A Generalization of Laplace Nonnegative Matrix Factorization and Its Multichannel Extension

Hiroki Tanji*, Takahiro Murakami*†, and Hiroyuki Kamata*

* Department of Electronics and Bioinformatics, School of Science and Technology, Meiji University, Japan

† Centre for Vision, Speech and Signal Processing (CVSSP), University of Surrey, UK

{htanji, tmrkm, kamata}@meiji.ac.jp

Abstract—The aim of this paper is to generalize the statistical models of nonnegative matrix factorization (NMF) and multichannel NMF (MNMF). For the NMF and its multichannel extensions, various statistical models have been proposed to improve the model flexibility in the literature on signal separation. However, few studies have been done on the generalization which includes the model based on the Laplace distribution. Thus, we propose the generalized models of the NMF and the MNMF, which include the models based on the Gaussian distribution and the two types of the Laplace distributions, using the Bessel function distribution. To estimate unknown model parameters, we derive the update rules based on the majorization-minimization algorithm. The performances of the proposed NMF and MNMF are evaluated in fitting synthetic data and music signal separation, respectively.

I. INTRODUCTION

Nonnegative matrix factorization (NMF) [1] have been widely exploited in single-channel signal separation [2]–[5]. In the NMF for audio signals, the observed nonnegative spectrogram $Y \in \mathbb{R}_+^{M \times N}$ ($\mathbb{R}_+ = [0, \infty)$) is decomposed into the set of frequently appearing spectral patterns $W \in \mathbb{R}_+^{M \times K}$ and time-varying activation weights $H \in \mathbb{R}_+^{K \times N}$ [6] as $Y \simeq WH$.

To obtain W and H, the reconstruction error between Y and WH is minimized. Fevotte derived the Itakura-Saito divergence for the reconstruction error, assuming the Gaussian distribution for each time-frequency bin of the observed complex spectrogram [7]. As alternatives of the complex Gaussian distribution, the Cauchy and the Levy distributions have been used for the divergence to model robust NMF against outliers [4], [8]. Moreover, we have been proposed Laplace-NMF [9], which is based on the Laplace distribution, motivated by the previous works [10], [11] that employ the Laplace distribution for acoustic signals. Note that the Laplace distribution has two types of definitions using the modified Bessel function of the second kind [10], [12] and the exponential function [11]. For convenience, we henceforth refer to them as the Bessel-function-type (BF) and the exponential-function-type (EF) Laplace distributions, respectively. In [9], using these definitions, two NMF named BF- and EF-Laplace-NMF are proposed.

To improve model flexibility, there have been several works on the generalization of the statistical model for the NMF. For a generalization which includes the NMF based on the Gaussian (Gaussian-NMF [7]) and the Cauchy distribution (Cauchy-NMF [4]), Simsekli has formulated α-stable-NMF [13] using the symmetric α-stable (SαS) distribution. Due to the difficulty of representing the probability density function (PDF) of the SαS distribution, in [5], the NMF based on the Student's t distribution has been proposed as another generalization of Gaussian- and Cauchy-NMF. Moreover, using the generalized Gaussian distribution (GGD), Kitamura has recently proposed GGD-NMF [14], which includes Gaussian- and EF-Laplace-NMF.

These statistical models for the NMF have been applied to audio denoising and semi-supervised signal separation. However, the NMF cannot be straightforwardly applied to blind signal separation, because the basis spectra are not grouped into each source. To overcome this drawback, multichannel NMF (MNMF) has been proposed in [15]. In the MNMF, using the spatial covariance matrix [16], spatial properties of the sources are introduced to the NMF. Using the spatial properties, the basis spectra are clustered. The MNMF has been applied to speech recognition [17] and unsupervised beamforming [18] in addition to blind signal separation. The optimization algorithm of the MNMF estimates the mixing system and the source spectrograms which minimize the cost function based on the Gaussian distribution. Also in the MNMF, the cost function based on the Student's t distribution has been proposed to improve flexibility [19].

Independent low-rank matrix analysis (ILRMA) [20], [21] has been recently growing up as another multichannel extension of the NMF. Owing to the restricted spatial model, the ILRMA achieves faster convergence and better robustness to initialization than the MNMF. However, the ILRMA is not executable when the number of microphone is less than the number of sources because it estimates the demixing system.

In this paper, towards flexible models of the NMF and the MNMF, we seek their generalized models which include the model based on the Laplace distribution. For our model, we focus on the Bessel function distribution [22]. In this distribution, its sharpness is controlled by the shape parameter $\eta > 0$. When η takes $\eta = 1$ and $\eta = 3/2$, the distribution of a univariate complex-valued random variable is reduced to the BF [10], [12] and the EF Laplace distributions [11], respectively. Moreover, when $\eta \to \infty$, the Bessel function distribution converges to the Gaussian distribution [22]. These characteristics allow to generalize the statistical models for the NMF and the MNMF.

Using the Bessel function distribution, we propose the generalized statistical models called Bessel-NMF and Bessel-MNMF. Thanks to the characteristics of the distribution, our statistical models include the models based on the Laplace and the Gaussian distributions. In this paper, we describe the cost function of Bessel-MNMF. Moreover, the convergence-guaranteed optimization algorithm is derived to minimize the cost function. Using the result in Bessel-MNMF, we present the cost function and the update rules of the proposed single-channel NMF.

