Acceleration of rank-constrained spatial covariance matrix estimation for blind speech extraction

Yuki Kubo[*], Norihiro Takamune[*], Daichi Kitamura[†], and Hiroshi Saruwatari[*]
[*] The University of Tokyo, Graduate School of Information Science and Technology, Tokyo, Japan
[†] National Institute of Technology, Kagawa College, Kagawa, Japan

Abstract—In this paper, we propose new accelerated update rules for rank-constrained spatial covariance model estimation, which efficiently extracts a directional target source in diffuse background noise. The naive update rule requires heavy computation such as matrix inversion or matrix multiplication. We resolve this problem by expanding matrix inversion to reduce computational complexity; in the parameter update step, we need neither matrix inversion nor multiplication. In an experiment, we show that the proposed accelerated update rule achieves 87 times faster calculation than the naive one.

I. Introduction

Blind source separation (BSS) [1] is a technique that separates an observed multichannel signal into each source signal without any prior information about each source or the mixing system. In a determined or overdetermined situation (number of sensors ≥ number of sources), frequency-domain independent component analysis (FDICA) [2]–[4], independent vector analysis (IVA) [5]–[7], and independent low-rank matrix analysis (ILRMA) [8]–[12] have been proposed for audio BSS problems.

In this paper, we address a situation where a directional target source and diffuse noise that arrives from all directions are mixed. In this case, FDICA, IVA, and ILRMA do not enable the extraction of only the target source in principle [13], and the estimated target source includes residual diffuse noise.

To model such diffuse (spatially spread) noise, a multichannel extension of nonnegative matrix factorization (multichannel NMF: MNMF) [14]–[16] was proposed. MNMF estimates a full-rank spatial covariance matrix (SCM) [17], which represents the time-invariant spatial characteristics of each source. However, since the number of parameters in MNMF is large, its optimization requires a huge computational cost and lacks robustness against the initialization [8].

FastMNMF [18], [19] enables computationally efficient estimation by introducing a jointly diagonalizable SCM into the MNMF model. Although this assumption greatly reduces the computational cost of the update algorithm, it still suffers from dependence on the initial values.

To solve the above-mentioned problems, we proposed rank-constrained SCM estimation [20], which is a postprocessing method for ILRMA. In this method, to effectively model diffuse noise in an overdetermined situation, a full-rank SCM of diffuse noise is recovered from the estimates obtained by ILRMA. Since ILRMA can precisely cancel the directional target source [21], an accurate rank-$(M-1)$ SCM of diffuse noise

can be obtained. Thus, the rank-constrained SCM estimation enables us to restore the lost spatial basis of diffuse noise, resulting in accurate extraction of the directional target source. By employing ILRMA as a preprocessing method, we can achieve both lower-cost computation and more initialization-robust estimation compared with MNMFs.

In this paper, we present a more computationally efficient algorithm for rank-constrained SCM estimation. The naive update rule in rank-constrained SCM estimation (hereafter referred to as *naive update*) requires matrix inversion at each time-frequency slot, which leads to a heavy computational load. To solve this problem, we propose an accelerated algorithm that expands the inversion of matrices using the Sherman–Morrison formula and pseudoinverse. The computational cost of each update is reduced from $O(IJM^3)$ to $O(IJ)$, where I and J are the numbers of frequency bins and time frames, respectively. The efficacy of the proposed algorithm is confirmed via BSS with directional target speech and diffuse babble noise.

II. Rank-Constrained SCM Estimation

A. Formulation

Let us denote a source signal and a multichannel observed signal as $\boldsymbol{s}_{ij} = (s_{ij,1}, \ldots, s_{ij,n}, \ldots, s_{ij,N})^{\mathsf{T}} \in \mathbb{C}^N$ and $\boldsymbol{x}_{ij} = (x_{ij,1}, \ldots, x_{ij,m}, \ldots, x_{ij,M})^{\mathsf{T}} \in \mathbb{C}^M$, where $i = 1, \ldots, I$, $j = 1, \ldots, J$, $n = 1, \ldots, N$, and $m = 1, \ldots, M$ are the indices of the frequency bins, time frames, microphones, and sources, respectively, N is the number of sources, and $^{\mathsf{T}}$ denotes the transpose.

If each source can be considered a point source and the reverberation time is sufficiently shorter than the window length in a short-time Fourier transform (STFT), there exists a mixing matrix $\mathbf{A}_i = (\boldsymbol{a}_{i,1} \cdots \boldsymbol{a}_{i,N}) \in \mathbb{C}^{M \times N}$ for each frequency bin and the following holds:

$$\boldsymbol{x}_{ij} = \mathbf{A}_i \boldsymbol{s}_{ij}, \tag{1}$$

where $\boldsymbol{a}_{i,n}$ is the steering vector of source n at frequency i.

When $M \geq N$, independence-based BSS, such as FDICA, IVA, and ILRMA, can be applied to estimate the demixing matrix $\mathbf{W}_i = \mathbf{A}_i^{-1}$, where dimensionality reduction is used so that $M = N$. In this work, we use ILRMA as a state-of-the-art BSS method. For the case of a mixture of one directional target source and diffuse noise, ILRMA outputs the directional target source estimate and $M - 1$ diffuse noise components, although the directional target source estimate contains many

diffuse noise components in the same direction. Therefore, we can calculate M rank-1 SCMs, namely, one rank-1 SCM of the directional target source and $M-1$ rank-1 SCMs of the diffuse noise components, from the estimated demixing matrix $\mathbf{W}_i$. One of these SCMs corresponds to the target source, and the others are components of diffuse noise. A rank-$(M-1)$ SCM of diffuse noise can be obtained by summing the $M-1$ rank-1 SCMs. Since the rank of the SCM of diffuse noise is $M-1$, we need to restore one lost spatial basis (steering vector) of diffuse noise to extract the directional target source with a multichannel Wiener filter.

