AN ALTERNATING LEAST-SQUARES ALGORITHM FOR APPROXIMATE JOINT DIAGONALIZATION AND ITS APPLICATION TO BLIND SOURCE SEPARATION

Shinya Saito† *Kunio Oishi*⋆ *Toshihiro Furukawa*†

† Department of Management Science, Tokyo University of Science
1-3 Kagurazaka, Shinjuku-ku, Tokyo 162-8601, Japan
⋆ School of Computer Science, Tokyo University of Technology
1404-1 Katakura, Hachioji, Tokyo 192-0982, Japan
Email: sinyasaito@ms.kagu.tus.ac.jp, kohishi@stf.teu.ac.jp, furukawa@ms.kagu.tus.ac.jp

ABSTRACT

This paper presents an iterative alternating least-squares (ALS) algorithm for alternately solving two different least-squares approximate joint diagonalization (LS-AJD) problems for application to convolutive frequency-domain blind source separation (BSS). The constrained forward-model LS-AJD criterion is minimized to estimate the mixing matrix by using the method of Lagrange multipliers. The other criterion, based on backward modeling, is to find the diagonal matrices by the method of least squares. The method of Lagrange multipliers is well suited for accelerating the convergence of the ALS algorithm. The correlation between the interfrequency power ratios is used to prevent misalignment permutation for the new BSS. Finally, we compare our results with those of conventional BSS in highly reverberant environments.

Index Terms— Blind source separation (BSS), convolutive audio mixture, joint diagonalization, alternating least-squares (ALS) algorithm, method of Lagrange multipliers

1. INTRODUCTION

Blind source separation (BSS) is a technique used to recover source signals from observed signals that are modeled as an unknown convolutive mixture of unknown quasistationary source signals. In such cases, the quasistationary signals are modeled as an approximately stationary behavior over a short time interval, known as an epoch. Minimization of the least-squares (LS) criterion requires the mixing matrix to be mathematically equivalent for approximate joint diagonalization (AJD) of the cross-spectral density matrices of the observed signals. The AJD problem [1, 2, 3, 4, 5] entails finding the diagonalizing matrix and diagonal matrices. The LS-AJD estimate is suitable for blind separation of quasistationary sources by estimating the epoch-by-epoch cross-spectral density matrices of the source signal and the mixing matrix simultaneously. An alternating least-squares with projection (ALSP) algorithm for convolutive BSS in the frequency domain has been recently developed [1]. A large number of observed signals are required to achieve a good separation performance in the Rahbar LS-AJD estimate. The main drawback, however, is its slow convergence. With the ALSP algorithm [1], after the unconstrained LS estimation problem is solved by the method of least squares, the constrained LS estimation problem is solved by projecting the unconstrained LS estimate onto the constraint set $\Omega \subset \mathbb{C}^{J^2\times 1}$, defined as $\Omega = \left\{\mathrm{vec}\{\mathbf{\Phi}\} | \mathbf{\Phi} = \mathbf{v}\mathbf{v}^H, \mathbf{v} \in \mathbb{C}^{J\times 1}, \|\mathbf{v}\|_2^2 = 1\right\}$, where $\mathrm{vec}\{\mathbf{A}\}$ forms a column vector by stacking the columns of the matrix $\mathbf{A}$. This operation is fulfilled by using the power method.

In this paper, we introduce two different LS criteria into the AJD problem. The first is a constrained forward-model LS-AJD criterion for estimating the mixing matrix by using the method of Lagrange multipliers. The second is a backward-model LS-AJD criterion for determining the diagonal matrices by using the method of least squares. The mixing matrix obtained from the former LS-AJD estimate must be nonsingular to find a full-rank separated matrix. Therefore, if the mixing matrix is not of full rank, it can be replaced by a full-rank matrix once every iteration before minimizing the latter LS-AJD criterion. The full-rank matrix in our ALS algorithm can lead to stability. The correlation between the interfrequency power ratios [6] is used to solve the permutation problem. The separation performance of the new BSS, in which the number of microphones is set to that of the sources, is demonstrated using artificial room impulse responses (RIRs).

