EFFICIENT ALGORITHMS FOR MULTICHANNEL EXTENSIONS OF ITAKURA-SAITO NONNEGATIVE MATRIX FACTORIZATION

Hiroshi Sawada, Hirokazu Kameoka, Shoko Araki, Naonori Ueda

NTT Communication Science Laboratories, NTT Corporation, Japan

ABSTRACT

This paper proposes new algorithms for multichannel extensions of nonnegative matrix factorization (NMF) with the Itakura-Saito (IS) divergence. We employ Hermitian positive definite matrices for modeling the covariance matrix of a multivariate complex Gaussian distribution. Such matrices are basically estimated for NMF bases, but a source separation task can be performed by introducing variables that relate NMF bases and sources. The new algorithms are derived by using a majorization scheme with properly designed auxiliary functions. The algorithms are in the form of multiplicative updates, and exhibit good convergence behavior. We have succeeded in separating a professionally produced music recording into its vocal and guitar components.

Index Terms— Nonnegative matrix factorization, Itakura-Saito divergence, Auxiliary function, Multichannel, Source separation

1. INTRODUCTION

Identifying frequent spectral patterns in sounds is an important technique for various kinds of audio signal applications, including audio scene analysis and music transcription. Nonnegative matrix factorization (NMF) [1] is widely used for such purposes (e.g., [2]). There are many choices for the distance/divergence measure used in the NMF cost function, such as the Euclidean distance [3], the generalized Kullback-Leibler (KL) divergence [3], and the Itakura-Saito (IS) divergence [4]. It is recognized that the IS divergence is often preferable for modeling audio signals.

On the other hand, multichannel extensions of NMF have been receiving attention with a view to realizing sound source separation with multiple microphones. In [5, 6], the covariance matrices of multivariate complex Gaussian distributions are modeled with the NMF scheme. This can be seen as a multichannel extension of IS-NMF (NMF with the IS divergence). Although expectation-maximization (EM) algorithms have been derived for source separation tasks, it was reported that the algorithms were sensitive to the initialization of parameters. In [7], we proposed a multichannel extension of Euclidean NMF, and derived multiplicative update rules [3] based on majorization [8] with a properly designed auxiliary function. The algorithm converges favorably.

In this paper, we propose new algorithms for multichannel extensions of IS-NMF. The algorithms are derived from the same majorization scheme that we employed for the multichannel extensions of Euclidean NMF [7]. We study two formulations in which a spatial property is considered for each NMF basis (the first one) and for each source (the second one). In the second formulation, a source separation task is performed by automatically clustering the NMF bases for each source. The derived algorithms are in the form of multiplicative updates [3], and they exhibited efficient convergence behaviors in the experiments.

2. NMF WITH IS DIVERGENCE

This section reviews the formulation and algorithm of IS-NMF [4, 9]. Let us assume that we have a single channel observation, and we apply a short-time Fourier transform (STFT) to the observation. Let $x_{ij} \in \mathbb{C}$ be the STFT coefficient at frequency i and time j. The generative model of all the STFT coefficients $\mathbf{X}$, $[\mathbf{X}]_{ij} = x_{ij}$, of size $I \times J$ can be written as

$$p(\mathbf{X}|\mathbf{T},\mathbf{V}) = \prod_{i=1}^{I} \prod_{j=1}^{J} \mathcal{N}(x_{ij}|0, \textstyle\sum_{k=1}^{K} t_{ik}v_{kj}) \qquad (1)$$

where $\mathcal{N}$ represents a complex Gaussian distribution, and K is the number of rank-1 basis matrices. Nonnegative matrices $\mathbf{T}$ and $\mathbf{V}$, whose elements are $t_{ik} \geq 0$ and $v_{kj} \geq 0$, have sizes of $I \times K$ and $K \times J$, respectively.

