A New Frequency Domain Method for Blind Source Separation of Convolutive Audio Mixtures

Kamran Rahbar, James P. Reilly*

Abstract

In this paper we propose a new frequency domain approach to blind source separation (BSS) of audio signals mixed in a reverberant environment. It is first shown that joint diagonalization of the cross power spectral density matrices of the signals at the output of the mixing system is sufficient to identify the mixing system at each frequency bin up to a scale and permutation ambiguity. The frequency domain joint diagonalization is performed using a new and quickly converging algorithm which uses an alternating least-squares (ALS) optimization method. The inverse of the mixing system, estimated using the joint diagonalization algorithm, is then used to separate the sources. An efficient diadic algorithm to resolve the frequency dependent permutation ambiguities that exploits the inherent non-stationarity of the sources is presented. The effect of the unknown scaling ambiguities is partially resolved using a novel initialization procedure for the ALS algorithm.

The performance of the proposed algorithm is demonstrated by experiments conducted in real reverberant rooms. The algorithm demonstrates good separation performance and enhanced output audio quality. The proposed algorithm is compared to that of Parra [1]. Audio results are available at "www.ece.mcmaster.ca/˜reilly/kamran/index.htm".

I. Introduction

In the blind source separation problem the objective is to separate multiple sources, mixed through an unknown mixing system(channel), using only the system output data (observed signals) and in particular without using any (or least amount of) information about the sources or the system. The blind source separation problem arises in many fields of studies, including speech processing, data communication, biomedical signal processing, etc.

An extensive part of the literature has been directed toward the simple case of instantaneous mixing; i.e., when the observed signals are a linear combination of the sources and no time-delays are involved in the mixing model [2–5]. A more challenging case is when the mixing system is convolutive; i.e., when the sources are mixed through a linear filtering operation and the observed signals are linear combinations of the sources and their corresponding delayed versions [6–9]. A difficult practical example for the convolutive BSS problem is separation of audio signals mixed in a reverberant environment.

The current literature on blind signal separation can be divided into the higher-order statistics (HOS) [4, 10–12] and second-order statistics (SOS) [13–16] methods. The criterion used most often in the SOS category is minimizing a correlation function of the observed signals subject to a constraint on the separating network or the power of the output signals. The main motivation behind the use of SOS methods is that estimating the correlation functions is easier and more robust compared with estimating higher-order moments, required in most HOS based methods.

Department of Electrical & Computer Engineering
McMaster University, 1280 Main St. W, Hamilton, Ontario, Canada L8S 4K1
* Corresponding Author: email: reillyj@mcmaster.ca, ph: 905 525 9140 x22895, fax: 905 523 4407.

The SOS methods usually have a simple implementation, require fewer data samples, and contrary to HOS methods, they can handle Gaussian distributed inputs. A significant disadvantage of SOS methods is that they require additional assumptions on both the channel and the source signals. For example, one cannot use SOS methods for blind separation of instantaneously mixed uncorrelated white signals unless additional constraints are available on the mixing system. For colored sources, the main assumption used by SOS methods is that the autocorrelation coefficients of the sources should be linearly independent [17]. Also refer to [15] and [18] for blind identifiability conditions of SOS methods applied to convolutive mixing.

In recent years a few blind source separation methods have been proposed that exploit the non-stationarity of the source signals. A non-stationarity assumption can be justified by realizing that most real world signals are inherently non-stationary (e.g., speech or biological signals). In [19, 20], methods have been proposed for blind source separation of instantaneously mixed, non-stationary signals (cyclostationary in the second reference), while [1] considers blind source separation of colored non-stationary signals when the mixing system is convolutive. Based on the non-stationarity assumption, the methods in [21, 22] can identify the underlying convolutive mixing system which later can be used to recover the sources.

For the convolutive mixing problem, we can also divide the existing methods into frequency domain [1, 12, 23]and time domain approaches [6, 24, 25]. The advantage of using frequency-domain methods is that a time domain estimation problem with a large number of parameters is decomposed into multiple, independent estimation problems, each with fewer parameters to be estimated. As a result, in general the frequency-domain estimation algorithms have a simpler implementation and better convergence properties over their time–domain counterparts. The main difficulties with frequency–domain blind source separation of convolutive mixtures however is the arbitrary permutation and scaling ambiguities of the estimated frequency response of the un-mixing system at each frequency bin.

The proposed method in this paper exploits second-order non-stationarity of the input signals to separate the convolved audio sources. Sufficient identifiability conditions under which non-stationary sources can be separated in a convolutive mixing environment using only second-order statistics of the observed signals has been discussed in [22] [26] [27]. These works propose a frequency-domain method based on joint approximate diagonalization of observed signals' cross spectral density matrices, estimated at different time epochs. In this paper, we extend this idea to concentrate on the problem of separating audio sources mixed in real reverberant environments.

Other methods, such as [1], also exploit non–stationarity of the observed signals in the frequency domain. However, the approach offered in this paper differs from previous works in at least three respects. First, the approach in [1] is based on a gradient descent procedure, while the method proposed in this paper is based on a more efficient alternating least–squares with projection (ALSP) method. Second, an efficient method for addressing the permutation problem inherent with frequency domain approaches is proposed. This algorithm exploits the non–stationarity of the observed signals, and the spectral correlation of the sources between adjacent frequency bins. Most audio signals of interest possess these properties. Thirdly, the scaling problem, also inherent with frequency domain approaches, is addressed in this paper by a novel initialization procedure that significantly improves the quality of the separated audio, as has been demonstrated in our real room audio experiments.

So far the results shown for most of the convolutive blind source separation methods, especially the HOS methods, are limited to computer simulations, using synthetically generated mixing channels with small orders. Among these methods there are some that only consider two-input two-output (TITO) mixing systems [7,8,15] and some that assume mixing systems with the same

number of inputs and outputs [28]. There are few methods that consider blind source separation of acoustic signals in real room situations [1,29]. Nevertheless the performance of these algorithms requires some improvement in highly reverberant environments, in order to achieve satisfactory subjective audio quality. In this paper we show our results for real room experiments with long reverberation times. The proposed method is not limited to mixing systems with fixed dimensions nor to ones with the same number of outputs and inputs. The only requirement is that the number of outputs for the mixing system be greater than or equal to the number of inputs.

The organization of this paper is as follows: The problem formulation including the set of required assumptions is presented in Section II. Section III discuses the joint diagonalization problem including some new identifiability results based only on the second–order statistics and the non-stationarity property of the input signals. In Section IV we present a frequency-domain algorithm for convolutive blind source separation. A novel solution to the permutation problem is discussed in Section V. Simulation results for synthetic convolutive mixing scenarios are described in Section VI and real room experimental results are described in Section VII. The first set of experiments are based on real room recordings done in a small office environment with a moderate reverberation time. The second set of experiments are done in a conference room with a highly reverberant characteristic. In all these experiments, the original sources (speech signals) are successfully recovered with good audio quality. We also compare our results with those obtained using the method in [1]. Conclusions and final remarks are presented in Section VIII.

II. Problem statement

A. Notation

We use bold upper and lower case letters to show matrices and vectors respectively. The remaining notational conventions are listed as follows:

$(\cdot)^T$	Transpose
$(\cdot)^\dagger$	Hermitian transpose
$E\{\cdot\}$	Expectation Operator
diag($\mathbf{a}$)	Forms a diagonal matrix from the vector $\mathbf{a}$.
diag($\mathbf{A}$)	Forms a column vector from the diagonal elements of $\mathbf{A}$.
vec$\{\mathbf{A}\}$	Forms a column vector by stacking the columns of $\mathbf{A}$.
mat$\{\mathbf{a}\}$	Forms a $J \times J$ matrix from a $J^2 \times 1$ column vector $\mathbf{a}$
$\mathbf{A}^+$	Pseudo Inverse of the matrix $\mathbf{A}$.

B. The Convolutive Mixing Problem

We consider the following N-source J-sensor multi–input multi–output (MIMO) linear model for the received signal for the convolutive mixing problem[1]:

$$\mathbf{x}(t) = [\mathbf{H}(z)]\mathbf{s}(t) + \mathbf{n}(t) \quad t \in \mathbb{Z} \tag{1}$$

where $\mathbf{x}(t) = (x_1(t), \cdots, x_J(t))^T$ is the vector of observed signals,
$\mathbf{s}(t) = (s_1(t), \cdots, s_N(t))^T$ is the vector of sources, $\mathbf{H}(z)$ is the $J \times N$ transfer function of the mixing system and $\mathbf{n}(t) = (n_1(t), \cdots, n_J(t))^T$ is the additive noise vector. We assume $h_{ij}(z)$,

[1]Here we use the notation $[\mathbf{H}(z)]\mathbf{s}(t)$ to denote the convolution between a system with z-transform $\mathbf{H}(z)$ and source vectors $\mathbf{s}(t)$.

the ij_{th} element of $\mathbf{H}(z)$, to be a rational function of z. For the special case where the $h_{ij}(z)$ are causal FIR filters, we have $\mathbf{H}(z) = \sum_{t=0}^{L} \mathbf{H}(t)z^{-t}$ where L is the highest polynomial degree of $h_{ij}(z)$ for all i, j. The objective of the blind source separation algorithm is to estimate the un-mixing filters $\mathbf{W}(z)$ from the observed signals $\mathbf{x}(t)$ such that

$$\mathbf{W}(z)\mathbf{H}(z) = \mathbf{\Pi}\mathbf{D}(z) \tag{2}$$

where $\mathbf{\Pi} \in \mathbb{R}^{N\times N}$ is a permutation matrix and $\mathbf{D}(z)$ is a constant diagonal matrix with diagonal elements which are rational functions of z. In the frequency-domain this is equivalent to finding a $\mathbf{W}(\omega) \in \mathbb{C}^{J\times N}$ such that:

$$\mathbf{W}(\omega)\mathbf{H}(\omega) = \mathbf{\Pi}\mathbf{D}(\omega) \quad \forall\, \omega \in [0, \pi) \tag{3}$$

where $\mathbf{H}(\omega)$ is the corresponding DTFT for $\mathbf{H}(z)$. Often with frequency-domain methods, the matrix $\mathbf{\Pi}$ is dependent on ω. However, to ensure adequate separation performance, $\mathbf{\Pi}$ must be made independent of frequency. Equation (3) corresponds to the case where the outputs of the un-mixing filter, although separated, are a filtered version of the original sources. To ensure satisfactory audio output quality, $\boldsymbol{D}(\omega)$ should at least be approximately constant with frequency.

