Experimental Analysis of EM and MU Algorithms for Optimizing Full-rank Spatial Covariance Model

Hiroshi Sawada, Rintaro Ikeshita, and Tomohiro Nakatani
NTT Communication Science Laboratories, NTT Corporation, Kyoto, Japan

***Abstract*—Full-rank spatial covariance analysis (FCA) is based on a flexible source model, and achieves high-quality results for blind source separation. An expectation-maximization (EM) algorithm as well as a multiplicative update (MU) algorithm are known to optimize the FCA model parameters. In this paper, we first investigate the behaviors of both algorithms. We observed that the MU algorithm minimizes the FCA objective function faster than the EM algorithm, but the separation performance at the converged point is better by the EM algorithm than the MU algorithm. We found that the MU algorithm tends to push the covariance matrices towards rank deficient. To mitigate this tendency, we propose a modified FCA model where the temporal parameters are shared within a time block. Experimental results show that the modified model provides better separation performance not only by the MU algorithm but also by the EM algorithm.**

***Index Terms*—blind source separation (BSS), full-rank spatial covariance analysis (FCA), expectation-maximization (EM) algorithm, multiplicative update (MU) algorithm, rank deficient**

I. Introduction

Blind source separation (BSS) has been studied for several decades [1–5] with inventing various methods. Independent component analysis (ICA) [2] is a well-established method in which the mixing system is assumed to be invertible. Full-rank spatial covariance analysis (FCA) [6, 7], on the other hand, models a more flexible mixing system than ICA does. The most crucial difference is that FCA can be applied to an underdetermined case where the number N of sources is larger than the number M of sensors (e.g., microphones in audio cases) while ICA cannot be because of the invertible assumption. FCA has several merits over ICA thanks to the flexibility.

An expectation-maximization (EM) algorithm [6, 7] and a multiplicative update (MU) algorithm [8–10] have been proposed for FCA. In both algorithms, the demanding computation for calculating many inverse matrices is a practical issue. Recently, two methods have been proposed to solve this issue. The first one [10–12] assumes that all the spatial covariance matrices can be jointly diagonalized. This assumption exactly holds when $N = 2$, but approximates the mixing model for a larger number $N \geq 3$ of sources. The second one [13] accelerates the computation by using single-instruction-multiple-data (SIMD) instructions executed on a graphics processing unit (GPU). Either of these two makes FCA easier to employ in a practical BSS task than before.

Real-world audio BSS tasks deal with convolutive mixtures with delays and reverberations. For efficient computation, the time-domain mixtures are typically transformed into frequency-domain time-series mixtures by using a short-time Fourier transform (STFT). There are two distinct approaches for frequency-domain BSS. The first one is a narrowband approach where all the model parameters are disjoint among different frequency bins. Post-processing is needed to align the permutation ambiguities that occurred in the BSS problem [14]. The original ICA and FCA are classified here. The second one is a broadband approach in which some model parameters are shared among frequency bins. Permutation ambiguities are expected to automatically align thanks to the shared parameters. As ICA has been extended to independent vector analysis (IVA) [15, 16] and others [17], FCA has also been extended to richer models [8, 18–22] in a broadband approach.

In this paper, we study the behaviors of the EM and MU algorithms for FCA in a narrowband approach, which attains good separations for speech mixtures by the post-processing permutation alignment [14]. Regarding the above-mentioned methods for computational efficiency, we employ the second one [13] to examine the model parameters precisely without approximation. In Sect. III, we focus on the rank-deficient problem of spatial covariance matrices frequently caused by the MU algorithm. We then propose a modified model and algorithms in Sect. IV to mitigate the rank-deficient problem. Experimental results show that the modified model contributes to better separation performance with both EM and MU algorithms.

