COMPLEX NMF WITH THE GENERALIZED KULLBACK-LEIBLER DIVERGENCE

Hirokazu Kameoka†*, Hideaki Kagami*‡*, Masahiro Yukawa*‡

† NTT Communication Science Laboratories, NTT Corporation, Japan.
‡ Dept. Electronics and Electrical Engineering, Keio University, Japan.

ABSTRACT

We previously introduced a phase-aware variant of the non-negative matrix factorization (NMF) approach for audio source separation, which we call the "Complex NMF (CNMF)." This approach makes it possible to realize NMF-like signal decompositions in the complex time-frequency domain. One limitation of the CNMF framework is that the divergence measure is limited to only the Euclidean distance. Some previous studies have revealed that for source separation tasks with NMF, the generalized Kullback-Leibler (KL) divergence tends to yield higher accuracy than when using other divergence measures. This motivated us to believe that CNMF could achieve even greater source separation accuracy if we could derive an algorithm for a KL divergence counterpart of CNMF. In this paper, we start by defining the notion of the "dual" form of the CNMF formulation, derived from the original Euclidean CNMF, and show that a KL divergence counterpart of CNMF can be developed based on this dual formulation. We call this "KL-CNMF". We further derive a convergence-guaranteed iterative algorithm for KL-CNMF based on a majorization-minimization scheme. The source separation experiments revealed that the proposed KL-CNMF yielded higher accuracy than the Euclidean CNMF and NMF with varying divergences.

Index Terms— Audio source separation, non-negative matrix factorization (NMF), Complex NMF, generalized Kullback-Leibler (KL) divergence

1. INTRODUCTION

Audio source separation has long been a challenging task in the field of audio signal processing. A deep neural network-based approach has recently proved powerful for supervised audio source separation tasks where the mixture consists of speech and noise [1]. Furthermore, a recently proposed approach called "deep clustering" [2] has made it possible to deal with "cocktail party" scenarios where the interference is also speech. Although these methods have been shown to work well when a large number of training samples are available, the non-negative matrix factorization (NMF) approach [3, 4] still remains attractive for audio source separation tasks particularly where only a limited amount of training data is available or when prior knowledge about the underlying sources is limited. In addition, since NMF is a generative approach, it can be convenient in semi-supervised scenarios.

With the NMF approach, the magnitude (or power) spectrogram of a mixture signal, interpreted as a non-negative matrix $\boldsymbol{Y}$, is modeled as the product of two non-negative matrices $\boldsymbol{H}$ and $\boldsymbol{U}$. This can be interpreted as approximating the observed spectrum at each time frame as a linear sum of basis spectra scaled by time-varying amplitudes, and amounts to approximating the observed spectrogram as the sum of rank-1 spectrograms. In a supervised/semi-supervised setting, NMF is first employed to train the basis spectra of each sound source using individually recorded audio samples. At test time, NMF is applied to the spectrogram of a test mixture signal, where the subsets of the basis spectra are fixed at the pretrained spectra.

Although the NMF approach has been shown to be successful, one drawback is that it assumes the additivity of magnitude (or power) spectra, which holds only approximately, and does not take account of phase information. To address this drawback, we have previously proposed a framework called the "Complex NMF (CNMF)" [5], where the complex spectrum observed at each time frame is modeled as the sum of components, each of which is described by the multiplication of a static basis spectrum, a time-varying amplitude and a time-varying phase spectrum. With a similar motivation, Parry and Essa [6] and Fevotte *et al*. [7] proposed a generative model of the complex spectrogram obtained with the short-time Fourier transform (STFT) of a mixture signal, where the power spectrogram of each component is modeled as a rank-1 matrix whereas the phase spectrogram is treated as uniformly distributed latent variables. It can be shown that when we assume that each element of the complex spectrogram independently follows a zero-mean complex normal distribution, the maximum likelihood estimation of the model parameters amounts to fitting the NMF model to an observed power spectrogram using the Itakura-Saito (IS) divergence as a goodness-of-fit criterion. This approach is called IS-NMF. A similar kind of generative model using a complex Cauchy distribution instead of a complex normal distribution has also been proposed [8]. IS-NMF and Cauchy-NMF treat the phase spectrogram of each underlying component as a latent variable to be marginalized out and the aim is to find the expectation of the complex spectrogram of each component taken over all possible phase spectrograms of all the components. Although this estimator is reasonable if the phase spectrograms are really stochastic, they are in fact deterministic and unique. Indeed, the phase part of this estimator is always given as that of the observed mixture signal, which certainly differs from the true value. By contrast, CNMF allows us to find the *jointly optimal* estimates of the power and phase spectrograms of all the components. Thus, we can expect CNMF to lead to higher source separation accuracy than the other NMF variants if the parameters can be properly estimated.

