Deficient-basis-complementary rank-constrained spatial covariance matrix estimation based on multivariate generalized Gaussian distribution for blind speech extraction

Yuto Kondo[1], Yuki Kubo[1], Norihiro Takamune[1*], Daichi Kitamura[2] and Hiroshi Saruwatari[1]

*Correspondence:
Norihiro_Takamune@ipc.i.u-tokyo.ac.jp

[1] The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8654, Japan
[2] National Institute of Technology, Kagawa College, 355 Chokushi, Takamatsu, Kagawa 761-8058, Japan

Abstract

Rank-constrained spatial covariance matrix estimation (RCSCME) is a blind speech extraction method utilized under the condition that one-directional target speech and diffuse background noise are mixed. In this paper, we propose a new model extension of RCSCME. RCSCME simultaneously conducts both the deficient rank-1 component complementation of the diffuse noise spatial covariance matrix, which is incompletely estimated by preprocessing methods such as independent low-rank matrix analysis, and the estimation of the source model parameters. In the conventional RCSCME, between the two parameters constituting the deficient rank-1 component, only the scale is estimated, whereas the other parameter, the deficient basis, is fixed in advance; however, how to choose the fixed deficient basis is not unique. In the proposed RCSCME model, we also regard the deficient basis as a parameter to estimate. As the generative model of an observed signal, we utilized the super-Gaussian generalized Gaussian distribution, which achieves better separation performance than the Gaussian distribution in the conventional RCSCME. Assuming the model, we derive new majorization-minimization (MM)- and majorization-equalization (ME)-algorithm-based update rules for the deficient basis. In particular, among innumerable ME-algorithm-based update rules, we successfully find an ME-algorithm-based update rule with a mathematical proof supporting the fact that the step of the update rule is larger than that of the MM-algorithm-based update rule. We confirm that the proposed method outperforms conventional methods under several simulated noise conditions and a real noise condition.

Keywords: Blind speech extraction, Blind source separation, Diffuse noise, Spatial covariance matrix, Rank-constrained spatial covariance matrix estimation

1 Introduction

Blind speech extraction (BSE) is a technique of extracting a target speech signal from observed noisy mixture signals without any prior information, e.g., spatial locations of the target speech, noise sources, or microphones. BSE can be interpreted as a particular case of blind source separation (BSS) [1]; BSS is a more widely applicable technique that separates not only the target source but also other sources. We focus on the BSE problem for the special case that an observed noisy mixture consists of directional target speech and diffuse background noise. BSE methods can be utilized for many applications, e.g., hearing aid systems and automatic speech recognition [2, 3].

For a determined or overdetermined case (number of microphones $\geq$ number of point sources), high-performance BSS methods such as frequency-domain independent component analysis (FDICA) [4–6], independent vector analysis [7, 8], and independent low-rank matrix analysis (ILRMA) [9–11] have been proposed. These methods assume that the frequency-wise acoustic path from each source to microphones can be modeled by a single time-invariant vector parameter, which is called *the steering vector*. In this model, the rank of a spatial covariance matrix (SCM) [12] becomes unity. Thus, hereafter, we call these BSS methods *rank-1 methods*. Under diffuse noise conditions, a directional target source cannot be cleanly separated by rank-1 methods in principle [2], and it is contaminated with a diffuse noise component remaining in the same direction. This is because steering vectors are not suitable for representing the nondirectional noise transmission.

As opposed to rank-1 methods, multichannel nonnegative matrix factorization (MNMF) [13–15] can represent nondirectional sources because MNMF utilizes a full-rank SCM of each source. However, the estimation of the full-rank SCM has a huge computational cost and lacks robustness against the parameter initialization [9]. Hence, FastMNMF [16–18], which is a BSS method whose spatial model is more severely constrained than that of MNMF, has been proposed and achieves efficient optimization with lower computational cost. However, its BSS performance still depends on the parameter initialization. Since SCMs are assumed to be full-rank matrices in these models, we call these BSS methods *full-rank methods*.

To overcome the lack of representation ability of rank-1 methods and of robustness of full-rank methods, rank-constrained SCM estimation (RCSCME) [19] has been proposed, which explicitly models a mixture of directional target speech and diffuse background noise. Figure 1 shows the process flow of RCSCME. First, a rank-1 method such as ILRMA is utilized as a preprocess of RCSCME. From the rank-1 method, M separated signals are obtained; one includes target speech components contaminated with
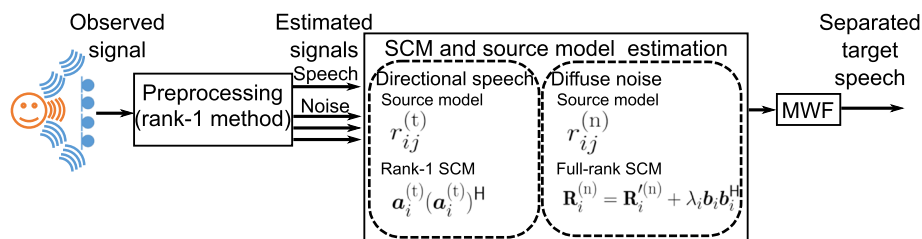

Fig. 1 Process flow of RCSCME

diffuse noise in the same direction and the other $M - 1$ signals consist of the only diffuse noise components in other directions [2], where M is the number of microphones. From these signals, useful spatial parameters, i.e., the steering vector of the directional speech and the rank-$(M - 1)$ component of the full-rank SCM of diffuse noise, are calculated. Subsequently, in the main part of RCSCME, both the deficient rank-1 component of the noise SCM and the source model parameters are estimated. Finally, the clean target speech signal can be obtained via multichannel Wiener filtering (MWF) constructed using the estimated spatial and source model parameters. Regarding speech extraction performance, it has been confirmed that RCSCME can outperform the above rank-1 methods [19]. Since the estimation of the deficient rank-1 component is valid and the number of parameters to estimate in RCSCME is much smaller than that in conventional full-rank methods, RCSCME also achieves better speech extraction performance than conventional full-rank methods.

In this work, we extend the spatial model of the conventional RCSCME. In the conventional RCSCME, the deficient rank-1 component of the diffuse noise SCM is represented by the scalar $\lambda \in \mathbb{R}_+$ and the direction vector $\boldsymbol{b} \in \mathbb{C}^M$ as $\lambda \boldsymbol{b} \boldsymbol{b}^{\mathsf{H}}$. For explanation, we refer to λ and $\boldsymbol{b}$ as *the scale* and *the deficient basis* of the deficient component, respectively. In the conventional RCSCME, the deficient basis $\boldsymbol{b}$ is fixed and only the scale λ is estimated. However, this deficient basis is not unique; any vector outside the space spanned by column vectors of the rank-$(M - 1)$ SCM is a possible candidate of the deficient basis. In the proposed method, we parameterize not only the scale but also the deficient basis itself to estimate the optimal full-rank noise SCM.

In many BSS methods, super-Gaussian distributions are often used as the generative model of an observed signal. For example, ILRMA based on the complex generalized Gaussian distribution (GGD) or the complex Student's t distribution [10, 11], MNMF based on the multivariate complex Student's t distribution [15], and FastMNMF based on the multivariate complex Student's t distribution [18] have been proposed. The complex GGD and the complex Student's t distribution are generalized versions of the complex Gaussian distribution and can represent many types of source. In the conventional RCSCME, the multivariate complex GGD is utilized as the generative model of the observed signal and the super-Gaussian multivariate complex GGD achieves better separation performance than the Gaussian distribution [19]. Then, in the proposed RCSCME model, we also use the super-Gaussian multivariate complex GGD as in the conventional RCSCME.

Assuming the GGD model, we derive new update rules of the deficient basis using auxiliary function techniques [20, 21] for the estimation. The proposed method is interpreted as the world's first *spatial* model extension of the conventional RCSCME; this extension had been considered as a difficult *vector* optimization problem, but we solve the problem with two types of auxiliary function technique, namely, the majorization-minimization (MM) algorithm [20] and majorization-equalization (ME) algorithm [21]. Whereas ME-algorithm-based update rules of scalar parameters are unique in many cases, those of vector parameters are innumerable. We find an ME-algorithm-based update rule of the deficient component. Additionally, we can successfully provide a mathematical proof supporting the fact that change in each target variable of the update rule is always larger than that of the MM-algorithm-based update rule. To the best of

our knowledge, regarding the scope of BSE methods, there has been no ME-algorithm-based *vector variable update rule* that has such a proof. The proof is the mathematical contribution of this paper.

The rest of this paper is organized as follows. In Sect. 2, we explain auxiliary function techniques and the conventional RCSCME. In Sect. 3, we propose a new model and derive MM- and ME-algorithm-based update rules. Additionally, we provide the proof supporting the advantage of the proposed ME-algorithm-based update rule over the proposed MM-algorithm-based update rule. In Sect. 4, we show the results of the experiments under simulated and real noise conditions. Finally, conclusions are presented in Sect. 5. Note that this paper is partially based on an international conference paper [22] we wrote. The major new contribution of this paper is that, whereas we derive an EM-algorithm-based update rule with the generative model using the multivariate complex Gaussian distribution in [22], we derive other MM- and ME-algorithm-based update rules with the generative model using the GGD because it is difficult to apply the EM algorithm to the GGD. Furthermore, in this paper, we provide a new mathematical proof supporting the fact that the step of the ME-algorithm-based update rule is always larger than that of the MM-algorithm-based update rule. We also present experiments conducted not only under simulated noise conditions but also a real noise condition.

2 Conventional RCSCME

2.1 Auxiliary function technique [20, 21]

In this section, we describe auxiliary function techniques, which are iterative optimization algorithms utilized in many BSS methods including the conventional RCSCME. Auxiliary function techniques are often used for the optimization problems that are difficult to solved directly.

We explain two types of auxiliary function technique, namely, the MM algorithm [20] and the ME algorithm [21]. Let Θ be a set of parameters of the objective function $\mathcal{F}$ and consider the optimization problem $\min_{\Theta} \mathcal{F}(\Theta)$. This technique uses the auxiliary function $\mathcal{F}^{\mathrm{U}}(\Theta, \Omega)$ that satisfies the following conditions:

(I)   It holds that $\mathcal{F}(\Theta) \leq \mathcal{F}^{\mathrm{U}}(\Theta, \Omega)$ for any Θ and Ω.
(II) For any Θ, there exists Ω such that $\mathcal{F}(\Theta) = \mathcal{F}^{\mathrm{U}}(\Theta, \Omega)$ holds.

Here, Ω is the set of auxiliary variables. Instead of the direct optimization of the objective function $\mathcal{F}(\Theta)$, Θ and Ω in the auxiliary function $\mathcal{F}^{\mathrm{U}}(\Theta, \Omega)$ are alternatively updated as follows. First, $\mathcal{F}^{\mathrm{U}}$ is minimized with respect to Ω as

$$\Omega^{(l+1)} \leftarrow \underset{\Omega}{\arg\min} \, \mathcal{F}^{\mathrm{U}}(\Theta^{(l)}, \Omega), \tag{1}$$

where $\Theta^{(l)}$ and $\Omega^{(l)}$ are the sets of parameters and auxiliary variables after the lth iteration, respectively. (1) is equivalent to the update to the set of auxiliary variables that satisfy condition (II). Second, in the MM algorithm, Θ is updated as

$$\Theta^{(l+1)} \leftarrow \underset{\Theta}{\arg\min} \, \mathcal{F}^{\mathrm{U}}(\Theta, \Omega^{(l+1)}). \tag{2}$$

On the other hand, in the ME algorithm, we discover a set of parameters $\tilde{\Theta}(\neq \Theta^{(l)})$ that satisfies

$$\mathcal{F}^{\mathrm{U}}(\tilde{\Theta}, \Omega^{(l+1)}) = \mathcal{F}^{\mathrm{U}}(\Theta^{(l)}, \Omega^{(l+1)}), \tag{3}$$

and instead of (2), Θ is updated as

$$\Theta^{(l+1)} \leftarrow \tilde{\Theta}. \tag{4}$$

The common advantage of the MM and ME algorithms is that the update rules guarantee a monotonic nonincrease in the objective function [20, 21]. In many cases, we design the auxiliary function to be convex for each variable. In such cases, for scalar variables, the ME-algorithm-based update rule is unique and always takes a larger step in the variable for each iteration than the MM-algorithm-based update rule, and the convergence of the ME algorithm is experimentally confirmed to be faster than that of the MM algorithm [21]. Note that if we design the nonconvex auxiliary function, the above-mentioned advantage is not always guaranteed.