2. PROBLEM FORMULATION AND PREVIOUS WORK ON LS-AJD-BASED BSS

In the convolutive mixing model between N sources $s_1(t), s_2(t), \cdots, s_N(t)$ and J microphones $x_1(t), x_2(t), \cdots, x_J(t)$ at time t, assuming that $h_{ij}(t)$ is a stable and causal non-minimum-phase mixing-filter impulse response from the jth source to the ith microphone without changing over the entire observation interval, we obtain the observed signal at the ith microphone as $x_i(t) = \sum_{j=1}^{N} h_{ij}(t) * s_j(t) + n_i(t)$, where the sources are zero mean, second-order quasistationary signals [1]. In addition, the sources are independent of each other, $J \geq N \geq 2$, the asterisk $*$ denotes time-domain convolution, and an additive white Gaussian noise (AWGN) $n_i(t)$ with mean zero and

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

variance σ^2 is independent of the sources. If the number of the discrete Fourier transform (DFT) points, as represented by K, is significantly larger than the length of the impulse response $h_{ij}(t)$, the time-domain convolution is approximately converted to multiplication by short-time Fourier transform (STFT) as

$$\mathbf{x}(\omega_k, m) \approx \mathbf{H}(\omega_k)\mathbf{s}(\omega_k, m) + \mathbf{n}(\omega_k, m) \tag{1}$$

where $\omega_k = 2\pi k/K$, $k = 0, 1, \cdots, K-1$, $x_i(\omega_k, m)$, $s_j(\omega_k, m)$, and $n_i(\omega_k, m)$ are the STFTs of $x_i(t)$, $s_j(t)$, and $n_i(t)$ at time epoch m, $h_{ij}(\omega_k)$ is the DFT of $h_{ij}(t)$, $\mathbf{s}(\omega_k, m) = [s_1(\omega_k, m), s_2(\omega_k, m), \cdots, s_N(\omega_k, m)]^T$ is the $N \times 1$ vector of sources, $\mathbf{H}(\omega_k)$ is the $J \times N$ mixing matrix of the transfer function from the N sources to the J microphones, the $J \times 1$ observed signal vector is defined by $\mathbf{x}(\omega_k, m) = [x_1(\omega_k, m), x_2(\omega_k, m), \cdots, x_J(\omega_k, m)]^T$, and $\mathbf{n}(\omega_k, m) = [n_1(\omega_k, m), n_2(\omega_k, m), \cdots, n_J(\omega_k, m)]^T$ is the $J \times 1$ vector of AWGN. All observed signals are available in the time epoch interval $1 \le m \le M$, where M is the total number of time epochs. The cross-spectral density matrix of the source signal $\mathbf{P}_s(\omega_k, m) = E[\mathbf{s}(\omega_k, m)\mathbf{s}(\omega_k, m)^H] \in \mathbb{R}^{N\times N}$ is diagonal, where $E[\cdot]$ and the superscript ^{H} denote expectation operation and Hermitian transpose, respectively.

To separate the sources at each frequency bin ω_k independently in BSS, premultiplication of $\mathbf{H}(\omega_k)$ by the $N \times J$ unmixing matrix $\mathbf{W}(\omega_k)$ yields

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

where $\mathbf{\Pi}(\omega_k) \in \mathbb{R}^{N\times N}$ is a frequency-dependent permutation matrix and $\mathbf{D}(\omega_k) \in \mathbb{C}^{N\times N}$ is a scale or phase arbitrary diagonal matrix. Let $\mathbf{P}_x(\omega_k, m) \in \mathbb{C}^{J\times J}$ define the cross-spectral density matrix of the observed signal at point (ω_k, m)

$$\mathbf{P}_x(\omega_k, m) = \mathbf{H}(\omega_k)\mathbf{P}_s(\omega_k, m)\mathbf{H}(\omega_k)^H + \sigma^2\mathbf{I} \tag{3}$$

where $\mathbf{I}$ denotes the $J \times J$ identity matrix, and $\mathbf{H}(\omega_k)$ and $\mathbf{P}_s(\omega_k, m)$ are assumed to be nonsingular. Although a noise-free cross-spectral density matrix can be obtained as $\mathbf{P}_x(\omega_k, m) - \sigma^2\mathbf{I} = \mathbf{H}(\omega_k)\mathbf{P}_s(\omega_k, m)\mathbf{H}(\omega_k)^H$ for the number of microphones being larger than that of sources, where σ^2 is the smallest eigenvalue of the matrix $\mathbf{P}_x(\omega_k, m)$, it cannot be obtained for the number of microphones being equal to that of the sources. If we find a diagonalizing matrix $\mathbf{B}(\omega_k) \in \mathbb{C}^{J\times N}$ and diagonal matrices $\mathbf{\Lambda}(\omega_k, m) \in \mathbb{R}^{N\times N}$ to satisfy

$$\mathbf{P}_x(\omega_k, m) - \sigma^2\mathbf{I} = \mathbf{B}(\omega_k)\mathbf{\Lambda}(\omega_k, m)\mathbf{B}(\omega_k)^H \tag{4}$$