II. Full-rank spatial Covariance Analysis

A. Model and objective function

Suppose that $n = 1, \ldots, N$ sources are mixed and observed at $m = 1, \ldots, M$ sensors. Let the sensor observations at time frame t and frequency bin f, $t = 1, \ldots, T$ and $f = 1, \ldots, F$, be denoted by an M-dimensional complex vector $\mathbf{x}_{tf} \in \mathbb{C}^M$ with $\mathbf{x}_{tf} = [x_{1tf}, \ldots, x_{Mtf}]^{\mathsf{T}}$. In FCA, a mixture vector $\mathbf{x}_{tf}$ follows a zero-mean multivariate complex Gaussian distribution ($\cdot^*$ denotes the Hermitian transpose of a vector)

$$p(\mathbf{x}_{tf} \mid \mathbf{0}, \hat{\mathsf{X}}_{tf}) \propto \frac{1}{\det \hat{\mathsf{X}}_{tf}} \exp\left(-\mathbf{x}_{tf}^* \hat{\mathsf{X}}_{tf}^{-1} \mathbf{x}_{tf}\right) \qquad (1)$$

with a covariance matrix

$$\hat{\mathsf{X}}_{tf} = \textstyle\sum_{n=1}^{N} v_{ntf} \mathsf{A}_{nf} + \epsilon\, \mathsf{I}\,, \qquad (2)$$

where A_{nf} is a spatial covariance matrix that encodes the time-invariant spatial property from source n to all M sensors and is assumed to be Hermitian and positive definite. A positive

scalar v_{ntf} represents the temporal power of source n at time frame t. The last term $\epsilon\,\mathsf{I}$ can be regarded as that of background noise and contributes to keeping $\hat{\mathsf{X}}_{tf}$ full-rank, where I is an identity matrix and ϵ is a very small positive number, e.g., 10^{-12} that is slightly larger than the machine epsilon 2.2204×10^{-16} for a double-precision floating-point number. The parameters for frequency bin f are summarized as

$$\theta_f = \{\mathsf{A}_{nf}, \{v_{ntf}\}_{t=1}^{T}\}_{n=1}^{N}\,, \tag{3}$$

and optimized independently for each frequency bin.

The parameters θ_f can be optimized in a maximum likelihood sense, equivalently by minimizing the negative log-likelihood $\mathcal{C}(\theta_f) = -\log p(\{\mathbf{x}_{tf}\}_{t=1}^{T} \mid \theta_f)$. We assume that the likelihood is decomposed into those of each time frame as

$$p(\{\mathbf{x}_{tf}\}_{t=1}^{T} \mid \theta_f) = \textstyle\prod_{t=1}^{T} p(\mathbf{x}_{tf} \mid \mathbf{0}, \hat{\mathsf{X}}_{tf})\,. \tag{4}$$

Consequently, we have the objective functions to be minimized

$$\mathcal{C}(\theta_f) = \textstyle\sum_{t=1}^{T} \left[\mathbf{x}_{tf}^{*}\hat{\mathsf{X}}_{tf}^{-1}\mathbf{x}_{tf} + \log\det\hat{\mathsf{X}}_{tf}\right] \tag{5}$$

with (2) and (3).

B. Expectation-Maximization (EM) algorithm

The objective function (5) can be minimized by an EM algorithm [6, 7], where the observation vector $\mathbf{x}_{tf}$ is expressed as a summation

$$\mathbf{x}_{tf} = \textstyle\sum_{n=1}^{N} \mathbf{y}_{ntf} \tag{6}$$

of latent variables $\{\mathbf{y}_{ntf}\}_{n=1}^{N}$. The EM algorithm iterates the following **E-step** and **M-step** for a convergence.

E-step calculates the conditional expectations $\mathbb{E}[\mathbf{y}_{ntf}\mathbf{y}_{ntf}^{*} \mid \mathbf{x}_{tf}, \theta]$ of the outer product of latent vectors $\mathbf{y}_{ntf}$ as

$$\tilde{\mathsf{Y}}_{ntf} = \hat{\mathsf{Y}}_{ntf} + \hat{\mathsf{Y}}_{ntf}\left(\hat{\mathsf{X}}_{tf}^{-1}\mathbf{x}_{tf}\mathbf{x}_{tf}^{*}\hat{\mathsf{X}}_{tf}^{-1} - \hat{\mathsf{X}}_{tf}^{-1}\right)\hat{\mathsf{Y}}_{ntf}\,, \tag{7}$$

where $\hat{\mathsf{Y}}_{ntf}$ is the n-th source component of $\hat{\mathsf{X}}_{tf}$ and defined as $\hat{\mathsf{Y}}_{ntf} = v_{ntf}\mathsf{A}_{nf}$.

