STUDENT'S T MULTICHANNEL NONNEGATIVE MATRIX FACTORIZATION FOR BLIND SOURCE SEPARATION

Koichi Kitamura, Yoshiaki Bando, Katsutoshi Itoyama, Kazuyoshi Yoshii

Graduate School of Informatics, Kyoto University, Japan

ABSTRACT

This paper presents a robust generalization of multichannel nonnegative matrix factorization (MNMF) for blind source separation of mixture audio signals recorded by a microphone array. In conventional MNMF, the complex spectra of observed mixture signals are assumed to be complex Gaussian distributed and are decomposed into the product of the power spectra, temporal activations, and spatial correlation matrices of individual sources in such a way that the complex Gaussian likelihood is maximized. Since the mixture spectra usually include outliers, we propose MNMF based on the complex Student's t likelihood, called t-MNMF, including the original MNMF as a special case. The parameters of t-MNMF can be iteratively optimized with an efficient multiplicative updating algorithm. Experiments showed that t-MNMF with a certain range of degrees of freedom tends to be insensitive to parameter initialization and outperform conventional MNMF.

Index Terms— Blind source separation, nonnegative matrix factorization, Student's t distribution

1. INTRODUCTION

Blind source separation (BSS) forms the basis of computational auditory scene analysis. It aims to decompose observed single or multi-channel mixture audio signals into individual source signals without using any prior knowledge on sound sources and microphone arrays. Since speech signals recorded in a real environment are contaminated by various kinds of sounds coming from different directions, speech enhancement is required for improved speech recognition.

Nonnegative matrix factorization (NMF) has widely been used for single-channel BSS based on the acoustic characteristics (low-rankness) of mixture spectrograms [1–3]. In NMF, a nonnegative matrix (a set of nonnegative vectors) is approximated as the product of two nonnegative matrices (a set of basis vectors and a set of activation vectors). If NMF is applied to the "power" spectrogram of a piano piece, for example, the basis vectors are expected to correspond to the average power spectra of different pitches (sources) used in the piece. The mixture spectrogram is then decomposed into the sum of source spectrograms via Wiener filtering according to the source proportions determined by the basis and activation vectors.

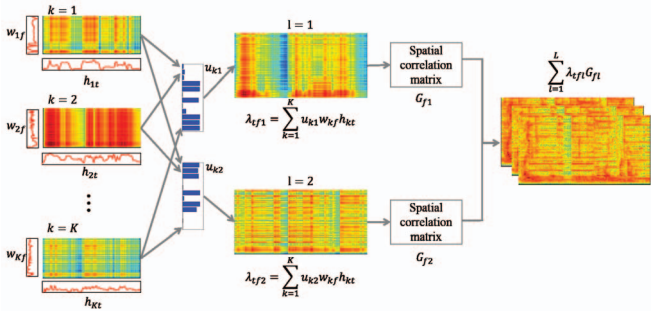

Fig. 1. The generative model of MNMF.

A multichannel extension of NMF (MNMF) dealing with both the acoustic and spatial characteristics of sound sources has recently been proposed [4], where the phase differences (spatial correlations) of multichannel "complex" spectra play a key role. A generative model underlying MNMF is shown in Fig 1. The power spectrogram of each source signal is generated as the product of basis vectors and activations vectors as in single-channel NMF. The multichannel complex spectra of observed mixture signals are then generated as the sum of the multichannel complex spectra of source signals obtained by applying spatial correlation matrices to the power spectrograms of the source signals. Given the multichannel complex spectra of observed mixture signals, MNMF tries to jointly estimate the basis vectors, activation vectors, spatial correlation matrices of individual sources.

A major problem of MNMF is that it tends to easily get stuck in bad local optima. One possible solution is to reduce the complexity of MNMF. In rank-1 MNMF [5], for example, each spatial correlation matrix is restricted to a rank-1 matrix because it must be equal to the product of a steering vector and its conjugate transpose in an idealized environment. Another solution is to formulate a probabilistic model that precisely represents the acoustic characteristics of target audio signals. In original MNMF, the complex spectra of source signals are assumed to be complex Gaussian distributed. If those spectra are propagated and superimposed in an instantaneous mixing process in the frequency domain, the complex spectra of observed mixture signals are complex Gaussian distributed due to the reproductive property of the Gaussian distribution. MNMF can thus be interpreted as maximum likelihood estimation of a probabilistic model with the complex Gaussian likelihood. In fact, however, source spectra tend to be heavy-

978-1-5090-2007-2/16/$31.00

tailed distributed. Although the central limit theorem suggest that mixture spectra are asymptotically Gaussian distributed regardless of source distributions, the Gaussian distribution is too sensitive to outliers and the initial parameters.