II. REVIEW OF MULTICHANNEL NMF

We review the conventional MNMF [15] in this section. Let $s_{mnl} \in \mathbb{C}$ be the complex spectrogram at the mth frequency bin and the nth frame of the lth source, where $l = 1, \ldots, L$, $m = 1, \ldots, M$, and $n = 1, \ldots, N$. The sources are observed using D microphones in a reverberant environment. For each observation point, the lth source is filtered by $\boldsymbol{g}_{ml} = [g_{ml1}, \ldots, g_{mlD}]^\top \in \mathbb{C}^D$. Therefore, using the spatial image [16] $\boldsymbol{\varsigma}_{mnl} = \boldsymbol{g}_{ml} s_{mnl}$, we have the observed multichannel complex spectrogram $\boldsymbol{y}_{mn} \in \mathbb{C}^D$ as $\boldsymbol{y}_{mn} = \sum_{l=1}^{L} \boldsymbol{\varsigma}_{mnl}$.

In the MNMF, the observed data $\boldsymbol{Y}_{mn} = \boldsymbol{y}_{mn} \boldsymbol{y}_{mn}^\dagger$ is approximated using the Hermitian positive semidefinite matrix $\hat{\boldsymbol{Y}}_{mn}$ as

$$\boldsymbol{Y}_{mn} \simeq \hat{\boldsymbol{Y}}_{mn} = \sum_{l=1}^{L} \boldsymbol{A}_{ml} v_{mnl}, \qquad (1)$$

where $\dagger$ is the conjugate transpose, $\boldsymbol{A}_{ml} \in \mathbb{C}^{D \times D}$ is the spatial covariance matrix [16], which represents the spatial paths, and $v_{mnl} \in \mathbb{R}_+$ represents the power spectrogram of the source. Using the basis spectra $\boldsymbol{W} = [w_{mk}]$ and their weights $\boldsymbol{H} = [h_{kn}]$ in the NMF, v_{mnl} is represented as

$$v_{mnl} = \sum_{k=1}^{K} q_{lk} w_{mk} h_{kn}, \qquad (2)$$

where K is the number of bases, and $q_{lk} \in [0, 1]$ is the contribution of the kth basis for lth source satisfying $\sum_{l=1}^{L} q_{lk} = 1$.

The cost function for the conventional MNMF is derived using the Gaussian distribution. When $\boldsymbol{\varsigma}_{mnl}$ follows the Gaussian distribution, $\boldsymbol{y}_{mn}$ also follows the Gaussian with zero mean and the covariance matrix $\hat{\boldsymbol{Y}}_{mn}$ written as

$$\mathcal{N}_{\mathbb{C}}(\boldsymbol{y}_{mn}; \boldsymbol{0}, \hat{\boldsymbol{Y}}_{mn}) = \pi^{-D} \det[\hat{\boldsymbol{Y}}_{mn}]^{-1} \exp(-\text{Tr}[\boldsymbol{Y}_{mn} \hat{\boldsymbol{Y}}_{mn}^{-1}]), \qquad (3)$$

where $\det[\cdot]$ is the determinant, and $\text{Tr}[\cdot]$ is the matrix trace. The cost function to be minimized in the conventional MNMF is therefore given as $-\sum_{m,n} \log \mathcal{N}_{\mathbb{C}}(\boldsymbol{y}_{mn}; \boldsymbol{0}, \hat{\boldsymbol{Y}}_{mn})$. The unknown parameters (i.e., $\boldsymbol{A}_{mn}$, q_{lk}, w_{mk}, and h_{kn}) are estimated using the multiplicative update rules.

III. STATISTICAL MODELS BASED ON THE BESSEL FUNCTION DISTRIBUTION

We propose Bessel-NMF and Bessel-MNMF in this section. In Sect. III-A and III-B, we describe the formulation of Bessel-MNMF, which is the generalization of the conventional
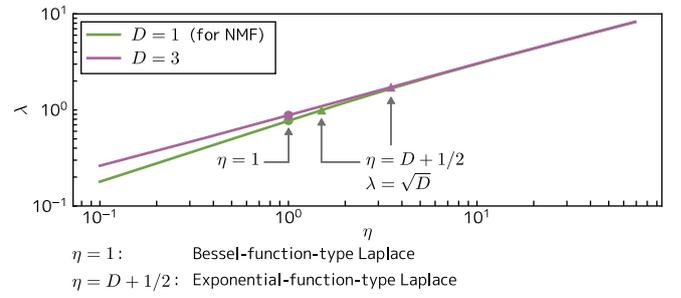

Fig. 1. Numerical solution of (7).

MNMF, and we derive the update rules for Bessel-MNMF, respectively. On the basis of the formulation of Bessel-MNMF, we present Bessel-NMF in Sect. III-C.

A. Cost function

To generalize the conventional cost function, we utilize the Bessel function distribution. Using the Gaussian scale mixture, the PDF [22] $\mathcal{BK}_{\mathbb{C}}(\boldsymbol{x})$ of a complex-valued random vector $\boldsymbol{x} \in \mathbb{C}^D$ is written as

$$\mathcal{BK}_{\mathbb{C}}(\boldsymbol{x}) = \int_{\mathbb{R}_+} \mathcal{N}_{\mathbb{C}}(\boldsymbol{x}; \boldsymbol{0}, z\boldsymbol{\Sigma}) \mathcal{G}(z; \eta, 1) dz, \qquad (4)$$

where $\mathcal{G}(z; \eta, 1)$ is the gamma distributions with shape η and unit scale defined as

$$\mathcal{G}(z; \eta, 1) = \frac{1}{\Gamma(\eta)} z^{\eta - 1} e^{-z}. \qquad (5)$$

When $\eta = 1$, $\mathcal{BK}_{\mathbb{C}}(\boldsymbol{x})$ is reduced to the BF Laplace distribution [10], [12]. Also, when $\eta = D + 1/2$, $\mathcal{BK}_{\mathbb{C}}(\boldsymbol{x})$ is equivalent to the EF Laplace distribution [11].

