Multichannel Extensions of Non-negative Matrix Factorization with Complex-valued Data

Hiroshi Sawada, *Senior Member, IEEE,* Hirokazu Kameoka, *Member, IEEE,*
Shoko Araki, *Senior Member, IEEE,* Naonori Ueda, *Member, IEEE*

***Abstract*— This paper presents new formulations and algorithms for multichannel extensions of non-negative matrix factorization (NMF). The formulations employ Hermitian positive semidefinite matrices to represent a multichannel version of non-negative elements. Multichannel Euclidean distance and multichannel Itakura-Saito (IS) divergence are defined based on appropriate statistical models utilizing multivariate complex Gaussian distributions. To minimize this distance/divergence, efficient optimization algorithms in the form of multiplicative updates are derived by using properly designed auxiliary functions. Two methods are proposed for clustering NMF bases according to the estimated spatial property. Convolutive blind source separation (BSS) is performed by the multichannel extensions of NMF with the clustering mechanism. Experimental results show that 1) the derived multiplicative update rules exhibited good convergence behavior, and 2) BSS tasks for several music sources with two microphones and three instrumental parts were evaluated successfully.**

***Index Terms*— Non-negative matrix factorization, multichannel, blind source separation, convolutive mixture, clustering**

I. Introduction

Non-negative matrix factorization (NMF) is an unsupervised learning technique with a wide range of applications such as parts-based image representation [3], document clustering [4], and music transcription [5]. As the top part of Fig. 1 shows, NMF decomposes a given non-negative matrix $\mathbf{X}$ into two smaller non-negative matrices $\mathbf{T}$ and $\mathbf{V}$. When we analyze an audio/music signal with NMF, we typically employ a short-time Fourier transform (STFT) to obtain complex-valued representations in the time-frequency domain. Then, we make them non-negative by calculating the (squared) absolute values (see Eq. 1) in order to apply NMF. The bottom part of Fig. 1 shows that NMF extracts frequent sound patterns as five NMF bases from an audio clip containing five different notes.

A typical issue with NMF-based audio signal analysis is how to cluster the extracted NMF bases for a higher-level interpretation of the audio signal. Various NMF models have been proposed for that purpose. Temporal continuity [6] is considered on activation gains $\mathbf{V}$. The sequential modeling of bases $\mathbf{T}$ is conducted with convolutive NMF [7] and hidden Markov models [8], [9]. Shifted NMF [10], [11] identifies the bases that correspond to the notes played by an identical instrument. All these methods are related to *single-channel* (monaural) source separation.

As humans/animals have two ears, *multichannel* processing is a way of realizing a more general source separation capability because the spatial properties (directions or locations) of source signals can be exploited [12]. Specifically with an NMF-based method, the bases can be clustered according the spatial property, as Fig. 2 shows. With that in mind, multichannel extensions of NMF have been studied with the aim of realizing sound source separation and localization. In an instantaneous mixture case, the authors of [13], [14] were interested in the gain of each source to each microphone. Thus, all the values are still non-negative, and the notion of non-negativity remains clear. However, in a convolutive mixture case, the phase difference between different microphones is

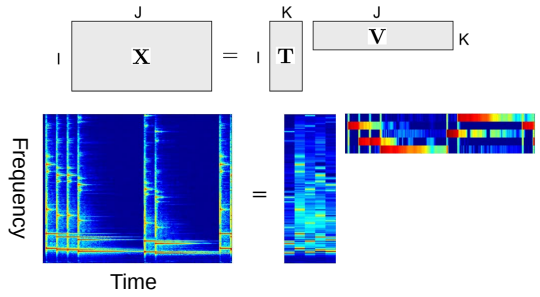

Fig. 1. Formulation of NMF (top) and its application to a music signal (bottom). Frequent sound patterns are identified in matrix $\mathbf{T}$ along with their activation periods and strengths shown in matrix $\mathbf{V}$.

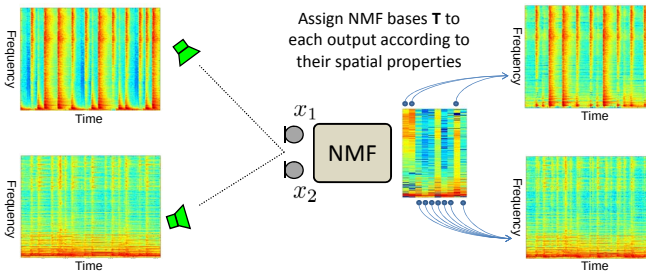

Fig. 2. Multichannel extensions of NMF associate the spatial property with each NMF basis. This enables us to cluster NMF bases according to the source location, and thus perform a source separation task.

Earlier versions of this work were presented at *the 2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 2011)* [1] and *the 2012 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2012)* [2] as workshop/conference papers. H. Sawada, H. Kameoka, S. Araki and N. Ueda are with NTT Communication Science Laboratories, NTT Corporation, 2-4 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0237, Japan (e-mail: sawada.hiroshi@lab.ntt.co.jp; kameoka.hirokazu@lab.ntt.co.jp; araki.shoko@lab.ntt.co.jp; ueda.naonori@lab.ntt.co.jp; phone: +81-774-93-5272, fax: +81-774-93-5155).

crucial information for source localization and separation. Thus, we need to handle complex-valued multichannel observations for these purposes, but the notion of non-negativity is not obvious. In [15], [16], the power spectrums of source signals are modeled with non-negative values, but there is no explicit description regarding non-negativity for mixing matrices or covariance matrices. In this paper, we propose that the *Hermitian positive semidefiniteness* of a matrix is a multichannel counterpart of *non-negativity*, and extend NMF to a multichannel case in a more generic way.

There are several choices available for the distance/divergence measures used in the NMF cost function including the Euclidean distance [17], the generalized Kullback-Leibler (KL) divergence [17], and the Itakura-Saito (IS) divergence [18]. In this paper, we define multichannel extensions of the Euclidean distance and the IS divergence, and extend the NMF model with these two definitions. We show that minimizing these distance/divergence is equivalent to maximizing the log-likelihood of the observations with appropriate statistical models utilizing multivariate complex Gaussian distributions.

The wide popularity of standard single-channel NMF comes from the fact that the algorithm is very easy to implement and works efficiently. In particular, *multiplicative update rules* [17] provide rapid convergence and have the attractive property of guaranteeing the non-negativity of all the matrix elements once they are initialized with non-negative values. In previous work on multichannel NMF [15], [16], however, expectation-maximization (EM) algorithms have been derived. It was reported that the algorithms were sensitive to parameter initialization, and thus they used the original source information for perturbed oracle initializations. This paper presents *multiplicative update rules* for multichannel extensions of NMF, and shows experimentally that their convergence behavior is similar to that of single-channel NMFs.

With the multichannel extensions of NMF presented in this paper, we have the estimations of the spatial properties for each basis. To perform a source separation task, we need to cluster the NMF bases for a source according to the similarity of the spatial properties. This paper proposes an automatic clustering mechanism that is built into the NMF model with cluster-indicator latent variables. The update rules are slightly changed but still multiplicative in form.

The main contributions of this paper are summarized as follows.

1) The notion of non-negativity is defined for a complex-valued vector (Section III-A).
2) Multiplicative update rules are derived for multichannel extensions of NMF (Section III-C). These updates provide faster convergence than the previous EM algorithms (Fig. 9).
3) Multichannel extensions are found for the Euclidean distance and IS divergence (Section III).
4) Methods for clustering NMF bases are proposed for a source separation task (Section IV).

In our previous work [1], [2], we succeeded merely in separating two sources and found it difficult to separate more sources. This paper newly proposes a bottom-up clustering method and a source separation procedure, in which redundant spatial properties are allowed. Consequently, we have succeeded in separating three sources for a variety of music sources. In addition, we changed the update rules for the multichannel Euclidean NMF from those shown in [1] to make the link to the standard single-channel Euclidean NMF clearer.

This paper is organized as follows. Section II explains the basics of the existing single-channel NMF. The proposed multichannel extensions and the clustering techniques are described in Sect. III and Sect. IV, respectively. Section V reports experimental results on the convergence behavior of the algorithms and source separation performance. Section VI concludes this paper.

II. Non-negative Matrix Factorization

This section reviews the formulation and algorithm of standard single-channel NMF [17]–[19]. Let us assume that we have a single-channel audio observation, to which we apply a short-time Fourier transform (STFT).

A. Formulation

Let $\tilde{x}_{ij} \in \mathbb{C}$ be the STFT coefficient at frequency bin i and time frame j. To apply NMF, we need to convert $\tilde{x}_{ij}$ to a non-negative value $x_{ij} \in \mathbb{R}_+$ via preprocessing. Typically, we take the absolute value or its squared value as

$$x_{ij} = \begin{cases} |\tilde{x}_{ij}| \\ |\tilde{x}_{ij}|^2 = \tilde{x}_{ij}\tilde{x}_{ij}^* \end{cases} \tag{1}$$

where $\cdot^*$ represents a complex conjugate. Then, a matrix $\mathbf{X}$, $[\mathbf{X}]_{ij} = x_{ij}$, is constructed with all the preprocessed values x_{ij} for $i = 1, \ldots, I$ and $j = 1, \ldots, J$.

NMF factorizes the $I \times J$ matrix $\mathbf{X}$ into the product of an $I \times K$ matrix $\mathbf{T}$ and a $K \times J$ matrix $\mathbf{V}$. The parameter K specifies the number of NMF bases, and is generally determined empirically by the user. All the elements of the two matrices, $t_{ik} = [\mathbf{T}]_{ik}$ $v_{kj} = [\mathbf{V}]_{kj}$, should be non-negative, i.e., $t_{ik} \in \mathbb{R}_+$ and $v_{kj} \in \mathbb{R}_+$.