In this paper we propose a generalization of MNMF called t-MNMF that assumes the complex spectra of observed mixture signals to follow complex Student's t distributions with a degree-of-freedom parameter ν adjustable to the tail heaviness of the spectra. t-MNMF reduces to MNMF when $\nu \to \infty$ because the t distribution converges to the Gaussian distribution. In single-channel BSS, a similar generalization called t-NMF has recently been proposed [6]. t-NMF includes NMF based on the Itakura-Saito divergence (complex Gaussian likelihood) [7] as its special case with $\nu = \infty$ and includes a new variant of NMF based on the complex Cauchy likelihood [8] as its special case with $\nu = 1$. Our contribution is to make MNMF stably work well regardless of parameter initialization in the context of multichannel BSS.

2. RELATED WORK

In multichannel source separation, linear-filtering-based methods are widely used to decompose mixture signals into individual source signals [9–11]. Independent component analysis (ICA) [12] is one of most widely used methods using a linear filter, but permutation alignment is needed because source separation is performed independently for each frequency. Independent vector analysis (IVA) [13, 14], which is an extension of ICA, solves the permutation problem by dealing with all frequency components in a vectorial manner. These methods can work only when the number of sources is equal to or less than the number of microphones.

MNMF is one promising blind source separation approach considering both the spatial characteristics and the structure of each source spectrogram. It works even when the number of sources is greater than the number of microphones. The optimization algorithm of MNMF can be derived by maximizing the likelihood for the mixture spectrogram. The sound source separation performance of MNMF is strongly dependent on initial values of parameters because MNMF has a lot of parameters that should be learned jointly.

3. MULTICHANNEL NMF

We review MNMF [4] designed for frequency-domain BSS. Suppose that L sources are observed with M microphones. Each time-frequency (TF) bin in the complex spectrograms of observed and source signals is given by

$$\boldsymbol{x}_{tf} = [x_{tf1}, \cdots, x_{tfM}]^T \in \mathbb{C}^M, \tag{1}$$

$$\boldsymbol{y}_{tf} = [y_{tf1}, \cdots, y_{tfl}]^T \in \mathbb{C}^l. \tag{2}$$

Assuming an instantaneous mixing process in the frequency domain, the observation $\boldsymbol{x}_{tf}$ is given by the sum of the spectra of *transferred* source signals, $\{\boldsymbol{x}_{tfl}\}_{l=1}^L$, as follows:

$$\boldsymbol{x}_{tf} = \textstyle\sum_{l=1}^L \boldsymbol{x}_{tfl} = \sum_{l=1}^L \boldsymbol{a}_{fl} y_{tfl}, \tag{3}$$

where $\boldsymbol{a}_{fl}$ is a steering vector of source l at frequency f. The complex spectrum y_{tfl} of source l is assumed to be complex Gaussian distributed as follows:

$$y_{tfl} \mid \lambda_{tfl} \sim \mathcal{N}_{\mathbb{C}}(y_{tfl} \mid 0, \lambda_{tfl}), \tag{4}$$

where λ_{tfl} is the power spectrum density of source l. Using Eq. (3) and Eq. (4), the complex spectrum $\boldsymbol{x}_{tf}$ is found to follow a complex Gaussian distribution as follows:

$$\boldsymbol{x}_{tf} \mid \boldsymbol{\lambda}, \boldsymbol{G} \sim \mathcal{N}_{\mathbb{C}}\left(\boldsymbol{x}_{tf} \;\middle|\; \boldsymbol{0}, \textstyle\sum_{l=1}^L \lambda_{tfl} \boldsymbol{G}_{fl}\right), \tag{5}$$

where $\boldsymbol{G}_{fl} = \boldsymbol{a}_{fl}\boldsymbol{a}_{fl}^H$ is a spatial correlation matrix for source l at frequency f. Since $\boldsymbol{G}_{fl}$ is a rank-1 matrix in an idealized environment, rank-1 MNMF [5], which restricts $\boldsymbol{G}_{fl}$ to a rank-1 matrix, is known to work well in practice.

