is a diagonal matrix with each entry on its diagonal being a filter.

3.2. Incorporating Minimal Filter Distortion

The filter indeterminacy in convolutive BSS is analogous to the trivial indeterminacy of nonlinear ICA: both of them are caused by the fact that independence amongst a set of variables does not change by component-wise transformations of these variables. This indeterminacy is troublesome since it may cause a strong distortion in the estimate of the sources. To eliminate it, some schemes have been proposed. For example, a feedback separation structure, instead of a feedforward one, was adopted in [8].

Usually the temporal structures of the sources are approximately preserved in the convolutive mixtures. Therefore, under the independence condition of the estimated sources $y_i(t)$, we expect the transformation from $y_i(t)$ to the observed mixtures $x_i(t)$ to be as close as possible to a linear instantaneous one; in this way the filter indeterminacy is eliminated. This is

called the MFD principle of the mixing system. To achieve MND, one just needs to minimize R_{MSE}, which is defined in Eq. 1, when making the outputs of the separation system spatially independent. After tedious derivations, one can find the relationship between the estimate of $s_j(t)$ produced by MFD and the contributions of $s_j(t)$ to all observed signals $x_i(t)$, as described by the following theorem, whose proof is given in Appendix.

Theorem 1: Let $b_{ij}(t)$ be the (i,j)th entry of the mixing system $\mathbf{B}_t$. Suppose that the sources $s_i(t)$ are zero-mean and that the separation system $\mathcal{W}(z)$ satisfies Eq. 5 and has enough freedom. Then when the MFD of the mixing system is achieved, i.e., R_{MSE} defined in Eq 1 is minimized, the separation result corresponding to the source s_j is a scaled version of the principal component (PC) of the contributions of $s_j(t)$ to all observations, i.e., $[b_{ij}(t) * s_j(t)]$, $i = 1, \cdots, M$.

In fact, another "minimal distortion" principle [9]–[11] has been incorporated for regularizing the separation system $\mathcal{W}(z)$ in the literature.[1] Originally, in [9], the authors proposed to achieve the minimal distortion of the separation system by minimizing $E\{||\mathbf{y}_t - \mathbf{x}_t||^2\}$. It was later changed to

$$\tilde{R}_{MFD} = E\{||\mathbf{y}(t) - \mathbf{Q}\mathbf{x}(t)||^2, \tag{6}$$

with the matrix $\mathbf{Q}$ pre-assigned [10]. In this method, the determination of $\mathbf{Q}$ requires certain prior knowledge. Moreover, the function $\tilde{R}_{MFD}$ is generally sensitive to the permutation of $y_i(t)$, i.e., different permutations of $y_i(t)$ may result in different estimates of the same source. With this regularization technique, the inherent permutation indeterminacy in the BSS problem would have some random effects on the recovered sources. Therefore, generally speaking, it would be better to let $\mathbf{Q}$ be the best-fitting linear transformation from $\mathbf{x}(t)$ to $\mathbf{y}(t)$, i.e., to use $\breve{\mathbf{Q}} = \arg\min_{\mathbf{Q}} \tilde{R}_{MFD}$ instead, just like the way we find $\breve{\mathbf{A}}$ in Section 2. In addition, minimization of $\tilde{R}_{MFD}$ tends to make the variance of the outputs $y_i(t)$ smaller and smaller.

Compared to the one exploiting $\tilde{R}_{MFD}$ given by Eq. 6 [10], the proposed scheme to enforce the MFD principle has some nice properties. Firstly, unlike $\tilde{R}_{MFD}$, which is sensitive to the matrix $\mathbf{Q}$, R_{MSE} (Eq. 1) is a faithful measure of the filter distortion level. It is also insensitive to the scaling of y_i. Moreover, the result of the proposed scheme is insensitive to the permutations of $x_i(t)$. Secondly, using the proposed scheme, we can easily incorporate any prior knowledge on the filter distortion level of the generating procedure of each $x_i(t)$. For instance, if we believe that the distortion in a particular observation $x_k(t)$ w.r.t. the original sources caused by the mixing filters is significant, we may reduce the variance of $x_k(t)$ in R_{MSE} or even drop it, to reduce the effect of $x_k(t)$. Thirdly, when $M > N$, the proposed scheme also enforces a minimal energy loss of the separation system, such that the sources which contribute more to the observations would be easier to be recovered. Finally, as shown in Subsection 4, convolutive BSS with MFD could not be achieved by regularizing MBD with the MFD condition; we will propose a simple two-stage procedure to do so.

