Blind Speech Extraction Based on Rank-Constrained Spatial Covariance Matrix Estimation With Multivariate Generalized Gaussian Distribution

Yuki Kubo, *Graduate Student Member, IEEE*, Norihiro Takamune, Daichi Kitamura, *Member, IEEE*, and Hiroshi Saruwatari, *Member, IEEE*

***Abstract*—In this article, we propose a new blind speech extraction (BSE) method that robustly extracts a directional speech from background diffuse noise by combining independent low-rank matrix analysis (ILRMA) and efficient rank-constrained spatial covariance matrix (SCM) estimation. To achieve more accurate BSE than ILRMA, which assumes each source to be a point source (rank-1 spatial model), the proposed method restores the lost spatial basis for the full-rank SCM of diffuse noise. We adopt the multivariate complex generalized Gaussian distribution (GGD) as the statistical generative model to express various types of observed signal. To estimate the model parameters for an arbitrary shape parameter of the multivariate GGD, we derive a new inequality for rank-constrained SCMs. Also, we propose new acceleration methods to accomplish much faster extraction than conventional blind source separation methods. In BSE experiments using simulated and real recorded data, we confirm that the proposed method achieves more accurate and faster speech extraction than conventional methods.**

***Index Terms*—Blind speech extraction, diffuse noise, spatial covariance matrix, multivariate complex generalized Gaussian distribution.**

I. Introduction

BLIND source separation (BSS) [1] is a technique for separating an observed multichannel signal, which is a mixture of multiple sources, into each source without any prior information about the sources or the mixing system. In a determined or overdetermined situation (number of sensors $\geq$ 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]–[13] have been proposed for audio BSS problems. In particular, ILRMA assumes low-rankness for the power spectrogram of each source using nonnegative matrix factorization (NMF) [14], [15] in addition to statistical independence between sources, and achieves efficient and accurate separation [8]. These methods assume a rank-1 spatial model; the frequency-wise acoustic path of each source can be represented by a single time-invariant spatial basis, which is often called a steering vector. Under this assumption, the determined BSS problem reduces to the estimation of a demixing matrix for each frequency. However, the assumption in the rank-1 spatial model becomes invalid in actual situations. For instance, when a target speech source (directional source) and diffuse noise that arrives from all directions are mixed, FDICA, IVA, and ILRMA cannot extract only the target speech in principle [16], and the estimated target speech includes residual diffuse noise. We often call this problem *blind speech extraction* (BSE). To address this problem, for example, blind spatial subtraction array (BSSA) has been proposed [17], where FDICA-based dynamic noise power estimation and spectral subtraction-based postfiltering are combined. In this method, since processing after FDICA is mainly performed in the single-channel domain, the extraction ability is limited, causing considerable speech distortion.

As a method expected to address the above-mentioned problem, multichannel NMF (MNMF) [18]–[20] has been proposed. MNMF is theoretically equivalent to ILRMA except for the mixing model, namely, MNMF employs a full-rank spatial covariance matrix (SCM) [21]. This model can represent not only the acoustic path but also the spatial spread of each source or diffuse noise, while its optimization has a huge computational cost and lacks robustness against the initialization [8]. To accelerate the parameter estimation, FastMNMF has been proposed [22], [23], although its performance still depends on the initial values of parameters. On the other hand, to increase the stability of its performance, ILRMA-based initialization was utilized for MNMF in [8], [24]. However, the improvement is still limited because of the complexity of optimization with a large number of parameters.

In this paper, we propose *rank-constrained SCM estimation* to extract only the directional target speech efficiently and robustly. Although the directional target speech can be expressed using a rank-1 (rank-constrained) SCM, diffuse noise requires a full-rank SCM because of its spatial spread. To achieve robust and computationally efficient extraction in this BSE situation, we propose a new approach partially using the BSS methods such as ILRMA. It utilizes the fact that the demixing filters for

Manuscript received August 7, 2019; revised January 14, 2020 and March 26, 2020; accepted June 7, 2020. Date of current version June 30, 2020. This work was supported in part by SECOM Science and Technology Foundation and JSPS KAKENHI under Grants 17H06101, 19H01116, 19H04131, and 19K20306, and in part by JSPS-CAS Joint Research Program, under Grant JPJSBP120197203. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Sven Erik Nordholm PhD. *(Corresponding author: Yuki Kubo.)*

Yuki Kubo, Norihiro Takamune, and Hiroshi Saruwatari are with the University of Tokyo, Tokyo 113-8654, Japan (e-mail: yuuki.initial.yk@gmail.com; norihiro_takamune@ipc.i.u-tokyo.ac.jp; hiroshi_saruwatari@ipc.i.u-tokyo.ac.jp).

Daichi Kitamura is with the National Institute of Technology, Kagawa College, Kagawa 7618058, Japan (e-mail: kitamura-d@t.kagawa-nct.ac.jp).

Digital Object Identifier 10.1109/TASLP.2020.3003165

diffuse noise can cancel the directional target speech in the BSS methods based on the rank-1 spatial model [17], resulting in the accurate estimation of a rank-$(M-1)$ diffuse noise SCM, where M denotes the number of microphones. We restore one lost spatial basis of diffuse noise by using maximum a posteriori estimation. Furthermore, we model the observed signal using the multivariate complex generalized Gaussian distribution (GGD). In past studies, GGD-based IVA [25]–[27] and GGD-based ILRMA have been reported [10]–[12]. The advantage of the GGD as a super-Gaussian distribution was discussed in [10], [11], [25]–[27] and that of the sub-Gaussian GGD was investigated in [12] in detail. Motivated by these works, we introduce the multivariate GGD into the statistical model of the full-rank SCM, which is the world's first attempt to the best of our knowledge. One of the notable points of the proposed method is the flexibility of its model; by changing the shape parameter of the multivariate GGD to an arbitrary positive value, we can model various types of observed signal with a super- or sub-Gaussian distribution.

Although the number of parameters to be estimated is smaller than that of conventional methods, the naive iterative algorithm of the rank-constrained SCM estimation requires inversions of matrices of order M at each time-frequency slot. This time-consuming operation becomes a bottleneck of the entire algorithm. The heavy computational load induced by such an operation restricts its implementation on low-resource hardware, such as hearing-aid devices and smartphones. In this paper, we also propose acceleration methods by making use of the expansion of matrix inversion. Although FastMNMF realizes acceleration by introducing an approximation on SCMs, the proposed accelerated update rules are analytically identical to the naive update rule. Thus, we can achieve much faster updates than by other conventional methods including FastMNMF while retaining its high-quality extraction.

The rest of this paper is organized as follows. Section II shows the formulation of BSS and outlines ILRMA, MNMF, and FastMNMF as conventional methods. In Section III, we propose framework of the rank-constrained SCM estimation and its parameter estimation method. In Section IV, accelerated algorithms for rank-constrained SCM estimation are described in detail. BSE experiments with simulated and real recorded data show the efficacy of rank-constrained SCM estimation in Sections V and VI. The conclusions of this paper are presented in Section VII. Note that this paper is partially based on international conference papers [28], [29] written by the authors. Additional contributions of this paper are that we generalize the statistical model using the multivariate GGD, employ the majorization-minimization (MM) [30] and majorization-equalization (ME) [31] algorithms for better parameter optimization, and conduct BSE experiments under extended acoustic conditions including real recorded data. It is worth mentioning that we cannot directly apply the method proposed in [28] to the multivariate GGD model because the method is based on the simple expectation-maximization (EM) algorithm. In this paper, we propose the MM and ME algorithms to address the GGD-related optimization problem that cannot be solved by the EM algorithm.

II. Conventional Methods

A. Definitions

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

B. ILRMA

As a BSS method for determined or overdetermined ($M \geq N$) situations, ILRMA [8], [9] assumes that each source is a point source and the reverberation time is sufficiently shorter than the window length in an STFT. Under these assumptions, 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 observed signal is approximated as

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

where $\boldsymbol{a}_{i,n}$ is the steering vector of source n at frequency i. After dimensionality reduction is used so that $M = N$, if $\mathbf{A}_i$ is invertible, the separated signal $\boldsymbol{y}_{ij} = (y_{ij,1}, \ldots, y_{ij,N})^{\mathsf{T}} \in \mathbb{C}^N$ can be obtained by estimating the demixing matrix $\mathbf{W}_i = (\boldsymbol{w}_{i,1} \ldots \boldsymbol{w}_{i,N})^{\mathsf{H}} = \mathbf{A}_i^{-1}$ as

$$\boldsymbol{y}_{ij} = \mathbf{W}_i \boldsymbol{x}_{ij}, \tag{2}$$

where $^{\mathsf{H}}$ denotes the Hermitian transpose.

In ILRMA, as the generative model of source signals, the following complex Gaussian distribution is assumed:

$$s_{ij,n} \sim \mathcal{N}_c\left(0, r_{ij,n}\right), \tag{3}$$

where $r_{ij,n}$ is the time-frequency-varying variance (power spectrogram model of $s_{ij,n}$). Also, $r_{ij,n}$ is modeled by NMF [32] as $r_{ij,n} = \sum_k t_{ik,n} v_{kj,n}$, where $t_{ik,n} \geq 0$ and $v_{kj,n} \geq 0$ are the NMF variables, $k = 1, \ldots, K$ is the index of the NMF bases, and K is the number of bases. Combining (1) and (3), the generative model of the observed signal becomes

$$\boldsymbol{x}_{ij} \sim \mathcal{N}_{\mathrm{c}}\left(\mathbf{0}, \sum_n r_{ij,n} \boldsymbol{a}_{i,n} \boldsymbol{a}_{i,n}^{\mathsf{H}}\right). \tag{4}$$

From the viewpoint of SCMs, each source has a rank-1 SCM expressed as $\boldsymbol{a}_{i,n}\boldsymbol{a}_{i,n}^{\mathsf{H}}$, whereby (1) is called a rank-1 spatial model.

The cost function in ILRMA is defined as the negative log-likelihood function of (4) as

$$\begin{aligned}
&\mathcal{L}(\{\mathbf{W}_i, t_{ik,n}, v_{kj,n}\}) \\
&\quad = \sum_{i,j,n} \left(\frac{|y_{ij,n}|^2}{\sum_k t_{ik,n} v_{kj,n}} + \log \sum_k t_{ik,n} v_{kj,n} \right) \\
&\quad = -2J \sum_i \log |\det \mathbf{W}_i| + \text{const.},
\end{aligned} \tag{5}$$

where $y_{ij,n} = \boldsymbol{w}_{i,n}^{\mathsf{H}} \boldsymbol{x}_{ij}$, $\{\mathbf{W}_i, t_{ik,n}, v_{kj,n}\}$ is the set of objective variables, and const. includes constant terms that do not depend on objective variables (we use this notation throughout the paper). The separation filter $\boldsymbol{w}_{i,n}$ and the NMF variables $t_{ik,n}$ and $v_{kj,n}$ can be estimated in the maximum likelihood sense (minimization of (5)) by iterating the following update rules [8]. For the separation filter, the method called iterative projection [7] is utilized, which has the update rule

$$\mathbf{G}_{i,n} = \frac{1}{J} \sum_j \frac{1}{r_{ij,n}} \boldsymbol{x}_{ij} \boldsymbol{x}_{ij}^{\mathsf{H}}, \tag{6}$$

$$\boldsymbol{w}_{i,n} \leftarrow (\mathbf{W}_i \mathbf{G}_{i,n})^{-1} \boldsymbol{e}_n, \tag{7}$$

$$\boldsymbol{w}_{i,n} \leftarrow \boldsymbol{w}_{i,n} (\boldsymbol{w}_{i,n}^{\mathsf{H}} \mathbf{G}_{i,n} \boldsymbol{w}_{i,n})^{-\frac{1}{2}}, \tag{8}$$

where $\boldsymbol{e}_n$ denotes the nth column vector of the $M \times M$ identity matrix. For the NMF variables, we minimize the Itakura–Saito divergence between $|y_{ij,n}|^2$ and $\sum_k t_{ik,n} v_{kj,n}$, which yields the update rule

$$t_{ik,n} \leftarrow t_{ik,n} \sqrt{\frac{\sum_j \frac{|y_{ij,n}|^2}{(\sum_{k'} t_{ik,'n} v_{k'j,n})^2} v_{kj,n}}{\sum_j \frac{1}{\sum_{k'} t_{ik,'n} v_{k'j,n}} v_{kj,n}}}, \tag{9}$$

$$v_{kj,n} \leftarrow v_{kj,n} \sqrt{\frac{\sum_i \frac{|y_{ij,n}|^2}{(\sum_{k'} t_{ik,'n} v_{k'j,n})^2} t_{ik,n}}{\sum_i \frac{1}{\sum_{k'} t_{ik,'n} v_{k'j,n}} t_{ik,n}}}. \tag{10}$$

The update rules in (6)–(10) provide convergence-guaranteed optimization, i.e., the value of the cost function does not increase via the iterative parameter update.

Although ILRMA achieves efficient and initialization-robust estimation, its performance is limited if some sources are not point sources. In theory, an SCM of diffuse noise has a rank of two or more and the rank-1 spatial model does not hold. This leads to a discrepancy from the model of ILRMA.

C. MNMF and FastMNMF

Compared with ILRMA, the full-rank SCM can model diffuse sources more appropriately [21]. MNMF [18], [19] unifies the NMF source model and full-rank SCM and improves BSS performance. However, it suffers from a heavy computational cost and dependence on parameter initialization owing to its many degrees of freedom. To accelerate its algorithm, FastMNMF has been proposed [22], [23]. FastMNMF assumes that SCMs are jointly diagonalizable to greatly reduce the computational complexity, but nevertheless it highly depends on the initial value. Unlike (2) in ILRMA, both MNMF and FastMNMF firstly estimate the full-rank SCMs and secondly apply the multichannel Wiener filter constructed with the SCM to the observed signal to output the separated signals.

III. Proposed Framework for BSE

A. Motivation and Strategy

In this paper, we deal with a BSE situation where one directional target speech and diffuse background noise are mixed. As mentioned in Section II-B, ILRMA cannot accurately express diffuse noise; instead, BSS based on a full-rank SCM, such as MNMF or FastMNMF, should be applied in this situation. However, the estimation of the full-rank SCM has a relatively large computational cost, and its performance is always more unstable than the performance of ILRMA [8] because of the large number of spatial parameters, INM^2, which can be reduced to INM using the rank-1 spatial model (ILRMA).

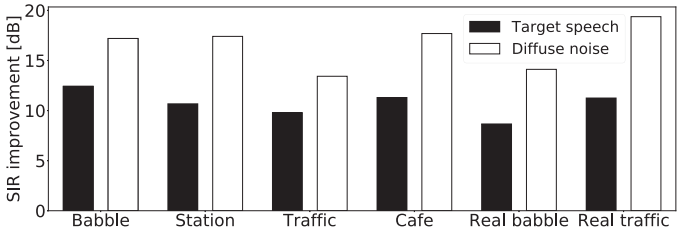

Fig. 1. SIR improvement obtained by ILRMA averaged over 10 parameter-initialization random seeds, four target directions, and six speech sources, where directional target speech and each diffuse noise were mixed and number of microphones was four (remaining experimental conditions are described in Section V-A and VI).

For this reason, to achieve efficient and robust BSE, we propose a new SCM estimation method, rank-constrained SCM estimation, as a multichannel-information-preserving postprocessing method of ILRMA. Although the sources are categorized into two groups (target and noise), we assume that one directional target speech and $M-1$ noise components are mixed ($N = M$) when ILRMA performs BSS. This assumption allows us to model diffuse noise using $M-1$ spatial bases (rank-$(M-1)$ SCM). The extraction of the directional target speech by ILRMA is still difficult because components of diffuse noise exist in the same direction as the target source. However, it is expected that ILRMA can separate diffuse noise with high accuracy even if one spatial basis for diffuse noise is lacking. Fig. 1 shows the average separation performance (source-to-interference ratio (SIR) [33]) obtained by ILRMA, where the directional target speech and diffuse noise are mixed and the experimental conditions are described in Section V-A and VI. It can be seen that diffuse noise is accurately estimated (almost perfectly with more than around 15 dB accuracy) rather than the directional target speech, where diffuse noise is modeled using the rank-$(M-1)$ SCM. This is because the demixing filters for diffuse noise can precisely cancel the directional target speech, which is a point source [17], meaning that the steering vector of the directional target speech can be estimated by ILRMA with high accuracy. This implies that we can fix some spatial parameters in the full-rank SCM for diffuse noise by utilizing the estimates obtained by ILRMA in advance.