C. Main Assumptions

A0*:* $J \geq N \geq 2$; i.e, we have at least as many sensors as sources and number of the sources are at least two.
A1*:* The sources $\mathbf{s}(t)$ are zero mean, second-order quasi-stationary signals and[2] The cross–spectral density matrices of the sources $\mathbf{P}_s(\omega, m)$ are diagonal for all ω and m where ω denotes frequency and m is the epoch index.
A2*:* The mixing system is modelled by a causal system of the form $\mathbf{H}(z) = [\mathbf{h}_1(z), ..., \mathbf{h}_N(z)]$ and does not change over the entire observation interval.
A3*:* $\mathbf{H}(\omega_k)$, the DFT of $\mathbf{H}(z)$, has full column rank for all ω_k, $k = 0, \ldots, K-1$, $\omega_k = \frac{2\pi k}{K}$. Also $||\mathbf{h}_i(\omega_k)||_2^2 = 1$ where $\mathbf{h}_i(\omega_k)$ is the i_{th} column of $\mathbf{H}(\omega_k)$.
A4*:* The noise $\mathbf{n}(t)$ is zero mean, *iid* across sensors, with power σ^2. The noise is assumed independent of the sources.

$A1$ is the core assumption. As is shown in [22], this non-stationarity assumption enables us to separate the convolved sources using only the second-order statistics of the observed signal. The first part of assumption $A3$ guarantees that $\mathbf{H}(\omega_k)$ is invertible for all $k = 0, \ldots, K-1$. Although there is no physical justification for the second part of $A3$, we use it to resolve the inherent scaling ambiguities that exist in our algorithm to identify $\boldsymbol{H}(\omega)$.

Let $\mathbf{P}_x(\omega, m)$ represent the cross-spectral density matrix of the observed signal at frequency ω and time epoch m. Based on the above assumptions we can write:

$$\mathbf{P}_x(\omega, m) = \mathbf{H}(\omega)\mathbf{P}_s(\omega, m)\mathbf{H}^{\dagger}(\omega) + \sigma^2\mathbf{I}, \tag{4}$$

where $\mathbf{P}_s(\omega, m)$ is a diagonal matrix which represents the cross-spectral density matrices of the sources at epoch m. In practice we use a discretized version of ω given as $\omega_k = (2\pi k/K)$ where

[2]By second order quasi-stationarity we mean the variances of the signals are slowly varying with time such that over some short time intervals they can be assumed approximately stationary. Further conditions on the non-stationary condition are given later in the statement of Theorem 1.

K is the total number of frequency samples. For $J > N$, σ^2 can be estimated from the smallest eigenvalue of the matrix $\mathbf{P}_x(\omega, m)$; so for now we consider the following noise free case:

$$\mathbf{P}_x(\omega_k, m) = \mathbf{H}(\omega_k)\mathbf{P}_s(\omega_k, m)\mathbf{H}^\dagger(\omega_k). \tag{5}$$

III. The Joint Diagonalization Problem

The first stage of the proposed algorithm employs joint diagonalization of the set of cross power spectral density matrices $P_x(\omega_k, m), m = 0, \ldots, M-1$ at each frequency ω_k, over M epochs, to estimate the mixing system up to a permutation and diagonal scaling ambiguity at each frequency bin.

The joint diagonalization problem first introduced by Flurry [30] and later on was used as a tool for solving the blind source separation problem by [14] [31], and [32]. The problem is expressed as finding a single matrix $\mathbf{Z}$ that jointly (approximately) diagonalizes the set of matrices $\mathbf{R}_1, \ldots, \mathbf{R}_M$. The most common criterion used for joint diagonalization is the one given as

$$\begin{aligned} &\min_{\mathbf{Z}} \sum_{m=1}^{M} \text{Off}(\mathbf{Z}^\dagger \mathbf{R}_m \mathbf{Z}) \\ &\text{subject to } \mathbf{Z} \in \mathcal{C} \end{aligned} \tag{6}$$

where $\text{Off}(\mathbf{X})$ for an arbitrary matrix $\mathbf{X}$ is defined as the sum squares of the off-diagonal values of the matrix $\mathbf{X}$ and $\mathcal{C}$ is the constraint set, which for example can be the set of square orthogonal matrices or the set of full column rank matrices with unit norm columns. Another common criterion is the following least-squares cost function

$$\min_{\mathbf{A}, \mathbf{\Lambda}(m)} ||\mathbf{R}_m - \mathbf{A}\mathbf{\Lambda}_m\mathbf{A}^\dagger||_F^2 \tag{7}$$

where $\mathbf{\Lambda}_m$ is diagonal for all m. The motivation behind using joint diagonalization as part of the our algorithm can be explained through the following theorem:

Theorem 1: Consider the set of matrices

$$\mathcal{R} = \{\mathbf{R}_m \in \mathbb{C}^{J\times J} | \mathbf{R}_m = \mathbf{A}\mathbf{\Lambda}_m\mathbf{A}^\dagger \;\; m = 0, \ldots, M-1\} \tag{8}$$

where $\mathbf{A} \in \mathbb{C}^{J\times N}$ is some full column rank matrix and $\mathbf{\Lambda}_m \in \mathbb{R}^{N\times N}$ are diagonal matrices such that the set of vectors $\boldsymbol{\lambda}_m = \text{diag}\{\mathbf{\Lambda}_m\}$ span $\mathbb{R}^N$. We also assume that $\mathbf{a}_i$, the i_{th} column of $\mathbf{A}$ has unit norm, i.e.; $||\mathbf{a}_i||_2 = 1$. Now if there is a matrix $\mathbf{B} \in \mathbb{C}^{J\times N}$ and diagonal matrices $\tilde{\mathbf{\Lambda}}_m$ such that

$$\mathbf{R}_m = \mathbf{B}\tilde{\mathbf{\Lambda}}_m\mathbf{B}^\dagger \tag{9}$$

and assuming that $\mathbf{B}$ has unit norm columns then $\mathbf{B}$ must be related to $\mathbf{A}$ as

$$\mathbf{B} = \mathbf{A}\mathbf{\Pi}e^{j\mathbf{D}} \tag{10}$$

where $\mathbf{\Pi}$ is a permutation matrix and $\mathbf{D} \in \Re^{N\times N}$ is a diagonal matrix.

Proof: See Appendix I.

Based on the above Theorem we can easily see that for the set of matrices $\mathbf{P}_x(\omega_k, m) \;\; m = 0, \ldots, M-1$, given by (5), if we find a matrix $\mathbf{B}(\omega_k)$ and diagonal matrices $\mathbf{\Lambda}(\omega_k, m)$ such that

$\mathbf{P}_x(\omega_k, m) = \mathbf{B}(\omega_k)\mathbf{\Lambda}(\omega_k, m)\mathbf{B}^{\dagger}(\omega_k)$ and with the constraint that $||\mathbf{b}_i(\omega_k)||_2^2 = 1$, where $\mathbf{b}_i(\omega_k)$ is the i_{th} column of $\mathbf{B}(\omega_k)$, then we should have

$$\mathbf{B}(\omega_k) = \mathbf{H}(\omega_k)\mathbf{\Pi}(\omega_k)e^{j\mathbf{D}_k} \tag{11}$$

where $\mathbf{D}_k$ are diagonal matrices. In other words, using a joint diagonalization procedure, we can estimate the mixing system $\mathbf{H}(\omega_k)$ up to a frequency dependent permutation $\mathbf{\Pi}(\omega_k)$ and frequency dependent phase ambiguity $e^{j\mathbf{D}_k}$.

In order to attain adequate separation performance, the permutation ambiguity must be resolved. In order to achieve good audio quality at the output, the phase ambiguity problem must be addressed. The permutation issue is addressed in Section V. A procedure to partially address the scaling problem is discussed in Section IV-A.

Having $\mathbf{B}(\omega_k)$ at each frequency bin we can calculate the separating matrix $\mathbf{W}(\omega_k)$ from

$$\mathbf{W}(\omega_k) = \mathbf{B}^{+}(\omega_k) \tag{12}$$

where $\mathbf{B}^{+}(\omega_k)$ is the pseudo inverse of matrix $\mathbf{B}(\omega_k)$.

IV. Algorithm

Based on Theorem 1 we can propose the following least-squares based joint diagonalization criterion for the case when a sample estimate of each $\mathbf{P}_x(\omega_k, m)$ is available:

$$\min_{\mathbf{B}(\omega_k), \mathbf{\Lambda}(m)} \sum_{k=0}^{K-1} \sum_{m=0}^{M-1} ||\hat{\mathbf{P}}_x(\omega_k, m) - \mathbf{B}(\omega_k)\mathbf{\Lambda}(\omega_k, m)\mathbf{B}^{\dagger}(\omega_k)||_F^2, \tag{13}$$

where $\mathbf{B}(\omega)$, which has unit-norm columns, is an estimate of the mixing system $\mathbf{H}(\omega_k)$, $\hat{\mathbf{P}}_x(\omega_k, m)$ is a sample estimate of the observed signal cross spectral density matrix at frequency bin ω_k and time epoch m, and $\mathbf{\Lambda}(\omega_k, m)$ is a diagonal matrix, representing the unknown cross-spectral density matrix of the sources at epoch m. In [1] a similar criterion has been proposed, that uses a backward model which directly estimates the separating matrix $\mathbf{W}(\omega_k)$. Using the criterion in (13) allows us to implement the ALS algorithm as will be described later in this section. In [1], an additional FIR constraint on the un-mixing matrix is required to prevent arbitrary frequency dependent permutations. As shown in [33] and [34] such a constraint is not effective for long reverberant environments and the performance of the algorithm may degrade as the length of the separating filter increases. In the proposed method we do not require an FIR constraint on the mixing model nor on the un-mixing system, mainly because we use a different approach for resolving the permutation problem. This is done following the same approach as in [23], by exploiting the inherent non-stationarity of the input signals in the second stage of the algorithm.

For the first stage of the algorithm we optimize the criterion given by (13) using an alternating least-squares (ALS) approach. The basic idea behind the ALS algorithm is that in the optimization process we divide the parameter space into multiple sets. At each iteration of the algorithm we minimize the criterion with respect to one set conditioned on the previously estimated sets of the parameters. The newly estimated set is then used to update the remaining sets. This process continues until convergence is achieved. Alternating least-squares methods have been used for blind source separation of finite alphabet signals in [35] and [36] and *parallel factor analysis* (PARAFAC) in [37]. The advantage of using ALS (rather than gradient based optimization methods) is that it usually has fast convergence (as will be demonstrated in the simulations) and there are no parameters to adjust. One disadvantage, which is also shared by

gradient techniques for non-linear optimization, is that unless it is properly initialized, it can fall into a local minimum. Later on in this section we introduce a procedure for initializing the algorithm to diminish this possibility.