2.2 Generative model and update rules of RCSCME [19]

In this section, we explain the conventional RCSCME. Let $\boldsymbol{x}_{ij} \in \mathbb{C}^M$ be the observed M-channel vector obtained by a short-time Fourier transform (STFT), where $i = 1, 2, \ldots, I$ and $j = 1, 2, \ldots, J$ are the indices of frequency bins and time frames, respectively. The generative model of the observed signal $\boldsymbol{x}_{ij}$ is defined using the zero-mean circularly symmetric multivariate GGD [11, 23] as

$$p(\boldsymbol{x}_{ij}; \boldsymbol{0}, \mathbf{R}_{ij}^{(\mathrm{x})}, \rho) = \frac{\Gamma(1 + M) \exp\left(-(\boldsymbol{x}_{ij}^{\mathsf{H}}(\mathbf{R}_{ij}^{(\mathrm{x})})^{-1}\boldsymbol{x}_{ij})^{\frac{\rho}{2}}\right)}{\pi^M \Gamma(1 + \frac{2M}{\rho}) \det \mathbf{R}_{ij}^{(\mathrm{x})}}, \tag{5}$$

where $\boldsymbol{0} \in \mathbb{C}^M$ is the zero vector, $\rho \in \mathbb{R}_+$ is the shape parameter of the GGD, $\mathbf{R}_{ij}^{(\mathrm{x})} \in \mathbb{C}^{M \times M}$ is the full-rank SCM of the observed signal, $\Gamma(\cdot)$ is the gamma function, and $^{\mathsf{H}}$ denotes the Hermitian transpose. The GGD can have the properties of Gaussian ($\rho = 2$), super-Gaussian ($\rho < 2$), and sub-Gaussian ($\rho > 2$). In particular, we discuss the Gaussian and super-Gaussian cases ($\rho \leq 2$). The SCM of the observed signal $\mathbf{R}_{ij}^{(\mathrm{x})}$ is modeled as the sum of the SCM of the directional target speech and that of diffuse noise as

$$\mathbf{R}_{ij}^{(\mathrm{x})} = r_{ij}^{(\mathrm{t})}\boldsymbol{a}_i^{(\mathrm{t})}(\boldsymbol{a}_i^{(\mathrm{t})})^{\mathsf{H}} + r_{ij}^{(\mathrm{n})}\mathbf{R}_i^{(\mathrm{n})}, \tag{6}$$

where $r_{ij}^{(\mathrm{t})}, r_{ij}^{(\mathrm{n})} \in \mathbb{R}_+$ are the time-variant variances of the directional target speech and diffuse noise, respectively, $\boldsymbol{a}_i^{(\mathrm{t})} \in \mathbb{C}^M$ is the steering vector of the target speech, i.e., the n_{t}th vector among the steering vectors $\boldsymbol{a}_{i,1}, \ldots, \boldsymbol{a}_{i,M}$ obtained by the preprocessing rank-1 method, with n_{t} being the index of the directional target speech, and $\mathbf{R}_i^{(\mathrm{n})}$ is the full-rank SCM of diffuse noise.

For the directional target speech, since the power spectrogram of the speech signal has the property of sparsity, we assume that $r_{ij}^{(\mathrm{t})}$ follows the inverse gamma distribution as the prior:

Kondo *et al. EURASIP Journal on Advances in Signal Processing* (2022) 2022:88

Page 6 of 24

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

where $\alpha \in \mathbb{R}_+$ is the shape parameter, and $\beta \in \mathbb{R}_+$ is the scale parameter. On the other hand, the full-rank SCM of diffuse noise $\mathbf{R}_i^{(n)}$ is modeled as

$$\mathbf{R}_i^{(n)} = \mathbf{R}_i'^{(n)} + \lambda_i \boldsymbol{b}_i \boldsymbol{b}_i^{\mathsf{H}}, \tag{8}$$

where $\mathbf{R}_i'^{(n)} \in \mathbb{C}^{M \times M}$ is the rank-$(M-1)$ SCM calculated by the rank-1 method in advance as

$$\mathbf{R}_i'^{(n)} = \frac{1}{J} \sum_j \hat{\boldsymbol{y}}_{ij}^{(n)} (\hat{\boldsymbol{y}}_{ij}^{(n)})^{\mathsf{H}}, \tag{9}$$

$$\hat{\boldsymbol{y}}_{ij}^{(n)} = \mathbf{W}_i^{-1} (\boldsymbol{w}_{i,1}^{\mathsf{H}} \boldsymbol{x}_{ij}, \dots, \boldsymbol{w}_{i,n_t-1}^{\mathsf{H}} \boldsymbol{x}_{ij}, 0, \boldsymbol{w}_{i,n_t+1}^{\mathsf{H}} \boldsymbol{x}_{ij}, \dots, \boldsymbol{w}_{i,M}^{\mathsf{H}} \boldsymbol{x}_{ij})^{\mathsf{T}}, \tag{10}$$

$$\mathbf{W}_i = (\boldsymbol{w}_{i,1}, \dots, \boldsymbol{w}_{i,m}, \dots, \boldsymbol{w}_{i,M})^{\mathsf{H}}. \tag{11}$$

Here, $\boldsymbol{w}_{i,m}$ is the demixing filter estimated by the rank-1 method, $\hat{\boldsymbol{y}}_{ij}^{(n)}$ is the sum of diffuse noise components whose scales are modified by a projection-back operation [24], and $^{\mathsf{T}}$ denotes the transpose. $\boldsymbol{b}_i \in \mathbb{C}^M$ in (8) is the deficient basis, which is introduced to make the SCM $\mathbf{R}_i^{(n)}$ full-rank, and $\lambda_i \in \mathbb{R}_+$ represents the scale of the deficient component. In the conventional RCSCME, the deficient basis $\boldsymbol{b}_i$ is fixed and only the scale λ_i is estimated. One possible candidate of the fixed vector $\boldsymbol{b}_i$ is an eigenvector of the zero eigenvalue in $\mathbf{R}_i'^{(n)}$.

In the conventional RCSCME [19], the parameters $\Theta_c = \{r_{ij}^{(t)}, r_{ij}^{(n)}, \lambda_i\}$ are estimated using maximum a posteriori estimation. The cost function is the following negative log posterior:

$$\mathcal{L}(\Theta_c) = \sum_{i,j} \left[(\boldsymbol{x}_{ij}^{\mathsf{H}} (\mathbf{R}_{ij}^{(x)})^{-1} \boldsymbol{x}_{ij})^{\frac{\rho}{2}} + \log \det \mathbf{R}_{ij}^{(x)} \right. \\ \left. + (\alpha + 1) \log r_{ij}^{(t)} + \frac{\beta}{r_{ij}^{(t)}} \right] + \text{const.}, \tag{12}$$

where const. includes the terms independent of Θ_c.

The parameters are estimated by designing an auxiliary function of $\mathcal{L}$ and optimizing the auxiliary function as written in [19]. For the Gaussian or super-Gaussian case ($\rho \leq 2$), the update rules are derived as

$$r_{ij}^{(t)} \leftarrow r_{ij}^{(t)} \left(\frac{\kappa_{ij} |\boldsymbol{x}_{ij}^{\mathsf{H}} (\mathbf{R}_{ij}^{(x)})^{-1} \boldsymbol{a}_i^{(t)}|^2 + \frac{\beta}{(r_{ij}^{(t)})^2}}{(\boldsymbol{a}_i^{(t)})^{\mathsf{H}} (\mathbf{R}_{ij}^{(x)})^{-1} \boldsymbol{a}_i^{(t)} + \frac{\alpha+1}{r_{ij}^{(t)}}} \right)^q, \tag{13}$$

$$r_{ij}^{(n)} \leftarrow r_{ij}^{(n)} \left(\frac{\kappa_{ij} \boldsymbol{x}_{ij}^{\mathsf{H}} (\mathbf{R}_{ij}^{(x)})^{-1} \mathbf{R}_i^{(n)} (\mathbf{R}_{ij}^{(x)})^{-1} \boldsymbol{x}_{ij}}{\text{tr}((\mathbf{R}_{ij}^{(x)})^{-1} \mathbf{R}_i^{(n)})} \right)^q, \tag{14}$$

$$\lambda_i \leftarrow \lambda_i \left(\frac{\sum_j \kappa_{ij} r_{ij}^{(n)} |\boldsymbol{b}_i^{\mathsf{H}} (\mathbf{R}_{ij}^{(x)})^{-1} \boldsymbol{x}_{ij}|^2}{\sum_j r_{ij}^{(n)} \boldsymbol{b}_i^{\mathsf{H}} (\mathbf{R}_{ij}^{(x)})^{-1} \boldsymbol{b}_i} \right)^q, \tag{15}$$

where q equals 1/2 for the MM-algorithm-based update rules and 1 for the ME-algorithm-based update rules, and

$$\kappa_{ij} = \frac{\rho}{2(\boldsymbol{x}_{ij}^{\mathsf{H}} (\mathbf{R}_{ij}^{(x)})^{-1} \boldsymbol{x}_{ij})^{1 - \frac{\rho}{2}}}. \tag{16}$$

It can be seen from (13)–(15) that the ME-algorithm-based update rules take larger steps in each iteration than the MM-algorithm-based update rules because q is larger in the ME-algorithm-based update rules than in the MM-algorithm-based update rules. It is experimentally confirmed in [19] that the ME-algorithm-based update rules achieve faster convergence than the MM-algorithm-based update rules.

3 Proposed basis-optimizing RCSCME

3.1 Motivation

In the conventional RCSCME, for spatial parameters, the deficient basis $\boldsymbol{b}_i$ is fixed and only the scale λ_i is estimated. However, how to choose the fixed deficient basis is not unique because the possible candidate of the deficient basis is any complex vector that is not included in the $(M - 1)$-dimensional hyperplane spanned by column vectors of $\mathbf{R}_i^{\prime(n)}$. In this work, we propose a new optimization scheme of RCSCME, *the basis-optimizing RCSCME*, whereas we call the conventional RCSCME *the fixed-basis RCSCME*. The theoretical assumptions of the proposed basis-optimizing RCSCME are the same as that of the fixed-basis RCSCME and are as follows [19]:

- Target speech source
  - spatial assumption: one point source
  - statistical assumption: a sparse power spectrogram
- Noise
  - spatial assumption: diffuse source

These assumptions are valid in many acoustic applications such as hearing aid systems and automatic speech recognition [2, 3]. The proposed basis-optimizing RCSCME estimates not only the scale but also the deficient basis itself. From the scale ambiguity between the scale and the deficient basis, it is natural for the propos6ed basis-optimizing RCSCME to parameterize the deficient basis and the scale with one variable simultaneously as

$$\mathbf{R}_i^{(n)} = \mathbf{R}_i^{\prime(n)} + \boldsymbol{c}_i \boldsymbol{c}_i^{\mathsf{H}}, \tag{17}$$

where $\boldsymbol{c}_i \in \mathbb{C}^M$ is the vector parameter that represents the deficient component in $\mathbf{R}_i^{\prime(n)}$. That is, we regard $\sqrt{\lambda_i} \boldsymbol{b}_i$ in the fixed-basis RCSCME as $\boldsymbol{c}_i$ in the basis-optimizing RCSCME. We model $r_{ij}^{(t)}$ and $r_{ij}^{(n)}$ in the same manner as in the conventional fixed-basis RCSCME.

Kondo *et al. EURASIP Journal on Advances in Signal Processing* (2022) 2022:88

Page 8 of 24

We apply the MM and ME algorithms to the estimation of the deficient basis. To derive MM-algorithm-based update rules, there remains a difficult vector optimization problem to minimize the auxiliary function. In addition, applying the ME algorithm to the vector parameter has a difficulty different from the case of applying the MM algorithm; the possible candidates of the ME-algorithm-based update rule of a vector parameter are innumerable because the problem finding an ME-algorithm-based update rule has only one equation but M variables. There can be ME-algorithm-based update rules whose steps are unfortunately smaller than the step of the MM-algorithm-based update rule, i.e., such an inappropriate ME algorithm updates the parameters only around the neighborhoods of the pre-update vector parameter, which result in slow convergence. In this paper, we find a specific ME-algorithm-based update rule of the deficient component and provide a mathematical proof supporting the fact that the step of the update rule is always larger than that of the MM-algorithm-based update rule.

