Joint-Diagonalizability-Constrained Multichannel Nonnegative Matrix Factorization Based on Multivariate Complex Student's t-distribution

Keigo Kamo*, Yuki Kubo*, Norihiro Takamune*, Daichi Kitamura†
Hiroshi Saruwatari*, Yu Takahashi‡ and Kazunobu Kondo‡
* The University of Tokyo, Graduate School of Information Science and Technology, Tokyo, Japan
† National Institute of Technology, Kagawa Collage, Kagawa, Japan
‡ Yamaha Corporation, Shizuoka, Japan

Abstract—**In this paper, we propose the model generalization of a fast version of multichannel nonnegative matrix factorization (FastMNMF). FastMNMF is a blind source separation (BSS) method under the assumption that the spatial covariance matrices of multiple sources are jointly diagonalizable. To further improve its source-separation performance, we introduce a multivariate complex Student's *t*-distribution as a generative model, which includes a multivariate complex Gaussian distribution used in conventional FastMNMF. We derive a new parameter update rule using the auxiliary-function-based method and show the validity of the proposed method on the basis of BSS experiments using music sources.**

I. Introduction

Blind source separation (BSS) [1] is a technique of separating each sound source from observed mixtures without any prior information about the sources or the mixing system. In previous studies, various BSS methods have been proposed to improve the performance under the determined or overdetermined situation, such as frequency-domain independent component analysis [2], [3], independent vector analysis (IVA) [4]–[6], and independent low-rank matrix analysis (ILRMA) [7], [8]. These methods assume that the mixing system is instantaneous mixing in the time-frequency domain (referred to as the rank-1 spatial model) and invertible, and estimate the demixing system. In particular, ILRMA is the state-of-the-art BSS method that can efficiently and stably estimate the demixing system under the above assumptions. However, the rank-1 spatial model does not hold in the case of spatially spread sources or strong reverberation.

Multichannel nonnegative matrix factorization (MNMF) [9], [10] is an extension of nonnegative matrix factorization (NMF) [11] to multichannel cases, which estimates the spatial covariance matrices (SCMs) of each source. MNMF employs full-rank SCMs [12], and this model can simulate situations where, e.g., the reverberation is longer than the length of time-frequency analysis. However, it has been reported that MNMF has a huge computational cost and its performance strongly depends on the initial values of parameters [7]. To accelerate the parameter optimization, *FastMNMF* [13] has been proposed. FastMNMF is an improved algorithm of MNMF under the assumption of jointly diagonalizable SCMs, and this assumption contributes to the considerable reduction of computation time. However, it has been reported that its source-separation performance is still almost the same as that of original MNMF [14].

Recently, to achieve further improvement of the source-separation performance of BSS methods, the generative model generalization from a Gaussian distribution to a Student's t-distribution has been proposed, e.g., t-MNMF [15] and t-ILRMA [16]. Since the Student's t-distribution is less sensitive to outliers than the Gaussian distribution, it is expected to model source or observed signals well. It has been reported that this model generalization improves their performance in some cases [15]–[18].

In this paper, we generalize the generative model of FastMNMF from the multivariate complex Gaussian distribution to the multivariate complex Student's t-distribution; this is hereafter referred to as *t-FastMNMF*. By generalizing the generative model to the multivariate complex Student's t-distribution, we can change the shape of the distribution parametrically, which is expected to be applicable to various types of sources and further improve the source-separation accuracy. Next, we derive the update rules of the proposed method using the auxiliary function technique [19] that guarantees the monotonic nonincrease in the cost function. Finally, we conduct BSS experiments under reverberant conditions, showing that proposed t-FastMNMF outperforms conventional methods in source-separation accuracy.

II. Conventional Methods

A. Auxiliary function technique [19]

The auxiliary function technique is a type of iterative method for minimizing a function. This optimization method has been used in the derivation of update rules for conventional methods, e.g., MNMF or FastMNMF, and is also used in the proposed method.

Here is an overview of the auxiliary function technique. Let Θ be the parameter space and consider finding a solution $\hat{\theta}$ that minimizes the cost function $\mathcal{J}(\theta)$.

$$\hat{\theta} = \arg\min_{\theta\in\Theta}\mathcal{J}(\theta) \tag{1}$$

978-988-14768-8-3/20/$31.00 ©2020 APSIPA

The auxiliary function technique is effective for cases in which it is difficult to optimize $\mathcal{J}(\theta)$ for this parameter θ. In this technique, the following auxiliary function $\mathcal{J}^+$ is designed and optimized.

- $\mathcal{J}(\theta) \leq \mathcal{J}^+(\theta|\tilde{\theta})$ holds for any $\theta \in \Theta$, $\tilde{\theta} \in \tilde{\Theta}$.
- For any $\theta \in \Theta$, a $\tilde{\theta} \in \tilde{\Theta}$ exists and $\mathcal{J}(\theta) = \mathcal{J}^+(\theta|\tilde{\theta})$ holds.