On the basis of the above motivation, we propose the following new estimation method for the full-rank SCM of diffuse noise: (a) the rank-1 SCM for the directional target speech and rank-$(M-1)$ SCM for diffuse noise are estimated by ILRMA and fixed, (b) the lost spatial basis for diffuse noise is restored to estimate noise components in the direction of the target speech, and (c) a multichannel Wiener filter is applied to suppress the noise components remaining in the separated directional target speech. Also, we re-estimate the variance of the directional

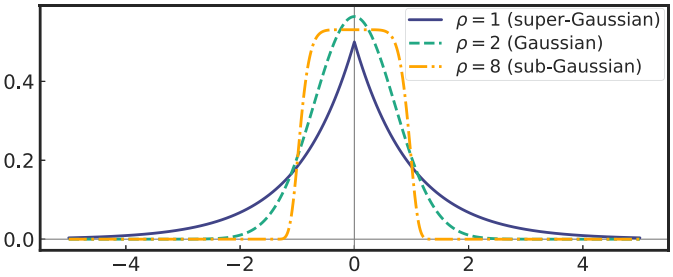

Fig. 2. Examples of univariate real-valued GGD with unit variance. By changing shape parameter of GGD to arbitrary positive value, we can represent various types of distribution, e.g., super- or sub-Gaussian distribution.

target speech and diffuse noise simultaneously because ILRMA-estimated directional target speech is contaminated with noise components (e.g., in Fig. 1, the SIR of the directional target speech is degraded to around 10 dB).

In addition, we introduce the multivariate GGD as the statistical generative model of the observed signal to represent various types of observed signal by changing its shape parameter. We propose MM-algorithm-based and ME-algorithm-based update rules for an arbitrary shape parameter, making use of newly introduced mathematical formulae.

B. Model and Speech Extraction

Here, we summarize the theoretical assumptions of the rank-constrained SCM estimation as follows:

- Target speech source
  - spatial assumption: a point source
  - statistical assumption: the power spectrogram is sparse
- Noise
  - spatial assumption: diffuse and not a point source
  - statistical assumption: no assumption

These assumptions are valid in many practical acoustic applications, e.g., general multichannel speech enhancement [34], hands-free speech recognition [17], and hearing-aid system [35].

We assume that the observed signal $\boldsymbol{x}_{ij}$ follows the multivariate GGD as

$$p(\boldsymbol{x}_{ij};\mathbf{0},\mathbf{R}_{ij}^{(x)},\rho)=\frac{C(\rho)}{\det\mathbf{R}_{ij}^{(x)}}\exp(-(\boldsymbol{x}_{ij}^{\mathsf{H}}(\mathbf{R}_{ij}^{(x)})^{-1}\boldsymbol{x}_{ij})^{\frac{\rho}{2}}), \tag{11}$$

where $\rho>0$ is the shape parameter of the multivariate GGD, $C(\rho)$ is the normalizing constant of the multivariate GGD, which only depends on ρ, and $\mathbf{R}_{ij}^{(x)}\in\mathbb{C}^{M\times M}$ is the SCM of the observed signal. Fig. 2 shows examples of GGD; we only show a univariate real-valued case for readers' easy understanding. For $\rho=2$, the multivariate GGD corresponds to the complex Gaussian distribution. $\mathbf{R}_{ij}^{(x)}$ is modeled as the sum of the SCMs of the directional target speech and diffuse noise as

$$\mathbf{R}_{ij}^{(x)}=r_{ij}^{(\mathrm{t})}\boldsymbol{a}_i^{(\mathrm{t})}(\boldsymbol{a}_i^{(\mathrm{t})})^{\mathsf{H}}+r_{ij}^{(\mathrm{n})}\mathbf{R}_i^{(\mathrm{n})}, \tag{12}$$

where $r_{ij}^{(\mathrm{t})}>0$ and $r_{ij}^{(\mathrm{n})}>0$ are the variances of the directional target speech and diffuse noise, respectively, $\boldsymbol{a}_i^{(\mathrm{t})}\in\mathbb{C}^M$ is the n_{t}th steering vector, i.e., $\boldsymbol{a}_i^{(\mathrm{t})}:=\boldsymbol{a}_{i,n_{\mathrm{t}}}$, where n_{t} denotes the index of the directional target speech, $\boldsymbol{a}_{i,1},\ldots,\boldsymbol{a}_{i,M}$ are the steering vectors estimated by ILRMA, and $\boldsymbol{a}_i^{(\mathrm{t})}(\boldsymbol{a}_i^{(\mathrm{t})})^{\mathsf{H}}$ represents the rank-1 SCM of the directional target speech. $\mathbf{R}_i^{(\mathrm{n})}\in\mathbb{C}^{M\times M}$ denotes the full-rank SCM of diffuse noise. To impose sparsity and improve the estimation performance for speech, we introduce an a priori distribution for $r_{ij}^{(\mathrm{t})}$ using the inverse gamma distribution [36]

$$p(r_{ij}^{(\mathrm{t})};\alpha,\beta)=\frac{\beta^{\alpha}}{\Gamma(\alpha)}\left(r_{ij}^{(\mathrm{t})}\right)^{-\alpha-1}\exp\left(-\frac{\beta}{r_{ij}^{(\mathrm{t})}}\right), \tag{13}$$

where $\alpha>0$, $\beta>0$, and $\Gamma(\cdot)$ are the shape parameter, scale parameter, and gamma function, respectively. In contrast, we do not assume any a priori distribution on the variance $r_{ij}^{(\mathrm{n})}$ to represent many kinds of noises. $\mathbf{R}_i^{(\mathrm{n})}$ is expressed by the sum of two components as

$$\mathbf{R}_i^{(\mathrm{n})}=\mathbf{R}_i^{\prime(\mathrm{n})}+\lambda_i\boldsymbol{b}_i\boldsymbol{b}_i^{\mathsf{H}}, \tag{14}$$

$$\mathbf{R}_i^{\prime(\mathrm{n})}=\frac{1}{J}\sum_j\hat{\boldsymbol{y}}_{ij}^{(\mathrm{n})}(\hat{\boldsymbol{y}}_{ij}^{(\mathrm{n})})^{\mathsf{H}}, \tag{15}$$

$$\hat{\boldsymbol{y}}_{ij}^{(\mathrm{n})}=\mathbf{W}_i^{-1}(\boldsymbol{w}_{i,1}^{\mathsf{H}}\boldsymbol{x}_{ij},\ldots,\boldsymbol{w}_{i,n_{\mathrm{t}}-1}^{\mathsf{H}}\boldsymbol{x}_{ij},0,\boldsymbol{w}_{i,n_{\mathrm{t}}+1}^{\mathsf{H}}\boldsymbol{x}_{ij},\ldots,\boldsymbol{w}_{i,N}^{\mathsf{H}}\boldsymbol{x}_{ij})^{\mathsf{T}}, \tag{16}$$

where $\mathbf{R}_i^{\prime(\mathrm{n})}\in\mathbb{C}^{M\times M}$ is the specific SCM of diffuse noise estimated by ILRMA; since $\mathbf{R}_i^{\prime(\mathrm{n})}$ consists of $M-1$ noise estimates, its rank is $M-1$. $\boldsymbol{b}_i\in\mathbb{C}^M$ is a vector satisfying the condition that the column vectors of $\mathbf{R}_i^{\prime(\mathrm{n})}$ and $\boldsymbol{b}_i$ are linearly independent, and λ_i is a scalar variable. For example, $\boldsymbol{b}_i$ can be set to $\boldsymbol{a}_i^{(\mathrm{t})}$ or the unit eigenvector of $\mathbf{R}_i^{\prime(\mathrm{n})}$ that corresponds to the zero eigenvalue. $\hat{\boldsymbol{y}}_{ij}^{(\mathrm{n})}$ is the source image of diffuse noise, whose scale is fixed using a back-projection operation [37]. To restore the lost spatial basis in $\mathbf{R}_i^{\prime(\mathrm{n})}$, we simultaneously estimate the scalar weight λ_i, the variance of the directional target speech $r_{ij}^{(\mathrm{t})}$, and the variance of diffuse noise $r_{ij}^{(\mathrm{n})}$ with $\boldsymbol{a}_i^{(\mathrm{t})}$, the rank-$(M-1)$ SCM $\mathbf{R}_i^{\prime(\mathrm{n})}$, and $\boldsymbol{b}_i$ fixed. In summary, the number of spatial parameters to be estimated in the proposed method is INM (for ILRMA) $+\,I$ (for λ_i), i.e., $I(NM+1)$, which is much less than those of MNMF (INM^2) and FastMNMF (IM^2+INM).

For the estimation of the parameters $r_{ij}^{(\mathrm{t})}$, $r_{ij}^{(\mathrm{n})}$, and λ_i, the cost function to be minimized is defined as the following negative log posterior of (11) with the prior distribution (13):

$$\begin{aligned}\mathcal{L}(\Theta)&=-\sum_{i,j}\log\left(p(\boldsymbol{x}_{ij}|\Theta;\rho)p(r_{ij}^{(\mathrm{t})};\alpha,\beta)\right)\\&=\sum_{i,j}\left[(\boldsymbol{x}_{ij}^{\mathsf{H}}(\mathbf{R}_{ij}^{(x)})^{-1}\boldsymbol{x}_{ij})^{\frac{\rho}{2}}+\log\det\mathbf{R}_{ij}^{(x)}\right.\\&\qquad\left.+\,(\alpha+1)\log r_{ij}^{(\mathrm{t})}+\frac{\beta}{r_{ij}^{(\mathrm{t})}}\right]+\mathrm{const.},\end{aligned} \tag{17}$$

where $\Theta = \{r_{ij}^{(\mathrm{t})}, r_{ij}^{(\mathrm{n})}, \lambda_i\}$ is the set of objective variables. The variables are estimated so that they minimize $\mathcal{L}(\Theta)$ in an iterative manner after the initialization using the ILRMA estimates, whose detail is given in Sections III-C, III-D, III-E, III-F, and III-G. We initialize $r_{ij}^{(\mathrm{t})}$ and $r_{ij}^{(\mathrm{n})}$ as follows:

$$r_{ij}^{(\mathrm{t})} = \sum_k t_{ik,n_\mathrm{t}} v_{kj,n_\mathrm{t}}, \tag{18}$$

$$r_{ij}^{(\mathrm{n})} = \frac{1}{M} (\hat{\boldsymbol{y}}_{ij}^{(\mathrm{n})})^{\mathsf{H}} (\mathbf{R}_i'^{(\mathrm{n})})^{+} \hat{\boldsymbol{y}}_{ij}^{(\mathrm{n})}, \tag{19}$$

where t_{ik,n_t} and v_{kj,n_t} are the components of the NMF source model of the directional target speech estimated by ILRMA and $^{+}$ denotes the pseudoinverse. Also, we initialize λ_i as the minimum nonzero eigenvalue of $\mathbf{R}_i'^{(\mathrm{n})}$. After the initialization and estimation of the parameters, we extract the source image of the directional target speech using the multichannel Wiener filter as

$$\hat{\boldsymbol{s}}_{ij}^{(t)} = r_{ij}^{(\mathrm{t})} \boldsymbol{a}_i^{(\mathrm{t})} (\boldsymbol{a}_i^{(\mathrm{t})})^{\mathsf{H}} (\mathbf{R}_{ij}^{(x)})^{-1} \boldsymbol{x}_{ij}. \tag{20}$$

The definition of the multichannel Wiener filter is the same as that in [21].

C. Optimization Framework

Since it is difficult to minimize (17) directly, we employ the MM and ME algorithms [30], [38], [31]. We prepare an auxiliary function $Q(\Theta, \Omega)$ that satisfies

$$\mathcal{L}(\Theta) \leq Q(\Theta, \Omega) \qquad ({}^{\forall}\Theta, {}^{\forall}\Omega), \tag{21}$$

$$\mathcal{L}(\Theta) = \min_{\Omega} Q(\Theta, \Omega) \qquad ({}^{\forall}\Theta), \tag{22}$$

where Ω is a set of auxiliary variables. Note that the definition of the auxiliary function here is identical to the one used in [38], which is a general definition and introduces auxiliary variables explicitly. In the MM and ME algorithms, we iterate the following two updates: First, we update the auxiliary variable using up-to-date objective variables $\tilde{\Theta}$ as

$$\tilde{\Omega} \leftarrow \arg\min_{\Omega} Q(\tilde{\Theta}, \Omega). \tag{23}$$

Next, we update the objective variables using up-to-date auxiliary variables in the MM algorithm as

$$\Theta \leftarrow \arg\min_{\hat{\Theta}} Q(\hat{\Theta}, \tilde{\Omega}) \tag{24}$$

and in the ME algorithm as

$$\Theta \leftarrow \hat{\Theta} \text{ s.t. } Q(\hat{\Theta}, \tilde{\Omega}) = Q(\tilde{\Theta}, \tilde{\Omega}),\ \hat{\Theta} \neq \tilde{\Theta}. \tag{25}$$

These algorithms guarantee the monotonic non-increase of the cost function $\mathcal{L}(\Theta)$. In particular, it has been experimentally found that the ME algorithm achieves faster convergence than the MM algorithm in a univariate model because the changes in parameters are larger [31]. However, it is difficult to determine or even to find $\hat{\Theta}$ that does not change the value of the auxiliary function in a multivariate case, and an ME-based update rule for MNMF or FastMNMF has not yet been proposed.

D. Generic Inequality and Identity for Rank-Constrained SCM Estimation

We first design auxiliary functions depending on ρ for use in the MM and ME algorithms. For each of the terms in (17) except the power of the quadratic term $(\boldsymbol{x}_{ij}^{\mathsf{H}} (\mathbf{R}_{ij}^{(x)})^{-1} \boldsymbol{x}_{ij})^{\frac{\rho}{2}}$, we can utilize inequalities proposed in [19], [39]. Regarding the power of the quadratic term, if we have an auxiliary function $Q(\Theta, \Omega)$ for the quadratic term satisfying

$$\boldsymbol{x}_{ij}^{\mathsf{H}} (\mathbf{R}_{ij}^{(x)})^{-1} \boldsymbol{x}_{ij} \leq Q(\Theta, \Omega), \tag{26}$$

then we can construct a new auxiliary function $\tilde{Q}(\Theta, \Omega)$ as

$$(\boldsymbol{x}_{ij}^{\mathsf{H}} (\mathbf{R}_{ij}^{(x)})^{-1} \boldsymbol{x}_{ij})^{\frac{\rho}{2}} \leq (Q(\Theta, \Omega))^{\frac{\rho}{2}} =: \tilde{Q}(\Theta, \Omega). \tag{27}$$

However, for the quadratic term $\boldsymbol{x}_{ij}^{\mathsf{H}} (\mathbf{R}_{ij}^{(x)})^{-1} \boldsymbol{x}_{ij}$, the inequality used in [19] assumes the SCM of each source to be full-rank (positive definite), which does not hold for the rank-constrained SCM estimation. Therefore, we derive the following *new generic* inequality so as to design the auxiliary function for the quadratic term with positive semi-definite SCMs.

Theorem 1: Suppose that positive semi-definite Hermitian matrices $\mathbf{R}_n \in \mathbb{C}^{M \times M}$ $(n = 1, \ldots, N')$ satisfy $\mathrm{rank}(\sum_n \mathbf{R}_n) = M$. Then, it holds that

$$\mathrm{tr}((\textstyle\sum_n \mathbf{R}_n)^{-1} \mathbf{X}) \leq \sum_n \mathrm{tr}\left(\boldsymbol{\Phi}_n^{\mathsf{H}} \mathbf{R}_n^{+} \boldsymbol{\Phi}_n \mathbf{X}\right) \tag{28}$$

for any positive semi-definite Hermitian matrix $\mathbf{X} \in \mathbb{C}^{M \times M}$ and matrices $\boldsymbol{\Phi}_n \in \mathbb{C}^{M \times M}$ that satisfy the following equations:

$$\mathrm{Ker}\, \boldsymbol{\Phi}_n = \mathrm{Ker}\, \mathbf{X} \quad (\forall n), \tag{29}$$

$$\mathrm{Im}\, \boldsymbol{\Phi}_n = \mathrm{Im}\, \mathbf{R}_n \quad (\forall n), \tag{30}$$

$$\sum_n \boldsymbol{\Phi}_n = \mathbf{P}, \tag{31}$$

where Ker and Im denote the kernel and column spaces, respectively. Here, $\mathbf{P} \in \mathbb{C}^{M \times M}$ is the projection matrix to the column space of $\mathbf{X}$.

Equality holds if and only if the following holds:

$$\boldsymbol{\Phi}_n = \mathbf{R}_n \left(\sum_{n'} \mathbf{R}_{n'} \right)^{-1} \mathbf{P} \quad (\forall n). \tag{32}$$

Proof: To prove the inequality (28) and the equality condition (32), we formulate the optimization problem as

$$\text{(P)} \left| \begin{array}{ll} \underset{\{\boldsymbol{\Phi}_n\}_{n=1}^{N'}}{\text{minimize}} & \sum_n \mathrm{tr}\left(\boldsymbol{\Phi}_n^{\mathsf{H}} \mathbf{R}_n^{+} \boldsymbol{\Phi}_n \mathbf{X}\right) \\ \text{subject to} & \sum_n \boldsymbol{\Phi}_n = \mathbf{P}, \\ & \mathrm{Ker}\, \boldsymbol{\Phi}_n = \mathrm{Ker}\, \mathbf{X} \quad (\forall n), \\ & \mathrm{Im}\, \boldsymbol{\Phi}_n = \mathrm{Im}\, \mathbf{R}_n \quad (\forall n). \end{array} \right.$$

The objective function is strictly convex as long as the kernel space of $\boldsymbol{\Phi}_n$ is equivalent to that of $\mathbf{X}$. Therefore, an optimal solution of (P) exists and is unique. In the following, we prove that the unique optimal solution is (32) and the optimal value is identical to the left-hand side of (28) using the method of Lagrange multipliers.