Referring back to criterion (13), and using the properties of Kronecker products [38], the quantity $\mathbf{B}(\omega_k)\mathbf{\Lambda}(\omega_k, m)\mathbf{B}^\dagger(\omega_k)$ in (13) can be written as

$$\text{vec}\{\mathbf{B}(\omega_k)\mathbf{\Lambda}(\omega_k, m)\mathbf{B}(\omega_k)\} = \mathbf{B}(\omega_k) \odot \mathbf{B}(\omega_k)\, \text{diag}\{\mathbf{\Lambda}(\omega_k, m)\} \tag{14}$$

where $\odot$ is the Khatri-Rao product and is defined as:

$$\mathbf{B}(\omega_k) \odot \mathbf{B}(\omega_k) = [\mathbf{b}_1(\omega_k) \otimes \mathbf{b}_1(\omega_k), \ldots, \mathbf{b}_N(\omega_k) \otimes \mathbf{b}_N(\omega_k)] \tag{15}$$

where $\mathbf{b}_i(\omega_k)$ is the i_{th} column of $\mathbf{B}(\omega_k)$ and $\otimes$ represents the Kronecker product. Setting $\mathbf{G}(\omega_k) = \mathbf{B}(\omega_k) \odot \mathbf{B}(\omega_k)$, $\mathbf{d}(\omega_k, m) = \text{diag}\{\mathbf{\Lambda}(\omega_k, m))\}$ and $\hat{\mathbf{p}}_x(\omega_k, m) = \text{vec}\{\hat{\mathbf{P}}_x(\omega_k, m)\}$ we can rewrite (13) as:

$$\min_{\mathbf{g}_i(\omega_k)\in\Omega, \mathbf{d}(\omega_k, m)} \sum_{k=0}^{K-1} \sum_{m=0}^{M-1} ||\hat{\mathbf{p}}_x(\omega_k, m) - \mathbf{G}(\omega_k)\mathbf{d}(\omega_k, m)||_2^2 \tag{16}$$

where $\mathbf{g}_i(\omega_k)$ is the i_{th} column of $\mathbf{G}(\omega_k)$. Based on assumption $A3$ we have $||\mathbf{b}_i(\omega_k)||_2^2 = 1$ and we define the constraint set $\Omega \subset \mathbb{C}^{J^2\times 1}$ as:

$$\Omega = \{\text{vec}\{\mathbf{\Phi}\} | \mathbf{\Phi} = \boldsymbol{\nu}\boldsymbol{\nu}^\dagger,\ \boldsymbol{\nu} \in \mathbb{C}^{J\times 1},\ ||\boldsymbol{\nu}||_2^2 = 1\}. \tag{17}$$

In defining the constraint set Ω we used the fact that for a column vector $\boldsymbol{\nu}$ we have $\boldsymbol{\nu} \otimes \boldsymbol{\nu} = \text{vec}\{\boldsymbol{\nu}\boldsymbol{\nu}^\dagger\}$. Following the discussion above, we can first minimize (16) with respect to $\mathbf{g}_i(\omega_k)$ conditioned on $\hat{\mathbf{d}}(\omega_k, m)$, the previously estimated values of $\mathbf{d}(\omega_k, m)$. To do this we form the matrices $\mathbf{T}(\omega_k) = [\hat{\mathbf{p}}(\omega_k, 0), ..., \hat{\mathbf{p}}(\omega_k, M-1)]$ and $\mathbf{F}(\omega_k) = [\hat{\mathbf{d}}(\omega_k, 0), ..., \hat{\mathbf{d}}(\omega_k, M-1)]$ and we write equation (16) as:

$$\min_{\mathbf{g}_i(\omega_k)\in\Omega} \sum_{k=0}^{K-1} ||\mathbf{T}(\omega_k) - \mathbf{G}(\omega_k)\mathbf{F}(\omega_k)||_F^2. \tag{18}$$

Notice that (18) is a constrained least-squares problem. One simple, although approximate, way to minimize (18) is to first find the unconstrained least-squares minimizer of (18) by setting

$$\tilde{\mathbf{G}}(\omega_k) = \mathbf{T}(\omega_k)\mathbf{F}^+(\omega_k). \tag{19}$$

We then project each column of $\tilde{\mathbf{G}}(\omega_k)$ onto Ω; i.e.,

$$\hat{\mathbf{g}}_i(\omega_k) = \text{proj}_{\,\Omega}[\tilde{\mathbf{g}}_i(\omega_k)] \tag{20}$$

where $\tilde{\mathbf{g}}_i(\omega_k)$ is the i_{th} column of $\tilde{\mathbf{G}}(\omega_k)$. We now discuss a convenient method of performing the projection operation. The projection operation can be effected by the following minimization:

$$\hat{\mathbf{g}}_i(\omega_k) = \arg \min_{\mathbf{g}_i(\omega_k)\in\Omega} ||\tilde{\mathbf{g}}_i(\omega_k) - \mathbf{g}_i(\omega_k)||_2^2. \tag{21}$$

Since $\mathbf{g}_i(\omega_k) = \text{vec}\{\mathbf{b}_i(\omega_k)\mathbf{b}_i^\dagger(\omega_k)\}$, by defining $\mathbf{Y}_i(\omega_k) = \text{mat}\{\tilde{\mathbf{g}}_i(\omega_k)\}$ we can write the above equation as:

$$\begin{aligned}
&\min_{||\mathbf{b}_i(\omega_k)||_2=1} ||\mathbf{Y}_i(\omega_k) - \mathbf{b}_i(\omega_k)\mathbf{b}_i^\dagger(\omega_k)||_F^2 = \\
&\min_{||\mathbf{b}_i(\omega_k)||_2=1} (\mathbf{b}_i^\dagger(\omega_k)\mathbf{b}_i(\omega_k))^2 + \text{Trace}\{\mathbf{Y}_i^\dagger(\omega_k)\mathbf{Y}_i(\omega_k)\} - 2\mathbf{b}_i^\dagger(\omega_k)\mathbf{Y}_i(\omega_k)\mathbf{b}_i(\omega_k) = \\
&\min_{||\mathbf{b}_i(\omega_k)||_2=1} C - 2\mathbf{b}_i^\dagger(\omega_k)\mathbf{Y}_i(\omega_k)\mathbf{b}_i(\omega_k)
\end{aligned} \tag{22}$$

where $C = 1 + \text{Trace}\{\mathbf{Y}_i^{\dagger}(\omega_k)\mathbf{Y}_i(\omega_k)\}$ is a constant term. The above minimization can be done easily by choosing $\hat{\mathbf{b}}_i(\omega_k)$, the estimated i_{th} column of $\mathbf{B}(\omega_k)$, to be the dominant eigenvector of $\mathbf{Y}_i(\omega_k)$. To find the dominant eigenvector of a matrix we can use the power iteration method described in [39]. Since an initial estimate of $\mathbf{b}_i(\omega_k)$ is available (as is explained later), $\mathbf{Y}_i(\omega_k)$ is nearly a rank one matrix. Hence, the ratio of the largest eigenvalue of $\mathbf{Y}_i(\omega_k)$ to the second-largest (this ratio determines the convergence of the power method), is large. Hence, we need to apply only few iterations of the power method to minimize (22) [3].

To minimize (13) with respect to $\mathbf{d}(\omega_k, m)$ conditioned on the previous estimate of $\mathbf{G}(\omega_k)$ we solve the following least-squares problem.

$$\hat{\mathbf{d}}(\omega_k, m) = \arg \min_{\mathbf{d}(\omega_k, m)} ||\hat{\mathbf{p}}_x(\omega_k, m) - \hat{\mathbf{G}}(\omega_k)\mathbf{d}(\omega_k, m)||_2^2 \tag{23}$$

Minimizing (23) with respect to $\mathbf{d}(\omega_k, m)$ we get:

$$\hat{\mathbf{d}}(\omega_k, m) = \hat{\mathbf{G}}^{+}(\omega_k)\hat{\mathbf{p}}(\omega_k, m) \quad m = 0, ..., M-1, \;\; k = 0, \ldots, K-1. \tag{24}$$

Using equation (19), (20) and (24) we can repeatedly update the values of $\mathbf{d}(m)$ and $\mathbf{G}(m)$ until convergence is achieved.

As mentioned previously, to avoid being trapped in local minima, and to ensure perceptually acceptable voice quality, we need to properly initialize the ALS algorithm. As outlined in the following discussion, we note that the initialization procedure has a significant impact on resolving the scaling ambiguity problem.

A. Initialization

To initialize the algorithm we have different options. The first option is that a rough estimate of the mixing system at each frequency bin (up to some scaling and permutation ambiguity) can be obtained using the following closed-form, exact joint diagonalization procedure [26].

Select two matrices $\hat{\mathbf{P}}(\omega_k, m_1)$, $\hat{\mathbf{P}}(\omega_k, m_2)$, $m_1 \neq m_2$. Then choose the initial estimate for $\mathbf{B}(\omega_k)$ to be the matrix consisting of the N dominant generalized eigenvectors of the matrix couple $(\hat{\mathbf{P}}(\omega_k, m_1), \hat{\mathbf{P}}(\omega_k, m_2))$. Although no optimum selection for m_1 and m_2 can be given at this stage, in our simulations we choose $\hat{\mathbf{P}}(\omega_k, m_1)$ and $\hat{\mathbf{P}}(\omega_k, m_2)$ such that their non-zero generalized eigenvalues are not all repeated. Since we need to use this initialization procedure at each frequency bin, one draw-back of above initialization method will be its computational complexity.

An alternative, novel, *ad hoc* initialization method which not only requires less computation, but also dramatically improves the quality of the separated audio signals, is described as follows. The main idea of this initialization procedure, is that first we choose the initial value of $\mathbf{B}(\omega_0)$, the first frequency bin, using the exact closed form joint diagonalization method described above. We then apply the ALS algorithm to find the final estimate of $\mathbf{B}(\omega_0)$. This final estimate is then used as an initial value for the next adjacent frequency bin, which is $\mathbf{B}(\omega_1)$. The outcome of this frequency bin is also used as an initial value for the next frequency bin and this procedure continues until all the frequency bins have been covered. Note that in this way we need to apply the exact closed form joint diagonalization algorithm only for one frequency bin.

As has been demonstrated in our simulation results, this initialization procedure significantly improves the quality of the separated audio signals. An intuitive explanation for this is as

[3]In our simulations we use only one power iteration per ALS iteration. Increasing the number of iterations beyond one did not noticeably improve the convergence nor the performance of the algorithm.

follows. We realize that the estimate of $\boldsymbol{B}$ is not unique, because each column $\boldsymbol{b}_i$ is subject to a multiplicative phase ambiguity of the form $e^{j\mathbf{D}_k}$, even though the condition $||\ \boldsymbol{b}_i\ ||_2 = 1$ in the solution of (13) has been enforced[4]. Fast variation of this phase ambiguity in frequency can cause the resulting time–domain estimate $\hat{\boldsymbol{H}}(t)$ of the channel to be excessively long. By initializing the algorithm in the manner proposed, this phase ambiguity varies smoothly with frequency, therefore creating an $\hat{\boldsymbol{H}}(t)$ which can be of moderate length. As is shown in our simulations, the resulting overall system (channel + inverse) is then much more localized in time. This property is known to minimize degradation in audio quality due to reverberative effects.

Summary of Stage I of the Algorithm for Blind Identification

1. Estimate the normalized observed signals' cross spectral density matrices, $\hat{\mathbf{P}}_x(\omega_k, m)$, based on the method given in Appendix B and set $\mathbf{T}(\omega_k) = [\hat{\mathbf{p}}_x(\omega_k, 0), ..., \hat{\mathbf{p}}_x(\omega_k, M-1)]$ where $\hat{\mathbf{p}}_x(\omega_k) = \text{vec}\{\hat{\mathbf{P}}_x(\omega_k, m)\}$
2. Compute $\hat{\mathbf{B}}^0(\omega_k)$[5], the initial value for $\mathbf{B}(\omega_k)$, according to the method described in Section IV-A.