A key idea of MNMF is to decompose the power spectrum density λ_{tfl} of source l by using a low-rank approximation technique (*e.g.*, CP decomposition [15]) as follows:

$$\lambda_{tfl} = \textstyle\sum_{k=1}^K u_{kl} w_{kf} h_{kt}, \tag{6}$$

where $\{w_{kf}\}_{f=1}^F$ is the power spectrum of basis k, $\{h_{kt}\}_{t=1}^T$ is the activation vector of basis k, and u_{kl} is a contribution of basis k to source l such that $\sum_{l=1}^L u_{kl} = 1$.

Given observed data $\boldsymbol{X}$, the unknown parameters $\boldsymbol{W}$, $\boldsymbol{H}$, $\boldsymbol{U}$, and $\boldsymbol{G}$ are jointly estimated by using a multiplicative updating algorithm such that the likelihood given by Eq. (5) is maximized. The latent source signals are then estimated with multichannel Wiener filtering as follows:

$$\mathbb{E}[\boldsymbol{x}_{tfl} | \boldsymbol{\lambda}, \boldsymbol{G}] = \lambda_{tfl}\boldsymbol{G}_{fl}\left(\textstyle\sum_{l'=1}^L \lambda_{tfl'}\boldsymbol{G}_{fl'}\right)^{-1}\boldsymbol{x}_{tf}. \tag{7}$$

4. PROPOSED METHOD

This section explains t-MNMF based on the Student's t likelihood. Since the complex Gaussian distribution is sensitive to outliers in observed data, we replace the complex Gaussian distribution given by Eq. (5) with a heavy-tailed t distribution. The parameters to be estimated are the same as those of original MNMF [4] and a generalized version of the original multiplicative updating (MU) algorithm can be derived.

4.1. Model formulation

The log-likelihood of t-MNMF is based on the multivariate complex t distribution, which is given by

$$\boldsymbol{x}_{tf} \mid \boldsymbol{\lambda}, \boldsymbol{G} \sim \mathcal{T}_{\mathbb{C}}\left(\boldsymbol{x}_{tf} \;\middle|\; \boldsymbol{0}, \textstyle\sum_{l=1}^L \lambda_{tfl} \boldsymbol{G}_{fl}\right), \tag{8}$$

where the multivariate complex t distribution is defined as

$$\mathcal{T}_\nu(\boldsymbol{x}|\boldsymbol{0}, \boldsymbol{\Sigma}) = \frac{\Gamma(\frac{2d+\nu}{2})}{\Gamma(\frac{\nu}{2})}\frac{2^d}{(\upsilon\pi)^d}\frac{1}{|\boldsymbol{\Sigma}|}\left(1 + \frac{2}{\nu}\boldsymbol{x}^H\boldsymbol{\Sigma}^{-1}\boldsymbol{x}\right)^{-\frac{2d+\nu}{2}}.$$

This distribution converges to the multivariate complex Gaussian distribution as $\nu \to \infty$ and reduces to the multivariate complex Cauchy distribution as $\nu = 1$. For readability, we let

$\hat{\boldsymbol{X}}_{tfl} = \lambda_{tfl}\boldsymbol{G}_{fl} \in \mathbb{C}^{M\times M}$, $\hat{\boldsymbol{X}}_{tf} = \sum_{l=1}^{L}\hat{\boldsymbol{X}}_{tfl} \in \mathbb{C}^{M\times M}$, and $\boldsymbol{X}_{tf} = \boldsymbol{x}\boldsymbol{x}^H \in \mathbb{C}^{M\times M}$. The log-likelihood given by Eq. (8) can be written as follows:

$$\begin{aligned}&\log p(\boldsymbol{X}_{tf}|\hat{\boldsymbol{X}}_{tf})\\ &\overset{c}{=} -\log|\hat{\boldsymbol{X}}_{tf}| - \tfrac{2M+\nu}{2}\log\left(1+\tfrac{2}{\nu}\mathrm{tr}(\hat{\boldsymbol{X}}_{tf}^{-1}\boldsymbol{X}_{tf})\right) \quad (9)\\ &\overset{\mathrm{def}}{=} \mathcal{L}(\boldsymbol{X}_{tf}|\hat{\boldsymbol{X}}_{tf}). \quad (10)\end{aligned}$$