A natural way to implement convolutive BSS with MFD is to adopt the mutual information minimization method [12] with MFD for regularization. Minimization of mutual information between the output sequences makes the outputs spatially independent, and MFD helps to preserve the temporal information of the sources. Unfortunately, the mutual information minimization method for convolutive BSS involved the estimation of some variants of the joint densities, which is computationally expensive and is not suitable when the data dimension is high [12]. Below we present an efficient two-stage procedure to perform convolutive BSS by combining a computationally appealing MBD approach and MFD.

4. By a Two-Stage Method

MBD using information maximization [8], [13] or the natural gradient [14], [15] is much easier to do and involves

[1] We would like to address that the "inverse minimal distortion" principle given in [10] is essentially different from our MFD criterion: the "inverse minimal distortion" principle minimizes the square error between the observed mixtures $\mathbf{x}(t)$ and the reconstructed ones from the outputs with the pseudo-inverse of the separation system, $\mathcal{W}^{\dagger}(z)\mathbf{y}(t)$. Consequently, this principle reduces the noise effect in the over-determined case ($N < M$), and it has no effect at all in the square case ($N = M$), since it is always zero from any invertible $\mathcal{W}(z)$.

lower computational loads than convolutive BSS. Unlike convolutive BSS, MBD makes the output sequences both spatially and temporally independent, and it does not have the filter indeterminacy. One should note that *convolutive BSS with MFD could not be achieved by regularizing MBD with MFD* – If we use the MFD to regularize the results of MBD, in order to avoid the temporal whitening effect, the regularization parameter must be large; a large regularization parameter would make the mixing system too close to a linear instantaneous transformation such that the spatial independence between output sequences may be violated.

However, we can achieve convolutive BSS by combining MBD and MFD in two separate stages. In the first stage we perform multichannel blind deconvolution on $\mathbf{x}(t)$. Inspired by the method for analyzing Granger causality with instantaneous effects [16], below we will propose a computationally very appealing approach to MBD. Denote by $\tilde{\mathbf{y}}(t) = (\tilde{y}_1(t), \cdots, \tilde{y}_N(t))^T$ the output of multichannel blind deconvolution. The expected outputs of convolutive BSS, $y_i(t)$, are a filtered version of $\tilde{y}_i(t)$, i.e., $y_i(t) = e_i(t) * \tilde{y}_i(t) = \sum_\tau e_i(\tau) \cdot \tilde{y}_i(t-\tau)$, where $e_i(t)$ can be considered as post-filters. In the second stage, one can determine these post-filters by making use of the MFD principle. Below we discuss these two stages in detail.

4.1. Stage 1 – VAR-ICA: An Efficient Approach to MBD

MBD of speech signals usually involves quite a lot of samples and parameters. Hence, a MBD algorithm is expected to be computationally efficient in practice. However, traditionally, in the MBD procedure, all parameters are tuned simultaneously, making the learning speed rather slow. Below we propose a simple and efficient approach to MBD, by combining the vector autoregression (VAR) model and instantaneous ICA. VAR transforms the mixtures to the error series which are temporally as independent as possible, and ICA further makes them instantaneously as independent as possible; as a consequence, the final outputs are independent both spatially and temporally.

Suppose that causal and minimum-phase finite impulse response (FIR) filters with a suitable length are used for MBD; the corresponding mixing system would be causal, stable, and minimum-phase. The MBD problem can be formulated as

$$\tilde{\mathbf{y}}(t) = \sum_{\tau=0}^{L} \widetilde{\mathbf{W}}_\tau \cdot \mathbf{x}(t-\tau), \tag{7}$$

where $\tilde{\mathbf{y}}(t) = (\tilde{y}_1(t), ..., \tilde{y}_N(t))^T$ are spatially and temporally as independent as possible. Here for simplicity we have assumed that the source number N and the observation number M are equal and that the data are zero-mean. Eq. 7 can be re-written as