The negative log likelihood $-\log p(\mathbf{X}|\mathbf{T},\mathbf{V})$ is given by

$$\mathcal{L}(\mathbf{T},\mathbf{V}) = \sum_{i,j} \left(\frac{|x_{ij}|^2}{\sum_k t_{ik}v_{kj}} + \log \sum_k t_{ik}v_{kj} \right) \qquad (2)$$

where constant terms are omitted. Minimizing $\mathcal{L}(\mathbf{T},\mathbf{V})$ is equivalent to minimizing the IS divergence

$$d_{IS}(|x_{ij}|^2, \textstyle\sum_{k=1}^{K} t_{ik}v_{kj}) = \frac{|x_{ij}|^2}{\sum_k t_{ik}v_{kj}} - \log \frac{|x_{ij}|^2}{\sum_k t_{ik}v_{kj}} - 1 \quad (3)$$

for all the STFT coefficients $\mathbf{X}$.

The negative log likelihood $\mathcal{L}(\mathbf{T},\mathbf{V})$ can be minimized in an iterative manner by majorization [8] with an auxiliary function

$$\mathcal{L}^+(\mathbf{T},\mathbf{V},\mathbf{R},\mathbf{U}) = \sum_{i,j} \left(|x_{ij}|^2 \sum_k \frac{r_{ijk}^2}{t_{ik}v_{kj}} + \log u_{ij} + \frac{\sum_k t_{ik}v_{kj} - u_{ij}}{u_{ij}} \right) \qquad (4)$$

where r_{ijk} and u_{ij} are auxiliary variables that satisfy $\sum_k r_{ijk} = 1$, $r_{ijk} > 0$ and $u_{ij} > 0$. It can be verified [10, 9] that the auxiliary function $\mathcal{L}^+$ cannot be smaller than the original function $\mathcal{L}(\mathbf{T},\mathbf{V}) \leq \mathcal{L}^+(\mathbf{T},\mathbf{V},\mathbf{R},\mathbf{U})$ and the equality $\mathcal{L} = \mathcal{L}^+$ is satisfied when

$$r_{ijk} = \frac{t_{ik}v_{kj}}{\sum_k t_{ik}v_{kj}}\,, \quad u_{ij} = \sum_k t_{ik}v_{kj}\,. \qquad (5)$$

Therefore, the negative log likelihood $\mathcal{L}$ is indirectly minimized by repeating the following two steps:

1. Minimizing $\mathcal{L}^+$ with respect to $\mathbf{R}$ and $\mathbf{U}$ by (5), which makes $\mathcal{L}(\mathbf{T},\mathbf{V}) = \mathcal{L}^+(\mathbf{T},\mathbf{V},\mathbf{R},\mathbf{U})$.
2. Minimizing $\mathcal{L}^+$ with respect to $\mathbf{T}$ or $\mathbf{V}$, which also minimizes $\mathcal{L}$. The update rules are derived from the partial derivative of $\mathcal{L}^+$ with respect to the corresponding variables as

$$t_{ik} \leftarrow \sqrt{\frac{\sum_j \frac{r_{ijk}^2 |x_{ij}|^2}{v_{kj}}}{\sum_j \frac{v_{kj}}{u_{ij}}}}\,, \quad v_{kj} \leftarrow \sqrt{\frac{\sum_i \frac{r_{ijk}^2 |x_{ij}|^2}{t_{ik}}}{\sum_i \frac{t_{ik}}{u_{ij}}}}\,. \quad (6)$$

By substituting (5) into (6), we obtain the update rules for IS-NMF:

$$t_{ik} \leftarrow t_{ik}\sqrt{\frac{\sum_j \frac{v_{kj}}{\hat{x}_{ij}}\frac{|x_{ij}|^2}{\hat{x}_{ij}}}{\sum_j \frac{v_{kj}}{\hat{x}_{ij}}}}\,, \quad v_{kj} \leftarrow v_{kj}\sqrt{\frac{\sum_i \frac{t_{ik}}{\hat{x}_{ij}}\frac{|x_{ij}|^2}{\hat{x}_{ij}}}{\sum_i \frac{t_{ik}}{\hat{x}_{ij}}}} \quad (7)$$

with $\hat{x}_{ij} = \sum_k t_{ik}v_{kj}$.