3.2 Design of auxiliary function and derivation of MM-algorithm-based update rule for deficient component

In the proposed basis-optimizing RCSCME, we estimate $\Theta_{\mathrm{p}} = \{r_{ij}^{(\mathrm{t})}, r_{ij}^{(\mathrm{n})}, \boldsymbol{c}_i\}$. The negative log posterior $\mathcal{L}(\Theta_{\mathrm{p}})$ is the same formula as (12), where it is notable that $\mathbf{R}_i^{(\mathrm{n})}$ in $\mathbf{R}_{ij}^{(\mathrm{x})}$ is expressed as (17) in the proposed basis-optimizing RCSCME, whereas that is expressed as (8) in the conventional fixed-basis RCSCME.

Using the inequalities that are proposed in [19] and hold in the Gaussian and super-Gaussian cases ($\rho \leq 2$), we can design the following auxiliary function $\mathcal{L}^{\mathrm{U}}$ (see "Appendix" for detail derivation):

$$
\begin{aligned}
\mathcal{L}(\Theta_{\mathrm{p}}) \leq \sum_{i,j} \Bigg[&\frac{\rho}{2\iota_{ij}^{1-\frac{\rho}{2}}} \left(\frac{|(\boldsymbol{a}_i^{(\mathrm{t})})^{\mathsf{H}} \boldsymbol{\Phi}_{ij}^{(\mathrm{t})} \boldsymbol{x}_{ij}|^2}{r_{ij}^{(\mathrm{t})} \|\boldsymbol{a}_i^{(\mathrm{t})}\|^4} + \frac{|\boldsymbol{u}_i^{\mathsf{H}} \boldsymbol{\Phi}_{ij}^{(\mathrm{n})} \boldsymbol{x}_{ij}|^2}{r_{ij}^{(\mathrm{n})} |\boldsymbol{u}_i^{\mathsf{H}} \boldsymbol{c}_i|^2} \right. \\
&+ \left. \frac{\boldsymbol{x}_{ij}^{\mathsf{H}} (\boldsymbol{\Phi}_{ij}^{(\mathrm{n})})^{\mathsf{H}} \breve{\mathbf{R}}_i^{(\mathrm{n})} \boldsymbol{\Phi}_{ij}^{(\mathrm{n})} \boldsymbol{x}_{ij}}{r_{ij}^{(\mathrm{n})}} \right) + \left(1 - \frac{\rho}{2}\right) \iota_{ij}^{\frac{\rho}{2}} \\
&+ \mathrm{tr}(\boldsymbol{\Psi}_{ij}^{-1} \mathbf{R}_{ij}^{(\mathrm{x})}) + \log \det \boldsymbol{\Psi}_{ij} - M \\
&+ (\alpha + 1) \frac{r_{ij}^{(\mathrm{t})}}{\zeta_{ij}} + (\alpha + 1)(\log \zeta_{ij} - 1) + \frac{\beta}{r_{ij}^{(\mathrm{t})}} \Bigg] \\
=: &\,\mathcal{L}^{\mathrm{U}}(\Theta_{\mathrm{p}}, \Omega_{\mathrm{p}}),
\end{aligned}
\tag{18}
$$

where $\boldsymbol{u}_i \in \mathbb{C}^M$ is an eigenvector of the zero eigenvalue in $\mathbf{R}_i^{\prime(\mathrm{n})}$, and $\breve{\mathbf{R}}_i^{(\mathrm{n})}$ is the matrix defined as

$$
\breve{\mathbf{R}}_i^{(\mathrm{n})} := \left(\mathbf{E} - \frac{\boldsymbol{u}_i \boldsymbol{c}_i^{\mathsf{H}}}{\boldsymbol{c}_i^{\mathsf{H}} \boldsymbol{u}_i} \right) (\mathbf{R}_i^{\prime(\mathrm{n})})^+ \left(\mathbf{E} - \frac{\boldsymbol{c}_i \boldsymbol{u}_i^{\mathsf{H}}}{\boldsymbol{u}_i^{\mathsf{H}} \boldsymbol{c}_i} \right),
\tag{19}
$$

with $(\mathbf{R}_i^{\prime(\mathrm{n})})^+$ being the Moore–Penrose inverse matrix of $\mathbf{R}_i^{\prime(\mathrm{n})}$, $\mathbf{E} \in \mathbb{R}^{M \times M}$ is the identity matrix, $\Omega_{\mathrm{p}} = \{\boldsymbol{\Phi}_{ij}^{(\mathrm{t})}, \boldsymbol{\Phi}_{ij}^{(\mathrm{n})}, \boldsymbol{\Psi}_{ij}, \iota_{ij}, \zeta_{ij}\}$ is the set of auxiliary variables, $\boldsymbol{\Phi}_{ij}^{(\mathrm{t})}, \boldsymbol{\Phi}_{ij}^{(\mathrm{n})} \in \mathbb{C}^{M \times M}$ are matrices that satisfy $\boldsymbol{\Phi}_{ij}^{(\mathrm{t})} + \boldsymbol{\Phi}_{ij}^{(\mathrm{n})} = \boldsymbol{x}_{ij} \boldsymbol{x}_{ij}^{\mathsf{H}} / \|\boldsymbol{x}_{ij}\|^2$, $\boldsymbol{\Psi}_{ij} \in \mathbb{C}^{M \times M}$ is a positive semidefinite matrix, and $\iota_{ij}, \zeta_{ij} \in \mathbb{R}$ are positive. The equality of (18) holds if and only if

$$\boldsymbol{\Phi}_{ij}^{(t)} = r_{ij}^{(t)} \boldsymbol{a}_i^{(t)} (\boldsymbol{a}_i^{(t)})^{\mathsf{H}} (\mathbf{R}_{ij}^{(x)})^{-1} \frac{\boldsymbol{x}_{ij} \boldsymbol{x}_{ij}^{\mathsf{H}}}{\|\boldsymbol{x}_{ij}\|^2}, \tag{20}$$

$$\boldsymbol{\Phi}_{ij}^{(n)} = r_{ij}^{(n)} \mathbf{R}_i^{(n)} (\mathbf{R}_{ij}^{(x)})^{-1} \frac{\boldsymbol{x}_{ij} \boldsymbol{x}_{ij}^{\mathsf{H}}}{\|\boldsymbol{x}_{ij}\|^2}, \tag{21}$$

$$\boldsymbol{\Psi}_{ij} = \mathbf{R}_{ij}^{(x)}, \tag{22}$$

$$
\begin{aligned}
\iota_{ij} = {} & \frac{|(\boldsymbol{a}_i^{(t)})^{\mathsf{H}} \boldsymbol{\Phi}_{ij}^{(t)} \boldsymbol{x}_{ij}|^2}{r_{ij}^{(t)} \|\boldsymbol{a}_i^{(t)}\|^4} + \frac{|\boldsymbol{u}_i^{\mathsf{H}} \boldsymbol{\Phi}_{ij}^{(n)} \boldsymbol{x}_{ij}|^2}{r_{ij}^{(n)} |\boldsymbol{u}_i^{\mathsf{H}} \boldsymbol{c}_i|^2} \\
& + \frac{\boldsymbol{x}_{ij}^{\mathsf{H}} (\boldsymbol{\Phi}_{ij}^{(n)})^{\mathsf{H}} \check{\mathbf{R}}_i^{(n)} \boldsymbol{\Phi}_{ij}^{(n)} \boldsymbol{x}_{ij}}{r_{ij}^{(n)}},
\end{aligned}
\tag{23}
$$

$$\zeta_{ij} = r_{ij}^{(t)}. \tag{24}$$

Note that whereas $\|\boldsymbol{a}_i^{(t)}\| = 1$ and $\boldsymbol{b}_i^{\mathsf{H}} \boldsymbol{u}_i = 1$ are assumed in [19] for simplifying the derivation of the update rules, we remove these restrictions and recalculate the auxiliary function. Thus, if we replace $\|\boldsymbol{a}_i^{(t)}\|$ with 1 and $\boldsymbol{c}_i$ with $\sqrt{\lambda_i} \boldsymbol{b}_i$ and assume $\boldsymbol{b}_i^{\mathsf{H}} \boldsymbol{u}_i = 1$, $\mathcal{L}^{\mathsf{U}}$ coincides with the auxiliary function written in [19].

By minimizing the auxiliary function $\mathcal{L}^{\mathsf{U}}$ with respect to Θ_{p}, we derive MM-algorithm-based update rules. We describe in detail the derivation of the update rule of $\boldsymbol{c}_i$ since the MM-algorithm-based update rules of $r_{ij}^{(t)}$ and $r_{ij}^{(n)}$ are the same as those in the conventional fixed-basis RCSCME. We try to obtain the update rule of $\boldsymbol{c}_i$ by finding the stationary point of $\mathcal{L}^{\mathsf{U}}$ for $\boldsymbol{c}_i$. However, the analytical calculation of the stationary point is difficult because the term $\boldsymbol{c}_i \boldsymbol{u}_i^{\mathsf{H}} / \boldsymbol{u}_i^{\mathsf{H}} \boldsymbol{c}_i$ in (19) has $\boldsymbol{c}_i$ in both the numerator and the denominator. By paying attention to the invariance of the term $\boldsymbol{c}_i \boldsymbol{u}_i^{\mathsf{H}} / \boldsymbol{u}_i^{\mathsf{H}} \boldsymbol{c}_i$ for the scale of $\boldsymbol{c}_i$, we again resolve $\boldsymbol{c}_i$ into the scalar λ_i and the vector $\boldsymbol{b}_i$ as $\boldsymbol{c}_i = \sqrt{\lambda_i} \boldsymbol{b}_i$ and restrict the vector $\boldsymbol{b}_i$ to the hyperplane $\boldsymbol{b}_i^{\mathsf{H}} \boldsymbol{u}_i = 1$[1]. Then, the optimization problem of $\boldsymbol{c}_i$ can be reconsidered as the optimization of λ_i and $\boldsymbol{b}_i$. That is, Θ_{p} is redefined as $\Theta_{\mathsf{p}} = \{r_{ij}^{(t)}, r_{ij}^{(n)}, \lambda_i, \boldsymbol{b}_i\}$; $\boldsymbol{b}_i$ is also variable in this paper while only $\Theta_{\mathsf{c}} = \{r_{ij}^{(t)}, r_{ij}^{(n)}, \lambda_i\}$ is the set of variables in the conventional fixed-basis RCSCME. We focus on the derivation of the update rule for $\boldsymbol{b}_i$ because the derivation for λ_i is the same as that in the conventional fixed-basis RCSCME when the other parameters $r_{ij}^{(t)}$, $r_{ij}^{(n)}$, and $\boldsymbol{b}_i$ are fixed. Using $\boldsymbol{c}_i \boldsymbol{u}_i^{\mathsf{H}} / \boldsymbol{u}_i^{\mathsf{H}} \boldsymbol{c}_i = \boldsymbol{b}_i \boldsymbol{u}_i^{\mathsf{H}}$, we simply express $\mathcal{L}^{\mathsf{U}}$ as

$$
\begin{aligned}
\mathcal{L}^{\mathsf{U}} = {} & \sum_i \left[\boldsymbol{b}_i^{\mathsf{H}} \mathbf{G}_i \boldsymbol{b}_i - \boldsymbol{h}_i^{\mathsf{H}} \boldsymbol{b}_i - \boldsymbol{b}_i^{\mathsf{H}} \boldsymbol{h}_i \right] + \text{const.} \\
& \text{s.t. } \boldsymbol{b}_i^{\mathsf{H}} \boldsymbol{u}_i = 1,
\end{aligned}
\tag{25}
$$

where const. includes the terms independent of $\boldsymbol{b}_i$, and

---

[1] In [19], the restriction $\boldsymbol{b}_i^{\mathsf{H}} \boldsymbol{u}_i = 1$ is introduced for simplifying the derivation of the update rules. In this paper, this restriction is necessary for obtaining the stationary point of the auxiliary function as a closed-form solution.

$$\mathbf{G}_i = \sum_j \left[\frac{\rho}{2\iota_{ij}^{1-\frac{\rho}{2}}} \frac{|\boldsymbol{u}_i^{\mathsf{H}} \boldsymbol{\Phi}_{ij}^{(\mathrm{n})} \boldsymbol{x}_{ij}|^2}{r_{ij}^{(\mathrm{n})}} (\mathbf{R}_i^{\prime(n)})^+ \right.$$
$$\left. + \lambda_i r_{ij}^{(\mathrm{n})} \boldsymbol{\Psi}_{ij}^{-1} \right], \tag{26}$$

$$\boldsymbol{h}_i = (\mathbf{R}_i^{\prime(n)})^+ \sum_j \frac{\rho}{2\iota_{ij}^{1-\frac{\rho}{2}}} \frac{\boldsymbol{\Phi}_{ij}^{(\mathrm{n})} \boldsymbol{x}_{ij} \boldsymbol{x}_{ij}^{\mathsf{H}} (\boldsymbol{\Phi}_{ij}^{(\mathrm{n})})^{\mathsf{H}}}{r_{ij}^{(\mathrm{n})}} \boldsymbol{u}_i. \tag{27}$$