NMF algorithms are designed to minimize the distance/divergence between the given matrix $\mathbf{X}$ and its factored form $\mathbf{TV}$. Let

$$\hat{x}_{ij} = \textstyle\sum_{k=1}^{K} t_{ik} v_{kj} \tag{2}$$

be the factorization approximation of x_{ij}. Then the distance/divergence can be defined in a general form

$$D_*(\mathbf{X}, \{\mathbf{T}, \mathbf{V}\}) = \sum_{i=1}^{I} \sum_{j=1}^{J} d_*(x_{ij}, \hat{x}_{ij}) \tag{3}$$

where d_* specifies an element-wise distance/divergence. The following three types of distance/divergence are widely used:

Squared Euclidean distance

$$d_{Eu}(x_{ij}, \hat{x}_{ij}) = |x_{ij} - \hat{x}_{ij}|^2 , \tag{4}$$

Generalized Kullback-Leibler (KL) divergence

$$d_{KL}(x_{ij}, \hat{x}_{ij}) = x_{ij} \log \frac{x_{ij}}{\hat{x}_{ij}} - x_{ij} + \hat{x}_{ij} , \tag{5}$$

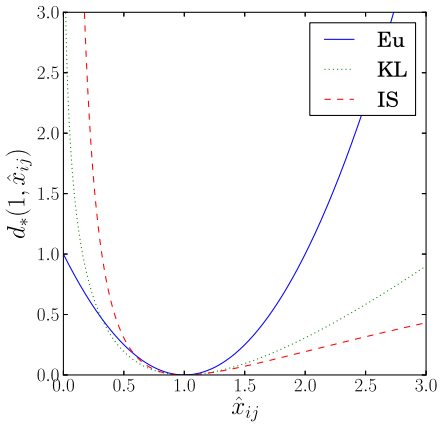

Fig. 3. Three types of distance/divergence with $x_{ij} = 1$: squared Euclid distance, generalized KL divergence and IS divergence

Itakura-Saito (IS) divergence

$$d_{IS}(x_{ij}, \hat{x}_{ij}) = \frac{x_{ij}}{\hat{x}_{ij}} - \log \frac{x_{ij}}{\hat{x}_{ij}} - 1 \,. \tag{6}$$

Figure 3 shows examples of these distance/divergences with $x_{ij} = 1$. We observe that KL and IS divergences are less sensitive to over-approximation than under-approximation. And from (6), we observe that IS divergence depends only on the ratio $x_{ij}/\hat{x}_{ij}$. Thus, for instance, $d_{IS}(900, 1000) = d_{IS}(9, 10)$. This property is favorable when analyzing most audio signals such as music and speech, where low frequency components have much higher energy than high frequency components. This is because low and high frequency components are treated equally with similar importance according to the property.

B. Algorithm: Multiplicative Update Rules

We can minimize the distance/divergence according to (3) together with (4), (5), or (6) in the following manner. First, the elements of $\mathbf{T}$ and $\mathbf{V}$ are randomly initialized with non-negative values. Then, the following update rules [17], [19] are iteratively applied until convergence.

Squared Euclidean distance

$$t_{ik} \leftarrow t_{ik} \frac{\sum_j x_{ij} v_{kj}}{\sum_j \hat{x}_{ij} v_{kj}} \,, \quad v_{kj} \leftarrow v_{kj} \frac{\sum_i x_{ij} t_{ik}}{\sum_i \hat{x}_{ij} t_{ik}} \tag{7}$$

KL divergence

$$t_{ik} \leftarrow t_{ik} \frac{\sum_j \frac{x_{ij}}{\hat{x}_{ij}} v_{kj}}{\sum_j v_{kj}} \,, \quad v_{kj} \leftarrow v_{kj} \frac{\sum_i \frac{x_{ij}}{\hat{x}_{ij}} t_{ik}}{\sum_i t_{ik}} \tag{8}$$

IS divergence

$$t_{ik} \leftarrow t_{ik} \sqrt{\frac{\sum_j \frac{x_{ij}}{\hat{x}_{ij}} \frac{v_{kj}}{\hat{x}_{ij}}}{\sum_j \frac{v_{kj}}{\hat{x}_{ij}}}} \,, \quad v_{kj} \leftarrow v_{kj} \sqrt{\frac{\sum_i \frac{x_{ij}}{\hat{x}_{ij}} \frac{t_{ik}}{\hat{x}_{ij}}}{\sum_i \frac{t_{ik}}{\hat{x}_{ij}}}} \tag{9}$$

These update rules are called *multiplicative*, since each element is updated by multiplying a scalar value, which is guaranteed to be non-negative.

C. Probability distributions related to distance/divergence

There are relations between the three distance/divergences (4)-(6) and specific probability distributions [18], [20], namely a Gaussian distribution $\mathcal{N}$, a complex Gaussian distribution $\mathcal{N}_c$ and a Poisson distribution $\mathcal{PO}$ as shown below. Studying these relationships helps us to consider multichannel extensions of NMF in the next section.

Minimizing the distance/divergence $D_*(\mathbf{X}, \{\mathbf{T}, \mathbf{V}\})$ is equivalent to maximizing the log-likelihood $\log p(\mathbf{X}|\mathbf{T}, \mathbf{V})$ or $\log p(\tilde{\mathbf{X}}|\mathbf{T}, \mathbf{V})$, where $\tilde{\mathbf{X}}$, $[\tilde{\mathbf{X}}]_{ij} = \tilde{x}_{ij}$, is a matrix of STFT coefficients.

Squared Euclidean distance

$$p(\mathbf{X}|\mathbf{T}, \mathbf{V}) = \prod_{i=1}^{I} \prod_{j=1}^{J} \mathcal{N}(x_{ij}|\hat{x}_{ij}, \frac{1}{2}) \,,$$
$$\mathcal{N}(x_{ij}|\hat{x}_{ij}, \frac{1}{2}) \propto \exp\left(-|x_{ij} - \hat{x}_{ij}|^2\right) \,. \tag{10}$$

KL divergence

$$p(\mathbf{X}|\mathbf{T}, \mathbf{V}) = \prod_{i=1}^{I} \prod_{j=1}^{J} \mathcal{PO}(x_{ij}|\hat{x}_{ij}) \,,$$
$$\mathcal{PO}(x_{ij}|\hat{x}_{ij}) = \frac{\hat{x}_{ij}^{x_{ij}}}{\Gamma(x_{ij}+1)} \exp\left(-\hat{x}_{ij}\right) \,. \tag{11}$$

where $\Gamma(x)$ is the Gamma function.

IS divergence

$$p(\tilde{\mathbf{X}}|\mathbf{T}, \mathbf{V}) = \prod_{i=1}^{I} \prod_{j=1}^{J} \mathcal{N}_c(\tilde{x}_{ij}|0, \hat{x}_{ij}) \,,$$
$$\mathcal{N}_c(\tilde{x}_{ij}|0, \hat{x}_{ij}) \propto \frac{1}{\hat{x}_{ij}} \exp\left(-\frac{|\tilde{x}_{ij}|^2}{\hat{x}_{ij}}\right) \,. \tag{12}$$

Regarding IS divergence, the likelihood $p(\tilde{\mathbf{X}}|\mathbf{T}, \mathbf{V})$ is calculated not for the matrix $\mathbf{X}$ of preprocessed non-negative values but for the matrix $\tilde{\mathbf{X}}$ of complex-valued STFT coefficients, and it is thus necessary to specify $x_{ij} = |\tilde{x}_{ij}|^2$ in a preprocessing step for the connection to (6).

When $x_{ij} = \hat{x}_{ij}$, the distance/divergence (4), (5) or (6) becomes 0 and each ij-term of the log-likelihood defined above is maximized. Therefore, the distance/divergence can be derived as the difference between the log-likelihoods of x_{ij} and $\hat{x}_{ij}$. We show the IS divergence case as an example:

$$\begin{aligned} d_{IS}(x_{ij}, \hat{x}_{ij}) &= \log \mathcal{N}_c(\tilde{x}_{ij}|0, x_{ij}) - \log \mathcal{N}_c(\tilde{x}_{ij}|0, \hat{x}_{ij}) \\ &= -\log x_{ij} - \frac{x_{ij}}{x_{ij}} - \left(-\log \hat{x}_{ij} - \frac{x_{ij}}{\hat{x}_{ij}}\right) \\ &= \frac{x_{ij}}{\hat{x}_{ij}} - \log \frac{x_{ij}}{\hat{x}_{ij}} - 1 \,. \end{aligned} \tag{13}$$

III. Multichannel Extensions of NMF

This section presents our multichannel extensions of NMF. Figure 4 shows an overview of the multichannel extensions (in red), in contrast with standard single-channel NMF (in blue). We begin with the multichannel extension of IS divergence (6), since this extension is the most natural. We then extend Euclidean distance (4) to a multichannel case. Unfortunately, we have not found a multichannel counterpart for generalized KL divergence (5).

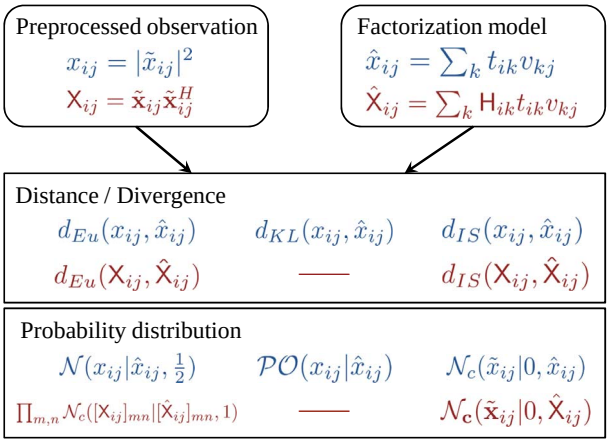

Fig. 4. Variations of NMF presented in this paper. Items that correspond to standard single-channel NMF are shown in blue. Three distance/divergences are discussed, and their corresponding probability distributions are presented. Items that correspond to multichannel extensions of NMF are shown in red. Two distance/divergence are extended to multichannel.

A. Formulation (IS divergence)

Let M be the number of microphones, and $\tilde{\mathbf{x}} = [\tilde{x}_1, \ldots, \tilde{x}_M]^T \in \mathbb{C}^M$ be a complex-valued vector for a time-frequency slot, with $\tilde{x}_m$ being the STFT coefficient at the m-th microphone. Let $\tilde{\mathbf{x}}_{ij}$ be such a vector at frequency bin i and time frame j. Now, let us introduce a multivariate complex Gaussian distribution $\mathcal{N}_{\mathbf{c}}$ that extends (12)

$$\mathcal{N}_{\mathbf{c}}(\tilde{\mathbf{x}}_{ij}|\mathbf{0}, \hat{\mathsf{X}}_{ij}) \propto \frac{1}{\det \hat{\mathsf{X}}_{ij}} \exp\left(-\tilde{\mathbf{x}}_{ij}^H \hat{\mathsf{X}}_{ij}^{-1} \tilde{\mathbf{x}}_{ij}\right), \tag{14}$$

where $\hat{\mathsf{X}}_{ij}$ is an $M \times M$ covariance matrix that should be Hermitian positive definite. Let $\mathsf{X}_{ij} = \tilde{\mathbf{x}}_{ij}\tilde{\mathbf{x}}_{ij}^H$ or

$$\mathsf{X} = \tilde{\mathbf{x}}\,\tilde{\mathbf{x}}^H = \begin{bmatrix} |\tilde{x}_1|^2 & \ldots & \tilde{x}_1\tilde{x}_M^* \\ \vdots & \ddots & \vdots \\ \tilde{x}_M\tilde{x}_1^* & \ldots & |\tilde{x}_M|^2 \end{bmatrix} \tag{15}$$

be the outer product of a complex-valued vector. We then define the **multichannel IS divergence** similarly to (13)

$$\begin{aligned} d_{IS}(\mathsf{X}_{ij}, \hat{\mathsf{X}}_{ij}) &= \log \mathcal{N}_{\mathbf{c}}(\tilde{\mathbf{x}}_{ij}|0, \mathsf{X}_{ij}) - \log \mathcal{N}_{\mathbf{c}}(\tilde{\mathbf{x}}_{ij}|0, \hat{\mathsf{X}}_{ij}) \\ &= -\log\det \mathsf{X}_{ij} - \mathrm{tr}(\mathsf{X}_{ij}\mathsf{X}_{ij}^{-1}) - \left[-\log\det \hat{\mathsf{X}}_{ij} - \mathrm{tr}(\mathsf{X}_{ij}\hat{\mathsf{X}}_{ij}^{-1})\right] \\ &= \mathrm{tr}(\mathsf{X}_{ij}\hat{\mathsf{X}}_{ij}^{-1}) - \log\det \mathsf{X}_{ij}\hat{\mathsf{X}}_{ij}^{-1} - M\,, \end{aligned} \tag{16}$$

where $\mathrm{tr}(\mathsf{X}) = \sum_{m=1}^M x_{mm}$ is the trace of a square matrix X.