Let $l_n := \mathrm{rank}\, \mathbf{R}_n$. Since $\mathbf{R}_n$ is a positive semi-definite Hermitian matrix, there exists $\mathbf{S}_n \in \mathbb{C}^{M \times l_n}$ that satisfies $\mathbf{R}_n =$

$\mathbf{S}_n\mathbf{S}_n^{\mathsf{H}}$. There also exists $\mathbf{\Xi}_n \in \mathbb{C}^{l_n \times M}$ that satisfies $\mathbf{\Phi}_n = \mathbf{S}_n\mathbf{\Xi}_n\mathbf{P}$ because $\mathrm{Im}\,\mathbf{\Phi}_n = \mathrm{Im}\,\mathbf{R}_n$. Using the pseudoinverse, we have $\mathbf{\Phi}_n^{\mathsf{H}}\mathbf{R}_n^{+}\mathbf{\Phi}_n = \mathbf{P}\mathbf{\Xi}_n^{\mathsf{H}}\mathbf{\Xi}_n\mathbf{P}$, and the objective function is deformed as $\sum_n \mathrm{tr}(\mathbf{P}\mathbf{\Xi}_n^{\mathsf{H}}\mathbf{\Xi}_n\mathbf{P}\mathbf{X}) = \sum_n \mathrm{tr}(\mathbf{\Xi}_n^{\mathsf{H}}\mathbf{\Xi}_n\mathbf{X})$. Thus, the optimization problem (P) is equivalent to the following problem:

$$\text{(P')} \left| \begin{array}{ll} \underset{\{\mathbf{\Xi}_n\}_{n=1}^{N'}}{\text{minimize}} & \sum_n \mathrm{tr}\left(\mathbf{\Xi}_n^{\mathsf{H}}\mathbf{\Xi}_n\mathbf{X}\right) \\ \text{subject to} & \sum_n \mathbf{S}_n\mathbf{\Xi}_n\mathbf{P} = \mathbf{P}. \end{array} \right.$$

Since the kernel and column spaces of $\mathbf{\Xi}_n$ are unconstrained, we can solve (P') by the method of Lagrange multipliers. We introduce a Lagrange multiplier $\mathbf{L} \in \mathbb{C}^{M\times M}$ and thus reduce the problem to the minimization of the function

$$g(\{\mathbf{\Xi}_n\}_{n=1}^{N'}, \mathbf{L}) = \sum_n \mathrm{tr}(\mathbf{\Xi}_n^{\mathsf{H}}\mathbf{\Xi}_n\mathbf{X}) - \mathrm{Re}\left[\mathrm{tr}\left(\left(\sum_n \mathbf{S}_n\mathbf{\Xi}_n\mathbf{P} - \mathbf{P}\right)^{\mathsf{H}}\mathbf{L}\right)\right] \tag{33}$$

with respect to $\{\mathbf{\Xi}_n\}_{n=1}^{N'}$ and $\mathbf{L}$, where $\mathrm{Re}[\cdot]$ denotes the real part of the argument. Since the objective function of (P') is a convex function of $\{\mathbf{\Xi}_n\}_{n=1}^{N'}$, the optimal points of (P') can be acquired by setting the derivative of g with respect to $\overline{\mathbf{\Xi}}_n$ to $\mathbf{O}$, where $\bar{\cdot}$ denotes the conjugate matrix. Therefore, we have the following equation that gives the optimal points of (P'):

$$\frac{\partial g}{\partial \overline{\mathbf{\Xi}}_n} = \mathbf{\Xi}_n\mathbf{X} - \mathbf{S}_n^{\mathsf{H}}\mathbf{L}\mathbf{P} = \mathbf{O}. \tag{34}$$

Left-multiplying $\mathbf{S}_n$ to both sides of the equation, we have

$$\mathbf{\Phi}_n\mathbf{X} = \mathbf{R}_n\mathbf{L}\mathbf{P} \tag{35}$$

using $\mathbf{P}\mathbf{X} = \mathbf{X}$ and the definitions of $\mathbf{S}_n$ and $\mathbf{\Xi}_n$. Summing both sides of (35) with respect to n leads to the following equations:

$$\mathbf{X} = \left(\sum_n \mathbf{R}_n\right)\mathbf{L}\mathbf{P} \tag{36}$$

$$\Leftrightarrow \left(\sum_n \mathbf{R}_n\right)^{-1}\mathbf{X} = \mathbf{L}\mathbf{P}. \tag{37}$$

Substituting (37) into (35), we have the unique optimal solution $\mathbf{\Phi}_n = \mathbf{R}_n(\sum_{n'}\mathbf{R}_{n'})^{-1}\mathbf{X}\mathbf{X}^{+} = \mathbf{R}_n(\sum_{n'}\mathbf{R}_{n'})^{-1}\mathbf{P}$. The optimal value of (P) coincides with $\mathrm{tr}((\sum_n \mathbf{R}_n)^{-1}\mathbf{X})$, which can be checked by substitution. ■

Here, we put $N' = 2$, $\mathbf{X} = \boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^{\mathsf{H}}$, $\mathbf{R}_1 = r_{ij}^{(\mathrm{t})}\boldsymbol{a}_i^{(\mathrm{t})}(\boldsymbol{a}_i^{(\mathrm{t})})^{\mathsf{H}}$, and $\mathbf{R}_2 = r_{ij}^{(\mathrm{n})}\mathbf{R}_i^{(\mathrm{n})}$. By introducing the auxiliary variables $\mathbf{\Phi}_{ij}^{(\mathrm{t})} \in \mathbb{C}^{M\times M}$ and $\mathbf{\Phi}_{ij}^{(\mathrm{n})} \in \mathbb{C}^{M\times M}$ that satisfy $\mathbf{\Phi}_{ij}^{(\mathrm{t})} + \mathbf{\Phi}_{ij}^{(\mathrm{n})} = \mathbf{P}_{ij}$, we have the following inequality:

$$\begin{aligned} &\boldsymbol{x}_{ij}^{\mathsf{H}}\left(\mathbf{R}_{ij}^{(x)}\right)^{-1}\boldsymbol{x}_{ij} \\ &= \boldsymbol{x}_{ij}^{\mathsf{H}}\left(r_{ij}^{(\mathrm{t})}\boldsymbol{a}_i^{(\mathrm{t})}\left(\boldsymbol{a}_i^{(\mathrm{t})}\right)^{\mathsf{H}} + r_{ij}^{(\mathrm{n})}\mathbf{R}_i^{(\mathrm{n})}\right)^{-1}\boldsymbol{x}_{ij} \\ &\le \frac{|\left(\boldsymbol{a}_i^{(\mathrm{t})}\right)^{\mathsf{H}}\mathbf{\Phi}_{ij}^{(\mathrm{t})}\boldsymbol{x}_{ij}|^2}{r_{ij}^{(\mathrm{t})}} + \frac{\boldsymbol{x}_{ij}^{\mathsf{H}}\left(\mathbf{\Phi}_{ij}^{(\mathrm{n})}\right)^{\mathsf{H}}\left(\mathbf{R}_i^{(\mathrm{n})}\right)^{-1}\mathbf{\Phi}_{ij}^{(\mathrm{n})}\boldsymbol{x}_{ij}}{r_{ij}^{(\mathrm{n})}}, \end{aligned} \tag{38}$$

where $\mathbf{P}_{ij} = \boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^{\mathsf{H}}/\|\boldsymbol{x}_{ij}\|_2^2$ is the projection matrix to the column space of $\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^{\mathsf{H}}$. Equality holds if and only if the following equations hold:

$$\mathbf{\Phi}_{ij}^{(\mathrm{t})} = r_{ij}^{(\mathrm{t})}\boldsymbol{a}_i^{(\mathrm{t})}(\boldsymbol{a}_i^{(\mathrm{t})})^{\mathsf{H}}(\mathbf{R}_{ij}^{(x)})^{-1}\mathbf{P}_{ij}, \tag{39}$$

$$\mathbf{\Phi}_{ij}^{(\mathrm{n})} = r_{ij}^{(\mathrm{n})}\mathbf{R}_i^{(\mathrm{n})}(\mathbf{R}_{ij}^{(x)})^{-1}\mathbf{P}_{ij}. \tag{40}$$

To construct auxiliary functions for any ρ, we expand $(\mathbf{R}_i^{(\mathrm{n})})^{-1}$ using the following claim.

Claim 1: Let $\boldsymbol{u}_i \in \mathbb{C}^M$ be a vector that satisfies $\mathbf{R}_i'^{(\mathrm{n})}\boldsymbol{u}_i = \mathbf{0}$ and $\boldsymbol{b}_i^{\mathsf{H}}\boldsymbol{u}_i = 1$. Then, the following holds:

$$(\mathbf{R}_i^{(\mathrm{n})})^{-1} = \breve{\mathbf{R}}_i^{(\mathrm{n})} + \frac{1}{\lambda_i}\boldsymbol{u}_i\boldsymbol{u}_i^{\mathsf{H}}. \tag{41}$$

Here, $\breve{\mathbf{R}}_i^{(\mathrm{n})} = (\mathbf{E}_M - \boldsymbol{u}_i\boldsymbol{b}_i^{\mathsf{H}})(\mathbf{R}_i'^{(\mathrm{n})})^{+}(\mathbf{E}_M - \boldsymbol{b}_i\boldsymbol{u}_i^{\mathsf{H}})$ and $\mathbf{E}_M$ is the identity matrix of order M.

Proof: We denote the eigenvalue decomposition of $\mathbf{R}_i'^{(\mathrm{n})}$ as

$$\mathbf{R}_i'^{(\mathrm{n})} = \begin{pmatrix} \mathbf{U}_i & \frac{\boldsymbol{u}_i}{\|\boldsymbol{u}_i\|_2} \end{pmatrix}\begin{pmatrix} \mathbf{D}_i & \mathbf{0} \\ \mathbf{0}^{\mathsf{T}} & 0 \end{pmatrix}\begin{pmatrix} \mathbf{U}_i^{\mathsf{H}} \\ \frac{\boldsymbol{u}_i^{\mathsf{H}}}{\|\boldsymbol{u}_i\|_2} \end{pmatrix}. \tag{42}$$

Here, $\mathbf{D}_i \in \mathbb{R}^{(M-1)\times(M-1)}$ is a diagonal matrix whose entries are nonzero and $\mathbf{U}_i \in \mathbb{C}^{M\times(M-1)}$ is a matrix satisfying the condition that $(\,\mathbf{U}_i\ \boldsymbol{u}_i/\|\boldsymbol{u}_i\|_2\,) \in \mathbb{C}^{M\times M}$ is unitary. Note that from the orthogonality $\mathbf{R}_i'^{(\mathrm{n})}\boldsymbol{u}_i = \mathbf{0}$, we have $\mathbf{U}_i^{\mathsf{H}}\boldsymbol{u}_i = \mathbf{0}$ and $\boldsymbol{u}_i$ is the eigenvector of $\mathbf{R}_i'^{(\mathrm{n})}$ that corresponds to the zero eigenvalue. Using this, we have the following equation:

$$\mathbf{R}_i^{(\mathrm{n})} = \begin{pmatrix} \mathbf{U}_i & \boldsymbol{b}_i \end{pmatrix}\begin{pmatrix} \mathbf{D}_i & \mathbf{0} \\ \mathbf{0}^{\mathsf{T}} & \lambda_i \end{pmatrix}\begin{pmatrix} \mathbf{U}_i^{\mathsf{H}} \\ \boldsymbol{b}_i^{\mathsf{H}} \end{pmatrix}. \tag{43}$$

To calculate $(\mathbf{R}_i^{(\mathrm{n})})^{-1}$, we use the following equation:

$$\begin{pmatrix} \mathbf{U}_i^{\mathsf{H}} \\ \boldsymbol{b}_i^{\mathsf{H}} \end{pmatrix}^{-1} = \begin{pmatrix} (\mathbf{E}_M - \boldsymbol{u}_i\boldsymbol{b}_i^{\mathsf{H}})\mathbf{U}_i & \boldsymbol{u}_i \end{pmatrix}. \tag{44}$$

This is confirmed by the following calculation using $\mathbf{U}_i^{\mathsf{H}}\boldsymbol{u}_i = \mathbf{0}$ and $\boldsymbol{b}_i^{\mathsf{H}}\boldsymbol{u}_i = 1$:

$$\begin{aligned} &\begin{pmatrix} \mathbf{U}_i^{\mathsf{H}} \\ \boldsymbol{b}_i^{\mathsf{H}} \end{pmatrix}\begin{pmatrix} (\mathbf{E}_M - \boldsymbol{u}_i\boldsymbol{b}_i^{\mathsf{H}})\mathbf{U}_i & \boldsymbol{u}_i \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{U}_i^{\mathsf{H}}\mathbf{U}_i - \mathbf{U}_i^{\mathsf{H}}\boldsymbol{u}_i\boldsymbol{b}_i^{\mathsf{H}}\mathbf{U}_i & \mathbf{U}_i^{\mathsf{H}}\boldsymbol{u}_i \\ \boldsymbol{b}_i^{\mathsf{H}}\mathbf{U}_i - \boldsymbol{b}_i^{\mathsf{H}}\boldsymbol{u}_i\boldsymbol{b}_i^{\mathsf{H}}\mathbf{U}_i & \boldsymbol{b}_i^{\mathsf{H}}\boldsymbol{u}_i \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{E}_{M-1} & \mathbf{0} \\ \mathbf{0}^{\mathsf{T}} & 1 \end{pmatrix} \\ &= \mathbf{E}_M. \end{aligned} \tag{45}$$

Using (44), we can calculate $(\mathbf{R}_i^{(\mathrm{n})})^{-1}$ as follows:

$$\begin{aligned} &(\mathbf{R}_i^{(\mathrm{n})})^{-1} \\ &= \begin{pmatrix} \mathbf{U}_i^{\mathsf{H}} \\ \boldsymbol{b}_i^{\mathsf{H}} \end{pmatrix}^{-1}\begin{pmatrix} \mathbf{D}_i^{-1} & \mathbf{0} \\ \mathbf{0}^{\mathsf{T}} & \frac{1}{\lambda_i} \end{pmatrix}\begin{pmatrix} \mathbf{U}_i & \boldsymbol{b}_i \end{pmatrix}^{-1} \\ &= \begin{pmatrix} (\mathbf{E}_M - \boldsymbol{u}_i\boldsymbol{b}_i^{\mathsf{H}})\mathbf{U}_i & \boldsymbol{u}_i \end{pmatrix}\begin{pmatrix} \mathbf{D}_i^{-1} & \mathbf{0} \\ \mathbf{0}^{\mathsf{T}} & \frac{1}{\lambda_i} \end{pmatrix} \end{aligned}$$

$$\cdot \begin{pmatrix} \mathbf{U}_i^{\mathsf{H}}(\mathbf{E}_M - \boldsymbol{b}_i\boldsymbol{u}_i^{\mathsf{H}}) \\ \boldsymbol{u}_i^{\mathsf{H}} \end{pmatrix}$$
$$= (\mathbf{E}_M - \boldsymbol{u}_i\boldsymbol{b}_i^{\mathsf{H}})\mathbf{U}_i\mathbf{D}_i^{-1}\mathbf{U}_i^{\mathsf{H}}(\mathbf{E}_M - \boldsymbol{b}_i\boldsymbol{u}_i^{\mathsf{H}}) + \frac{1}{\lambda_i}\boldsymbol{u}_i\boldsymbol{u}_i^{\mathsf{H}}. \tag{46}$$

From the positive semi-definiteness and Hermitian property of $\mathbf{R}_i'^{(\mathrm{n})}$, the pseudoinverse $(\mathbf{R}_i'^{(\mathrm{n})})^+$ equals $\mathbf{U}_i\mathbf{D}_i^{-1}\mathbf{U}_i^{\mathsf{H}}$ because its eigenvalue decomposition and singular value decomposition coincide. ■

Finally, we can combine the identity (41) and inequalities (27) and (38) to obtain the auxiliary function $\tilde{Q}(\Theta, \Omega)$ as

$$\tilde{Q}(\Theta,\Omega) = \left[\frac{|(\boldsymbol{a}_i^{(\mathrm{t})})^{\mathsf{H}}\boldsymbol{\Phi}_{ij}^{(\mathrm{t})}\boldsymbol{x}_{ij}|^2}{r_{ij}^{(\mathrm{t})}} + \frac{1}{r_{ij}^{(\mathrm{n})}}\left(\frac{|\boldsymbol{u}_i^{\mathsf{H}}\boldsymbol{\Phi}_{ij}^{(\mathrm{n})}\boldsymbol{x}_{ij}|^2}{\lambda_i} + \boldsymbol{x}_{ij}^{\mathsf{H}}(\boldsymbol{\Phi}_{ij}^{(\mathrm{n})})^{\mathsf{H}}\breve{\mathbf{R}}_i^{(\mathrm{n})}\boldsymbol{\Phi}_{ij}^{(\mathrm{n})}\boldsymbol{x}_{ij}\right)\right]^{\frac{\rho}{2}}, \tag{47}$$

where $\Omega = \{\boldsymbol{\Phi}_{ij}^{(\mathrm{t})}, \boldsymbol{\Phi}_{ij}^{(\mathrm{n})}\}$ is the set of auxiliary variables. To construct the auxiliary function of the cost function (17), we further bound the right-hand side of (47) for the cases $\rho > 2$ (sub-Gaussian case) and $\rho \leq 2$ (Gaussian or super-Gaussian case) separately and construct auxiliary functions in Sections III-E and III-F.