3. for $k = 0$ to $K-1$
- if $k \neq 0$ set $\hat{\mathbf{B}}^0(\omega_k) = \hat{\boldsymbol{H}}(\omega_{k-1})$, where $\hat{\boldsymbol{H}}(\omega_{k-1})$ is the final estimate from the previous bin.
- for $\nu = 0$ to Max_itr
  - Set $\hat{\mathbf{G}}^\nu(\omega_k) = [\hat{\mathbf{b}}_1^\nu(\omega_k) \otimes \hat{\mathbf{b}}_1^\nu(\omega_k), \ldots, \hat{\mathbf{b}}_N^\nu(\omega_k) \otimes \hat{\mathbf{b}}_N^\nu(\omega_k)]$
  - Calculate $\hat{\mathbf{d}}^\nu(\omega_k, m) = (\hat{\mathbf{G}}^\nu(\omega_k))^+ \hat{\mathbf{p}}(\omega_k, m) \quad m = 0, ..., M-1$
  - Set $\mathbf{F}^\nu(\omega_k) = [\hat{\mathbf{d}}^\nu(\omega_k, 0), ..., \hat{\mathbf{d}}^\nu(\omega_k, M-1)]$
  - Calculate $\tilde{\mathbf{G}}^\nu(\omega_k) = \mathbf{T}(\omega_k)(\mathbf{F}^\nu(\omega_k))^+$
  - for $i = 1$ to N
    - $\mathbf{Y} = \text{mat}\{\tilde{\mathbf{g}}_i^\nu(\omega_k)\}$
    - $\mathbf{q} = \mathbf{Y}\mathbf{b}_i^\nu(\omega_k)$
    - $\hat{\mathbf{b}}_i^{\nu+1}(\omega_k) = \dfrac{\mathbf{q}}{||\mathbf{q}||_2}$
  - end
  - Calculate the cost value $C_k^\nu = ||\mathbf{T}(\omega_k) - \hat{\mathbf{G}}^\nu(\omega_k)\mathbf{F}^\nu(\omega_k)||_F^2$
  - if $\dfrac{|C_k^\nu - C_k^{\nu-1}|}{C_k^\nu} < \epsilon$, where $0 < \epsilon \ll 1$, then stop, go to the next frequency bin
  - end
- end

It is worthy of note, as a passing comment, that in the case where $J > N$, the number of sources N can be estimated from the number of elements of $\boldsymbol{\Lambda}$ which are significantly different from zero.

[4]Note that placing a constraint on both the norm and the phase of the columns $\boldsymbol{b}_i$ would lead to a less computationally efficient algorithm.

[5]In this summary, the superscript refers to the iteration index.

V. Resolving Permutations

One potential problem with the cost function in (13) is that it is insensitive to permutations of the columns of $\mathbf{B}(\omega_k)$. More specifically if $\mathbf{B}_{opt}(\omega_k)$ is an optimum solution to (13) then $\mathbf{B}_{opt}(\omega_k)\mathbf{\Pi}_k$, where $\mathbf{\Pi}_k$ is an arbitrary permutation matrix for each ω_k, will also be a optimum solution. Since in general $\mathbf{\Pi}_k$ can vary with frequency, poor overall separation performance will result.

In this section we suggest a novel solution for solving the permutation problem which exploits the cross-frequency correlation between diagonal values of $\mathbf{\Lambda}(\omega_k, m)$ and $\mathbf{\Lambda}(\omega_{k+1}, m)$ given in (13). Notice that $\mathbf{\Lambda}(\omega_k, m)$ can be considered an estimate of the sources' cross-power spectral density at epoch m. When the sources are speech signals, the temporal trajectories of the power spectral density of speech, known as spectrum modulation of speech, are correlated across the frequency spectrum. Using this correlation we can correct wrong permutations as shown in the following example for the two source case. This same principle was also used in [40]; however, it has only been applied to the two sources case. As we see later, the proposed method for resolving permutations can be extended to an arbitrary number of sources.

Assume that $\mathbf{\Lambda}(\omega_k, m), m = 0, \ldots, M-1$, represents the estimated cross-spectral density of the two sources at frequency bin ω_k. We want to adjust the permutation at frequency ω_j such that it has the same permutation as frequency bin ω_k. To do so we first calculate the cross frequency correlation between the diagonal elements of $\mathbf{\Lambda}(\omega_j, m)$ and $\mathbf{\Lambda}(\omega_k, m)$ using following measure:

$$\rho_{qp}(\omega_k, \omega_j) = \frac{\sum_{m=0}^{M-1} \lambda_q(\omega_k, m)\lambda_p(\omega_j, m)}{\sqrt{\sum_{m=0}^{M-1} \lambda_q^2(\omega_k, m)}\sqrt{\sum_{m=0}^{M-1} \lambda_p^2(\omega_j, m)}} \tag{25}$$

where $\rho_{qp}(\omega_k, \omega_j)$ represents the cross frequency correlation between$\lambda_q(\omega_k, m)$, the q_{th} diagonal element of $\mathbf{\Lambda}(\omega_k, m)$, and $\lambda_p(\omega_j, m)$, the p_{th} diagonal element of $\mathbf{\Lambda}(\omega_j, m)$. To determine whether frequency bins ω_k and ω_j have the same permutation, we perform the following hypothesis test [41]. Let H_0 be the hypothesis that frequency bins ω_k and ω_j have the same permutation, and H_1 be the hypothesis to the contrary. Then, we decide between H_0 and H_1 according to

$$r = \frac{\rho_{11}(\omega_k, \omega_j) + \rho_{22}(\omega_k, \omega_j)}{\rho_{12}(\omega_k, \omega_j) + \rho_{21}(\omega_k, \omega_j)} \underset{H_1}{\overset{H_0}{\gtrless}} T, \tag{26}$$

where T is a threshold which is dependent on the relative weight between a type I and a type II error [41], and the probability in the tails of the probability distributions of the statistics in the numerator and denominator of (26). Since the probability distributions of these statistics are difficult to determine, we empirically set $T = 1$. If we declare H_1, the permutation at one of the frequency bins ω_k or ω_j is changed. We apply the above hypothesis test to all frequency bins to detect and correct wrong permutations.

In general when number of sources is greater than two, the permutation matrix can be estimated by solving the following discrete optimization problem.

$$\max_{\mathbf{\Pi}_k \in \mathcal{P}} \quad \text{trace}\big(\mathbf{\Pi}_k\ \mathbf{E}(\omega_k)\mathbf{E}^T(\omega_j)\big) \tag{27}$$

where $\mathcal{P}$ is the set of $N \times N$ permutation matrices including the identity matrix, and $\mathbf{E}(\omega_k)$ is

an $N \times M$ matrix given as

$$\mathbf{E}(\omega_k) = \left(\sum_{m=0}^{M-1} \mathbf{\Lambda}^2(\omega_k, m) \right)^{-\frac{1}{2}} \begin{pmatrix} \lambda_1(\omega_k, 0) & \dots & \lambda_1(\omega_k, M-1) \\ \vdots & \ddots & \vdots \\ \lambda_N(\omega_k, 0) & \dots & \lambda_N(\omega_k, M-1) \end{pmatrix}. \tag{28}$$

The discrete optimization criterion in (27) can be solved by enumerating over all possible selections for $\mathbf{\Pi}_k$. This means for a set of $N \times N$ permutation matrices the criterion in (27) must be calculated $N!$ times to find the optimum solution for $\mathbf{\Pi}_k$. For large values of N this may not be computationally efficient. A more computationally efficient but suboptimal approach to estimate the permutation matrix between the two frequency bins is given by the following algorithm.

Adjusting Permutations

1. initialize the $N \times N$ matrix $\mathbf{\Pi}_k$ to an all zeros matrix.
2. For $i = j, k$ set up matrices

$$\mathbf{E}(\omega_i) = \left(\sum_{m=0}^{M-1} \mathbf{\Lambda}^2(\omega_i, m) \right)^{-\frac{1}{2}} \begin{pmatrix} \lambda_1(\omega_i, 0) & \dots & \lambda_1(\omega_i, M-1) \\ \vdots & \ddots & \vdots \\ \lambda_N(\omega_i, 0) & \dots & \lambda_N(\omega_i, M-1) \end{pmatrix} \tag{29}$$

3. Form the multiplication $\mathbf{T}_{kj} = \mathbf{E}(\omega_k)\mathbf{E}^T(\omega_j)$
4. Find the row number r_{max} and column number c_{max} corresponding to the element of $\mathbf{T}$ with largest absolute value. Zero all elements of $\mathbf{T}$ corresponding to this row and column numbers and set $\mathbf{\Pi}_k(c_{max}, r_{max}) = 1$.
5. Recursively repeat the previous step for the remaining elements of matrix $\mathbf{T}_{kj}$ until only one non-zero element remains. Set

$$\mathbf{\Pi}_k(c_f, r_f) = 1 \tag{30}$$

where r_f and c_f are the corresponding row and column numbers of the remaining non-zero element.

Notice that the above algorithm calculates the permutation matrix $\mathbf{\Pi}_k$ between two frequency bins ω_k and ω_j. To obtain a uniform permutation across the whole frequency spectrum, we need to apply the above algorithm repeatedly to all pairs of frequency bins. One way of doing this is to adjust the permutation between adjacent frequency bins in a sequential order where, for example, starting from frequency bin ω_0 we adjust the permutation of each bin relative to it's previous bin. This approach, although simple, has a major drawback as explained as follows. Consider the situation where an error is made in estimating the correct permutation matrix for frequency bin ω_k. In this case, all frequency bins placed after ω_k will receive a different permutation than the ones placed before ω_k. In the worst case scenario we will have half of the frequency bins with one permutation and the other half with a different permutation, which will result in very poor or no separation. To prevent such a catastrophic situation we propose following hierarchical sorting scheme to sort the permutations across all frequency bins. For clarity we explain the algorithm

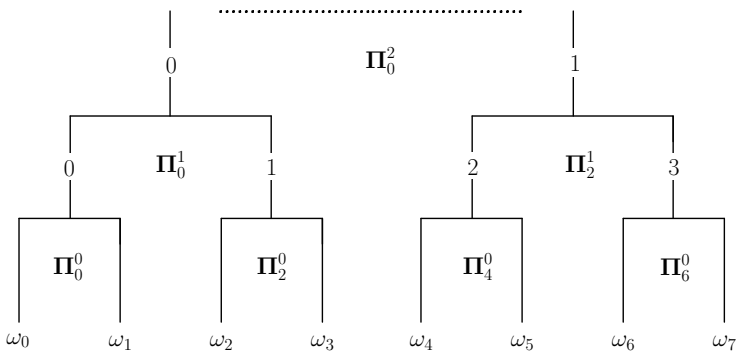

Fig. 1. An example of the diadic permutation sorting algorithm for the case when the total number of frequency bins is eight.

for the case when we have only eight frequency bins (Figure 1). Extension to general case for an arbitrary number of frequency bins then can be easily deduced.