3. PROPOSED MULTICHANNEL EXTENSIONS

This section presents our proposed extensions of IS-NMF for a multichannel case, where we have a complex-valued vector $\mathbf{x} = [x_1, \ldots, x_M]^T \in \mathbb{C}^M$ for a time-frequency slot, with x_m being the m-th microphone observation. Let $\mathbf{x}_{ij}$ be such a vector at frequency i and time j. Then, the generative model of all the STFT coefficients $\mathbf{X}$, $[\mathbf{X}]_{ij} = \mathbf{x}_{ij} \in \mathbb{C}^M$, can be written as

$$p(\mathbf{X}|\theta) = \prod_{i=1}^{I}\prod_{j=1}^{J}\mathcal{N}(\mathbf{x}_{ij}|0, \hat{\mathsf{X}}_{ij}) \quad (8)$$

where $\mathcal{N}$ represents a multivariate complex Gaussian distribution, and $\hat{\mathsf{X}}_{ij}$ is an $M \times M$ Hermitian positive definite covariance matrix that models the multichannel observation at frequency i and time j. We will detail the parameter set θ and the covariance matrix $\hat{\mathsf{X}}_{ij}$ in the next two subsections.

3.1. NMF basis-wise spatial property H_{ik}

We start with a simple case where each NMF basis has its own spatial property in this multichannel scenario. Let H_{ik} be an $M \times M$ Hermitian positive definite matrix that models the spatial property of the k-th NMF basis at frequency i. Then, we detail the covariance matrix as

$$\hat{\mathsf{X}}_{ij} = \sum_{k=1}^{K}\mathsf{H}_{ik}t_{ik}v_{kj}\,, \quad (9)$$

where t_{ik} and v_{kj} are nonnegative scalars that constitute NMF bases $\mathbf{T}$ and $\mathbf{V}$. To fix the scaling ambiguity between H_{ik} and t_{ik}, a constraint $||\mathsf{H}_{ik}||_F = 1$ is introduced.

The negative log likelihood $-\log p(\mathbf{X}|\theta)$ with $\theta = \{\mathbf{T}, \mathbf{V}, \mathbf{H}\}$ is given by (omitting constant terms)

$$\mathcal{L}(\mathbf{T}, \mathbf{V}, \mathbf{H}) = \sum_{i,j}\left[\mathrm{tr}(\mathsf{X}_{ij}\hat{\mathsf{X}}_{ij}^{-1}) + \log\det\hat{\mathsf{X}}_{ij}\right] \quad (10)$$

with $\mathsf{X}_{ij} = \mathbf{x}_{ij}\mathbf{x}_{ij}^H$ and $\mathrm{tr}(\cdot)$ being the trace of a matrix. This negative log-likelihood can be seen as a multichannel extension of (2). Therefore, we consider the minimization of $\mathcal{L}(\mathbf{T}, \mathbf{V}, \mathbf{H})$ as a multichannel extension of IS-NMF.

Following the optimization framework explained in Sect. 2, we introduce an auxiliary function

$$\mathcal{L}^+(\mathbf{T}, \mathbf{V}, \mathbf{H}, \mathbf{R}, \mathbf{U}) = \sum_{i,j} \quad (11)$$

$$\left[\sum_k \frac{\mathrm{tr}(\mathsf{X}_{ij}\mathsf{R}_{ijk}\mathsf{H}_{ik}^{-1}\mathsf{R}_{ijk})}{t_{ik}v_{kj}} + \log\det\mathsf{U}_{ij} + \frac{\det\hat{\mathsf{X}}_{ij} - \det\mathsf{U}_{ij}}{\det\mathsf{U}_{ij}}\right]$$

where R_{ijk} and U_{ij} are auxiliary variables that satisfy Hermitian positive definiteness and $\sum_k \mathsf{R}_{ijk} = \mathsf{I}$, with I being the identity matrix of size M. As shown in Appendix 6.1, the auxiliary function $\mathcal{L}^+$ cannot be smaller than the original function $\mathcal{L}(\mathbf{T}, \mathbf{V}, \mathbf{H}) \le \mathcal{L}^+(\mathbf{T}, \mathbf{V}, \mathbf{H}, \mathbf{R}, \mathbf{U})$ and the equality $\mathcal{L} = \mathcal{L}^+$ is satisfied when

$$\mathsf{R}_{ijk} = t_{ik}v_{kj}\mathsf{H}_{ik}\hat{\mathsf{X}}_{ij}^{-1}\,, \quad \mathsf{U}_{ij} = \hat{\mathsf{X}}_{ij}\,. \quad (12)$$

