IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 16, 2020, accepted December 6, 2020, date of publication December 18, 2020,
date of current version December 31, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3045791

Majorization-Minimization Algorithm for Discriminative Non-Negative Matrix Factorization

LI LI[1], (Student Member, IEEE), HIROKAZU KAMEOKA[2], (Senior Member, IEEE), AND SHOJI MAKINO[1], (Fellow, IEEE)

[1]Graduate School of Systems and Information Engineering, University of Tsukuba, Ibaraki 3050821, Japan
[2]NTT Communication Science Laboratories, Kanagawa 2430198, Japan

Corresponding author: Li Li (lili@mmlab.cs.tsukuba.ac.jp)

This work was supported in part by the JSPS KAKENHI under Grant 18J20059, in part by the JST CREST under Grant JPMJCR19A3, and in part by the SECOM Science and Technology Foundation.

ABSTRACT This paper proposes a basis training algorithm for discriminative non-negative matrix factorization (NMF) with applications to single-channel audio source separation. With an NMF-based approach to supervised audio source separation, NMF is first applied to train the basis spectra of each source using training examples and then applied to the spectrogram of a mixture signal using the pretrained basis spectra at test time. The source signals can then be separated out using a Wiener filter. Here, a typical way to train the basis spectra is to minimize the dissimilarity measure between the observed spectrogram and the NMF model. However, obtaining the basis spectra in this way does not ensure that the separated signal will be optimal at test time due to the inconsistency between the objective functions for training and separation (Wiener filtering). To address this mismatch, a framework called discriminative NMF (DNMF) has recently been proposed. While this framework is noteworthy in that it uses a common objective function for training and separation, the objective function becomes more analytically complex than that of regular NMF. In the original DNMF work, a multiplicative update algorithm was proposed for the basis training; however, the convergence of the algorithm is not guaranteed and can be very slow. To overcome this weakness, this paper proposes a convergence-guaranteed algorithm for DNMF based on a majorization-minimization principle. Experimental results show that the proposed algorithm outperform the conventional DNMF algorithm as well as the regular NMF algorithm in terms of both the signal-to-distortion and signal-to-interference ratios.

INDEX TERMS Discriminative non-negative matrix factorization (NMF), majorization-minimization, single-channel signal processing, speech enhancement, source separation.

I. INTRODUCTION

Single-channel audio source separation is a challenging task of extracting individual source signals from a monaural recording of a mixture signal. Since the presence of noise or interference can severely degrade the performance of many audio applications such as automatic transcription of music, speech recognition, voice conversion, many attempts have been made to address this problem [1]–[8]. One successful approach for monaural audio source separation involves applications of non-negative matrix factorization (NMF) [6], [10]. Although deep neural networks-based methods [7]–[9] have been shown to work impressively in recent years, the NMF approach still remains attractive when only a limited amount of training data is available.

The basic idea of the NMF approach is to interpret the observed magnitude (or power) spectrogram of a signal as a non-negative matrix and factorize it into the product of non-negative matrices. This amounts to approximating the observed spectra by a linear sum of basis spectra scaled by time-varying amplitudes. In a supervised/semi-supervised source separation problem setting, NMF is first used 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 each subset of the basis spectra is fixed at the pretrained spectra. The source signals can then be separated out using a Wiener filter constructed by employing the estimated power spectrogram

The associate editor coordinating the review of this manuscript and approving it for publication was Manuel Rosa-Zurera.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

of each source. A typical way to train the basis spectra of each source is to minimize a divergence measure between the NMF model and the spectrogram of the training samples of that source. However, the basis spectra obtained in this way do not ensure that the separated signal at test time will be optimal since the objective functions for training and separation are inconsistent, namely a divergence measure for training and Wiener filtering for separation.