E. Auxiliary Functions for Sub-Gaussian Case ($\rho > 2$)

For $\rho > 2$, to bound $\tilde{Q}(\Theta, \Omega)$, we introduce the auxiliary variables $\eta_{ij} > 0$ and $\varphi_{ij} > 0$ that satisfy $\eta_{ij} + \varphi_{ij} < 1$ and use Jensen's inequality as follows:

$$\begin{aligned}
&\tilde{Q}(\Theta,\Omega) \\
&\leq \eta_{ij}\left(\frac{|(\boldsymbol{a}_i^{(\mathrm{t})})^{\mathsf{H}}\boldsymbol{\Phi}_{ij}^{(\mathrm{t})}\boldsymbol{x}_{ij}|^2}{\eta_{ij}r_{ij}^{(\mathrm{t})}}\right)^{\frac{\rho}{2}} + \varphi_{ij}\left(\frac{|\boldsymbol{u}_i^{\mathsf{H}}\boldsymbol{\Phi}_{ij}^{(\mathrm{n})}\boldsymbol{x}_{ij}|^2}{r_{ij}^{(\mathrm{n})}\lambda_i\varphi_{ij}}\right)^{\frac{\rho}{2}} \\
&\quad + (1-\eta_{ij}-\varphi_{ij})\left(\frac{\boldsymbol{x}_{ij}^{\mathsf{H}}(\boldsymbol{\Phi}_{ij}^{(\mathrm{n})})^{\mathsf{H}}\breve{\mathbf{R}}_i^{(\mathrm{n})}\boldsymbol{\Phi}_{ij}^{(\mathrm{n})}\boldsymbol{x}_{ij}}{r_{ij}^{(\mathrm{n})}(1-\eta_{ij}-\varphi_{ij})}\right)^{\frac{\rho}{2}} \\
&= \eta_{ij}^{1-\frac{\rho}{2}}\left(\frac{|(\boldsymbol{a}_i^{(\mathrm{t})})^{\mathsf{H}}\boldsymbol{\Phi}_{ij}^{(\mathrm{t})}\boldsymbol{x}_{ij}|^2}{r_{ij}^{(\mathrm{t})}}\right)^{\frac{\rho}{2}} \\
&\quad + \frac{1}{(r_{ij}^{(\mathrm{n})})^{\frac{\rho}{2}}}\left\{\varphi_{ij}^{1-\frac{\rho}{2}}\left(\frac{|\boldsymbol{u}_i^{\mathsf{H}}\boldsymbol{\Phi}_{ij}^{(\mathrm{n})}\boldsymbol{x}_{ij}|^2}{\lambda_i}\right)^{\frac{\rho}{2}}\right. \\
&\quad \left. + (1-\eta_{ij}-\varphi_{ij})^{1-\frac{\rho}{2}}\left(\boldsymbol{x}_{ij}^{\mathsf{H}}(\boldsymbol{\Phi}_{ij}^{(\mathrm{n})})^{\mathsf{H}}\breve{\mathbf{R}}_i^{(\mathrm{n})}\boldsymbol{\Phi}_{ij}^{(\mathrm{n})}\boldsymbol{x}_{ij}\right)^{\frac{\rho}{2}}\right\},
\end{aligned} \tag{48}$$

where equality holds if and only if the following equations hold:

$$\eta_{ij} = \frac{\frac{|(\boldsymbol{a}_i^{(\mathrm{t})})^{\mathsf{H}}\boldsymbol{\Phi}_{ij}^{(\mathrm{t})}\boldsymbol{x}_{ij}|^2}{r_{ij}^{(\mathrm{t})}}}{\frac{|(\boldsymbol{a}_i^{(\mathrm{t})})^{\mathsf{H}}\boldsymbol{\Phi}_{ij}^{(\mathrm{t})}\boldsymbol{x}_{ij}|^2}{r_{ij}^{(\mathrm{t})}} + \frac{\boldsymbol{x}_{ij}^{\mathsf{H}}(\boldsymbol{\Phi}_{ij}^{(\mathrm{n})})^{\mathsf{H}}(\mathbf{R}_i^{(\mathrm{n})})^{-1}\boldsymbol{\Phi}_{ij}^{(\mathrm{n})}\boldsymbol{x}_{ij}}{r_{ij}^{(\mathrm{n})}}}, \tag{49}$$

$$\varphi_{ij} = \frac{\frac{|\boldsymbol{u}_i^{\mathsf{H}}\boldsymbol{\Phi}_{ij}^{(\mathrm{n})}\boldsymbol{x}_{ij}|^2}{r_{ij}^{(\mathrm{n})}\lambda_i}}{\frac{|(\boldsymbol{a}_i^{(\mathrm{t})})^{\mathsf{H}}\boldsymbol{\Phi}_{ij}^{(\mathrm{t})}\boldsymbol{x}_{ij}|^2}{r_{ij}^{(\mathrm{t})}} + \frac{\boldsymbol{x}_{ij}^{\mathsf{H}}(\boldsymbol{\Phi}_{ij}^{(\mathrm{n})})^{\mathsf{H}}(\mathbf{R}_i^{(\mathrm{n})})^{-1}\boldsymbol{\Phi}_{ij}^{(\mathrm{n})}\boldsymbol{x}_{ij}}{r_{ij}^{(\mathrm{n})}}}. \tag{50}$$

Next, by introducing an auxiliary variable $\boldsymbol{\Psi}_{ij} \in \mathbb{C}^{M\times M}$, we have

$$\log\det\mathbf{R}_{ij}^{(x)} \leq \mathrm{tr}(\boldsymbol{\Psi}_{ij}^{-1}(\mathbf{R}_{ij}^{(x)} - \boldsymbol{\Psi}_{ij})) + \log\det\boldsymbol{\Psi}_{ij}, \tag{51}$$

where equality holds if and only if $\boldsymbol{\Psi}_{ij} = \mathbf{R}_{ij}^{(x)}$, which evaluates the second term of (17). Moreover, to derive the update rule of the objective variables, we introduce the auxiliary variables $\chi_{ij}^{(\mathrm{t})} > 0$, $\chi_{ij}^{(\mathrm{n})} > 0$, and $\delta_i > 0$, and bound the first term of the right-hand side of (51) in the same manner as in [39] as

$$\begin{aligned}
&\mathrm{tr}(\boldsymbol{\Psi}_{ij}^{-1}\mathbf{R}_{ij}^{(x)}) \\
&= r_{ij}^{(\mathrm{t})}(\boldsymbol{a}_i^{(\mathrm{t})})^{\mathsf{H}}\boldsymbol{\Psi}_{ij}^{-1}\boldsymbol{a}_i^{(\mathrm{t})} + r_{ij}^{(\mathrm{n})}(\lambda_i\boldsymbol{b}_i^{\mathsf{H}}\boldsymbol{\Psi}_{ij}^{-1}\boldsymbol{b}_i + \mathrm{tr}(\boldsymbol{\Psi}_{ij}^{-1}\mathbf{R}_i'^{(\mathrm{n})})) \\
&\leq \left(\frac{(\chi_{ij}^{(\mathrm{t})})^{1-\nu_1}(r_{ij}^{(\mathrm{t})})^{\nu_1}}{\nu_1} + \left(1-\frac{1}{\nu_1}\right)\chi_{ij}^{(\mathrm{t})}\right)(\boldsymbol{a}_i^{(\mathrm{t})})^{\mathsf{H}}\boldsymbol{\Psi}_{ij}^{-1}\boldsymbol{a}_i^{(\mathrm{t})} \\
&\quad + \left(\frac{(\chi_{ij}^{(\mathrm{n})})^{1-\nu_1}(r_{ij}^{(\mathrm{n})})^{\nu_1}}{\nu_1} + \left(1-\frac{1}{\nu_1}\right)\chi_{ij}^{(\mathrm{n})}\right) \\
&\quad \cdot \left(\lambda_i\boldsymbol{b}_i^{\mathsf{H}}\boldsymbol{\Psi}_{ij}^{-1}\boldsymbol{b}_i + \mathrm{tr}(\boldsymbol{\Psi}_{ij}^{-1}\mathbf{R}_i'^{(\mathrm{n})})\right) \\
&\leq \left(\frac{(\chi_{ij}^{(\mathrm{t})})^{1-\nu_1}(r_{ij}^{(\mathrm{t})})^{\nu_1}}{\nu_1} + \left(1-\frac{1}{\nu_1}\right)\chi_{ij}^{(\mathrm{t})}\right)(\boldsymbol{a}_i^{(\mathrm{t})})^{\mathsf{H}}\boldsymbol{\Psi}_{ij}^{-1}\boldsymbol{a}_i^{(\mathrm{t})} \\
&\quad + \left(\frac{(\chi_{ij}^{(\mathrm{n})})^{1-\nu_1}(r_{ij}^{(\mathrm{n})})^{\nu_1}}{\nu_1} + \left(1-\frac{1}{\nu_1}\right)\chi_{ij}^{(\mathrm{n})}\right) \\
&\quad \cdot \left(\left(\frac{\delta_i^{1-\nu_1}\lambda_i^{\nu_1}}{\nu_1} + \left(1-\frac{1}{\nu_1}\right)\delta_i\right)\boldsymbol{b}_i^{\mathsf{H}}\boldsymbol{\Psi}_{ij}^{-1}\boldsymbol{b}_i\right. \\
&\quad \left. + \mathrm{tr}\left(\boldsymbol{\Psi}_{ij}^{-1}\mathbf{R}_i'^{(\mathrm{n})}\right)\right),
\end{aligned} \tag{52}$$

where $\nu_1 \geq 1$ is set to 1 for the MM algorithm and $\frac{\rho}{2}$ for the ME algorithm. For $\nu_1 > 1$, equality holds if and only if $\chi_{ij}^{(\mathrm{t})} = r_{ij}^{(\mathrm{t})}$, $\chi_{ij}^{(\mathrm{n})} = r_{ij}^{(\mathrm{n})}$, and $\delta_i = \lambda_i$.

To bound the third and fourth terms of the right-hand side of (17), we introduce the auxiliary variables $\zeta_{ij} > 0$ and $\kappa_{ij} > 0$, and obtain the following inequality using the tangent-line inequality and the inequality in [39]:

$$\begin{aligned}
&(\alpha+1)\log r_{ij}^{(\mathrm{t})} + \frac{\beta}{r_{ij}^{(\mathrm{t})}} \\
&\leq \frac{\alpha+1}{\nu_1}\frac{(r_{ij}^{(\mathrm{t})})^{\nu_1} - (\zeta_{ij})^{\nu_1}}{(\zeta_{ij})^{\nu_1}} + \frac{\alpha+1}{\nu_1}\log(\zeta_{ij})^{\nu_1} \\
&\quad + \frac{\beta}{\kappa_{ij}}\left(\frac{1}{\nu_2}\frac{\kappa_{ij}^{\nu_2}}{(r_{ij}^{(\mathrm{t})})^{\nu_2}} + 1 - \frac{1}{\nu_2}\right),
\end{aligned} \tag{53}$$

where $\nu_2 \geq 1$ is set to $\frac{\rho}{2}$. Equality holds if and only if $\zeta_{ij} = r_{ij}^{(\mathrm{t})}$ and $\kappa_{ij} = r_{ij}^{(\mathrm{t})}$ for $\nu_2 > 1$.

We can bring together the above-mentioned inequalities to obtain the auxiliary functions $Q_{\rho>2,\mathrm{MM}}$ and $Q_{\rho>2,\mathrm{ME}}$. For the MM algorithm, we have the auxiliary function

$$
\begin{aligned}
&Q_{\rho>2,\mathrm{MM}}(\Theta, \Omega_{\rho>2,\mathrm{MM}}) \\
&= \left[\eta_{ij}^{1-\frac{\rho}{2}} \left(\frac{|(\boldsymbol{a}_i^{(\mathrm{t})})^{\mathsf{H}} \boldsymbol{\Phi}_{ij}^{(\mathrm{t})} \boldsymbol{x}_{ij}|^2}{r_{ij}^{(\mathrm{t})}} \right)^{\frac{\rho}{2}} \right. \\
&\quad + \frac{1}{(r_{ij}^{(\mathrm{n})})^{\frac{\rho}{2}}} \left\{ \varphi_{ij}^{1-\frac{\rho}{2}} \left(\frac{|\boldsymbol{u}_i^{\mathsf{H}} \boldsymbol{\Phi}_{ij}^{(\mathrm{n})} \boldsymbol{x}_{ij}|^2}{\lambda_i} \right)^{\frac{\rho}{2}} \right. \\
&\quad \left. + (1 - \eta_{ij} - \varphi_{ij})^{1-\frac{\rho}{2}} \left(\boldsymbol{x}_{ij}^{\mathsf{H}} \left(\boldsymbol{\Phi}_{ij}^{(\mathrm{n})} \right)^{\mathsf{H}} \breve{\mathbf{R}}_i^{(\mathrm{n})} \boldsymbol{\Phi}_{ij}^{(\mathrm{n})} \boldsymbol{x}_{ij} \right)^{\frac{\rho}{2}} \right\} \\
&\quad + \mathrm{tr}\left(\boldsymbol{\Psi}_{ij}^{-1} \mathbf{R}_{ij}^{(x)} \right) + \log\det \boldsymbol{\Psi}_{ij} - M + (\alpha+1) \frac{r_{ij}^{(\mathrm{t})}}{\zeta_{ij}} \\
&\quad \left. + (\alpha+1)(\log \zeta_{ij} - 1) + \frac{\beta}{\kappa_{ij}} \left(\frac{2\kappa_{ij}^{\frac{\rho}{2}}}{\rho (r_{ij}^{(\mathrm{t})})^{\frac{\rho}{2}}} + 1 - \frac{2}{\rho} \right) \right],
\end{aligned} \tag{54}
$$

where $\Omega_{\rho>2,\mathrm{MM}} = \{\boldsymbol{\Phi}_{ij}^{(\mathrm{t})}, \boldsymbol{\Phi}_{ij}^{(\mathrm{n})}, \eta_{ij}, \varphi_{ij}, \boldsymbol{\Psi}_{ij}, \zeta_{ij}, \kappa_{ij}\}$ is the set of auxiliary variables. Also, we have the auxiliary function for the ME algorithm