We assume that the source locations are time-invariant in a source separation task (see Fig. 2). Therefore, we introduce a matrix H_{ik} that models the spatial property of the k-th NMF basis at frequency bin i. The matrix is of size $M \times M$ to be matched with the size of $\hat{\mathsf{X}}_{ij}$. Also, the matrix H_{ik} is Hermitian positive semidefinite to possess the non-negativity in a multichannel sense. Then, we model $\hat{\mathsf{X}}_{ij}$ with a sum-of-product form

$$\hat{\mathsf{X}}_{ij} = \sum_{k=1}^K \mathsf{H}_{ik}t_{ik}v_{kj}\,, \tag{17}$$

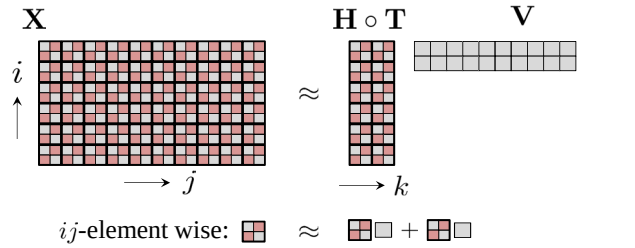

Fig. 5. Illustrative example of multichannel NMF: $I = 6$, $J = 10$, $K = 2$, $M = 2$. Non-negative values are shown in gray and complex values are shown in red.

where t_{ik} and v_{kj} are non-negative scalars as in the single-channel case. To solve the scaling ambiguity between H_{ik} and t_{ik}, let H_{ik} have a unit trace $\mathrm{tr}(\mathsf{H}_{ik}) = 1$.

In a matrix-wise notation, let $\mathbf{X}$ and $\mathbf{H}$ be $I \times J$ and $I \times K$ hierarchical matrices whose elements are $M \times M$ matrices, i.e., $[\mathbf{X}]_{ij} = \mathsf{X}_{ij}$ and $[\mathbf{H}]_{ik} = \mathsf{H}_{ij}$. Figure 5 provides an illustrative example in which multichannel NMF factorizes a hierarchically structured matrix $\mathbf{X}$ into the product of $\mathbf{H} \circ \mathbf{T}$ and $\mathbf{V}$, where $\circ$ represents the Hadamard product, i.e., $[\mathbf{H} \circ \mathbf{T}]_{ik} = \mathsf{H}_{ik}t_{ik}$. The multichannel NMF is formulated to minimize the total multichannel divergence similar to (3)

$$D_*(\mathbf{X}, \{\mathbf{T}, \mathbf{V}, \mathbf{H}\}) = \sum_{i=1}^I \sum_{j=1}^J d_*(\mathsf{X}_{ij}, \hat{\mathsf{X}}_{ij}) \tag{18}$$

where d_* is the element-wise multichannel divergence such as (16) in the IS divergence case.

B. Formulation (Squared Euclidean distance)

In this subsection, we consider a multichannel extension of Euclidean NMF. Thanks to the versatility of a univariate complex Gaussian distribution $\mathcal{N}_c$, we can model the preprocessed observations $\mathbf{X}$, $[\mathbf{X}]_{ij} = \mathsf{X}_{ij}$, as

$$\begin{aligned} p(\mathbf{X}|\mathbf{T}, \mathbf{V}, \mathbf{H}) &= \prod_{i=1}^I \prod_{j=1}^J \prod_{m=1}^M \prod_{n=1}^M \mathcal{N}_c([\mathsf{X}_{ij}]_{mn}|[\hat{\mathsf{X}}_{ij}]_{mn}, 1) \\ &\propto \prod_{i=1}^I \prod_{j=1}^J \exp(-||\mathsf{X}_{ij} - \hat{\mathsf{X}}_{ij}||_F^2)\,, \end{aligned} \tag{19}$$

where $||\mathsf{B}||_F^2 = \sum_{m=1}^M \sum_{n=1}^M |b_{mn}|^2$ is the squared Frobenius norm of matrix B. Maximizing the log of the likelihood (19) is equivalent to minimizing the distance (18) with the element-wise multichannel distance

$$d_{Eu}(\mathsf{X}_{ij}, \hat{\mathsf{X}}_{ij}) = ||\mathsf{X}_{ij} - \hat{\mathsf{X}}_{ij}||_F^2\,. \tag{20}$$

Therefore, multichannel Euclidean NMF has been formulated as minimizing (18) with (20).

When applying standard single-channel Euclidean NMF, it is typical that the absolute value $x_{ij} = |\tilde{x}_{ij}|$ in (1) is employed rather than the squared value to prevent some observations from being unnecessarily enhanced. In the same sense, an amplitude square-rooted version of the outer product (15)

would be useful in the multichannel Euclidean NMF:

$$\mathsf{X} = \begin{bmatrix} |x_1| & \dots & |x_1 x_M|^{\frac{1}{2}} \mathrm{sign}(x_1 x_M^*) \\ \vdots & \ddots & \vdots \\ |x_M x_1|^{\frac{1}{2}} \mathrm{sign}(x_M x_1^*) & \dots & |x_M| \end{bmatrix}, \tag{21}$$

where $\mathrm{sign}(x) = \frac{x}{|x|}$.

C. Algorithm: (Multiplicative Update Rules)

As shown in the next two subsections, the following multiplicative update rules are derived to minimize the total distance/divergence (18) with (16) or (20). These update rules reduce to their single channel counterparts (9) and (7) if $M = 1$, $\mathsf{X}_{ij} = x_{ij}$ and $\mathsf{H}_{ik} = 1$. Therefore, sets of these updates constitute multichannel extensions of NMF.

IS-NMF (IS divergence)

$$t_{ik} \leftarrow t_{ik} \sqrt{\frac{\sum_j v_{kj} \mathrm{tr}(\hat{\mathsf{X}}_{ij}^{-1} \mathsf{X}_{ij} \hat{\mathsf{X}}_{ij}^{-1} \mathsf{H}_{ik})}{\sum_j v_{kj} \mathrm{tr}(\hat{\mathsf{X}}_{ij}^{-1} \mathsf{H}_{ik})}} \tag{22}$$

$$v_{kj} \leftarrow v_{kj} \sqrt{\frac{\sum_i t_{ik} \mathrm{tr}(\hat{\mathsf{X}}_{ij}^{-1} \mathsf{X}_{ij} \hat{\mathsf{X}}_{ij}^{-1} \mathsf{H}_{ik})}{\sum_i t_{ik} \mathrm{tr}(\hat{\mathsf{X}}_{ij}^{-1} \mathsf{H}_{ik})}} \tag{23}$$

To update H_{ik}, we solve an algebraic Riccati equation (see Appendix I)

$$\mathsf{H}_{ik} \mathsf{A} \mathsf{H}_{ik} = \mathsf{B} \tag{24}$$

with

$$\mathsf{A} = \sum_j v_{kj} \hat{\mathsf{X}}_{ij}^{-1}, \quad \mathsf{B} = \mathsf{H}'_{ik} \left(\sum_j v_{kj} \hat{\mathsf{X}}_{ij}^{-1} \mathsf{X}_{ij} \hat{\mathsf{X}}_{ij}^{-1} \right) \mathsf{H}'_{ik}$$

where H'_{ik} is the target matrix before the update.

EU-NMF (Squared Euclidean distance)

$$t_{ik} \leftarrow t_{ik} \frac{\sum_j v_{kj} \mathrm{tr}(\mathsf{X}_{ij} \mathsf{H}_{ik})}{\sum_j v_{kj} \mathrm{tr}(\hat{\mathsf{X}}_{ij} \mathsf{H}_{ik})} \tag{25}$$

$$v_{kj} \leftarrow v_{kj} \frac{\sum_i t_{ik} \mathrm{tr}(\mathsf{X}_{ij} \mathsf{H}_{ik})}{\sum_i t_{ik} \mathrm{tr}(\hat{\mathsf{X}}_{ij} \mathsf{H}_{ik})} \tag{26}$$

$$\mathsf{H}_{ik} \leftarrow \mathsf{H}_{ik} \left(\sum_j v_{kj} \hat{\mathsf{X}}_{ij} \right)^{-1} \left(\sum_j v_{kj} \mathsf{X}_{ij} \right). \tag{27}$$

Post-processing is needed to make H_{ik} Hermitian and positive semidefinite. This can be accomplished by $\mathsf{H}_{ik} \leftarrow \frac{1}{2}(\mathsf{H}_{ik} + \mathsf{H}_{ik}^H)$ and then by performing eigenvalue decomposition as $\mathsf{H}_{ik} = \mathsf{U}\mathsf{D}\mathsf{U}^H$, setting all the negative elements of D at zero, and updating $\mathsf{H}_{ik} \leftarrow \mathsf{U}\mathsf{D}\mathsf{U}^H$ with the new D. We confirmed empirically that the update (27) followed by the post-processing always decreases the squared Euclidean distance. However, we have not yet found a theoretical guarantee.

For both the IS and Euclidean cases, unit-trace normalization $\mathsf{H}_{ik} \leftarrow \mathsf{H}_{ik} / \mathrm{tr}(\mathsf{H}_{ik})$ should follow.