$$\begin{aligned} & \widetilde{\mathbf{W}}_0 \mathbf{x}(t) = -\sum_{\tau=1}^{L} \widetilde{\mathbf{W}}_\tau \mathbf{x}(t-\tau) + \tilde{\mathbf{y}}(t) \\ \Rightarrow \quad & \mathbf{x}(t) = -\sum_{\tau=1}^{L} \widetilde{\mathbf{W}}_0^{-1} \widetilde{\mathbf{W}}_\tau \mathbf{x}(t-\tau) + \widetilde{\mathbf{W}}_0^{-1} \tilde{\mathbf{y}}(t) \\ \Rightarrow \quad & \mathbf{x}(t) = -\sum_{\tau=1}^{L} \mathbf{M}_\tau \mathbf{x}(t-\tau) + \boldsymbol{\epsilon}(t), \end{aligned} \tag{8}$$

where

$$\mathbf{M}_\tau \triangleq \widetilde{\mathbf{W}}_0^{-1} \widetilde{\mathbf{W}}_\tau, \text{ and } \boldsymbol{\epsilon}(t) \triangleq \widetilde{\mathbf{W}}_0^{-1} \tilde{\mathbf{y}}(t). \tag{9}$$

As $\tilde{\mathbf{y}}(t)$ are assumed to be spatially and temporally independent, the errors $\boldsymbol{\epsilon}(t)$, as a linear instantaneous transformation of $\mathbf{y}(t)$, are temporally independent. Thus Eq. 8 is exactly a vector autoregression (VAR) model, and all parameters involved in Eq. 8 can be conveniently estimated by multivariate least squares (MLS) [17]. In this step we implicitly assumed that the data are normally distributed, but the estimator is consistent in the statistical asymptotic sense. Moreover, $\mathbf{M}_\tau$ in Eq. 8 are estimated in closed form, rather than in an iterative manner, so this step involves light computational loads, and there are no local optimum issues.

Once we obtain the estimate of $\mathbf{M}_\tau$, the errors $\boldsymbol{\epsilon}(t)$ are easily constructed. As $\boldsymbol{\epsilon}(t)$ are a linear instantaneous mixture of the independent signals $\tilde{y}_i(t)$, by applying traditional ICA to the estimate of $\boldsymbol{\epsilon}(t)$, one can find the estimate of $\widetilde{\mathbf{W}}_0$ and $\tilde{\mathbf{y}}(t)$ in Eq. 9. In our implementation, FastICA [18] is adopted. If needed, all $\widetilde{\mathbf{W}}_\tau$ can then be estimated by making use of Eq. 9.

Finally, MBD is achieved in the two separate steps given above, and both steps are computationally attractive. As the

MBD algorithm consists of two separate steps, the estimate of the parameters may not be statistically efficient; however, in practice it would not be a serious problem since we usually have a large number of samples for MBD.

4.2. Stage 2 – Learning the Post-Filters by Enforcing MFD

In the second stage, we need to find the post-filters $e_i(\tau)$, $i = 1, ..., N$, using the MFD principle. The choice of the form and order of the filters $e_i(\tau)$ depends on the auto-correlation properties of the sources. Here we let $e_i(\tau)$ be finite impulse response (FIR) filters.

The learning rule for $e_i(\tau)$ can be derived by minimizing the linear reconstruction error R_{MSE} (Eq. 3). Noting $y_i(t) = e_i(t) * \tilde{y}_i(t)$, we have

$$\frac{\partial R_{MSE}}{\partial e_i(\tau)} = E_t\{K_i(t) \cdot \tilde{y}_i(t-\tau)\} \tag{10}$$

where $K_i(t)$ is defined in Eq. 4 (the only difference is that K_i in Eq. 4 does not have the time index t). In this way, MFD provides a method to learn the post-filters in convolutive BSS.

We can easily find a rough estimate for $e_i(\tau)$, which can be used for initializing the gradient-based method (Eq. 10). $\tilde{y}_i(t)$, the outputs of MBD, have been estimated in the first stage. For simplicity, we assume that $\tilde{y}_i(t)$ have been made of unit variance. Let us find the filters $e_i(\tau)$ such that $y_i(t) = e_i(t) * \tilde{y}_i(t)$ can best reconstruct a particular observation, denoted by $x_k(t)$, instead of all observations, with a linear instantaneous transformation. That is, we aim to minimize $R_k = E_t\{||x_k(t) - \breve{\mathbf{a}}^T\mathbf{y}(t)||^2\} = \{||x_k(t) - \sum_{j=1}^N \breve{a}_j y_j(t)||^2\}$, where $\breve{\mathbf{a}} = \arg_{\mathbf{a}} \min E_t\{||x_k(t) - \mathbf{a}^T\mathbf{y}(t)||^2\}$. Without loss of generality, we absorb $\breve{a}_j$ into $e_j(\tau)$, or equivalently set $\breve{a}_j = 1$. The minimum is achived when $\frac{\partial R_k}{e_i(\tau)} = 0$, i.e.,

$$E_t\Big\{\big(x_k(t) - \sum_{j=1}^{N}[e_j(t) * \tilde{y}_j(t)]\big) \cdot \tilde{y}_i(t-\tau)\Big\} = 0.$$