However, one limitation with the CNMF framework is that the divergence measure used to measure the difference between an observed spectrogram and the model is limited to only the Euclidean distance (squared error) due to the fact that the arguments are complex numbers unlike NMF. This is in contrast to the conventional NMF framework where effi-

This work was supported by JSPS KAKENHI Grant Numbers 26730100, 15K06081, 15K13986, 15H02757.

cient algorithms have been proposed for different divergence measures [7–14]. Some previous studies have revealed that for source separation tasks with NMF, using the generalized Kullback-Leibler (KL) divergence tends to yield higher accuracy than using other divergence measures [14] (though consensus has yet to be reached [15]). This motivated us to expect that we can achieve even higher source separation accuracy if we can derive an algorithm for a KL divergence counterpart of CNMF. In this paper, we start by defining the notion of the "dual" form of the CNMF formulation, derived from the auxiliary function of the original Euclidean CNMF, and show that a KL divergence counterpart of CNMF can be developed based on this dual formulation. We call this "KL-CNMF". We further derive a convergence-guaranteed algorithm for KL-CNMF based on a majorization-minimization scheme.

2. EUCLIDEAN COMPLEX NMF

Let $Y_{k,m} \in \mathbb{C}$ denote the complex spectrogram of an observed mixture signal, where k and m denote the frequency and time (frame) indices, respectively. While the NMF approach approximates the magnitude (or power) spectrum $|Y_{k,m}|$ at each frame m as a linear sum of static basis spectra $H_{k,1}, \ldots, H_{k,L}$ scaled by time-varying amplitudes $U_{1,m}, \ldots, U_{L,m}$

$$|Y_{k,m}| \simeq \sum_{l=1}^{L} H_{k,l} U_{l,m}, \tag{1}$$

CNMF approximates the complex spectrum $Y_{k,m}$ at each frame m as a linear sum of static basis spectra $H_{k,1}, \ldots, H_{k,L}$ scaled by time-varying amplitudes $U_{1,m}, \ldots, U_{L,m}$ and multiplied by time-varying phase spectra $e^{j\phi_{1,k,m}}, \ldots, e^{j\phi_{L,k,m}}$

$$Y_{k,m} \simeq \sum_{l} H_{k,l} U_{l,m} e^{j\phi_{l,k,m}}. \tag{2}$$

Here, it is important to emphasize that $\phi_{l,k,m}$ is indexed by m, meaning that this model allows the phase spectrum of each component to vary freely over time. For simplicity of notation, let us hereafter put $c_{l,k,m} = e^{j\phi_{l,k,m}}$. In [5], we considered an objective function

$$\mathcal{I}_{\mathrm{EU}}(\theta) = \sum_{k,m} \Big| Y_{k,m} - \sum_{l} H_{k,l} U_{l,m} c_{l,k,m} \Big|^2 + \mathcal{R}(\boldsymbol{U}), \tag{3}$$

where $\theta = \{\boldsymbol{H}, \boldsymbol{U}, \boldsymbol{C}\}$ and $\mathcal{R}(\boldsymbol{U})$ is a regularization term for $\boldsymbol{U}$. It is important to note that the complex NMF model allows the components to cancel each other out, and so some constraint is needed to induce the sparsity of $\boldsymbol{U}$. For this purpose, we define $\mathcal{R}(\boldsymbol{U})$ using the ℓ_p norm

$$\mathcal{R}(\boldsymbol{U}) = 2\lambda \sum_{l,m} |U_{l,m}|^p, \tag{4}$$

where $\lambda > 0$ weighs the importance of the sparsity cost relative to the fitting cost. When $0 < p < 2$, $\mathcal{R}(\boldsymbol{U})$ promotes sparsity if the norm of $\boldsymbol{U}$ is bounded. To bound $\boldsymbol{U}$, we assume $\sum_k H_{k,l}^2 = 1$ or $\sum_k H_{k,l} = 1$.