Here, $\tilde{\theta}$ is an auxiliary variable. θ and $\tilde{\theta}$ in the auxiliary function $\mathcal{J}^+(\theta|\tilde{\theta})$ are alternatively optimized instead of θ in $\mathcal{J}(\theta)$. Thus, the parameter is updated as follows. First, let $\theta^{(l)}$ be the parameter of the lth iteration, and $\theta^{(0)}$ is initialized by a random point. Then, the following operations are performed on $l = 0, 1, 2 \ldots$ to update the parameters:

$$\tilde{\theta}^{(l+1)} \leftarrow \arg\min_{\tilde{\theta}\in\tilde{\Theta}} \mathcal{J}^+(\theta^{(l)}|\tilde{\theta}), \tag{2}$$

$$\theta^{(l+1)} \leftarrow \arg\min_{\theta\in\Theta} \mathcal{J}^+(\theta|\tilde{\theta}^{(l+1)}). \tag{3}$$

This update rule of parameters θ and $\tilde{\theta}$ is an optimization method that guarantees a monotonic nonincrease in the cost function.

B. MNMF [9], [10]

MNMF unifies the NMF-based source model and the full-rank SCM [12] and improves the BSS performance in a reverberant environment. The short-time Fourier transform (STFT) of the observed multichannel signal is defined as

$$\boldsymbol{x}_{ij} = (x_{ij,1}, \ldots, x_{ij,m}, \ldots, x_{ij,M})^{\mathsf{T}} \in \mathbb{C}^M, \tag{4}$$

where $i = 1, 2, \ldots, I$, $j = 1, 2, \ldots, J$, and $m = 1, 2, \ldots, M$ are the indices of the frequency bins, time frames, and channels, respectively, and $\cdot^{\mathsf{T}}$ denotes the transpose. The MNMF model assumes that the observed signal $\boldsymbol{x}_{ij}$ is distributed by the zero-mean multivariate complex Gaussian distribution as follows:

$$\boldsymbol{x}_{ij} \sim \mathcal{N}\Big(\mathbf{0}_M, \sum_n \sigma_{ijn}\boldsymbol{G}_{in}\Big), \tag{5}$$

where $\mathbf{0}_M \in \mathbb{C}^M$ is an M-dimensional zero vector and $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ indicates the multivariate complex Gaussian distribution whose mean is $\boldsymbol{\mu}$ and the covariance matrix is $\boldsymbol{\Sigma}$. $\boldsymbol{G}_{in}$ is the time-invariant SCM of the nth source at the ith frequency, which represents spatial characteristics of the source. The source model σ_{ijn} is the time-varying spectrogram of the nth source at the ith frequency and jth time frame, where $n = 1, 2, \ldots, N$ is the index of the sources. The source model σ_{ijn} has a low-rank spectral structure and can be factorized using NMF as follows:

$$\sigma_{ijn} = \sum_k t_{ik}v_{kj}z_{kn}, \tag{6}$$

where $k = 1, 2, \ldots, K$ is the index of the NMF basis, and $t_{ik} \in \mathbb{R}_{\geq 0}$ and $v_{kj} \in \mathbb{R}_{\geq 0}$ represent the ith frequency component of the kth basis and the jth time-frame activation component of the kth basis, respectively. In addition, $z_{kn} \in \mathbb{R}_{\geq 0}$ is a latent variable that indicates whether the kth basis belongs to the nth source. MNMF estimates the parameters t_{ik}, v_{kj}, z_{kn}, and $\boldsymbol{G}_{in}$ that minimize a cost function, which is the negative log-likelihood of the observed signal $\boldsymbol{x}_{ij}$, as

$$\mathcal{L}_{\mathrm{MNMF}} \overset{c}{=} \sum_{i,j}\Big(\boldsymbol{x}_{ij}^{\mathsf{H}}\hat{\boldsymbol{X}}_{ij}^{-1}\boldsymbol{x}_{ij} + \log\det\hat{\boldsymbol{X}}_{ij}\Big), \tag{7}$$

where $\hat{\boldsymbol{X}}_{ij} = \sum_n \sigma_{ijn}\boldsymbol{G}_{in} = \sum_{k,n} t_{ik}v_{kj}z_{kn}\boldsymbol{G}_{in}$ and $\cdot^{\mathsf{H}}$ denotes the Hermitian transpose and $\overset{c}{=}$ denotes equality up to a constant. All parameters can be optimized by using the auxiliary function technique and solving the Riccati equation (details of these update rules are described in [10]). After the updates, we can estimate the separated signal $\hat{\boldsymbol{s}}_{ijn}$ using the multichannel Wiener filter.

$$\hat{\boldsymbol{s}}_{ijn} = \Big(\sum_k t_{ik}v_{kj}z_{kn}\boldsymbol{G}_{in}\Big)\hat{\boldsymbol{X}}_{ij}^{-1}\boldsymbol{x}_{ij} \tag{8}$$

MNMF assumes that $\boldsymbol{G}_{in}$ is a full-rank matrix [12], which increases versatility for various types of spatial conditions. However, this optimization is sensitive to parameter initialization and requires a large amount of computation because the SCM $\boldsymbol{G}_{in}$ does not have any restrictions other than the constraint that it is a positive semidefinite Hermitian matrix and has a large number of parameters.