M-step updates the model parameters by

$$v_{ntf} \leftarrow \max\left(\frac{1}{M}\mathrm{tr}\left(\mathsf{A}_{nf}^{-1}\tilde{\mathsf{Y}}_{ntf}\right), \epsilon\right)\,, \tag{8}$$

$$\mathsf{A}_{nf} \leftarrow \frac{1}{T}\sum_{t=1}^{T}\frac{1}{v_{ntf}}\tilde{\mathsf{Y}}_{ntf} + \epsilon\,\mathsf{I}\,. \tag{9}$$

In (8) and (9) together with upcoming (10) and (11), the ϵ-terms are used for ensuring the positivity of v and the positive definiteness of A.

C. Multiplicative Update (MU) algorithm

The objective function (5) can also be minimized by multiplicative update rules [8–10], which are popularly used for nonnegative matrix factorization (NMF) [23, 24]. Actually, the FCA model can be regarded as a special case of multichannel Itakura-Saito NMF [8].

The MU algorithm iterates the following **V-update** and **A-update** for a convergence.

V-update

$$v_{ntf} \leftarrow \max\left(v_{ntf}\sqrt{\frac{\mathbf{x}_{tf}^{*}\hat{\mathsf{X}}_{tf}^{-1}\mathsf{A}_{nf}\hat{\mathsf{X}}_{tf}^{-1}\mathbf{x}_{tf}}{\max(\mathrm{tr}(\hat{\mathsf{X}}_{tf}^{-1}\mathsf{A}_{nf}), \epsilon)}}, \epsilon\right) \tag{10}$$

A-update

$$\mathsf{A}_{nf} \leftarrow \mathsf{F}^{-1}\#(\mathsf{A}_{nf}\mathsf{G}\mathsf{A}_{nf}) + \epsilon\,\mathsf{I}\,, \tag{11}$$

where $\#$ calculates the geometric mean [25, 26]

$$\mathsf{B}\#\mathsf{C} = \mathsf{B}^{\frac{1}{2}}\left(\mathsf{B}^{-\frac{1}{2}}\mathsf{C}\mathsf{B}^{-\frac{1}{2}}\right)^{\frac{1}{2}}\mathsf{B}^{\frac{1}{2}} \tag{12}$$

of two positive definite matrices. The matrices F and G are defined as

$$\mathsf{F} = \textstyle\sum_{t=1}^{T} v_{ntf}\hat{\mathsf{X}}_{tf}^{-1}\,,\quad \mathsf{G} = \sum_{t=1}^{T} v_{ntf}\hat{\mathsf{X}}_{tf}^{-1}\mathbf{x}_{tf}\mathbf{x}_{tf}^{*}\hat{\mathsf{X}}_{tf}^{-1}\,. \tag{13}$$

Regarding the geometric mean calculation, the Cholesky-Schur method in [25] is employed in this paper for ensuring the resultant A_{nf} to be Hermitian.

D. Initialization of parameters

Before performing the EM or MU algorithm, the parameters $\{\theta_f\}_{f=1}^{F}$ should be initialized. In the experiments explained in the following sections, the covariance matrices A_{nf} were initialized by the online clustering-based method [13] that provides us with a good initialization. The temporal powers were initialized as $v_{ntf} = 1$.

E. Source separation

Once the parameters θ_f are optimized, separated signals are obtained typically by the multichannel Wiener filter as

$$\tilde{\mathbf{y}}_{ntf} = \mathbb{E}[\mathbf{y}_{ntf} \mid \mathbf{x}_{tf}, \theta_f] = \hat{\mathsf{Y}}_{ntf}\hat{\mathsf{X}}_{tf}^{-1}\mathbf{x}_{tf}\,. \tag{14}$$

III. Algorithm Behaviors

In this section, we study the behaviors of the EM and MU algorithms for FCA. One actual experimental example is taken based on the experimental conditions explained in Sect. V. Specifically, the number N of sources was two, and the room reverberation time was 270 ms.