Although it is difficult to solve the above optimization problem analytically, a convergence-guaranteed algorithm for finding a stationary point can be developed based on the auxiliary function concept. To derive the algorithm, we first introduce the general principle of the auxiliary function approach (majorization-minimization approach) [9, 16, 17].

We use $\mathcal{F}(\theta)$ to denote an objective function that we want to minimize with respect to θ. $\mathcal{F}^+(\theta, \alpha)$ is defined as an auxiliary function for $\mathcal{F}(\theta)$ if it satisfies

$$\mathcal{F}(\theta) = \min_{\alpha} \mathcal{F}^+(\theta, \alpha). \tag{5}$$

We call α an auxiliary variable. By using $\mathcal{F}^+(\theta, \alpha)$, $\mathcal{F}(\theta)$ can be iteratively decreased according to the following theorem:

Theorem 1. *$\mathcal{F}(\theta)$ is non-increasing under the updates, $\theta \leftarrow \mathrm{argmin}_\theta \mathcal{F}^+(\theta, \alpha)$ and $\alpha \leftarrow \mathrm{argmin}_\alpha \mathcal{F}^+(\theta, \alpha)$.*

It can be shown [5] that

$$\mathcal{I}^+_{\mathrm{EU}}(\theta, \alpha) = \sum_{l,k,m} \frac{|X_{l,k,m} - H_{k,l} U_{l,m} c_{l,k,m}|^2}{\beta_{l,k,m}} + \lambda \sum_{l,m} \big\{ p|V_{l,m}|^{p-2} U_{l,m}^2 + (2-p)|V_{l,m}|^p \big\}, \tag{6}$$

is an auxiliary function of $\mathcal{I}_{\mathrm{EU}}(\theta)$ where $\alpha = \{\boldsymbol{X}, \boldsymbol{V}\}$, $\beta_{l,k,m}$ can be any positive number satisfying $\sum_l \beta_{l,k,m} = 1$, and $X_{l,k,m} \in \mathbb{C}$ and $V_{l,m} \geq 0$ are auxiliary variables satisfying

$$\sum_{l} X_{l,k,m} = Y_{k,m}. \tag{7}$$

By using $\mathcal{I}^+_{\mathrm{EU}}(\theta, \alpha)$, the update rules for α are derived as

$$X_{l,k,m} = H_{k,l} U_{l,m} c_{l,k,m} + \beta_{l,k,m} \Big(Y_{k,m} - \sum_{l} H_{k,l} U_{l,m} c_{l,k,m} \Big), \tag{8}$$

$$V_{l,m} = U_{l,m}, \tag{9}$$

and the update rules for θ are derived as

$$H_{k,l} = \frac{\sum_m U_{l,m} |X_{l,k,m}| / \beta_{l,k,m}}{\sqrt{\sum_k (\sum_m U_{l,m} |X_{l,k,m}| / \beta_{l,k,m})^2}}, \tag{10}$$

$$U_{l,m} = \frac{\sum_k H_{k,l} |X_{l,k,m}| / \beta_{l,k,m}}{\sum_k H_{k,l}^2 / \beta_{l,k,m} + \lambda p V_{l,m}^{p-2}}, \tag{11}$$

$$c_{l,k,m} = X_{l,k,m} / |X_{l,k,m}|. \tag{12}$$

The CNMF algorithm can thus be summarized as follows:

1. Initialize $\boldsymbol{H}$, $\boldsymbol{U}$ and $\boldsymbol{C}$.
2. Update $\boldsymbol{X}$ and $\boldsymbol{V}$ using (8) and (9).
3. Update $\boldsymbol{H}$, $\boldsymbol{U}$ and $\boldsymbol{C}$ using (10)–(12) and return to 2.