Substituting $\boldsymbol{y}_{mn}$, $\lambda^{-2} \hat{\boldsymbol{Y}}_{mn}$, and z_{mn} into $\boldsymbol{x}$, $\boldsymbol{\Sigma}$, and z of (4), respectively, we obtain the generalized cost function as

$$\mathcal{F}(\boldsymbol{\Theta}) = -\sum_{m,n} \log p(\boldsymbol{y}_{mn}; \hat{\boldsymbol{Y}}_{mn})$$

$$\stackrel{c}{=} -\sum_{m,n} \log \left(\det[\hat{\boldsymbol{Y}}_{mn}]^{-1} \text{Tr}[\boldsymbol{Y}_{mn} \hat{\boldsymbol{Y}}_{mn}^{-1}]^{\frac{\eta - D}{2}} \right.$$

$$\left. \mathcal{K}_{\eta - D} \left(2\lambda \sqrt{\text{Tr}[\boldsymbol{Y}_{mn} \hat{\boldsymbol{Y}}_{mn}^{-1}]} \right) \right), \qquad (6)$$

where λ is a positive constant, $\boldsymbol{\Theta} = \{\{\boldsymbol{A}_{mn}\}, \{q_{lk}\}, \boldsymbol{W}, \boldsymbol{H}\}$ is the set of unknown parameters, $\stackrel{c}{=}$ is equality up to constant terms, and $\mathcal{K}_\nu(x)$ ($\nu \in \mathbb{R}$, $x > 0$) is the modified Bessel function of the second kind defined as $\mathcal{K}_\nu(x) = \frac{1}{2} \int_0^\infty u^{-\nu - 1} \exp\left(-\frac{x}{2}(u + u^{-1})\right) du$. In (6), we have to choose λ so that $\boldsymbol{Y}_{mn} = \hat{\boldsymbol{Y}}_{mn}, \forall m, n$ minimizes the cost function. Therefore, λ is given as the solution of

$$\left. \frac{\partial \mathcal{F}(\boldsymbol{\Theta})}{\partial \hat{\boldsymbol{Y}}_{mn}} \right|_{\hat{\boldsymbol{Y}}_{mn} = \boldsymbol{Y}_{mn}} = 0. \qquad (7)$$

The solution of (7) can be obtained without $\boldsymbol{Y}_{mn}$ and $\hat{\boldsymbol{Y}}_{mn}$. We have $\lambda = \sqrt{D}$, especially when $p(\boldsymbol{y}_{mn}; \hat{\boldsymbol{Y}}_{mn})$ is equivalent to the EF Laplace distribution (i.e., $\eta = D + 1/2$). The

solutions for various values of η are depicted in Fig. 1. Note that the green line shows λ for the cost function of Bessel-NMF described in Sect. III-C.

Owing to introducing λ, $p(\boldsymbol{y}_{mn}; \hat{\boldsymbol{Y}}_{mn})$ converges to the Gaussian distribution when $\eta \to \infty$ [22]. Our statistical model therefore includes the model based on the Laplace and the Gaussian distributions.

B. Optimization algorithm

Given the observed data $\boldsymbol{Y}_{mn}$, (6) is minimized with respect to $\boldsymbol{\Theta}$. To estimate the parameters, we derive the multiplicative update rules based on the majorization-minimization (MM) algorithm [23], [24], using the upper bound of $\mathcal{F}(\boldsymbol{\Theta})$.

We introduce three inequalities to derive the upper bound in advance. For $\log(\boldsymbol{y}_{mn}; \boldsymbol{\Theta})$, the probabilistic form of the Jensen's inequality is given as

$$\log(\boldsymbol{y}_{mn}; \boldsymbol{\Theta}) \geq$$
$$\int_{\mathbb{R}_+} p(z_{mn}|\boldsymbol{y}_{mn}; \tilde{\boldsymbol{\Theta}}) \log \frac{p(\boldsymbol{y}_{mn}|z_{mn}; \boldsymbol{\Theta})p(z_{mn})}{p(z_{mn}|\boldsymbol{y}_{mn}; \tilde{\boldsymbol{\Theta}})} dz_{mn}, \quad (8)$$

where $\tilde{\boldsymbol{\Theta}}$ is the latest value of $\boldsymbol{\Theta}$, $\mathbb{E}_{p(x)}[x]$ is the expectation defined as $\mathbb{E}_{p(x)}[x] = \int x p(x) dx$, and

$$p(\boldsymbol{y}_{mn}|z_{mn}; \boldsymbol{\Theta}) = \mathcal{N}_{\mathbb{C}}(\boldsymbol{y}_{mn}; \boldsymbol{0}, \lambda^{-2} z_{mn} \hat{\boldsymbol{Y}}_{mn}) \quad (9)$$
$$p(z_{mn}) = \mathcal{G}(z_{mn}; \eta, 1). \quad (10)$$