For optimizing (25), we can use the method of Lagrange multipliers. The Lagrangian is defined as

$$\mathcal{L}^{\mathrm{L}} = \mathcal{L}^{\mathrm{U}} + \sum_i \left(\eta_i (\boldsymbol{b}_i^{\mathsf{H}} \boldsymbol{u}_i - 1) + \eta_i^* (\boldsymbol{u}_i^{\mathsf{H}} \boldsymbol{b}_i - 1) \right), \tag{28}$$

where η_i is the Lagrangian multiplier and η^* denotes the complex conjugate of η. From $\partial \mathcal{L}^{\mathrm{L}} / \partial \boldsymbol{b}_i^* = 0$ and $\partial \mathcal{L}^{\mathrm{L}} / \partial \eta_i^* = 0$, we have

$$\mathbf{G}_i \boldsymbol{b}_i - \boldsymbol{h}_i + \eta_i \boldsymbol{u}_i = 0, \tag{29}$$

$$\boldsymbol{u}_i^{\mathsf{H}} \boldsymbol{b}_i - 1 = 0. \tag{30}$$

By solving (29) and (30), we derive the MM-algorithm-based update rule of $\boldsymbol{b}_i$ as

$$\boldsymbol{b}_i^{(\mathrm{MM})} = \mathbf{G}_i^{-1} \boldsymbol{h}_i - \frac{\boldsymbol{u}_i^{\mathsf{H}} \mathbf{G}_i^{-1} \boldsymbol{h}_i - 1}{\boldsymbol{u}_i^{\mathsf{H}} \mathbf{G}_i^{-1} \boldsymbol{u}_i} \mathbf{G}_i^{-1} \boldsymbol{u}_i. \tag{31}$$

The update rules of all parameters other than $\boldsymbol{b}_i$ can be obtained in the same manner as those in the conventional fixed-basis RCSCME. By substituting (20)–(24) in (31) and rearranging the formula, we can obtain the MM-algorithm-based update rules as follows:

$$r_{ij}^{(\mathrm{t})} \leftarrow r_{ij}^{(\mathrm{t})} \sqrt{\frac{\kappa_{ij} |(\boldsymbol{a}_i^{(\mathrm{t})})^{\mathsf{H}} (\mathbf{R}_{ij}^{(\mathrm{x})})^{-1} \boldsymbol{x}_{ij}|^2 + \frac{\beta}{(r_{ij}^{(\mathrm{t})})^2}}{(\boldsymbol{a}_i^{(\mathrm{t})})^{\mathsf{H}} (\mathbf{R}_{ij}^{(\mathrm{x})})^{-1} \boldsymbol{a}_i^{(\mathrm{t})} + \frac{\alpha+1}{r_{ij}^{(\mathrm{t})}}}}, \tag{32}$$

$$r_{ij}^{(\mathrm{n})} \leftarrow r_{ij}^{(\mathrm{n})} \sqrt{\frac{\kappa_{ij} \boldsymbol{x}_{ij}^{\mathsf{H}} (\mathbf{R}_{ij}^{(\mathrm{x})})^{-1} \mathbf{R}_i^{(\mathrm{n})} (\mathbf{R}_{ij}^{(\mathrm{x})})^{-1} \boldsymbol{x}_{ij}}{\mathrm{tr}((\mathbf{R}_{ij}^{(\mathrm{x})})^{-1} \mathbf{R}_i^{(\mathrm{n})})}}, \tag{33}$$

$$\lambda_i \leftarrow \lambda_i \sqrt{\frac{\sum_j \kappa_{ij} r_{ij}^{(\mathrm{n})} |\boldsymbol{b}_i^{\mathsf{H}} (\mathbf{R}_{ij}^{(\mathrm{x})})^{-1} \boldsymbol{x}_{ij}|^2}{\sum_j r_{ij}^{(\mathrm{n})} \boldsymbol{b}_i^{\mathsf{H}} (\mathbf{R}_{ij}^{(\mathrm{x})})^{-1} \boldsymbol{b}_i}}, \tag{34}$$

$$\boldsymbol{b}_i \leftarrow (\mathbf{R}_i^{\prime(\mathrm{n})} \boldsymbol{\Xi}_i + \mu_i \mathbf{E})^{-1} (\mathbf{R}_i^{\prime(\mathrm{n})} \boldsymbol{\Upsilon}_i + \mu_i \mathbf{E}) \boldsymbol{b}_i, \tag{35}$$

where

$$\kappa_{ij} = \frac{\rho}{2(\boldsymbol{x}_{ij}^{\mathsf{H}}(\mathbf{R}_{ij}^{(\mathrm{x})})^{-1}\boldsymbol{x}_{ij})^{1-\frac{\rho}{2}}}, \tag{36}$$

$$\boldsymbol{\Xi}_i = \sum_j r_{ij}^{(\mathrm{n})}(\mathbf{R}_{ij}^{(\mathrm{x})})^{-1}, \tag{37}$$

$$\boldsymbol{\Upsilon}_i = \sum_j \kappa_{ij} r_{ij}^{(\mathrm{n})}(\mathbf{R}_{ij}^{(\mathrm{x})})^{-1}\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^{\mathsf{H}}(\mathbf{R}_{ij}^{(\mathrm{x})})^{-1}, \tag{38}$$

$$\mu_i = \lambda_i \boldsymbol{b}_i^{\mathsf{H}} \boldsymbol{\Upsilon}_i \boldsymbol{b}_i. \tag{39}$$

3.3 ME-algorithm-based update rule of deficient component

We focus on the derivation of the ME-algorithm-based update rule for $\boldsymbol{b}_i$ because the derivation for λ_i is the same as that in the conventional fixed-basis RCSCME when the other parameters $r_{ij}^{(\mathrm{t})}$, $r_{ij}^{(\mathrm{n})}$ and $\boldsymbol{b}_i$ are fixed. We use the same auxiliary function and restriction on parameters as in Sect. 3.2.

We heuristically discover the vector $\boldsymbol{b}_i^{(\mathrm{ME})} \in \mathbb{C}^M$ as one of the innumerable possible candidates of the ME-algorithm-based update rule of $\boldsymbol{b}_i$:

$$\boldsymbol{b}_i^{(\mathrm{ME})} := 2\boldsymbol{b}_i^{(\mathrm{MM})} - \tilde{\boldsymbol{b}}_i, \tag{40}$$

where $\tilde{\boldsymbol{b}}_i \in \mathbb{C}^M$ is the pre-update vector of $\boldsymbol{b}_i$, and $\boldsymbol{b}_i^{(\mathrm{ME})}$ satisfies $(\boldsymbol{b}_i^{(\mathrm{ME})})^{\mathsf{H}}\boldsymbol{u}_i = 1$ because $\tilde{\boldsymbol{b}}_i^{\mathsf{H}}\boldsymbol{u}_i = 1$ and $(\boldsymbol{b}_i^{(\mathrm{MM})})^{\mathsf{H}}\boldsymbol{u}_i = 1$. First, we present Claim 1 guaranteeing that $\boldsymbol{b}_i \leftarrow \boldsymbol{b}_i^{(\mathrm{ME})}$ is an ME-algorithm-based update rule. That is, $\boldsymbol{b}_i^{(\mathrm{ME})}$ provides the same value of the auxiliary function as the pre-update vector.

Claim 1 *We define the function $\mathcal{L}_i^{\mathrm{U}}$ as*

$$\mathcal{L}_i^{\mathrm{U}}(\boldsymbol{b}_i) := \boldsymbol{b}_i^{\mathsf{H}}\mathbf{G}_i\boldsymbol{b}_i - \boldsymbol{h}_i^{\mathsf{H}}\boldsymbol{b}_i - \boldsymbol{b}_i^{\mathsf{H}}\boldsymbol{h}_i. \tag{41}$$

Then, the following equation holds:

$$\mathcal{L}_i^{\mathrm{U}}(\boldsymbol{b}_i^{(\mathrm{ME})}) = \mathcal{L}_i^{\mathrm{U}}(\tilde{\boldsymbol{b}}_i). \tag{42}$$

Proof First, $\mathcal{L}_i^{\mathrm{U}}$ is deformed as

$$\begin{aligned}
\mathcal{L}_i^{\mathrm{U}}(\boldsymbol{b}_i) &= (\boldsymbol{b}_i - \boldsymbol{b}_i^{(\mathrm{MM})})^{\mathsf{H}}\mathbf{G}_i(\boldsymbol{b}_i - \boldsymbol{b}_i^{(\mathrm{MM})}) \\
&\quad + \frac{|1 - \boldsymbol{u}_i^{\mathsf{H}}\mathbf{G}_i^{-1}\boldsymbol{h}_i|^2}{\boldsymbol{u}_i^{\mathsf{H}}\mathbf{G}_i^{-1}\boldsymbol{u}_i} - \boldsymbol{h}_i^{\mathsf{H}}\mathbf{G}_i^{-1}\boldsymbol{h}_i.
\end{aligned} \tag{43}$$

Then, by using the definition of $\boldsymbol{b}_i^{(\mathrm{ME})}$, we can calculate the following:

$$
\begin{aligned}
&\mathcal{L}_i^{\mathrm{U}}(\boldsymbol{b}_i^{(\mathrm{ME})}) - \mathcal{L}_i^{\mathrm{U}}(\tilde{\boldsymbol{b}}_i) \\
&= (\boldsymbol{b}_i^{(\mathrm{MM})} - \tilde{\boldsymbol{b}}_i)^{\mathsf{H}}\mathbf{G}_i(\boldsymbol{b}_i^{(\mathrm{MM})} - \tilde{\boldsymbol{b}}_i) \\
&\quad - (\tilde{\boldsymbol{b}}_i - \boldsymbol{b}_i^{(\mathrm{MM})})^{\mathsf{H}}\mathbf{G}_i(\tilde{\boldsymbol{b}}_i - \boldsymbol{b}_i^{(\mathrm{MM})}) \\
&= 0.
\end{aligned}
\tag{44}
$$

□

Next, we present Claim 2 supporting the fact that the step of the ME-algorithm-based update rule is always larger in some sense (sense of the LogDet divergence [25] between the pre/post-update noise SCMs) than that of the MM-algorithm-based update rule proposed in Sect. 3.2. This larger step is expected to yield fast convergence; indeed, [21] reports the improvement in convergence and we will experimentally show the improvement in Sect. 4.2. Since the LogDet divergence is often used as the measure between two covariance matrices, we also introduce this divergence in this study. For the justification of Claim 2, we prepare Lemma 1.

Lemma 1 *Let $\mathbf{R}' \in \mathbb{C}^{M \times M}$ be a rank-$(M-1)$ Hermitian matrix and $\boldsymbol{u} \in \mathbb{C}^M$ be a unit eigenvector of the zero eigenvalue in $\mathbf{R}'$. For all $\boldsymbol{b} \in \mathbb{C}^M$ that satisfy $\boldsymbol{b}^{\mathsf{H}}\boldsymbol{u} = 1$ and $\lambda \in \mathbb{R}_+$, it holds that*

$$
\det(\mathbf{R}' + \lambda \boldsymbol{b}\boldsymbol{b}^{\mathsf{H}}) = \det(\mathbf{R}' + \lambda \boldsymbol{u}\boldsymbol{u}^{\mathsf{H}}).
\tag{45}
$$

Proof From $\boldsymbol{b}^{\mathsf{H}}\boldsymbol{u} = 1$,

$$
\boldsymbol{b} = \boldsymbol{v} + \boldsymbol{u}
\tag{46}
$$

holds, where $\boldsymbol{v} \in \mathbb{C}^M$ satisfies $\boldsymbol{u}^{\mathsf{H}}\boldsymbol{v} = 0$. It holds that

$$
\det(\mathbf{E} - \boldsymbol{v}\boldsymbol{u}^{\mathsf{H}}) = 1 - \boldsymbol{u}^{\mathsf{H}}\boldsymbol{v} = 1,
\tag{47}
$$

which is derived from the matrix determinant lemma [26]. Utilizing

$$
(\mathbf{E} - \boldsymbol{v}\boldsymbol{u}^{\mathsf{H}})\boldsymbol{b} = \boldsymbol{b} - \boldsymbol{v} = \boldsymbol{u},
\tag{48}
$$

$$
(\mathbf{E} - \boldsymbol{v}\boldsymbol{u}^{\mathsf{H}})\mathbf{R}' = \mathbf{R}',
\tag{49}
$$

we deform $\det(\mathbf{R}' + \lambda \boldsymbol{b}\boldsymbol{b}^{\mathsf{H}})$ as

$$
\begin{aligned}
&\det(\mathbf{R}' + \lambda \boldsymbol{b}\boldsymbol{b}^{\mathsf{H}}) \\
&= \det((\mathbf{E} - \boldsymbol{v}\boldsymbol{u}^{\mathsf{H}})(\mathbf{R}' + \lambda \boldsymbol{b}\boldsymbol{b}^{\mathsf{H}})(\mathbf{E} - \boldsymbol{u}\boldsymbol{v}^{\mathsf{H}})) \\
&= \det(\mathbf{R}' + \lambda \boldsymbol{u}\boldsymbol{u}^{\mathsf{H}}).
\end{aligned}
\tag{50}
$$