A. Convergence and separation performance

Figure 1 shows a typical example of how the EM and MU algorithms behave for FCA. Generally, the MU algorithm minimizes the objective function faster than the EM algorithm. With enough number of iterations, e.g., 100, both algorithms converge to similar objective function values. Such behaviors have been recognized, especially in the multichannel NMF [8]. On the other hand, the separation performances measured in signal-to-distortion ratios (SDRs) [27] tend to be higher by the EM algorithm than by the MU algorithm, even if the objective function values are similar. This tendency might be specific to FCA and have not been well recognized in FCA extensions to richer models [8, 18–22].

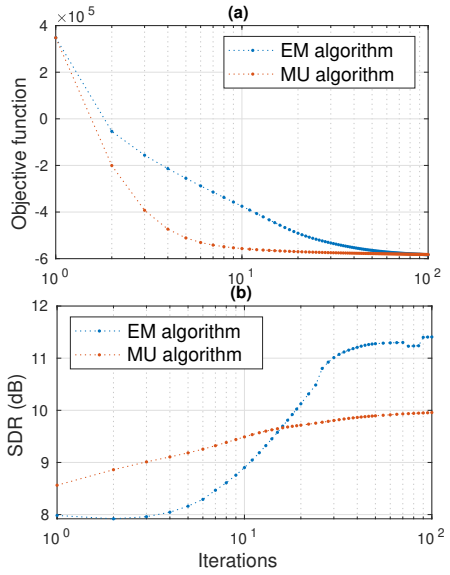

Fig. 1. Convergence examples of the EM and MU algorithms. (a) The EM algorithm converged slower than the MU algorithm. (b) The separation performance measured in signal-to-distortion ratios (SDRs) [27]. The MU algorithm improved SDRs rapidly but hit the ceiling at around 10 dB. The EM algorithm improved SDRs slowly and eventually reached a higher SDR value than that of the MU algorithm.

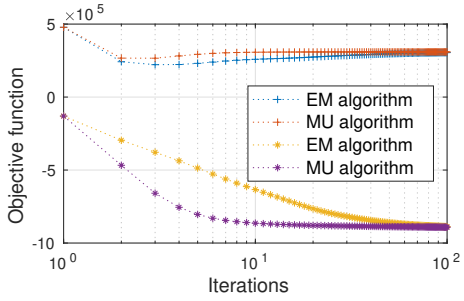

Fig. 2. (continued from Fig. 1) (a) The objective function values are decomposed into the $\mathbf{x}_{tf}^* \hat{\mathsf{X}}_{tf}^{-1} \mathbf{x}_{tf}$ terms ('+') and the $\log\det \hat{\mathsf{X}}_{tf}$ terms ('*'). The fast convergence of the MU algorithm attained by the fast minimization of the $\log\det \hat{\mathsf{X}}_{tf}$ terms.

B. Rank deficiency

To study the reason for the tendency mentioned above, we display the objective functions separately in Fig. 2. We notice that the $\log\det \hat{\mathsf{X}}_{tf}$ terms of the objective function (5) were minimized faster by the MU algorithm than by the EM algorithm. Minimization of $\log\det \hat{\mathsf{X}}_{tf}$ favors $\hat{\mathsf{X}}_{tf}$ to be rank deficient. When $\hat{\mathsf{X}}_{tf}$ is modeled with a plenty number of parameters and can be optimized without any restriction, it is natural that $\hat{\mathsf{X}}_{tf}$ changes towards rank deficient because $\hat{\mathsf{X}}_{tf}$ is responsible for only one observation $\mathbf{x}_{tf}$ whose covariance matrix is rank-1. The FCA model (2) falls into such a case because there are too many parameters $N \cdot T + N \cdot M^2/2$ for $T \cdot M$ observations of each frequency bin f. In contrast to this, the aforementioned FCA extensions [8, 18–22] do not exhibit such rank deficiency because the parameters are structurally shared.