For $\log \det[\hat{\boldsymbol{Y}}_{mn}]$ and $\mathrm{Tr}[\boldsymbol{Y}_{mn} \hat{\boldsymbol{Y}}_{mn}^{-1}]$, the first-order Taylor expansion [19] at any Hermitian positive semidefinite matrix $\boldsymbol{\Phi}_{mn} \in \mathbb{C}^{D \times D}$ and the Sawada's inequality [15] are

$$\det[\hat{\boldsymbol{Y}}_{mn}] \leq \log \det[\boldsymbol{\Phi}_{mn}] + \mathrm{Tr}[\boldsymbol{\Phi}_{mn}^{-1} \hat{\boldsymbol{Y}}_{mn}] - D \quad (11)$$
$$\mathrm{Tr}[\boldsymbol{Y}_{mn} \hat{\boldsymbol{Y}}_{mn}^{-1}] \leq \sum_{k,l} \frac{1}{q_{lk} w_{mk} h_{kn}} \mathrm{Tr}[\boldsymbol{Y}_{mn} \boldsymbol{R}_{mnlk}^{\dagger} \boldsymbol{A}_{ml}^{-1} \boldsymbol{R}_{mnlk}], \quad (12)$$

respectively, where $\boldsymbol{R}_{mnlk} \in \mathbb{C}^{D \times D}$ is the Hermitian positive semidefinite matrix that satisfies $\sum_{k,l} \boldsymbol{R}_{mnlk} = \boldsymbol{I}_D$, and $\boldsymbol{I}_D$ is the identity matrix of size D. In (11) and (12), the equalities hold if and only if $\boldsymbol{\Phi}_{mn} = \hat{\boldsymbol{Y}}_{mn}$ and $\boldsymbol{R}_{mnlk} = \boldsymbol{A}_{ml} \hat{\boldsymbol{Y}}_{mn}^{-1} q_{lk} w_{mk} h_{kn}$, respectively.

Applying (8) to $\mathcal{F}(\boldsymbol{\Theta})$, we obtain the upper bound $\mathcal{Q}(\boldsymbol{\Theta}; \tilde{\boldsymbol{\Theta}}) \geq \mathcal{F}(\boldsymbol{\Theta})$ as

$$\mathcal{Q}(\boldsymbol{\Theta}; \tilde{\boldsymbol{\Theta}}) \overset{c}{=} \sum_{m,n} \left(\det[\hat{\boldsymbol{Y}}_{mn}] + \zeta_{mn} \mathrm{Tr}[\boldsymbol{Y}_{mn} \hat{\boldsymbol{Y}}_{mn}^{-1}] \right), \quad (13)$$

where $\zeta_{mn} = \lambda^2 \mathbb{E}_{p(z_{mn}|\boldsymbol{y}_{mn}; \tilde{\boldsymbol{\Theta}})}[z_{mn}^{-1}]$. Equation (13) has simple representation, but it is still intractable with respect to $\hat{\boldsymbol{Y}}_{mn}$. Thus, we majorize $\mathcal{Q}(\boldsymbol{\Theta}; \tilde{\boldsymbol{\Theta}})$ using (11) and (12). Where $\boldsymbol{\Psi} = \{\{\boldsymbol{\Phi}_{mn}\}, \{\boldsymbol{R}_{mnlk}\}\}$, the upper bound $\mathcal{Q}^+(\boldsymbol{\Theta}; \tilde{\boldsymbol{\Theta}}, \boldsymbol{\Psi}) \geq \mathcal{Q}(\boldsymbol{\Theta}; \tilde{\boldsymbol{\Theta}})$ is written as

$$\mathcal{Q}^+(\boldsymbol{\Theta}; \tilde{\boldsymbol{\Theta}}, \boldsymbol{\Psi}) \overset{c}{=} \sum_{m,n} \left(\mathrm{Tr}[\boldsymbol{\Phi}_{mn}^{-1} \hat{\boldsymbol{Y}}_{mn}] \right.$$
$$\left. + \sum_{k,l} \frac{\zeta_{mn}}{q_{lk} w_{mk} h_{kn}} \mathrm{Tr}[\boldsymbol{Y}_{mn} \boldsymbol{R}_{mnlk}^{\dagger} \boldsymbol{A}_{ml}^{-1} \boldsymbol{R}_{mnlk}] \right). \quad (14)$$

Using the partial derivatives of $\mathcal{Q}^+(\boldsymbol{\Theta}; \tilde{\boldsymbol{\Theta}}, \boldsymbol{\Psi})$ with respect to $\boldsymbol{A}_{ml}$, q_{lk}, w_{mk}, and h_{kn}, we get the update rules as

$$\boldsymbol{A}_{ml} \leftarrow \boldsymbol{A}_{ml} \boldsymbol{\Delta}_{ml}^{\frac{1}{2}} \left(\boldsymbol{\Delta}_{ml}^{\frac{1}{2}} \boldsymbol{A}_{ml} \boldsymbol{\Lambda}_{ml} \boldsymbol{A}_{ml} \boldsymbol{\Delta}_{ml}^{\frac{1}{2}} \right)^{-\frac{1}{2}} \boldsymbol{\Delta}_{ml}^{\frac{1}{2}} \boldsymbol{A}_{ml}$$
$$(15)$$

$$q_{lk} \leftarrow q_{lk} \sqrt{\frac{\sum_{m,n} \zeta_{mn} \tau_{mnl} w_{mk} h_{kn}}{\sum_{m,n} \varrho_{mnl} w_{mk} h_{kn}}} \quad (16)$$

$$w_{mk} \leftarrow w_{mk} \sqrt{\frac{\sum_{l,n} \zeta_{mn} \tau_{mnl} q_{lk} h_{kn}}{\sum_{l,n} \varrho_{mnl} q_{lk} h_{kn}}} \quad (17)$$