D. Derivation of algorithm (IS-NMF)

This subsection explains the derivation of the multiplicative update rules (22)-(24) for IS divergence. For a given observation $\mathbf{X}$, the total distance (18) together with (16) can be written as

$$f(\mathbf{T}, \mathbf{V}, \mathbf{H}) = \sum_{i,j} \left[\mathrm{tr}(\mathsf{X}_{ij} \hat{\mathsf{X}}_{ij}^{-1}) + \log \det \hat{\mathsf{X}}_{ij} \right], \tag{28}$$

where constant terms are omitted. To minimize this function $f(\mathbf{T}, \mathbf{V}, \mathbf{H})$, we follow the optimization scheme of majorization [21], [22], in which an auxiliary (majorization) function is used. Let us define an auxiliary function

$$f^+(\mathbf{T}, \mathbf{V}, \mathbf{H}, \mathbf{R}, \mathbf{U}) = \sum_{i,j} \left[\sum_k \frac{\mathrm{tr}(\mathsf{X}_{ij} \mathsf{R}_{ijk}^H \mathsf{H}_{ik}^{-1} \mathsf{R}_{ijk})}{t_{ik} v_{kj}} + \log \det \mathsf{U}_{ij} + \frac{\det \hat{\mathsf{X}}_{ij} - \det \mathsf{U}_{ij}}{\det \mathsf{U}_{ij}} \right] \tag{29}$$

where R_{ijk} and U_{ij} are auxiliary variables that satisfy positive definiteness, $\sum_k \mathsf{R}_{ijk} = \mathsf{I}$ with I being an identity matrix of size M, and $\mathsf{U}_{ij} = \mathsf{U}_{ij}^H$ (Hermitian). It can be verified that the auxiliary function f^+ has two properties:

1) $f(\mathbf{T}, \mathbf{V}, \mathbf{H}) \leq f^+(\mathbf{T}, \mathbf{V}, \mathbf{H}, \mathbf{R}, \mathbf{U})$
2) $f(\mathbf{T}, \mathbf{V}, \mathbf{H}) = \min_{\mathbf{R}, \mathbf{U}} f^+(\mathbf{T}, \mathbf{V}, \mathbf{H}, \mathbf{R}, \mathbf{U})$

and the equality $f = f^+$ is satisfied when

$$\mathsf{R}_{ijk} = t_{ik} v_{kj} \mathsf{H}_{ik} \hat{\mathsf{X}}_{ij}^{-1}, \quad \mathsf{U}_{ij} = \hat{\mathsf{X}}_{ij} \tag{30}$$

(see Appendix II for the proof).

The function f is indirectly minimized by repeating the following two steps:

1) Minimizing f^+ with respect to $\mathbf{R}$ and $\mathbf{U}$ by (30), which makes $f(\mathbf{T}, \mathbf{V}, \mathbf{H}) = f^+(\mathbf{T}, \mathbf{V}, \mathbf{H}, \mathbf{R}, \mathbf{U})$.
2) Minimizing f^+ with respect to $\mathbf{T}$, $\mathbf{V}$ or $\mathbf{H}$, which also minimizes f.

For the second step, we calculate the partial derivatives of f^+ w.r.t. the variables $\mathbf{T}, \mathbf{V}$ and $\mathbf{H}$. Setting these derivatives at zero, we have the following equations.

$$t_{ik}^2 \leftarrow \frac{\sum_j \frac{1}{v_{kj}} \mathrm{tr}(\mathsf{R}_{ijk}^H \mathsf{H}_{ik}^{-1} \mathsf{R}_{ijk} \mathsf{X}_{ij})}{\sum_j \frac{\det \hat{\mathsf{X}}_{ij}}{\det \mathsf{U}_{ij}} v_{kj} \mathrm{tr}(\hat{\mathsf{X}}_{ij}^{-1} \mathsf{H}_{ik})} \tag{31}$$

$$v_{kj}^2 \leftarrow \frac{\sum_i \frac{1}{t_{ik}} \mathrm{tr}(\mathsf{R}_{ijk}^H \mathsf{H}_{ik}^{-1} \mathsf{R}_{ijk} \mathsf{X}_{ij})}{\sum_i \frac{\det \hat{\mathsf{X}}_{ij}}{\det \mathsf{U}_{ij}} t_{ik} \mathrm{tr}(\hat{\mathsf{X}}_{ij}^{-1} \mathsf{H}_{ik})} \tag{32}$$

$$\mathsf{H}_{ik} \left(t_{ik} \sum_j \hat{\mathsf{X}}_{ij}^{-1} v_{kj} \right) \mathsf{H}_{ik} = \sum_j \frac{\mathsf{R}_{ijk} \mathsf{X}_{ij} \mathsf{R}_{ijk}^H}{t_{ik} v_{kj}} \tag{33}$$

By substituting (30) into these equations, we obtain the multiplicative update rules (22)-(24).

E. Derivation of algorithm (EU-NMF)

The EU-NMF updates (25)-(27) can be derived in a similar manner. For a given observation $\mathbf{X}$, the total distance (18)

together with (20) and (17) can be written as

$$f(\mathbf{T},\mathbf{V},\mathbf{H}) = \sum_{i,j} \mathrm{tr}\left[(\sum_k \mathsf{H}_{ik} t_{ik} v_{kj})(\sum_k \mathsf{H}_{ik} t_{ik} v_{kj})^H \right] - \sum_{i,j,k} t_{ik} v_{kj} \mathrm{tr}(\mathsf{X}_{ij} \mathsf{H}_{ik}^H) - \sum_{i,j,k} t_{ik} v_{kj} \mathrm{tr}(\mathsf{H}_{ik} \mathsf{X}_{ij}^H)\,, \quad (34)$$

where constant terms are omitted. To minimize this function $f(\mathbf{T},\mathbf{V},\mathbf{H})$, we again follow the optimization scheme of majorization [21], [22]. Let us define an auxiliary function

$$f^+(\mathbf{T},\mathbf{V},\mathbf{H},\mathbf{R}) = \sum_{i,j,k} t_{ik}^2 v_{kj}^2 \mathrm{tr}(\mathsf{H}_{ik} \mathsf{R}_{ijk}^{-1} \mathsf{H}_{ik}^H) - \sum_{i,j,k} t_{ik} v_{kj} \mathrm{tr}(\mathsf{X}_{ij} \mathsf{H}_{ik}^H) - \sum_{i,j,k} t_{ik} v_{kj} \mathrm{tr}(\mathsf{H}_{ik} \mathsf{X}_{ij}^H) \quad (35)$$

with auxiliary variables R_{ijk} that satisfy Hermitian positive definiteness and $\sum_k \mathsf{R}_{ijk} = \mathsf{I}$. It can be verified that the auxiliary function f^+ has two properties:

1) $f(\mathbf{T},\mathbf{V},\mathbf{H}) \le f^+(\mathbf{T},\mathbf{V},\mathbf{H},\mathbf{R})$
2) $f(\mathbf{T},\mathbf{V},\mathbf{H}) = \min_{\mathbf{R}} f^+(\mathbf{T},\mathbf{V},\mathbf{H},\mathbf{R})$

and the equality $f = f^+$ is satisfied when

$$\mathsf{R}_{ijk} = \hat{\mathsf{X}}_{ij}^{-1} \mathsf{H}_{ik} t_{ik} v_{kj} \quad (36)$$

(see Appendix III for the proof).

The function f is indirectly minimized by repeating the following two steps:

1) Minimizing f^+ with respect to $\mathbf{R}$ by (36), which makes $f(\mathbf{T},\mathbf{V},\mathbf{H}) = f^+(\mathbf{T},\mathbf{V},\mathbf{H},\mathbf{R})$.
2) Minimizing f^+ with respect to $\mathbf{T}$, $\mathbf{V}$ or $\mathbf{H}$, which also minimizes f.

For the second step, we calculate the partial derivatives of f^+ w.r.t. the variables $\mathbf{T}, \mathbf{V}$ and $\mathbf{H}$. Setting these derivatives at zero, we have the following equations.

$$t_{ik} = \frac{\sum_j v_{kj} \mathrm{tr}(\mathsf{X}_{ij} \mathsf{H}_{ik})}{\sum_j v_{kj}^2 \mathrm{tr}(\mathsf{H}_{ik} \mathsf{R}_{ijk}^{-1} \mathsf{H}_{ik})} \quad (37)$$

$$v_{kj} = \frac{\sum_i t_{ik} \mathrm{tr}(\mathsf{X}_{ij} \mathsf{H}_{ik})}{\sum_i t_{ik}^2 \mathrm{tr}(\mathsf{H}_{ik} \mathsf{R}_{ijk}^{-1} \mathsf{H}_{ik})} \quad (38)$$

$$\mathsf{H}_{ik} = \left(t_{ik} \sum_j v_{kj}^2 \mathsf{R}_{ijk}^{-1} \right)^{-1} \left(\sum_j v_{kj} \mathsf{X}_{ij} \right) \quad (39)$$

By substituting (36) into these equations, we obtain the multiplicative update rules (25)-(27).

F. Interpretation of learned matrices $\mathbf{H}$

The multichannel NMF algorithm, (22)-(24) or (25)-(27), learns matrices $\mathbf{T}$, $\mathbf{V}$ and $\mathbf{H}$. The interpretation of the matrices $\mathbf{T}$ and $\mathbf{V}$ is the same as with standard single-channel NMF. This subsection provides an interpretation of the matrices $\mathbf{H}$, which are particular to the multichannel NMF.

To understand and interpret H_{ik}, we detail it by using the rank-1 convolutive model [23] as

$$\mathsf{H}_{ik} = \mathbf{h}_{ik} \mathbf{h}_{ik}^H + \epsilon_{ik} \mathsf{I}$$

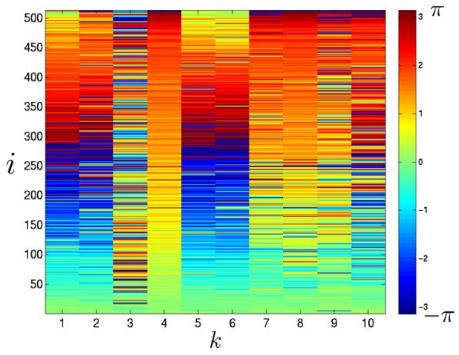

Fig. 6. An example of learned spatial properties. They are represented as $\arg([\mathsf{H}_{ik}]_{12})$, the phase difference between the first and second microphones, for each frequency bin i and NMF basis k.

where $\mathbf{h} = [h_1, \ldots, h_M]^T \in \mathbb{C}^M$ is a mixing vector whose m-th element h_m is the Fourier transform of a windowed impulse response from the source to the m-th microphone, ϵ_{ik} is a small positive scalar, and I is an $M \times M$ identity matrix. Then, we see that the diagonal elements of H_{ik} represent the power gain of the k-th basis at the i-th frequency bin to each microphone. And the off-diagonal elements represent the phase differences between microphones.