$$
\begin{aligned}
&Q_{\rho>2,\mathrm{ME}}(\Theta, \Omega_{\rho>2,\mathrm{ME}}) \\
&= \left[\eta_{ij}^{1-\frac{\rho}{2}} \left(\frac{|\left(\boldsymbol{a}_i^{(\mathrm{t})} \right)^{\mathsf{H}} \boldsymbol{\Phi}_{ij}^{(\mathrm{t})} \boldsymbol{x}_{ij}|^2}{r_{ij}^{(\mathrm{t})}} \right)^{\frac{\rho}{2}} \right. \\
&\quad + \frac{1}{(r_{ij}^{(\mathrm{n})})^{\frac{\rho}{2}}} \left\{ \varphi_{ij}^{1-\frac{\rho}{2}} \left(\frac{|\boldsymbol{u}_i^{\mathsf{H}} \boldsymbol{\Phi}_{ij}^{(\mathrm{n})} \boldsymbol{x}_{ij}|^2}{\lambda_i} \right)^{\frac{\rho}{2}} \right. \\
&\quad \left. + (1 - \eta_{ij} - \varphi_{ij})^{1-\frac{\rho}{2}} \left(\boldsymbol{x}_{ij}^{\mathsf{H}} \left(\boldsymbol{\Phi}_{ij}^{(\mathrm{n})} \right)^{\mathsf{H}} \breve{\mathbf{R}}_i^{(\mathrm{n})} \boldsymbol{\Phi}_{ij}^{(\mathrm{n})} \boldsymbol{x}_{ij} \right)^{\frac{\rho}{2}} \right\} \\
&\quad + \left(\frac{2}{\rho} \left(\chi_{ij}^{(\mathrm{n})} \right)^{1-\frac{\rho}{2}} (r_{ij}^{(\mathrm{n})})^{\frac{\rho}{2}} + \left(1 - \frac{2}{\rho} \right) \chi_{ij}^{(\mathrm{n})} \right) \\
&\quad \cdot \left(\left(\frac{2}{\rho} \delta_i^{1-\frac{\rho}{2}} \lambda_i^{\frac{\rho}{2}} + \left(1 - \frac{2}{\rho} \right) \delta_i \right) \boldsymbol{b}_i^{\mathsf{H}} \boldsymbol{\Psi}_{ij}^{-1} \boldsymbol{b}_i \right. \\
&\quad \left. + \mathrm{tr}(\boldsymbol{\Psi}_{ij}^{-1} \mathbf{R}_i'^{(\mathrm{n})}) \right) + \left(\frac{2}{\rho} \left(\chi_{ij}^{(\mathrm{t})} \right)^{1-\frac{\rho}{2}} \left(r_{ij}^{(\mathrm{t})} \right)^{\frac{\rho}{2}} \right. \\
&\quad \left. + \left(1 - \frac{2}{\rho} \right) \chi_{ij}^{(\mathrm{t})} \right) (\boldsymbol{a}_i^{(\mathrm{t})})^{\mathsf{H}} \boldsymbol{\Psi}_{ij}^{-1} \boldsymbol{a}_i^{(\mathrm{t})} + \log\det \boldsymbol{\Psi}_{ij} \\
&\quad - M + \frac{2(\alpha+1)}{\rho} \frac{(r_{ij}^{(\mathrm{t})})^{\frac{\rho}{2}}}{\zeta_{ij}^{\frac{\rho}{2}}} + \frac{2(\alpha+1)}{\rho} (\log(\zeta_{ij})^{\frac{\rho}{2}} - 1) \\
&\quad \left. + \frac{\beta}{\kappa_{ij}} \left(\frac{2\kappa_{ij}^{\frac{\rho}{2}}}{\rho (r_{ij}^{(\mathrm{t})})^{\frac{\rho}{2}}} + 1 - \frac{2}{\rho} \right) \right],
\end{aligned} \tag{55}
$$

where $\Omega_{\rho>2,\mathrm{ME}} = \{\boldsymbol{\Phi}_{ij}^{(\mathrm{t})}.\boldsymbol{\Phi}_{ij}^{(\mathrm{n})}, \eta_{ij}, \varphi_{ij}, \boldsymbol{\Psi}_{ij}, \chi_{ij}^{(\mathrm{t})}, \chi_{ij}^{(\mathrm{n})}, \delta_i, \zeta_{ij}, \kappa_{ij}\}$ is the set of auxiliary variables.

F. Auxiliary Function for Gaussian or Super-Gaussian Case ($\rho \leq 2$)

For $\rho \leq 2$, we use the following inequality by introducing an auxiliary variable $\gamma_{ij} > 0$:

$$
\begin{aligned}
\tilde{Q}(\Theta, \Omega) &\leq \frac{\rho}{2\gamma_{ij}^{1-\frac{\rho}{2}}} \\
&\quad \times \left[\frac{|(\boldsymbol{a}_i^{(\mathrm{t})})^{\mathsf{H}} \boldsymbol{\Phi}_{ij}^{(\mathrm{t})} \boldsymbol{x}_{ij}|^2}{r_{ij}^{(\mathrm{t})}} + \frac{1}{r_{ij}^{(\mathrm{n})}} \left(\frac{|\boldsymbol{u}_i^{\mathsf{H}} \boldsymbol{\Phi}_{ij}^{(\mathrm{n})} \boldsymbol{x}_{ij}|^2}{\lambda_i} \right. \right. \\
&\quad \left. \left. + \boldsymbol{x}_{ij}^{\mathsf{H}} (\boldsymbol{\Phi}_{ij}^{(\mathrm{n})})^{\mathsf{H}} \breve{\mathbf{R}}_i^{(\mathrm{n})} \boldsymbol{\Phi}_{ij}^{(\mathrm{n})} \boldsymbol{x}_{ij} \right) - \gamma_{ij} \right] + \gamma_{ij}^{\frac{\rho}{2}}.
\end{aligned} \tag{56}
$$

Equality holds if and only if the following equation holds:

$$
\begin{aligned}
\gamma_{ij} &= \frac{|(\boldsymbol{a}_i^{(\mathrm{t})})^{\mathsf{H}} \boldsymbol{\Phi}_{ij}^{(\mathrm{t})} \boldsymbol{x}_{ij}|^2}{r_{ij}^{(\mathrm{t})}} + \frac{1}{r_{ij}^{(\mathrm{n})}} \left(\frac{|\boldsymbol{u}_i^{\mathsf{H}} \boldsymbol{\Phi}_{ij}^{(\mathrm{n})} \boldsymbol{x}_{ij}|^2}{\lambda_i} \right. \\
&\quad \left. + \boldsymbol{x}_{ij}^{\mathsf{H}} (\boldsymbol{\Phi}_{ij}^{(\mathrm{n})})^{\mathsf{H}} \breve{\mathbf{R}}_i^{(\mathrm{n})} \boldsymbol{\Phi}_{ij}^{(\mathrm{n})} \boldsymbol{x}_{ij} \right).
\end{aligned} \tag{57}
$$

Combining this inequality with inequalities (51) and (53) with $\nu_1 = 1$ and $\nu_2 = 1$, we have the auxiliary function for both of the MM and ME algorithms as

$$
\begin{aligned}
&Q_{\rho\leq2,\mathrm{MM/ME}}(\Theta, \Omega_{\rho\leq2,\mathrm{MM/ME}}) \\
&= \sum_{i,j} \left[\frac{\rho}{2\gamma_{ij}^{1-\frac{\rho}{2}}} \left(\frac{|(\boldsymbol{a}_i^{(\mathrm{t})})^{\mathsf{H}} \boldsymbol{\Phi}_{ij}^{(\mathrm{t})} \boldsymbol{x}_{ij}|^2}{r_{ij}^{(\mathrm{t})}} + \frac{|\boldsymbol{u}_i^{\mathsf{H}} \boldsymbol{\Phi}_{ij}^{(\mathrm{n})} \boldsymbol{x}_{ij}|^2}{r_{ij}^{(\mathrm{n})} \lambda_i} \right. \right. \\
&\quad \left. + \frac{\boldsymbol{x}_{ij}^{\mathsf{H}} (\boldsymbol{\Phi}_{ij}^{(\mathrm{n})})^{\mathsf{H}} \breve{\mathbf{R}}_i^{(\mathrm{n})} \boldsymbol{\Phi}_{ij}^{(\mathrm{n})} \boldsymbol{x}_{ij}}{r_{ij}^{(\mathrm{n})}} \right) + \left(1 - \frac{\rho}{2} \right) \gamma_{ij}^{\frac{\rho}{2}} \\
&\quad + \mathrm{tr}(\boldsymbol{\Psi}_{ij}^{-1} \mathbf{R}_{ij}^{(x)}) + \log\det \boldsymbol{\Psi}_{ij} - M \\
&\quad \left. + (\alpha+1) \frac{r_{ij}^{(\mathrm{t})}}{\zeta_{ij}} + (\alpha+1)(\log \zeta_{ij} - 1) + \frac{\beta}{r_{ij}^{(\mathrm{t})}} \right],
\end{aligned} \tag{58}
$$

where $\Omega_{\rho\leq2,\mathrm{MM/ME}} = \{\boldsymbol{\Phi}_{ij}^{(\mathrm{t})}, \boldsymbol{\Phi}_{ij}^{(\mathrm{n})}, \gamma_{ij}, \boldsymbol{\Psi}_{ij}, \zeta_{ij}\}$ is the set of auxiliary variables.

G. MM-Algorithm-Based and ME-Algorithm-Based Update Rules

To derive the update rules, we minimize $Q_{\rho>2,\mathrm{MM}}$ and $Q_{\rho\leq2,\mathrm{MM/ME}}$, or equalize $Q_{\rho>2,\mathrm{ME}}$ and $Q_{\rho\leq2,\mathrm{MM/ME}}$. $Q_{\rho>2,\mathrm{MM}}$ has the form

$$
Q_{\rho>2,\mathrm{MM}}(x, \Omega_{\rho>2,\mathrm{MM}}) = \frac{A}{x^{\frac{\rho}{2}}} + Bx + C \tag{59}
$$

with respect to each objective variable x (such as $r_{ij}^{(\mathrm{t})}$ in (54)), where $A > 0$, $B > 0$, and C are numbers independent of x but

dependent on $\Omega_{\rho>2,\mathrm{MM}}$ and the other objective variables. By setting the derivative of $Q_{\rho>2,\mathrm{MM}}$ to zero, we have the update rule

$$x \leftarrow \arg\min_x Q_{\rho>2,\mathrm{MM}}(x) = \left(\frac{\rho A}{2B}\right)^{\frac{2}{\rho+2}}. \tag{60}$$

Next, we consider the ME-algorithm-based update rule for $\rho > 2$. The auxiliary function $Q_{\rho>2,\mathrm{ME}}$ has the form

$$Q_{\rho>2,\mathrm{ME}}(x) = \frac{A}{x^{\frac{\rho}{2}}} + Bx^{\frac{\rho}{2}} + C. \tag{61}$$

Therefore, we have the following update rule derivation in the same manner as in [39]:

$$\begin{aligned} &Q_{\rho>2,\mathrm{ME}}(x) = Q_{\rho>2,\mathrm{ME}}(\tilde{x}) \\ &\Leftrightarrow (x^{\frac{\rho}{2}} - \tilde{x}^{\frac{\rho}{2}})\left(x^{\frac{\rho}{2}} - \frac{A}{B\tilde{x}^{\frac{\rho}{2}}}\right) = 0 \qquad (62) \\ &\Leftrightarrow x = \tilde{x}, \frac{1}{\tilde{x}}\left(\frac{A}{B}\right)^{\frac{2}{\rho}}, \qquad (63) \end{aligned}$$

for arbitrary $x > 0$ and $\tilde{x} > 0$, which leads to the update rule

$$x \leftarrow \frac{1}{x}\left(\frac{A}{B}\right)^{\frac{2}{\rho}}. \tag{64}$$

For the MM-algorithm-based and ME-algorithm-based update rules for $\rho \leq 2$, the auxiliary function $Q_{\rho\leq2,\mathrm{MM/ME}}$ has the form

$$Q_{\rho\leq2,\mathrm{MM/ME}}(x) = \frac{A}{x} + Bx + C, \tag{65}$$

and therefore, we can acquire the update rules in the same manner as (60) and (64).

Consequently, the update rule is given by

$$r_{ij}^{(\mathrm{t})} \leftarrow r_{ij}^{(\mathrm{t})}\left(\frac{c_{ij}|\boldsymbol{x}_{ij}^{\mathsf{H}}(\mathbf{R}_{ij}^{(x)})^{-1}\boldsymbol{a}_i^{(\mathrm{t})}|^2 + \frac{\beta}{(r_{ij}^{(\mathrm{t})})^2}}{(\boldsymbol{a}_i^{(\mathrm{t})})^{\mathsf{H}}(\mathbf{R}_{ij}^{(x)})^{-1}\boldsymbol{a}_i^{(\mathrm{t})} + \frac{\alpha+1}{r_{ij}^{(\mathrm{t})}}}\right)^q, \tag{66}$$

$$r_{ij}^{(\mathrm{n})} \leftarrow r_{ij}^{(\mathrm{n})}\left(\frac{c_{ij}\boldsymbol{x}_{ij}^{\mathsf{H}}(\mathbf{R}_{ij}^{(x)})^{-1}\mathbf{R}_i^{(\mathrm{n})}(\mathbf{R}_{ij}^{(x)})^{-1}\boldsymbol{x}_{ij}}{\mathrm{tr}((\mathbf{R}_{ij}^{(x)})^{-1}\mathbf{R}_i^{(\mathrm{n})})}\right)^q, \tag{67}$$

$$\lambda_i \leftarrow \lambda_i\left(\frac{\sum_j c_{ij}r_{ij}^{(\mathrm{n})}|\boldsymbol{b}_i^{\mathsf{H}}(\mathbf{R}_{ij}^{(x)})^{-1}\boldsymbol{x}_{ij}|^2}{\sum_j r_{ij}^{(\mathrm{n})}\boldsymbol{b}_i^{\mathsf{H}}(\mathbf{R}_{ij}^{(x)})^{-1}\boldsymbol{b}_i}\right)^q, \tag{68}$$

where q equals $\min(1/2, 2/(\rho+2))$ for the MM-algorithm-based update rule and $\min(1, 2/\rho)$ for the ME-algorithm-based update rule, and $c_{ij} = \rho/(2(\boldsymbol{x}_{ij}^{\mathsf{H}}(\mathbf{R}_{ij}^{(x)})^{-1}\boldsymbol{x}_{ij})^{1-\frac{\rho}{2}})$. Note that $\mathbf{R}_{ij}^{(x)}$ and c_{ij} are updated only after the update of $r_{ij}^{(\mathrm{n})}$ and λ_i. It is expected that the ME-algorithm-based update will be faster than the MM-algorithm-based one because q for the ME-algorithm-based update rule is always larger than that for the MM-algorithm-based one, i.e.,

$$\min\left(\frac{1}{2}, \frac{2}{\rho+2}\right) < \min\left(1, \frac{2}{\rho}\right) \quad (\forall \rho > 0), \tag{69}$$

which makes the multiplication coefficients more distant from unity in the multiplicative update. Also, for any shape parameter ρ, we can always apply the ME algorithm to minimization of $\mathcal{L}(\Theta)$ because the solution that satisfies $Q_{\rho>2,\mathrm{ME}}(x) = Q_{\rho>2,\mathrm{ME}}(\tilde{x})$ or $Q_{\rho\leq2,\mathrm{MM/ME}}(x) = Q_{\rho\leq2,\mathrm{MM/ME}}(\tilde{x})$ is found in a closed form.

IV. Acceleration of Parameter Update

A. Motivation

The update rule (66)–(68) involves an inverse matrix operation of $M \times M$ matrices at each time-frequency slot. Thus, its computational complexity is $O(IJM^3)$. Such a heavy computational load restricts its implementation on low-resource hardware. 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 (*first-stage acceleration*) and (ii) expanding matrix inversion using the pseudoinverse of matrices (*second-stage acceleration*). After these expansions, we propose a new update rule that involves only scalar operations; thus, we can update parameters with neither matrix inversions nor matrix multiplications. The proposed update rules described in this chapter are analytically identical to the naive update rule because they are based on equivalent expansions of matrix inversions. We can reduce the computational complexity in each iteration of the MM and ME algorithms described in Section III.

B. Key Concept

For the first-stage acceleration, we use the Sherman–Morrison formula to expand the inverse of matrix $\mathbf{R}_{ij}^{(x)} = r_{ij}^{(\mathrm{t})}\boldsymbol{a}_i^{(\mathrm{t})}(\boldsymbol{a}_i^{(\mathrm{t})})^{\mathsf{H}} + r_{ij}^{(\mathrm{n})}\mathbf{R}_i^{(\mathrm{n})}$ as

$$\begin{aligned} &(\mathbf{R}_{ij}^{(x)})^{-1} \\ &= \frac{1}{r_{ij}^{(\mathrm{n})}}\left((\mathbf{R}_i^{(\mathrm{n})})^{-1} - \frac{r_{ij}^{(\mathrm{t})}}{r_{ij}^{(\mathrm{n})} + r_{ij}^{(\mathrm{t})}(\boldsymbol{a}_i^{(\mathrm{t})})^{\mathsf{H}}(\mathbf{R}_i^{(\mathrm{n})})^{-1}\boldsymbol{a}_i^{(\mathrm{t})}}\right. \\ &\qquad \left.\cdot (\mathbf{R}_i^{(\mathrm{n})})^{-1}\boldsymbol{a}_i^{(\mathrm{t})}(\boldsymbol{a}_i^{(\mathrm{t})})^{\mathsf{H}}(\mathbf{R}_i^{(\mathrm{n})})^{-1}\right). \end{aligned} \tag{70}$$

Note that $\mathbf{R}_i^{(\mathrm{n})} = \mathbf{R}_i'^{(\mathrm{n})} + \lambda_i\boldsymbol{b}_i\boldsymbol{b}_i^{\mathsf{H}}$ is invertible. We only need to calculate $(\mathbf{R}_i^{(\mathrm{n})})^{-1}$ at each frequency bin instead of calculating $(\mathbf{R}_{ij}^{(x)})^{-1}$ at each time-frequency slot. This makes it possible to reduce the computational complexity of the update rule from $O(IJM^3)$ to $O(IM^3 + IJM^2)$. The first term $O(IM^3)$ corresponds to matrix inversion at each frequency bin and the second term $O(IJM^2)$ corresponds to the multiplication of a matrix and a vector at each time-frequency slot.