with the scale constraint $\|\mathbf{b}_j(\omega_k)\|_2 = 1$, from (2), the relationship between $\mathbf{B}(\omega_k)$ and $\mathbf{H}(\omega_k)$ becomes $\mathbf{B}(\omega_k) = \mathbf{H}(\omega_k)\mathbf{D}(\omega_k)\mathbf{\Pi}(\omega_k)$, where $\mathbf{b}_j(\omega_k)$ is the jth column of $\mathbf{B}(\omega_k)$, $\|\cdot\|_2$ denotes Euclidean norm, and

$$\mathbf{W}(\omega_k)\mathbf{B}(\omega_k) = \mathbf{I}. \tag{5}$$

In the frequency-domain LS-AJD-based BSS [1], the Welch periodogram method [7] is used to approximate the cross-spectral density matrix of the observed signal. After the M estimated power spectral density matrices are obtained by dividing all observed signals into M time epochs, the estimation value is normalized. By using the normalized estimation value, the measurement error at point (ω_k, m) is given by

$$\mathbf{E}(\omega_k, m) = \tilde{\mathbf{P}}_x(\omega_k, m) - \mathbf{B}(\omega_k)\mathbf{\Lambda}(\omega_k, m)\mathbf{B}(\omega_k)^H. \tag{6}$$

The diagonalizing matrix $\mathbf{B}(\omega_k)$ and M associated diagonal matrices $\mathbf{\Lambda}(\omega_k, 1), \mathbf{\Lambda}(\omega_k, 2), \cdots, \mathbf{\Lambda}(\omega_k, M)$ are estimated by minimizing the sum of the measurement error squares

$$\hat{\mathbf{B}}(\omega_k), \hat{\mathbf{\Lambda}}(\omega_k, m) = \underset{\mathbf{B}(\omega_k),\ \mathbf{\Lambda}(\omega_k, m)}{\operatorname{argmin}} \sum_{m=1}^{M} \|\mathbf{E}(\omega_k, m)\|_F^2 \tag{7}$$

subject to the scale constraint $\left\|\mathbf{b}_j(\omega_k)\right\|_2 = 1$ over a significant number of time epochs M at each frequency bin ω_k.

3. LS-AJD ESTIMATE FOR BSS AND DERIVATION

The new LS-AJD estimate also differs from the conventional LS-AJD estimates in that the former includes two different LS criteria. The first is used to estimate the mixing matrix and the second is used to estimate the cross-spectral density diagonal matrix of the source signal. The ALS algorithm alternates between minimization of the constrained LS criterion with respect to the mixing matrix on the previously obtained estimate of the diagonal matrix and the minimization of the second LS criterion with respect to the diagonal matrix on that of the mixing matrix.

The combined response of the mixing filter with the unmixing filter in tandem must satisfy a minimization of the following error square

$$e(\omega_k) = \|\mathbf{I} - \mathbf{W}(\omega_k)\mathbf{B}(\omega_k)\|_F^2. \tag{8}$$

Minimizing (8) with respect to the unmixing matrix $\mathbf{W}(\omega_k)$ yields the unmixing matrix

$$\mathbf{W}(\omega_k) = \left(\mathbf{B}(\omega_k)^H\mathbf{B}(\omega_k)\right)^{-1}\mathbf{B}(\omega_k)^H. \tag{9}$$

The minimum error square $e_{\min}(\omega_k)$ is found to be $e_{\min}(\omega_k) = 0$ only if $\mathbf{B}(\omega_k)^H\mathbf{B}(\omega_k)$ has full-rank N. On the contrary, if the rank of $\mathbf{B}(\omega_k)^H\mathbf{B}(\omega_k)$ is $r < N$, the value of the error square is greater than $e_{\min}(\omega_k)$. Therefore, in addition to the scale constraint $\left\|\mathbf{b}_j(\omega_k)\right\|_2 = 1$, we impose the constraint $\operatorname{rank}\left(\mathbf{B}(\omega_k)^H\mathbf{B}(\omega_k)\right) = N$ on (7), where rank($\mathbf{A}$) denotes the rank of matrix $\mathbf{A}$. That is, $\mathbf{B}(\omega_k)^H\mathbf{B}(\omega_k)$ obtained from minimizing (7) must be nonsingular to find $e_{\min}(\omega_k) = 0$. In the new BSS, $\mathbf{B}(\omega_k)$ can be replaced by a full-rank matrix only once every iteration before estimating the diagonal matrix, but only if $\mathbf{B}(\omega_k)^H\mathbf{B}(\omega_k)$ is not of full rank.