With a small microphone array, phase differences among microphones are typically more visible than gain differences. Figure 6 shows an example of phase differences that appeared on learned matrices $\mathbf{H}$. We interpret that the bases of number $k = 1, 2, 5, 6$ have similar spatial properties, and thus these are coming from the same direction. There is another group of NMF bases regarding $k = 4, 7, 8, 9$ that constitute another source coming from a different direction. How to cluster NMF bases according to such spatial properties will be discussed in the next section.

IV. Clustering NMF bases

This section presents two techniques for clustering NMF bases for a source separation task. The first is a top-down approach whose clustering mechanism is built in the NMF model. The second is a bottom-up approach that performs sequential pair-wise merges. Later, in Section V-C, we describe a robust source separation procedure that combines these two techniques.

A. Top-down approach with modified NMF model

Let us consider clustering K matrices $\mathsf{H}_{i1}, \ldots, \mathsf{H}_{iK}$ into $L < K$ classes $l = 1, \ldots, L$, and let z_{lk} indicate whether the k-th matrix belongs to the l-th cluster ($z_{lk} = 1$) or not ($z_{lk} = 0$). Then, H_{ik} in the NMF model can be replaced with $\sum_{l=1}^L \mathsf{H}_{il} z_{lk}$, and the sum-of-product form (17) is changed to

$$\hat{\mathsf{X}}_{ij} = \sum_{k=1}^K \left(\sum_{l=1}^L \mathsf{H}_{il} z_{lk} \right) t_{ik} v_{kj}\,. \quad (40)$$

The multichannel Euclidean distance (20) and IS divergence (16) can still be employed in the NMF model.

Now, we want to optimize the cluster-indicator latent variables $\mathbf{Z}$, $[\mathbf{Z}]_{lk} = z_{lk}$, in the same manner as $\mathbf{T}$ and $\mathbf{V}$. For that

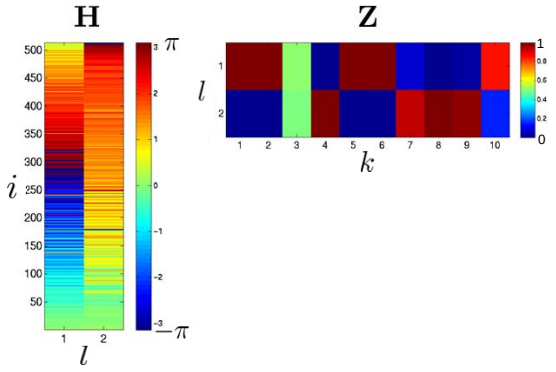

Fig. 7. The matrices $\mathbf{H}$ shown in Fig. 6 are clustered into two sets of matrices. They are represented as $\arg([\mathsf{H}_{il}]_{12})$. Latent variables $\mathbf{Z}$ show that k-th basis belongs to the first or second cluster in a soft sense.

purpose, let us allow z_{lk} to have a continuous value such that $z_{lk} \geq 0$ and $\sum_{l=1}^{L} z_{lk} = 1$. We consider that this relaxation corresponds to estimating the expectation of z_{lk} according to the posterior probability $p(z_{lk}|\mathbf{X}, \mathbf{T}, \mathbf{V}, \mathbf{H})$ instead of the value z_{lk} itself. But for simplicity, we do not change the notation of z_{lk} in (40) even with this relaxation. Figure 7 shows an example, where ten sets of matrices shown in Fig. 6 are clustered into two sets of matrices.

The algorithms to minimize

$$D_*(\mathbf{X}, \{\mathbf{T}, \mathbf{V}, \mathbf{H}, \mathbf{Z}\}) = \sum_{i=1}^{I}\sum_{j=1}^{J} d_*(\mathsf{X}_{ij}, \hat{\mathsf{X}}_{ij}) \tag{41}$$

with the element-wise distance/divergence (20) or (16) and with the model (40) can be derived in a similar manner as explained in Sect. III-E. The update rules are in the following forms.

IS-NMF (IS divergence)

$$t_{ik} \leftarrow t_{ik}\sqrt{\frac{\sum_l z_{lk}\sum_j v_{kj}\mathrm{tr}(\hat{\mathsf{X}}_{ij}^{-1}\mathsf{X}_{ij}\hat{\mathsf{X}}_{ij}^{-1}\mathsf{H}_{il})}{\sum_l z_{lk}\sum_j v_{kj}\mathrm{tr}(\hat{\mathsf{X}}_{ij}^{-1}\mathsf{H}_{il})}}, \tag{42}$$

$$v_{kj} \leftarrow v_{kj}\sqrt{\frac{\sum_l z_{lk}\sum_i t_{ik}\mathrm{tr}(\hat{\mathsf{X}}_{ij}^{-1}\mathsf{X}_{ij}\hat{\mathsf{X}}_{ij}^{-1}\mathsf{H}_{il})}{\sum_l z_{lk}\sum_i t_{ik}\mathrm{tr}(\hat{\mathsf{X}}_{ij}^{-1}\mathsf{H}_{il})}}, \tag{43}$$

$$z_{lk} \leftarrow z_{lk}\sqrt{\frac{\sum_{i,j} t_{ik}v_{kj}\mathrm{tr}(\hat{\mathsf{X}}_{ij}^{-1}\mathsf{X}_{ij}\hat{\mathsf{X}}_{ij}^{-1}\mathsf{H}_{il})}{\sum_{i,j} t_{ik}v_{kj}\mathrm{tr}(\hat{\mathsf{X}}_{ij}^{-1}\mathsf{H}_{il})}}. \tag{44}$$

For H_{il}, we solve an algebraic Riccati equation

$$\mathsf{H}_{il}\mathsf{A}\mathsf{H}_{il} = \mathsf{B} \tag{45}$$

with

$$\mathsf{A} = \textstyle\sum_k z_{lk}t_{ik}\sum_j v_{kj}\hat{\mathsf{X}}_{ij}^{-1}, \tag{46}$$

$$\mathsf{B} = \mathsf{H}'_{il}\left(\textstyle\sum_k z_{lk}t_{ik}\sum_j v_{kj}\hat{\mathsf{X}}_{ij}^{-1}\mathsf{X}_{ij}\hat{\mathsf{X}}_{ij}^{-1}\right)\mathsf{H}'_{il}. \tag{47}$$

where H'_{il} is the target matrix before the update.

EU-NMF (Squared Euclidean distance)

$$t_{ik} \leftarrow t_{ik}\frac{\sum_l z_{lk}\sum_j v_{kj}\mathrm{tr}(\mathsf{X}_{ij}\mathsf{H}_{il})}{\sum_l z_{lk}\sum_j v_{kj}\mathrm{tr}(\hat{\mathsf{X}}_{ij}\mathsf{H}_{il})} \tag{48}$$

Algorithm 1 Multichannel NMF with bottom-up clustering

procedure MCHNMF_BOTTOMUPCLUSTERING
  $iteration \leftarrow 0$
  while $L > finalClusterSize$ **do**
    $iteration \leftarrow iteration + 1$
    update $\mathbf{T}$ by (42) or (48)
    update $\mathbf{V}$ by (43) or (49)
    if $mod(iteration, interval) = 1$ **then**
      $(\mathbf{H}, \mathbf{Z}) \leftarrow$ PAIRWISEMERGE$(\mathbf{H}, \mathbf{Z})$
      $L \leftarrow L - 1$
    end if
    update $\mathbf{H}$ by (45) or (51)
    update $\mathbf{Z}$ by (44) or (50)
  end while
end procedure

procedure PAIRWISEMERGE$(\mathbf{H}, \mathbf{Z})$
  $(l_1, l_2) \leftarrow findPair(\mathbf{H})$
  $w_1 \leftarrow \sum_k z_{l_1 k}$
  $w_2 \leftarrow \sum_k z_{l_2 k}$
  $\{\mathsf{H}_1, \ldots, \mathsf{H}_I\} \leftarrow weightedMean(\mathbf{H}, l_1, l_2, w_1, w_2)$
  $\mathbf{H} \leftarrow removeAdd(\mathbf{H}, l_1, l_2, \{\mathsf{H}_1, \ldots, \mathsf{H}_I\})$
  $\mathbf{Z} \leftarrow merge(\mathbf{Z}, l_1, l_2)$
end procedure

$$v_{kj} \leftarrow v_{kj}\frac{\sum_l z_{lk}\sum_i t_{ik}\mathrm{tr}(\mathsf{X}_{ij}\mathsf{H}_{il})}{\sum_l z_{lk}\sum_i t_{ik}\mathrm{tr}(\hat{\mathsf{X}}_{ij}\mathsf{H}_{il})} \tag{49}$$

$$z_{lk} \leftarrow z_{lk}\frac{\sum_{i,j} t_{ik}v_{kj}\mathrm{tr}(\mathsf{X}_{ij}\mathsf{H}_{il})}{\sum_{i,j} t_{ik}v_{kj}\mathrm{tr}(\hat{\mathsf{X}}_{ij}\mathsf{H}_{il})} \tag{50}$$

$$\mathsf{H}_{il} \leftarrow \mathsf{H}_{il}\left(\textstyle\sum_k z_{lk}t_{ik}\sum_j v_{kj}\hat{\mathsf{X}}_{ij}\right)^{-1}\left(\sum_k z_{lk}t_{ik}\sum_j v_{kj}\mathsf{X}_{ij}\right) \tag{51}$$

For both the IS and Euclidean cases, unit-trace normalization $\mathsf{H}_{ik} \leftarrow \mathsf{H}_{ik}/\mathrm{tr}(\mathsf{H}_{ik})$ and unit-sum normalization $z_{lk} \leftarrow z_{lk} / (\sum_l z_{lk})$ should follow.

The clustering result obtained with the top-down approach heavily depends on the initial values of the cluster-indicator latent variables $\mathbf{Z}$. To prevent an important cluster from disappearing by chance, it is a good idea to have some redundant clusters by setting the cluster number L at a larger than expected. Then later, the redundant clusters can be merged by employing the bottom-up clustering shown in the next subsection.