C. FastMNMF [13], [20]

To reduce the computational complexity of the update algorithm of MNMF, FastMNMF additionally assumes that the SCMs $\boldsymbol{G}_{i1}, \ldots, \boldsymbol{G}_{im}, \ldots, \boldsymbol{G}_{iN}$ are jointly diagonalizable by $\boldsymbol{Q}_i = (\boldsymbol{q}_{i1}, \ldots, \boldsymbol{q}_{im}, \ldots, \boldsymbol{q}_{iM})^{\mathsf{H}}$, which does not depend on the source index n, as

$$\begin{cases} \boldsymbol{Q}_i\boldsymbol{G}_{i1}\boldsymbol{Q}_i^{\mathsf{H}} = \boldsymbol{\mathcal{G}}_{i1} \\ \qquad\vdots \\ \boldsymbol{Q}_i\boldsymbol{G}_{in}\boldsymbol{Q}_i^{\mathsf{H}} = \boldsymbol{\mathcal{G}}_{in} \\ \qquad\vdots \\ \boldsymbol{Q}_i\boldsymbol{G}_{iN}\boldsymbol{Q}_i^{\mathsf{H}} = \boldsymbol{\mathcal{G}}_{iN}, \end{cases} \tag{9}$$

where $\boldsymbol{\mathcal{G}}_{in}$ is a diagonal matrix. From (7) and (9), the negative log-likelihood of the observed signal is given by

$$\mathcal{L}_{\mathrm{F}} \overset{c}{=} \sum_{i,j,m}\left[\frac{|\boldsymbol{q}_{im}^{\mathsf{H}}\boldsymbol{x}_{ij}|^2}{\sum_{n,k} t_{ik}v_{kj}z_{kn}g_{inm}} + \log\sum_{n,k} t_{ik}v_{kj}z_{kn}g_{inm}\right] - 2J\sum_i \log|\det\boldsymbol{Q}_i|, \tag{10}$$

where g_{inm} is the mth diagonal element of $\boldsymbol{\mathcal{G}}_{in}$. The joint-diagonalization matrix $\boldsymbol{Q}_i$ in (10) can be optimized by iterative projection (IP) [21], and the remaining parameters are updated by using the auxiliary function technique. IP is one of the coordinate-decent algorithm, which can minimize the cost function that has the sum of the quadratic form of $\boldsymbol{q}_{im}$ and the log-determinant of $\boldsymbol{Q}_i$. Similarly to the auxiliary function technique, IP guarantees the monotonic nonincrease in the cost function and provides efficient optimization for matrix variables. After the updates, we can estimate the separated signal using the multichannel Wiener filter as follows:

$$\hat{\boldsymbol{s}}_{ijn} = \boldsymbol{Q}_i^{-1}\mathrm{diag}\Big(h_{ijn1},\ldots,h_{ijnM}\Big)\boldsymbol{Q}_i\boldsymbol{x}_{ij}, \tag{11}$$

$$h_{ijnm} = \frac{\sum_k t_{ik}v_{kj}z_{kn}g_{inm}}{\sum_{k,n'} t_{ik}v_{kj}z_{kn'}g_{in'm}}. \tag{12}$$

This change to the model provides a different separation performance; it is reported that FastMNMF is almost the same as or slightly better than MNMF [14], [20].

III. Proposed Method

A. FastMNMF based on Student's t-distribution

In this paper, we propose the generalization of the model of FastMNMF to the multivariate complex Student's t-distribution (t-FastMNMF). The Student's t-distribution is a heavy-tailed distribution and its shape can be changed parametrically. This model generalization increases versatility for various types of sources, and we can expect to further improve the source-separation accuracy.