The minimization updates with respect to the main variables $\mathbf{T}$ and $\mathbf{V}$ are derived from the partial derivatives of $\mathcal{L}^+$ (see Appendix 6.2), and given by

$$t_{ik} \leftarrow t_{ik}\sqrt{\frac{\sum_j v_{kj}\mathrm{tr}(\hat{\mathsf{X}}_{ij}^{-1}\mathsf{X}_{ij}\hat{\mathsf{X}}_{ij}^{-1}\mathsf{H}_{ik})}{\sum_j v_{kj}\mathrm{tr}(\hat{\mathsf{X}}_{ij}^{-1}\mathsf{H}_{ik})}} \quad (13)$$

$$v_{kj} \leftarrow v_{kj}\sqrt{\frac{\sum_i t_{ik}\mathrm{tr}(\hat{\mathsf{X}}_{ij}^{-1}\mathsf{X}_{ij}\hat{\mathsf{X}}_{ij}^{-1}\mathsf{H}_{ik})}{\sum_i t_{ik}\mathrm{tr}(\hat{\mathsf{X}}_{ij}^{-1}\mathsf{H}_{ik})}} \quad (14)$$

We observe that these update rules reduce to (7) if $M = 1$, $\mathsf{X}_{ij} = |x_{ij}|^2$ and $\mathsf{H}_{ik} = 1$. Thus, we understand that the set of these updates constitutes a multichannel extension of IS-NMF.

The minimization update with respect to H_{ik} is performed by solving an algebraic Riccati equation as shown in Appendix 6.2. The procedure is as follows. First, two matrices

$$\mathsf{A} = \sum_j v_{kj}\hat{\mathsf{X}}_{ij}^{-1}\,, \quad (15)$$

$$\mathsf{B} = \mathsf{H}_{ik}\left(\sum_j v_{kj}\hat{\mathsf{X}}_{ij}^{-1}\mathsf{X}_{ij}\hat{\mathsf{X}}_{ij}^{-1}\right)\mathsf{H}_{ik} \quad (16)$$

are calculated. Then, we perform an eigenvalue decomposition of a $2M \times 2M$ matrix

$$\begin{bmatrix} 0 & -\mathsf{A} \\ -\mathsf{B} & 0 \end{bmatrix} \quad (17)$$

and let $\mathbf{e}_1, \ldots, \mathbf{e}_M$ be eigenvectors with negative eigenvalues. It is guaranteed that there are exactly M negative eigenvalues. Then, let us decompose the $2M$-dimensional eigenvectors as

$$\mathbf{e}_m = \begin{bmatrix} \mathbf{f}_m \\ \mathbf{g}_m \end{bmatrix} \quad \text{for} \quad m = 1, \ldots, M \quad (18)$$

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

$$\mathsf{H}_{ik} \leftarrow \mathsf{G}\mathsf{F}^{-1} \quad (19)$$

with $\mathsf{F} = [\mathbf{f}_1, \ldots, \mathbf{f}_M]$ and $\mathsf{G} = [\mathbf{g}_1, \ldots, \mathbf{g}_M]$.

In summary, the negative log-likelihood (10) with (9) is iteratively minimized by the updates (13), (14) and (19).

3.2. Source-wise spatial property H_{io}

When multichannel NMF is applied to a source separation task, NMF bases originating from the same source should be clustered together. To realize such functionality, this subsection proposes another model for the covariance matrix

$$\hat{\mathsf{X}}_{ij} = \sum_{k=1}^{K}\sum_{o=1}^{N} z_{ko}\mathsf{H}_{io}t_{ik}v_{kj} \quad (20)$$

by extending (9). In this model, spatial property matrices H_{io} depend on frequency i and source o, where $o = 1, \ldots, N$ with N being the number of sources. To relate NMF bases and sources, we introduce variables z_{ko} that satisfy $\sum_{o=1}^{N} z_{ko} = 1$ and $z_{ko} \ge 0$. It is interpreted that the k-th NMF basis belongs to the o-th source if z_{ko} is close to 1.