Bearing in mind that $\tilde{y}_j(t)$ are spatially and temporally approximately independent, finally we can see

$$e_i(\tau) = E_t\{x_k(t)\tilde{y}_i(t-\tau)\},$$

which is very easy to calculate. One can then use Eq. 10 to refine the result if needed.

The MATLAB source code implementing the proposed two-stage method for convolutive BSS is available at http://www.cs.helsinki.fi/u/kunzhang/cbss_mfd.html. In practice this method is very fast; we found that it takes about 20 seconds to separate two sources with 1.2×10^5 samples on a 2.0GHz PC, with the filter length of about 1000.

5. EXPERIMENTS

To illustrate the performance of the proposed method for convolutive BSS with MFD, we report some experimental results on the separation of convolutive mixtures of speech signals. The two-stage method given in Section 4 was used, and in the first stage, to do MBD, the proposed VAR-ICA approach was adopted. We used both the signal-to-interference ratio (SIR) and the signal-to-noise ratio (SNR) to measure the separation performance. Suppose that y_i provides the estimate of s_i. Its SIR is defined as $\text{SIR}_i = 10\log_{10}\frac{E\{y_i^2|_{s_j=0,\forall j\neq i}\}}{E\{y_i^2|_{s_i=0}\}}$, where $y_i|_{s_j=0}$ stands for what is at the i-th output, when the source $s_j(t)$ is zero, so $y_i|_{s_i=0}$ is the interference at the i-th output caused by other sources. A high SIR means that the corresponding source is recovered up to the filter indeterminacy with good performance. In addition, considering the recovered signal $y_i(t)$ as a sum of a scaled version of the original source $s_i(t)$ and some noise, we can define the SNR of the ith channel. A high SNR means that both the interference of other sources and the distortion caused by the filter indeterminacy are small.

5.1. On Artificially Mixed Convolutive Mixtures

We first used the artificially mixed convolutive mixtures of speech signals. The three source speech signals, as well as the magnitude of their Fourier transforms, are shown in Figure 2. Each signal has 9000 samples. Each entry of the convolutive mixing system $\mathcal{B}(z)$ has a length of 60, and it was generated randomly from the uniform distribution $\mathcal{U}(-c, c)$. To make $\mathcal{B}(z)$ tend to be invertible by a causal system, we decreased c

as the time lag increases. The convolutive mixtures, together with their Fourier transforms, are shown in Figure 3.

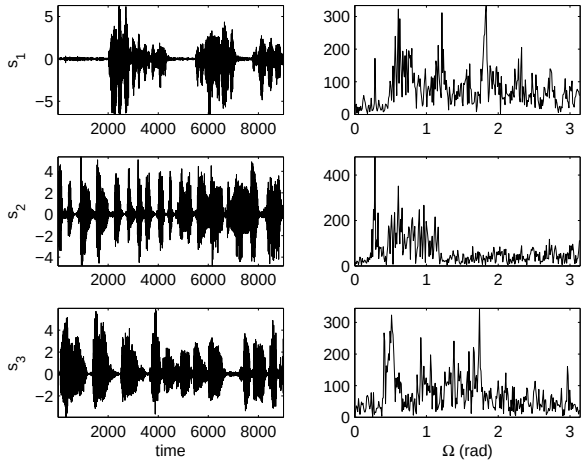

Fig. 2. Sources used for blind separation of convolutive mixtures with MFD. Left: sources. Right: magnitude of their Fourier transforms.