The observed signal $\boldsymbol{x}_{ij}$ is assumed to be generated by the multivariate complex Student's t-distribution as follows:

$$\boldsymbol{x}_{ij} \sim \mathcal{T}_\nu(\boldsymbol{0}_M, \sum_n \sigma_{ijn}\boldsymbol{G}_{in}), \tag{13}$$

where $\mathcal{T}_\nu(\boldsymbol{\mu},\boldsymbol{\Sigma})$ is the multivariate complex Student's t-distribution with mean $\boldsymbol{\mu}$, scale matrix $\boldsymbol{\Sigma}$, and the degree-of-freedom parameter $\nu > 0$. Here, we substitute the model parameter $\hat{\boldsymbol{X}}_{ij}$ for $\boldsymbol{\Sigma}$. The probability density function of the multivariate complex Student's t-distribution is given by

$$p(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu) = \frac{2^M\Gamma(\frac{\nu+2M}{2})}{(\nu\pi)^M\Gamma(\frac{\nu}{2})\det\boldsymbol{\Sigma}}\left(1+\frac{2}{\nu}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathsf{H}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)^{-\frac{\nu+2M}{2}}. \tag{14}$$

When ν is set to $\nu = 1$ and $\nu \to \infty$, the Student's t-distribution corresponds to the Cauchy distribution and the Gaussian distribution, respectively. The negative log-likelihood function based on (13) can be obtained as

$$\mathcal{L}_{\mathrm{St}} \stackrel{c}{=} \sum_{i,j}\Bigg[\frac{2M+\nu}{2}\log\Big(1+\frac{2}{\nu}\sum_m\frac{|\boldsymbol{q}_{im}^{\mathsf{H}}\boldsymbol{x}_{ij}|^2}{\sum_{k,n}t_{ik}v_{kj}z_{kn}g_{inm}}\Big) + \sum_m\log\sum_{k,n}t_{ik}v_{kj}z_{kn}g_{inm}\Bigg] - 2J\sum_i\log|\det\boldsymbol{Q}_i|. \tag{15}$$

B. Derivation of update rules

First, we derive the update rule of the joint-diagonalization matrix $\boldsymbol{Q}_i$ by modifying (15) to an IP-applicable form. Note that we cannot directly use IP to optimize $\boldsymbol{Q}_i$ for t-FastMNMF because the cost function (15) does not have the sum of the quadratic form of $\boldsymbol{q}_{im}$. To optimize the cost function (15) by IP, we apply a tangent line inequality to the first logarithm term in (15). The tangent line inequality for the logarithm function can be represented as

$$\log y \le \frac{1}{\tilde{\xi}}(y-\tilde{\xi}) + \log\tilde{\xi}, \tag{16}$$

where $y > 0$ is the original variable and $\tilde{\xi} > 0$ is an auxiliary variable. The equality of (16) holds if and only if $\tilde{\xi} = y$. Hence, by applying (16) to the first term in (15), we can design the following auxiliary function.

$$\begin{aligned}\mathcal{L}_{\mathrm{St}} \le& \sum_{i,j}\Bigg[\frac{2M+\nu}{2}\Bigg(\frac{1}{\tilde{\xi}_{ij}}\Big(1+\frac{2}{\nu}\sum_m\frac{|\boldsymbol{q}_{im}^{\mathsf{H}}\boldsymbol{x}_{ij}|^2}{\sum_{k,n}t_{ik}v_{kj}z_{kn}g_{inm}} - \tilde{\xi}_{ij}\Big) + \log\tilde{\xi}_{ij}\Bigg) + \sum_m\log\sum_{k,n}t_{ik}v_{kj}z_{kn}g_{inm}\Bigg] \\ &- 2J\sum_i\log|\det\boldsymbol{Q}_i| \qquad (17)\\ :=& \mathcal{L}_{\mathrm{St}}^+, \qquad (18)\end{aligned}$$

where $\tilde{\xi}_{ij} > 0$ is the auxiliary variable and the equality of (17) holds if and only if the $\tilde{\xi}_{ij}$ is set as

$$\tilde{\xi}_{ij} = 1 + \frac{2}{\nu}\sum_m\frac{|\boldsymbol{q}_{im}^{\mathsf{H}}\boldsymbol{x}_{ij}|^2}{\sum_{k,n}t_{ik}v_{kj}z_{kn}g_{inm}}. \tag{19}$$

Since the auxiliary function (17) is the sum of the quadratic form of $\boldsymbol{q}_{im}$ and the negative log-determinant of $\boldsymbol{Q}_i$, we can apply IP to (17) and obtain the update rule of $\boldsymbol{Q}_i$ as

$$\boldsymbol{q}_{im} \leftarrow (\boldsymbol{Q}_i\boldsymbol{F}_{im})^{-1}\boldsymbol{e}_m, \tag{20}$$

$$\boldsymbol{q}_{im} \leftarrow \frac{\boldsymbol{q}_{im}}{\sqrt{\boldsymbol{q}_{im}^{\mathsf{H}}\boldsymbol{F}_{im}\boldsymbol{q}_{im}}}, \tag{21}$$

where $\boldsymbol{e}_m$ denotes the one-hot vector in which the mth element equals unity and the others are zero, and

$$\boldsymbol{F}_{im} = \frac{1}{J}\sum_j\frac{\alpha_{ij}}{\chi_{ijm}}\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^{\mathsf{H}}, \tag{22}$$

$$\chi_{ijm} = \sum_{k,n}t_{ik}v_{kj}z_{kn}g_{inm}, \tag{23}$$

$$\alpha_{ij} = \frac{2M+\nu}{\nu + 2\sum_m\frac{|\boldsymbol{q}_{im}^{\mathsf{H}}\boldsymbol{x}_{ij}|^2}{\chi_{ijm}}}. \tag{24}$$