To address this mismatch between the training and test objectives, a framework called discriminative NMF (DNMF) has recently been proposed [11]. While many methods called ''discriminative NMF'' [12]–[17] have been proposed with the aim of enhancing the discriminative power of the basis spectra, in this paper, we use this term in relation to the work done by Weninger [11]. Note that the term ''discriminative'' is used in association with the discriminative models for classification and regression. The central idea of DNMF is that the basis spectra are trained in such a way that the output of the Wiener filter becomes as close to the spectrogram of each of the training examples as possible so that the separated signals become optimal at test time. This approach differs from the conventional supervised NMF framework in that it uses the training examples of all the sources to train the basis spectra for each of the sources. This is important since it helps to enhance the discriminative power of the basis spectra. However, the training criterion for DNMF becomes analytically more complex than the typical divergence measures used in the standard NMF framework, which causes difficulty as regards optimization of the basis spectra. In [11], Weninger proposed a multiplicative update (MU) algorithm for the basis training, where the multiplicative factor is obtained by dividing the negative parts by the positive parts of the partial derivative of the objective function as done in [18]. Although this way of obtaining update rules is indeed convenient in that it is applicable as long as an objective function is differentiable, one drawback is that the algorithm is generally not guaranteed to converge to a stationary point. To overcome this weakness, this paper proposes using a majorization-minimization (MM) principle to derive a convergence-guaranteed basis training algorithm for DNMF. We show in Sec. IV that using the present basis training algorithm instead of the conventional MU algorithm leads to notable improvements in source separation performance.

The rest of this paper is organized as follows. Section Sec. II reviews the standard NMF and DNMF approaches for single-channel source separation and the multiplicative update algorithm. In section Sec. III, we introduce the MM principle, on which basis we derive the proposed algorithm. We show the experimental results in Sec. IV and conclude the paper in Sec. V.

II. DISCRIMINATIVE NON-NEGATIVE MATRIX FACTORIZATION

A. STANDARD NMF APPROACH

We start by reviewing the standard NMF approach for single-channel source separation. Let the number of sources be L. We use $\mathbf{Y} = (y_{\omega,t})_{\Omega\times T} \in \mathbb{R}^{\geq 0,\Omega\times T}$ to denote the power spectrogram of a mixture signal obtained using the short-term Fourier transform (STFT), where ω and t are the frequency and time indices, respectively. With the supervised NMF approach, we factorize $\mathbf{Y}$, interpreted as a non-negative matrix, into the product of a non-negative basis matrix $\tilde{\mathbf{W}} = [\tilde{\mathbf{W}}^1, \tilde{\mathbf{W}}^2, \ldots \tilde{\mathbf{W}}^L]$ and a non-negative coefficient (activation) matrix $\hat{\mathbf{H}} = [\hat{\mathbf{H}}^1; \hat{\mathbf{H}}^2; \ldots; \hat{\mathbf{H}}^L]$, where $\tilde{\mathbf{W}}^l = (\tilde{w}^l_{\omega,k})_{\Omega\times K^l} \in \mathbb{R}^{\geq 0,\Omega\times K^l}$ is assumed to be pretrained using the spectrogram of a training sample $\mathbf{S}^l = (s^l_{\omega,t})_{\Omega\times T}$ for each $l = 1, 2, \ldots, L$. A common way to train $\tilde{\mathbf{W}}^l$ is to solve

$$(\tilde{\mathbf{W}}^l, \tilde{\mathbf{H}}^l) = \underset{\mathbf{W}^l,\mathbf{H}^l}{\operatorname{argmin}}\, \mathcal{D}(\mathbf{S}^l|\mathbf{W}^l\mathbf{H}^l) + \mu||\mathbf{H}^l||_1, \tag{1}$$

where $\mathcal{D}$ is a cost function that measures the dissimilarity of $\mathbf{S}^l$ and $\mathbf{W}^l\mathbf{H}^l$. Here, we have assumed $\mu\|\mathbf{H}^l\|_1$ is used as a regularization term for promoting sparsity of $\|\mathbf{H}^l\|_1$, where μ is a regularization parameter that weighs the importance of the regularization term. Note that we can use other kinds of regularization terms, but here we omit them for simplicity. At test time, the concatenated basis matrix $\tilde{\mathbf{W}}$ is fixed at the pretrained basis spectra and the activation matrix $\mathbf{H}$ is estimated by solving

$$\hat{\mathbf{H}} = \underset{\mathbf{H}}{\operatorname{argmin}}\, \mathcal{D}(\mathbf{Y}|\tilde{\mathbf{W}}\mathbf{H}) + \mu||\mathbf{H}||_1, \tag{2}$$

subject to non-negativity. Typical choices for $\mathcal{D}(\mathbf{Y}|\mathbf{X})$ include the Euclidean distance, the generalized Kullback-Leibler (KL) divergence, and the Itakura-Saito (IS) divergence:

$$\mathcal{D}_{EU}(\mathbf{Y}|\mathbf{X}) = \|\mathbf{Y} - \mathbf{X}\|_F^2 = \sum_{\omega,t} |y_{\omega,t} - x_{\omega,t}|^2, \tag{3}$$

$$\mathcal{D}_{KL}(\mathbf{Y}|\mathbf{X}) = \sum_{\omega,t} \left(y_{\omega,t} \log \frac{y_{\omega,t}}{x_{\omega,t}} - y_{\omega,t} + x_{\omega,t} \right), \tag{4}$$

$$\mathcal{D}_{IS}(\mathbf{Y}|\mathbf{X}) = \sum_{\omega,t} \left(\frac{y_{\omega,t}}{x_{\omega,t}} - \log \frac{y_{\omega,t}}{x_{\omega,t}} - 1 \right), \tag{5}$$

where $y_{\omega,t}$ and $x_{\omega,t}$ are the (ω, t)th elements of $\mathbf{Y}$ and $\mathbf{X}$.

A naïve way of obtaining the time-domain signal of the lth source is to simply use $\tilde{\mathbf{W}}^l\hat{\mathbf{H}}^l$ and the phase spectrogram of the mixture signal to obtain the complex spectrogram and perform the inverse STFT. However, the signals obtained in this way usually contain artifacts and often sound artificial. Another widely used way involves using the Wiener filter. Namely, once $\tilde{\mathbf{W}}$ and $\hat{\mathbf{H}}$ are obtained, the magnitude spectrogram of the lth source can be refined using the Wiener filter constructed using the estimated power spectrogram

$$\mathbf{C}^l = \frac{\tilde{\mathbf{W}}^l\hat{\mathbf{H}}^l}{\tilde{\mathbf{W}}\hat{\mathbf{H}}} \odot \mathbf{Y} \tag{6}$$

so that $\mathbf{C}^1, \ldots, \mathbf{C}^L$ are ensured to sum to the magnitude spectrogram $\mathbf{Y}$ of the test mixture signal, where $\odot$ and $\frac{\cdot}{\cdot}$ denote elementwise multiplication and division. Note that here we have used sans serif fonts to express magnitude spectrograms, $\mathbf{Y} = \sqrt{\mathbf{Y}}$ and $\mathbf{C}^l = \sqrt{\mathbf{C}^l}$, where $\sqrt{\cdot}$ denotes the element-wise square-root.

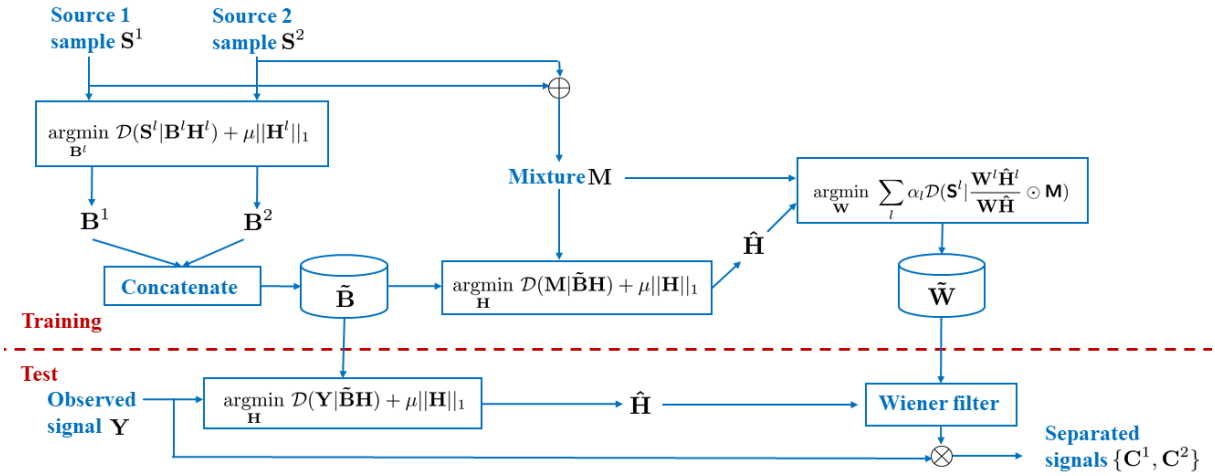

FIGURE 1. Flowchart of DNMF in two-source case.