□

Claim 2 *We define the positive-definite matrices as*

$$\tilde{\mathbf{R}}_i^{(n)} := \mathbf{R}_i'^{(n)} + \lambda_i \tilde{\boldsymbol{b}}_i \tilde{\boldsymbol{b}}_i^{\mathsf{H}}, \tag{51}$$

$$\mathbf{R}_i^{(n,MM)} := \mathbf{R}_i'^{(n)} + \lambda_i \boldsymbol{b}_i^{(MM)} (\boldsymbol{b}_i^{(MM)})^{\mathsf{H}}, \tag{52}$$

$$\mathbf{R}_i^{(n,ME)} := \mathbf{R}_i'^{(n)} + \lambda_i \boldsymbol{b}_i^{(ME)} (\boldsymbol{b}_i^{(ME)})^{\mathsf{H}}. \tag{53}$$

We denote the LogDet divergence [25] *defined between two positive-definite matrices* $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2 \in \mathbb{C}^{M \times M}$ *as*

$$\mathcal{D}(\boldsymbol{\Sigma}_1; \boldsymbol{\Sigma}_2) = \mathrm{tr}(\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2^{-1}) - \log \det (\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2^{-1}) - M. \tag{54}$$

Then, it holds that

$$\mathcal{D}(\mathbf{R}_i^{(n,ME)}; \tilde{\mathbf{R}}_i^{(n)}) \geq \mathcal{D}(\mathbf{R}_i^{(n,MM)}; \tilde{\mathbf{R}}_i^{(n)}). \tag{55}$$

Proof We subtract the right side from the left side as

$$
\begin{aligned}
&\mathcal{D}(\mathbf{R}_i^{(n,ME)}; \tilde{\mathbf{R}}_i^{(n)}) - \mathcal{D}(\mathbf{R}_i^{(n,MM)}; \tilde{\mathbf{R}}_i^{(n)}) \\
&= \mathrm{tr}\left(\mathbf{R}_i^{(n,ME)} (\tilde{\mathbf{R}}_i^{(n)})^{-1} \right) - \mathrm{tr}\left(\mathbf{R}_i^{(n,MM)} (\tilde{\mathbf{R}}_i^{(n)})^{-1} \right) \\
&\quad + \log \det \mathbf{R}_i^{(n,MM)} - \log \det \mathbf{R}_i^{(n,ME)}
\end{aligned}
\tag{56}
$$

$$
\begin{aligned}
&= \mathrm{tr}\left(\left(\mathbf{R}_i'^{(n)} + \lambda_i \boldsymbol{b}_i^{(ME)} (\boldsymbol{b}_i^{(ME)})^{\mathsf{H}} \right) (\tilde{\mathbf{R}}_i^{(n)})^{-1} \right) \\
&\quad - \mathrm{tr}\left(\left(\mathbf{R}_i'^{(n)} + \lambda_i \boldsymbol{b}_i^{(MM)} (\boldsymbol{b}_i^{(MM)})^{\mathsf{H}} \right) (\tilde{\mathbf{R}}_i^{(n)})^{-1} \right) \\
&\quad + \log \det \mathbf{R}_i^{(n,MM)} - \log \det \mathbf{R}_i^{(n,ME)}
\end{aligned}
\tag{57}
$$

$$
\begin{aligned}
&= \lambda_i \Big((\boldsymbol{b}_i^{(ME)})^{\mathsf{H}} (\tilde{\mathbf{R}}_i^{(n)})^{-1} \boldsymbol{b}_i^{(ME)} \\
&\quad - (\boldsymbol{b}_i^{(MM)})^{\mathsf{H}} (\tilde{\mathbf{R}}_i^{(n)})^{-1} \boldsymbol{b}_i^{(MM)} \Big) \\
&\quad + \log \det \mathbf{R}_i^{(n,MM)} - \log \det \mathbf{R}_i^{(n,ME)}.
\end{aligned}
\tag{58}
$$

First, since $\boldsymbol{b}_i^{(MM)}$ and $\boldsymbol{b}_i^{(ME)}$ satisfy the condition of $\boldsymbol{b}$ in Lemma 1, $\det \mathbf{R}_i^{(n,MM)} = \det \mathbf{R}_i^{(n,ME)} = \det (\mathbf{R}_i'^{(n)} + \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^{\mathsf{H}})$ holds, resulting in $\log \det \mathbf{R}_i^{(n,MM)} - \log \det \mathbf{R}_i^{(n,ME)} = 0$. Next, $(\tilde{\mathbf{R}}_i^{(n)})^{-1}$ is expanded as

$$
\begin{aligned}
(\tilde{\mathbf{R}}_i^{(n)})^{-1} &= (\mathbf{E} - \boldsymbol{u}_i \tilde{\boldsymbol{b}}_i^{\mathsf{H}})(\mathbf{R}_i'^{(n)})^{+}(\mathbf{E} - \tilde{\boldsymbol{b}}_i \boldsymbol{u}_i^{\mathsf{H}}) \\
&\quad + \frac{1}{\lambda_i} \boldsymbol{u}_i \boldsymbol{u}_i^{\mathsf{H}},
\end{aligned}
\tag{59}
$$

which is described in [19]. Utilizing $(\boldsymbol{b}_i^{(MM)})^{\mathsf{H}} \boldsymbol{u}_i = (\boldsymbol{b}_i^{(ME)})^{\mathsf{H}} \boldsymbol{u}_i = 1$, we obtain

$$(\mathbf{E} - \tilde{\boldsymbol{b}}_i \boldsymbol{u}_i^{\mathsf{H}}) \boldsymbol{b}_i^{(\mathrm{MM})} = \boldsymbol{b}_i^{(\mathrm{MM})} - \tilde{\boldsymbol{b}}_i, \tag{60}$$

$$\begin{aligned}
(\mathbf{E} - \tilde{\boldsymbol{b}}_i \boldsymbol{u}_i^{\mathsf{H}}) \boldsymbol{b}_i^{(\mathrm{ME})} &= \boldsymbol{b}_i^{(\mathrm{ME})} - \tilde{\boldsymbol{b}}_i \\
&= 2(\boldsymbol{b}_i^{(\mathrm{MM})} - \tilde{\boldsymbol{b}}_i).
\end{aligned} \tag{61}$$

Then, we can calculate the following:

$$\begin{aligned}
\mathcal{D}(\mathbf{R}_i^{(\mathrm{n,ME})}; \tilde{\mathbf{R}}_i^{(\mathrm{n})}) &- \mathcal{D}(\mathbf{R}_i^{(\mathrm{n,MM})}; \tilde{\mathbf{R}}_i^{(\mathrm{n})}) \\
&= \lambda_i \Big((\boldsymbol{b}_i^{(\mathrm{ME})})^{\mathsf{H}} (\tilde{\mathbf{R}}_i^{(\mathrm{n})})^{-1} \boldsymbol{b}_i^{(\mathrm{ME})} \\
&\quad - (\boldsymbol{b}_i^{(\mathrm{MM})})^{\mathsf{H}} (\tilde{\mathbf{R}}_i^{(\mathrm{n})})^{-1} \boldsymbol{b}_i^{(\mathrm{MM})} \Big) \\
&= 3\lambda_i (\boldsymbol{b}_i^{(\mathrm{MM})} - \tilde{\boldsymbol{b}}_i)^{\mathsf{H}} (\mathbf{R}_i^{\prime(\mathrm{n})})^{+} (\boldsymbol{b}_i^{(\mathrm{MM})} - \tilde{\boldsymbol{b}}_i) \\
&\geq 0,
\end{aligned} \tag{62}$$

because $(\mathbf{R}_i^{\prime(\mathrm{n})})^{+}$ is a positive semidefinite matrix. □

The update rules of all parameters other than $\boldsymbol{b}_i$ can be obtained in the same manner as those in the conventional fixed-basis RCSCME. Finally, the ME-algorithm-based update rules are derived as

$$r_{ij}^{(\mathrm{t})} \leftarrow r_{ij}^{(\mathrm{t})} \frac{|(\boldsymbol{a}_i^{(\mathrm{t})})^{\mathsf{H}} (\mathbf{R}_{ij}^{(\mathrm{x})})^{-1} \boldsymbol{x}_{ij}|^2 + \frac{\beta}{(r_{ij}^{(\mathrm{t})})^2}}{(\boldsymbol{a}_i^{(\mathrm{t})})^{\mathsf{H}} (\mathbf{R}_{ij}^{(\mathrm{x})})^{-1} \boldsymbol{a}_i^{(\mathrm{t})} + \frac{\alpha+1}{r_{ij}^{(\mathrm{t})}}}, \tag{63}$$

$$r_{ij}^{(\mathrm{n})} \leftarrow r_{ij}^{(\mathrm{n})} \frac{\boldsymbol{x}_{ij}^{\mathsf{H}} (\mathbf{R}_{ij}^{(\mathrm{x})})^{-1} \mathbf{R}_i^{(\mathrm{n})} (\mathbf{R}_{ij}^{(\mathrm{x})})^{-1} \boldsymbol{x}_{ij}}{\mathrm{tr}((\mathbf{R}_{ij}^{(\mathrm{x})})^{-1} \mathbf{R}_i^{(\mathrm{n})})}, \tag{64}$$

$$\lambda_i \leftarrow \lambda_i \frac{\sum_j r_{ij}^{(\mathrm{n})} |\boldsymbol{b}_i^{\mathsf{H}} (\mathbf{R}_{ij}^{(\mathrm{x})})^{-1} \boldsymbol{x}_{ij}|^2}{\sum_j r_{ij}^{(\mathrm{n})} \boldsymbol{b}_i^{\mathsf{H}} (\mathbf{R}_{ij}^{(\mathrm{x})})^{-1} \boldsymbol{b}_i}, \tag{65}$$

$$\begin{aligned}
\boldsymbol{b}_i \leftarrow\ &2(\mathbf{R}_i^{\prime(\mathrm{n})} \boldsymbol{\Xi}_i + \mu_i \mathbf{E})^{-1} (\mathbf{R}_i^{\prime(\mathrm{n})} \boldsymbol{\Upsilon}_i + \mu_i \mathbf{E}) \boldsymbol{b}_i \\
&- \boldsymbol{b}_i,
\end{aligned} \tag{66}$$

where κ_{ij}, $\boldsymbol{\Xi}_i$, $\boldsymbol{\Upsilon}_i$, and μ_i are the same as those in Sect. 3.2.

4 Experimental results and discussion

4.1 Experimental conditions

To confirm the efficacy of the proposed basis-optimizing RCSCME, we conducted BSE experiments under simulated noise conditions. We simulated a mixture of a target speech source and diffuse noise by convoluting dry sources with impulse responses from each position to four equally spaced microphones as shown in Fig. 2. The diffuse noise was simulated by reproduction from 19 positions on the same circumference. The target speech originates from the loudspeaker located $0°, 10°, 20°$, and $30°$ clockwise from the
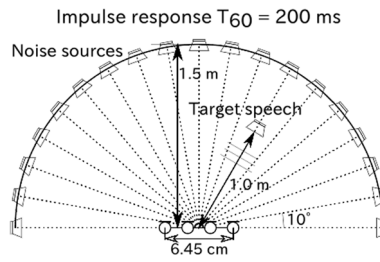
Fig. 2 Recording situation of impulse responses

normal to the microphone array and is closer to the microphone array than each diffuse noise loudspeaker. As the target speech source, we utilized six JNAS clean speech corpus sources [27]. We used four diffuse noises, namely, the babble, station, traffic, and cafe noises. For the babble noise, we simulated diffuse noise by reproducing 19 other JNAS speech corpus sources from each loudspeaker. For station, traffic, and cafe noises, noise signals in DEMAND [28] are split into 19 fragments, which are then reproduced from each position. An STFT was performed by using a 64-ms-long Hamming window with a 32-ms-long shift. The input signal-to-noise ratio was set to 0 dB.