Next, we derive the update rules of t_{ik}, v_{kj}, z_{kn}, and g_{inm}. To obtain the auxiliary function for these parameters, we can use a tangent line inequality and Jensen's inequality in (17). Jensen's inequality for the reciprocal function can be represented as

$$\Big(\sum_\tau y_\tau\Big)^{-1} = \Big(\sum_\tau\tilde{\eta}_\tau\frac{y_\tau}{\tilde{\eta}_\tau}\Big)^{-1} \le \sum_\tau\tilde{\eta}_\tau\Big(\frac{y_\tau}{\tilde{\eta}_\tau}\Big)^{-1} = \sum_\tau\frac{\tilde{\eta}_\tau^2}{y_\tau}, \tag{25}$$

where $y_\tau > 0$ is the original variable and $\tilde{\eta}_\tau > 0$ is an auxiliary variable that satisfies $\sum_\tau\tilde{\eta}_\tau = 1$, and the equality of (25) holds if and only if $\tilde{\eta}_\tau = y_\tau / \sum_{\tau'} y_{\tau'}$. Then, we can design the following auxiliary function.

$$\begin{aligned}
\mathcal{L}_{\mathrm{St}}^{+} &\overset{c}{=} \sum_{i,j}\Bigg[\frac{2M+\nu}{2}\Bigg(\frac{1}{\tilde{\xi}_{ij}}\Big(1+\frac{2}{\nu}\sum_m \frac{|\boldsymbol{q}_{im}^{\mathsf{H}}\boldsymbol{x}_{ij}|^2}{\sum_{k,n} t_{ik}v_{kj}z_{kn}g_{inm}}\Big)\Bigg) \\
&\quad + \sum_m \log\sum_{k,n} t_{ik}v_{kj}z_{kn}g_{inm}\Bigg] - 2J\sum_i \log|\det \boldsymbol{Q}_i| \\
&\le \sum_{i,j}\Bigg[\frac{2M+\nu}{2}\Bigg(\frac{1}{\tilde{\xi}_{ij}}\Big(1+\frac{2}{\nu}\sum_m\sum_{k,n} \frac{|\boldsymbol{q}_{im}^{\mathsf{H}}\boldsymbol{x}_{ij}|^2\tilde{\eta}_{ijknm}^2}{t_{ik}v_{kj}z_{kn}g_{inm}}\Big)\Bigg) \\
&\quad + \sum_m\Bigg(\frac{1}{\tilde{\zeta}_{ijm}}\Big(\sum_{k,n} t_{ik}v_{kj}z_{kn}g_{inm} - \tilde{\zeta}_{ijm}\Big) + \log\tilde{\zeta}_{ijm}\Bigg)\Bigg] \\
&\quad - 2J\sum_i \log|\det \boldsymbol{Q}_i| \qquad (26) \\
&:= \mathcal{L}_{\mathrm{St}}^{++}, \qquad (27)
\end{aligned}$$

where $\tilde{\eta}_{ijknm} > 0$ and $\tilde{\zeta}_{ijm} > 0$ are auxiliary variables and $\tilde{\eta}_{ijknm}$ satisfies $\sum_{k,n}\tilde{\eta}_{ijknm} = 1$. The equality of (26) holds only when

$$\tilde{\eta}_{ijknm} = \frac{t_{ik}v_{kj}z_{kn}g_{inm}}{\sum_{k',n'} t_{ik'}v_{k'j}z_{k'n'}g_{in'm}}, \qquad (28)$$

$$\tilde{\zeta}_{ijm} = \sum_{k,n} t_{ik}v_{kj}z_{kn}g_{inm}. \qquad (29)$$

From $\partial\mathcal{L}_{\mathrm{St}}^{++}/\partial t_{ik} = 0$, we obtain the update rule of t_{ik} as

$$t_{ik} \leftarrow t_{ik}\sqrt{\frac{\sum_{j,n} v_{kj}z_{kn}\alpha_{ij}\beta_{ijn}}{\sum_{j,n} v_{kj}z_{kn}\gamma_{ijn}}}, \qquad (30)$$

where

$$\beta_{ijn} = \sum_m \frac{|\boldsymbol{q}_{im}^{\mathsf{H}}\boldsymbol{x}_{ij}|^2 g_{inm}}{\chi_{ijm}^2}, \qquad (31)$$