The total log-likelihood over all TF bins, $\mathcal{L}(\boldsymbol{X}|\hat{\boldsymbol{X}})$, is given by $\mathcal{L}(\boldsymbol{X}|\hat{\boldsymbol{X}}) = \sum_{t=1}^{T}\sum_{f=1}^{F}\mathcal{L}(\boldsymbol{X}_{tf}|\hat{\boldsymbol{X}}_{tf})$.

4.2. Log-likelihood lower bound

Given observed data $\boldsymbol{X}$, our goal is to estimate the parameters $\hat{\boldsymbol{X}}$ (*i.e.*, $\boldsymbol{W}$, $\boldsymbol{H}$, $\boldsymbol{U}$, and $\boldsymbol{G}$) such that the multivariate complex t log-likelihood $\mathcal{L}(\boldsymbol{X}|\hat{\boldsymbol{X}})$ is maximized. Since it is hard to analytically calculate the optimal parameters, we use an iterative optimization method [16]. More specifically, we derive an auxiliary function $\mathcal{F}(\boldsymbol{X}|\hat{\boldsymbol{X}}, \boldsymbol{\Theta})$ that forms the lower bound of $\mathcal{L}(\boldsymbol{X}|\hat{\boldsymbol{X}})$ as follows:

$$\mathcal{L}(\boldsymbol{X}|\hat{\boldsymbol{X}}) = \max_{\boldsymbol{\Theta}}\mathcal{F}(\boldsymbol{X}|\hat{\boldsymbol{X}}, \boldsymbol{\Theta}) \quad (11)$$

and iteratively maximize $\mathcal{F}(\boldsymbol{X}|\hat{\boldsymbol{X}}, \boldsymbol{\Theta})$ with respect to $\hat{\boldsymbol{X}}$ and $\boldsymbol{\Theta}$, where $\boldsymbol{\Theta}$ is a set of newly-introduced auxiliary parameters. This procedure is proven to indirectly maximize $\mathcal{L}(\boldsymbol{X}|\hat{\boldsymbol{X}})$ with respect to $\hat{\boldsymbol{X}}$ and to converge.

In preparation for deriving the lower bound $\mathcal{F}(\boldsymbol{X}|\hat{\boldsymbol{X}}, \boldsymbol{\Theta})$, we introduce two inequalities. For a convex function $f(\boldsymbol{Z}) = -\log|\boldsymbol{Z}|$ ($\boldsymbol{Z} \succeq \boldsymbol{0} \in \mathbb{C}^{M\times M}$), the first-order Taylor expansion at arbitrary $\boldsymbol{\Omega} \succeq \boldsymbol{0}$ is given by

$$-\log|\boldsymbol{Z}| \geq -\log|\boldsymbol{\Omega}| - \mathrm{tr}(\boldsymbol{\Omega}^{-1}\boldsymbol{Z}) + M, \quad (12)$$

where the equality is attained if and only if $\boldsymbol{\Omega} = \boldsymbol{Z}$.

For a concave function $g(\boldsymbol{Z}) = -\mathrm{tr}(\boldsymbol{Z}^{-1}\boldsymbol{A})$ ($\boldsymbol{Z} \succeq \boldsymbol{0} \in \mathbb{C}^{M\times M}$) with any positive semidefinite matrix $\boldsymbol{A} \succeq \boldsymbol{0}$, the following inequality holds [17]:

$$-\mathrm{tr}\left(\left(\textstyle\sum_{k=1}^{K}\boldsymbol{Z}_k\right)^{-1}\boldsymbol{A}\right) \geq -\mathrm{tr}\left(\boldsymbol{Z}_k^{-1}\boldsymbol{\Phi}_k\boldsymbol{A}\boldsymbol{\Phi}_k^H\right), \quad (13)$$

where $\{\boldsymbol{Z}_k \succeq \boldsymbol{0}\}_{k=1}^{K}$ is a set of arbitrary positive semidefinite matrices, $\{\boldsymbol{\Phi}_k\}_{k=1}^{K}$ is a set of auxiliary matrices such that $\sum_k \boldsymbol{\Phi}_k = \boldsymbol{I}$. The equality is attained if and only if $\boldsymbol{\Phi}_k = \boldsymbol{Z}_k(\sum_{k'}\boldsymbol{Z}_{k'})^{-1}$.