We call this "Euclidean CNMF (EU-CNMF)".

Here, the auxiliary variable $X_{l,k,m}$ can be viewed as an estimate of the complex spectrogram of the l-th signal component. At step 2, $X_{l,k,m}$ is updated by adding the portion of the error between the observed spectrogram and the model to the current estimate of $H_{k,l} U_{l,m} c_{l,k,m}$. $c_{l,k,m}$ is then updated at its argument $X_{l,k,m}/|X_{l,k,m}|$, and $\boldsymbol{H}$ and $\boldsymbol{U}$ are updated using its magnitude $|X_{l,k,m}|$. Although the details are omitted owing to space limitations, it can be shown that (8) is an MMSE estimator of $X_{l,k,m}$, i.e., $\mathbb{E}[X_{l,k,m}|\boldsymbol{Y}, \boldsymbol{H}, \boldsymbol{U}, \boldsymbol{C}]$, when seen from a generative model perspective. This is in contrast to the Wiener filter $\mathbb{E}[X_{l,k,m}|\boldsymbol{Y}, \boldsymbol{H}, \boldsymbol{U}]$, which only uses the estimate of the power spectrogram.

3. KULLBACK-LEIBLER COMPLEX NMF

Interestingly, it can be shown that the algorithm presented above also converges to a stationary point of the following optimization problem:

$$\begin{aligned} &\text{minimize} \;\; \mathcal{J}_{\mathrm{EU}}(\bar{\theta}) \\ &\text{subject to} \;\; \sum_{l} X_{l,k,m} = Y_{k,m}, \end{aligned} \tag{13}$$

where $\bar{\theta} = \{\boldsymbol{H}, \boldsymbol{U}, \boldsymbol{X}\}$ and

$$\mathcal{J}_{\mathrm{EU}}(\bar{\theta}) = \sum_{l,k,m} \frac{(|X_{l,k,m}| - H_{k,l}U_{l,m})^2}{\beta_{l,k,m}} + \mathcal{R}(\boldsymbol{U}). \tag{14}$$

This can be confirmed as follows. The first term of the auxiliary function $\mathcal{I}^{+}(\theta, \alpha)$ of EU-CNMF can be written as

$$\begin{aligned} &|X_{l,k,m} - H_{k,l}U_{l,m}c_{l,k,m}|^2 \\ =&|X_{l,k,m}|^2 - 2H_{k,l}U_{l,m}\mathrm{Re}[c^{*}_{l,k,m}X_{l,k,m}] + H^2_{k,l}U^2_{l,m}. \end{aligned} \tag{15}$$

Now, by using the fact that a function $g(z) = -\mathrm{Re}(c^{*}z)$ with complex arguments z and $|c| = 1$ is a tangent plane to a cone $f(z) = -|z|$, where c indicates the direction of the tangent line, we obtain an inequality

$$-\mathrm{Re}(c^{*}z) \geq -|z|. \tag{16}$$

We therefore obtain

$$\begin{aligned} &|X_{l,k,m}|^2 - 2H_{k,l}U_{l,m}\mathrm{Re}[c^{*}_{l,k,m}X_{l,k,m}] + H^2_{k,l}U^2_{l,m} \\ \geq&|X_{l,k,m}|^2 - 2H_{k,l}U_{l,m}|X_{l,k,m}| + H^2_{k,l}U^2_{l,m} \\ =&(|X_{l,k,m}| - H_{k,l}U_{l,m})^2. \end{aligned}$$

Hence, $\mathcal{J}_{\mathrm{EU}}(\bar{\theta}) \leq \mathcal{I}^{+}(\theta, \alpha)$. This implies that (6) is also an auxiliary function of $\mathcal{J}_{\mathrm{EU}}(\bar{\theta})$. Here, the difference is that the roles of $\boldsymbol{X} = \{X_{l,k,m}\}$ and $\boldsymbol{C} = \{c_{l,k,m}\}$ are reversed: $\boldsymbol{X}$ is a model parameter and $\boldsymbol{C}$ is an auxiliary variable, while $\boldsymbol{X}$ is an auxiliary variable and $\boldsymbol{C}$ is a model parameter in the original CNMF. Thus, there is a duality between the optimization problem of the original CNMF and (13). (13) can be interpreted as a problem of decomposing an observed complex spectrogram $Y_{k,m}$ into the sum of L components $X_{1,k,m}, \ldots, X_{L,k,m}$ such that the magnitude spectrogram of each component is as close as possible to a rank-1 structure. This gives a different explanation to the objective of CNMF. We call this formulation technique "dual formulation."