The negative log-likelihood $\mathcal{L}(\mathbf{T}, \mathbf{V}, \mathbf{H}, \mathbf{Z})$, $-\log p(\mathbf{X}|\theta)$ with $\theta = \{\mathbf{T}, \mathbf{V}, \mathbf{H}, \mathbf{Z}\}$, in this model is literally the same as (10) but the definition of $\hat{\mathsf{X}}_{ij}$ is different and should follow (20). $\mathcal{L}(\mathbf{T}, \mathbf{V}, \mathbf{H}, \mathbf{Z})$ can be minimized in a similar manner to that shown in the previous subsection.

We introduce an auxiliary function

$$\mathcal{L}^{+}(\mathbf{T},\mathbf{V},\mathbf{H},\mathbf{Z},\mathbf{R},\mathbf{U}) = \sum_{i,j} \tag{21}$$

$$\left[\sum_{k,o}\frac{\mathrm{tr}(\mathsf{X}_{ij}\mathsf{R}_{ijko}\mathsf{H}_{io}^{-1}\mathsf{R}_{ijko})}{z_{ko}t_{ik}v_{kj}} + \log\det\mathsf{U}_{ij} + \frac{\det\hat{\mathsf{X}}_{ij}-\det\mathsf{U}_{ij}}{\det\mathsf{U}_{ij}}\right]$$

where R_{ijko} and U_{ij} are auxiliary variables that satisfy Hermitian positive definiteness and $\sum_{k,o}\mathsf{R}_{ijko} = \mathsf{I}$. It is verified that $\mathcal{L}(\mathbf{T},\mathbf{V},\mathbf{H},\mathbf{Z}) \le \mathcal{L}^{+}(\mathbf{T},\mathbf{V},\mathbf{H},\mathbf{Z},\mathbf{R},\mathbf{U})$ and the equality is satisfied when

$$\mathsf{R}_{ijko} = z_{ko}t_{ik}v_{kj}\mathsf{H}_{io}\hat{\mathsf{X}}_{ij}^{-1}\,, \quad \mathsf{U}_{ij} = \hat{\mathsf{X}}_{ij}\,. \tag{22}$$

The minimization updates with respect to the main variables $\mathbf{T}$, $\mathbf{V}$ and $\mathbf{Z}$ are given by

$$t_{ik} \leftarrow t_{ik}\sqrt{\frac{\sum_j v_{kj}\sum_o z_{ko}\mathrm{tr}(\hat{\mathsf{X}}_{ij}^{-1}\mathsf{X}_{ij}\hat{\mathsf{X}}_{ij}^{-1}\mathsf{H}_{io})}{\sum_j v_{kj}\sum_o z_{ko}\mathrm{tr}(\hat{\mathsf{X}}_{ij}^{-1}\mathsf{H}_{io})}}\,, \tag{23}$$

$$v_{kj} \leftarrow v_{kj}\sqrt{\frac{\sum_i t_{ik}\sum_o z_{ko}\mathrm{tr}(\hat{\mathsf{X}}_{ij}^{-1}\mathsf{X}_{ij}\hat{\mathsf{X}}_{ij}^{-1}\mathsf{H}_{io})}{\sum_i t_{ik}\sum_o z_{ko}\mathrm{tr}(\hat{\mathsf{X}}_{ij}^{-1}\mathsf{H}_{io})}}\,, \tag{24}$$

$$z_{ko} \leftarrow z_{ko}\sqrt{\frac{\sum_{i,j} t_{ik}v_{kj}\mathrm{tr}(\hat{\mathsf{X}}_{ij}^{-1}\mathsf{X}_{ij}\hat{\mathsf{X}}_{ij}^{-1}\mathsf{H}_{io})}{\sum_{i,j} t_{ik}v_{kj}\mathrm{tr}(\hat{\mathsf{X}}_{ij}^{-1}\mathsf{H}_{io})}}\,. \tag{25}$$

Regarding z_{ko}, a normalization $z_{ko} \leftarrow z_{ko}\,/\,(\sum_o z_{ko})$ should be conducted after the updates to satisfy the constraint $\sum_{o=1}^{N} z_{ko} = 1$. Again, we observe that these updates reduce to (7) if $N = 1$, $M = 1$, $\mathsf{X}_{ij} = |x_{ij}|^2$ and $\mathsf{H}_{ik} = 1$. Thus, we understand that the set of these updates constitutes another multichannel extension of IS-NMF.