Diadic Sorting Algorithm

1. Divide the frequency bins into groups of two bins[6] each with group index p.
2. Let k and $k+1$ be the indices to the frequency bins inside the group p and $\mathbf{\Pi}_k^0$ be the permutation matrix estimated from the criterion given in (27). Also let $\mathbf{\Sigma}_k^0(m) = \mathbf{\Lambda}(\omega_k, m)$ and $\mathbf{\Sigma}_{k+1}^0(m) = \mathbf{\Lambda}(\omega_{k+1}, m)$. Then for all $p = 0, \ldots, 3$, update the order of diagonal values of $\mathbf{\Sigma}_k^0(m)$ using

$$\mathbf{\Sigma}_k^0(m) = \mathbf{\Pi}_k^0 \mathbf{\Sigma}_k^0(m) \mathbf{\Pi}_k^{0^T} \quad k = 0, 2, 4, 6$$

3. Update the order of columns of $\mathbf{B}(\omega_k)$ using

$$\mathbf{B}(\omega_k) = \mathbf{\Pi}_k^0 \mathbf{B}(\omega_k) \quad k = 0, 2, 4, 6$$

4. For each group calculate

$$\mathbf{\Sigma}_p^1(m) = \mathbf{\Sigma}_k^0(m) + \mathbf{\Sigma}_{k+1}^0(m) \quad k = 0, 2, 4, 6, \quad p = 0, 1, 2, 3 \tag{31}$$

5. Divide the set of $\mathbf{\Sigma}_0^1(m), \ldots, \mathbf{\Sigma}_3^1(m)$ into groups of two elements and for each of the new groups estimate the permutation matrices $\mathbf{\Pi}_p^1 \quad p = 0, 2$ using the diagonal values of $\mathbf{\Sigma}_p^1(m)$ based on the criterion given in (27). Also for all $p = 0, 2$, update the order of the diagonal values of $\mathbf{\Sigma}_p^1(m)$ using

$$\mathbf{\Sigma}_p^1(m) = \mathbf{\Pi}_p^1 \mathbf{\Sigma}_p^1(m) \mathbf{\Pi}_p^{1^T} \quad p = 0, 2$$

6. Update the order of columns of $\mathbf{B}(\omega_k)$ using

$$\mathbf{B}(\omega_{2p}) = \mathbf{\Pi}_p^1 \mathbf{B}(\omega_{2p})$$
$$\mathbf{B}(\omega_{2p+1}) = \mathbf{\Pi}_p^1 \mathbf{B}(\omega_{2p+1}) \quad p = 0, 2$$

[6]Here we assume that K, the total number of the frequency bins, is a multiple integer of 2.

7. For the new groups calculate

$$\mathbf{\Sigma}_q^2(m) = \mathbf{\Sigma}_p^1(m) + \mathbf{\Sigma}_{p+1}^1(m) \quad p = 0, 2 \quad q = 0, 1 \tag{32}$$

8. Finally calculate $\mathbf{\Pi}_0^2$, by substituting $\mathbf{\Sigma}_0^2(m)$ and $\mathbf{\Sigma}_1^2(m)$ in (27). Update the columns of $\mathbf{B}(\omega_k) \quad k = 0, \ldots, 3$ using

$$\mathbf{B}(\omega_k) = \mathbf{\Pi}_0^2 \mathbf{B}(\omega_k) \quad k = 0, \ldots, 3$$

The success of the proposed diadic sorting algorithm can be explained for the $N = 2$ case as follows. As the process moves up the hierarchy, the statistics comprising the numerator and denominator of (26) accumulate more data samples, as is made evident by (31) and (32). Thus, the probability distribution governing these statistics decreases in variance, with the result that the probability of error in (26) diminishes as the hierarchy is ascended. This argument is easily extended to the $N > 2$ case. Thus, we can say that the dominant error mechanism is at the lower levels of the hierarchy, where the impact of a permutation error is not catastrophic.

VI. Simulation Results

A. Example I, FIR Convolutive Mixing

The objective of this first simulation is to characterize the performance of the algorithm under a controlled mixing environment. For this purpose we use an 8-tap FIR convolutive mixing system where the impulse responses of its elements are selected randomly from a uniform distribution with zero mean and unit variance. For the sources, we use two independent white Gaussian processes, multiplied by slowly varying sine and cosine signals respectively to create the required non–stationary effect. In this example, the epoch size was kept at 500 and the total data length was varied between 10000 and 50000 samples, corresponding to M, the number of epochs, ranging between 20 to 100 epochs. White Gaussian noise was added to the output of the system at a level corresponding to the desired value of averaged SNR over all epochs[7]. At each epoch, 128–point FFTs, applied to time segments overlapping by 50%, weighted by Hanning windows were used to estimate the cross-spectral density matrices. Further details regarding the construction of the cross-spectral density matrices is given in Appendix II. For the joint diagonalization algorithm, for all frequency bins for this specific case only, we chose the initial estimate of the channel to be an identity matrix[8].

Let $\mathbf{C}(\omega_k)$ represent the global system frequency response

$$\mathbf{C}(\omega_k) = \mathbf{W}(\omega_k)\mathbf{H}(\omega_k) \tag{33}$$

whose ij_{th} element is $c_{ij}(\omega_k)$. To measure the separation performance we can use following formula:

$$\mathrm{SIR}(i) = \frac{\sum_{q=0}^{M_c-1} \max_j(S_{ij}^q)}{\sum_{q=0}^{M_c-1} \left\{ \sum_{j=1}^{N} S_{ij}^q - \max_j(S_{ij}^q) \right\}} \tag{34}$$

[7]The power of the noise was kept constant at all epochs.

[8]The initialization method described in previous section can of course be used in this example. Nevertheless the motivation for using an identity matrix for initialization in this specific example is to demonstrate the desirable convergence properties of the algorithm, even if we do not know a good initialization point.

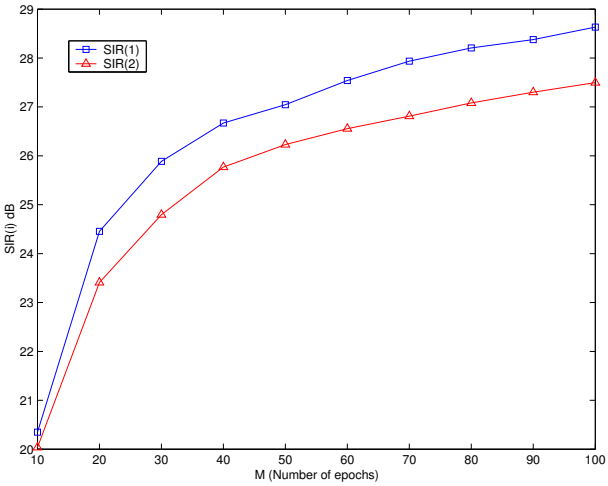

Fig. 2. Example I: SIR versus M, the number of epochs, for SNR=20dB, using $M_c = 50$ Monte Carlo runs.

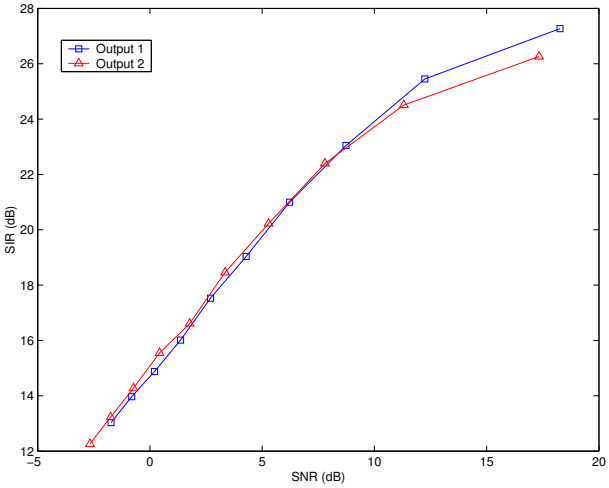

Fig. 3. Example I: SIR versus SNR, for M=50, using $M_c = 50$ Monte Carlo runs.

where $S_{ij}^q = \sum_{k=0}^{K-1} |c_{ij}^q(\omega_k)|^2$, q is an index to the q_{th} Monte Carlo run and the quantity M_c is the total number of Monte Carlo runs. Notice that (34) implicitly measures the signal to interference ratio (SIR) for each output of the separating system. Here, at each output, the signal is the separated source that has the maximum power and the interference is the contribution from the other sources. Figure 2 shows the variation of each outputs' SIR with M, the number of epochs for a fixed signal to noise ratio (SNR=20dB). As can be seen from this figure, by increasing the number of epochs, which corresponds to increasing the data length, the output SIR improves (increases). Also Figure 3 shows how the separation performance changes with observed signals' signal to noise ratio for a fixed number of epochs, $M = 50$.

To demonstrate the effectiveness of the proposed permutation algorithm, Table I shows the separation performance before and after the permutation ambiguities have been resolved. Also refer to Figures 4 and 5 for a graphical visualization of the effects of arbitrary permutations at different frequency bins, and the improvement resulting from the proposed permutation algo-

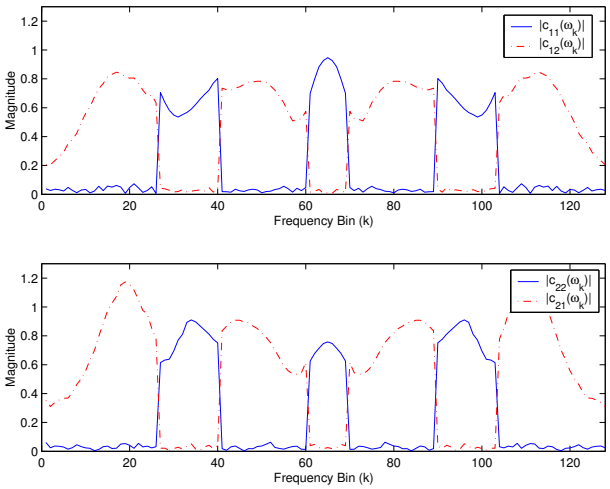

Fig. 4. Example 1: Effects of permutation ambiguities in the frequency-domain before applying the permutation algorithm (M=50, SNR=20dB). The quantity $\mathbf{c}_{ij}(\omega_k)$ is the ij_{th} element of the global system $\mathbf{C}(\omega_k) = \mathbf{W}(\omega_k)\mathbf{H}(\omega_k)$.

rithm. As can be seen from the table and also the figures, the frequency dependent permutation ambiguity can severely degrade the overall separation performance. Nevertheless the proposed algorithm is able to significantly improve the results by resolving these permutation ambiguities.

TABLE I

EXAMPLE I, OUTPUT SIR BEFORE AND AFTER APPLYING THE PERMUTATION ALGORITHM. SNR=$20dB$, M=50 AND $M_c = 50$ MONTE CARLO RUNS.