Speech is characterized by spectrum envelope peaks, known as formants, most of which are in the low-frequency range. By normalizing the estimated point-by-point cross-spectral density matrix of the observed signal $\hat{\mathbf{P}}_x(\omega_k, m)$ by its magnitude [1], the spectrum envelope becomes flat in the frequency band. Our epoch-by-epoch normalization differs from the point-by-point normalization used in [1]. Without losing the characteristics of such a peak produced in speech,

our objective is to normalize $\hat{\mathbf{P}}_x(\omega_k, m)$ by the maximum magnitude of all observed signals at the time epoch m

$$\tilde{\mathbf{P}}_x(\omega_k, m) = \frac{\hat{\mathbf{P}}_x(\omega_k, m)}{\max_{\omega_k} \left\| \hat{\mathbf{P}}_x(\omega_k, m) \right\|_F} \tag{10}$$

where $\hat{\mathbf{P}}_x(\omega_k, m)$ is estimated by the Welch periodogram method. The spectrum envelope derived by our normalization can approximate the true spectrum envelope of the observed signal.

The LS-AJD problem is used to find the diagonalizing matrix $\mathbf{B}(\omega_k)$ and M associated diagonal matrices $\mathbf{\Lambda}(\omega_k, 1), \mathbf{\Lambda}(\omega_k, 2), \cdots, \mathbf{\Lambda}(\omega_k, M)$ by minimizing the cost function

$$\xi(\omega_k) = \sum_{m=1}^{M} \|\mathbf{E}(\omega_k, m)\|_F^2 + 2 \sum_{i=1}^{N} \gamma_i \left(\mathbf{b}_i(\omega_k)^H \mathbf{b}_i(\omega_k) - 1 \right) \tag{11}$$

at each frequency bin ω_k over M time epochs, where γ_i is the Lagrange multiplier. By using the Khatri-Rao (KR) product, we can rewrite (11) as

$$\xi(\omega_k) = \sum_{m=1}^{M} \|\boldsymbol{\varepsilon}(\omega_k, m)\|_2^2 + 2 \sum_{i=1}^{N} \gamma_i \left(\text{vec}\{\mathbf{I}\}^T \text{Re}\left(\mathbf{g}_i(\omega_k)\right) - 1 \right) \tag{12}$$

where

$$\boldsymbol{\varepsilon}(\omega_k, m) = \tilde{\mathbf{r}}_x(\omega_k, m) - \mathbf{G}(\omega_k)\mathbf{d}(\omega_k, m) \tag{13}$$

$$\tilde{\mathbf{r}}_x(\omega_k, m) = \text{vec}\left\{\tilde{\mathbf{P}}_x(\omega_k, m)\right\} \tag{14}$$

$$\mathbf{G}(\omega_k) = [\mathbf{b}_1(\omega_k)^* \otimes \mathbf{b}_1(\omega_k), \cdots, \mathbf{b}_N(\omega_k)^* \otimes \mathbf{b}_N(\omega_k)] \tag{15}$$

$$\mathbf{d}(\omega_k, m) = [\lambda_1(\omega_k, m), \lambda_2(\omega_k, m), \cdots, \lambda_N(\omega_k, m)]^T \tag{16}$$

$$\begin{aligned} \mathbf{G}(\omega_k)\mathbf{d}(\omega_k, m) &= \text{vec}\left\{\mathbf{B}(\omega_k)\mathbf{\Lambda}(\omega_k, m)\mathbf{B}(\omega_k)^H\right\} \\ &= [\mathbf{B}(\omega_k)^* \odot \mathbf{B}(\omega_k)] \cdot \mathbf{d}(\omega_k, m), \end{aligned} \tag{17}$$

the superscript *, $\otimes$, and $\odot$ denote the complex conjugate, Kronecker product, and KR product respectively; $\lambda_i(\omega_k, m)$ denotes ith diagonal element of $\mathbf{\Lambda}(\omega_k, m)$; and $\mathbf{g}_j(\omega_k)$ denotes jth column of $\mathbf{G}(\omega_k)$.

To apply the ALS algorithm, $\mathbf{z}_i(\omega_k)$ and $\mathbf{T}(\omega_k)$ are defined by

$$\mathbf{z}_i(\omega_k) = [\lambda_i(\omega_k, 1), \lambda_i(\omega_k, 2), \cdots, \lambda_i(\omega_k, M)]^T \tag{18}$$

$$\mathbf{T}(\omega_k) = [\tilde{\mathbf{r}}_x(\omega_k, 1), \tilde{\mathbf{r}}_x(\omega_k, 2), \cdots, \tilde{\mathbf{r}}_x(\omega_k, M)]. \tag{19}$$