B. Bottom-up clustering by sequential merge operation

This subsection explains another way to cluster NMF bases. It is based on a pair-wise merge operation, in which a pair with the minimum distance is identified and merged. The pair-wise distance between the l_1-th set and the l_2-th set is defined by using the Frobenius norm as

$$d_{\mathsf{H}}(l_1, l_2) = \textstyle\sum_{i=1}^{I} ||\mathsf{H}_{il_1} - \mathsf{H}_{il_2}||_F. \tag{52}$$

Algorithm 1 shows the whole procedure. Inside the basic updates for NMF, the pair-wise merge operation is interleaved at a rate specified by a variable $interval$. In the pair-wise

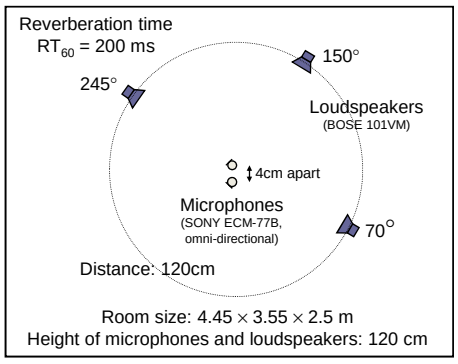

Fig. 8. Experimental setup for room impulse responses

merge operation, a pair (l_1, l_2) with the minimum distance is found, and a new set $\{\mathsf{H}_{1l}, \ldots, \mathsf{H}_{Il}\}$ is calculated as element-wise weighted means of the two sets. Then, the number of clusters is decreased by 1 as the new set is added and the l_1-th and l_2-th sets are removed. The main loop is repeated until the number of clusters becomes a specified number $finalClusterSize$.

C. Source separation

If NMF bases are appropriately clustered, source separation can be performed by using Wiener filters. Remember that $\tilde{\mathbf{x}}_{ij}$ is an M-dimensional complex vector representing STFT coefficients at frequency bin i and time frame j, and let $\tilde{\mathbf{y}}_{ij}^{(l)}$ be the STFT coefficient vector for the l-th separated signal. Then, the separated signals are obtained by the single-channel Wiener filter for the m-th channel

$$[\tilde{\mathbf{y}}_{ij}^{(l)}]_m = \frac{[\mathsf{H}_{il}]_{mm} \sum_{k=1}^{K} z_{lk} t_{ik} v_{kj}}{\sum_{l=1}^{L} [\mathsf{H}_{il}]_{mm} \sum_{k=1}^{K} z_{lk} t_{ik} v_{kj}} [\tilde{\mathbf{x}}_{ij}]_m \,, \tag{53}$$

or by the multichannel Wiener filter

$$\tilde{\mathbf{y}}_{ij}^{(l)} = \left(\sum_{k=1}^{K} z_{lk} t_{ik} v_{kj} \right) \mathsf{H}_{il} \hat{\mathsf{X}}_{ij}^{-1} \tilde{\mathbf{x}}_{ij} \,, \tag{54}$$

where $\hat{\mathsf{X}}_{ij}$ is the sum-of-product form defined in (40).

V. Experiments

A. Experimental Setups

We examined the proposed multichannel extensions of NMF with stereo ($M = 2$) music mixtures that contained three music parts. Sets of stereo mixtures were generated by convolving the music parts and the impulse responses measured in a real room whose conditions are shown in Fig. 8. The impulse responses were measured by using a maximum length sequence generated by a 17-th order polynomial over GF(2). We made four sets of mixtures using the music sources listed in Table I, which can be found at the professionally produced music recordings page of the Signal Separation Evaluation Campaign (SiSEC 2011) [24]. The mixtures were down-sampled to 16 kHz. The STFT frame size was 64 ms and the frame shift was 16 ms. The algorithms were coded with Matlab and run on an Intel Xeon W3690 (3.46GHz) processor.

TABLE I
Music sources

ID	Author / Song	Snip	Part
1	Bearlin Roads	85-99 (14 sec)	piano ambient+windchimes vocals
2	Another Dreamer The Ones We Love	69-94 (25 sec)	drums vocals guitar
3	Fort Minor Remember the Name	54-78 (24 sec)	drums vocals violins_synth
4	Ultimate NZ Tour	43-61 (18 sec)	drums guitar synth

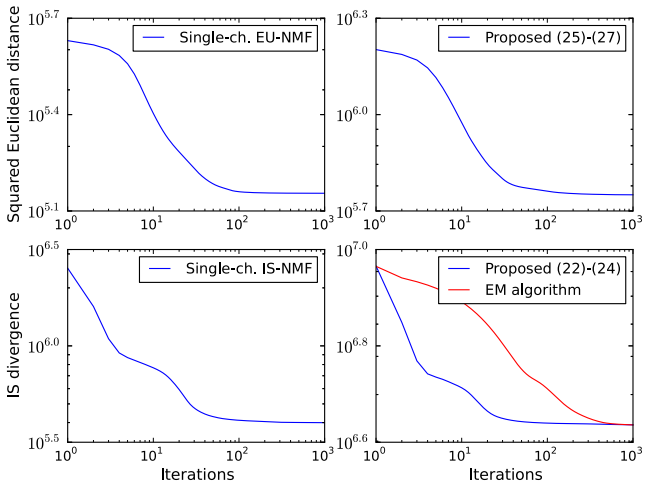

Fig. 9. Convergence behavior shown in log-log plots: EU-NMF (top) and IS-NMF (bottom), single-channel (left) and multichannel (right).

B. Convergence Behavior

Let us first show the convergence behavior of the multichannel NMF algorithms proposed in Sect. III. For comparison, we run the algorithms of single-channel NMF (7) and (9), multichannel NMF (25)-(27) and (22)-(24), and the EM algorithm for multichannel IS-NMF shown in Appendix IV. We set the number of NMF bases $K = 10$, and used a music mixture with ID = 1.

Figure 9 shows the convergence behavior for 1000 iterations, and Table II shows the computational time. NMF algorithms with IS divergence generally take more time than EU-NMF. We observe that the convergence behavior of the single-channel NMF algorithms is similar to that of the proposed multichannel NMF algorithms. With respect to multichannel IS-NMF, the proposed algorithms show faster convergence and computation than the EM algorithm.

TABLE II
Computational time (in seconds) for 1000 iterations with 14-second signals

	Single-ch.	Proposed	EM
EU-NMF	3.09	291.27	-
IS-NMF	13.24	1303.30	2958.28

C. Source separation procedure

This subsection explains the source separation procedure with the multichannel NMF proposed in Sect. IV. We adopted the following procedure so that the spatial properties of sources were well extracted.

1) Preprocessing: For EU-NMF: we normalized the power of the observation vectors at each frequency bin such that $\sum_j \tilde{\mathbf{x}}_{ij}\tilde{\mathbf{x}}_{ij}^H = 1$. This was to prevent the low frequency components from being dominant in the total distance (18). Then, we generated matrices X_{ij} as the amplitude square-rooted outer products (21). For IS-NMF: we generated matrices X_{ij} by the outer products (15). To prevent the determinant of multichannel IS divergence (16) from being zero, we then added a regularization term to each matrix as $\mathsf{X}_{ij} \leftarrow \mathsf{X}_{ij} + \epsilon\mathsf{I}$ with $\epsilon = 10^{-10}$ and I being an identity matrix.

2) Initialization: Matrices $\mathbf{T}$ and $\mathbf{V}$ were randomly initialized with non-negative entries. The diagonal elements of H_{il} were initially all set at $1/M$, and the off-diagonal elements were initially all set at zero. The elements of matrix $\mathbf{Z}$ were initialized with random values around $1/L$.

3) Multichannel NMF: We set the number of NMF bases K at 30 (ten times the number of sources). As for clustering the bases for each source, we employ both top-down and bottom-up approaches as follows.

i) 20 iterations to update $\mathbf{T}$ and $\mathbf{V}$.
ii) 200 iterations to update $\mathbf{T}$, $\mathbf{V}$, $\mathbf{H}$ and $\mathbf{Z}$ by the top-down approach with $L = L_{init} = 9$.
iii) Bottom-up clustering with $interval = 10$ until $L = 3$.
iv) 200 iterations to update $\mathbf{T}$, $\mathbf{V}$, $\mathbf{H}$ and $\mathbf{Z}$ by the top-down approach with $L = 3$.

Having redundant ($L = 9$) spatial properties (step ii) followed by the bottom-up clustering (step iii) contributes to robust estimations of the spatial properties. Some results will be shown in the next subsection with Fig. 12.

4) Separation: For EU-NMF, a single-channel Wiener filter (53) was used for each channel. For IS-NMF, a multichannel Wiener filter (54) was used. We selected these configurations because each of these produced better results empirically.

Figure 10 shows an example in which the IS divergence was minimized by the procedure. The blue line shows how steps i)–iv) work, especially when parameter L is decreased from 9 to 3. The red line shows the case where only the top-down approach was employed. In this example, the IS divergence was better minimized by having redundant ($L_{init} = 9$) spatial properties.

D. Source separation results

The separation performance was numerically evaluated in terms of the signal-to-distortion ratio (SDR) [25]. We need to know all the source images s_{ml}^{img} for all microphones $m = 1, \ldots, M$ and sources $l = 1, \ldots, L$. To calculate SDR_l for the l-th source, we first decompose the time-domain multichannel signals $y_{1l}, \ldots, y_{Ml}$ as

$$y_{ml}(t) = s_{ml}^{\mathrm{img}}(t) + y_{ml}^{\mathrm{spat}}(t) + y_{ml}^{\mathrm{int}}(t) + y_{ml}^{\mathrm{artif}}(t) \qquad (55)$$

where $y_{ml}^{\mathrm{spat}}(t)$, $y_{ml}^{\mathrm{int}}(t)$, and $y_{ml}^{\mathrm{artif}}(t)$ are unwanted error components that correspond to spatial (filtering) distortion, interferences, and artifacts, respectively. These are calculated by using a least-squares projection [25]. Then, SDR_l is calculated as the power ratio between the wanted and unwanted components

$$\mathrm{SDR}_l = 10\log_{10}\frac{\sum_{m=1}^{M}\sum_t s_{ml}^{\mathrm{img}}(t)^2}{\sum_{m=1}^{M}\sum_t \left[y_{ml}^{\mathrm{spat}}(t) + y_{ml}^{\mathrm{int}}(t) + y_{ml}^{\mathrm{artif}}(t)\right]^2}.$$

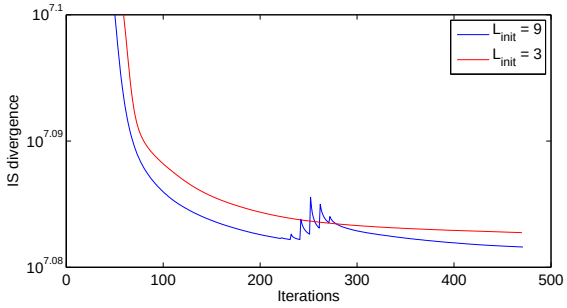

Fig. 10. Convergence behavior of two different clustering strategies. The blue line ($L_{init} = 9$) corresponds to the procedure described in Section V-C. It used both top-down and bottom-up approaches. The zigzag pattern (iterations 220–280) shows that the IS divergence was increased by a pair-wise merge operation. The red line ($L_{init} = 3$) corresponds to using only the top-down approach.