Output SIR (dB)	SIR(1)	SIR(2)
Before applying the permutation algorithm	3.5 dB	4.2 dB
After applying the permutation algorithm	27 dB	26 dB

VII. REAL ROOM EXPERIMENTS

In this section we present the results of applying our algorithm to blind source septation of speech signals in a real reverberant environment. All recordings were done using an 8.0 Khz, 16 bits sampling format.

A. Real Room Experiment I

The first set of experiments were conducted in an office room using $N = 2$ loudspeakers for the sources and $J = 4$ omnidirectional microphones for the sensors. The configuration of the experiment, showing all distances etc., is shown in Figure 6. The two sources were created by catenating independent multiple speech segments from the TIMIT speech database. The speech signals then were played simultaneously through the two speakers with approximately the same sound volume.

To quantify the separation performance, we measure the signal–to–interference ratio at each microphone (i.e., the input to the algorithm) and at each output of the algorithm. To achieve

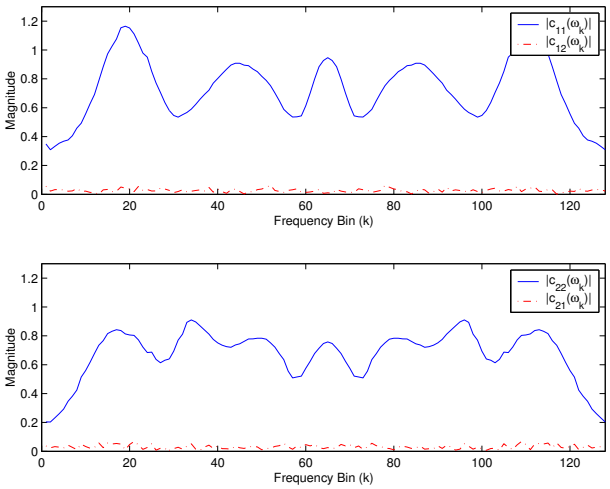

Fig. 5. Example 1: Results after applying the permutation algorithm (M=50, SNR=20dB)

.

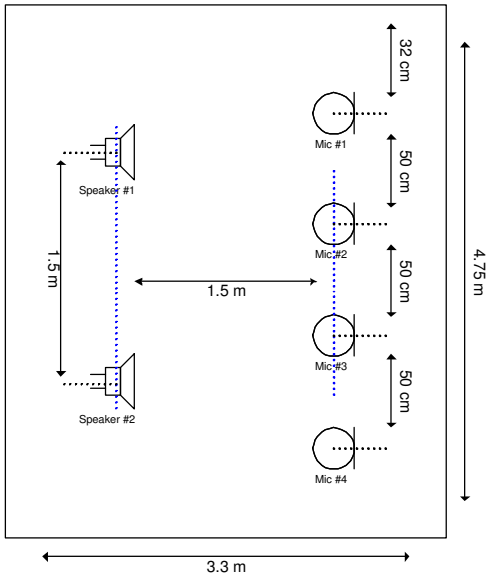

Fig. 6. Real Room Experiments, Recording setup in an office room.

this, using the same setup, white noise signals were played through each speaker one at a time (i.e., only one source was active at each time). Let $\hat{\sigma}_x^2(x_i, s_j) = \sum_{t=0}^{T-1} x_i^2(t)$ represent the power of the recorded signal at the i_{th} microphone when only speaker j is active and all other speakers (sources) are inactive. The signal–to–interference ratio of the recorded signal at the i_{th} microphone can be estimated using

$$SIR_x(i) = \frac{\max_j \hat{\sigma}_x^2(x_i, s_j)}{\sum_{j=1}^{N} \hat{\sigma}_x^2(x_i, s_j) - \max_j \hat{\sigma}_x^2(x_i, s_j)}. \tag{35}$$

Using above formula, we can also measure $\text{SIR}_y(i)$, the SIR of the ith output of the separating network, by substituting $\hat{\sigma}_x^2(x_i, s_j)$ with $\hat{\sigma}_y^2(y_i, s_j)$, the power of the signal at the i_{th} output when only source j is active.

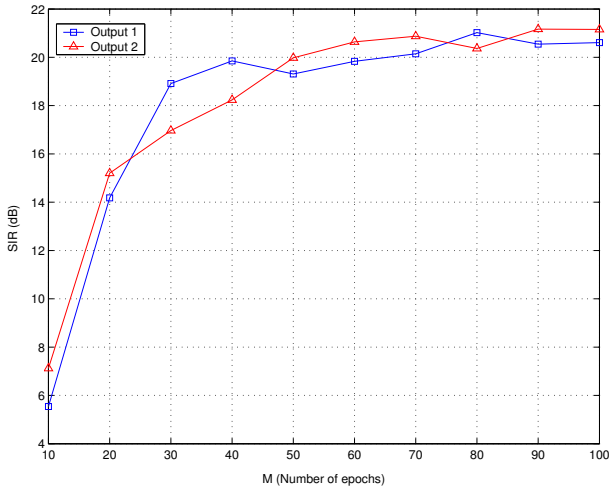

Fig. 7. Results of separation performance for recordings in an office room: SIR$_y i$ versus M, the number of epochs, for $K = 4096$, $i = 1, 2$.

To perform the separation, in a manner similar to examples I and II, we first divided the recorded signal into multiple time segments (epochs), where we chose the size of each epoch to be 10000 data samples long. For each epoch we calculated the cross spectral density matrices according to Appendix II, with the number of FFT-points K equal to 4096, and with the overlap-percentage equal to 80%. Figure 7 shows the output SIR$_y(i)$ versus M, the number of epochs, for $i = 1, 2$. As can be seen for $M = 100$, an average SIR of more than 20dB is reached for each output. As a reference, Table II shows the input SIRs corresponding to the recorded signals. By comparing the SIRs before and after applying the separating algorithm, it can be seen that the output SIRs have been improved by roughly 19 to 20dB. In this experiment, on average, at each frequency bin the ALS algorithm converged within 30 to 40 iterations.

As discussed in Section II, we can recover the sources only up to a frequency dependent scaling ambiguity using the proposed joint diagonalization algorithm. The effect of this unknown scaling ambiguity can significantly deteriorate the quality of the separated audio signals. However, our listening tests reveal that the sequential initialization procedure discussed in Section IV-A dramatically improves the quality of the separated speech signals in these experiments[9]. To demonstrate the effect of the initialization procedure on the impulse response of the separating network, we calculate the impulse response of the separating matrix with and without using the proposed initialization method. Figure 8 shows the results for one of the impulse responses of the separating network. As can be seen from the figure, the tail of the impulse response of the separating matrix obtained by using the new initialization procedure is more suppressed

[9]To hear the recordings and also the separation results, please refer to the following website: "www.ece.mcmaster.ca/~reilly/kamram/index.htm".

TABLE II
INPUT SIRS FOR THE RECORDED SIGNALS IN AN OFFICE ENVIRONMENT

	MIC 1	MIC 2	MIC 3	MIC 4
SIR	2.6945 dB	1.2282 dB	0.3266 dB	1.4031 dB

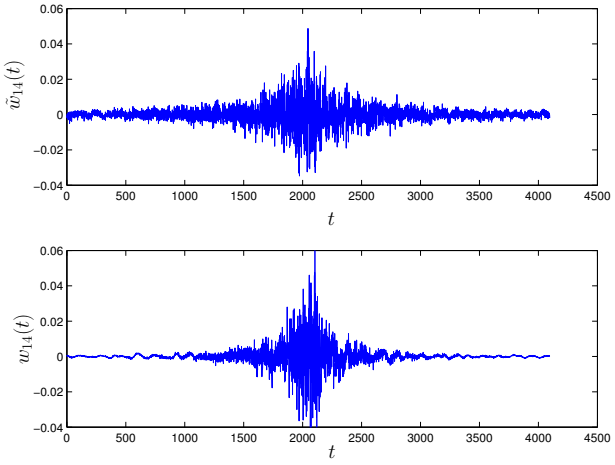

Fig. 8. The impulse response of the separating matrix before ($\tilde{w}_{14}(t)$) and after ($w_{14}(t)$) the sequential initialization procedure has been applied.

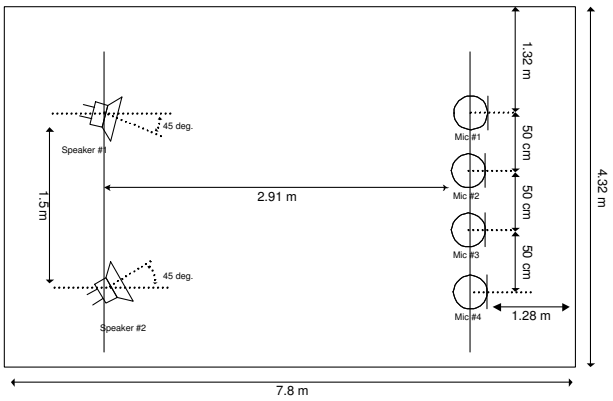

Fig. 9. Real room experiments: Recording setup in a conference room.

compared to the tail of the impulse response of the separating matrix calculated without using the sequential initialization procedure. Our hearing experiments show that in fact the tails do contribute to the distortion of the separated audio signal and the more suppressed they are, the better is the perceptual quality of the separated signals.

B. Real Room Experiment II

In the next set of experiments, we perform real recordings in a highly reverberant conference room. The recording setup is similar to the previous experiment, except that the room dimensions and the distance between the microphones and speakers are increased for this experiment (Figure 9). The reverberation characteristics of the two rooms used in these experiments are shown in Figure 10, which shows the reverberation time[10] of the room vs. frequency. As can be seen from the figure, on average the reverberation time of the conference room, used in this experiment, is much higher than the reverberation time of the office room, used in the previous experiment. Due to the long reverberation time of the conference room, we expect that in this case more frequency

[10]The reverberation time in a room at a given frequency is the time required for the mean-square sound pressure in that room to decay from a steady state value by 60dB after the sound suddenly ceases(see also [42]).

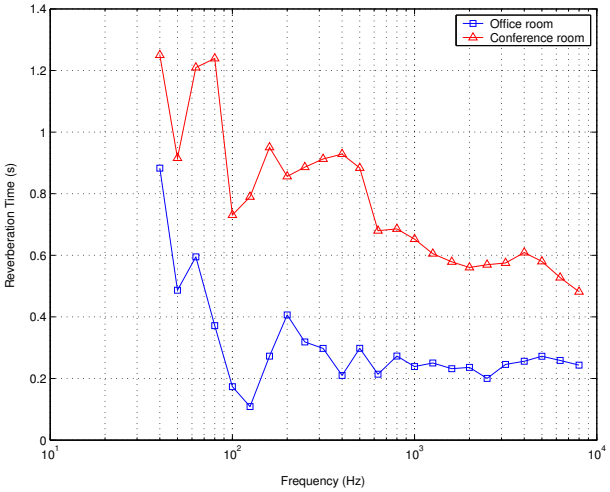

Fig. 10. Reverberation times of the rooms used in experiments I and II.