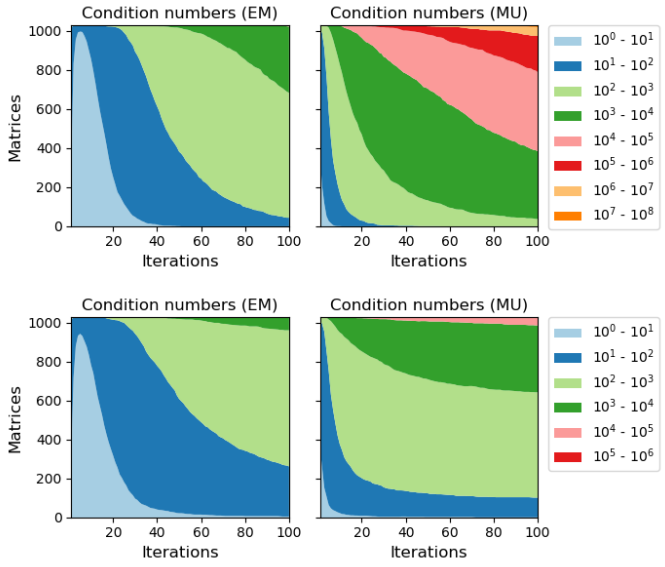

Fig. 3. Distributions of the condition numbers of matrices A_{nf} as the iteration went on. Upper: the original model (2). Lower: the modified model (15). The vertical axis represents the number of matrices whose condition numbers fall into the ranges shown in the legend.

We then examined how the spatial covariance matrices A_{nf} changed as iterations went on by looking at the distributions of matrix condition numbers. The upper half of Fig. 3 shows that the MU algorithm pushed a large portion of the A_{nf} matrices towards rank deficient with large condition numbers. On the other hand, the EM algorithm kept them stable with moderate condition numbers. Such situations might explain the EM and MU algorithms' behaviors regarding the minimization of the $\log\det \hat{\mathsf{X}}_{tf}$ terms.

IV. Modified Model and Algorithms

In this section, we propose a modified model for FCA to mitigate the tendency towards rank deficient. Based on the previous section's discussion, one approach is to impose some restrictions on the model. We here propose to share the parameters along the time axis. Thus the number of parameters is reduced from (2). Let the total T time frames are partitioned $\bigcup_{b=1}^{B} \mathcal{T}_b = \{1, \ldots, T\}$ into B time blocks $\mathcal{T}_b$, $b = 1, \ldots, B$. In the modified model, the temporal power v_{nbf} of source n should be the same in a time block b:

$$\hat{\mathsf{X}}_{tf} = \textstyle\sum_{n=1}^{N} v_{nbf} \mathsf{A}_{nf} + \epsilon \mathsf{I}, \quad t \in \mathcal{T}_b \,. \tag{15}$$

The EM and MU algorithms are easily modified to reflect this model modification from (2) to (15). In the EM algorithm, **E-step** calculates (7) with $\hat{\mathsf{Y}}_{ntf} = v_{nbf} \mathsf{A}_{nf}$, $t \in \mathcal{T}_b$, and **M-step** updates the parameters by

$$v_{nbf} \leftarrow \max\left(\frac{1}{M \cdot |\mathcal{T}_b|} \sum_{t \in \mathcal{T}_b} \mathrm{tr}\left(\mathsf{A}_{nf}^{-1} \tilde{\mathsf{Y}}_{ntf} \right), \epsilon \right), \tag{16}$$

$$\mathsf{A}_{nf} \leftarrow \frac{1}{T} \sum_{t=1}^{T} \frac{1}{v_{nbf}} \tilde{\mathsf{Y}}_{ntf} + \epsilon \mathsf{I} \,. \tag{17}$$

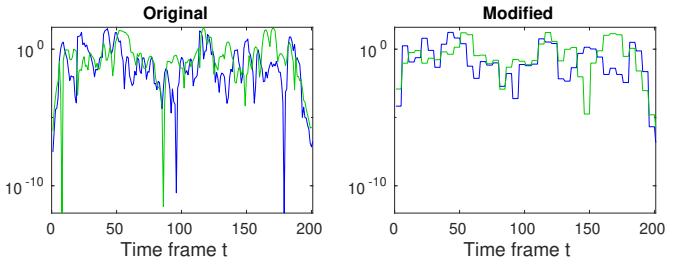

Fig. 4. Temporal powers v_{ntf} (original) and v_{nbf} (modified, $|\mathcal{T}_b| = 5$) of two sources optimized by the MU algorithm with 100 iterations for some frequency bin f.