Using Eq. (12) and Eq. (13), $\mathcal{F}(\boldsymbol{X}|\hat{\boldsymbol{X}}, \boldsymbol{\Omega}, \boldsymbol{\Psi}, \boldsymbol{\Phi})$ can be derived as follows:

$$\begin{aligned}\mathcal{L}(\boldsymbol{X}|\hat{\boldsymbol{X}}) &\overset{\mathrm{c}}{\geq} -\sum_{t,f,l,k}\left(\lambda_{tflk}\mathrm{tr}(\hat{\boldsymbol{\Omega}}_{tf}^{-1}\boldsymbol{G}_{fl})\right)\\ &-\sum_{t,k}\left(\psi_{tf}^{-1}\pi\right)\sum_{l,k}\frac{1}{\lambda_{tflk}}\mathrm{tr}\left(\boldsymbol{G}_{fl}^{-1}\boldsymbol{\Phi}_{tfkl}\boldsymbol{X}_{tf}\boldsymbol{\Phi}_{tfkl}^H\right)\\ &\overset{\mathrm{def}}{=} \mathcal{F}(\boldsymbol{X}|\hat{\boldsymbol{X}}, \boldsymbol{\Omega}, \boldsymbol{\Psi}, \boldsymbol{\Phi}), \quad (14)\end{aligned}$$

where $\pi = \frac{2M+\nu}{2}$, $\lambda_{tflk} = u_{kl}w_{kf}h_{kt}$, and the equality is attained, *i.e.*, $\mathcal{F}(\boldsymbol{X}|\hat{\boldsymbol{X}}, \boldsymbol{\Omega}, \boldsymbol{\Psi}, \boldsymbol{\Phi})$ is maximized, when $\hat{\boldsymbol{\Omega}}_{tf} = \hat{\boldsymbol{X}}_{tf}$, $\psi_{tf} = \frac{\nu+2\mathrm{tr}(\boldsymbol{X}_{tf}^{-1}\boldsymbol{X}_{tf})}{\nu}$, and $\boldsymbol{\Phi}_{tfkl} = \lambda_{tflk}\boldsymbol{G}_{fl}\hat{\boldsymbol{X}}_{tf}^{-1}$.

4.3. Multiplicative updating

To derive the convergence-guaranteed update rules of $\boldsymbol{W}$, $\boldsymbol{H}$, $\boldsymbol{U}$, and $\boldsymbol{G}$ that maximize $\mathcal{F}(\boldsymbol{X}|\hat{\boldsymbol{X}}, \boldsymbol{\Omega}, \boldsymbol{\Psi}, \boldsymbol{\Phi})$, we let the partial derivatives of $\mathcal{F}(\boldsymbol{X}|\hat{\boldsymbol{X}}, \boldsymbol{\Omega}, \boldsymbol{\Psi}, \boldsymbol{\Phi})$ with respect to w_{kf}, h_{kt}, u_{kl}, and $\boldsymbol{G}_{fl}$ be zero, respectively. Finally, we get the update rules of w_{kf}, h_{kt}, u_{kl}, and $\boldsymbol{G}_{fl}$ as follows:

$$w_{kf} \leftarrow w_{kf}\sqrt{\frac{\sum_t h_{kt}\frac{2M+\nu}{\nu+2\mathrm{tr}\left(\hat{\boldsymbol{X}}_{tf}^{-1}\boldsymbol{X}_{tf}\right)}\sum_l u_{kl}\alpha_{tfl}}{\sum_t h_{kt}\sum_l u_{kl}\beta_{tfl}}}, \quad (15)$$