The ALS algorithm alternates the following two phases. In phase one, to minimize (12) with respect to $\mathbf{g}_j(\omega_k)$ while keeping its other columns $\mathbf{g}_i(\omega_k)$ and $\mathbf{z}_i(\omega_k)$ fixed and defining

$$\mathbf{F}_j(\omega_k) = \mathbf{T}(\omega_k) - \sum_{i=1, i\neq j}^{N} \mathbf{g}_i(\omega_k)\mathbf{z}_i(\omega_k)^H, \tag{20}$$

by using (18) and (19), we have the constrained LS estimation problem

$$\xi(\omega_k) = \left\| \mathbf{F}_j(\omega_k) - \mathbf{g}_j(\omega_k)\mathbf{z}_j(\omega_k)^H \right\|_F^2 + 2 \sum_{i=1}^{N} \gamma_i \left(\text{vec}\{\mathbf{I}\}^T \text{Re}(\mathbf{g}_i(\omega_k)) - 1 \right). \tag{21}$$

This problem is solved by the method of Lagrange multipliers as follows:

$$\hat{\mathbf{g}}_j(\omega_k) = \frac{1}{\left\|\mathbf{z}_j(\omega_k)\right\|_2^2} \left[\mathbf{F}_j(\omega_k) \text{Re}\left(\mathbf{z}_j(\omega_k)\right) - \frac{1}{J} \left(\text{vec}\{\mathbf{I}\}^T \text{Re}\left(\mathbf{F}_j(\omega_k)\right) \text{Re}\left(\mathbf{z}_j(\omega_k)\right) - \left\|\mathbf{z}_j(\omega_k)\right\|_2^2 \right) \text{vec}\{\mathbf{I}\} \right]. \tag{22}$$

That is, differentiating (21) with respect to $\mathbf{g}_j(\omega_k)$, setting the derivative to zero, then substituting the result into the constraint $\text{vec}\{\mathbf{I}\}^T \text{Re}\left(\mathbf{g}_j(\omega_k)\right) - 1$ yields $\hat{\mathbf{g}}_j(\omega_k)$. After minimizing (12) with respect to each column of $\mathbf{G}(\omega_k)$ successively while keeping the others fixed, (22) is repeated until $\mathbf{G}(\omega_k)$ changes by less than ϵ_G between iterations. After the convergence criterion is satisfied, $\left\| \text{unvec}\left\{\hat{\mathbf{g}}_j(\omega_k)\right\} - \mathbf{b}_j(\omega_k)\mathbf{b}_j(\omega_k)^H \right\|_F^2$ is minimized using the power method [8] to find $\mathbf{b}_j(\omega_k)$ for $j = 1, 2, \cdots, N$ only once every iteration, where unvec$\{\mathbf{a}\}$ forms a matrix from the column vector $\mathbf{a}$.

The singular value decomposition (SVD) of the matrix $\mathbf{B}(\omega_k)$ is given by

$$\mathbf{B}(\omega_k) = \mathbf{V}_r(\omega_k)\mathbf{\Sigma}_r(\omega_k)\mathbf{U}_r(\omega_k)^H \tag{23}$$

where $\mathbf{V}_r(\omega_k)$, $\mathbf{U}_r(\omega_k)$, and $\mathbf{\Sigma}_r(\omega_k)$ denote unitary matrices and the diagonal matrix

$$\mathbf{V}_r(\omega_k) = [\mathbf{v}_1(\omega_k), \mathbf{v}_2(\omega_k), \cdots, \mathbf{v}_r(\omega_k)] \tag{24}$$

$$\mathbf{U}_r(\omega_k) = [\mathbf{u}_1(\omega_k), \mathbf{u}_2(\omega_k), \cdots, \mathbf{u}_r(\omega_k)] \tag{25}$$

$$\mathbf{\Sigma}_r(\omega_k) = \text{diag}\left(\sigma_1(\omega_k), \sigma_2(\omega_k), \cdots, \sigma_r(\omega_k)\right) \tag{26}$$

$$\sigma_1(\omega_k) \geq \sigma_2(\omega_k) \geq \cdots \geq \sigma_r(\omega_k) > 0. \tag{27}$$