B. Rank-constrained SCM Estimation

1) Generative model: We assume the observed signal $\boldsymbol{x}_{ij}$ to be the sum of two components, as

$$\boldsymbol{x}_{ij} = \boldsymbol{h}_{ij} + \boldsymbol{u}_{ij}, \tag{2}$$

where $\boldsymbol{h}_{ij} = (h_{ij,1}, \ldots, h_{ij,M})^{\mathsf{T}} \in \mathbb{C}^M$ and $\boldsymbol{u}_{ij} = (u_{ij,1}, \ldots, u_{ij,M})^{\mathsf{T}} \in \mathbb{C}^M$ are spatial images of the directional target source and diffuse noise, respectively. $\boldsymbol{h}_{ij}$ is modeled as

$$\boldsymbol{h}_{ij} = \boldsymbol{a}_i^{(h)} s_{ij}^{(h)}, \tag{3}$$

$$s_{ij}^{(h)} | r_{ij}^{(h)} \sim \mathcal{N}_c(0, r_{ij}^{(h)}), \tag{4}$$

$$r_{ij}^{(h)} \sim \mathcal{IG}(\alpha, \beta), \tag{5}$$

where $\boldsymbol{a}_i^{(h)}$ is the n_hth steering vector $\boldsymbol{a}_{i,n_h}$, $s_{ij}^{(h)}$ is the dry source of the directional target source s_{ij,n_h}, $r_{ij}^{(h)}$ is the power spectrogram of the n_hth source, and n_h denotes the index of the directional target source. Here, $\mathcal{N}_c$ and $\mathcal{IG}$ denote the circularly symmetric complex Gaussian distribution and the inverse gamma distribution, respectively:

$$\mathcal{N}_c(s_{ij}^{(h)} | 0, r_{ij}^{(h)}) = \frac{1}{\pi r_{ij}^{(h)}} \exp\left(-\frac{|s_{ij}^{(h)}|^2}{r_{ij}^{(h)}}\right), \tag{6}$$

$$\mathcal{IG}(r_{ij}^{(h)}; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} (r_{ij}^{(h)})^{-\alpha-1} \exp\left(-\frac{\beta}{r_{ij}^{(h)}}\right), \tag{7}$$

where $\alpha > 0$ and $\beta > 0$ are the shape and scale parameter of the inverse gamma distribution, respectively. Introducing the above prior distribution improves the estimation performance.

The generative model of the diffuse noise $\boldsymbol{u}_{ij}$ is assumed to be the multivariate complex Gaussian distribution:

$$\boldsymbol{u}_{ij} \sim \mathcal{N}_c(\mathbf{0}, r_{ij}^{(u)} \mathbf{R}_i^{(u)}), \tag{8}$$

$$p(\boldsymbol{u}_{ij}) = \frac{1}{\pi^M (r_{ij}^{(u)})^M |\det \mathbf{R}_i^{(u)}|} \exp\left(-\frac{\boldsymbol{u}_{ij}^{\mathsf{H}} (\mathbf{R}_i^{(u)})^{-1} \boldsymbol{u}_{ij}}{r_{ij}^{(u)}}\right), \tag{9}$$

where $r_{ij}^{(u)}$ and $\mathbf{R}_i^{(u)}$ are the variance and full-rank SCM of diffuse noise, respectively. In rank-constrained SCM estimation, $\mathbf{R}_i^{(u)}$ is expressed using separation filter $\mathbf{W}_i =$

$(\boldsymbol{w}_{i,1} \cdots \boldsymbol{w}_{i,M})^{\mathsf{H}}$ estimated by ILRMA as follows:

$$\mathbf{R}_i^{(u)} = \mathbf{R}_i^{\prime(u)} + \lambda_i \boldsymbol{b}_i \boldsymbol{b}_i^{\mathsf{H}}, \tag{10}$$

$$\mathbf{R}_i^{\prime(u)} = \frac{1}{J} \sum_j \mathbf{W}_i^{-1} (|\boldsymbol{w}_{i,1}^{\mathsf{H}} \boldsymbol{x}_{ij}|^2, \ldots, |\boldsymbol{w}_{i,n_h-1}^{\mathsf{H}} \boldsymbol{x}_{ij}|^2, 0,$$

$$|\boldsymbol{w}_{i,n_h+1}^{\mathsf{H}} \boldsymbol{x}_{ij}|^2, \ldots, |\boldsymbol{w}_{i,M}^{\mathsf{H}} \boldsymbol{x}_{ij}|^2)(\mathbf{W}_i^{-1})^{\mathsf{H}}, \tag{11}$$

where $\mathbf{R}_i^{\prime(u)}$ is the rank-$(M-1)$ SCM estimated by ILRMA, $\boldsymbol{b}_i$ is the unit eigenvector of $\mathbf{R}_i^{\prime(u)}$ that corresponds to the zero eigenvalue, λ_i is a scalar variable, and H denotes the Hermitian transpose. Equation (11) applies back-projection [22] to $M-1$ diffuse noise components (estimates) and calculates the sum of the $M-1$ resultant SCMs. Thus, we fix both the spatial basis of the directional target source $\boldsymbol{a}_i^{(h)}$ and the rank-$(M-1)$ SCM $\mathbf{R}_i^{\prime(u)}$. The full-rank SCM $\mathbf{R}_i^{(u)}$ can be recovered by estimating the residual eigenvalue λ_i. Also, the variances of the directional target source and diffuse noise, $r_{ij}^{(h)}$ and $r_{ij}^{(u)}$, respectively, are simultaneously estimated.

2) Parameter Estimation: We use the EM algorithm to estimate parameters $r_{ij}^{(h)}$, $r_{ij}^{(u)}$, and λ_i by maximum a posteriori estimation. A Q function is defined by the expectation of the complete log-likelihood with respect to the posterior probability of latent variables $s_{ij}^{(h)}$ and $\boldsymbol{u}_{ij}$ as

$$Q(\Theta; \tilde{\Theta}) = \sum_{i,j} \left[-(\alpha+2) \log r_{ij}^{(h)} - \frac{\hat{r}_{ij}^{(h)} + \beta}{r_{ij}^{(h)}} - M \log r_{ij}^{(u)} \right.$$

$$\left. - \log \det \mathbf{R}_i^{(u)} - \frac{\operatorname{tr}((\mathbf{R}_i^{(u)})^{-1} \hat{\mathbf{R}}_{ij}^{(u)})}{r_{ij}^{(u)}} \right] + \text{const.}, \tag{12}$$

where const. includes the constant terms independent of the parameters, $\Theta = \{r_{ij}^{(h)}, r_{ij}^{(u)}, \lambda_i\}$ is the set of parameters to be updated, $\tilde{\Theta} = \{\tilde{r}_{ij}^{(h)}, \tilde{r}_{ij}^{(u)}, \tilde{\lambda}_i\}$ is the set of up-to-date parameters, and $\hat{r}_{ij}^{(h)}$ and $\hat{\mathbf{R}}_{ij}^{(u)}$ are the sufficient statistics obtained by the following E-step update rules:

$$\tilde{\mathbf{R}}_i^{(u)} = \mathbf{R}_i^{\prime(u)} + \tilde{\lambda}_i \boldsymbol{b}_i \boldsymbol{b}_i^{\mathsf{H}} \tag{13}$$

$$\tilde{\mathbf{R}}_{ij}^{(x)} = \tilde{r}_{ij}^{(h)} \boldsymbol{a}_i^{(h)} (\boldsymbol{a}_i^{(h)})^{\mathsf{H}} + \tilde{r}_{ij}^{(u)} \tilde{\mathbf{R}}_i^{(u)}, \tag{14}$$

$$\hat{r}_{ij}^{(h)} = \tilde{r}_{ij}^{(h)} - (\tilde{r}_{ij}^{(h)})^2 (\boldsymbol{a}_i^{(h)})^{\mathsf{H}} (\tilde{\mathbf{R}}_{ij}^{(x)})^{-1} \boldsymbol{a}_i^{(h)}$$

$$+ |\tilde{r}_{ij}^{(h)} \boldsymbol{x}_{ij}^{\mathsf{H}} (\tilde{\mathbf{R}}_{ij}^{(x)})^{-1} \boldsymbol{a}_i^{(h)}|^2, \tag{15}$$

$$\hat{\mathbf{R}}_{ij}^{(u)} = \tilde{r}_{ij}^{(u)} \tilde{\mathbf{R}}_i^{(u)} - (\tilde{r}_{ij}^{(u)})^2 \tilde{\mathbf{R}}_i^{(u)} (\tilde{\mathbf{R}}_{ij}^{(x)})^{-1} \tilde{\mathbf{R}}_i^{(u)}$$

$$+ (\tilde{r}_{ij}^{(u)})^2 \tilde{\mathbf{R}}_i^{(u)} (\tilde{\mathbf{R}}_{ij}^{(x)})^{-1} \boldsymbol{x}_{ij} \boldsymbol{x}_{ij}^{\mathsf{H}} (\tilde{\mathbf{R}}_{ij}^{(x)})^{-1} \tilde{\mathbf{R}}_i^{(u)}. \tag{16}$$

In the M-step, the parameters in Θ are updated as follows:

$$r_{ij}^{(h)} \leftarrow \frac{\hat{r}_{ij}^{(h)} + \beta}{\alpha + 2}, \tag{17}$$

$$\lambda_i \leftarrow \frac{1}{J} \sum_j \frac{1}{\tilde{r}_{ij}^{(u)}} \boldsymbol{b}_i^{\mathsf{H}} \hat{\mathbf{R}}_{ij}^{(u)} \boldsymbol{b}_i, \tag{18}$$

$$\mathbf{R}_i^{(u)} \leftarrow \mathbf{R}_i^{\prime(u)} + \lambda_i \boldsymbol{b}_i \boldsymbol{b}_i^{\mathsf{H}}, \tag{19}$$

$$r_{ij}^{(u)} \leftarrow \frac{1}{M} \mathrm{tr}((\mathbf{R}_i^{(u)})^{-1} \hat{\mathbf{R}}_{ij}^{(u)}). \tag{20}$$

3) Initialization of parameters: We employ ILRMA estimates to initialize the variances $r_{ij}^{(h)}$ and $r_{ij}^{(u)}$ as

$$r_{ij}^{(h)} = \sum_k t_{ik,n_h} v_{kj,n_h}, \tag{21}$$

$$r_{ij}^{(u)} = \frac{1}{M} (\hat{\boldsymbol{y}}_{ij}^{(u)})^{\mathsf{H}} (\mathbf{R}_i^{\prime(u)})^+ \hat{\boldsymbol{y}}_{ij}^{(u)}, \tag{22}$$

where t_{ik,n_h} and v_{kj,n_h} are the NMF parameters in the low-rank source model obtained by ILRMA, $k = 1, \ldots, K$ is the index of the NMF bases, K is the number of NMF bases, $^+$ denotes the pseudoinverse, and $\hat{\boldsymbol{y}}_{ij}^{(u)}$ is the scale-fixed source image of diffuse noise calculated as

$$\hat{\boldsymbol{y}}_{ij}^{(u)} = \mathbf{W}_i^{-1} (\boldsymbol{w}_{i,1}^{\mathsf{H}} \boldsymbol{x}_{ij}, \ldots, \boldsymbol{w}_{i,n_h-1}^{\mathsf{H}} \boldsymbol{x}_{ij}, 0,$$
$$\boldsymbol{w}_{i,n_h+1}^{\mathsf{H}} \boldsymbol{x}_{ij}, \ldots, \boldsymbol{w}_{i,M}^{\mathsf{H}} \boldsymbol{x}_{ij})^{\mathsf{T}}. \tag{23}$$

The parameter λ_i is initialized by the minimum nonzero eigenvalue of $\mathbf{R}_i^{\prime(u)}$.

III. PROPOSED METHOD

A. Motivation

Update rules (15) and (16) involve an inverse matrix operation of $M \times M$ matrices at each time-frequency slot. Thus, their computational complexity is $O(IJM^3)$. Such a heavy computational load restricts their implementation on low-resource hardware, such as hearing-aid devices. To avoid this problem, we propose an efficient update algorithm that greatly accelerates the estimation of parameters by expanding matrix inversion. The acceleration consists of two steps: (i) expanding matrix inversion using the Sherman–Morrison formula and (ii) expanding matrix inversion using the pseudoinverse of matrices.

B. Key Concept for Acceleration

1) First-stage acceleration: Using the Sherman–Morrison formula, we can expand the inverse of matrix (14) as follows:

$$(\tilde{\mathbf{R}}_{ij}^{(x)})^{-1}$$

$$= \frac{1}{\tilde{r}_{ij}^{(u)}} (\tilde{\mathbf{R}}_i^{(u)})^{-1} - \frac{\frac{\tilde{r}_{ij}^{(h)}}{(\tilde{r}_{ij}^{(u)})^2} (\tilde{\mathbf{R}}_i^{(u)})^{-1} \boldsymbol{a}_i^{(h)} (\boldsymbol{a}_i^{(h)})^{\mathsf{H}} (\tilde{\mathbf{R}}_i^{(u)})^{-1}}{1 + \frac{\tilde{r}_{ij}^{(h)}}{\tilde{r}_{ij}^{(u)}} (\boldsymbol{a}_i^{(h)})^{\mathsf{H}} (\tilde{\mathbf{R}}_i^{(u)})^{-1} \boldsymbol{a}_i^{(h)}}$$
$$\tag{24}$$

$$= \frac{1}{\tilde{r}_{ij}^{(u)}} \left((\tilde{\mathbf{R}}_i^{(u)})^{-1} - \frac{\tilde{r}_{ij}^{(h)}}{\tilde{r}_{ij}^{(u)} + \tilde{r}_{ij}^{(h)} (\boldsymbol{a}_i^{(h)})^{\mathsf{H}} (\tilde{\mathbf{R}}_i^{(u)})^{-1} \boldsymbol{a}_i^{(h)}} \right.$$
$$\left. \cdot (\tilde{\mathbf{R}}_i^{(u)})^{-1} \boldsymbol{a}_i^{(h)} (\boldsymbol{a}_i^{(h)})^{\mathsf{H}} (\tilde{\mathbf{R}}_i^{(u)})^{-1} \right). \tag{25}$$