We compared 11 BSE methods, namely, ILRMA [9], independent vector extraction (IVE) [3], blind spatial subtraction array (BSSA) [2], MWF with single-channel noise power estimation (MWF1) [29], MWF with multichannel noise power estimation (MWF2) [30], original MNMF [14], MNMF initialized by ILRMA (ILRMA + MNMF) [9, 31], original FastMNMF [17], FastMNMF initialized by ILRMA (ILRMA + FastMNMF), the conventional fixed-basis RCSCME [19], and the proposed basis-optimizing RCSCME. In ILRMA, which was used as the preprocessing for each method, the number of bases was 10, the number of iterations was 50, and the observed signal was preprocessed using sphering transformation with principal component analysis. In IVE, the separation filter for the target speech was initialized by ILRMA. As for BSSA, ILRMA was used in place of FDICA utilized in [2] and the oversubtraction and flooring parameters were set to 1.4 and 0, respectively. In MWF1 and MWF2, the a priori speech-to-noise ratio was estimated by a decision-directed approach [32]. In MWF1, we used a minima controlled recursive averaging noise estimation approach [29] for estimating the noise power spectrum. In MWF2, we estimated the noise power spectrum using $M - 1$ outputs of ILRMA excluding the n_tth signal. For ILRMA, MNMF, and FastMNMF, the source model variables were initialized using nonnegative random values following a uniform distribution on [0, 1]. As for ILRMA + MNMF and ILRMA + FastMNMF, the source model variables were handed over from ILRMA to MNMF and FastMNMF, respectively. For ILRMA, the demixing matrix was initialized by the identity matrix $\mathbf{E}$. The SCM was initialized by $\mathbf{E}$ for both MNMF and FastM-NMF, $\boldsymbol{a}_{i,n_\mathrm{t}} \boldsymbol{a}_{i,n_\mathrm{t}}^{\mathsf{H}} + \epsilon \mathbf{E}$ for ILRMA + MNMF, and $\boldsymbol{a}_{i,n_\mathrm{t}} \boldsymbol{a}_{i,n_\mathrm{t}} + \epsilon \sum_{n \neq n_\mathrm{t}} \boldsymbol{a}_{i,n} \boldsymbol{a}_{i,n}^{\mathsf{H}}$ for ILRMA + FastMNMF, where ϵ was set to 10^{-5}. For MNMF and FastMNMF, we blindly selected the separated signal whose kurtosis was maximum from four separated signals as the target source. For all methods except IVE, MNMF, and FastMNMF, we blindly determined the index n_t of the target source by selecting the demixed signal whose kurtosis was maximum from the M demixed signals of ILRMA. In both RCSCMEs, we utilized

the minimum positive eigenvalue σ_i of $\mathbf{R}_i'^{(n)}$ as the initial value of λ_i. As both the fixed $\boldsymbol{b}_i$ in the conventional fixed-basis RCSCME and the initial value of $\boldsymbol{b}_i$ in the proposed basis-optimizing RCSCME, we used a unit eigenvector $\boldsymbol{u}_i$ of the zero eigenvalue in $\mathbf{R}_i'^{(n)}$. In the conventional fixed-basis RCSCME, we utilized $\alpha = 2.5$ and $\beta = 10^{-16}$, which are the parameters of the inverse gamma distribution as the prior and showed the best separation performance at the preliminary experiment in [19]. We experimentally chose $\alpha = 0.01$ and $\beta = 10^{-16}$ in the proposed basis-optimizing RCSCME. For both RCSC-MEs, the shape parameters of GGD $\rho = 0.5, 1, 2$ were utilized. For the evaluation of BSE performance, we used the source-to-distortion ratio (SDR) improvement [33]. The SDR improvement was averaged over 10 parameter-initialization random seeds, four target directions, and six target speech sources (totally 240 trials).

4.2 Comparison between three types of basis-optimizing RCSCME

We conducted a preliminary experiment under the babble noise condition. We compared the MM-algorithm-based fixed-basis RCSCME, the ME-algorithm-based fixed-basis RCSCME, the proposed MM-algorithm-based basis-optimizing RCSCME, and the proposed ME-algorithm-based basis-optimizing RCSCME whose generative models are the Gaussian distribution ($\rho = 2$) and the super-Gaussian distribution ($\rho = 0.5$). The reason why the shape parameter of the GGD $\rho = 0.5$ was used was that the fixed-basis RCSCME showed the best performance in the shape parameter $\rho = 0.5$ in experiments in [19]. In the Gaussian case, we also compared the EM-algorithm-based basis-optimizing RCSCME proposed in our conference paper [22], which is a method that can only be applied to the Gaussian case.

Figure 3 shows the SDR improvements of RCSCMEs for each iteration. As a reference, we also show the SDR improvement after 50 iterations of preprocessing ILRMA. The SDR improvements of all the RCSCMEs reach a peak followed by a decrease, which is caused by the sparsity of the speech signal [19]. Regarding the proposed basis-optimizing RCSCMEs, ME-algorithm-based update rules outperform MM-algorithm-based update rules in terms of both separation performance (the peak of the SDR curve) and convergence, which is a trend that can also be seen in conventional fixed-basis RCSC-MEs. The advantage of ME-algorithm-based update rules over MM-algorithm-based update rules in terms of convergence is consistent with the description in Sect. 3.1. To show this advantage in convergence of the cost function, we additionally compared the proposed MM- and ME-algorithm-based basis-optimizing RCSCMEs in the behavior of cost functions (12) except for the constant term. Figure 4 shows the average of the cost function over 10 parameter-initialization random seeds under the condition that the target speech originated from the loudspeaker located $0°$ clockwise from the normal to the microphone array and the shape parameter ρ was set to 2. From Fig. 4, ME-algorithm-based basis-optimizing RCSCME achieved faster convergence in the cost function than MM-algorithm-based basis-optimizing RCSCME, which provides support for the advantage of ME-algorithm-based basis-optimizing RCSCME against MM-algorithm-based basis-optimizing RCSCME regarding convergence speed. Furthermore, since the difference of the computational complexity between the proposed MM- and ME-algorithm-based update rules is caused by (40) and (40) has the much less computational
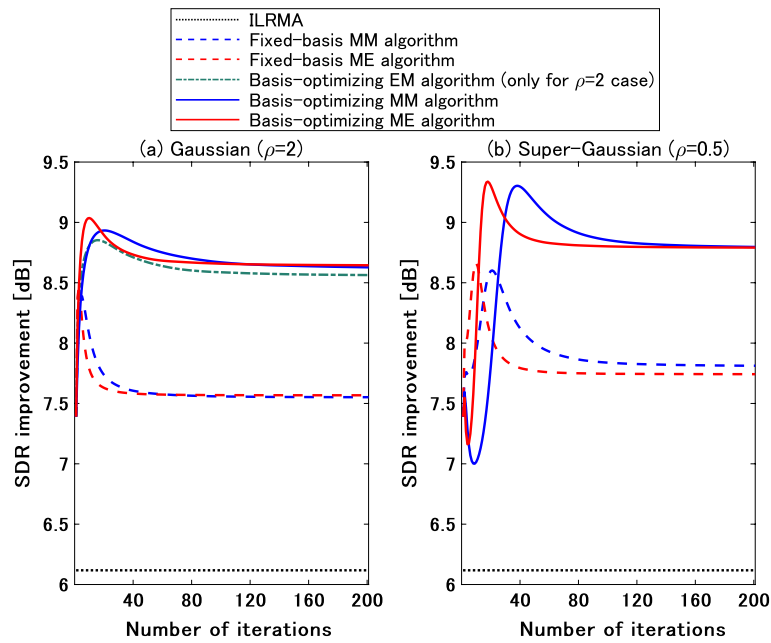
Fig. 3 Behavior of SDR improvements under the babble noise condition. SDR improvements are averaged over six speech sources, four target directions, and 10 parameter-initialization random seeds. The shape parameter ρ in GGD is set to (**a**) 2 (Gaussian) and (**b**) 0.5 (super-Gaussian)
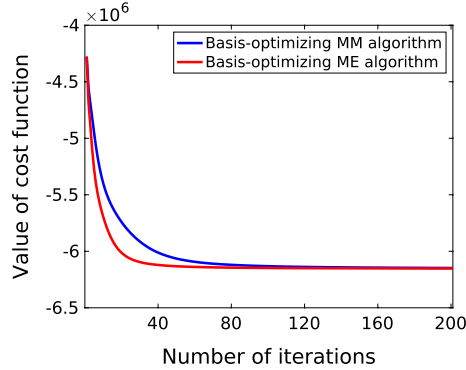

Fig. 4 Behavior of the cost functions (12) except for the constant term under the babble noise condition in the proposed basis-optimizing RCSCMEs. The cost functions are averaged over 10 parameter-initialization random seeds. The shape parameter ρ in GGD is set to 2

complexity than (35), the computational times of the proposed MM- and ME-algorithm-based basis-optimizing RCSCMEs are expected to be almost the same. In fact, according to the experiment, the averages of the execution times per step in MM- and ME-algorithm-based update rules over 2000 trials, which consist of 10 parameter-initialization random seeds and 200 iterations, were 10.3 s and 10.2 s, respectively. This time measurement was executed under the condition that the target speech originated from the loudspeaker located 0° clockwise from the normal to the microphone array and the shape parameter ρ was set to 2. The code for the time measurement was implemented in MATLAB (R2022a), and the computation was performed on an Intel Core i9-9980XE (3.00

Table 1 SDR improvements for each method under each simulated noise condition

Methods	Babble noise	Station noise	Traffic noise	Cafe noise
ILRMA	6.1/–	6.2/–	4.7/–	6.4/–
IVE	6.2/5.8	6.2/5.8	5.4/5.2	6.4/5.9
BSSA	6.8/–	6.9/–	5.7/–	7.2/–
MWF1	6.1/–	6.9/–	5.8/–	7.0/–
MWF2	6.9/–	7.2/–	5.6/–	7.4/–
MNMF	2.7/2.7	3.5/3.5	2.6/2.6	2.8/2.8
ILRMA + MNMF	6.5/6.5	7.0/7.0	6.2/6.2	8.1/8.1
FastMNMF	1.7/1.7	2.6/2.5	2.9/2.8	2.6/2.6
ILRMA + FastMNMF	6.5/6.2	6.6/6.6	5.4/5.4	7.3/7.3
Fixed-basis RCSCME				
$\rho = 0.5$	8.7/7.7	10.2/9.5	8.3/7.8	10.1/9.3
$\rho = 1$	8.6/7.7	9.9/9.4	8.0/7.7	9.9/9.2
$\rho = 2$	8.5/7.6	9.8/9.2	7.9/7.5	9.8/9.1
Proposed basis-optimizing RCSCME				
$\rho = 0.5$	**9.3/8.8**	**11.5/11.0**	**9.4/9.2**	**11.1/10.5**
$\rho = 1$	9.1/8.6	11.2/10.8	9.1/9.0	10.9/10.4
$\rho = 2$	9.0/8.6	11.1/10.7	9.0/8.9	10.8/10.4

Each term represents "best-iteration score [dB]/after-200-iteration (except for IVE; after-4000-iteration for IVE) score [dB]."
Bold text indicates the maximum value

GHz, 18 cores) CPU. Thus, the advantage of convergence in computational time had the same tendency as that in iteration shown in Figs. 3 and 4. On the basis of these results, we employ only ME-algorithm-based update rules for both the conventional fixed-basis RCSCME and the proposed basis-optimizing RCSCME.

4.3 Comparison between the proposed basis-optimizing RCSCME and conventional methods under simulated noise conditions

Table 1 shows the SDR improvements averaged over 240 cases for each method under each simulated noise condition. For MNMF, ILRMA + MNMF, FastMNMF, ILRMA + FastMNMF, the conventional fixed-basis RCSCME, and the proposed basis-optimizing RCSCME, we show both the peak SDR improvement and the SDR improvement after 200 iterations. For IVE, which is slow to converge, we show both the peak SDR improvement and the SDR improvement after 4000 iterations, which is the number of iterations recommended to achieve sufficient separation performance [3]. The proposed basis-optimizing RCSCME outperforms all the conventional methods under all the noise conditions. Furthermore, for the shape parameter ρ in the proposed basis-optimizing RCSCME, $\rho = 0.5$ provides the best SDR improvement. This confirms the efficacy of the super-Gaussian GGD.