$$h_{kn} \leftarrow h_{kn} \sqrt{\frac{\sum_{l,m} \zeta_{mn} \tau_{mnl} q_{lk} w_{mk}}{\sum_{l,m} \varrho_{mnl} q_{lk} w_{mk}}}, \quad (18)$$

where

$$\zeta_{mn} = \frac{\lambda \mathcal{K}_{D-\eta+1}(2\lambda \gamma_{mn})}{\gamma_{mn} \mathcal{K}_{D-\eta}(2\lambda \gamma_{mn})} \quad (19)$$

$$\gamma_{mn} = \sqrt{\mathrm{Tr}[\boldsymbol{Y}_{mn} \hat{\boldsymbol{Y}}_{mn}^{-1}]} \quad (20)$$

$$\tau_{mnl} = \mathrm{Tr}[\hat{\boldsymbol{Y}}_{mn}^{-1} \boldsymbol{Y}_{mn} \hat{\boldsymbol{Y}}_{mn}^{-1} \boldsymbol{A}_{ml}] \quad (21)$$

$$\varrho_{mnl} = \mathrm{Tr}[\hat{\boldsymbol{Y}}_{mn}^{-1} \boldsymbol{A}_{ml}] \quad (22)$$

$$\boldsymbol{\Lambda}_{ml} = \sum_{n,k} \hat{\boldsymbol{Y}}_{mn}^{-1} q_{lk} w_{mk} h_{kn} \quad (23)$$

$$\boldsymbol{\Delta}_{ml} = \sum_{n,k} \hat{\boldsymbol{Y}}_{mn}^{-1} \boldsymbol{Y}_{mn} \hat{\boldsymbol{Y}}_{mn}^{-1} \zeta_{mn} q_{lk} w_{mk} h_{kn}. \quad (24)$$

To eliminate the scale ambiguity, we normalize $\boldsymbol{A}_{ml}$, q_{lk}, and w_{mk} such that $\mathrm{Tr}[\boldsymbol{A}_{ml}]$, $\sum_l q_{lk}$, and $\sum_m w_{mk}$ are unity.

C. Bessel-NMF

On the basis of the discussion in the previous subsections, we describe the formulation and the optimization algorithm of Bessel-NMF. In the NMF, the time-frequency bin of the observed single-channel complex spectrogram $y_{mn}^{\mathbb{C}} \in \mathbb{C}$ is approximated as

$$|y_{mn}^{\mathbb{C}}|^\delta \simeq \hat{y}_{mn} = \sum_{k=1}^K w_{mk} h_{kn}, \quad (25)$$

where $\delta > 0$ is the domain parameter [14]. To obtain the approximate of the amplitude and the power spectra, δ is set to $\delta = 1$ and $\delta = 2$, respectively.

Replacing $\boldsymbol{Y}_{mn}$ and $\hat{\boldsymbol{Y}}_{mn}$ in (6) with $\boldsymbol{Y}_{mn} \leftarrow |y_{mn}^{\mathbb{C}}|^2$ and $\hat{\boldsymbol{Y}}_{mn} \leftarrow \hat{y}_{mn}^{2/\delta}$, respectively, we define the cost function of Bessel-NMF as

$$f(\boldsymbol{W}, \boldsymbol{H}) \overset{c}{=} -\sum_{m,n} \log \left(\hat{y}_{mn}^{-(\eta+1)/\delta} \mathcal{K}_{\eta-1} \left(2\lambda \frac{|y_{mn}^{\mathbb{C}}|}{\hat{y}_{mn}^{1/\delta}} \right) \right).$$
$$(26)$$

We show the divergence and the PDF with respect to $|y_{mn}^{\mathbb{C}}|^2$ in Fig. 2. The Bessel function distribution has heavy-tail compared to the Gaussian distribution. Moreover, we can see that the divergence and the PDF approach to the Gaussian, as η is increased.
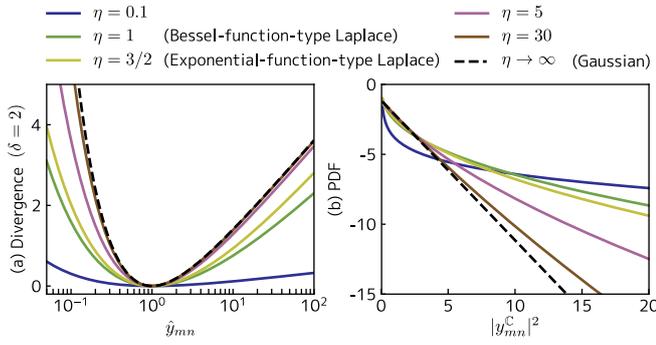
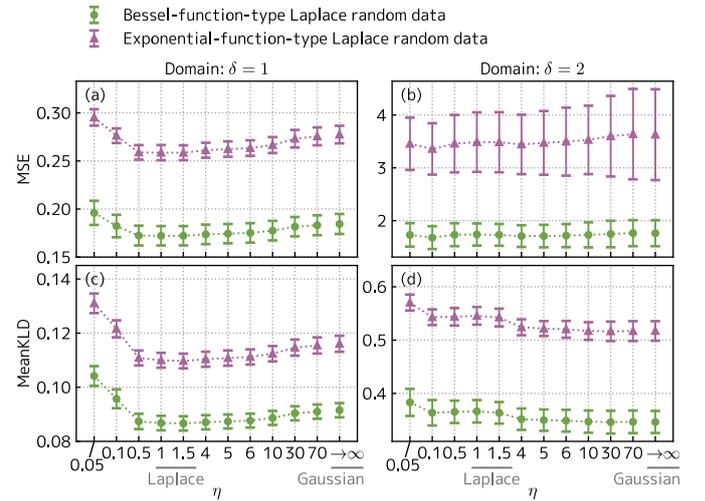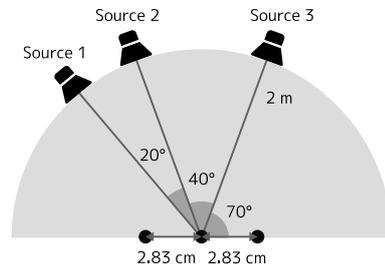Fig. 2. Cost function and PDF of Bessel-NMF.