$$h_{kt} \leftarrow h_{kt}\sqrt{\frac{\sum_f w_{kf}\frac{2M+\nu}{\nu+2\mathrm{tr}\left(\hat{\boldsymbol{X}}_{tf}^{-1}\boldsymbol{X}_{tf}\right)}\sum_l u_{kl}\alpha_{tfl}}{\sum_f w_{kf}\sum_l u_{kl}\beta_{tfl}}}, \quad (16)$$

$$u_{kl} \leftarrow u_{kl}\sqrt{\frac{\sum_{tf} w_{kf}h_{kt}\frac{2M+\nu}{\nu+2\mathrm{tr}\left(\hat{\boldsymbol{X}}_{tf}^{-1}\boldsymbol{X}_{tf}\right)}\sum_l \alpha_{tfl}}{\sum_{tf} w_{kf}h_{kt}\sum_l \beta_{tfl}}}, \quad (17)$$

$$\boldsymbol{G}_{fl} \leftarrow \boldsymbol{G}_{fl}\boldsymbol{B}_{fl}^{\frac{1}{2}}\left(\boldsymbol{B}_{fl}^{\frac{1}{2}}\boldsymbol{G}_{fl}\boldsymbol{A}_{fl}\boldsymbol{G}_{fl}\boldsymbol{B}_{fl}^{\frac{1}{2}}\right)^{-\frac{1}{2}}\boldsymbol{B}_{fl}^{\frac{1}{2}}\boldsymbol{G}_{fl}, \quad (18)$$

where $\alpha_{tfl} = \mathrm{tr}\left(\hat{\boldsymbol{X}}_{tf}^{-1}\boldsymbol{X}_{tf}\hat{\boldsymbol{X}}_{tf}^{-1}\boldsymbol{G}_{fl}\right)$, $\beta_{tfl} = \mathrm{tr}\left(\hat{\boldsymbol{X}}_{tf}^{-1}\boldsymbol{G}_{fl}\right)$, and $\boldsymbol{A}_{fl}$ and $\boldsymbol{B}_{fl}$ are given by

$$\boldsymbol{A}_{fl} = \sum_{t,k}\lambda_{tflk}\hat{\boldsymbol{X}}_{tf}^{-1}, \quad (19)$$

$$\boldsymbol{B}_{fl} = \sum_{t,k}\frac{2M+\nu}{\nu+2\mathrm{tr}(\hat{\boldsymbol{X}}_{tf}^{-1}\boldsymbol{X}_{tf})}\lambda_{tflk}\hat{\boldsymbol{X}}_{tf}^{-1}\boldsymbol{X}_{tf}\hat{\boldsymbol{X}}_{tf}^{-1}, \quad (20)$$

We disambiguate the scales of $\boldsymbol{U}$ and $\boldsymbol{G}$ by $u_{kl} \leftarrow \frac{u_{kl}}{\sum_l u_{kl}}$ and $\boldsymbol{G}_{fl} \leftarrow \frac{\boldsymbol{G}_{fl}}{\mathrm{tr}(\boldsymbol{G}_{fl})}$.

5. EVALUATION

This section reports experiments conducted to compare the source separation performance of the conventional MNMF method and that of the proposed t-MNMF method.

5.1. Experimental conditions

We used three-channel audio signals of musical pieces ($M = 3$). Those signals were synthesized using an impulse response named E2A in the RWCP database [18]. The reverberation time (RT_{60}) of the recording environment was 300 ms. The sound sources of each recording are listed in Table 1 ($L = 3$). These data were obtained in the Signal Separation Evaluation Campaign (SiSEC) [19]. The recordings were down-sampled from 44.1 kHz to 16 kHz. The STFT frame size and frame shift were 512 and 160 samples, respectively.