Since the KL divergence only allows non-negative arguments, it cannot be straightforwardly used to measure the difference between $Y_{k,m}$ and $\sum_l H_{k,l}U_{l,m}c_{l,k,m}$. However, since the dual CNMF formulation uses the Euclidean distance between non-negative values, $|X_{l,k,m}|$ and $H_{k,l}U_{l,m}$, as the cost function, we can also use the KL divergence to measure their difference, which leads us to an objective function

$$\mathcal{J}_{\mathrm{KL}}(\bar{\theta}) = \sum_{l,k,m} \mathcal{D}_{\mathrm{KL}}(|X_{l,k,m}| \| H_{k,l}U_{l,m}) + \mathcal{R}(\boldsymbol{U}), \tag{17}$$

where $\mathcal{D}_{\mathrm{KL}}(x\|y) = x\log\frac{x}{y} - x + y$. Thus, we can consider the following optimization problem

$$\begin{aligned} &\text{minimize} \;\; \mathcal{J}_{\mathrm{KL}}(\bar{\theta}) \\ &\text{subject to} \;\; \sum_{l} X_{l,k,m} = Y_{k,m}, \end{aligned} \tag{18}$$

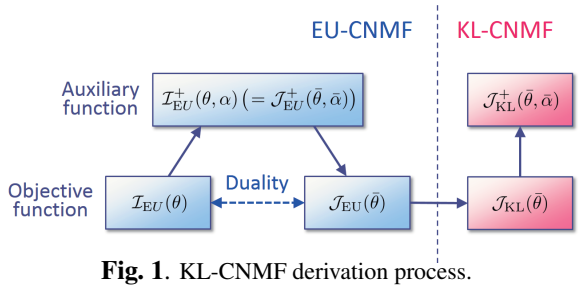

Fig. 1. KL-CNMF derivation process.

which we call "KL-CNMF."

The objective $\mathcal{J}_{\mathrm{KL}}(\bar{\theta})$ is non-differentiable with respect to $X_{l,k,m}$. In the following, we construct an easy-to-optimize auxiliary function to obtain a closed form update rule for $X_{l,k,m}$. First, we can show that

$$\begin{aligned} &|X_{l,k,m}|\log|X_{l,k,m}| - |X_{l,k,m}|\log H_{k,l}U_{l,m} - |X_{l,k,m}| \\ \leq&|X_{l,k,m}|\left\{\tfrac{|X_{l,k,m}| - Z_{l,k,m}}{Z_{l,k,m}} + \log Z_{l,k,m}\right\} \\ &- |X_{l,k,m}|\log H_{k,l}U_{l,m} - |X_{l,k,m}| \\ =&\tfrac{|X_{l,k,m}|^2}{Z_{l,k,m}} + |X_{l,k,m}|\left(\log\tfrac{Z_{l,k,m}}{H_{k,l}U_{l,m}} - 2\right), \end{aligned} \tag{19}$$

by using the fact that a logarithmic function is a concave function and that for any concave function $f(x)$, $f(x)$ is below or equal to its tangent at any point, namely $f(x) \leq f'(z)(x - z) + f(z)$. The equality of (19) holds when $Z_{l,k,m} = |X_{l,k,m}|$. The right-hand side of (19) still involves a non-differentiable term $|X_{l,k,m}|$. Here, the coefficient

$$d_{l,k,m} = \log\tfrac{Z_{l,k,m}}{H_{k,l}U_{l,m}} - 2, \tag{20}$$

can be either non-negative or negative. According to the sign of $d_{l,k,m}$, we can use the following inequalities