$$\gamma_{ijn} = \sum_m \frac{g_{inm}}{\chi_{ijm}}. \qquad (32)$$

Similarly to t_{ik}, we obtain the update rules of v_{kj}, z_{kn}, and g_{inm} as

$$v_{kj} \leftarrow v_{kj}\sqrt{\frac{\sum_{i,n} t_{ik}z_{kn}\alpha_{ij}\beta_{ijn}}{\sum_{i,n} t_{ik}z_{kn}\gamma_{ijn}}}, \qquad (33)$$

$$z_{kn} \leftarrow z_{kn}\sqrt{\frac{\sum_{i,j} t_{ik}v_{kj}\alpha_{ij}\beta_{ijn}}{\sum_{i,j} t_{ik}v_{kj}\gamma_{ijn}}}, \qquad (34)$$

$$g_{inm} \leftarrow g_{inm}\sqrt{\frac{\sum_{j,k}\frac{|\boldsymbol{q}_{im}^{\mathsf{H}}\boldsymbol{x}_{ij}|^2}{\chi_{ijm}^2} t_{ik}v_{kj}z_{kn}\alpha_{ij}}{\sum_{j,k}\frac{1}{\chi_{ijm}} t_{ik}v_{kj}z_{kn}}}. \qquad (35)$$

Note that these update rules have already been substituted under the equality conditions (19), (28), and (29) and rearranged. We can confirm that the update rules of *t*-FastMNMF (20), (21), (30), and (33)–(35) are equal to those of original Gaussian FastMNMF (described by [20]) when $\nu \to \infty$. Finally, we estimate the separated signal using the multichannel Wiener filter, similarly to (11). The algorithm for *t*-FastMNMF is summarized in Algorithm 1.

Algorithm 1: Algorithm for t-FastMNMF

Initialize $\boldsymbol{Q}_i$ and $\boldsymbol{\mathcal{G}}_{in}$ with identity matrix and t_{ik}, v_{kj}, and z_{kn} with positive random values for all i, j, k, and n;
Calculate χ_{ijm}, α_{ij}, β_{ijn}, γ_{ijn} by (23), (24), (31), and (32) for all i, j, n, and m;
repeat
    Calculate t_{ik} by (30) for all i and k;
    Calculate χ_{ijm}, α_{ij}, β_{ijn}, γ_{ijn} by (23), (24), (31), and (32) for all i, j, n, and m;
    Calculate v_{kj} by (33) for all k and j;
    Calculate χ_{ijm}, α_{ij}, β_{ijn}, γ_{ijn} by (23), (24), (31), and (32) for all i, j, n, and m;
    Calculate z_{kn} by (34) for all k and n;
    Calculate χ_{ijm}, α_{ij}, β_{ijn}, γ_{ijn} by (23), (24), (31), and (32) for all i, j, n, and m;
    Calculate g_{inm} by (35) for all i, n, and m;
    Calculate χ_{ijm}, α_{ij}, β_{ijn}, γ_{ijn} by (23), (24), (31), and (32) for all i, j, n, and m;
    Calculate $\boldsymbol{q}_{im}$ by (20) and (21) for all i and m;
until *converge*;
Calculate $\hat{\boldsymbol{s}}_{ijn}$ by (11) for all i, j, and n;

Fig. 1. Scores of each part.

IV. Experiments

A. Experimental conditions

To confirm the validity of the proposed *t*-FastMNMF, we conducted a BSS experiment with simulated mixtures. We compared the following seven methods: IVA [21], ILRMA [7], *t*-ILRMA [16], MNMF [10], *t*-MNMF [15], FastMNMF [13], [20], and proposed *t*-FastMNMF. All the NMF variables were initialized by nonnegative random values. For IVA, ILRMA, and *t*-ILRMA, the demixing matrices were initialized by the identity matrices. For MNMF and *t*-MNMF, SCMs were initialized by the identity matrices. For FastMNMF and proposed *t*-FastMNMF, the joint-diagonalized SCMs and joint-diagonalization matrices were initialized by the identity matrices.

The dry sources of four melody parts depicted in Fig. 1 were obtained from [22], [23]. Eight combinations of instruments with two different melody parts were selected, as shown in Table I. To simulate reverberant mixing, two-channel mixed

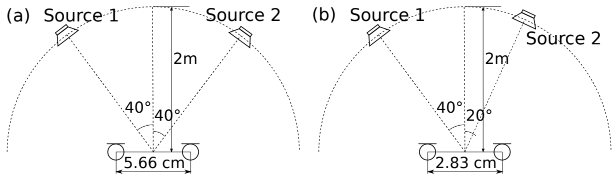

Fig. 2. Spatial arrangements of impulse responses used in experiments.