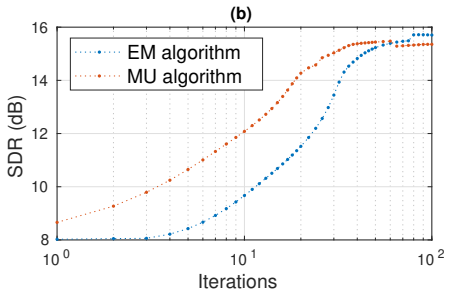

Fig. 5. (compared with Fig. 1 (b)) The separation performance by the modified EM and MU algorithms with $|\mathcal{T}_b| = 5$. Both EM and MU algorithm attained higher SDR values than the original ones.

The MU algorithm is described by **V-update**

$$v_{nbf} \leftarrow \max\left(v_{nbf}\sqrt{\frac{\sum_{t\in\mathcal{T}_b}\mathbf{x}_{tf}^*\hat{\mathsf{X}}_{tf}^{-1}\mathsf{A}_{nf}\hat{\mathsf{X}}_{tf}^{-1}\mathbf{x}_{tf}}{\max(\sum_{t\in\mathcal{T}_b}\mathrm{tr}(\hat{\mathsf{X}}_{tf}^{-1}\mathsf{A}_{nf}),\epsilon)}},\epsilon\right) \tag{18}$$

and **A-update** (11) with

$$\mathsf{F} = \sum_{t=1}^{T} v_{nbf}\hat{\mathsf{X}}_{tf}^{-1}\,,\quad \mathsf{G} = \sum_{t=1}^{T} v_{nbf}\hat{\mathsf{X}}_{tf}^{-1}\mathbf{x}_{tf}\mathbf{x}_{tf}^*\hat{\mathsf{X}}_{tf}^{-1}\,. \tag{19}$$

For (17) and (19), in the summation from $t = 1$ to T, the index b is specified so that $t \in \mathcal{T}_b$.

Figure 4 shows examples of converged temporal powers v_{ntf} and v_{nbf}. In the original case with (2), the temporal powers took various values and some went down to the floor $\epsilon = 10^{-12}$. In the modified case with (15), the temporal powers were averaged within a block of size 5. The lower half of Fig. 3 shows that the condition numbers became stable with the modified model (15) even when the MU algorithm was employed. Figure 5 shows that the separation performances for the same mixtures with those of Fig. 1 were considerably improved by the modified EM and MU algorithms.

Empirically speaking, the modified EM or MU algorithm itself was effective for cases with enough number $M \geq N$ of sensors. However, for underdetermined cases $M < N$, time-invariant parameters v_{nbf} within a block were not optimal. Therefore, we further update the v_{ntf} parameters by the original EM (8) or MU (10) algorithm with several iterations while keeping the A_{nf} parameters unchanged.

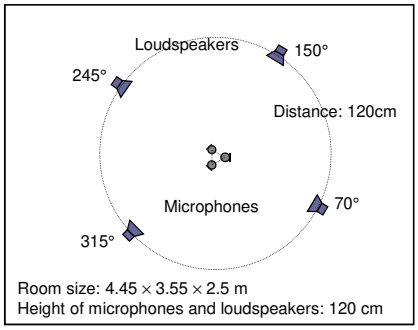

Fig. 6. Experimental setup

TABLE I
COMPUTATIONAL TIME (IN SECONDS) FOR 6-SECOND SPEECH MIXTURES

	$N=2$	$N=3$	$N=4$
EM	5.15	5.72	7.04
MU	6.46	7.57	9.29

V. EXPERIMENTS

We performed experiments to separate from two to four speech sources ($N = 2, 3, 4$) with three microphones ($M = 3$). We measured the impulse responses from the sources (the loudspeakers) to the microphones under the room conditions shown in Fig. 6. The room reverberation time was varied from 130 ms to 450 ms. The mixtures at the microphones were constructed by convolving the impulse responses and 6-second English speech sources. No noise was added to the mixtures. The sampling frequency was 8 kHz. The frame width and shift of STFT were 128 ms and 32 ms, respectively. Consequently, the numbers of time frames and frequency bins were $T = 201$ and $F = 513$, respectively.