$$|X_{l,k,m}| \leq \tfrac{|X_{l,k,m}|^2}{2W_{l,k,m}} + \tfrac{W_{l,k,m}}{2}, \tag{21}$$

$$-|X_{l,k,m}| \leq -\mathrm{Re}[c^{*}_{l,k,m}X_{l,k,m}], \tag{22}$$

to obtain

$$\begin{aligned} &\tfrac{|X_{l,k,m}|^2}{Z_{l,k,m}} + d_{l,k,m}|X_{l,k,m}| \\ \leq&A_{l,k,m}|X_{l,k,m}|^2 - 2\mathrm{Re}[B^{*}_{l,k,m}X_{l,k,m}] + D_{l,k,m}, \end{aligned} \tag{23}$$

where

$$A_{l,k,m} := \begin{cases} \frac{d_{l,k,m}}{2W_{l,k,m}} + \frac{1}{Z_{l,k,m}} & (d_{l,k,m} \geq 0) \\ \frac{1}{Z_{l,k,m}} & (d_{l,k,m} < 0) \end{cases}, \tag{24}$$

$$B_{l,k,m} := \begin{cases} 0 & (d_{l,k,m} \geq 0) \\ -d_{l,k,m}c_{l,k,m}/2 & (d_{l,k,m} < 0) \end{cases}. \tag{25}$$

$D_{l,k,m}$ is given by $d_{l,k,m}W_{l,k,m}/2$ when $d_{l,k,m} \geq 0$ and 0 otherwise. The equalities of (21) and (22) hold when $W_{l,k,m} = |X_{l,k,m}|$ and $c_{l,k,m} = X_{l,k,m}/|X_{l,k,m}|$, respectively. Thus, we obtain an auxiliary function of $\mathcal{J}_{\mathrm{KL}}(\bar{\theta})$ as

$$\mathcal{J}^{+}_{\mathrm{KL}}(\bar{\theta}, \bar{\alpha}) = A_{l,k,m}|X_{l,k,m}|^2 - 2\mathrm{Re}[B^{*}_{l,k,m}X_{l,k,m}]$$

$$+ D_{l,k,m} + H_{k,l}U_{l,m} + \mathcal{R}^{+}(\boldsymbol{U}, \boldsymbol{V}), \quad (26)$$

where $\bar{\theta} = \{\boldsymbol{H}, \boldsymbol{U}, \boldsymbol{X}\}$ is a set of parameters, $\bar{\alpha} = \{\boldsymbol{Z}, \boldsymbol{W}, \boldsymbol{C}, \boldsymbol{V}\}$ is a set of auxiliary variables, and $\mathcal{R}^{+}(\boldsymbol{U}, \boldsymbol{V})$ is an auxiliary function of $\mathcal{R}(\boldsymbol{U})$, given by

$$\mathcal{R}^{+}(\boldsymbol{U}, \boldsymbol{V}) \quad (27)$$
$$= \begin{cases} 2\lambda \sum_{l,m} \{pV_{l,m}^{p-1}(U_{l,m} - V_{l,m}) + V_{l,m}^{p}\} & (0 < p \leq 1) \\ \lambda \sum_{l,m} \{pV_{l,m}^{p-2}U_{l,m}^{2} + (2-p)V_{l,m}^{p}\} & (1 < p \leq 2) \end{cases}.$$

As can be seen from (26), $\mathcal{J}_{\mathrm{KL}}^{+}(\bar{\theta}, \bar{\alpha})$ is a quadratic function of $X_{l,k,m}$. Thus, by using the method of Lagrange multipliers, we obtain the update rule for $X_{l,k,m}$ analytically as

$$X_{l,k,m} = \frac{1}{A_{l,k,m}} \left(B_{l,k,m} + \frac{Y_{k,m} - \sum_{l} \frac{B_{l,k,m}}{A_{l,k,m}}}{\sum_{l} \frac{1}{A_{l,k,m}}} \right). \quad (28)$$