TABLE I
DRY SOURCES USED IN EXPERIMENTS

	Part name	Source (1/2)
Music 1	Midrange/Melody 2	Flute/Piano
Music 2	Melody 1/Melody 2	Flute/Oboe
Music 3	Melody 2/Midrange	Harpsichord/Violin
Music 4	Melody 2/Bass	Cello/Violin
Music 5	Melody 1/Bass	Cello/Oboe
Music 6	Melody 2/Melody 1	Trumpet/Violin
Music 7	Bass/Melody 2	Flute/Bassoon
Music 8	Bass/Melody 1	Trumpet/Bassoon

signals were produced by convoluting the impulse response E2A ($T_{60} = 300$ ms) in the RWCP database [24]. Fig. 2 shows the recording conditions of E2A used in our experiments. In these mixtures, the input signal-to-noise ratio was 0 dB. The other experimental conditions are shown in Table II.

B. Results

We used the source-to-distortion ratio (SDR) improvement [25] to evaluate the total separation performance. Fig. 3 shows the average SDR improvements over the recording conditions, the source pairs, and 10-trial initialization. The result for t-ILRMA using $\nu = 10^5$ is shown, which is the best parameter setting. Since the Student's t-distribution corresponds to the Gaussian distribution when the degree-of-freedom parameter in the Student's t-distribution $\nu = \infty$, the first and second rightmost bars correspond to FastMNMF and MNMF, respectively. IVA, ILRMA, and t-ILRMA assume the rank-1 spatial covariance matrix, and since this assumption does not hold under the situations where reverberation is longer than the STFT window length, their SDR improvement values are considered to be low. By comparing proposed t-FastMNMF and t-MNMF, we can see that proposed t-FastMNMF shows higher performance overall. It can be confirmed that proposed t-FastMNMF outperforms conventional FastMNMF for the highest degree-of-freedom parameter $\nu = 8$. Therefore, it can be said that the proposed method is superior to the conventional methods.

V. Conclusions

In this paper, we propose the generalization of the generative model of FastMNMF from the multivariate complex Gaussian distribution to the multivariate complex Student's t-distribution. We derive a new parameter update rule using the auxiliary-function-based method that guarantees a monotonic nonincrease in the cost function, and we confirmed the validity of the proposed method from BSS experiments using music sources.

TABLE II
EXPERIMENTAL CONDITIONS

Sampling frequency	16 kHz
STFT	64-ms Hamming window with 16-ms shift
Number of bases in low-rank model	20
Number of iterations	300
Degree-of-freedom parameter of Student's t-distribution for t-MNMF and t-FastMNMF	$\nu = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100$
Number of trials	10

Acknowledgment

This work was partly supported by SECOM Science and Technology Foundation and JSPS KAKENHI Grant Numbers JP19H01116, JP19H04131, and JP19K20306, and JSPS-CAS Joint Research Program, Grant number JPJSBP120197203.

References

[1] H. Sawada, N. Ono, H. Kameoka, D. Kitamura, and H. Saruwatari, "A review of blind source separation methods: two converging routes to ILRMA originating from ICA and NMF," *APSIPA Trans. on Signal and Information Processing*, vol. 8, no. e12, pp. 1–14, 2019.

[2] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol. 22, no. 1–3, pp. 21–34, 1998.

[3] H. Saruwatari, T. Kawamura, T. Nishikawa, A. Lee, and K. Shikano, "Blind source separation based on a fast-convergence algorithm combining ICA and beamforming," *IEEE Trans. on ASLP*, vol. 14, no. 2, pp. 666–678, 2006.

[4] A. Hiroe, "Solution of permutation problem in frequency domain ICA, using multivariate probability density functions," in *Proc. ICA*, 2006, pp. 601–608.

[5] T. Kim, T. Eltoft, and T.-W. Lee, "Independent vector analysis: An extension of ICA to multivariate components," in *Proc. ICA*, 2006, pp. 165–172.

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

[7] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, "Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization," *IEEE/ACM Trans. on ASLP*, vol. 24, no. 9, pp. 1626–1641, 2016.

[8] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, "Determined blind source separation with independent low-rank matrix analysis," in chapter of *Audio Source Separation, (S. Makino Ed.)*, pp. 125–155. Springer, Cham, 2018.

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

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

[11] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.

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

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

[14] Y. Kubo, N. Takamune, D. Kitamura, and H. Saruwatari, "Efficient full-rank spatial covariance estimation using independent low-rank matrix analysis for blind source separation," in *Proc. EUSIPCO*, 5 pages, 2019.

[15] K. Kitamura, Y. Bando, K. Itoyama, and K. Yoshii, "Student's t multichannel nonnegative matrix factorization for blind source separation," in *Proc. IWAENC*, 5 pages, 2016.