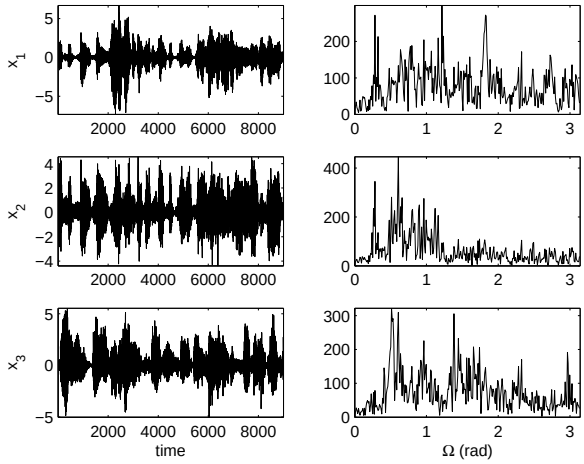

Fig. 3. Convolutive mixtures. Left: mixtures. Right: magnitude of their Fourier transforms.

We adopted the two-stage method given in Section 4 for source separation. Each entry of the deconvolution system $\mathcal{W}(z)$ has a length of 101. The MBD results $\tilde{y}_i(t)$ produced by the first stage are shown in Figure 4. From their Fourier transforms, we can see that $\tilde{y}_i(t)$ are approximately white. In the second stage, we then applied the post-filters $e_i(t)$ to reduce the temporal distortion in the outputs. We adopted FIR filters with the length 200 for $e_i(t)$. $e_i(t)$ were learned using the rule Eq. 10. The final outputs $y_i(t) = e_i(t) * \tilde{y}_i(t)$ are given in Figure 5. Comparing their Fourier transforms as well as their waveforms to those of the original sources shown in Figure 2, one can see that the temporal structure of the sources is approximately recovered.

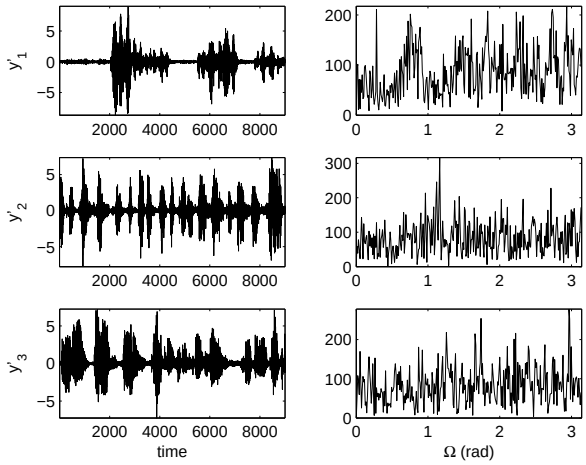

Fig. 4. Results of MBD by VAR-ICA (Subsection 4.1). Left: recovered sources. Right: magnitude of their Fourier transforms. Corresponding SIR's are 22.84dB, 20.49dB, and 19.75dB, respectively.

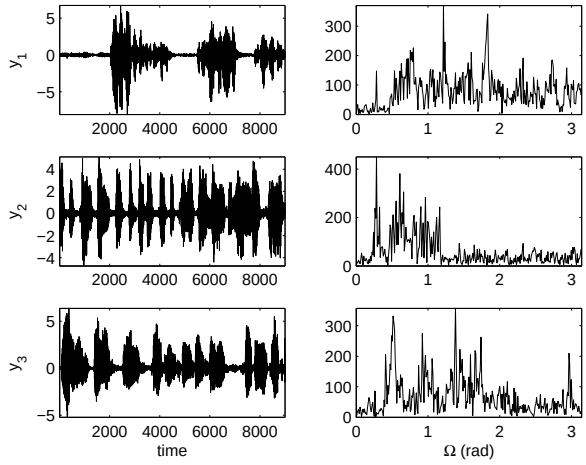

Fig. 5. Recovered sources by convolutive BSS with MFD. Left: recovered sources. Right: magnitude of their Fourier transforms.

The SIR's in the three channels are 22.84dB, 20.49dB, and 19.75dB, respectively, meaning that each source is recovered with very little interference from others. Figure 6 (left) shows the scatter plot of each recovered source $y_j(t)$ versus the PC of $[b_{ij}(t) * s_j(t)]$ ($i = 1, 2, 3$), the contributions of the corresponding source to all observations. It is almost a straight

line. The SNR's of $y_i(t)$ w.r.t. the corresponding PC's are 15.89dB, 17.44dB, and 17.18dB, respectively, which are very high. This is consistent with Theorem 1, which states that the recovered sources with MFD are approximately the PC's of their contributions to all observations. Figure 6 (right) plots the scatter plot of $y_i(t)$ versus the original source $s_i(t)$. The SNR's of $y_i(t)$ w.r.t. $s_i(t)$ in the three channels are 7.39dB, 13.35dB, and 7.60dB, respectively.