The minimization procedure for H_{io} is basically the same as that shown in the previous subsection, except that the calculation of the A, B matrices should be performed as follows

$$\mathsf{A} = \textstyle\sum_{j,k} z_{ko}t_{ik}v_{kj}\hat{\mathsf{X}}_{ij}^{-1}\,, \tag{26}$$

$$\mathsf{B} = \mathsf{H}_{io}\left(\textstyle\sum_{j,k} z_{ko}t_{ik}v_{kj}\hat{\mathsf{X}}_{ij}^{-1}\mathsf{X}_{ij}\hat{\mathsf{X}}_{ij}^{-1}\right)\mathsf{H}_{io}\,. \tag{27}$$

4. EXPERIMENTS

We used a sound mixture available at the Signal Separation Evaluation Campaign (SiSEC 2010) [11]. More specifically, we used `dev1__tamy-que_pena_tanto_faz__snip_6_19__mix.wav` found in the development data of the Professionally produced music recording dataset. The sound file has 2 channels (microphones) and is a mixture of 2 sources (vocal and guitar). We down-sampled the mixture from 44.1 to 16 kHz, and applied STFT (with a 64 ms frame size and a 16 ms frame shift) to the mixture.

Figure 1 shows the convergence behavior when we applied NMF algorithms to the mixture. For comparison, we ran four different algorithms: 1) single-channel NMF described in Sect. 2 applied to the first microphone observation, 2) multichannel NMF with a basis-wise spatial property proposed in Sect. 3.1 (Basis-wise H), 3) the same multichannel NMF with a basis-wise spatial property but optimized by the EM algorithm proposed in [6] (EM algorithm), and 4) multichannel NMF with a source-wise spatial property proposed in Sect. 3.2 (Source-wise H). The number K of rank-1 basis matrices was set at 10 in all cases. For each algorithm, we performed

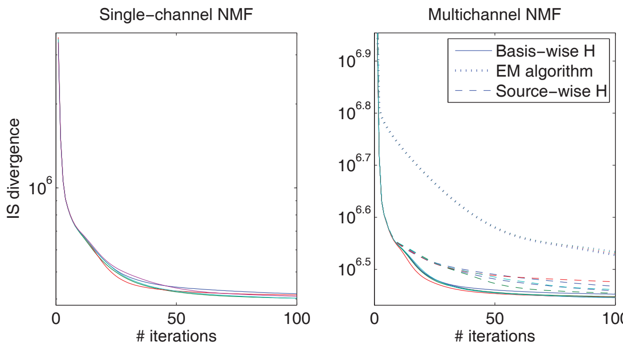

Fig. 1. Convergence behavior

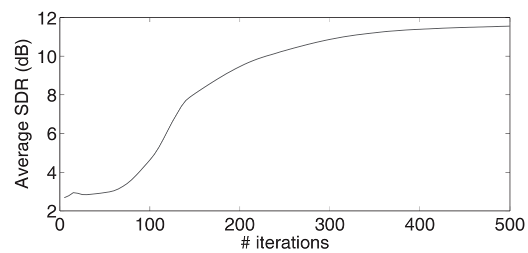

Fig. 2. Source separation performance

five runs starting from different random initial matrices $\mathbf{T}$, $\mathbf{V}$, and $\mathbf{Z}$. The diagonal elements of $\mathbf{H}$ were initially all set to $1/\sqrt{M}$, with M being the number of microphones, and the off-diagonal elements were initially all set to zero.