Note that $\tilde{\mathbf{R}}_i^{(u)} = \mathbf{R}_i^{\prime(u)} + \tilde{\lambda}_i \boldsymbol{b}_i \boldsymbol{b}_i^{\mathsf{H}}$ is invertible. This expansion enables us to reduce the computational complexity of (15) and (16) from $O(IJM^3)$ to $O(IM^3 + IJM^2)$, where the two terms respectively correspond to matrix inversion at each frequency bin and multiplication of a matrix and vector at each time-frequency slot.

2) Second-stage acceleration: Using $\mathbf{R}_i^{\prime(u)} \boldsymbol{b}_i = \mathbf{0}$ and $\|\boldsymbol{b}_i\|_2 = 1$, we can expand the inversion $(\mathbf{R}_i^{(u)})^{-1} = (\mathbf{R}_i^{\prime(u)} + \lambda_i \boldsymbol{b}_i \boldsymbol{b}_i^{\mathsf{H}})^{-1}$ using the pseudoinverse of $\mathbf{R}_i^{\prime(u)}$ as

$$(\mathbf{R}_i^{(u)})^{-1} = (\mathbf{R}_i^{\prime(u)})^+ + \frac{1}{\lambda_i} \boldsymbol{b}_i \boldsymbol{b}_i^{\mathsf{H}}. \tag{26}$$

Since $\mathbf{R}_i^{\prime(u)}$ is fixed in rank-constrained SCM estimation, neither matrix inversion nor pseudoinversion is necessary in the parameter update step if the pseudoinverse $(\mathbf{R}_i^{\prime(u)})^+$ is calculated in advance. Hence, the number of matrix inversions is reduced from I to zero with expression (26). Furthermore, no multiplication involving a matrix or vector is required, as shown in Sect. III-C2.

C. Accelerated Update Rule

1) First-stage acceleration: For the sake of simplicity, we define the following variables:

$$\rho_i^{(aa)} := (\boldsymbol{a}_i^{(h)})^{\mathsf{H}} (\mathbf{R}_i^{(u)})^{-1} \boldsymbol{a}_i^{(h)}, \tag{27}$$

$$\tilde{\rho}_i^{(aa)} := (\boldsymbol{a}_i^{(h)})^{\mathsf{H}} (\tilde{\mathbf{R}}_i^{(u)})^{-1} \boldsymbol{a}_i^{(h)}, \tag{28}$$

$$\rho_{ij}^{(ax)} := (\boldsymbol{a}_i^{(h)})^{\mathsf{H}} (\mathbf{R}_i^{(u)})^{-1} \boldsymbol{x}_{ij}, \tag{29}$$

$$\tilde{\rho}_{ij}^{(ax)} := (\boldsymbol{a}_i^{(h)})^{\mathsf{H}} (\tilde{\mathbf{R}}_i^{(u)})^{-1} \boldsymbol{x}_{ij}, \tag{30}$$

$$\rho_{ij}^{(xx)} := \boldsymbol{x}_{ij}^{\mathsf{H}} (\mathbf{R}_i^{(u)})^{-1} \boldsymbol{x}_{ij}, \tag{31}$$

$$\sigma_i^{(ab)} := (\boldsymbol{a}_i^{(h)})^{\mathsf{H}} \boldsymbol{b}_i, \tag{32}$$

$$\sigma_{ij}^{(bx)} := \boldsymbol{b}_i^{\mathsf{H}} \boldsymbol{x}_{ij}, \tag{33}$$

$$\tilde{\gamma}_{ij} := \frac{\tilde{r}_{ij}^{(h)}}{\tilde{r}_{ij}^{(u)} + \tilde{r}_{ij}^{(h)} \tilde{\rho}_i^{(aa)}}. \tag{34}$$

From (25), we obtain

$$(\boldsymbol{a}_i^{(h)})^{\mathsf{H}} (\tilde{\mathbf{R}}_{ij}^{(x)})^{-1} \boldsymbol{a}_i^{(h)} = \frac{1}{\tilde{r}_{ij}^{(u)}} \left(\tilde{\rho}_i^{(aa)} - \tilde{\gamma}_{ij} (\tilde{\rho}_i^{(aa)})^2 \right) \tag{35}$$

$$= \frac{\tilde{\rho}_i^{(aa)}}{\tilde{r}_{ij}^{(u)} + \tilde{r}_{ij}^{(h)} \tilde{\rho}_i^{(aa)}}, \tag{36}$$

$$(\boldsymbol{a}_i^{(h)})^{\mathsf{H}} (\tilde{\mathbf{R}}_{ij}^{(x)})^{-1} \boldsymbol{x}_{ij} = \frac{\tilde{\rho}_{ij}^{(ax)}}{\tilde{r}_{ij}^{(u)} + \tilde{r}_{ij}^{(h)} \tilde{\rho}_i^{(aa)}}. \tag{37}$$

The update rule (15) in the E-step can be rewritten as

$$\hat{r}_{ij}^{(h)} = \tilde{\gamma}_{ij} (\tilde{r}_{ij}^{(u)} + \tilde{\gamma}_{ij} |\tilde{\rho}_{ij}^{(ax)}|^2). \tag{38}$$

Hence, we update $r_{ij}^{(h)}$ in the M-step as

$$r_{ij}^{(h)} \leftarrow \frac{\tilde{\gamma}_{ij} \left(\tilde{r}_{ij}^{(u)} + \tilde{\gamma}_{ij} |\tilde{\rho}_{ij}^{(ax)}|^2 \right) + \beta}{\alpha + 2}. \tag{39}$$