We compared t-MNMF ($\nu = 0.5, 1, 2, 5, 10, 20$) with conventional MNMF. The number of t-MNMF bases K was set to 60. $\boldsymbol{W}$ and $\boldsymbol{H}$ were initialized randomly the diagonal values of $\boldsymbol{G}_{fl}$ were initially all set to $1/M$, and the off-diagonal values were initially all set to zero. The diagonal elements of $\boldsymbol{U}$ were initialized with values calculated by adding

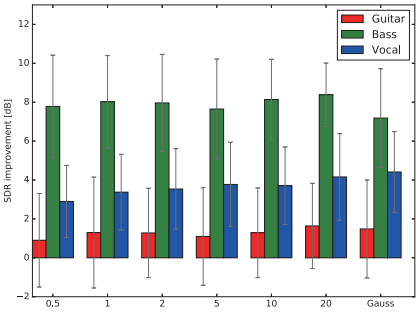

Fig. 2. Source separation performance for recording ID1.

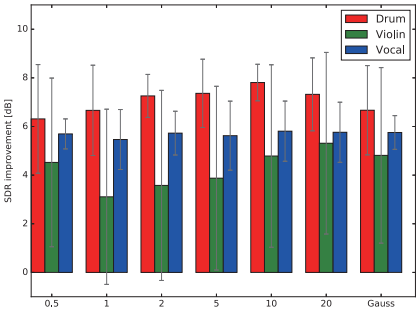

Fig. 3. Source separation performance for recording ID2.

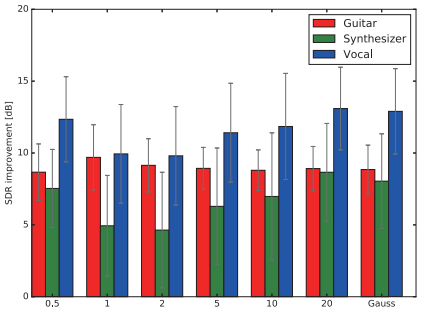

Fig. 4. Source separation performance for recording ID3.

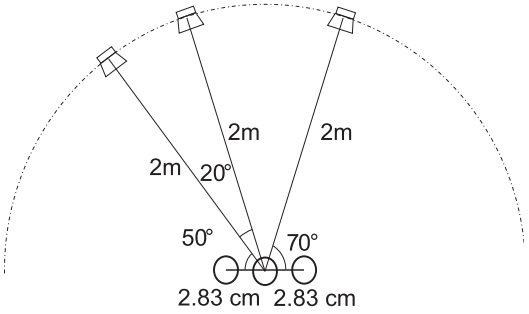

Fig. 5. Conditions for recording room impulse response.

to $1/L$ random numbers uniformly distributed between -0.01 and 0.01. The updating of $\boldsymbol{W}, \boldsymbol{H}, \boldsymbol{U}$, and $\boldsymbol{G}$ was conducted as follows. First, we updated $\boldsymbol{U}$ and $\boldsymbol{W}$ 20 times by using Eqs. (15) and (16) in turns. Then, we updated $\boldsymbol{W}, \boldsymbol{H}, \boldsymbol{U}$, and $\boldsymbol{G}$ 500 times by using Eqs. (15), (16), (17), and (18) in turn.

The separation performance was measured with the source-to-distortion ratio (SDR) [20] improvement that indicates the difference between the SDR of a target mixture signal and that of a separated signal. The SDR was calculated by using the BSS EVAL Toolbox [20]. Since the methods used involved randomness, we conducted 100 trials. In all trials, each variant of t-MNMF and MNMF was initialized in the same way.

5.2. Experimental results

Figs. 2, 3, and 4 show the SDR improvements of the separated signals estimated by using the proposed method (when the parameter ν was set to $0.5, 1, 2, 5, 10, 20$) and the conventional MNMF method. For each recording (ID1, ID2, and ID3) the SDR improvements of the proposed method ($\nu = 10, 20$) were superior to those of the conventional method. However, the differences between maximum values and minimum values of the SDRs of the proposed method and the conventional method were larger than 10 dB in many cases. This indicates that both methods are considered to have initial value dependency with regard to $\boldsymbol{W}$ and $\boldsymbol{H}$. A promising solution to this problem is to put rank-1 constraint on the estimation of spatial correlation matrices [5].

The SDR improvement for the bass guitar in ID1 and that for the drums in ID2 obtained by t-MNMF were significantly higher than those obtained by Gaussian MNMF. This indicates that t-MNMF works well especially when source spectrograms have low-rank structures. Since the sounds of a bass guitar and those of drums tend to be repeated many times in a musical piece, the spectrograms of these musical instruments can be approximated well as low-rank matrices.