Because of the condition $\left\|\mathbf{b}_j(\omega_k)\right\|_2 = 1$ for $j = 1, 2, \cdots, N$, the trace of $\mathbf{\Sigma}_r(\omega_k)$ is equal to $\sqrt{N}$, that is, $\text{tr}[\mathbf{\Sigma}_r(\omega_k)] = \sqrt{N}$. We form an orthonormal set $\mathbf{v}_{r+1}(\omega_k), \mathbf{v}_{r+2}(\omega_k), \cdots, \mathbf{v}_N(\omega_k)$ orthogonal to the orthonormal set $\mathbf{v}_1(\omega_k), \mathbf{v}_2(\omega_k), \cdots, \mathbf{v}_r(\omega_k)$. Similarly, we form an orthonormal set $\mathbf{u}_{r+1}(\omega_k), \mathbf{u}_{r+2}(\omega_k), \cdots, \mathbf{u}_N(\omega_k)$ orthogonal to the orthonormal set $\mathbf{u}_1(\omega_k), \mathbf{u}_2(\omega_k), \cdots, \mathbf{u}_r(\omega_k)$. To realize the minimum error square $e_{\min}(\omega_k)$, if the rank of $\mathbf{B}(\omega_k)^H\mathbf{B}(\omega_k)$ is $r < N$, $\mathbf{B}(\omega_k)$ can always be replaced by the following full-rank matrix

$$\begin{aligned} \mathbf{V}(\omega_k)\mathbf{\Sigma}(\omega_k)\mathbf{U}(\omega_k)^H &= \frac{\sqrt{N}}{\sqrt{N} + \delta(\omega_k)N} \\ &\cdot \left[\mathbf{V}_r(\omega_k), \mathbf{V}_f(\omega_k)\right] \left(\begin{bmatrix} \mathbf{\Sigma}_r(\omega_k) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \delta(\omega_k)\mathbf{I} \right) \left[\mathbf{U}_r(\omega_k), \mathbf{U}_f(\omega_k)\right]^H \end{aligned}$$

constructed by the Gram-Schmidt orthonormalization process, where we choose $\delta(\omega_k)$ in the range $\delta(\omega_k) > 0$; $\sqrt{N} / \left(\sqrt{N} + \delta(\omega_k)N\right)$ in the right-hand side is required to satisfy the condition $\text{tr}[\mathbf{\Sigma}(\omega_k)] = \sqrt{N}$, and

$$\mathbf{V}_f(\omega_k) = [\mathbf{v}_{r+1}(\omega_k), \mathbf{v}_{r+2}(\omega_k), \cdots, \mathbf{v}_N(\omega_k)] \tag{28}$$

$$\mathbf{U}_f(\omega_k) = [\mathbf{u}_{r+1}(\omega_k), \mathbf{u}_{r+2}(\omega_k), \cdots, \mathbf{u}_N(\omega_k)]. \tag{29}$$

Because our LS-AJD estimate guarantees (5), it follows from (4) that pre- and post-multiplying $\mathbf{\Lambda}(\omega_k, m)$ by $\mathbf{W}(\omega_k)\mathbf{B}(\omega_k)$ and $(\mathbf{W}(\omega_k)\mathbf{B}(\omega_k))^H$ yields

$$\begin{aligned} \mathbf{\Lambda}(\omega_k, m) &= \mathbf{W}(\omega_k)\mathbf{B}(\omega_k)\mathbf{\Lambda}(\omega_k, m)\mathbf{B}(\omega_k)^H\mathbf{W}(\omega_k)^H \\ &= \mathbf{W}(\omega_k)\left(\mathbf{P}_x(\omega_k, m) - \sigma^2\mathbf{I}\right)\mathbf{W}(\omega_k)^H. \end{aligned} \tag{30}$$

This equation is satisfied under the correct estimates of $\mathbf{W}(\omega_k)$ and $\mathbf{P}_x(\omega_k, m) - \sigma^2\mathbf{I}$. By using the estimate of $\mathbf{P}_x(\omega_k, m)$, let $\mathbf{\Psi}(\omega_k, m)$ define a measurement error

$$\mathbf{\Psi}(\omega_k, m) = \mathbf{W}(\omega_k)\tilde{\mathbf{P}}_x(\omega_k, m)\mathbf{W}(\omega_k)^H - \mathbf{\Lambda}(\omega_k, m). \tag{31}$$

Table 1. Procedure for applying the LS-AJD-based BSS.