4.4 Comparison between proposed basis-optimizing RCSCME and conventional methods under real noise condition

To confirm the efficacy of the proposed method under a more realistic noise condition, we conducted a BSE experiment using real-world sounds. We used a parking

Table 2 SDR improvements for each method under real parking noise condition

Methods	
ILRMA	6.0/–
BSSA	5.1/–
MWF1	6.7/–
MWF2	6.2/–
FastMNMF	2.5/2.5
ILRMA + FastMNMF	8.4/8.4
Fixed-basis RCSCME	
$\rho = 0.5$	9.1/8.8
$\rho = 1$	9.0/8.8
$\rho = 2$	8.9/8.8
Proposed basis-optimizing RCSCME	
$\rho = 0.5$	10.3/10.0
$\rho = 1$	10.4/10.0
$\rho = 2$	**10.4/10.1**

Each term represents "best-iteration score [dB]/after-200-iteration score [dB]." Bold text indicates the maximum value

noise, which was a diffuse noise recorded outdoors. The reverberation time was about 90 ms. For the directional target speech, the same dry source was reproduced from the loudspeaker located $0°, 10°, 20°,$ and $30°$ clockwise from the normal to the microphone array at a distance of 1.0 m. An STFT was performed by using a 256-ms-long Hamming window with a 32-ms-long shift.

We compared eight methods, ILRMA, BSSA, MWF1, MWF2, FastMNMF, ILRMA + FastMNMF, the conventional fixed-basis RCSCME, and the proposed basis-optimizing RCSCME. We excluded IVE, MNMF, and ILRMA + MNMF for this experiment because the convergence of these methods is significantly slow. In the conventional fixed-basis RCSCME, we used $\alpha = 2.3$ and $\beta = 10^{-16}$, which are the parameters of the inverse gamma distribution as the prior. The other conditions of the methods compared were the same as those described in Sect. 4.1.

Table 2 shows SDR improvements averaged over 240 cases for each method under the parking noise condition. The proposed basis-optimizing RCSCME outperforms all the conventional methods. This result shows the efficacy of the proposed basis-optimizing RCSCME in a practical situation.

5 Conclusions

In this paper, we proposed a new model extension of RCSCME, which is a blind speech extraction method utilized under the condition that one-directional target speech and diffuse background noise are mixed. In the conventional fixed-basis RCSCME, between two parameters constituting the deficient rank-1 component, only the scale is estimated, whereas the deficient basis is fixed in advance. In the proposed basis-optimizing RCSCME model, we regarded the deficient basis as a parameter to estimate. We derived new MM- and ME-algorithm-based update rules for the deficient basis in the GGD model, which achieves better separation performance than the Gaussian distribution in the conventional RCSCME. In particular, among innumerable ME-algorithm-based update rules, we successfully found an ME-algorithm-based update rule with a

mathematical proof supporting the fact that the step of the update rule is larger than that of the MM-algorithm-based update rule. We confirmed that the proposed method outperforms conventional methods under several simulated noise conditions and a real noise condition.

Appendix

In this section, we describe the derivation of the auxiliary function $\mathcal{L}^{\mathsf{U}}(\Theta_{\mathrm{p}}, \Omega_{\mathrm{p}})$ in the Gaussian and super-Gaussian cases ($\rho \leq 2$) for the following negative log posterior in the same manner as in [19]:

$$
\begin{aligned}
\mathcal{L}(\Theta_{\mathrm{p}}) = \sum_{i,j} \bigg[& (\boldsymbol{x}_{ij}^{\mathsf{H}}(\mathbf{R}_{ij}^{(\mathrm{x})})^{-1}\boldsymbol{x}_{ij})^{\frac{\rho}{2}} + \log \det \mathbf{R}_{ij}^{(\mathrm{x})} \\
& + (\alpha + 1)\log r_{ij}^{(\mathrm{t})} + \frac{\beta}{r_{ij}^{(\mathrm{t})}} \bigg] + \mathrm{const.}
\end{aligned}
\tag{67}
$$

First, when $\rho \leq 2$ holds, we can apply the following tangent inequality for the first term in the right-hand side of (67):

$$
(\boldsymbol{x}_{ij}^{\mathsf{H}}(\mathbf{R}_{ij}^{(\mathrm{x})})^{-1}\boldsymbol{x}_{ij})^{\frac{\rho}{2}} \leq \frac{\rho}{2\iota_{ij}^{1-\frac{\rho}{2}}}\boldsymbol{x}_{ij}^{\mathsf{H}}(\mathbf{R}_{ij}^{(\mathrm{x})})^{-1}\boldsymbol{x}_{ij} + \left(1 - \frac{\rho}{2}\right)\iota_{ij}^{\frac{\rho}{2}},
\tag{68}
$$

where ι_{ij} is positive and the equality of (68) holds if and only if

$$
\iota_{ij} = \boldsymbol{x}_{ij}^{\mathsf{H}}(\mathbf{R}_{ij}^{(\mathrm{x})})^{-1}\boldsymbol{x}_{ij}.
\tag{69}
$$

For the first term in the right-hand side of (68), the inequality described in the following theorem [19] can be applied again.

Theorem 1 ([19]) *Let K be any natural number, $\mathbf{R}_k \in \mathbb{C}^{M \times M}(k = 1, \ldots, K)$ be a positive semidefinite Hermitian matrix satisfying $\mathrm{rank}(\sum_k \mathbf{R}_k) = M$, $\mathbf{X} \in \mathbb{C}^{M \times M}$ be any positive semidefinite Hermitian matrix, and $\mathcal{X} \in \mathbb{C}^{M \times M}$ be the projection matrix to the image space of $\mathbf{X}$. For any matrix $\boldsymbol{\Phi}_k \in \mathbb{C}^{M \times M}$ satisfying the conditions*

$$
\mathrm{Ker}\,\boldsymbol{\Phi}_k \supseteq \mathrm{Ker}\,\mathbf{X},
\tag{70}
$$

$$
\mathrm{Im}\,\boldsymbol{\Phi}_k \subseteq \mathrm{Im}\,\mathbf{R}_k,
\tag{71}
$$

$$
\sum_k \boldsymbol{\Phi}_k = \mathcal{X},
\tag{72}
$$

it holds that

$$
\mathrm{tr}\left(\left(\sum_k \mathbf{R}_k\right)^{-1}\mathbf{X}\right) \leq \sum_k \mathrm{tr}(\boldsymbol{\Phi}_k^{\mathsf{H}}\mathbf{R}_k^{+}\boldsymbol{\Phi}_k\mathbf{X}),
\tag{73}
$$

where Ker and Im, respectively, represent a kernel space and an image space. The equality of (73) holds if and only if

$$\mathbf{\Phi}_k = \mathbf{R}_{k'}(\sum_{k'} \mathbf{R}_{k'})^{-1}\mathcal{X}. \tag{74}$$

By replacing $K = 2$, $\mathbf{X} = \boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^{\mathsf{H}}$, $\mathbf{R}_1 = r_{ij}^{(\text{t})}\boldsymbol{a}_i^{(\text{t})}(\boldsymbol{a}_i^{(\text{t})})^{\mathsf{H}}$, $\mathbf{R}_2 = r_{ij}^{(\text{n})}\mathbf{R}_i^{(\text{n})}$, $\mathbf{\Phi}_1 = \mathbf{\Phi}_{ij}^{(\text{t})}$, and $\mathbf{\Phi}_2 = \mathbf{\Phi}_{ij}^{(\text{n})}$ at Theorem 1, it holds that

$$\begin{aligned}
\boldsymbol{x}_{ij}^{\mathsf{H}}(\mathbf{R}_{ij}^{(\text{x})})^{-1}\boldsymbol{x}_{ij} &= \boldsymbol{x}_{ij}^{\mathsf{H}}(r_{ij}^{(\text{t})}\boldsymbol{a}_i^{(\text{t})}(\boldsymbol{a}_i^{(\text{t})})^{\mathsf{H}} + r_{ij}^{(\text{n})}\mathbf{R}_i^{(\text{n})})^{-1}\boldsymbol{x}_{ij} \\
&\leq \frac{|(\boldsymbol{a}_i^{(\text{t})})^{\mathsf{H}}\mathbf{\Phi}_{ij}^{(\text{t})}\boldsymbol{x}_{ij}|^2}{r_{ij}^{(\text{t})}\|\boldsymbol{a}_i^{(\text{t})}\|^4} + \frac{\boldsymbol{x}_{ij}^{\mathsf{H}}(\mathbf{\Phi}_{ij}^{(\text{n})})^{\mathsf{H}}(\mathbf{R}_i^{(\text{n})})^{-1}\mathbf{\Phi}_{ij}^{(\text{n})}\boldsymbol{x}_{ij}}{r_{ij}^{(\text{n})}},
\end{aligned} \tag{75}$$

where $\mathbf{\Phi}_{ij}^{(\text{t})}, \mathbf{\Phi}_{ij}^{(\text{n})} \in \mathbb{C}^{M \times M}$ are the auxiliary variables satisfying $\mathbf{\Phi}_{ij}^{(\text{t})} + \mathbf{\Phi}_{ij}^{(\text{n})} = \boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^{\mathsf{H}}/\|\boldsymbol{x}_{ij}\|_2^2$. The equality of (75) holds if and only if

$$\mathbf{\Phi}_{ij}^{(\text{t})} = r_{ij}^{(\text{t})}\boldsymbol{a}_i^{(\text{t})}(\boldsymbol{a}_i^{(\text{t})})^{\mathsf{H}}(\mathbf{R}_{ij}^{(\text{x})})^{-1}\frac{\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^{\mathsf{H}}}{\|\boldsymbol{x}_{ij}\|^2}, \tag{76}$$

$$\mathbf{\Phi}_{ij}^{(\text{n})} = r_{ij}^{(\text{n})}\mathbf{R}_i^{(\text{n})}(\mathbf{R}_{ij}^{(\text{x})})^{-1}\frac{\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^{\mathsf{H}}}{\|\boldsymbol{x}_{ij}\|^2}. \tag{77}$$

Since the second term of the right-hand side of (75) contains $(\mathbf{R}_i^{(\text{n})})^{-1}$, it is difficult to differentiate the term with respect to $\boldsymbol{c}_i$. Then, we represent $(\mathbf{R}_i^{(\text{n})})^{-1}$ as an explicit expression about $\boldsymbol{c}_i$ using the following claim.

Claim 3 ([19]) *Let $\mathbf{R}' \in \mathbb{C}^{M \times M}$ be a rank-$(M-1)$ positive semidefinite Hermitian matrix, λ be positive, $\boldsymbol{u} \in \mathbb{C}^M$ be an eigenvector of the zero eigenvalue in $\mathbf{R}'$, and $\boldsymbol{b} \in \mathbb{C}^M$ be any vector that is not included in the $(M-1)$-dimensional hyperplane spanned by column vectors of $\mathbf{R}'$. We define the matrix $\mathbf{R} \in \mathbb{C}^{M \times M}$ as*

$$\mathbf{R} := \mathbf{R}' + \lambda\boldsymbol{b}\boldsymbol{b}^{\mathsf{H}}. \tag{78}$$

Then, it holds that

$$\mathbf{R}^{-1} = \check{\mathbf{R}} + \frac{1}{\lambda|\boldsymbol{b}^{\mathsf{H}}\boldsymbol{u}|^2}\boldsymbol{u}\boldsymbol{u}^{\mathsf{H}}, \tag{79}$$

where

$$\check{\mathbf{R}} = \left(\mathbf{E} - \frac{\boldsymbol{u}\boldsymbol{b}^{\mathsf{H}}}{\boldsymbol{b}^{\mathsf{H}}\boldsymbol{u}}\right)(\mathbf{R}')^{+}\left(\mathbf{E} - \frac{\boldsymbol{b}\boldsymbol{u}^{\mathsf{H}}}{\boldsymbol{u}^{\mathsf{H}}\boldsymbol{b}}\right). \tag{80}$$

Note that whereas $\boldsymbol{b}^{\mathsf{H}}\boldsymbol{u} = 1$ are assumed in [19] for simplifying the derivation, we remove this restriction and reformulate Claim 3. From Claim 3, it holds that

$$(\mathbf{R}_i^{(\text{n})})^{-1} = \check{\mathbf{R}}_i^{(\text{n})} + \frac{1}{|\boldsymbol{c}_i^{\mathsf{H}}\boldsymbol{u}_i|^2}\boldsymbol{u}_i\boldsymbol{u}_i^{\mathsf{H}}, \tag{81}$$

where

$$\check{\mathbf{R}}_i^{(\mathrm{n})} = \left(\mathbf{E} - \frac{\boldsymbol{u}_i \boldsymbol{c}_i^{\mathsf{H}}}{\boldsymbol{c}_i^{\mathsf{H}} \boldsymbol{u}_i} \right) (\mathbf{R}_i^{'(\mathrm{n})})^+ \left(\mathbf{E} - \frac{\boldsymbol{c}_i \boldsymbol{u}_i^{\mathsf{H}}}{\boldsymbol{u}_i^{\mathsf{H}} \boldsymbol{c}_i} \right). \tag{82}$$