B. DISCRIMINATIVE NMF

If we assume the Wiener filter is used to obtain source signals, the training and test objectives become inconsistent. Namely, the basis spectra are not necessarily trained in such a way that the separated signals at test time will be optimal. With the standard NMF approach, at test time, the basis matrix $\mathbf{W}$ is used not only for estimating $\mathbf{H}$ from $\mathbf{Y}$ but also for constructing the Wiener filter in Eq. (6). To make the training objective consistent with this test inference procedure, Weninger [11] proposed introducing two separate basis matrices for these different purposes, $\mathbf{B}$ and $\mathbf{W}$, and formulating a bilevel optimization problem

$$(\tilde{\mathbf{B}}^l,\tilde{\mathbf{H}}^l)=\underset{\mathbf{B}^l,\mathbf{H}^l}{\operatorname{argmin}}\ \mathcal{D}(\mathbf{S}^l|\mathbf{B}^l\mathbf{H}^l)+\mu||\mathbf{H}^l||_1, \tag{7}$$

$$\hat{\mathbf{H}}=\underset{\mathbf{H}}{\operatorname{argmin}}\ \mathcal{D}(\mathbf{M}|\tilde{\mathbf{B}}\mathbf{H})+\mu||\mathbf{H}||_1, \tag{8}$$

$$\tilde{\mathbf{W}}=\underset{\mathbf{W}}{\operatorname{argmin}}\ \sum_l \alpha_l \mathcal{D}\left(\mathbf{S}^l\middle|\frac{\mathbf{W}^l\hat{\mathbf{H}}^l}{\mathbf{W}\hat{\mathbf{H}}}\odot\mathbf{M}\right) \tag{9}$$

for training $\mathbf{B}$ and $\mathbf{W}$ so that $\mathbf{B}$ will be optimized for estimating $\mathbf{H}$ from $\mathbf{Y}$ and $\mathbf{W}$ will be optimized for obtaining $\mathbf{C}^1,\ldots,\mathbf{C}^L$ based on the Wiener filter. Here, $\alpha_l \geq 0$ is a constant that weighs the importance of source l. $\mathbf{M}=(m_{\omega,t})_{\Omega\times T}\in\mathbb{R}^{\geq 0,\Omega\times T}$ denotes the power spectrogram of a mixture signal, which can be simply constructed by mixing the training samples $\mathbf{S}^1=(s^1_{\omega,t})_{\Omega,T},\ldots,\mathbf{S}^L=(s^L_{\omega,t})_{\Omega,T}$. $\mathbf{M}=(\mathsf{m}_{\omega,t})_{\Omega,T}$ and $\mathbf{S}^l=(\mathsf{s}^l_{\omega,t})_{\Omega,T}$ denote the magnitude spectrograms $\sqrt{\mathbf{M}}$ and $\sqrt{\mathbf{S}^l}$, respectively. When our goal is to reconstruct a single-source l only, we shall set α_l at 1 and 0 for other sources $l'\neq l$. Fig. 1 illustrates the training and test processes of DNMF using two sources.

C. MULTIPLICATIVE UPDATE ALGORITHM

An inspection of Eqs. (1) and (9) shows that the training criterion for DNMF is more analytically complex than the objective function of standard NMF. In [11], Weninger proposed a two-stage iterative algorithm for solving the above optimization problem: First, $\mathbf{B}$ and $\mathbf{H}$ are obtained by solving Eq. (8) using a standard NMF algorithm. Second, by using the obtained $\mathbf{H}$, the basis matrix $\mathbf{W}$ is iteratively updated according to multiplicative update rules. Here, we set $\alpha_l=1$ and $\alpha_{l'}=0$ $(l'\neq l)$ and define $\mathbf{W}^{\bar{l}}=[\mathbf{W}^1,\cdots,\mathbf{W}^{l-1},\mathbf{W}^{l+1},\cdots,\mathbf{W}^L]$ and $\mathbf{H}^{\bar{l}}=[\mathbf{H}^1;\cdots;\mathbf{H}^{l-1};\mathbf{H}^{l+1};\cdots;\mathbf{H}^L]$. When $\mathcal{D}$ is defined as the KL divergence, the update rules are given by