1) Normalize $\hat{\mathbf{P}}_x(\omega_k, m)$ by (10) for $m = 1, 2, \cdots, M$.
2) Minimize (21) to find $\hat{\mathbf{g}}_j(\omega_k)$ by the method of Lagrange multipliers for $j = 1, 2, \cdots, N$.
3) Go to step 2 until $\mathbf{G}(\omega_k)$ changes by less than ϵ_G between iterations.
4) Minimize $\left\|\text{unvec}\left\{\hat{\mathbf{g}}_j(\omega_k)\right\} - \mathbf{b}_j(\omega_k)\mathbf{b}_j(\omega_k)^H\right\|_F^2$ to find $\mathbf{b}_j(\omega_k)$ by the power method for $j = 1, 2, \cdots, N$.
5) If rank $\left(\mathbf{B}(\omega_k)^H\mathbf{B}(\omega_k)\right) = r < N$, construct the orthonormal sets $\mathbf{v}_{r+1}(\omega_k), \mathbf{v}_{r+2}(\omega_k), \cdots, \mathbf{v}_N(\omega_k)$ and $\mathbf{u}_{r+1}(\omega_k), \mathbf{u}_{r+2}(\omega_k), \cdots, \mathbf{u}_N(\omega_k)$ by the Gram-Schmidt process. Then replace $\mathbf{B}(\omega_k)$ with $\mathbf{V}(\omega_k)\mathbf{\Sigma}(\omega_k)\mathbf{U}(\omega_k)^H$.
6) Compute $\mathbf{W}(\omega_k)$ from $\mathbf{B}(\omega_k)$.
7) Minimize $\sum_{m=1}^{M} \|\mathbf{\Psi}(\omega_k, m)\|_F^2$ to find $\hat{\mathbf{\Lambda}}(\omega_k, m)$ by the method of least squares.
8) Go to step 2 until (21) changes by less than ϵ_C between iterations.
9) Solve the scale problem to find $\mathbf{D}(\omega_k)^{-1}\mathbf{W}(\omega_k)$.
10) Resolve the permutation ambiguity to find $\mathbf{\Pi}(\omega_k)^{-1}\mathbf{D}(\omega_k)^{-1}\mathbf{W}(\omega_k)$.
11) Convert $e^{-j\pi k}\mathbf{\Pi}(\omega_k)^{-1}\mathbf{D}(\omega_k)^{-1}\mathbf{W}(\omega_k)$ in the time domain by the inverse FFT to realize a delayed unmixing filter.

While keeping $\mathbf{W}(\omega_k)$ fixed, we find

$$\hat{\mathbf{\Lambda}}(\omega_k, m) = \text{diag}\left[\mathbf{W}(\omega_k)\tilde{\mathbf{P}}_x(\omega_k, m)\mathbf{W}(\omega_k)^H\right] \qquad (32)$$

by minimizing the LS criterion $\sum_{m=1}^{M} \|\mathbf{\Psi}(\omega_k, m)\|_F^2$ with respect to $\mathbf{\Lambda}(\omega_k, m)$ [9, 10], where diag[$\mathbf{A}$] denotes the diagonal matrix of the matrix $\mathbf{A}$ and $\mathbf{W}(\omega_k)$ is calculated by (9). The ALS algorithm is repeated until (21) changes by less than ϵ_C between iterations. The procedure of the LS-AJD estimate is shown in Table 1.

4. SIMULATION RESULTS

We generated artificial RIRs from three sources to three microphones in a room by using the image method [11, 12] at a sampling rate of 8 kHz. The room size was 4.45×3.55×2.5 m. The three loudspeakers were located at [3.35, 1.36, 1.2], [2.83, 2.81, 1.2], and [1.14, 2.28, 1.2]. The three microphones were placed at [2.34, 1.78, 1.2], [2.17, 1.88, 1.2], and [2.17, 1.68, 1.2]. The speech dataset duration was 1 000 s [13]. The Hanning window was used for STFT. The parameter was chosen empirically as an 8192-point FFT, $\epsilon_G = \epsilon_C = 10^{-6}$, and $\delta(\omega_k) = \sigma_r(\omega_k)$. In the new, Rahbar, and Parra BSSs [1, 9], the scale problem was solved by normalizing each row vector of $\mathbf{W}(\omega_k)$ at each frequency bin. The new BSS solved the permutation problem by using the approach based on the correlation between the interfrequency power ratios [6]. To estimate $\mathbf{P}_x(\omega_k, m)$, the number of 80% overlapping frames within each epoch was set to 2 for the new and the Rahbar BSSs. The signal-to-noise ratio (SNR) was determined by the ratio of the desired signal power and the power of interference plus the noise component in the output signals [3]. When the optimum permutation and the optimum unmixing matrix were available, the optimum output signal-to-interference ratio (SIR) was equal to the SNR [3]. The simulation program was coded in C programming language and run on an Intel Core i7-2600 3.4 GHz processor.

Table 2. Comparison of the average output SIR, the average CPU time per frequency bin, and the average number of iterations per frequency bin with the conventional BSS for SNR ≈ 20 dB, $N = J = 3$, and 8192-point FFT.