In summary, we can design the following inequality for the first term of (67):

$$\begin{aligned}
(\boldsymbol{x}_{ij}^{\mathsf{H}} (\mathbf{R}_{ij}^{(\mathrm{x})})^{-1} \boldsymbol{x}_{ij})^{\frac{\rho}{2}} \\
\leq \frac{\rho}{2 \iota_{ij}^{1-\frac{\rho}{2}}} &\left[\frac{|(\boldsymbol{a}_i^{(\mathrm{t})})^{\mathsf{H}} \boldsymbol{\Phi}_{ij}^{(\mathrm{t})} \boldsymbol{x}_{ij}|^2}{r_{ij}^{(\mathrm{t})} \|\boldsymbol{a}_i^{(\mathrm{t})}\|^4} \right. \\
&+ \frac{1}{r_{ij}^{(\mathrm{n})}} \left. \left(\frac{|\boldsymbol{u}_i^{\mathsf{H}} \boldsymbol{\Phi}_{ij}^{(\mathrm{n})} \boldsymbol{x}_{ij}|^2}{|\boldsymbol{c}_i^{\mathsf{H}} \boldsymbol{u}_i|^2} + \boldsymbol{x}_{ij}^{\mathsf{H}} (\boldsymbol{\Phi}_{ij}^{(\mathrm{n})})^{\mathsf{H}} \check{\mathbf{R}}_i^{(\mathrm{n})} \boldsymbol{\Phi}_{ij}^{(\mathrm{n})} \boldsymbol{x}_{ij} \right) \right] + \left(1 - \frac{\rho}{2} \right) \iota_{ij}^{\frac{\rho}{2}}.
\end{aligned} \tag{83}$$

The equality of (83) holds if and only if

$$\iota_{ij} = \boldsymbol{x}_{ij}^{\mathsf{H}} (\mathbf{R}_{ij}^{(\mathrm{x})})^{-1} \boldsymbol{x}_{ij}, \tag{84}$$

$$\boldsymbol{\Phi}_{ij}^{(\mathrm{t})} = r_{ij}^{(\mathrm{t})} \boldsymbol{a}_i^{(\mathrm{t})} (\boldsymbol{a}_i^{(\mathrm{t})})^{\mathsf{H}} (\mathbf{R}_{ij}^{(\mathrm{x})})^{-1} \frac{\boldsymbol{x}_{ij} \boldsymbol{x}_{ij}^{\mathsf{H}}}{\|\boldsymbol{x}_{ij}\|^2}, \tag{85}$$

$$\boldsymbol{\Phi}_{ij}^{(\mathrm{n})} = r_{ij}^{(\mathrm{n})} \mathbf{R}_i^{(\mathrm{n})} (\mathbf{R}_{ij}^{(\mathrm{x})})^{-1} \frac{\boldsymbol{x}_{ij} \boldsymbol{x}_{ij}^{\mathsf{H}}}{\|\boldsymbol{x}_{ij}\|^2}. \tag{86}$$

Next, since $\log \det (\cdot)$ is a concave function for any positive definite Hermitian matrix, the following inequality derived from the relationship between a concave function and its tangent plane can be applied to the second term of (67):

$$\log \det \mathbf{R}_{ij}^{(\mathrm{x})} \leq \log \det \boldsymbol{\Psi}_{ij} + \mathrm{tr}(\boldsymbol{\Psi}_{ij}^{-1} (\mathbf{R}_{ij}^{(\mathrm{x})} - \boldsymbol{\Psi}_{ij})), \tag{87}$$

where $\boldsymbol{\Psi}_{ij} \in \mathbb{C}^{M \times M}$ is a positive semidefinite Hermitian matrix and the equality of (87) holds if and only if $\boldsymbol{\Psi}_{ij} = \mathbf{R}_{ij}^{(\mathrm{x})}$.

For the third term of (67), from the relationship between the logarithmic function and its tangent, it holds that

$$\log r_{ij}^{(\mathrm{t})} \leq \frac{r_{ij}^{(\mathrm{t})} - \zeta_{ij}}{\zeta_{ij}} + \log \zeta_{ij}, \tag{88}$$

where ζ_{ij} is positive and the equality of (88) holds if and only if $\zeta_{ij} = r_{ij}^{(\mathrm{t})}$.

By combining (83), (87), and (88), we can design the auxiliary function $\mathcal{L}^{\mathrm{U}}(\Theta_{\mathrm{p}}, \Omega_{\mathrm{p}})$ as (18).

Abbreviations

BSE	Blind speech extraction
BSS	Blind source separation
FDICA	Frequency-domain independent component analysis
ILRMA	Independent low-rank matrix analysis
SCM	Spatial covariance matrix
MNMF	Multichannel non-negative matrix factorization
RCSCME	Rank-constrained spatial covariance matrix estimation

MWF	Multichannel Wiener filtering
GGD	Generalized Gaussian distribution
MM	Majorization-minimization
ME	Majorization-equalization
STFT	Short-time Fourier transform
IVE	Independent vector extraction
BSSA	Blind spatial subtraction array
SDR	Source-to-distortion ratio

Availability of data and materials
Not available online. Please contact author for data requests.

Declarations

Ethical approval and consent to participate
All experiments are the computer simulations that used the acoustical databases, and do not relate with human and animals in this work.

Competing interests
The authors declare that they have no competing interests.

References
1. H. Sawada, N. Ono, H. Kameoka, D. Kitamura, H. Saruwatari, A review of blind source separation methods: two converging routes to ILRMA originating from ICA and NMF. APSIPA Trans. Signal Inf. Process. **8**(e12), 1–14 (2019)
2. Y. Takahashi, T. Takatani, K. Osako, H. Saruwatari, K. Shikano, Blind spatial subtraction array for speech enhancement in noisy environment. IEEE Trans. ASLP **17**(4), 650–664 (2009)
3. Z. Koldovský, P. Tichavský, Gradient algorithms for complex non-Gaussian independent component/vector extraction, question of convergence. IEEE Trans. SP **67**(4), 1050–1064 (2019)
4. P. Smaragdis, Blind separation of convolved mixtures in the frequency domain. Neurocomputing **22**(1–3), 21–34 (1998)
5. S. Araki, R. Mukai, S. Makino, T. Nishikawa, H. Saruwatari, The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech. IEEE Trans. ASP **11**(2), 109–116 (2003)
6. H. Saruwatari, T. Kawamura, T. Nishikawa, A. Lee, K. Shikano, Blind source separation based on a fast-convergence algorithm combining ICA and beamforming. IEEE Trans. ASLP **14**(2), 666–678 (2006)
7. A. Hiroe, Solution of permutation problem in frequency domain ICA using multivariate probability density functions, in *Proceedings of ICA* (2006), pp. 601–608
8. T. Kim, H.T. Attias, S.-Y. Lee, T.-W. Lee, Blind source separation exploiting higher-order frequency dependencies. IEEE Trans. ASLP **15**(1), 70–79 (2007)
9. D. Kitamura, N. Ono, H. Sawada, H. Kameoka, H. Saruwatari, Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization. IEEE/ACM Trans. ASLP **24**(9), 1626–1641 (2016)
10. D. Kitamura, S. Mogami, Y. Mitsui, N. Takamune, H. Saruwatari, N. Ono, Y. Takahashi, K. Kondo, Generalized independent low-rank matrix analysis using heavy-tailed distributions for blind source separation. EURASIP J. Adv. Signal Process. **2018**(1), 1–28 (2018)
11. R. Ikeshita, Y. Kawaguchi, Independent low-rank matrix analysis based on multivariate complex exponential power distribution, in *Proceedings of ICASSP* (2018), pp. 741–745
12. N.Q.K. Duong, E. Vincent, R. Gribonval, Under-determined reverberant audio source separation using a full-rank spatial covariance model. IEEE Trans. ASLP **18**(7), 1830–1840 (2010)
13. A. Ozerov, C. Févotte, Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation. IEEE Trans. ASLP **18**(3), 550–563 (2010)
14. H. Sawada, H. Kameoka, S. Araki, N. Ueda, Multichannel extensions of non-negative matrix factorization with complex-valued data. IEEE Trans. ASLP **21**(5), 971–982 (2013)

15. K. Kitamura, Y. Bando, K. Itoyama, K. Yoshii, Student's *t* multichannel nonnegative matrix factorization for blind source separation, in *Proceedings of IWAENC* (2016)
16. N. Ito, T. Nakatani, FastMNMF: joint diagonalization based accelerated algorithms for multichannel nonnegative matrix factorization, in *Proceedings of ICASSP* (2019), pp. 371–375
17. K. Sekiguchi, Y. Bando, A.A. Nugraha, K. Yoshii, T. Kawahara, Fast multichannel nonnegative matrix factorization with directivity-aware jointly-diagonalizable spatial covariance matrices for blind source separation. IEEE Trans. ASLP **28**, 2610–2625 (2020)
18. K. Kamo, Y. Kubo, N. Takamune, D. Kitamura, H. Saruwatari, Y. Takahashi, K. Kondo, Joint-diagonalizability-constrained multichannel nonnegative matrix factorization based on multivariate complex Student's *t*-distribution, in *Proceedings of APSIPA* (2020)
19. Y. Kubo, N. Takamune, D. Kitamura, H. Saruwatari, Blind speech extraction based on rank-constrained spatial covariance matrix estimation with multivariate generalized Gaussian distribution. IEEE/ACM Trans. ASLP **28**, 1948–1963 (2020)
20. D.R. Hunter, K. Lange, Quantile regression via an MM algorithm. J. Comput. Graph. Stat. **9**(1), 60–77 (2000)
21. C. Févotte, J. Idier, Algorithms for nonnegative matrix factorization with the *β*-divergence. Neural Comput. **23**(9), 2421–2456 (2011)
22. Y. Kondo, Y. Kubo, N. Takamune, D. Kitamura, H. Saruwatari, Deficient basis estimation of noise spatial covariance matrix for rank-constrained spatial covariance matrix estimation method in blind speech extraction, in *Proceedings of ICASSP* (2021), pp. 806–810
23. E. Gómez, M. Gomez-Viilegas, J.M. Marín, A multivariate generalization of the power exponential family of distributions. Commun. Stat. Theory Methods **27**(3), 589–600 (1998)
24. N. Murata, S. Ikeda, A. Ziehe, An approach to blind source separation based on temporal structure of speech signals. Neurocomputing **41**(1–4), 1–24 (2001)
25. B. Kulis, M.A. Sustik, I.S. Dhillon, Low-rank kernel learning with Bregman matrix divergences. J. Mach. Learn. Res. **10**, 341–376 (2009)
26. J. Ding, A. Zhou, Eigenvalues of rank-one updated matrices with some applications. Appl. Math. Lett. **20**(12), 1223–1226 (2007)
27. K. Itou, M. Yamamoto, K. Takeda, T. Takezawa, T. Matsuoka, T. Kobayashi, K. Shikano, S. Itahashi, JNAS: Japanese speech corpus for large vocabulary continuous speech recognition research. J. Acoust. Soc. Jpn. (E) **20**(3), 199–206 (1999)
28. J. Thiemann, N. Ito, E. Vincent, DEMAND: a collection of multi-channel recordings of acoustic noise in diverse environments (2013). https://doi.org/10.5281/zenodo.1227121
29. I. Cohen, B. Berdugo, Speech enhancement for non-stationary noise environments. Signal Process. **81**(11), 2403–2418 (2001)
30. R. Miyazaki, H. Saruwatari, R. Wakisaka, K. Shikano, T. Takatani, Theoretical analysis of parametric blind spatial subtraction array and its application to speech recognition performance prediction, in *Proceedings of HSCMA* (2011), pp. 19–24
31. K. Shimada, Y. Bando, M. Mimura, K. Itoyama, K. Yoshii, T. Kawahara, Unsupervised beamforming based on multichannel nonnegative matrix factorization for noisy speech recognition, in *Proceedings of ICASSP* (2018), pp. 5734–5738
32. Y. Ephraim, D. Malah, Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. IEEE Trans. ASSP **32**(6), 1109–1121 (1984)
33. E. Vincent, R. Gribonval, C. Févotte, Performance measurement in blind audio source separation. IEEE Trans. ASLP **14**(4), 1462–1469 (2006)

Publisher's Note