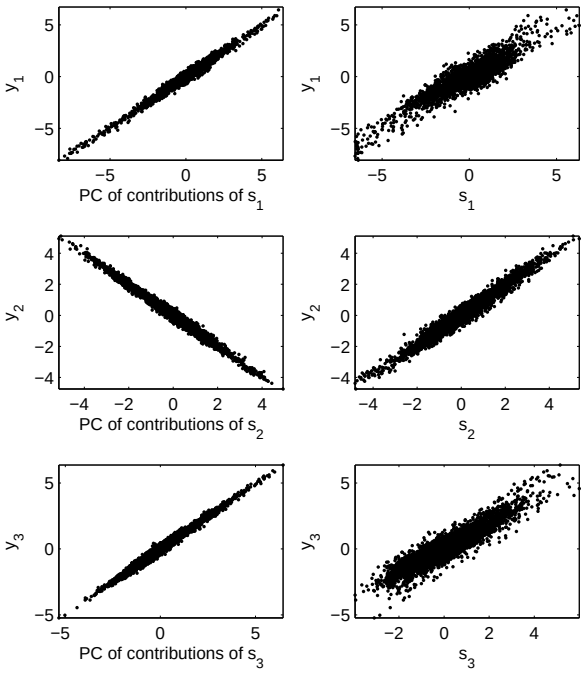

Fig. 6. Scatter plot of each recovered source by convolutive BSS with MFD versus the PC of the contributions of the source, as well as the original source. Left: recovered source versus the PC of the contributions of the original source to all mixtures. Corresponding SNR's are 15.89dB, 17.44dB, and 17.18dB, respectively. Right: recovered source versus the original one. Corresponding SNR's are 7.39dB, 13.35dB, and 7.60dB, respectively.

For comparison, we also used Parra & Spence's method [19] to separate the convolutive mixtures.[2] This method exploits the inherent non-stationarity of the acoustic sources and uses cross-correlations at multiple times for separation. As mentioned in a recent survey of convolutive BSS methods [5], it gives comparatively good performance for real room recordings. The SIR's of the outputs produced by this method are 12.43dB, 14.12dB, and 13.66dB, respectively, which are lower than those produced by the two-stage method. To see if the temporal structure of the sources is preserved, we give the scatter plot of each recovered signal versus the PC's of the corresponding source to the observations (just shown for completeness) and the original source; see Figure 7. The original sources are recovered with the SNR's 5.06dB, 7.96dB, and 7.50dB, respectively. Compared to the two-stage method, in this experiment this method caused larger temporal distortion in the recovered sources.

[2]Thanks to Stefan Harmeling for sharing the MATLAB source code.

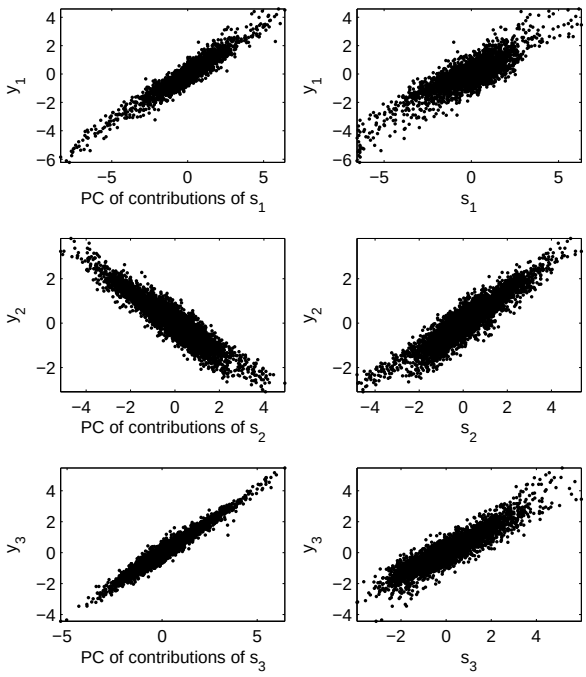

Fig. 7. Scatter plot of each recovered source by Parra & Spence's method [19] versus the PC of the contributions of the source and the original source. SIR's in the three channels are 12.43dB, 14.12dB, and 13.66dB, respectively. Left: recovered source versus the PC of the contributions of the corresponding source to all mixtures. Corresponding SNR's are 12.43dB, 14.12dB, and 13.66dB, respectively. Right: recovered source versus the original one. Corresponding SNR's are 5.06dB, 7.96dB, and 7.50dB, respectively.