Method	Reverberation time [ms]				
	100	300	500	700	900
	Overall input SIR [dB]				
	−2.59	−3.39	−3.63	−3.66	−3.62
New					
Output SIR [dB]	**17.43**	**17.23**	**15.27**	**13.62**	**11.93**
CPU time [s]	0.124	0.179	0.182	0.190	0.206
Iterations	**4.77**	**5.25**	**6.04**	**6.33**	**7.03**
Rahbar [1]					
Output SIR [dB]	10.13	4.65	3.25	2.67	1.56
CPU time [s]	3.544	13.905	14.979	16.336	17.246
Iterations	110.41	147.05	157.12	158.18	162.81
Parra [9, 14]					
Output SIR [dB]	3.69	1.37	0.29	−0.46	−0.67
CPU time [s]	**0.073**	**0.080**	**0.085**	**0.108**	**0.117**
Iterations	127.38	136.72	146.21	158.38	171.48

We applied the new method in reverberant environments. Table 2 shows the average output SIR, the average CPU time per frequency bin, and the average number of iterations per frequency bin until convergence for the batch implementations of these BSSs was achieved. We achieved good separation performance, although the performance was degraded for reverberation times longer than 300 ms. Compared with the Rahbar BSS, the new BSS increased to a maximum of 12.58 dB in a 300-ms reverberant environment. The reverberation dominates the separation performance in the long reverberation time range. Therefore, the separation performance worsens as reverberation time increases. The convergence of the ALSP algorithm was very slow. A striking improvement was apparent by using our algorithm such that convergence to an LS-AJD estimate was made within approximately 5.25 iterations per frequency bin in a 300-ms reverberant environment. We note that the complexity of the new BSS is at most 0.01 times that of the Rahbar BSS.

5. CONCLUSION

We have exploited an iterative ALS algorithm for the convolutive frequency-domain blind separation of quasistationary sources to include two different LS criteria. The ALS algorithm alternates between the minimization of the constrained LS criterion with respect to the mixing matrix on the previously obtained estimate of the diagonal matrices and the minimization of the second LS criterion with respect to the diagonal matrices on that of the mixing matrix. In numerical examples, we showed that the ALS algorithm provides superior convergence properties and offers an improvement in highly reverberant environments.

6. REFERENCES

[1] K. Rahbar and J. P. Reilly, "A frequency domain method for blind source separation of convolutive audio mixtures," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 5, pp. 832–844, Sept. 2005.

[2] D. Nion, K. N. Mokios, N. D. Sidiropoulos, and A. Potamianos, "Batch and adaptive PARAFAC-based blind separation of convolutive speech mixtures," *IEEE Trans. Audio Speech, Lang. Process.*, vol. 18, no. 6, pp. 1193–1207, Aug. 2010.

[3] S. Saito, K. Oishi, and T. Furukawa, "An approach to convolutive backward-model blind source separation based on joint diagonalization," in *Proc. European Signal Process. Conference (EUSIPCO2012)*, Bucharest, Romania, Aug. 2012, pp. 579–583.

[4] M. Ammar, K. Abed-Meraim, and A. Belouchrani, "Comparative performance analysis of non orthogonal joint diagonalization algorithms," in *The 8th Int. Workshop on Syst., Signal Process. and their Applications*, Zeralda, Algiers, May 2013, pp. 50–54.

[5] K.-K. Lee, W.-K. Ma, X. Fu, T.-H. Chan, and C.-Y. Chi, "A Khatri-Rao subspace approach to blind identification of mixtures of quasi-stationary sources," *Signal Processing*, vol. 93, pp. 3515–3527, Apr. 2013.

[6] H. Sawada, S. Araki, and S. Makino, "Measuring dependence of bin-wise separated signals for permutation alignment in frequency-domain bss," in *IEEE Int. Symp. Circuits Syst.*, May 2007, pp. 3247–3250.

[7] P. D. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Trans. Audio Electroacoust.*, vol. 15, no. 2, pp. 70–73, June 1967.

[8] G. H. Golub and C. F. Van loan, "Matrix computations," *The Johns Hopkins University Press*, 1991.

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

[10] K. Tachibana, H. Saruwatari, Y. Mori, S. Miyabe, K. Shikano, and A. Tanaka, "Efficient blind source separation combining closed-form second-order ICA and nonclosed-form higher-order ICA," in *Proc. 2007 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2007)*, Honolulu, HI, April 2007, vol. 1, pp. 45–48.

[11] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoustical Society of America*, pp. 943–950, Apr. 1979.

[12] "Room impulse response generator for MATLAB," http://home.tiscali.nl/ehabets/rir_generator.html.

[13] "ASJ continuous speech corpus for research (ASJ-JIPDEC)," *National Institute of Information*, http://research.nii.ac.jp/src/en/ASJ-JIPDEC.html.

[14] M. Z. Ikram and D. R. Morgan, "Permutation inconsistency in blind speech separation: investigation and solutions," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 1, pp. 1–13, Jan. 2005.