For the second-stage acceleration, we can expand the inversion $(\mathbf{R}_i^{(\mathrm{n})})^{-1} = (\mathbf{R}_i'^{(\mathrm{n})} + \lambda_i\boldsymbol{b}_i\boldsymbol{b}_i^{\mathsf{H}})^{-1}$ as described by **Claim 1** in Section III-D. Since $\mathbf{R}_i'^{(\mathrm{n})}$ is fixed in the rank-constrained SCM estimation, neither matrix inversion nor pseudoinversion is necessary in the parameter update step. Hence, it follows that the number of matrix inversions is reduced from $O(I)$ to zero.

C. First-Stage Acceleration

Let us define scalar terms as

$$\sigma_i^{(aRa)} := \left(\boldsymbol{a}_i^{(\mathrm{t})}\right)^{\mathsf{H}} \left(\mathbf{R}_i^{(\mathrm{n})}\right)^{-1} \boldsymbol{a}_i^{(\mathrm{t})}, \tag{71}$$

$$\sigma_{ij}^{(aRx)} := \left(\boldsymbol{a}_i^{(\mathrm{t})}\right)^{\mathsf{H}} \left(\mathbf{R}_i^{(\mathrm{n})}\right)^{-1} \boldsymbol{x}_{ij}, \tag{72}$$

$$\sigma_{ij}^{(xRx)} := \boldsymbol{x}_{ij}^{\mathsf{H}} \left(\mathbf{R}_i^{(\mathrm{n})}\right)^{-1} \boldsymbol{x}_{ij}, \tag{73}$$

$$\sigma_i^{(aRu)} := \left(\boldsymbol{a}_i^{(\mathrm{t})}\right)^{\mathsf{H}} \left(\mathbf{R}_i^{(\mathrm{n})}\right)^{-1} \boldsymbol{u}_i, \tag{74}$$

$$\sigma_{ij}^{(uRx)} := \boldsymbol{u}_i^{\mathsf{H}} \left(\mathbf{R}_i^{(\mathrm{n})}\right)^{-1} \boldsymbol{x}_{ij}, \tag{75}$$

$$\sigma_i^{(aRb)} := \left(\boldsymbol{a}_i^{(\mathrm{t})}\right)^{\mathsf{H}} \left(\mathbf{R}_i^{(\mathrm{n})}\right)^{-1} \boldsymbol{b}_i, \tag{76}$$

$$\sigma_i^{(bRb)} := \boldsymbol{b}_i^{\mathsf{H}} \left(\mathbf{R}_i^{(\mathrm{n})}\right)^{-1} \boldsymbol{b}_i, \tag{77}$$

$$\sigma_i^{(au)} := \left(\boldsymbol{a}_i^{(\mathrm{t})}\right)^{\mathsf{H}} \boldsymbol{u}_i, \tag{78}$$

$$\sigma_{ij}^{(ux)} := \boldsymbol{u}_i^{\mathsf{H}} \boldsymbol{x}_{ij}, \tag{79}$$

$$\sigma_i^{(uu)} := \boldsymbol{u}_i^{\mathsf{H}} \boldsymbol{u}_i, \tag{80}$$

$$\mu_{ij} := \frac{r_{ij}^{(\mathrm{t})}}{r_{ij}^{(\mathrm{n})} + r_{ij}^{(\mathrm{t})} \sigma_i^{(aRa)}}. \tag{81}$$

Using (70), we can calculate the following:

$$(\boldsymbol{a}_i^{(\mathrm{t})})^{\mathsf{H}} (\mathbf{R}_{ij}^{(x)})^{-1} \boldsymbol{a}_i^{(\mathrm{t})} = \frac{\sigma_i^{(aRa)}}{r_{ij}^{(\mathrm{n})} + r_{ij}^{(\mathrm{t})} \sigma_i^{(aRa)}}, \tag{82}$$

$$(\boldsymbol{a}_i^{(\mathrm{t})})^{\mathsf{H}} (\mathbf{R}_{ij}^{(x)})^{-1} \boldsymbol{x}_{ij} = \frac{\sigma_{ij}^{(aRx)}}{r_{ij}^{(\mathrm{n})} + r_{ij}^{(\mathrm{t})} \sigma_i^{(aRa)}}, \tag{83}$$

$$\boldsymbol{x}_{ij}^{\mathsf{H}} (\mathbf{R}_{ij}^{(x)})^{-1} \boldsymbol{x}_{ij} = \frac{1}{r_{ij}^{(\mathrm{n})}} (\sigma_{ij}^{(xRx)} - \mu_{ij} |\sigma_{ij}^{(aRx)}|^2), \tag{84}$$

$$\mathrm{tr}((\mathbf{R}_{ij}^{(x)})^{-1} \mathbf{R}_i^{(\mathrm{n})}) = \frac{1}{r_{ij}^{(\mathrm{n})}} \left(M - \mu_{ij} \sigma_i^{(aRa)}\right), \tag{85}$$

$$\begin{aligned} &\boldsymbol{x}_{ij}^{\mathsf{H}} (\mathbf{R}_{ij}^{(x)})^{-1} \mathbf{R}_i^{(\mathrm{n})} (\mathbf{R}_{ij}^{(x)})^{-1} \boldsymbol{x}_{ij} \\ &= \frac{1}{(r_{ij}^{(\mathrm{n})})^2} \left(\sigma_{ij}^{(xRx)} - 2\mu_{ij} |\sigma_{ij}^{(aRx)}|^2 + \mu_{ij}^2 \sigma_i^{(aRa)} |\sigma_{ij}^{(aRx)}|^2\right), \end{aligned} \tag{86}$$

$$\boldsymbol{u}_i^{\mathsf{H}} (\mathbf{R}_{ij}^{(x)})^{-1} \boldsymbol{x}_{ij} = \frac{1}{r_{ij}^{(\mathrm{n})}} \left(\sigma_{ij}^{(uRx)} - \mu_{ij} \overline{\sigma_i^{(aRu)}} \sigma_{ij}^{(aRx)}\right), \tag{87}$$

$$\boldsymbol{b}_i^{\mathsf{H}} (\mathbf{R}_{ij}^{(x)})^{-1} \boldsymbol{b}_i = \frac{1}{r_{ij}^{(\mathrm{n})}} \left(\sigma_i^{(bRb)} - \mu_{ij} |\sigma_i^{(aRb)}|^2\right), \tag{88}$$

where $\bar{\cdot}$ denotes the complex conjugate. Applying these equations to (66)–(68), we obtain the first-stage accelerated update rule as follows:

$$r_{ij}^{(\mathrm{t})} \leftarrow r_{ij}^{(\mathrm{t})} \left(\frac{c_{ij} \left| \frac{\sigma_{ij}^{(aRx)}}{r_{ij}^{(\mathrm{n})} + r_{ij}^{(\mathrm{t})} \sigma_i^{(aRa)}} \right|^2 + \frac{\beta}{(r_{ij}^{(\mathrm{t})})^2}}{\frac{\sigma_i^{(aRa)}}{r_{ij}^{(\mathrm{n})} + r_{ij}^{(\mathrm{t})} \sigma_i^{(aRa)}} + \frac{\alpha+1}{r_{ij}^{(\mathrm{t})}}} \right)^q, \tag{89}$$

$$\begin{aligned} r_{ij}^{(\mathrm{n})} \leftarrow r_{ij}^{(\mathrm{n})} &\left(c_{ij} (\sigma_{ij}^{(xRx)} - 2\mu_{ij} |\sigma_{ij}^{(aRx)}|^2 \right. \\ &\left. + \mu_{ij}^2 \sigma_i^{(aRa)} |\sigma_{ij}^{(aRx)}|^2)\right)^q \left(r_{ij}^{(\mathrm{n})} (M - \mu_{ij} \sigma_i^{(aRa)})\right)^{-q}, \end{aligned} \tag{90}$$

$$\lambda_i \leftarrow \lambda_i \left(\frac{\sum_j \frac{c_{ij}}{r_{ij}^{(\mathrm{n})}} |\sigma_{ij}^{(uRx)} - \mu_{ij} \overline{\sigma_i^{(aRu)}} \sigma_{ij}^{(aRx)}|^2}{\sum_j (\sigma_i^{(bRb)} - \mu_{ij} |\sigma_i^{(aRb)}|^2)} \right)^q. \tag{91}$$

Note that μ_{ij} and c_{ij} are updated only after the update of $r_{ij}^{(\mathrm{n})}$ and λ_i.

D. Second-Stage Acceleration

Calculating the quadratic terms without matrix inversion enables further acceleration. We define some other quadratic terms as follows:

$$\tau_i^{(aa)} := (\boldsymbol{a}_i^{(\mathrm{t})})^{\mathsf{H}} \breve{\mathbf{R}}_i^{(\mathrm{n})} \boldsymbol{a}_i^{(\mathrm{t})}, \tag{92}$$

$$\tau_{ij}^{(ax)} := (\boldsymbol{a}_i^{(\mathrm{t})})^{\mathsf{H}} \breve{\mathbf{R}}_i^{(\mathrm{n})} \boldsymbol{x}_{ij}, \tag{93}$$

$$\tau_{ij}^{(xx)} := \boldsymbol{x}_{ij}^{\mathsf{H}} \breve{\mathbf{R}}_i^{(\mathrm{n})} \boldsymbol{x}_{ij}, \tag{94}$$

$$\tau_i^{(au)} := (\boldsymbol{a}_i^{(\mathrm{t})})^{\mathsf{H}} \breve{\mathbf{R}}_i^{(\mathrm{n})} \boldsymbol{u}_i, \tag{95}$$

$$\tau_{ij}^{(ux)} := \boldsymbol{u}_i^{\mathsf{H}} \breve{\mathbf{R}}_i^{(\mathrm{n})} \boldsymbol{x}_{ij}. \tag{96}$$

These terms and $\sigma_i^{(au)}$, $\sigma_{ij}^{(ux)}$, and $\sigma_i^{(uu)}$ do not depend on the variables $r_{ij}^{(\mathrm{t})}$, $r_{ij}^{(\mathrm{n})}$, and λ_i and can be calculated before the update iteration. The quadratic terms $\sigma_i^{(aRa)}$, $\sigma_{ij}^{(aRx)}$, $\sigma_{ij}^{(xRx)}$, $\sigma_i^{(aRu)}$, $\sigma_{ij}^{(uRx)}$, $\sigma_i^{(aRb)}$, and $\sigma_i^{(bRb)}$, which involve matrix inversion, can be transformed using (41) as follows:

$$\sigma_i^{(aRa)} = \tau_i^{(aa)} + \frac{|\sigma_i^{(au)}|^2}{\lambda_i}, \tag{97}$$

$$\sigma_{ij}^{(aRx)} = \tau_{ij}^{(ax)} + \frac{\sigma_i^{(au)} \sigma_{ij}^{(ux)}}{\lambda_i}, \tag{98}$$

$$\sigma_{ij}^{(xRx)} = \tau_{ij}^{(xx)} + \frac{|\sigma_{ij}^{(ux)}|^2}{\lambda_i}, \tag{99}$$

$$\sigma_i^{(aRu)} = \tau_i^{(au)} + \frac{\sigma_i^{(au)} \sigma_i^{(uu)}}{\lambda_i}, \tag{100}$$

$$\sigma_{ij}^{(uRx)} = \tau_{ij}^{(ux)} + \frac{\sigma_i^{(uu)} \sigma_{ij}^{(ux)}}{\lambda_i}, \tag{101}$$

$$\sigma_i^{(aRb)} = \frac{\sigma_i^{(au)}}{\lambda_i}, \tag{102}$$

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.	$O(IM^3 + IJM^2)$	$O(IJ)$

$$\sigma_i^{(b\mathrm{R}b)} = \frac{1}{\lambda_i}. \tag{103}$$

We can calculate these terms only by scalar operations and update parameters by substituting them into the update rule consisting of (89)–(91).

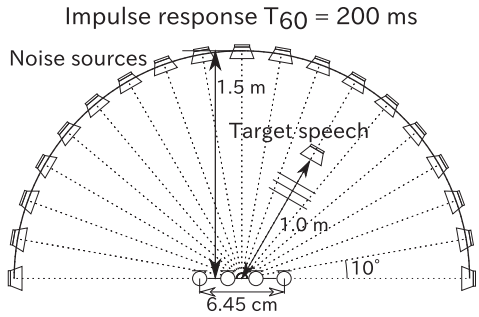

Fig. 3. Recording conditions of impulse responses (when directional target speech is located at 30°).

E. Advantage of Proposed Accelerated Update Rules

In general, the complexity of order-M matrix inversions 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 (18) and (19) is required for all methods. The proposed algorithms can reduce the complexity via the use of (70) and (41). In particular, the second-stage acceleration has greatly improved efficiency because its computational cost does not depend on the number of microphones, M.

For example, $I = 513$ and $J = 275$ for the conditions described in Section V-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.

V. BSE EXPERIMENT ON SIMULATED DATA

A. Experimental Condition

To confirm the efficacy of the proposed method, we conducted BSE experiments using simulated mixtures of directional target speech and diffuse noise. We compared eleven methods, namely, ILRMA [8], independent vector extraction (IVE) [34], BSSA [17], the multichannel Wiener filter with single-channel noise estimation (MWF1) [40], the multichannel Wiener filter with multichannel noise estimation (MWF2) [41], MNMF [19], MNMF initialized by ILRMA (ILRMA+MNMF) [8], [24], FastMNMF [23], FastMNMF initialized by ILRMA (ILRMA+FastMNMF), and the rank-constrained SCM estimation initialized by ILRMA with the proposed MM-algorithm-based and ME-algorithm-based updates. MWF1 and MWF2 consist of a minimum variance distortionless response (MVDR) beamformer, which is constructed using ILRMA estimates (the noise SCM and the steering vector of the directional target speech), and single-channel Wiener filtering. In MWF1, the noise power spectrum is estimated using a minima controlled recursive averaging noise estimation approach [40]. In MWF2, the noise power spectrum is estimated using $M - 1$ outputs of ILRMA that correspond to noise components. In both MWF1 and MWF2, the a priori speech-to-noise ratio is estimated by a decision-directed (DD) approach [42].

In ILRMA, the observed signal $\boldsymbol{x}_{ij}$ is preprocessed via a sphering transformation using principal component analysis. For IVE, the separation filter for the directional target speech was initialized by ILRMA. For BSSA, we used ILRMA instead of FDICA in [17] and set the oversubtraction and flooring parameters to 1.4 and 0, respectively. For MWF1, we used the same parameter settings as in [40] except for the weighting factor of the DD approach. For both MWF1 and MWF2, the weighting factor was set to 0.9. For ILRMA, MNMF, and FastMNMF, all the NMF variables were initialized by nonnegative random values. The demixing matrix $\mathbf{W}_i$ in ILRMA and the SCMs in MNMF and FastMNMF were initialized by the identity matrix. For ILRMA+MNMF and ILRMA+FastMNMF, the NMF variables were taken from ILRMA. Also, SCM was initialized using $\boldsymbol{a}_{i,n}\boldsymbol{a}_{i,n}^{\mathsf{H}} + \varepsilon \mathbf{E}_M$ for ILRMA+MNMF and $\boldsymbol{a}_{i,n}\boldsymbol{a}_{i,n}^{\mathsf{H}} + \varepsilon \sum_{n' \neq n} \boldsymbol{a}_{i,n'}\boldsymbol{a}_{i,n'}^{\mathsf{H}}$ for ILRMA+FastMNMF, where $\boldsymbol{a}_{i,n}$ was estimated by ILRMA and ε was set to 10^{-5}. For the rank-constrained SCM estimation, the hyperparameters in (13) were set to $\alpha = 2.5$ and $\beta = 10^{-16}$, which were selected experimentally, the shape parameter ρ of the multivariate GGD was set to 0.5, 1, 2, 4, and 8, and $\boldsymbol{b}_i$ was set to the unit eigenvector of $\mathbf{R}_i'^{(\mathrm{n})}$ that corresponds to the zero eigenvalue. For all methods, the index of the directional target speech was blindly determined by finding the separated signal that has the largest kurtosis value among all the separated signals [43].