In addition, for update rule (18) in the M-step, we can utilize the following equations, which are obtained using (25):

$$\tilde{\mathbf{R}}_i^{(u)}(\tilde{\mathbf{R}}_{ij}^{(x)})^{-1}$$
$$= \frac{1}{\tilde{r}_{ij}^{(u)}}\left(\mathbf{I} - \tilde{\gamma}_{ij}\boldsymbol{a}_i^{(h)}(\boldsymbol{a}_i^{(h)})^{\mathsf{H}}(\tilde{\mathbf{R}}_i^{(u)})^{-1}\right), \qquad (40)$$

$$\boldsymbol{b}_i^{\mathsf{H}}\tilde{\mathbf{R}}_i^{(u)}(\tilde{\mathbf{R}}_{ij}^{(x)})^{-1}\tilde{\mathbf{R}}_i^{(u)}\boldsymbol{b}_i$$
$$= \frac{1}{\tilde{r}_{ij}^{(u)}}\left(\tilde{\lambda}_i - \tilde{\gamma}_{ij}|\sigma_i^{(ab)}|^2\right), \qquad (41)$$

$$\boldsymbol{b}_i^{\mathsf{H}}\tilde{\mathbf{R}}_i^{(u)}(\tilde{\mathbf{R}}_{ij}^{(x)})^{-1}\boldsymbol{x}_{ij}$$
$$= \frac{1}{\tilde{r}_{ij}^{(u)}}\left(\sigma_{ij}^{(bx)} - \tilde{\gamma}_{ij}\overline{\sigma_i^{(ab)}}\tilde{\rho}_{ij}^{(ax)}\right), \qquad (42)$$

where $\bar{\cdot}$ denotes the complex conjugate. Using (40)–(42), we obtain

$$\lambda_i \leftarrow \frac{1}{J}\sum_j\left(\tilde{\gamma}_{ij}|\sigma_i^{(ab)}|^2 + \frac{1}{\tilde{r}_{ij}^{(u)}}\left|\sigma_{ij}^{(bx)} - \tilde{\gamma}_{ij}\overline{\sigma_i^{(ab)}}\tilde{\rho}_{ij}^{(ax)}\right|^2\right). \qquad (43)$$

Moreover, for update rule (20) in the M-step, the following equation can be used to derive an efficient update rule:

$$\hat{\mathbf{R}}_{ij}^{(u)} = \tilde{\gamma}_{ij}(\tilde{r}_{ij}^{(u)} + \tilde{\gamma}_{ij}|\tilde{\rho}_{ij}^{(ax)}|^2)\boldsymbol{a}_i^{(h)}(\boldsymbol{a}_i^{(h)})^{\mathsf{H}} + \boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^{\mathsf{H}}$$
$$- \tilde{\gamma}_{ij}\tilde{\rho}_{ij}^{(ax)}\boldsymbol{a}_i^{(h)}\boldsymbol{x}_{ij}^{\mathsf{H}} - \tilde{\gamma}_{ij}\overline{\tilde{\rho}_{ij}^{(ax)}}\boldsymbol{x}_{ij}(\boldsymbol{a}_i^{(h)})^{\mathsf{H}}. \qquad (44)$$

The parameter $r_{ij}^{(u)}$ can be updated as

$$r_{ij}^{(u)} \leftarrow \frac{1}{M}\left(\tilde{\gamma}_{ij}\rho_i^{(aa)}(\tilde{r}_{ij}^{(u)} + \tilde{\gamma}_{ij}|\tilde{\rho}_{ij}^{(ax)}|^2)\right.$$
$$\left. + \rho_{ij}^{(xx)} - 2\tilde{\gamma}_{ij}\mathrm{Re}[\tilde{\rho}_{ij}^{(ax)}\overline{\rho_{ij}^{(ax)}}]\right). \qquad (45)$$

The update algorithm for the first-stage acceleration is summarized in Algorithm 1. Note that we must update $\rho_i^{(aa)}$, $\rho_{ij}^{(ax)}$, and $\rho_{ij}^{(xx)}$ using up-to-date $\mathbf{R}_i^{(u)}$. In summary, in the first-stage acceleration, only the matrix inversions of $\mathbf{R}_i^{(u)}$ and $\tilde{\mathbf{R}}_i^{(u)}$ are necessary for each frequency bin, requiring a calculation cost of $O(IM^3)$.

2) Second-stage acceleration: On the basis of (26), we can further accelerate the update algorithm. We define the quadratic terms

$$\tau_i^{(aa)} := (\boldsymbol{a}_i^{(h)})^{\mathsf{H}}(\mathbf{R}_i'^{(u)})^+\boldsymbol{a}_i^{(h)}, \qquad (46)$$
$$\tau_{ij}^{(ax)} := (\boldsymbol{a}_i^{(h)})^{\mathsf{H}}(\mathbf{R}_i'^{(u)})^+\boldsymbol{x}_{ij}, \qquad (47)$$
$$\tau_{ij}^{(xx)} := \boldsymbol{x}_{ij}^{\mathsf{H}}(\mathbf{R}_i'^{(u)})^+\boldsymbol{x}_{ij}. \qquad (48)$$

These terms do not depend on the variables $r_{ij}^{(h)}$, $r_{ij}^{(u)}$, and λ_i and can be calculated before the update iteration. Matrix inversions appearing in $\rho_i^{(aa)}$, $\tilde{\rho}_i^{(aa)}$, $\rho_{ij}^{(ax)}$, $\tilde{\rho}_{ij}^{(ax)}$, and $\rho_{ij}^{(xx)}$

Algorithm 1: Algorithm for the first-stage acceleration update rule of the rank-constrained SCM estimation