For $N = 2, 3$ ($\leq M$) sources, the modified model (15) was optimized with EM or MU 90 iterations. For $N = 4$ ($> M$) sources, (15) was optimized with 90 EM or MU iterations and then the v_{ntf} parameters were updated by the original EM (8) or MU (10) algorithm with 10 iterations while keeping the A_{nf} parameters unchanged. The algorithms were coded with Matlab R2019b and run on an Intel Core i7-8700K (3.70GHz) processor together with GeForce GTX 1080 Ti as a GPU. Table I shows that the computational times were practical with the help of the GPU.

Figure 7 reports how the modified model (15) contributed to better BSS results. We observe that the modification generally improved the results compared to the original model (2), which corresponds to the one with block size $|\mathcal{T}_b| = 1$. Although we intended to solve the rank deficient problem of the MU algorithm, both EM and MU algorithms improved the results, especially for the $N = 2$ cases. We observe that the block size of $|\mathcal{T}_b| = 5$ is an appropriate choice for all the cases.

Figure 8 shows the separation results for $N = 4$ sources with $M = 3$ microphones under various room reverberation times. The modified EM and MU algorithms improved the results in almost every reverberant condition over the existing algorithms EM (1) and MU (1).

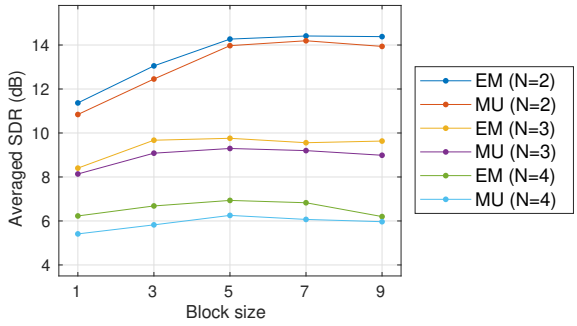

Fig. 7. Source separation performance by the modified EM and MU algorithms. The horizontal axis shows the block size $|\mathcal{T}_b|$, and $|\mathcal{T}_b| = 1$ corresponds to the original model (2). SDRs are averaged over eight combinations of sources. The room reverberation time was 270 ms.

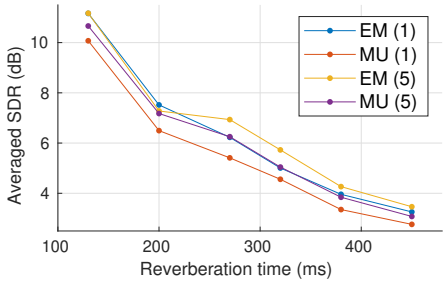

Fig. 8. Results for separating $N = 4$ sources under various room reverberation times. The numbers 1 or 5 in the parentheses indicate the block size $|\mathcal{T}_b|$. Therefore, EM (5) and MU (5) correspond to the modified algorithms. SDRs are averaged over eight combinations of sources.

VI. Conclusion

In this paper, we examined the behaviors of the EM and MU algorithms for FCA. We found that the rank deficient problem occurred with the MU algorithm. To solve this problem, we have modified the FCA model so that the parameters regarding temporal powers are shared in a time block. Experimental results show that the modified model led to better separation performance by both algorithms. However, still the EM algorithm generally performs better than the MU algorithm. For further investigation, more experiences by various researchers would be desired.

References

[1] C. Jutten and J. Herault, "Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture," *Signal processing*, vol. 24, no. 1, pp. 1–10, 1991.

[2] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. John Wiley & Sons, 2001.

[3] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing*. John Wiley & Sons, 2002.

[4] S. Makino, T.-W. Lee, and H. Sawada, Eds., *Blind Speech Separation*. Springer, 2007.

[5] E. Vincent, T. Virtanen, and S. Gannot, *Audio source separation and speech enhancement*. John Wiley & Sons, 2018.

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

[7] S. Arberet, A. Ozerov, N. Duong, E. Vincent, R. Gribonval, F. Bimbot, and P. Vandergheynst, "Nonnegative matrix factorization and spatial covariance model for under-determined reverberant audio source separation," in *Proc. ISSPA 2010*, May 2010, pp. 1–4.