Table 1. Music sources.

ID1	Song: bearlin_roads_snip_85_99 Sources: Guitar/Bass/Vocal
ID2	Song: for_minor_remember_the_name_snip_54_78 Sources: Drum/Violin/Vocal
ID3	Song: ultimate_nz_tour_snip_43_61 Sources: Guitar/Synthesizer/Vocal

6. CONCLUSION

This paper presented t-MNMF for blind source separation of multichannel audio signals. Using the multivariate complex t likelihood instead of the conventional multivariate complex Gaussian likelihood, t-MNMF includes MNMF as its special case. The experimental results showed that t-MNMF with an appropriate degree of freedom achieved better source separation performance than conventional MNMF.

To improve source separation performance and reduce the initial-value dependently, we plan to formulate an efficient extension of t-MNMF that forces spatial correlation matrices to have rank-1 structures [5]. This contributes to the reduction of the computational cost.

Acknowledgment: This study was partially supported by Grant-in-Aid for Scientific Research No. 24220006 and No. 15K12063.

7. REFERENCES

[1] P. Smaragdis, C. Févotte, G. Mysore, N. Mohammadiha, and M. Hoffman, "Static and dynamic source separation using nonnegative factorizations: A unified view," *Signal Processing Magazine*, vol. 31, no. 3, pp. 66–75, 2014.

[2] A. T. Cemgil, "Bayesian inference for nonnegative matrix factorisation models," *Computational Intelligence and Neuroscience*, 2009.

[3] A. Ozerov, A. Liutkus, R. Badeau, and G. Richard, "Coding-based informed source separation: Nonnegative tensor factorization approach," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 8, pp. 1699–1712, 2013.

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

[5] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, "Relaxation of rank-1 spatial constraint in overdetermined blind source separation," in *European Signal Processing Conference (EUSIPCO)*, 2015, pp. 1261–1265.

[6] K. Yoshii, K. Itoyama, and M. Goto, "Student's t nonnegative matrix factorization and positive semidefinite tensor factorization for single-channel audio source separation," in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2016, pp. 51–55.

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

[8] A. Liutkus, D. Fitzgerald, and R. Badeau, "Cauchy nonnegative matrix factorization," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2015.

[9] H. Sawada, , R. Mukai, S. Araki, and S. Makino, "A robust and precise method for solving the permutation problem of frequency-domain blind source separation," *IEEE Transactions on Speech and Audio Processing*, vol. 12, no. 5, pp. 530–538, 2004.

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

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

[12] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent component analysis and applications*, Academic press, 2010.

[13] N. Ono, "Stable and fast update rules for independent vector analysis based on auxiliary function technique," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2011, pp. 189–192.

[14] I. Lee, T. Kim, and T.-W. Lee, "Fast fixed-point independent vector analysis algorithms for convolutive blind source separation," *Signal Processing*, vol. 87, no. 8, pp. 1859–1871, 2007.

[15] P. Comon, "Tensors: a brief introduction," *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 44–53, 2014.

[16] M. Nakano, H. Kameoka, J. Le Roux, Y. Kitano, N. Ono, and S. Sagayama, "Convergence-guaranteed multiplicative algorithms for non-negative matrix factorization with beta divergence," in *International Workshop on Machine Learning for Signal Processing (MLSP)*, 2010, pp. 283–288.

[17] H. Sawada, H. Kameoka, S. Araki, and N. Ueda, "Efficient algorithms for multichannel extensions of Itakura-Saito nonnegative matrix factorization," in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2012, pp. 261–264.

[18] F. Asano T. Nishiura S. Nakamura, K. Hiyane and T. Yamada, "Acoustical sound database in real environments for sound scene understanding and hands-free speech recognition," *LREC2000*, pp. 965–968, 2000.

[19] S. Araki, F. Nesta, E. Vincent, Z. Koldovskỳ, G. Nolte, A. Ziehe, and A. Benichoux, "The 2011 signal separation evaluation campaign (sisec2011): audio source separation," in *Latent Variable Analysis and Signal Separation*, 2012, pp. 414–422.

[20] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 4, pp. 1462–1469, 2006.