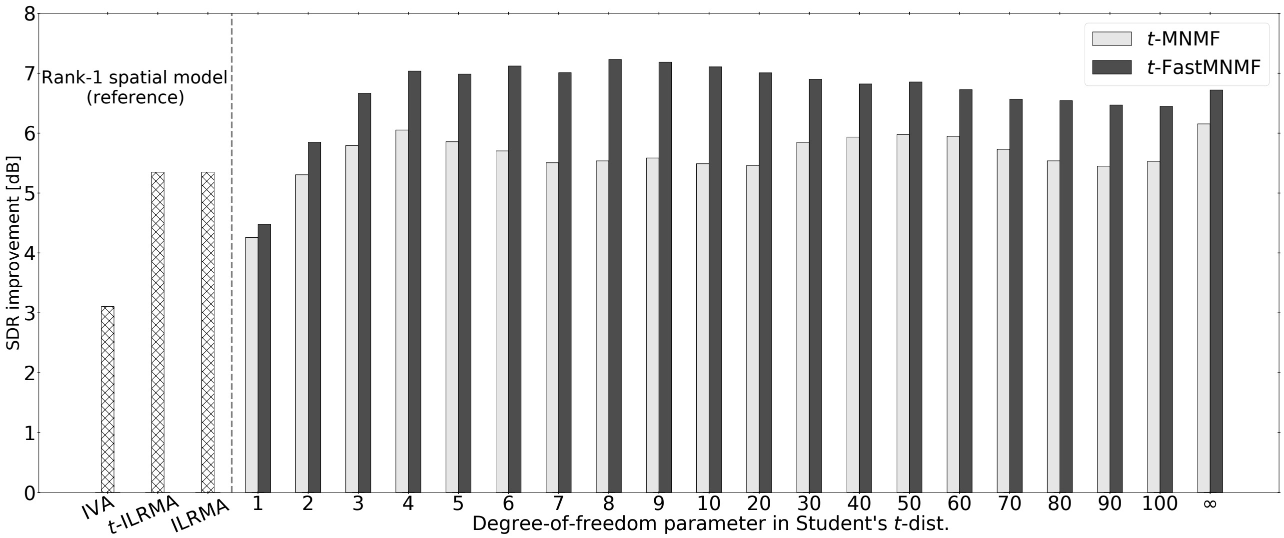

Fig. 3. Average SDR improvements for each method. The result for t-ILRMA using $\nu = 10^5$ is shown, which is the best parameter setting. For $\nu = 8$, the proposed method achieves the best SDR improvement and outperforms conventional methods.

[16] S. Mogami, D. Kitamura, Y. Mitsui, N. Takamune, H. Saruwatari, and N. Ono, "Independent low-rank matrix analysis based on complex Student's t-distribution for blind audio source separation," in *Proc. MSLP,* 6 pages, 2017.

[17] D. Kitamura, S. Mogami, Y. Mitsui, N. Takamune, H. Saruwatari, N. Ono, Y. Takahashi, and K. Kondo, "Generalized independent low-rank matrix analysis using heavy-tailed distributions for blind source separation," *EURASIP Journal on Advances in Signal Processing*, vol. 2018, no. 28, 25 pages, 2018.

[18] N. Makishima, S. Mogami, N. Takamune, D. Kitamura, H. Sumino, S. Takamichi, H. Saruwatari, and N. Ono, "Independent deeply learned matrix analysis for determined audio source separation," *IEEE/ACM Trans. on ASLP*, vol. 27, no. 10, pp. 1601–1615, 2019.

[19] D. R. Hunter and K. Lange, "Quantile regression via an MM algorithm," *Journal of Computational and Graphical Statistics*, vol. 9, no. 1, pp. 60–77, 2000.

[20] K. Sekiguchi, A. A. Nugraha, Y. Bando, and K. Yoshii, "Fast multichannel source separation based on jointly diagonalizable spatial covariance matrices," in *Proc. EUSIPCO,* 5 pages, 2019.

[21] N. Ono, "Stable and fast update rules for independent vector analysis based on auxiliary function technique," in *Proc. WASPAA*, 2011, pp. 189–192.

[22] D. Kitamura, H. Saruwatari, H. Kameoka, Y. Takahashi, K. Kondo, and S. Nakamura, "Multichannel signal separation combining directional clustering and nonnegative matrix factorization with spectrogram restoration," *IEEE/ACM Trans. on ASLP*, vol. 23, no. 4, pp. 654–669, 2015.

[23] D. Kitamura, "Open dataset: songKitamura," http://d-kitamura.net/dataset.html, Accessed 1 Feb. 2020.

[24] S. Nakamura, K. Hiyane, F. Asano, T. Nishiura, and T. Yamada, "Acoustical sound database in real environments for sound scene understanding and hands-free speech recognition," in *Proc. LREC*, 2000, pp. 965–968.

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