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

[9] K. Yoshii, R. Tomioka, D. Mochihashi, and M. Goto, "Infinite positive semidefinite tensor factorization for source separation of mixture signals," in *Proc. ICML*, Jun. 2013, pp. 576–584.

[10] N. Ito and T. Nakatani, "Multiplicative updates and joint diagonalization based acceleration for under-determined BSS using a full-rank spatial covariance model," in *Proc. GlobalSIP*, 2018, pp. 231–235.

[11] N. Ito, S. Araki, and T. Nakatani, "FastFCA: Joint diagonalization based acceleration of audio source separation using a full-rank spatial covariance model," in *Proc. EUSIPCO*, 2018, pp. 1667–1671.

[12] N. Ito and T. Nakatani, "Fastfca-as: Joint diagonalization based acceleration of full-rank spatial covariance analysis for separating any number of sources," in *Proc. IWAENC*, 2018.

[13] H. Sawada, R. Ikeshita, N. Ito, and T. Nakatani, "Computational acceleration and smart initialization of full-rank spatial covariance analysis," in *Proc. EUSIPCO*, 2019, pp. 1–5.

[14] H. Sawada, S. Araki, and S. Makino, "Underdetermined convolutive blind source separation via frequency bin-wise clustering and permutation alignment," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 19, no. 3, pp. 516–527, Mar. 2011.

[15] T. Kim, H. T. Attias, S.-Y. Lee, and T.-W. Lee, "Blind source separation exploiting higher-order frequency dependencies," *IEEE Trans. Audio, Speech and Language Processing*, vol. 15, no. 1, pp. 70–79, Jan. 2007.

[16] N. Ono and S. Miyabe, "Auxiliary-function-based independent component analysis for super-Gaussian sources," in *Proc. International Conference on Latent Variable Analysis and Signal Separation*. Springer, 2010, pp. 165–172.

[17] H. Buchner, R. Aichner, and W. Kellermann, "A generalization of blind source separation algorithms for convolutive mixtures based on second-order statistics," *IEEE trans. speech and audio processing*, vol. 13, no. 1, pp. 120–134, 2004.

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

[19] M. Togami, Y. Kawaguchi, R. Takeda, Y. Obuchi, and N. Nukaga, "Optimized speech dereverberation from probabilistic perspective for time varying acoustic transfer function," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 7, pp. 1369–1380, 2013.

[20] T. Otsuka, K. Ishiguro, H. Sawada, and H. G. Okuno, "Bayesian nonparametrics for microphone array processing," *IEEE/ACM Trans. Audio, Speech, and Language Processing*, vol. 22, no. 2, pp. 493–504, 2014.

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

[22] A. Liutkus, D. Fitzgerald, Z. Rafii, B. Pardo, and L. Daudet, "Kernel additive models for source separation," *IEEE Transactions on Signal Processing*, vol. 62, no. 16, pp. 4298–4310, 2014.

[23] D. Lee and H. Seung, "Algorithms for non-negative matrix factorization," in *Advances in Neural Information Processing Systems*, vol. 13, 2001, pp. 556–562.

[24] M. Nakano, H. Kameoka, J. L. Roux, Y. Kitano, N. Ono, and S. Sagayama, "Convergence-guaranteed multiplicative algorithms for non-negative matrix factorization with beta-divergence," in *Proc. MLSP*, Aug. 2010, pp. 283–288.

[25] B. Iannazzo, "The geometric mean of two matrices from a computational viewpoint," *Numerical Linear Algebra with Applications*, vol. 23, no. 2, pp. 208–229, 2016.

[26] K. Yoshii, K. Kitamura, Y. Bando, E. Nakamura, and T. Kawahara, "Independent low-rank tensor analysis for audio source separation," in *Proc. EUSIPCO*, Sep. 2018.

[27] E. Vincent, S. Araki, F. Theis, G. Nolte, P. Bofill, H. Sawada, A. Ozerov, V. Gowreesunker, D. Lutter, and N. Duong, "The signal separation evaluation campaign (2007–2010): Achievements and remaining challenges," *Signal Processing*, vol. 92, no. 8, pp. 1928–1936, Aug. 2012.