From the result shown in Fig. 1, we observe the followings. The convergence behavior of Basis-wise H multichannel NMF was favorable, and appears similar to that of the single-channel IS-NMF. With Source-wise H, the convergence was slowed slightly by the constraints caused by sharing the spatial property H_{io} among some NMF bases, but its behavior was still satisfactory. If we compare the convergence of Basis-wise H and EM algorithm, the effectiveness of the proposed algorithms over the conventional EM algorithms becomes clear. The algorithms were coded with Matlab and run on an Intel Core i7 965 (3.2 GHz) processor. The average computation time for 100 iterations by Basis-wise H, EM algorithm and Source-wise H were about 138, 267 and 89 seconds, respectively.

Figure 2 shows the source separation performance in terms of the average signal-to-distortion ratio (SDR) [11]. The separation was performed by the multichannel NMF (source-wise H_{io}) proposed in Sect. 3.2. We obtained the separated signals by Wiener filtering

$$\mathbf{y}_{ijo} = \left(\textstyle\sum_{k=1}^{K} z_{ko}t_{ik}v_{kj}\right)\mathsf{H}_{io}\hat{\mathsf{X}}_{ij}^{-1}\mathbf{x}_{ij} \tag{28}$$

for each source o and every time-frequency j,i slot. A fairly high SDR of around 11 dB was attained with the proposed multichannel NMF, starting from randomly initialized parameters.

Figure 3 shows the behavior of variables z_{ko} in the source separation task. It can be seen that randomly initialized variables z_{ko} approached 1 or 0 as the algorithm converged. From the final status of z_{ko}, we understood that six NMF bases constituted the first output (vocal) and the remaining four NMF bases mainly constituted the second output (guitar).

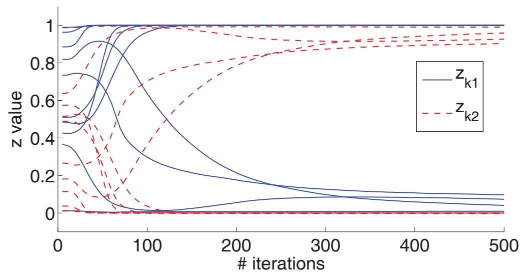

Fig. 3. Behavior of variables z_{ko}, $k = 1, \ldots, 10$, $o = 1, 2$

5. CONCLUSION

To extend IS-NMF to a multichannel case, we have designed an auxiliary function (11) and derived efficient multiplicative update rules (13), (14) and (19). We then proposed a second covariance model (20) by introducing variables z_{ko} that relate NMF bases $t_{ik}v_{kj}$ and a source specific spatial property H_{io}. The model forces the learned NMF bases to cluster into the sources, and contributes to solving a source separation task. Experimental results show that the proposed algorithms converged favorably, and source separation was accomplished with a high SDR value. Future work will include intensive experiments with a variety of sound sources.

6. APPENDIX

6.1. Proof for the auxiliary function condition of (11)

Let us consider the minimization of $\mathcal{L}^+$ defined in (11) with respect to $\mathbf{R}$ subject to the constraint $\sum_k \mathsf{R}_{ijk} = \mathsf{I}$, and also with respect to $\mathbf{U}$. By introducing Lagrange multipliers Λ_{ij} of size $M \times M$, we have

$$\mathcal{F} = \mathcal{L}^+ + \sum_{ij} \operatorname{tr}\left[(\textstyle\sum_k \mathsf{R}_{ijk} - \mathsf{I})^H \Lambda_{ij}\right] . \tag{29}$$

The partial derivative of $\mathcal{F}$ with respect to R^*_{ijk} is given as

$$\frac{\partial \mathcal{F}}{\partial \mathsf{R}^*_{ijk}} = 2(\mathsf{H}_{ik} t_{ik} v_{kj})^{-1} \mathsf{R}_{ijk} \mathsf{X}_{ij} + \Lambda_{ij} . \tag{30}$$

Setting this to zero gives $2\mathsf{R}_{ijk} = -t_{ik}v_{kj}\mathsf{H}_{ik}\Lambda_{ij}\mathsf{X}_{ij}^{-1}$ and adding these for $k = 1, \ldots, K$ gives $\Lambda_{ij} = -2\hat{\mathsf{X}}_{ij}^{-1}\mathsf{X}_{ij}$. By eliminating Λ_{ij}, we have $\mathsf{R}_{ijk} = t_{ik}v_{kj}\mathsf{H}_{ik}\hat{\mathsf{X}}_{ij}^{-1}$.