5.2. On Real Room Recordings

Thanks to the computational efficiency, the proposed two-stage method for convolutive BSS can be applied to relatively large scale problems. We applied this method to the various

mixture signals recorded by Lee et al.[3] For real room recordings, the SNR is not measurable due to the unavailability of the original individual speech or music signals. One can evaluate the results by listening to the separated signals.[4] When listening, one needs to pay attention to two aspects: the quality of the separation between different sources and the sound quality of each recovered source. One can see that the original signals are successfully recovered, and that the separation results are clearly better than, or at least as good as, those separated by [20] and [21], especially for the sound quality (indicating the temporal distortion in the estimated sources).

6. Conclusion

In this paper we considered the problem of convolutive BSS, and proposed the minimal filter distortion (MFD) principle to avoid the filter indeterminacy in the output and improve the separation results. MFD makes the filter distortion in the estimated mixing procedure as weak as possible, and is implemented by the least linear reconstruction error constraint of the separation system. We showed that the recovered source with this principle is approximately the principal component of the contributions of this source to all observations. We gave a two-stage method for convolutive BSS by combining multichannel blind deconvolution and MFD. In particular, as speech signals usually have a large sample size, we proposed a computationally very efficient two-step approach to perform multichannel blind deconvolution in the first stage. The two-stage method for convolutive BSS is computationally appealing, and its good performance has been demonstrated on both synthetic data and real room recordings.

Appendix: Proof of Theorem 1

Proof: We first present the following lemma, as it will be used in the proof.

[3]available at
`http://www.cnl.salk.edu/~tewon/Blind/blind_audio.html`.

[4]The separated results by our method, as well as the MATLAB source code, are available at
`http://www.cs.helsinki.fi/u/kunzhang/cbss_mfd.html`.

Lemma 1: Suppose we are given the random vector $\mathbf{d} = (d_1, d_2, \cdots, d_n)^T$. Let R_y be the mean square error of reconstructing $\mathbf{d}$ from the variable y with the best-fitting linear transformation, i.e., $R_y = \min_{\mathbf{a}} E\{||\mathbf{d} - \mathbf{a} \cdot y||^2\}$, where $\mathbf{a} = (a_1, a_2, \cdots, a_n)^T$. The variable y which gives the minimum R_y is a scaled version of the non-centered principal component (PC) of $\mathbf{d}$, and if y is constrained to be zero-mean, it is a scaled version of the PC of $\mathbf{d}$.

The proof is given in [6]. Note that this lemma is not straightforward. In fact, compared to the definition of PCA [22], here y is not constrained to be a linear combination of d_i, although finally it turns out to be so.

Now we are ready to prove Theorem 1. Since $\mathcal{W}(z)$ satisfies Eq. 5, we can denote by $\iota_j(t) * s_j(t)$ the estimate of $s_j(t)$. Let $\breve{a}_{ij}$ denote the (i,j)th entry of $\breve{\mathbf{A}}$. R_{MSE} defined in Eq 1 is

$$\begin{aligned}
& R_{MSE} \\
= & \sum_i E_t\Big\{x_i(t) - \sum_j \breve{a}_{ij} \cdot [\iota_j(t) * s_j(t)]\Big\}^2 \\
= & \sum_i E_t\Big\{\sum_j \Big[[b_{ij}(t) * s_j(t)] - \breve{a}_{ij} \cdot [\iota_j(t) * s_j(t)]\Big]\Big\}^2 \\
= & \sum_i \Big\{\sum_j E_t\big([b_{ij}(t) * s_j(t)] - \breve{a}_{ij} \cdot [\iota_j(t) * s_j(t)]\big)^2 \\
& + \sum_{k\neq l} E_t\Big[\big([b_{ik}(t) * s_k(t)] - \breve{a}_{ik} \cdot [\iota_j(t) * s_k(t)]\big) \\
& \qquad \cdot \big([b_{il}(t) * s_l(t)] - \breve{a}_{il} \cdot [\iota_j(t) * s_l(t)]\big)\Big]\Big\}.
\end{aligned}$$