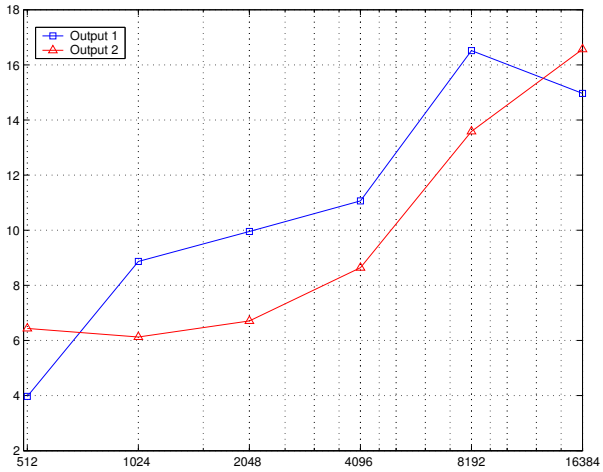

Fig. 11. Separation performance versus number of frequency bins (K) for recordings in a conference room, $M = 140$.

bins are needed to estimate the cross spectral density matrices of the recorded signals. In this experiment, the epoch size was kept the same as in the previous experiment (10,000 samples). Figure 11 shows how the performance of the algorithm improves by increasing the number of frequency bins used to estimate the cross spectral density matrices. (Zero-padding was used if K was greater than the number of samples in an epoch (10,000)). As can be seen, a total number of 16384 frequency bins is needed to achieve a separation performance of approximately 16dB. Also, similarly to the previous examples, the algorithm shows consistency with regards to improving the separation performance versus increasing M, the number of the epochs (Figure 12). Also in this case, convergence of the ALS algorithm was obtained within 30 – 40 iterations.

C. Comparisons with existing methods

In this section we compare the performance of our method with that of Parra [1]. As with the proposed algorithm, Parra's method uses the estimated CPSD matrices over different time

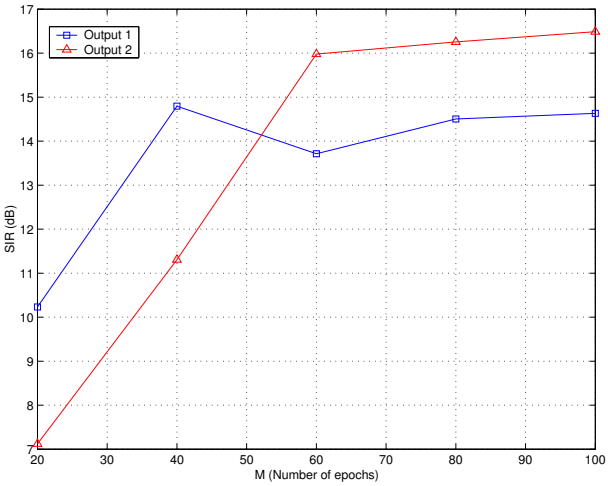

Fig. 12. Results of separation for recordings in an conference room: SIR versus M, the number of epochs, for $K = 16384$.

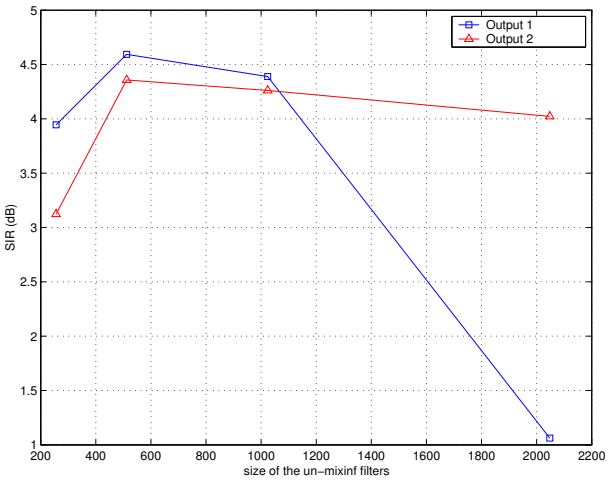

Fig. 13. Separation results for Parra's method for the set of recordings in an office room: SIR versus Q, the size of un-mixing filters, $M = 20$.

segments to calculate the un-mixing filters. Because of this, we use the set of the CPSD matrices, evaluated previously in real room experiment I, as the input to both the proposed method and the algorithm of [1]. Notice since the algorithm in [1] uses a finite length constraint on the size of un-mixing filters, the length of the un-mixing filters needs to be set beforehand in the program. Figure 13 shows the separation performance results for Parra's method versus the size of the un-mixing filters. As can be seen from the figure, the maximum SIR, which is around 4.5 dB, happens when the length of the un-mixing filters is about 512. Increasing the filter lengths after this value degrades the separation performance. The comparative results, for the proposed method and Parra's method, are shown in Figure 14. For this experiment, based on the previous simulation, we chose the length of un-mixing filter to be 512 for Parra's algorithm. As can be seen from the figure, the proposed algorithm outperforms the method in [1] by more than 15dB for $M > 50$.

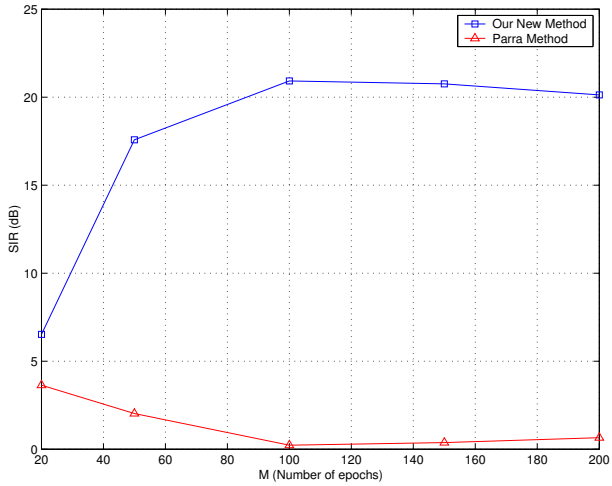

Fig. 14. Comparison of the proposed algorithm with Parra's method: Averaged output SIR versus M, number of epochs.

VIII. Conclusions

In this paper we discussed a new recursive algorithm for blind source separation of convolved non-stationary sources. It was proved that the set of the observed signals' cross spectral density matrices evaluated over different time segments is sufficient to recover the sources up to a frequency dependent scaling and permutation ambiguity, using a joint diagonalization procedure. A two stage frequency-domain algorithm to estimate the un-mixing filters was proposed. A novel procedure for resolving the frequency-dependent permutation ambiguities, and a novel initialization method which significantly improves the perceptual quality of the separated audio signals was also proposed.

The performance of the new algorithm using computer generated sources with real mixing systems under different mixing scenarios was demonstrated. It was shown that the algorithm performs well in real environments with approximately 20dB improvement in signal to interference ratio for a moderately reverberant office area. The results were compared to those of [1].

A significant feature of the proposed method, is that unlike most of the existing algorithms, no assumptions on the structure of the mixing system were made; i.e., the elements of the mixing system can be FIR or IIR filters. Another important feature of the algorithm is use of a recursive least–squares algorithm which has the advantage of fast convergence with no required parameter tuning.

IX. Acknowledgements

The authors wish to acknowledge the support of following institutions: Mitel Corporation, Canada; The Centre for Information Technology Ontario (CITO); and the Natural Sciences and Engineering Research Council of Canada (NSERC).

Appendix

I. Proof of Theorem 1

We want to show that if

$$\mathbf{B}\tilde{\mathbf{\Lambda}}(m)\mathbf{B}^{\dagger} = \boldsymbol{A}\mathbf{\Lambda}(m)\boldsymbol{A}^{\dagger} \quad \forall m = 0, \ldots, M-1 \tag{36}$$

then $\mathbf{B} = \boldsymbol{A}\mathbf{\Pi}e^{\jmath\mathbf{D}}$ for some diagonal matrix $\mathbf{D}$ and permutation matrix $\mathbf{\Pi}$.

First for any sequence of scalars $a = (a_0, ..., a_{M-1})$ we can define the diagonal matrices

$$\mathbf{\Sigma}_a = \sum_{m=0}^{M-1} a_m \mathbf{\Lambda}(m), \;\; \tilde{\mathbf{\Sigma}}_a = \sum_{m=0}^{M-1} a_m \tilde{\mathbf{\Lambda}}(m), \quad a_m \in \mathbb{R}. \tag{37}$$

Therefore an arbitrary linear combination of (36) over m is given as

$$\mathbf{B}\tilde{\mathbf{\Sigma}}_a\mathbf{B}^{\dagger} = \boldsymbol{A}\mathbf{\Sigma}_a\boldsymbol{A}^{\dagger}. \tag{38}$$

Since by assumption the vectors consisting of the diagonal values of $\mathbf{\Lambda}(m)$ span $\mathbb{R}^N$, $\mathbf{\Sigma}_a$ can be made equal to any real valued diagonal matrix by appropriate choice of a. In this instance we choose a such that $\mathbf{\Sigma}_a$ becomes an identity matrix. If $\mathbf{\Sigma}_a$ in (38) is substituted with an identity matrix, since $\boldsymbol{A}$ has full column rank, the LHS of (38) has rank N implying $\mathbf{B}$ has full column rank. Therefore we have $\mathbf{B}^{+}\mathbf{B} = \mathbf{I}$ where $\mathbf{B}^{+}$ is the pseudo inverse of $\mathbf{B}$. Multiplying both sides of (38) by $\mathbf{B}^{+}$ and setting $\mathbf{C} = \mathbf{B}^{+}\boldsymbol{A}$ we have

$$\tilde{\mathbf{\Sigma}}_a = \mathbf{C}\mathbf{\Sigma}_a\mathbf{C}^{\dagger}. \tag{39}$$

Notice that since $\boldsymbol{A}$ and $\mathbf{B}$ are full rank, $\mathbf{C}$ is also a full rank matrix. Now for any i, choose a such that all elements of $\mathbf{\Sigma}_a$ in (39) are zero except for the i_{th} diagonal element which is *one*. Then

$$\mathbf{C}\mathbf{\Sigma}_a\mathbf{C}^{\dagger} = \mathbf{c}_i\mathbf{c}_i^{\dagger} \tag{40}$$

where $\mathbf{c}_i$ is the i_{th} column of $\mathbf{C}$. Moreover, since the LHS of (39) is diagonal, all the off-diagonal elements of $\mathbf{c}_i\mathbf{c}_i^{\dagger}$ are zero. Because $\mathbf{c}_i\mathbf{c}_i^{\dagger}$ has rank one at most, it can have at most one non-zero diagonal element. It follows immediately that every column of $\mathbf{C}$ has precisely one non-zero element, and moreover, because $\mathbf{C}$ is full rank, every row has precisely one non-zero element. Clearly the same is true for $\mathbf{C}^{-1}$; in other words every column and row of $\mathbf{C}^{-1}$ has precisely one non-zero element, or

$$\mathbf{C}^{-1} = \mathbf{\Pi}\tilde{\mathbf{D}} \tag{41}$$

where $\mathbf{\Pi}$ is a permutation matrix and $\tilde{\mathbf{D}}$ is a diagonal matrix. Substituting $\mathbf{C}$ with $\mathbf{B}^{+}\boldsymbol{A}$ and rearranging the terms in (41) we get

$$\mathbf{B} = \mathbf{B}\mathbf{B}^{+}\boldsymbol{A}\mathbf{\Pi}\tilde{\mathbf{D}}. \tag{42}$$