For simulated experiments, we prepared four types of diffuse noise: babble, station, traffic, and cafe noises. As the dry source of the directional target speech, we selected six different speech signals from the JNAS speech corpus [44] and separately used each of them at each speech extraction trial. As the dry sources of 19 speakers employed as babble noise components, we used other speech signals obtained from the same corpus. For station, traffic, and cafe noises, we obtained noise signals from DEMAND [45] and split them into 19 short-time periods. These dry sources were convoluted with the impulse responses shown in Fig. 3. The directional target speech was located 0°, 10°, 20°, or 30° clockwise from the normal to a microphone array, the number of microphones was 2, 3, or 4, 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 × 3.5 m, its reverberation time was about 200 ms, and an STFT was performed using a 64-ms-long Hamming window with a 32-ms-long shift. The speech-to-noise ratio was set to 0 dB. The source-to-distortion ratio (SDR) [33] was used as a total evaluation score in terms of separation performance and

TABLE II
EXPERIMENTAL CONDITIONS

Sampling frequency	16 kHz
Number of NMF bases K	10 for each source model
Number of iterations in methods except ILRMA and IVE	200
Number of iterations in ILRMA	50
Number of iterations in IVE	4000

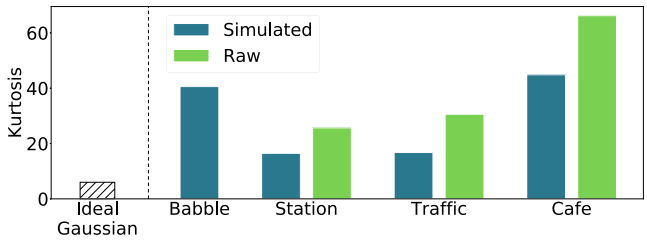

Fig. 4. Average frequency-subband-wise kurtosis of power spectra for ideal Gaussian, babble, station, traffic, and cafe noises. For station, traffic, and cafe noises, kurtosis values of simulated and raw noises are shown.

sound distortion. The SDR improvement, defined by the output SDR minus the input SDR, was obtained in each experiment. The other conditions are shown in Table II. Note that the number of iterations for IVE was set to 4000, which is recommended to ensure sufficient performance owing to the slow convergence [34].

To assess the realness, we calculated and compared the frequency-subband-wise kurtosis of power spectra [46] of simulated and raw noises used in the experiments. Fig. 4 shows the average frequency-subband-wise kurtosis for each noise. Each simulated noise has a sufficiently larger kurtosis value than ideal Gaussian noise, whereas the kurtosis becomes smaller than that of raw noise owing to the central limit theorem. At our listening, all of the simulated noises maintain the realness.

B. Comparison Between MM and ME Algorithms

We compared the proposed MM-algorithm-based and ME-algorithm-based update rules with $M = 4$ for the Gaussian ($\rho = 2$) and super-Gaussian ($\rho = 0.5$) cases. Also, for reference, we compared the EM-algorithm-based update rule proposed in our conference paper [28], ILRMA+MNMF, and ILRMA+FastMNMF only for the Gaussian case. Note that the EM algorithm cannot be applied to the super-Gaussian and sub-Gaussian cases, whereas the proposed MM and ME algorithms can be used for any cases. The SDR behaviors of these methods are shown in Fig. 5. Consistent with the supposition in Section III-C, the ME-algorithm-based update rule provided faster convergence than the MM-algorithm-based one. For the Gaussian case, the peak SDR improvement of the ME-algorithm-based update rule exceeded those of the other methods. The same tendency was observed for the other ρ and M. Considering these results, we hereafter use only the ME-algorithm-based update as the proposed method.

Here, SDR improvement reached the highest value and then decreased. To clarify the reason of this phenomenon, we compared the SDR improvement of the proposed ME-algorithm-based update rule by changing the value of hyperparameter α.

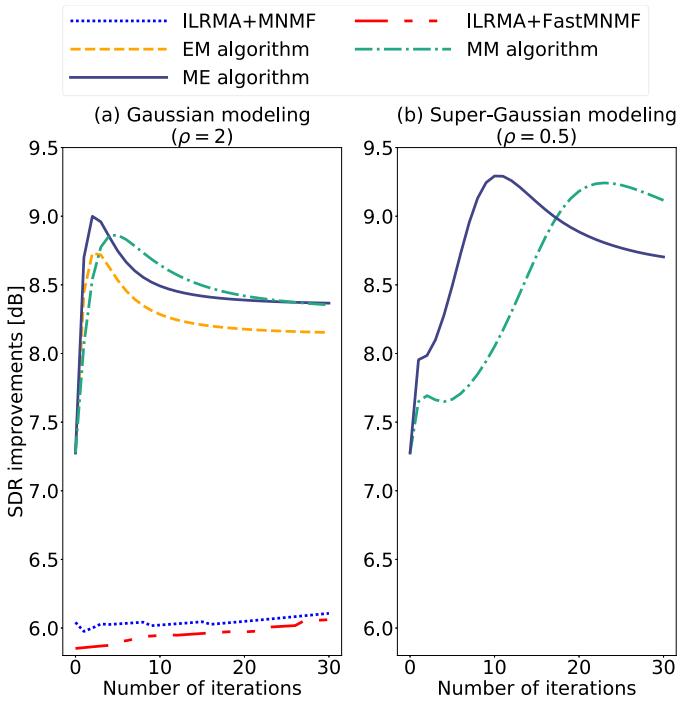

Fig. 5. SDR behaviors of conventional methods and proposed MM-algorithm-based and ME-algorithm-based update rules averaged over 10 parameter-initialization random seeds, four target directions, six speech sources, and four noises for cases of (a) $\rho = 2$ and (b) $\rho = 0.5$ in GGD.

TABLE III
SDR IMPROVEMENTS FOR EACH SHAPE PARAMETER ρ AND HYPERPARAMETER α. EACH TERM REPRESENTS "BEST-ITERATION SCORE [DB] / AFTER-30-ITERATION SCORE [DB]"

	$\alpha = 0.5$	$\alpha = 1.5$	$\alpha = 2.5$
$\rho = 2.0$ (Gaussian)	8.6 / 8.6	8.9 / 8.7	9.0 / 8.4
$\rho = 0.5$ (super-Gaussian)	8.7 / 8.7	9.1 / 8.9	9.3 / 8.7

Table III shows the best and after-30-iteration SDR improvements of the proposed method for each value of ρ and α. We can see that the larger value of α that increases sparsity in $r_{ij}^{(t)}$ makes the difference between the best SDR and the after-30-iteration SDR. This would be because the prior distribution for the directional target speech (13) acted as a sparsifier, which does not always contribute to the SDR improvement.

C. SDR and SCM Behavior Comparison Between Proposed and Conventional Methods

We compared the speech extraction performance characteristics of the conventional methods and the proposed method with various ρ values and numbers of microphones. Tables IV, V, VI, and VII show the best-iteration SDR improvement and the after-200-iteration SDR improvement for each method under babble, station, traffic, and cafe noise conditions, respectively. These SDR improvements were averaged over 10 parameter-initialization random seeds, four target directions, and six target speech sources; thus each SDR improvement score in these table entries is the average value over 240 trials.

The results in Tables IV–VII show that the proposed method outperformed the other methods. In particular, the model of the

TABLE IV
SDR IMPROVEMENTS FOR EACH METHOD AND NUMBER OF MICROPHONES UNDER BABBLE NOISE CONDITION. EACH TERM REPRESENTS "BEST-ITERATION SCORE [DB] / AFTER-200-ITERATION (EXCEPT FOR IVE; AFTER-4000-ITERATION FOR IVE) SCORE [DB]"

Methods		2 mics.	3 mics.	4 mics.
ILRMA		1.5 / -	5.3 / -	6.1 / -
IVE		1.6 / 1.6	5.3 / 4.9	6.2 / 5.8
BSSA		3.7 / -	6.0 / -	6.8 / -
MWF1		1.5 / -	5.3 / -	6.1 / -
MWF2		2.0 / -	6.1 / -	6.9 / -
MNMF		0.9 / 0.9	2.4 / 2.3	2.7 / 2.7
ILRMA+MNMF		1.9 / 1.9	5.6 / 5.5	6.5 / 6.5
FastMNMF		0.5 / 0.5	1.3 / 1.3	1.7 / 1.7
ILRMA+FastMNMF		2.0 / 2.0	5.8 / 5.7	6.5 / 6.2
Proposed method	$\rho = 0.5$	4.1 / 4.1	**7.3** / **6.7**	**8.7** / **7.7**
	$\rho = 1$	**4.2** / **4.2**	7.1 / 6.5	8.6 / **7.7**
	$\rho = 2$	**4.2** / **4.2**	7.0 / 6.4	8.5 / 7.6
	$\rho = 4$	**4.2** / **4.2**	6.8 / 6.4	8.4 / 7.6
	$\rho = 8$	**4.2** / **4.2**	6.7 / 6.4	8.3 / 7.6

TABLE V
SDR IMPROVEMENTS FOR EACH METHOD AND NUMBER OF MICROPHONES UNDER STATION NOISE CONDITION. EACH TERM REPRESENTS "BEST-ITERATION SCORE [DB] / AFTER-200-ITERATION (EXCEPT FOR IVE; AFTER-4000-ITERATION FOR IVE) SCORE [DB]"

Methods		2 mics.	3 mics.	4 mics.
ILRMA		1.0 / -	4.6 / -	6.2 / -
IVE		1.1 / 1.1	4.9 / 4.9	6.2 / 5.8
BSSA		3.1 / -	5.6 / -	6.9 / -
MWF1		1.9 / -	5.4 / -	6.9 / -
MWF2		1.8 / -	5.6 / -	7.2 / -
MNMF		3.2 / 3.2	4.1 / 4.1	3.5 / 3.5
ILRMA+MNMF		1.0 / 1.0	5.1 / 5.1	7.0 / 7.0
FastMNMF		2.9 / 2.9	3.0 / 3.0	2.6 / 2.5
ILRMA+FastMNMF		1.3 / 1.3	5.0 / 5.0	6.6 / 6.6
Proposed method	$\rho = 0.5$	**3.5** / **3.5**	**7.6** / **7.0**	**10.2** / **9.5**
	$\rho = 1$	**3.5** / **3.5**	7.2 / 6.7	9.9 / 9.4
	$\rho = 2$	**3.5** / **3.5**	6.9 / 6.5	9.8 / 9.2
	$\rho = 4$	**3.5** / **3.5**	6.7 / 6.3	9.7 / 9.2
	$\rho = 8$	3.4 / 3.4	6.6 / 6.3	9.6 / 9.2

TABLE VI
SDR IMPROVEMENTS FOR EACH METHOD AND NUMBER OF MICROPHONES UNDER TRAFFIC NOISE CONDITION. EACH TERM REPRESENTS "BEST-ITERATION SCORE [DB] / AFTER-200-ITERATION (EXCEPT FOR IVE; AFTER-4000-ITERATION FOR IVE) SCORE [DB]"

Methods		2 mics.	3 mics.	4 mics.
ILRMA		1.2 / -	4.3 / -	4.7 / -
IVE		1.2 / 1.2	4.3 / 4.3	5.4 / 5.2
BSSA		3.2 / -	5.7 / -	5.7 / -
MWF1		2.6 / -	5.4 / -	5.8 / -
MWF2		1.9 / -	5.3 / -	5.6 / -
MNMF		1.4 / 1.4	2.2 / 2.2	2.6 / 2.6
ILRMA+MNMF		1.5 / 1.5	5.4 / 5.4	6.2 / 6.2
FastMNMF		1.5 / 1.5	2.0 / 1.8	2.9 / 2.8
ILRMA+FastMNMF		1.4 / 1.4	4.9 / 4.9	5.4 / 5.4
Proposed method	$\rho = 0.5$	**4.0** / **4.0**	**7.8** / **7.5**	**8.3** / **7.8**
	$\rho = 1$	**4.0** / **4.0**	7.5 / 7.3	8.0 / 7.7
	$\rho = 2$	**4.0** / **4.0**	7.4 / 7.1	7.9 / 7.5
	$\rho = 4$	**4.0** / **4.0**	7.2 / 7.0	7.8 / 7.5
	$\rho = 8$	3.8 / 3.8	7.1 / 7.0	7.7 / 7.5

TABLE VII
SDR IMPROVEMENTS FOR EACH METHOD AND NUMBER OF MICROPHONES UNDER CAFE NOISE CONDITION. EACH TERM REPRESENTS "BEST-ITERATION SCORE [DB] / AFTER-200-ITERATION (EXCEPT FOR IVE; AFTER-4000-ITERATION FOR IVE) SCORE [DB]"

Methods		2 mics.	3 mics.	4 mics.
ILRMA		1.4 / -	5.7 / -	6.4 / -
IVE		1.5 / 1.5	5.7 / 5.5	6.4 / 5.9
BSSA		3.4 / -	6.5 / -	7.2 / -
MWF1		2.4 / -	6.3 / -	7.0 / -
MWF2		2.1 / -	6.7 / -	7.4 / -
MNMF		2.5 / 2.5	3.7 / 3.7	2.8 / 2.8
ILRMA+MNMF		1.6 / 1.6	7.3 / 7.3	8.1 / 8.1
FastMNMF		2.2 / 2.2	3.4 / 3.4	2.6 / 2.6
ILRMA+FastMNMF		1.8 / 1.8	6.5 / 6.5	7.3 / 7.3
Proposed method	$\rho = 0.5$	**4.1** / **4.0**	**8.6** / **7.7**	**10.1** / **9.3**
	$\rho = 1$	4.0 / **4.0**	8.3 / 7.3	9.9 / 9.2
	$\rho = 2$	4.0 / **4.0**	8.0 / 7.1	9.8 / 9.1
	$\rho = 4$	4.0 / **4.0**	7.7 / 7.0	9.7 / 9.0
	$\rho = 8$	3.9 / 3.9	7.5 / 6.9	9.6 / 9.0

full-rank SCM in the proposed method showed an improvement of more than 3 dB compared with the rank-1 spatial model in ILRMA that cannot express diffuse noise appropriately, and the efficacy of the proposed spatial model extension was confirmed. Also, we reveal that, even with the assistance of ILRMA-based initialization, the SDRs of the conventional MNMFs and FastMNMFs with the full-rank SCM cannot reach that of the proposed method. As the setting of the shape parameter ρ, $\rho = 0.5$ (super-Gaussian modeling) was the best for the three- and four-microphone cases. On the other hand, for the two-microphone case, $\rho = 1$ (super-Gaussian modeling), 2 (Gaussian modeling), and 4 (sub-Gaussian modeling) provided the best speech extraction performance.

Generally speaking, in MWF1 and MWF2, the postfilters try to predict the noise component in the direction of the target speech to increase the SDRs. However, the experimental results in Tables IV–VII show that the proposed method achieved better SDR improvement compared with MWF1 and MWF2. This may imply that the explicit statistical modeling of one lost spatial basis $\lambda_i \boldsymbol{b}_i \boldsymbol{b}_i^{\mathsf{H}}$ of the noise component enables us to estimate the full-rank SCM of noise to some extent and to better reduce the noise component in the direction of the target speech compared with the conventional methods.

Compared with the conventional methods based on the full-rank SCM, e.g., ILRMA+MNMF and ILRMA+FastMNMF, we can confirm that the proposed method provided better SDR improvement in Tables IV–VII. To reveal the reason, we analyzed the accuracy of estimation for SCMs of the directional target speech and diffuse noise. We calculated the LogDet divergence [47] between the estimated and true SCMs, where $M = 4$, diffuse noise was babble noise, the directional target speech was located normal to the microphone array, and the shape parameter ρ was set to 2. Here, we simulated source images of the target speech and diffuse noise and calculated their SCMs as true SCMs. LogDet divergence values of SCMs were summed over all $I = 513$ frequency bins after dividing SCMs by their trace values. Fig. 6 shows the LogDet divergence behaviors of SCMs along with iterations for ILRMA+MNMF, ILRMA+FastMNMF, and the proposed method. Although the conventional methods drift their SCMs, the proposed method can estimate the SCMs more accurately and consistently, especially for noise SCM (average divergence of noise SCMs in

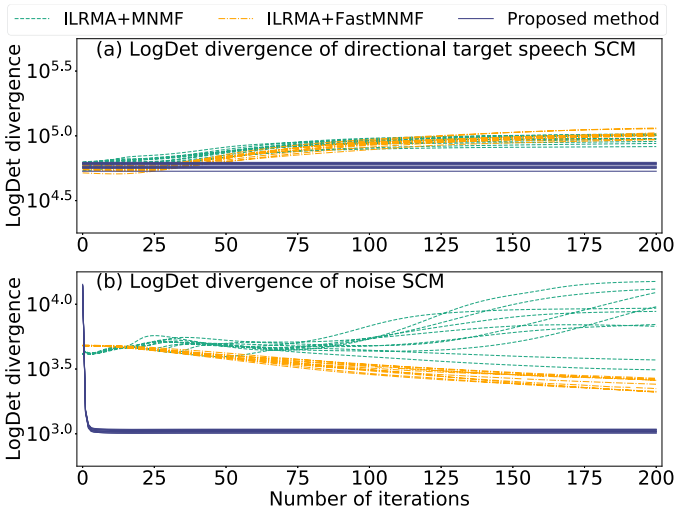

Fig. 6. LogDet divergence behaviors between estimated and true SCMs of (a) directional target speech and (b) noise in ILRMA+MNMF, ILRMA+FastMNMF, and proposed method with 10 parameter-initialization random seeds under babble noise condition.

each frequency is only $10^3/513 \approx 2$), which leads to better noise reduction.