This provides yet another form of filter that optimally decomposes $Y_{k,m}$ into L complex spectrograms $X_{1,k,m}, \ldots, X_{L,k,m}$ using the estimates of $\boldsymbol{H}$, $\boldsymbol{U}$ and $\boldsymbol{c}$. When $0 < p \leq 1$, the update rules for $\boldsymbol{H}$ and $\boldsymbol{U}$ can be derived as

$$H_{k,l} = \frac{\sum_{m} |X_{l,k,m}|}{\sum_{k} \sum_{m} |X_{l,k,m}|}, \quad (29)$$

$$U_{l,m} = \frac{\sum_{k} |X_{l,k,m}|}{1 + \lambda p V_{l,m}^{p-1}}. \quad (30)$$

Note that here we have used the ℓ_1 norm constraint $\sum_{k} H_{k,l} = 1$ for $\boldsymbol{H}$. The update rule for $\boldsymbol{U}$ when $1 < p \leq 2$ can also be derived in closed form. The update rules for the auxiliary variables $\bar{\alpha}$ are derived as

$$Z_{l,k,m} = |X_{l,k,m}|, \quad (31)$$
$$W_{l,k,m} = |X_{l,k,m}|, \quad (32)$$
$$c_{l,k,m} = X_{l,k,m}/|X_{l,k,m}|, \quad (33)$$
$$V_{l,m} = U_{l,m}. \quad (34)$$

Overall, the optimization algorithm for KL-CNMF can be summarized as follows:

1. Initialize $\boldsymbol{H}$, $\boldsymbol{U}$ and $\boldsymbol{X}$.
2. Update $\boldsymbol{Z}$, $\boldsymbol{W}$, $\boldsymbol{C}$ and $\boldsymbol{V}$ using (31)–(34).
3. Update $\boldsymbol{H}$, $\boldsymbol{U}$ and $\boldsymbol{X}$ using (28)–(30) and return to 2.

For the initialization (step 1), we can use conventional NMF algorithms followed by Wiener filtering or the EU-CNMF algorithm to obtain the estimates of $\boldsymbol{H}$, $\boldsymbol{U}$ and $\boldsymbol{X}$.

The derivation process of KL-CNMF and its relationship to EU-CNMF are illustrated in Fig. 1.

4. EXPERIMENTS

We conducted supervised source separation experiments to compare the source separation accuracy of KL-CNMF (proposed), EU-CNMF, KL-NMF, EU-NMF and IS-NMF. We used three music recordings from the SiSEC 2013 database, available at [18], as the experimental data. Each recording is a mixture of 5 tracks, each of which is produced by a single instrument or singer. The separated tracks are also available.

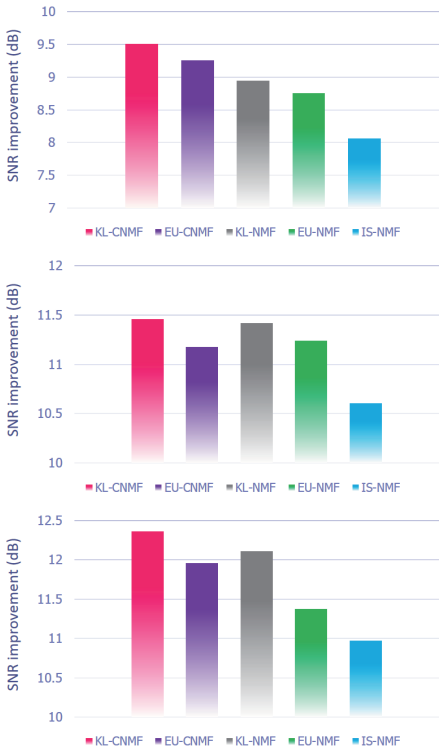

Fig. 2. Average SNR improvements obtained with KL-CNMF (proposed), EU-CNMF, KL-NMF, EU-NMF and IS-NMF. We used three music recordings from the SiSEC 2013 database [18], "Ultimate NZ Tour"(top), "Bearlin - Roads"(middle) and "Fort Minor - Remember the Name"(bottom), for the test data.