1 Run ILRMA and calculate $\boldsymbol{a}_i^{(h)}$, $\mathbf{R}_i'^{(u)}$, and $\boldsymbol{b}_i$;
2 Initialize $r_{ij}^{(h)}$, $r_{ij}^{(u)}$ and λ_i as described in Sect. II-B3;
3 Calculate $\sigma_i^{(ab)}$ and $\sigma_{ij}^{(bx)}$ by (32) and (33) for all i and j;
4 Calculate $\rho_i^{(aa)}$ and $\rho_{ij}^{(ax)}$ by (27) and (29) for all i and j;
5 **repeat**
6 $\quad$ $\tilde{r}_{ij}^{(h)} \leftarrow r_{ij}^{(h)}$ for all i and j;
7 $\quad$ $\tilde{r}_{ij}^{(u)} \leftarrow r_{ij}^{(u)}$ for all i and j;
8 $\quad$ $\tilde{\lambda}_i \leftarrow \lambda_i$ for all i;
9 $\quad$ $\tilde{\rho}_i^{(aa)} \leftarrow \rho_i^{(aa)}$ for all i;
10 $\quad$ $\tilde{\rho}_{ij}^{(ax)} \leftarrow \rho_{ij}^{(ax)}$ for all i and j;
11 $\quad$ Calculate $\tilde{\gamma}_{ij}$ by (34) for all i and j;
12 $\quad$ Update $r_{ij}^{(h)}$ by (39) for all i and j;
13 $\quad$ Update λ_i by (43) for all i;
14 $\quad$ Calculate $\rho_i^{(aa)}$, $\rho_{ij}^{(ax)}$, and $\rho_{ij}^{(xx)}$ by (27), (29), and (31) for all i and j;
15 $\quad$ Update $r_{ij}^{(u)}$ by (45) for all i and j;
16 **until** *converge*;

---

can be transformed using (26) as follows:

$$\rho_i^{(aa)} = \tau_i^{(aa)} + \frac{|\sigma_i^{(ab)}|^2}{\lambda_i}, \qquad (49)$$

$$\tilde{\rho}_i^{(aa)} = \tau_i^{(aa)} + \frac{|\sigma_i^{(ab)}|^2}{\tilde{\lambda}_i}, \qquad (50)$$

$$\rho_{ij}^{(ax)} = \tau_{ij}^{(ax)} + \frac{\sigma_i^{(ab)}\sigma_{ij}^{(bx)}}{\lambda_i}, \qquad (51)$$

$$\tilde{\rho}_{ij}^{(ax)} = \tau_{ij}^{(ax)} + \frac{\sigma_i^{(ab)}\sigma_{ij}^{(bx)}}{\tilde{\lambda}_i}, \qquad (52)$$

$$\rho_{ij}^{(xx)} = \tau_{ij}^{(xx)} + \frac{|\sigma_{ij}^{(bx)}|^2}{\lambda_i}. \qquad (53)$$

Thus, the update rules with the second-stage acceleration consist of only scalar operations and can be obtained as follows:

$$r_{ij}^{(h)} \leftarrow \frac{\tilde{\gamma}_{ij}\left(\tilde{r}_{ij}^{(u)} + \tilde{\gamma}_{ij}\left|\tau_{ij}^{(ax)} + \frac{\sigma_i^{(ab)}\sigma_{ij}^{(bx)}}{\tilde{\lambda}_i}\right|^2\right) + \beta}{\alpha + 2}, \qquad (54)$$

$$\lambda_i \leftarrow \frac{1}{J}\sum_j\left(\tilde{\gamma}_{ij}|\sigma_i^{(ab)}|^2\right.$$
$$\left. + \frac{1}{\tilde{r}_{ij}^{(u)}}\left|\sigma_{ij}^{(bx)} - \tilde{\gamma}_{ij}\overline{\sigma_i^{(ab)}}\left(\tau_{ij}^{(ax)} + \frac{\sigma_i^{(ab)}\sigma_{ij}^{(bx)}}{\tilde{\lambda}_i}\right)\right|^2\right), \qquad (55)$$

Algorithm 2: Algorithm for second-stage acceleration update rule of the rank-constrained SCM estimation

---

1 Run ILRMA and calculate $a_i^{(h)}$, $\mathbf{R}_i'^{(u)}$, and b_i;

2 Initialize $r_{ij}^{(h)}$, $r_{ij}^{(u)}$ and λ_i as described in Sect. II-B3;

3 Calculate $\sigma_i^{(ab)}$ and $\sigma_{ij}^{(bx)}$ by (32) and (33) for all i and j;

4 Calculate $\tau_i^{(aa)}$, $\tau_{ij}^{(ax)}$, and $\tau_{ij}^{(xx)}$ by (46), (47), and (48) for all i and j;

5 repeat

6 $\tilde{r}_{ij}^{(h)} \leftarrow r_{ij}^{(h)}$ for all i and j;

7 $\tilde{r}_{ij}^{(u)} \leftarrow r_{ij}^{(u)}$ for all i and j;

8 $\tilde{\lambda}_i \leftarrow \lambda_i$ for all i;

9   Calculate $\tilde{\gamma}_{ij}$ by (34) and (50) for all i and j;

10   Update $r_{ij}^{(h)}$ by (39) for all i and j;

11   Update λ_i by (43) for all i;

12   Update $r_{ij}^{(u)}$ by (45) for all i and j;

13 until *converge*;

$$
\begin{aligned}
r_{ij}^{(u)} \leftarrow \frac{1}{M} \Bigg(& \tilde{\gamma}_{ij} \left(\tau_i^{(aa)} + \frac{|\sigma_i^{(ab)}|^2}{\lambda_i} \right) \\
& \cdot \left(\tilde{r}_{ij}^{(u)} + \tilde{\gamma}_{ij} \left| \tau_{ij}^{(ax)} + \frac{\sigma_i^{(ab)}\sigma_{ij}^{(bx)}}{\tilde{\lambda}_i} \right|^2 \right) + \tau_{ij}^{(xx)} \\
& + \frac{|\sigma_{ij}^{(bx)}|^2}{\lambda_i} - 2\tilde{\gamma}_{ij}\mathrm{Re}\left[\left(\tau_{ij}^{(ax)} + \frac{\sigma_i^{(ab)}\sigma_{ij}^{(bx)}}{\tilde{\lambda}_i} \right) \right. \\
& \left. \cdot \left(\overline{\tau_{ij}^{(ax)}} + \overline{\frac{\sigma_i^{(ab)}\sigma_{ij}^{(bx)}}{\lambda_i}} \right) \right] \Bigg).
\end{aligned}
\tag{56}
$$

The update algorithm for the second-stage acceleration is summarized in Algorithm 2.

D. Advantage of Proposed Accelerated Update Rules

In general, the complexity of size-M matrix inversions and matrix multiplications is $O(M^3)$, and that of multiplications of a matrix and a vector is $O(M^2)$. The iteration-wise computational complexities of the naive update rule and the proposed update rules with the first-stage and second-stage accelerations are summarized in Table I. Note that initialization with (21) and (22) is required for all the methods. The proposed algorithms can reduce the complexity via the use of (25) and (26). In particular, the second-stage acceleration has greatly improved efficiency because its computational cost does not depend on the number of microphones M, thus enabling us to apply BSS with a large-scale microphone array.