TABLE III
COMPUTATIONAL TIME (IN SECONDS) FOR THE SEPARATION PROCEDURE DESCRIBED IN SECT. V-C AND AN EXISTING METHOD [26].

	ID=1	ID=2	ID=3	ID=4
EU-NMF	201.89	280.43	273.34	231.11
IS-NMF	576.73	867.72	838.30	677.55
UBSS [26]	9.98	13.72	12.60	14.42

Figure 11 shows the source separation results obtained with the procedure described in the last subsection (EU-NMF and IS-NMF). For comparison, results obtained with the Underdetermined Blind Source Separation (UBSS) method [26] are also shown. Four sets of mixtures whose sources are listed in Table I were examined. The source separation result obtained with the NMF-based method depends on the initial values of $\mathbf{T}$, $\mathbf{V}$ and $\mathbf{Z}$. Therefore, we conducted ten trials with different initializations for each set of mixtures. Table III shows the computational time. Sound examples can be found at our web page [27].

From these results, we observe the following. For music recordings with frequent sound patterns, the new NMF-based methods generally performed better than the existing method [26] that relies on the spatial property and simple time-wise activity information of each source. However, the computational burden of the NMF-based methods was heavy. This was because many operations related to matrix inversions and eigenvalue decompositions were involved in the NMF updates. Among the NMF-based methods, IS-NMF produced clearly better separation results than EU-NMF with increased computational effort. This result supports the superiority of IS divergence for audio signal modeling [18] also in a multichannel scenario.

Figure 12 show the effect of having the redundant spatial properties mentioned in the previous subsection. We observe

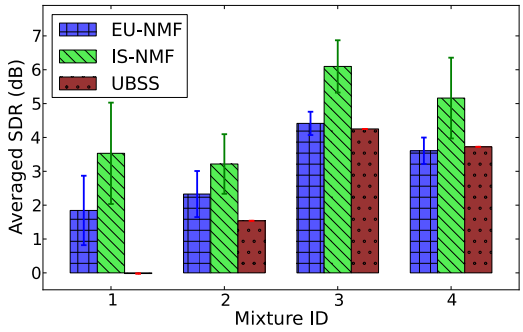

Fig. 11. Source separation performance evaluated in terms of SDRs averaged over all the three sources. With NMF-based methods (EU-NMF and IS-NMF), ten trials were conducted for each mixture ID. The error bars represent one standard deviation. For comparison, the results obtained with an existing underdetermined blind source separation method [26] (UBSS) are also shown.

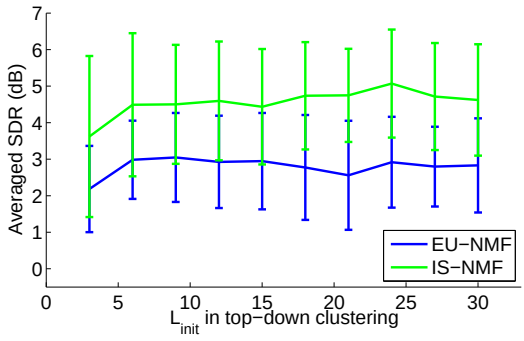

Fig. 12. Source separation performance with various numbers of redundant spatial properties L_{init}. Ten trials were conducted for each mixture ID, and thus the error bars represent one standard deviation over 40 averaged SDRs.

that even a small number of redundant spatial properties contributed to a better separation result than not having the redundancy ($L_{init} = 3$). Having too much redundancy also worked well, but the computational demands also increased. Therefore, we employed $L_{init} = 9$ for the experiments whose results are shown in Fig. 11.

VI. Conclusion

We have presented new formulations and algorithms for multichannel NMF. We started with a simple model (17) where each NMF basis t_{ik}, $i = 1, \ldots, I$, has its own spatial properties H_{ik}. Then, to cluster the NMF bases according to their spatial properties, we introduced the second model (40), which includes cluster-indicator variables $\mathbf{Z}$. Multiplicative update rules were derived to minimize the multichannel IS divergence (16) or multichannel Euclidean distance (20). Experimental results show that the derived multichannel algorithms were as efficient as the standard single-channel algorithms in terms of the number of iterations to converge. Multichannel NMF with IS divergence produced better source separation results than with Euclidean distance for several stereo music mixtures. Future work will include the automatic determination of such model complexity parameters as the number of NMF bases K and sources (spatial properties) L, for example by employing Bayesian nonparametrics [28], [29]. Computationally efficient implementations of the multichannel NMF algorithms, such as one using a general-purpose graphics processing unit (GPGPU), will also constitute future work.

Appendix I
Solving an Algebraic Riccati Equation

To solve (24), we perform an eigenvalue decomposition of a $2M \times 2M$ matrix

$$\begin{bmatrix} 0 & -\mathsf{A} \\ -\mathsf{B} & 0 \end{bmatrix} \tag{56}$$

and let $\mathbf{e}_1, \ldots, \mathbf{e}_M$ be eigenvectors with negative eigenvalues. It is theoretically guaranteed that there are exactly M negative eigenvalues. But in reality there may be computer arithmetic errors. Thus, we actually sort the $2M$ eigenvectors according to the corresponding eigenvalues in ascending order and employ the first M eigenvectors.

Then, let us decompose the $2M$-dimensional M eigenvectors as

$$\mathbf{e}_m = \begin{bmatrix} \mathbf{f}_m \\ \mathbf{g}_m \end{bmatrix} \quad \text{for} \quad m = 1, \ldots, M \tag{57}$$

with $\mathbf{f}_m$ and $\mathbf{g}_m$ being M-dimensional vectors. The new H_{ik} is calculated by

$$\mathsf{H}_{ik} \leftarrow \mathsf{G}\mathsf{F}^{-1} \tag{58}$$

with $\mathsf{F} = [\mathbf{f}_1, \ldots, \mathbf{f}_M]$ and $\mathsf{G} = [\mathbf{g}_1, \ldots, \mathbf{g}_M]$. Again to compensate for computer arithmetic errors, we ensure H_{ik} is Hermitian by $\mathsf{H}_{ik} \leftarrow \frac{1}{2}(\mathsf{H}_{ik} + \mathsf{H}_{ik}^H)$.

Appendix II
Proof for the Auxiliary Function (29)

Let us consider the minimization of f^+ defined in (29) with respect to $\mathbf{R}$ and $\mathbf{U}$ subject to the constraint $\sum_k \mathsf{R}_{ijk} = \mathsf{I}$. By introducing Lagrange multipliers Λ_{ij} of size $M \times M$, we have

$$\mathcal{F} = f^+ + \textstyle\sum_{ij} \mathrm{Re}\left\{ \mathrm{tr}\left[(\sum_k \mathsf{R}_{ijk} - \mathsf{I})^H \Lambda_{ij} \right] \right\} . \tag{59}$$

By setting the partial derivative of $\mathcal{F}$ with respect to R_{ijk}^* at zero

$$\frac{\partial \mathcal{F}}{\partial \mathsf{R}_{ijk}^*} = (t_{ik} v_{kj} \mathsf{H}_{ik})^{-1} \mathsf{R}_{ijk} \mathsf{X}_{ij} + \Lambda_{ij} = 0 , \tag{60}$$

we have $\mathsf{R}_{ijk} = -(t_{ik} v_{kj} \mathsf{H}_{ik}) \Lambda_{ij} \mathsf{X}_{ij}^{-1}$. Adding this for $k = 1, \ldots, K$ gives $\Lambda_{ij} = -\hat{\mathsf{X}}_{ij}^{-1} \mathsf{X}_{ij}$ with the fact that $\sum_k \mathsf{R}_{ijk} = \mathsf{I}$. Therefore, the minimum of the auxiliary function f^+ is obtained when

$$\mathsf{R}_{ijk} = t_{ik} v_{kj} \mathsf{H}_{ik} \hat{\mathsf{X}}_{ij}^{-1} \tag{61}$$

and the minimum value is equal to f defined in (28).

The partial derivative of f^+ with respect to Hermitian matrix U_{ij}^* is given by [30]

$$\frac{\partial f^+}{\partial \mathsf{U}_{ij}^*} = \mathsf{U}_{ij}^{-1} - \frac{\det \hat{\mathsf{X}}_{ij}}{\det \mathsf{U}_{ij}} \mathsf{U}_{ij}^{-1} . \tag{62}$$

Setting this zero gives $\mathsf{U}_{ij} = \hat{\mathsf{X}}_{ij}$.

APPENDIX III
PROOF FOR THE AUXILIARY FUNCTION (35)

Let us consider the minimization of f^+ defined in (35) with respect to $\mathbf{R}$ subject to the constraint $\sum_k \mathsf{R}_{ijk} = \mathsf{I}$. By introducing Lagrange multipliers $\mathsf{\Lambda}_{ij}$ of size $M \times M$, we have

$$\mathcal{F} = f^+ + \textstyle\sum_{ij} \mathrm{Re}\left\{\mathrm{tr}\left[(\sum_k \mathsf{R}_{ijk} - \mathsf{I})^H \mathsf{\Lambda}_{ij}\right]\right\} . \quad (63)$$

The partial derivative of $\mathcal{F}$ with respect to Hermitian matrix R^*_{ijk} is given by [30]

$$\frac{\partial \mathcal{F}}{\partial \mathsf{R}^*_{ijk}} = -t_{ik}^2 v_{kj}^2 \mathsf{R}_{ijk}^{-1} \mathsf{H}_{ik}^H \mathsf{H}_{ik} \mathsf{R}_{ijk}^{-1} + \mathsf{\Lambda}_{ij} . \quad (64)$$

Setting this zero and introducing a matrix U_{ij} we have

$$\mathsf{\Lambda}_{ij} = (t_{ik} v_{kj} \mathsf{R}_{ijk}^{-1} \mathsf{H}_{ik})(t_{ik} v_{kj} \mathsf{R}_{ijk}^{-1} \mathsf{H}_{ik})^H = \mathsf{U}_{ij} \mathsf{U}_{ij}^H . \quad (65)$$

A solution for this equation is given by

$$\mathsf{U}_{ij} = t_{ik} v_{kj} \mathsf{R}_{ijk}^{-1} \mathsf{H}_{ik} \Leftrightarrow \mathsf{R}_{ijk} = t_{ik} v_{kj} \mathsf{H}_{ik} \mathsf{U}_{ij}^{-1} , \quad (66)$$

and adding this for $k = 1, \ldots, K$ gives

$$\mathsf{U}_{ij} = \textstyle\sum_k \mathsf{H}_{ik} t_{ik} v_{kj} = \hat{\mathsf{X}}_{ij} \quad (67)$$

with the fact that $\sum_k \mathsf{R}_{ijk} = \mathsf{I}$. Therefore, the minimum of the auxiliary function f^+ is obtained when

$$\mathsf{R}_{ijk} = t_{ik} v_{kj} \mathsf{H}_{ik} \hat{\mathsf{X}}_{ij}^{-1} \quad (68)$$

and the minimum value is equal to f defined in (34).