As $s_j(t)$ are zero-mean spatially independent stochastic sequences, the above equation further becomes

$$R_{MSE} = \sum_i \sum_j E\big([b_{ij}(t) * s_j(t)] - \breve{a}_{ij} \cdot [\iota_j(t) * s_j(t)]\big)^2.$$

As $\mathcal{W}(z)$ has enough freedom, $\iota_j(t)$ also has enough freedom. Consequently, according to Lemma 1, when $\iota_j(t)$ minimizes R_{MSE}, one can see that $[\iota_j(t) * s_j(t)]$ is the PC of $[b_{ij}(t) * s_j(t)]$, $i = 1, \cdots, M$, multiplied by a constant. ■

References

[1] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. John Wiley & Sons, Inc, 2001.

[2] A. Cichocki and S. Amari. *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. John Wiley & Sons, UK, corrected and revisited edition edition, 2003.

[3] P. Comon. Independent component analysis – a new concept? *Signal Processing*, 36:287–314, 1994.

[4] A. Hyvärinen and P. Pajunen. Nonlinear independent component analysis: Existence and uniqueness results. *Neural Networks*, 12(3):429–439, 1999.

[5] U. Kjems L.C. Parra M.S. Pedersen, J. Larsen. A survey of convolutive blind source separation methods. In *Springer Handbook of Speech Processing*, 2007.

[6] K. Zhang and L. Chan. Minimal nonlinear distortion principle for nonlinear independent component analysis. *Journal of Machine Learning Research*, 9:2455–2487, 2008.

[7] K. Zhang and L. Chan. Kernel-based nonlinear independent component analysis. In *Proc. Int. Workshop on Independent Component Analysis and Signal Separation (ICA2007)*, pages 301–308, London, UK, Sept. 2007.

[8] K. Torkkola. Blind separation of convolved sources based on information maximization. In *IEEE Workshop on Neural Networks for Signal Processing*, pages 423–432, Kyoto, Japan, 1996.

[9] K. Matsuoka and S. Nakashima. Minimal distortion principle for blind source separation. In *Proc. Int. Conf. Independent Component Analysis and Blind Signal Separation (ICA 2001)*, pages 722–727, 2001.

[10] K. Matsuoka. Minimal distortion principle for blind source separation. In *Proceedings of the 41st SICE Annual Conference (SICE 2002)*, pages 2138–2143, 2002.

[11] K. Matsuoka. Elimination of filtering indeterminacy in blind source separation. *Neurocomputing*, 71:2113–2126, 2008.

[12] M. Babaie-Zadeh, C. Jutten, and K. Nayebi. Separating convolutive mixtures by mutual information minimization. In *Proc. IWANN*, volume 2, pages 834–842, Granada, Spain, June 2001.

[13] A.J. Bell and T.J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159, 1995.

[14] S. Amari, S.C. Douglas, A. Cichocki, and H.H. Yang. Multichannel blind deconvolution and equalization using the natural gradient. In *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications*, pages 101–104, Paris, France, 1997.

[15] S. Choi, S. Amari, A. Cichocki, and R. Liu. Natural gradient learning with a nonholonomic constraint for blind deconvolution of multiple channels. In *First International Workshop on Independent Component Analysis and Signal Separation*, pages 371–376, Aussois, France, 1999.

[16] A. Hyvärinen, S. Shimizu, and P. O. Hoyer. Causal modelling combining instantaneous and lagged effects: an identifiable model based on non-gaussianity. In *Proceedings of the 25th International Conference on Machine Learning (ICML2008)*, pages 424–431, Helsinki, Finland, 2008.

[17] J.D. Hamilton. *Time Series Analysis*. Princeton University Press, 1995.

[18] A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634, 1999.

[19] L. Parra and C. Spence. Convolutive blind separation of non-stationary sources. *IEEE Transactions on Speech and Audio Processing*, 8(3):320–327, 2000.

[20] N. Murata, S. Ikeda, and A. Ziehe. An approach to blind source separation based on temporal structure of speech signals. *Neurocomputing*, 41:1–24, 2001.

[21] T.W. Lee, A. Ziehe, R. Orglmeister, and T. Sejnowski. Combining time-delayed decorrelation and ica: Towards solving the cocktail party problem. In *In Proc. ICASSP98*, pages 1249–1252, 1998.

[22] I. J. Jolliffe. *Principal Component Analysis*. Springer series in Statistics. Springer Verlag, 2nd edition, 2002.