$$\mathbf{W}^l\leftarrow\mathbf{W}^l\odot\frac{\dfrac{\mathbf{S}^l\odot(\mathbf{W}^{\bar{l}}\mathbf{H}^{\bar{l}})}{(\mathbf{WH})\odot(\mathbf{W}^{\bar{l}}\mathbf{H}^{\bar{l}})}\mathbf{H}^{l\mathsf{T}}}{\dfrac{\mathbf{M}\odot(\mathbf{W}^{\bar{l}}\mathbf{H}^{\bar{l}})}{(\mathbf{WH})\odot(\mathbf{WH})}\mathbf{H}^{l\mathsf{T}}}, \tag{10}$$

$$\mathbf{W}^{\bar{l}}\leftarrow\mathbf{W}^{\bar{l}}\odot\frac{\dfrac{\mathbf{M}\odot(\mathbf{W}^l\mathbf{H}^l)}{(\mathbf{WH})\odot(\mathbf{WH})}\mathbf{H}^{\bar{l}\mathsf{T}}}{\dfrac{\mathbf{S}^l}{\mathbf{WH}}\mathbf{H}^{\bar{l}\mathsf{T}}}. \tag{11}$$

Here, the multiplicative factors are given by dividing the negative parts by the positive parts of the partial derivatives of the objective function in Eq. (9) with respect to the elements of $\mathbf{W}^l$ and $\mathbf{W}^{\bar{l}}$, as done in [18]. Although this way of obtaining update rules is convenient in that it is generally applicable as long as an objective function is differentiable, one downside is that the algorithm is not guaranteed to converge to a stationary point.

III. DNMF WITH MAJORIZATION-MINIMIZATION

A. MAJORIZATION-MINIMIZATION PRINCIPLE

To overcome the weakness of the conventional MU algorithm, in this paper, we propose employing an MM principle to derive a novel convergence-guaranteed algorithm for solving Eq. (9). When constructing an MM algorithm to minimize

a certain objective function, the main issue is how to design an auxiliary function called a "majorizer" that is guaranteed to never be below the objective function. The following lemma shows that once we obtain an auxiliary function, we can develop an iterative algorithm such that the objective function is guaranteed to be non-increasing at each iteration.

Lemma 1: If we use $F(\Theta)$ to denote an objective function that we want to minimize with respect to Θ and use $F^+(\Theta, \Lambda)$ to denote its auxiliary function, satisfying $F(\Theta) = \min_\Lambda F^+(\Theta, \Lambda)$, then $F(\Theta)$ is non-increasing under the following updates of Λ and Θ:

$$\hat{\Lambda} = \arg\min_{\Lambda} F^+(\Theta, \Lambda), \tag{12}$$

$$\hat{\Theta} = \arg\min_{\Theta} F^+(\Theta, \Lambda). \tag{13}$$

Thus, if $F(\Theta)$ is bounded below, a stationary point of $F(\Theta)$ can be found by iteratively performing these updates.

Proof of Lemma 1: Suppose we set Θ to an arbitrary value $\tilde{\Theta}$. We will prove that $F(\Theta)$ is non-increasing after the update Eq. (12) and Eq. (13). From Eq. (12), one obtains $F(\tilde{\Theta}) = F^+(\tilde{\Theta}, \hat{\Lambda})$, and it is obvious from Eq. (13) that $F^+(\tilde{\Theta}, \hat{\Lambda}) \geq F^+(\hat{\Theta}, \hat{\Lambda})$. By definition, one sees from Eq. (12) that $F^+(\hat{\Theta}, \hat{\Lambda}) \geq F(\hat{\Lambda})$. Therefore, we can immediately prove that $F(\tilde{\Theta}) = F^+(\tilde{\Theta}, \hat{\Lambda}) \geq F^+(\hat{\Theta}, \hat{\Lambda}) \geq F(\hat{\Theta})$. □

It should be noted that this concept is adopted in many existing algorithms. For example, the expectation-maximization (EM) algorithm [19] builds a surrogate for a likelihood function of latent variable models using Jensen's inequality. It is also well known for its use in devising an algorithm for standard NMF [10], [20]. In general, if we can build a tight majorizer that is easy to optimize for the objective function of some optimization problems, we can expect to obtain a fast-converging algorithm. Another advantage of MM-based algorithms is that they have no hyperparameters to tune. This is in contrast to gradient-based methods, which usually require step-size settings.