Notice that $\mathbf{B}\mathbf{B}^{+}$ is a projector onto the range space of $\mathbf{B}$. Choosing $\mathbf{\Sigma}_a$ to be the identity matrix in (38) reveals that the range space of $\mathbf{B}$ must contain the range space of $\boldsymbol{A}$. Therefore,

$$\mathbf{B}\mathbf{B}^{+}\boldsymbol{A} = \boldsymbol{A} \tag{43}$$

and from (43) and (42) we can write

$$\mathbf{B} = \mathbf{A}\mathbf{\Pi}\tilde{\mathbf{D}} \tag{44}$$

Since $\mathbf{B}$ and $\mathbf{A}$ both have unit norm columns then the absolute values of the diagonal elements of $\tilde{\mathbf{D}}$ must be one, i.e; $|\tilde{\mathbf{D}}| = \mathbf{I}$ or

$$\tilde{\mathbf{D}} = e^{j\mathbf{D}} \tag{45}$$

where $\mathbf{D}$ is a diagonal matrix. Equation (10) follows immediately from (44) and (45). □

II. Computation of the Cross-Spectral Density Matrices

To estimate the observed signals' cross-spectral density matrices, $\mathbf{P}_x(\omega_k, m)$, $m = 0, \ldots, M-1$, $k = 0, \ldots, K-1$, we first divide the observed sequence into M epochs, where stationarity can be assumed within the epoch but not over more than one epoch. We then apply the following formula to estimate the cross-spectral density matrix at the m_{th} epoch:

$$\hat{\mathbf{P}}(\omega_k, m) = \frac{1}{N_s} \sum_{i=0}^{N_s-1} \mathbf{x}_i(\omega_k, m)\mathbf{x}_i^{\dagger}(\omega_k, m) \tag{46}$$

where

$$\mathbf{x}_i(\omega_k, m) = \sum_{n=-\infty}^{\infty} \mathbf{x}(n)w(n - iT_s - mT_b)e^{-j\omega_k n} \quad k = 0, \ldots, K-1 \tag{47}$$

where N_s is the number of overlapping windows inside each epoch, T_b is the size of each epoch, T_s is the time shift between two overlapping windows, K is the number of frequency bins and $w(n)$ is the windowing sequence. Note that $\mathbf{x}_i(\omega_k, m)$ in (47) is computed using the FFT. For convenience, the estimated cross spectral density matrices can be normalized, and so for our computations we set

$$\hat{\mathbf{P}}_x(\omega_k, m) = \frac{\hat{\mathbf{P}}(\omega_k, m)}{||\hat{\mathbf{P}}(\omega_k, m)||_F}. \tag{48}$$

References

[1] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE Transactions on Speech and Audio Processing*, vol. 8, pp. 320–327, May. 2000.

[2] P. Comon, "Independent component analysis, A new concept?," *SIGNAL PROCESSING*, vol. 36, pp. 287–314, 1994.

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

[4] J. Cardoso and B. Laheld, "Equivariant adaptive source separation," *IEEE Transactions on Signal Processing*, vol. 44, pp. 3017–3030, Dec. 1996.

[5] L. Tong, R. Liu, V. Soon, and Y. Huang, "Indeterminacy and identifiability of blind identification," vol. 38, pp. 499–509, May 1991.

[6] H. Sahlin and H. Broman, "MIMO signal separation for FIR channels: a criterion and performance analysis," *IEEE Transactions on Signal Processing*, vol. 48, pp. 642–649, March. 2000.

[7] D. Yellin and E. Weinstein, "Criteria for multichannel signal separation," *IEEE Transactions on Signal Processing*, vol. 42, pp. 2158–2168, Aug. 1994 1994.

[8] E. Weinstein, M. Feder, and A. V. Oppenheim, "Multi-channel signal separation by decorrelation," *IEEE Transactions on Signal Processing*, vol. 1, pp. 404–413, Oct. 1993.

[9] J. K. Tugnait, "Adaptive blind separation of convolutive mixtures of independent linear signals," *SIGNAL PROCESSING*, vol. 73, pp. 139–152, 1999.

[10] J. Pesquet, B. Chen, and A. P. Petropulu, "Frequency-domain contrast functions for separation of convolutive mixtures," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 2765–2768, May 2001.

[11] J. Cardoso and A. S. Soulomiac, "Jacobi angles for simultaneous diagonalization," *SIAM Journal on Matrix Analysis and Applications*, vol. 17, pp. 161–164, 1 1996.

[12] D. Yellin and E. Weinstein, "Multichannel signal separation: methods and analysis," *IEEE Transactions on Signal Processing*, vol. 44, pp. 106–118, Jan. 1996.

[13] K. Rahbar and J. Reilly, "Geometric optimization methods for blind source separation of signals," in *International Workshop on Independent Component Analysis and Signal Separation*, pp. 375–380, June 2000.

[14] A. Belouchrani, K. Abed_Meraim, and J. Cardoso, "A blind source separation technique using second order statistics," *IEEE Transactions on Signal Processing*, vol. 45, pp. 434–444, Feb. 1997.

[15] U. A. Lindgren and H. Broman, "Source separation using a criterion based on second-order statistics," *IEEE Transactions on Signal Processing*, vol. 46, pp. 1837–1850, July 1998.

[16] D. Chan, P. Rayner, and S. Godsill, "Multi-channel signal separation," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 649–652, May 1996.

[17] K. Meraim, Y. Xiang, and Y. Hua, "Generalized second order identifiability condition and relevant testing technique," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 2989–2992, May 2000.

[18] Y. Hua and J. Tugnait, "Blind Identifiability of FIR-MIMO systems with colored input using second order statistics ," *IEEE Signal Processing Letters*, vol. 7, pp. 348–350, 12 2000.

[19] D. T. Pham and J. Cardoso, "Blind source separation of instantaneous mixtures of nonstationary sources," *IEEE Transactions on Signal Processing*, vol. 49, pp. 1837–1848, Sept. 2001.

[20] K. Abed_Meraim, Y. Xiang, J. H. Manton, and Y. Hua, "Blind Source Separation Using Second-Order Cyclostationary Statistics," *IEEE Transactions on Signal Processing*, vol. 49, pp. 694–701, April 2001.

[21] K. Rahbar and J. Reilly, "Blind source separation of convolved sources by joint approximate diagonalization of cross-spectral density matrices," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 2745–2748, May 2001.

[22] K. Rahbar, J. Reilly, and J. Manton, "A Frequency Domain Approach to Blind Identification of MIMO FIR Systems Driven By Quasi-Stationary Signals," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 1717–1720, May 2002.

[23] K. Rahbar and J. Reilly, "Blind source separation algorithm for MIMO convolutive mixtures," in *International Workshop on Independent Component Analysis and Signal Separation*, pp. 242–247, Dec 2001.

[24] A. Gorokhov and P. Loubaton, "Subspace based techniques for second order blind separation of convolutive mixtures with temporally correlated sources," *IEEE Transactions on Circuits and Systems—I: Fundamental Theory and Applications*, vol. 44, pp. 813–820, September 1997.

[25] C. T. Ma, Z. Ding, and S. F. Yau, "A Two-stage Algorithm for MIMO Blind Deconvolution of Nonstationary Colored Signals," *IEEE Transactions on Signal Processing*, vol. 48, pp. 1187–1192, 4 2000.

[26] K. Rahbar, *Multichannel blind estimation techniques: Blind system identification and blind source separation*. PhD thesis, Dept. of Elect. and Comp. Eng, McMaster University, Hamilton, Ontario, Canada, Nov., 2002.

[27] K. Rahbar, J. Reilly, and J. H. Manton, "Blind identification of mimo fir systems driven by quasi-stationary sources using second order statistics: A frequency domain approach," *Submitted to IEEE Transaction on Signal Processing*, 2002.

[28] T.-W. Lee, A. J. Bell, and R. H. Lambert, "Blind separation of delayed and convolved sources," in *Advances in Neural Information Processing System*, pp. 758–764, MIT Press, 1997.

[29] D. Schobben and P. Sommen, "A new blind signal separation algorithm based on second order statistics," in *IASTED International Conference on Signal and Image Processing*, pp. 564–569, Oct. 1998.

[30] B. Flury and W. Gautschi, "An algorithm for the simultaneous orthogonal transformation of several positive definite symmetric matrices to nearly orthogonal form," *Siam J. of Sci. Stat. Comp.*, vol. 7, pp. 169–184, 1986.

[31] J. Cardoso and A. Souloumiac, "Blind beamforming for non Gaussian signals," in *IEE-F*, vol. 140, pp. 362–370, Dec 1993.

[32] D. Pham, "Joint approximate diagonalization of positive definite hermitian matrices ," in *Technical Report*, (Grenoble University), 2000.

[33] M. Z. Ikram and D. R. Morgan, "Exploring Permutation Inconsistency in Blind Separation Of Signals In a Reverberant Environment," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 1041–1044, May 2000.

[34] S. Araki, S. Makino, T. Nishikawa, and H. Saruwatari, "Fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 2737–2740, May 2001.

[35] S. Talwar, M. Viberg, and A. Paulraj, "Blind separation of synchronous co-channel digital signals using an antenna array-Part I: Algorithms," *IEEE Transactions on Signal Processing*, vol. 44, pp. 1184–1197, May 1996.

[36] T. Li and N. Sidiropoulos, "Blind Digital Signal Separation Using Successive Interference Cancellation Iterative Least Squares," *IEEE Transactions on Signal Processing*, vol. 48, pp. 3146–3152, 11 2000.

[37] N. Sidiropoulos, G. Giannakis, and R. Bro, "Blind PARAFAC Receivers for DS-CDMA Systems," *IEEE Transactions on Signal Processing*, vol. 48, pp. 810–823, 3 2000.

[38] J. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Transactions on Circuits and Systems*, vol. 25, pp. 772–780, Sept. 1979.

[39] G. Golub and C. VanLoan, *Matrix Computations*. Baltimore and London: John Hopkins, third ed., 1996.

[40] N. Mitianoudis and M. Davies, "New fixed-point ica algorithms for convolved mixtures," in *International Workshop on Independent Component Analysis and Signal Separation*, pp. 633–638, Dec 2001.

[41] H. L. Van Trees, *Detection, Estimation and Modulation Theory*. New York, USA: John Wiley & Sons, 1968.

[42] M. Schroeder, "New method for measuring reverberation time," *Journal Acoustic Society America*, vol. 37, pp. 409–412, 1965.