For example, $I = 513$, $J = 275$, and $M = 4$ for the conditions described in Sect. IV-A. In such a case, the naive update rule requires $IJ = 141075$ inverse matrix operations per iteration, which is no longer necessary for the second-stage acceleration.

TABLE I
COMPUTATIONAL COMPLEXITY OF INITIALIZATION
AND ITERATIVE UPDATE FOR EACH METHOD

Method	Initialization	Iterative update
Naive	$O(IM^3 + IJM^2)$	$O(IJM^3)$
First-stage accel.	$O(IM^3 + IJM^2)$	$O(IM^3 + IJM^2)$
Second-stage accel.	$O(IM^3 + IJM^2)$	$O(IJ)$

IV. EXPERIMENTS

A. Experimental Conditions

To compare the efficacy of the proposed algorithms and the separation quality, we conducted an audio BSS experiment with simulated mixtures of a directional target speech and diffuse babble noise. We compared three methods: FastMNMF [19], FastMNMF initialized by ILRMA (ILRMA+FastMNMF), and rank-constrained SCM estimation initialized by ILRMA [20]. For rank-constrained SCM estimation, we compared three update algorithms, namely, the naive update rule (*Naive*) and the proposed update rules with the first and second accelerations (*Proposed 1st-stage accel.* and *Proposed 2nd-stage accel.*, respectively). In ILRMA, the observation x_{ij} was preprocessed via a sphering transformation using principal component analysis. For ILRMA and FastMNMF, all the NMF variables were initialized by nonnegative random values. The demixing matrix $\mathbf{W}_i$ in ILRMA and the spatial covariance matrix in FastMNMF were initialized by the identity matrix. For ILRMA+FastMNMF, the NMF variables were initialized using the estimates of ILRMA. Also, the spatial covariance matrix was initialized using $a_{i,n}a_{i,n}^{\mathsf{H}} + \varepsilon \sum_{n' \neq n} a_{i,n'}a_{i,n'}^{\mathsf{H}}$ for ILRMA+FastMNMF, where $a_{i,n}$ was estimated by ILRMA and ε was set to 10^{-5}. These methods were implemented in MATLAB (R2019a), and the computation was performed on an Intel Core i9-7900X (3.30 GHz, 10 cores) CPU. The dry sources of the directional target speech and diffuse babble noise were obtained from the JNAS speech corpus [23]. They were convoluted with the impulse responses shown in Fig. 1 to simulate a mixture of 8.7 s length. The target source was located 30° clockwise from the normal to a microphone array, the 19 loudspeakers used to simulate diffuse noise were arranged at intervals of 10°, the size of the recording room for these impulse responses was 3.9 m × 3.9 m, and its reverberation time was about 200 ms. Note that the diffuse babble noise was produced by convoluting 19 independent speeches with each impulse response. The speech-to-noise ratio was set to 0 dB. The other conditions are shown in Table II.

B. Results

We evaluated each method using the source-to-distortion ratio (SDR) [24] of the directional target speech, which measures the separation quality and the absence of distortion. For each of the methods, the SDR behaviors with respect to the elapsed time are shown in Fig. 2. Note that all methods except for FastMNMF are initialized using ILRMA. Multichannel Wiener filtering using ILRMA estimates improves the SDR

TABLE II
EXPERIMENTAL CONDITIONS

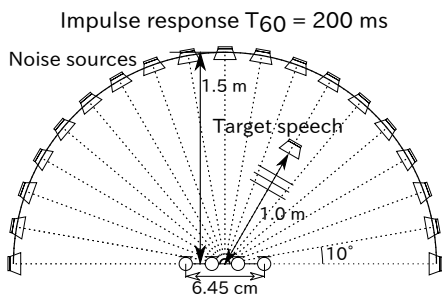| Sampling frequency	16 kHz
STFT	64-ms-long Hamming window with 32 ms shift
Number of NMF bases K	10 for each source
Number of iterations in ILRMA	50
Maximum number of iterations in methods except ILRMA	200
Shape and scale parameters	$\alpha = 1.1$ and $\beta = 10^{-16}$

Impulse response T_{60} = 200 ms

Fig. 1. Recording conditions of impulse responses.

value to some extent, and FastMNMF or rank-constrained SCM estimation further improves the separation quality. From Fig. 2, we can confirm that the second-stage acceleration achieves the fastest and most efficient target speech extraction.

Fig. 3 shows the average computational time of one iteration for each method. The proposed algorithm with the second-stage acceleration was 87 times faster than the naive update rule and 12 times faster than FastMNMF with the four-microphone condition.

V. CONCLUSION

In this paper, we presented new accelerated update rules for rank-constrained SCM estimation, enabling effective extraction of a directional target source contaminated by diffuse noise. We derived update rules by expanding matrix inversion in the naive update rule. The experiment showed that the proposed method achieved an 87 times faster update than the naive update rule and a 12 times faster update than FastMNMF.

REFERENCES

[1] H. Sawada, N. Ono, H. Kameoka, D. Kitamura, and H. Saruwatari, "A review of blind source separation methods: two converging routes to ILRMA originating from ICA and NMF," *APSIPA Trans. Signal and Inf. Process.*, vol. 8, no. e12, pp. 1–14, 2019.
[2] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol. 22, no. 1, pp. 21–34, 1998.
[3] S. Araki, R. Mukai, S. Makino, T. Nishikawa, and H. Saruwatari, "The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech," *IEEE Trans. ASP*, vol. 11, no. 2, pp. 109–116, 2003.
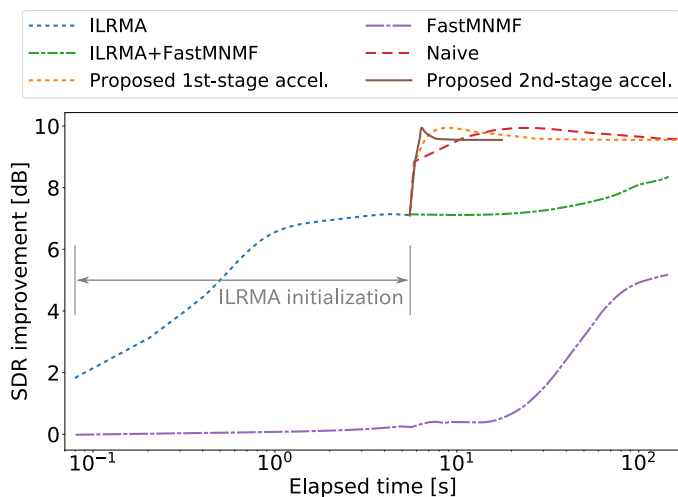