Fig. 3. Reconstruction qualities by Bessel-NMF in fitting Laplace distributed data. Lower is better.

Fig. 4. Recording condition of the impulse response E2A (reverberation time $T_{60} = 300$ [ms]) extracted from RWCP database [25].

Equation (26) is also minimized using the MM algorithm to obtain $\boldsymbol{W}$ and $\boldsymbol{H}$. We use (8), (11), and the Jensen's inequality

$$\hat{y}_{mn}^{-2/\delta} \leq \sum_{k=1}^{K} \frac{r_{mnk}^{2/\delta+1}}{(w_{mk}h_{kn})^{2/\delta}}, \tag{27}$$

to derive the upper bound which is minimized instead of $f(\boldsymbol{W}, \boldsymbol{H})$, where $r_{mnk} > 0, \forall m, n, k$ and $\sum_k \rho_{mnk} = 1$. Using auxiliary variables $\boldsymbol{\varphi} = \{\varphi_{mn}\}$ ($\varphi_{mn} > 0$) and $\boldsymbol{r} = \{r_{mnk}\}$, the upper bound $f^+(\boldsymbol{W}, \boldsymbol{H}, \boldsymbol{\varphi}, \boldsymbol{r})$ is represented as

$$f^+(\boldsymbol{W}, \boldsymbol{H}, \boldsymbol{\varphi}, \boldsymbol{r}) = \sum_{m,n} \left(\frac{2\hat{y}_{mn}}{\delta \varphi_{mn}} + \sum_{k=1}^{K} \frac{\zeta_{mn}|y_{mn}|^2 r_{mnk}^{2/\delta+1}}{(w_{mk}h_{kn})^{2/\delta}} \right). \tag{28}$$

By taking the derivatives of $f^+(\boldsymbol{W}, \boldsymbol{H}, \boldsymbol{\varphi}, \boldsymbol{r})$ with respect to w_{mk} and h_{kn}, we get following update rules:

$$w_{mk} \leftarrow w_{mk} \left(\frac{\sum_n \frac{\zeta_{mn}|y_{mn}|^2}{\hat{y}_{mn}^{2/\delta+1}} h_{kn}}{\sum_n h_{kn}/\hat{y}_{mn}} \right)^{\delta/(\delta+2)} \tag{29}$$

$$h_{kn} \leftarrow h_{kn} \left(\frac{\sum_m \frac{\zeta_{mn}|y_{mn}|^2}{\hat{y}_{mn}^{2/\delta+1}} w_{mk}}{\sum_m w_{mk}/\hat{y}_{mn}} \right)^{\delta/(\delta+2)}. \tag{30}$$

Note that ζ_{mn} is updated using (19), where $D = 1$ and $\gamma_{mn} = |y_{mn}|/\hat{y}_{mn}^{1/\delta}$.

IV. SIMULATIONS

In this section, we present the results of the fitting data using Bessel-NMF. Moreover, we report the separation performance of Bessel-MNMF.

A. Fitting synthetic data using Bessel-NMF

To evaluate the fitting ability of Bessel-NMF, we generated the plaid random matrix [9] following the BF and the EF Laplace distributions. The size of the observed matrix and the number of basis were 50×50 and 5, respectively. The domain parameter was set to $\delta = 1$ and 2 to obtain the nonnegative observation. We conducted 500 iterations of the proposed algorithm of Bessel-NMF using the various value of η. Moreover, we performed the MM algorithm for Gaussian-NMF [14] to compare. To measure performance, the mean squared error

(MSE) and the mean of the generalized Kullback-Leibler (KL) divergence [1] were used, where they are defined as

$$\text{MSE} = \frac{1}{MN} \sum_{m,n} (|y_{mn}|^{\delta} - \hat{y}_{mn})^2 \tag{31}$$

$$\text{MeanKLD} = \frac{1}{MN} \sum_{m,n} \left(|y_{mn}|^{\delta} \log \frac{|y_{mn}|^{\delta}}{\hat{y}_{mn}} - |y_{mn}|^{\delta} + \hat{y}_{mn} \right), \tag{32}$$

respectively. We obtained the evaluation indices using 20 different sets of randomly generated initial values for 3 different observed matrices.

The results are depicted in Fig. 3. Since the cost function of Bessel-NMF converges the Gaussian cost function when $\eta \to \infty$, the fitting ability of Bessel-NMF approaches that of Gaussian-NMF as η is increased in Figs. 3(a), (c), and (d). Although the update rules of Bessel-NMF are derived using different upper bound from Gaussian-NMF, the algorithms show similar dependence on initialization.