The partial derivative of $\mathcal{L}^+$ with respect to U^*_{ij} is given as

$$\frac{\partial \mathcal{L}^+}{\partial \mathsf{U}^*_{ij}} = \mathsf{U}_{ij}^{-1} - \frac{\det \hat{\mathsf{X}}_{ij}}{\det \mathsf{U}_{ij}} \mathsf{U}_{ij}^{-1} . \tag{31}$$

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

In summary, the minimum $\mathcal{L}^+$ is obtained by (12) and the minimum value is equal to $\mathcal{L}$ defined in (10).

6.2. Derivation of update rules

The partial derivatives of $\mathcal{L}^+$ defined in (11) with respect to t_{ik} and v_{kj} are

$$\frac{\partial \mathcal{L}^+}{\partial t_{ik}} = \sum_j \left[\frac{\det \hat{\mathsf{X}}_{ij}}{\det \mathsf{U}_{ij}} \operatorname{tr}(\hat{\mathsf{X}}_{ij}^{-1}\mathsf{H}_{ik}) v_{kj} - \frac{\operatorname{tr}(\mathsf{R}_{ijk}\mathsf{H}_{ik}^{-1}\mathsf{R}_{ijk}\mathsf{X}_{ij})}{t_{ik}^2 v_{kj}}\right] , \tag{32}$$

$$\frac{\partial \mathcal{L}^+}{\partial v_{kj}} = \sum_i \left[\frac{\det \hat{\mathsf{X}}_{ij}}{\det \mathsf{U}_{ij}} \operatorname{tr}(\hat{\mathsf{X}}_{ij}^{-1}\mathsf{H}_{ik}) t_{ik} - \frac{\operatorname{tr}(\mathsf{R}_{ijk}\mathsf{H}_{ik}^{-1}\mathsf{R}_{ijk}\mathsf{X}_{ij})}{t_{ik} v_{kj}^2}\right] . \tag{33}$$

Setting these derivatives to zero and solving them with respect to the corresponding variables yields update rules that are similar to (6). Then, by substituting R_{ijk} and U_{ij} with (12), we have the updates (13) and (14).

The partial derivative with respect to H_{ik} is

$$\frac{\partial \mathcal{L}^+}{\partial \mathsf{H}_{ik}} = \sum_j \left[\frac{\det \hat{\mathsf{X}}_{ij}}{\det \mathsf{U}_{ij}} \hat{\mathsf{X}}_{ij}^{-1} t_{ik} v_{kj} - \frac{\mathsf{H}_{ik}^{-1}\mathsf{R}_{ijk}\mathsf{X}_{ij}\mathsf{R}_{ijk}\mathsf{H}_{ik}^{-1}}{t_{ik} v_{kj}}\right] . \tag{34}$$

Setting the derivative to zero, solving it with respect to H_{ik}, and substituting R_{ijk} and U_{ij} with (12) yields

$$\mathsf{H}_{ik}\mathsf{A}\mathsf{H}_{ik} = \mathsf{B} \tag{35}$$

where matrices A and B are defined in (15) and (16). Equation (35) is a special case of an algebraic Riccati equation, and it can be solved by using the procedure described in the main text.

7. REFERENCES

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

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

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

[4] 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.

[5] 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.

[6] S. Arberet, A. Ozerov, N.Q.K. 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.

[7] 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.

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

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

[10] H. Kameoka, M. Goto, and S. Sagayama, "Selective amplifier of periodic and non-periodic components in concurrent audio signals with spectral control envelopes," in *IPSJ SIG Technical Reports, 2006-MUS-66-13*, Aug. 2006, pp. 77–84, in Japanese.

[11] S. Araki, A. Ozerov, V. Gowreesunker, H. Sawada, F. Theis, G. Nolte, D. Lutter, and N. Duong, "The 2010 signal separation evaluation campaign (SiSEC2010): Audio source separation," in *Latent Variable Analysis and Signal Separation*, vol. 6365 of *Lecture Notes in Computer Science*, pp. 114–122. 2010.