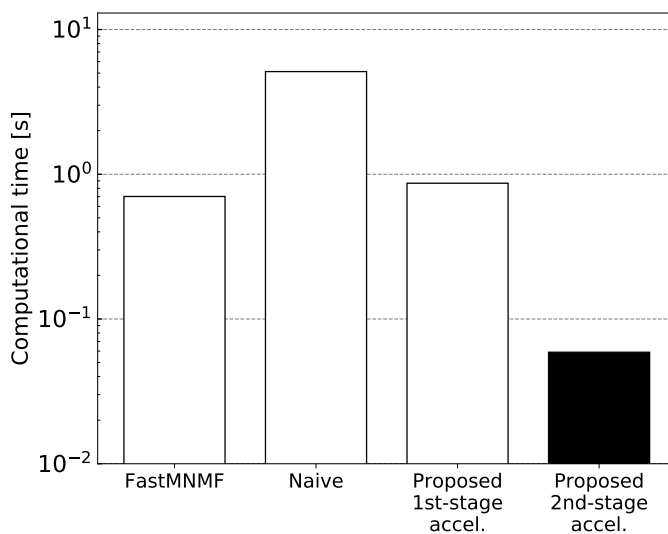Fig. 2. SDR behaviors with respect to elapsed time.

Fig. 3. Average computational time of one iteration for each method.

[4] H. Saruwatari, T. Kawamura, T. Nishikawa, A. Lee, and K. Shikano, "Blind source separation based on a fast-convergence algorithm combining ICA and beamforming," *IEEE Trans. ASLP*, vol. 14, no. 2, pp. 666–678, 2006.
[5] A. Hiroe, "Solution of permutation problem in frequency domain ICA using multivariate probability density functions," in *Proc. ICA*, 2006, pp. 601–608.
[6] T. Kim, H. T. Attias, S.-Y. Lee, and T.-W. Lee, "Blind source separation exploiting higher-order frequency dependencies," *IEEE Trans. ASLP*, vol. 15, no. 1, pp. 70–79, 2007.
[7] N. Ono, "Stable and fast update rules for independent vector analysis based on auxiliary function technique," in *Proc. WASPAA*, 2011, pp. 189–192.
[8] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, "Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization," *IEEE/ACM Trans. ASLP*, vol. 24, no. 9, pp. 1626–1641, 2016.
[9] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, "Determined blind source separation with independent low-rank matrix analysis," in *Audio Source Separation*, S. Makino, Ed., pp. 125–155. Springer, Cham, 2018.
[10] D. Kitamura, S. Mogami, Y. Mitsui, N. Takamune, H. Saruwatari, N. Ono, Y. Takahashi, and K. Kondo, "Generalized independent low-rank matrix analysis using heavy-tailed distributions for blind source

separation," *EURASIP JASP*, vol. 2018, no. 1, pp. 1–28, 2018.

[11] S. Mogami, N. Takamune, D. Kitamura, H. Saruwatari, Y. Takahashi, K. Kondo, H. Nakajima, and N. Ono, "Independent low-rank matrix analysis based on time-variant sub-Gaussian source model," in *Proc. APSIPA*, 2018, pp. 1684–1691.

[12] S. Mogami, Y. Mitsui, N. Takamune, D. Kitamura, H. Saruwatari, Y. Takahashi, K. Kondo, H. Nakajima, and H. Kameoka, "Independent low-rank matrix analysis based on generalized Kullback-Leibler divergence," *IEICE Trans. Fundamentals*, vol. E102-A, no. 2, pp. 458–463, 2019.

[13] S. Araki, S. Makino, Y. Hinamoto, R. Mukai, T. Nishikawa, and H. Saruwatari, "Equivalence between frequency-domain blind source separation and frequency-domain adaptive beamforming for convolutive mixtures," *EURASIP JASP*, vol. 2003, no. 11, pp. 1–10, 2003.

[14] A. Ozerov and C. Févotte, "Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation," *IEEE Trans. ASLP*, vol. 18, no. 3, pp. 550–563, 2010.

[15] H. Sawada, H. Kameoka, S. Araki, and N. Ueda, "Multichannel extensions of non-negative matrix factorization with complex-valued data," *IEEE Trans. ASLP*, vol. 21, no. 5, pp. 971–982, 2013.

[16] J. Nikunen and T. Virtanen, "Direction of arrival based spatial covariance model for blind sound source separation," *IEEE Trans. ASLP*, vol. 22, no. 3, pp. 727–739, 2014.

[17] N. Q. K. Duong, E. Vincent, and R. Gribonval, "Under-determined reverberant audio source separation using a full-rank spatial covariance model," *IEEE Trans. ASLP*, vol. 18, no. 7, pp. 1830–1840, 2010.

[18] N. Ito and T. Nakatani, "FastMNMF: Joint diagonalization based accelerated algorithms for multichannel nonnegative matrix factorization," in *Proc. ICASSP*, 2019, pp. 371–375.

[19] K. Sekiguchi, A. A. Nugraha, Y. Bando, and K. Yoshii, "Fast multichannel source separation based on jointly diagonalizable spatial covariance matrices," *CoRR*, vol. abs/1903.03237, 2019.

[20] Y. Kubo, N. Takamune, D. Kitamura, and H. Saruwatari, "Efficient full-rank spatial covariance estimation using independent low-rank matrix analysis for blind source separation," in *Proc. EUSIPCO*, 2019.

[21] Y. Takahashi, T. Takatani, K. Osako, H. Saruwatari, and K. Shikano, "Blind spatial subtraction array for speech enhancement in noisy environment," *IEEE Trans. ASLP*, vol. 17, no. 4, pp. 650–664, 2009.

[22] N. Murata, S. Ikeda, and A. Ziehe, "An approach to blind source separation based on temporal structure of speech signals," *Neurocomputing*, vol. 41, no. 1–4, pp. 1–24, 2001.

[23] K. Itou, M. Yamamoto, K. Takeda, T. Takezawa, T. Matsuoka, T. Kobayashi, K. Shikano, and S. Itahashi, "JNAS: Japanese speech corpus for large vocabulary continuous speech recognition research," *J. Acoust. Soc. Jpn. (E)*, vol. 20, no. 3, pp. 199–206, 1999.

[24] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," *IEEE Trans. ASLP*, vol. 14, no. 4, pp. 1462–1469, 2006.