We performed 3-fold cross validation. We partitioned each recording into three segments, used one segment as the test data and the other two segments as the training data, repeated signal-to-noise (SNR) evaluations three times with different test segments, and took the average of the SNR improvements obtained with the three repeated rounds. With all these methods, the basis spectra were pretrained using the individual tracks of the training data, and then source separation was performed on the test data. 6 basis spectra were assigned to each track. Thus, a total of 30 basis spectra were used for the separation. All the audio samples were monaural and sampled at 22.05kHz. An STFT was computed using a square-root Hanning window that was 32ms long with a 16ms overlap. Fig. 2 shows the SNR improvements after the separations with all the methods. From these results, we confirmed that KL-CNMF outperformed the other methods.

5. CONCLUSIONS

CNMF is a phase-aware variant of the NMF approach for audio source separation, which makes it possible to realize NMF-like signal separations in the complex time-frequency domain. One limitation of the conventional CNMF is that the divergence measure is limited to the Euclidean distance. This paper proposed a KL divergence counterpart of CNMF, which we call "KL-CNMF," and derived an algorithm for finding a locally optimal solution. We confirmed through supervised source separation experiments that KL-CNMF outperformed other NMF variants.

6. REFERENCES

[1] Y. Wang and D. Wang, "Towards scaling up classification-based speech separation," *IEEE Trans. Audio, Speech, Language Process.*, vol. 21, no. 7, pp. 1381–1390, 2013.

[2] J. R. Hershey, Z. Chen, J. Le Roux, and S. Watanabe, "Deep clustering: Discriminative embeddings for segmentation and separation," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 31–35.

[3] P. Smaragdis, B. Raj, and M Shashanka, "Supervised and semi-supervised separation of sounds from single-channel mixtures," in *Proceedings of the International Conference on Independent Component Analysis and Signal Separation (ICA)*, 2007, pp. 414–421.

[4] F. Weninger, J. Le Roux, J. R. Hershey, and S. Watanabe, "Discriminative NMF and its application to single-channel source separation," in *Proceedings of the Annual Conference of the International Speech Communication Association (Interspeech)*, 2014, pp. 536–543.

[5] H. Kameoka, N. Ono, K. Kashino, and S. Sagayama, "Complex NMF: A new sparse representation for acoustic signals," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2009, pp. 3437–3440.

[6] R. M. Parry and I. Essa, "Phase-aware non-negative spectrogram factorization," in *Proceedings of the International Conference on Independent Component Analysis and Signal Separation (ICA)*, 2007, pp. 536–543.

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

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

[9] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in Neural Information Processing Systems (NIPS)*. 2000, pp. 556–562, MIT Press.

[10] S. Dhillon and S. Sra, "Generalized nonnegative matrix approximations with Bregman divergences," in *Advances in Neural Information Processing Systems (NIPS)*, 2005, pp. 283–290.

[11] A. Cichocki, R. Zdunek, and S. Amari, "Csiszar's divergences for non-negative matrix factorization: Family of new algorithms," in *Proceedings of the International Conference on Independent Component Analysis and Signal Separation (ICA)*, 2006, pp. 32–39.

[12] 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 *Proceedings of the IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, 2010, pp. 283–288.

[13] C. Févotte and J. Idier, "Algorithms for nonnegative matrix factorization with the β-divergence," *Neural Computation*, vol. 23, no. 9, 2011.

[14] D FitzGerald, M Cranitch, and E Coyle, "On the use of the beta divergence for musical source separation," in *Proceedings of the Irish Signals and Systems Conference (ISSC)*, 2009.

[15] B. King, C. Févotte, and P. Smaragdis, "Optimal cost and magnitude power for nmf-based speech separation and music interpolation," in *Proceedings of the IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, 2012, pp. 23–26.

[16] J. M. Ortega and W. C. Rheinboldt, *Iterative solutions of nonlinear equations in several variables*, Academic Press, New York, 1970.

[17] D. R. Hunter and K. Lange, "Quantile regression via an MM algorithm," *Journal of Computational and Graphical Statistics*, vol. 9, pp. 60–77, 2000.

[18] "The main page for the fourth community-based signal separation evaluation campaign (SiSEC 2013)," `https://sisec.wiki.irisa.fr/`.