D. Computational Time Comparison

To see the efficacy of the proposed acceleration method, we compared three methods: MNMF, FastMNMF, and the proposed rank-constrained SCM estimation. For the 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). 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 signal used in this experiment was 8.7 s long and was mixed in the manner described in Section V-A.

The average computational time of one iteration for each method is shown in Fig. 7. The proposed algorithm with the second-stage acceleration achieved fast computation with complexity independent of the number of microphones. In particular, under the four-microphone condition, the proposed second-stage acceleration with $\rho = 2$ was 92 times faster than the naive update rule and 10 times faster than FastMNMF. Also, the proposed second-stage acceleration with $\rho = 0.5$, which achieved the best SDR improvement, was 57 times faster than the naive update rule and 6.2 times faster than FastMNMF, showing a slight trade-off between SDR improvement and computational time.

VI. BSE Experiment on Real Recorded Data

To evaluate the proposed method in a more realistic situation, we recorded directional target speech and diffuse babble and traffic noises. Diffuse babble noise was recorded in a real room and diffuse traffic noise was recorded in outdoor space. A microphone array composed of four microphones with an interval of 3 cm was located in the room and the outdoor space. For the room, its reverberation time was about 400 ms and size was 3.5 m $\times$ 6.0 m $\times$ 3.3 m. For the outdoor space, its reverberation time was about 90 ms. As the directional target speech, the same dry source signal was emitted from a loudspeaker located 0°, 10°, 20°, or 30° clockwise from the normal to the microphone array and at a distance of 1.0 m. As diffuse babble noise, 10 people talked freely at a distance of about 1.5 m from the microphone array, and their voices were simultaneously recorded. An STFT was performed using a 256-ms-long Hamming window with a 32-ms-long shift. The other conditions for each method were the same as those described in Section V-A.

We compared seven methods, namely, ILRMA, BSSA, MWF1, MWF2, FastMNMF, ILRMA+FastMNMF, and rank-constrained SCM estimation with the ME-algorithm-based update and second-stage acceleration, where IVE and MNMF were omitted owing to their slow convergence and huge computational complexity. For rank-constrained SCM estimation, the hyperparameters were set to $\alpha = 0.1$ and $\beta = 10^{-16}$. Tables VIII and IX show the best-iteration SDR improvement and after-200-iteration SDR improvement for each method under babble and traffic noise conditions, respectively. The proposed method performed the best for each number of microphones and noise condition. Fig. 8 shows the SDR behaviors with respect to the elapsed time in the four-microphone and traffic noise case, where the shape parameter ρ of the GGD was set to 0.5. The proposed method achieved the most efficient speech extraction

TABLE VIII
SDR Improvements for Each Method and Number of Microphones Under Real Recorded Babble Noise Condition. Each Term Represents "Best-Iteration Score [dB] / After-200-Iteration Score [dB]"

Methods		2 mics.	3 mics.	4 mics.
ILRMA		1.7 / -	3.7 / -	5.1 / -
BSSA		3.1 / -	4.4 / -	5.4 / -
MWF1		2.0 / -	3.9 / -	5.4 / -
MWF2		2.1 / -	4.1 / -	5.6 / -
FastMNMF		0.1 / -0.2	0.0 / -1.8	0.2 / -2.8
ILRMA+FastMNMF		1.6 / 1.2	4.0 / 3.6	4.8 / 3.5
Proposed method	$\rho = 0.5$	4.1 / 3.7	**6.0** / 5.6	**7.3** / 6.7
	$\rho = 1$	**4.2** / 3.7	**6.0** / 5.6	**7.3** / 6.7
	$\rho = 2$	**4.2** / 3.7	**6.0** / 5.6	**7.3** / 6.7
	$\rho = 4$	**4.2** / 3.7	**6.0** / 5.6	**7.3** / 6.7
	$\rho = 8$	**4.2** / **3.8**	**6.0** / **5.7**	**7.3** / **6.8**

TABLE IX
SDR Improvements for Each Method and Number of Microphones Under Real Recorded Traffic Noise Condition. Each Term Represents "Best-Iteration Score [dB] / After-200-Iteration Score [dB]"

Methods		2 mics.	3 mics.	4 mics.
ILRMA		3.2 / -	5.3 / -	6.2 / -
BSSA		3.7 / -	5.2 / -	5.3 / -
MWF1		3.8 / -	6.0 / -	6.8 / -
MWF2		3.5 / -	5.5 / -	6.4 / -
FastMNMF		4.7 / 4.7	4.2 / 4.2	2.1 / 2.1
ILRMA+FastMNMF		4.3 / 4.3	6.5 / 6.5	8.5 / 8.4
Proposed method	$\rho = 0.5$	**7.9** / **7.9**	10.2 / 10.1	10.4 / 10.0
	$\rho = 1$	**7.9** / 7.8	**10.3** / 10.1	**10.5** / 10.0
	$\rho = 2$	**7.9** / 7.8	**10.3** / 10.1	**10.5** / 10.0
	$\rho = 4$	**7.9** / 7.8	**10.3** / **10.2**	**10.5** / 10.0
	$\rho = 8$	**7.9** / **7.9**	**10.3** / **10.2**	**10.5** / **10.2**

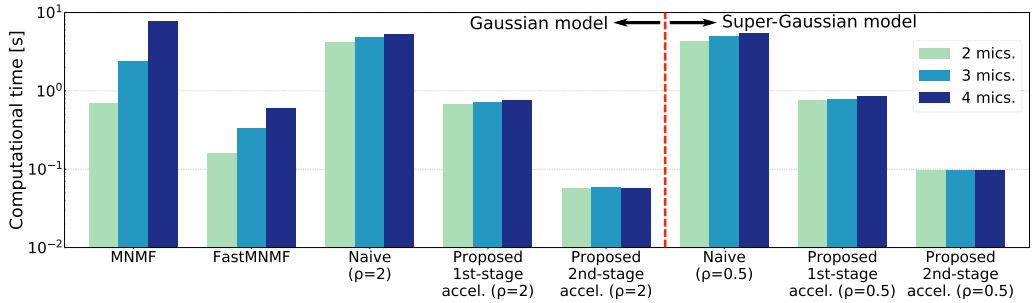

Fig. 7. Average computational time of one iteration for each method in cases of 2, 3, and 4 microphones under babble noise condition where directional target speech is located at 30°.

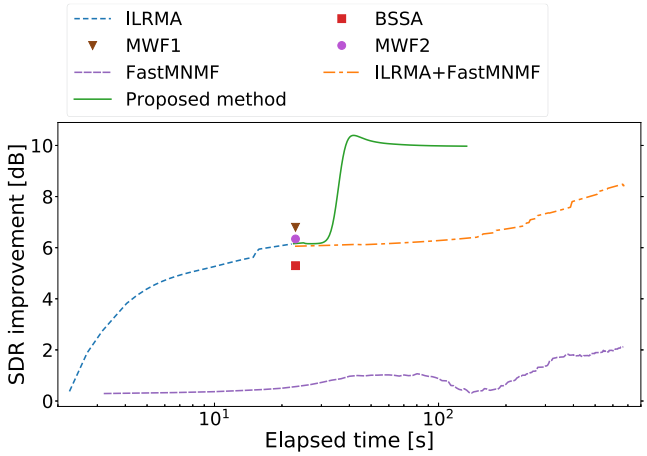

Fig. 8. SDR behaviors with respect to elapsed time averaged over 10 parameter-initialization random seeds, four target directions, and six speech sources.

while providing the highest SDR improvement. From these results, we confirmed the efficacy of the proposed method in realistic situations.

By comparing the results of the real and simulated noise cases, we can discuss when the proposed method can work well. For example, the best SDR improvement of the proposed method for simulated babble noise was 8.7 dB, while that for real recorded babble noise was 7.3 dB. Also, for traffic noise, the best SDR improvements were 8.3 dB (simulated) and 10.5 dB (real). Such a difference is due to the accuracy of the rank-$(M-1)$ noise SCM estimated by ILRMA. SIR improvement for each noise in Fig. 1 has the same tendency as the best SDR improvement of the proposed method for each noise. In the experiment using real recorded babble noise, the reverberation time was about 400 ms and the rank-1 spatial model assumption for the target speech may be violated (for traffic noise, the reverberation times were 200 ms (simulated) and 90 ms (real)). When the reverberation time greatly exceeds the STFT window size, the target speech no longer becomes a point source and the estimated rank-$(M-1)$ noise SCM becomes inaccurate, which leads to degradation of speech extraction performance of the proposed method. From these facts, whether the target speech can be regarded as a point source affects the performance of the proposed method.

VII. Conclusion

In this paper, we proposed a new BSE method to extract only the directional target speech from background diffuse noise. The proposed method utilizes ILRMA, which is based on the rank-1 spatial model, as a preprocessing method and further improves the extraction performance by restoring the lost spatial basis for the full-rank SCM of diffuse noise. We introduced the multivariate GGD into the statistical model of the observed signal and proved a new inequality to derive the MM-algorithm-based and ME-algorithm-based parameter update rules for an arbitrary shape parameter of the multivariate GGD. We also derived new acceleration algorithms for the proposed framework and realized a highly efficient update with its computational complexity independent of the number of microphones.

Through BSE experiments with simulated data, we confirmed that the proposed method with the ME-algorithm-based parameter update outperformed the conventional ILRMA, IVE, multichannel Wiener filter, MNMFs, and their combined methods in terms of SDR improvement, and the proposed method achieved faster computation than the conventional methods. Also, the applicability of the proposed method to realistic situations was shown by experiments with real recorded data.

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 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. Speech Audio Process.*, vol. 11, no. 2, pp. 109–116, Mar. 2003.

[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. Audio, Speech, Lang. Process.*, vol. 14, no. 2, pp. 666–678, Mar. 2006.

[5] A. Hiroe, "Solution of permutation problem in frequency domain ICA using multivariate probability density functions," in *Proc. 6th Int. Conf. Independent Compon. Anal. Blind Signal Separation*, 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. Audio, Speech, Lang. Process.*, vol. 15, no. 1, pp. 70–79, Jan. 2007.

[7] N. Ono, "Stable and fast update rules for independent vector analysis based on auxiliary function technique," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, 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. Audio, Speech, Lang. Process.*, vol. 24, no. 9, pp. 1626–1641, Sep. 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., Berlin, Germany: Springer, 2018, pp. 125–155.

[10] D. Kitamura *et al.*, "Generalized independent low-rank matrix analysis using heavy-tailed distributions for blind source separation," *EURASIP J. Adv. Signal Process.*, vol. 2018, no. 1, pp. 1–28, 2018.

[11] R. Ikeshita and Y. Kawaguchi, "Inpendent low-rank matrix analysis based on multivariate complex exponential power distribution," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2018, pp. 741–745.

[12] S. Mogami *et al.*, "Independent low-rank matrix analysis based on time-variant sub-Gaussian source model," in *Proc. Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf.*, 2018, pp. 1684–1691.

[13] S. Mogami *et al.*, "Independent low-rank matrix analysis based on generalized Kullback-Leibler divergence," *IEICE Trans. Fundam.*, vol. E102-A, no. 2, pp. 458–463, 2019.

[14] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[15] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. Neural Inf. Process. Syst.*, 2000, pp. 556–562.

[16] 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 J. Adv. Signal Process.*, vol. 2003, no. 11, pp. 1–10, 2003.

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

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

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

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

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

[22] N. Ito and T. Nakatani, "FastMNMF: Joint diagonalization based accelerated algorithms for multichannel nonnegative matrix factorization," in *Proc. Int. Conf. Acoust., Speech Signal Process.*, 2019, pp. 371–375.

[23] K. Sekiguchi, A. A. Nugraha, Y. Bando, and K. Yoshii, "Fast multichannel source separation based on jointly diagonalizable spatial covariance matrices," in *Proc. EUSIPCO*, 2019.

[24] K. Shimada, Y. Bando, M. Mimura, K. Itoyama, K. Yoshii, and T. Kawahara, "Unsupervised beamforming based on multichannel nonnegative matrix factorization for noisy speech recognition," in *Proc. Int. Conf. Acoust., Speech Signal Process.*, 2018, pp. 5734–5738.

[25] N. Ono, "Auxiliary-function-based independent vector analysis with power of vector-norm type weighting functions," in *Proc. Asia Pacific Signal Inf. Process. Assoc. Annu. Summit Conf.*, 2012, pp. 1–4.

[26] Y. Liang, J. Harris, S. Naqvi, G. Chen, and J. Chambers, "Independent vector analysis with a generalized multivariate Gaussian source prior for frequency domain blind source separation," *Signal Process.*, vol. 105, pp. 175–184, 2014.

[27] Z. Boukouvalas, G.-S. Fu, and T. Adalı, "An efficient multivariate generalized Gaussian distribution estimator: Application to IVA," in *Proc. Annu. Conf. Info. Sci. Syst.*, 2015.

[28] 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.

[29] Y. Kubo, N. Takamune, D. Kitamura, and H. Saruwatari, "Acceleration of rank-constrained spatial covariance matrix estimation for blind speech extraction," in *Proc. Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf.*, 2019, pp. 332–338.

[30] D. R. Hunter and K. Lange, "Quantile regression via an MM algorithm," *J. Comput. Graph. Stat.*, vol. 9, no. 1, pp. 60–77, 2000.

[31] C. Févotte and J. Idier, "Algorithms for nonnegative matrix factorization with the β-divergence," *Neural Comput.*, vol. 23, no. 9, pp. 2421–2456, 2011.

[32] C. Févotte, N. Bertin, and J.-L. Durrieu, "Nonnegative matrix factorization with the Itakura–Saito divergence: With application to music analysis," *Neural Comput.*, vol. 21, no. 3, pp. 793–830, 2009.

[33] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 4, pp. 1462–1469, Jul. 2006.

[34] Z. Koldovský and P. Tichavský, "Gradient algorithms for complex non-gaussian independent component/vector extraction, question of convergence," *IEEE Trans. Signal Process.*, vol. 67, no. 4, pp. 1050–1064, Feb. 2019.

[35] S. Doclo, S. Gannot, M. Moonen, and A. Spriet, "Acoustic beamforming for hearing aid applications," in *Handbook on Array Processing and Sensor Networks*, S. Haykin and K. J. R. Liu, Eds., Hoboken, NJ, USA: Wiley, 2010, pp. 269–302.

[36] H. Kameoka and K. Kashino, "Composite autoregressive system for sparse source-filter representation of speech," in *Proc. Int. Symp. Circuits Syst.*, 2009, pp. 2477–2480.

[37] 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.

[38] H. Kameoka, T. Nishimoto, and S. Sagayama, "A multipitch analyzer based on harmonic temporal structured clustering," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 3, pp. 982–994, Mar. 2007.

[39] Y. Mitsui, D. Kitamura, N. Takamune, H. Saruwatari, Y. Takahashi, and K. Kondo, "Independent low-rank matrix analysis based on parametric majorization-equalization algorithm," in *Proc. 7th Int. Workshop Comput. Advances Multi-Sensor Adaptive Process.*, 2017, pp. 98–102.

[40] I. Cohen and B. Berdugo, "Speech enhancement for non-stationary noise environments," *Signal Process.*, vol. 81, no. 11, pp. 2403–2418, 2001.

[41] R. Miyazaki, H. Saruwatari, R. Wakisaka, K. Shikano, and T. Takatani, "Theoretical analysis of parametric blind spatial subtraction array and its application to speech recognition performance prediction," in *Proc. Joint Workshop Hands-Free Speech Commun. Microphone Arrays*, 2011, pp. 19–24.

[42] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, no. 6, pp. 1109–1121, Dec. 1984.

[43] Y. Fujihara, Y. Takahashi, S. Miyabe, H. Saruwatari, K. Shikano, and A. Tanaka, "Performance improvement of higher-order ICA using learning period detection based on closed-form second-order ICA and kurtosis," in *Proc. Int. Workshop Acoust. Echo Noise Control*, 2008.

[44] K. Itou *et al.*, "JNAS: Japanese speech corpus for large vocabulary continuous speech recognition research," *J. Acoust. Soc. Jpn. (E)*, vol. 20, no. 3, pp. 199–206, 1999.

[45] J. Thiemann, N. Ito, and E. Vincent, "DEMAND: A collection of multi-channel recordings of acoustic noise in diverse environments," Zenodo, Jun. 2013, doi: 10.5281/zenodo.1227121.

[46] H. Saruwatari, Y. Ishikawa, Y. Takahashi, T. Inoue, K. Shikano, and K. Kondo, "Musical noise controllable algorithm of channelwise spectral subtraction and adaptive beamforming based on higher order statistics," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 6, pp. 1457–1466, Aug. 2011.

[47] B. Kulis, M. A. Sustik, and I. S. Dhillon, "Low-rank Kernel learning with bregman matrix divergences," *J. Mach. Learn. Res.*, vol. 10, pp. 341–376, Jun. 2009.