B. Music signal separation

We compared the performances in musical signal separation. In this experiment, we synthesized three-channel music mixture using three sources extracted from SiSEC database [26]. The sets of sources are listed in Table I. The sampling
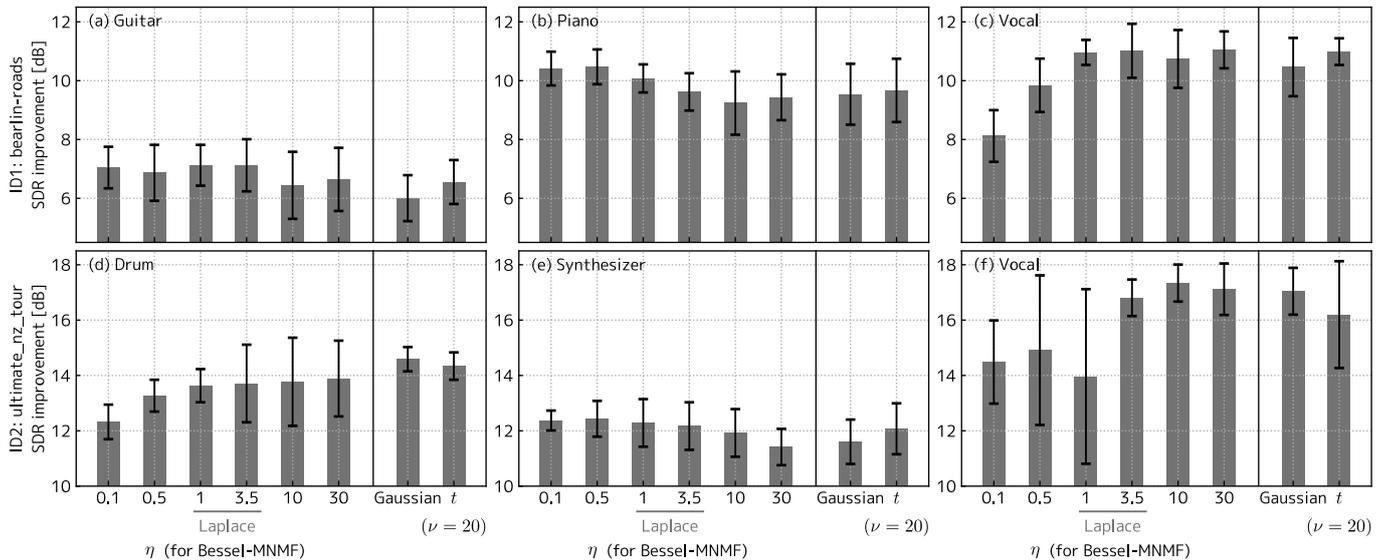
Fig. 5. Average SDR improvements of the conventional and the proposed MNMF. Higher is better.

TABLE I
DRY SOURCES FOR MUSIC SIGNAL SEPARATION EXTRACTED FROM
SiSEC2011 DATABASE [26].

ID	Name	Source (1, 2, 3)
ID1	bearlin-roads	guitar, piano, vocal (male)
ID2	ultimate_nz_tour	drum, synthesizer, vocal (female)

frequency was 16 [kHz]. We convoluted the impulse response "E2A" in RWCP database [25] with the sources to get the observed signal. The impulse response which we used was obtained under the conditions shown in Fig. 4. To obtain the multichannel spectrogram, we performed the short-time Fourier transform (STFT) using a 64-ms-long Hamming window with a 16-ms-long shift. We executed the MM algorithm for Bessel-MNMF with $\eta = 0.1, 0.5, 1, 3.5, 10$, and 30. Note again that Bessel-MNMF assumes the BF and the EF Laplace distributions when $\eta = 1$ and 3.5, respectively. To evaluate separation performances, we measured the source-to-distortion ratio (SDR) [27], using 10 different sets of the initial values. The SDRs were computed using the Python implementation [28] of the BSS Eval Matlab toolbox [27]. To compare, we performed Gaussian-MNMF [15] and t-MNMF [19]. In t-MNMF, the degree-of-freedom was set to $\nu = 20$, according to our best assessment. In this experiment, we assumed that the number of sources was known. The number of bases was set to 30. We initialized w_{mk} and h_{kn} using nonnegative random values. q_{lk} was initially set to random values around $1/L$. To reduce dependence on initial values, using the true value of the spatial images, we initialized A_{ml} such as $A_{ml} \leftarrow \frac{1}{N}\sum_{n=1}^{N}\varsigma_{mnl}\varsigma_{mnl}^{\dagger}$. The initial values of A_{ml}, q_{lk}, and w_{mk} were normalized as described in Sect. III-B. We updated the parameters using the following steps:

1) Update w_{mk} and h_{kn} 20 times using (17) and (18) while A_{ml} and q_{lk} are fixed.

2) Update A_{ml}, q_{lk}, w_{mk}, and h_{kn} 500 times using (15)–(18).

The evaluation result is shown in Fig. 5. The separation performances of Bessel-MNMF are better than that of the conventional methods excepted for Drum in ID2 (Fig. 5(d)). An informal listening test for ID1 shows that Bessel-MNMF provided the separated signals with reduced residual interference compared to the conventional MNMF. However, for ID2, Bessel-MNMF with $\eta = 0.1$ often provided unfavorable estimates for q_{lk} due to slow convergence. Although we used oracle initialization in this simulation, this results indicate that acceleration of the algorithm is an open problem.