APPENDIX IV
EM ALGORITHM FOR COMPARISON

This appendix shows an EM algorithm designed to minimize the total multichannel IS divergence, (18) with (16), according to the NMF model (17). The EM algorithm shown here is a simplification of the EM algorithm shown in [16]. For the STFT coefficient vectors $\tilde{\mathbf{x}}_{ij} \in \mathbb{C}^M$, let $\tilde{\mathbf{y}}_{ijk} \in \mathbb{C}^M$ be latent vectors that satisfy $\tilde{\mathbf{x}}_{ij} = \sum_k \tilde{\mathbf{y}}_{ijk}$.

1) E-step: calculate the expectation of the outer product of $\tilde{\mathbf{y}}_{ijk}$ by

$$\mathbb{E}[\tilde{\mathbf{y}}_{ijk} \tilde{\mathbf{y}}_{ijk}^H] = \hat{\mathsf{Y}}_{ijk} + \hat{\mathsf{Y}}_{ijk} \left(\hat{\mathsf{X}}_{ij}^{-1} \mathsf{X}_{ij} \hat{\mathsf{X}}_{ij}^{-1} - \hat{\mathsf{X}}_{ij}^{-1}\right) \hat{\mathsf{Y}}_{ijk} \quad (69)$$

with

$$\hat{\mathsf{X}}_{ij} = \sum_k \hat{\mathsf{Y}}_{ijk} , \quad \hat{\mathsf{Y}}_{ijk} = \mathsf{H}_{ik} t_{ik} v_{kj} . \quad (70)$$

2) M-step: update the NMF model parameters by

$$t_{ik} \leftarrow \frac{1}{JM} \sum_j \frac{1}{v_{kj}} \mathrm{tr}\left(\mathsf{H}_{ik}^{-1} \mathbb{E}[\tilde{\mathbf{y}}_{ijk} \tilde{\mathbf{y}}_{ijk}^H]\right) , \quad (71)$$

$$v_{kj} \leftarrow \frac{1}{IM} \sum_i \frac{1}{t_{ik}} \mathrm{tr}\left(\mathsf{H}_{ik}^{-1} \mathbb{E}[\tilde{\mathbf{y}}_{ijk} \tilde{\mathbf{y}}_{ijk}^H]\right) , \quad (72)$$

$$\mathsf{H}_{ik} \leftarrow \frac{1}{t_{ik} J} \sum_j \frac{1}{v_{kj}} \mathbb{E}[\tilde{\mathbf{y}}_{ijk} \tilde{\mathbf{y}}_{ijk}^H] . \quad (73)$$

REFERENCES

[1] H. Sawada, H. Kameoka, S. Araki, and N. Ueda, "New formulations and efficient algorithms for multichannel NMF," in *Proc. WASPAA 2011*, Oct. 2011, pp. 153–156.

[2] ——, "Efficient algorithms for multichannel extensions of Itakura-Saito nonnegative matrix factorization," in *Proc. ICASSP 2012*, Mar. 2012, pp. 261–264.

[3] D. D. Lee and H. S. Seung, "Learning the parts of objects with nonnegative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.

[4] W. Xu, X. Liu, and Y. Gong, "Document clustering based on non-negative matrix factorization," in *Proc. ACM SIGIR*, 2003, pp. 267–273.

[5] P. Smaragdis and J. C. Brown, "Non-negative matrix factorization for polyphonic music transcription," in *Proc. WASPAA 2003*, Oct. 2003, pp. 177–180.

[6] T. Virtanen, "Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 15, no. 3, pp. 1066–1074, 2007.

[7] P. Smaragdis, "Convolutive speech bases and their application to supervised speech separation," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 15, pp. 1–12, 2007.

[8] M. Nakano, J. Le Roux, H. Kameoka, Y. Kitano, N. Ono, and S. Sagayama, "Nonnegative matrix factorization with Markov-chained bases for modeling time-varying patterns in music spectrograms," *Latent Variable Analysis and Signal Separation*, pp. 149–156, 2010.

[9] G. Mysore, P. Smaragdis, and B. Raj, "Non-negative hidden Markov modeling of audio with application to source separation," in *Latent Variable Analysis and Signal Separation*. Springer, 2010, pp. 140–148.

[10] D. Fitzgerald, M. Cranitch, and E. Coyle, "Shifted non-negative matrix factorisation for sound source separation," in *IEEE/SP 13th Workshop on Statistical Signal Processing*, 2005, pp. 1132–1137.

[11] R. Jaiswal, D. FitzGerald, D. Barry, E. Coyle, and S. Rickard, "Clustering NMF basis functions using shifted NMF for monaural sound source separation," in *Proc. ICASSP 2011*, May 2011, pp. 245–248.

[12] J. Blauert, *Spatial hearing: the psychophysics of human sound localization*. MIT Press, 1997.

[13] D. FitzGerald, M. Cranitch, and E. Coyle, "Non-negative tensor factorisation for sound source separation," in *Proc. Irish Signals Syst. Conf.*, Sept. 2005, pp. 8–12.

[14] R. M. Parry and I. A. Essa, "Estimating the spatial position of spectral components in audio," in *Proc. ICA 2006*. Springer, Mar. 2006, pp. 666–673.

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

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

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

[18] 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.

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

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

[21] J. de Leeuw, "Block-relaxation methods in statistics," in *Information Systems and Data Analysis*, H. H. Bock, W. Lenski, and M. M. Richter, Eds. Springer Verlag, 1994, pp. 308–324.

[22] A. Marshall, I. Olkin, and B. Arnold, *Inequalities: theory of majorization and its applications*. Springer Verlag, 2010.

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

[24] 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," *Latent Variable Analysis and Signal Separation*, pp. 414–422, 2012.

[25] E. Vincent, H. Sawada, P. Bofill, S. Makino, and J. Rosca, "First stereo audio source separation evaluation campaign: Data, algorithms and results," in *Proc. ICA*, 2007, pp. 552–559. [Online]. Available: http://www.irisa.fr/metiss/SASSEC07/
[26] H. Sawada, S. Araki, and S. Makino, "Underdetermined convolutive blind source separation via frequency bin-wise clustering and permutation alignment," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 19, no. 3, pp. 516–527, 2011.
[27] [Online]. Available: http://www.kecl.ntt.co.jp/icl/signal/sawada/demo/mchnr
[28] M. Hoffman, D. Blei, and P. Cook, "Bayesian nonparametric matrix factorization for recorded music," in *Proc. ICML*, 2010, pp. 641–648.
[29] M. Nakano, J. Roux, H. Kameoka, T. Nakamura, N. Ono, and S. Sagayama, "Bayesian nonparametric spectrogram modeling based on infinite factorial infinite hidden Markov model," in *Proc. WASPAA 2011*, Oct. 2011, pp. 325–328.
[30] A. Hjorungnes, *Complex-valued matrix derivatives*. Cambridge University Press, 2011.

Hiroshi Sawada (M'02–SM'04) received the B.E., M.E. and Ph.D. degrees in information science from Kyoto University, Kyoto, Japan, in 1991, 1993 and 2001, respectively.

He joined NTT Corporation in 1993. He is now the group leader of Learning and Intelligent Systems Research Group at the NTT Communication Science Laboratories, Kyoto, Japan. His research interests include statistical signal processing, audio source separation, array signal processing, machine learning, latent variable model, graph-based data structure, and computer architecture.

From 2006 to 2009, he served as an associate editor of the IEEE Transactions on Audio, Speech & Language Processing. He is a member of the Audio and Acoustic Signal Processing Technical Committee of the IEEE SP Society. He received the 9th TELECOM System Technology Award for Student from the Telecommunications Advancement Foundation in 1994, the Best Paper Award of the IEEE Circuit and System Society in 2000, and the MLSP Data Analysis Competition Award in 2007. Dr. Sawada is a member of the IEICE and the ASJ.

Hirokazu Kameoka (M'07) received B.E., M.E. and Ph.D. degrees all from the University of Tokyo, Japan, in 2002, 2004 and 2007, respectively. He is currently a research scientist at the NTT Communication Science Laboratories and an Adjunct Associate Professor at the University of Tokyo. His research interests include computational auditory scene analysis, statistical signal processing, speech and music processing, and machine learning. He is a member of the IEICE, the IPSJ and the ASJ. He received 13 awards over the past 9 years, including the IEEE Signal Processing Society 2008 SPS Young Author Best Paper Award.

Shoko Araki (M'01–SM'12) is with NTT Communication Science Laboratories, NTT Corporation, Japan. She received the B. E. and the M. E. degrees from the University of Tokyo, Japan, in 1998 and 2000, respectively, and the Ph. D degree from Hokkaido University, Japan in 2007.

Since she joined NTT in 2000, she has been engaged in research on acoustic signal processing, array signal processing, blind source separation (BSS) applied to speech signals, meeting diarization and auditory scene analysis.

She was a member of the organizing committee of the ICA 2003, the finance chair of IWAENC 2003, the co-chair of a special session on undetermined sparse audio source separation in EUSIPCO 2006, the registration chair of WASPAA 2007, and the evaluation co-chair of SiSEC2008, 2010 and 2011. She received the 19th Awaya Prize from Acoustical Society of Japan (ASJ) in 2001, the Best Paper Award of the IWAENC in 2003, the TELECOM System Technology Award from the Telecommunications Advancement Foundation in 2004, the Academic Encouraging Prize from the Institute of Electronics, Information and Communication Engineers (IEICE) in 2006, and the Itakura Prize Innovative Young Researcher Award from (ASJ) in 2008. She is a member of the IEICE and the ASJ.

Naonori Ueda (M'02) received BS, MS, and PhD degrees in communication engineering from Osaka University, Osaka, Japan, in 1982, 1984, and 1992, respectively. In 1984, he joined the NTT Electrical Communication Laboratories, Kanagawa, Japan. In 1991, he joined the NTT Communication Science Laboratories, Kyoto, Japan, as a senior research scientist. His current research interests are statistical machine learning and its applications to pattern recognition, signal processing and data mining. He is the director of the NTT Communication Science Laboratories. He is a Guest Professor of Nara Advanced Institute of Science and Technology (NAIST) and National Institute of Informatics (NII). He is also the sub-project leader of Funding Program for World-Leading Innovative R&D on Science and Technology (First Program), cabinet office, government of Japan, March 2010 - February 2014. From 1993 to 1994, he was a visiting scholar at Purdue University, West Lafayette, Indiana. He is an associate editor of Neurocomputing, and is a member of the IPSJ. He is a fellow of the IEICE.