V. CONCLUSIONS

In this paper, we proposed the generalizations of the NMF and the MNMF, named Bessel-NMF and Bessel-MNMF, based on the Bessel function distribution. Moreover, we derived the MM algorithms for the proposed cost functions. The evaluation results in music signal separation demonstrated the Bessel-MNMF gives promising results. Future works includes further evaluation of the proposed methods to find the optimal value of η and the acceleration of the algorithms.

ACKNOWLEDGMENT

REFERENCES

[1] D.D. Lee and H.S. Seung, "Learning the parts of objects with nonnegative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, Oct. 1999.

[2] D. Kitamura, N. Ono, H. Saruwatari, Y. Takahashi, and K. Kondo, "Discriminative and reconstructive basis training for audio source separation with semi-supervised nonnegative matrix factorization," in *Proc. 2016 IEEE International Workshop on Acoustic Signal Enhancement (IWAENC)*, Sep. 2016, pp. 1–5.

[3] P. Smaragdis, "Convolutive speech bases and their application to supervised speech separation," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 15, no. 1, pp. 1–12, Jan. 2007.

[4] A. Liutkus, D. FitzGerald, and R. Badeau, "Cauchy nonnegative matrix factorization," in *Proc. 2015 IEEE International Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New York, USA, Oct. 2015, pp. 1–5.

[5] 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 *Proc. 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, Mar. 2016, pp. 51–55.

[6] P. Smaragdis and J.C. Brown, "Non-negative matrix factorization for polyphonic music transcription," in *Proc. 2003 IEEE International Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New York, USA, Oct. 2003, pp. 177–180.

[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, Sep. 2008.

[8] P. Magron, R. Badeau, and A. Liutkus, "Lévy NMF for robust nonnegative source separation," in *Proc. 2017 IEEE International Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New York, USA, Oct. 2017, pp. 259–263.

[9] H. Tanji, T. Murakami, and H. Kamata, "Laplace nonnegative matrix factorization with application to semi-supervised audio denoising," in *Proc. 27th European Signal Processing Conference (EUSIPCO)*, A Coruña, Spain, Sep. 2019.

[10] B. Lee, T. Kaler, and R.W. Schafer, "Maximum-likelihood sound source localization with a multivariate complex Laplacian distribution," in *Proc. 11th International Workshop on Acoustic Echo and Noise Control (IWAENC)*, WA, USA, Sep. 2008.

[11] 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.

[12] T. Eltoft, T. Kim, and T.W. Lee, "On the multivariate Laplace distribution," *IEEE Signal Processing Letters*, vol. 13, no. 5, pp. 300–303, May 2006.

[13] U. Şimşekli, A. Liutkus, and A.T. Cemgil, "Alpha-stable matrix factorization," *IEEE Signal Processing Letters*, vol. 22, no. 12, pp. 2289–2293, Dec. 2015.

[14] D. Kitamura, "Nonnegative matrix factorization based on complex generative model," *Acoustical Science and Technology*, vol. 40, no. 3, pp. 155–161, May 2019.

[15] 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.

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

[17] Y. Tachioka, T. Narita, I. Miura, T. Uramoto, N. Monta, S. Uenohara, K. Furuya, S. Watanabe, and J. Le Roux, "Coupled initialization of multi-channel non-negative matrix factorization based on spatial and spectral information," in *Proc. 18th Annual Conference of the International Speech Communication Association (INTERSPEECH)*, Stockholm, Sweden, Aug. 2017, pp. 2461–2465.

[18] K. Shimada, Y. Bando, M. Mimura, K. Itoyama, K. Yoshii, and T. Kawahara, "Unsupervised beamforming based on multichannel nonnegative matrix factorization for noisy speech recognition," in *Proc. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2018, pp. 5734–5738.

[19] K. Kitamura, Y. Bando, K. Itoyama, and K. Yoshii, "Student's t multichannel nonnegative matrix factorization for blind source separation," in *Proc. 2016 IEEE International Workshop on Acoustic Signal Enhancement (IWAENC)*, Xian, China, Sep. 2016, pp. 1–5.

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

[21] 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, pp. 1–25, May 2018.

[22] S. Kotz, T.J. Kozubowski, and K. Podgórski, *The Laplace distribution and generalizations*, Birkhauser-Verlag, 2001.

[23] D.R. Hunter and K. Lange, "A tutorial on MM algorithms," *The American Statistician*, vol. 58, no. 1, pp. 30–37, Feb. 2004.

[24] M. Nakano, H. Kameoka, J. Le Roux, Y. Kitano, N. Ono, and S. Sagayama, "Convergence-guaranteed multiplicative algorithms for nonnegative matrix factorization with β-divergence," in *Proc. 2010 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, Kittila, Finland, Aug. 2010, pp. 283–288.

[25] 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. 2nd International Conference on Language Resources and Evaluation (LREC)*, Athens, Greece, May 2000, pp. 965–968.

[26] S. Araki, F. Nesta, E. Vincent, Z. Koldovsky, G. Nolte, A. Ziehe, and A. Benichoux, "The 2011 signal separation evaluation campaign (SiSEC2011): - audio source separation -," in *Proc. 10th International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA)*, Tel Aviv, Israel, Mar. 2012, pp. 414–422.

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

[28] C. Raffel, B. McFee, E.J. Humphrey, J. Salamon, O. Nieto, D. Liang, and D.P.W. Ellis, "MIR_EVAL: a transparent implementation of common MIR metrics," in *Proc. 15th International Society for Music Information Retrieval Conference (ISMIR)*, Taipei, Taiwan, Oct. 2014, pp. 367-372.