B. DERIVATION OF MAJORIZERS

Here, we derive majorizers for the objective function where $\mathcal{D}$ is defined as the KL divergence and IS divergence. When $\mathcal{D}$ is defined as the KL divergence, the objective function in Eq. (9) is given by

$$\begin{aligned} & f_{KL}(\mathbf{W}) \\ &= \sum_l \alpha_l \mathcal{D}_{KL}\left(\mathbf{S}^l \middle| \frac{\mathbf{W}^l\mathbf{H}^l}{\mathbf{W}\mathbf{H}} \odot \mathbf{M}\right) \\ &=^c \sum_l \alpha_l \sum_{\omega,t} \left(-\mathsf{s}^l_{\omega,t} \log g^l_{\omega,t} + \mathsf{s}^l_{\omega,t} \log g_{\omega,t} + \frac{g^l_{\omega,t}}{g_{\omega,t}} \mathsf{m}_{\omega,t}\right), \end{aligned} \tag{14}$$

where we have used $g^l_{\omega,t}$ and $g_{\omega,t}$ to represent

$$g^l_{\omega,t} = \sum_{k=1}^{K^l} w^l_{\omega,k} h^l_{k,t}, \tag{15}$$

$$g_{\omega,t} = \sum_{k=1}^{K} w_{\omega,k} h_{k,t} \tag{16}$$

and $=^c$ to denote equality up to a constant term. First, let us focus on the term $g^l_{\omega,t}/g_{\omega,t}$. To construct a majorizer for this term, we can use the following inequality:

Lemma 2: For $a > 0$ and $b > 0$, we have

$$\frac{a}{b} \leq \frac{\lambda a^2}{2} + \frac{1}{2\lambda b^2}.$$

The equality holds if and only if

$$\lambda = \frac{1}{ab}.$$

Proof of Lemma 2: For $a, b, \lambda > 0$,

$$\begin{aligned} \lambda\left(a - \frac{1}{\lambda b}\right)^2 &= \lambda\left(a^2 - 2\frac{a}{\lambda b} + \frac{1}{\lambda^2 b^2}\right) \geq 0 \\ \Rightarrow \frac{a}{b} &\leq \frac{\lambda a^2}{2} + \frac{1}{2\lambda b^2}. \end{aligned} \tag{17}$$

The equality holds if and only if $a - \frac{1}{\lambda b} = 0$. □

Since $\mathsf{m}_{\omega,t}$ is non-negative, we can construct an upper bound for $g^l_{\omega,t}\mathsf{m}_{\omega,t}/g_{\omega,t}$ according to the above lemma,

$$\begin{aligned} f_{KL}(\mathbf{W}) \leq \sum_l \alpha_l \sum_{\omega,t} \Bigg(& -\mathsf{s}^l_{\omega,t} \log g^l_{\omega,t} + \mathsf{s}^l_{\omega,t} \log g_{\omega,t} \\ & + \frac{\lambda^l_{\omega,t} \mathsf{m}_{\omega,t} {g^l_{\omega,t}}^2}{2} + \frac{\mathsf{m}_{\omega,t}}{2\lambda^l_{\omega,t} {g_{\omega,t}}^2}\Bigg). \end{aligned} \tag{18}$$

The equality of Eq. (18) holds if and only if

$$\lambda^l_{\omega,t} = \frac{1}{g^l_{\omega,t} g_{\omega,t}}. \tag{19}$$

In the following, we construct a majorizer for each of the terms on the right-hand side of Eq. (18).

We notice that the function $-\log x$ is convex. Since $\mathsf{s}^l_{\omega,t}$ is positive, $-\mathsf{s}^l_{\omega,t} \log g^l_{\omega,t}$ is convex in $g^l_{\omega,t}$. Hence, we can use Jensen's inequality to obtain a majorizer for this term as

$$-\log g^l_{\omega,t} \leq -\sum_{k=1}^{K^l} \gamma^l_{k,\omega,t} \log \frac{w^l_{\omega,k} h^l_{k,t}}{\gamma^l_{k,\omega,t}}, \tag{20}$$

where $\gamma^l_{k,\omega,t}$ is a positive weight that sums to unity:

$$\sum_{k=1}^{K^l} \gamma^l_{k,\omega,t} = 1. \tag{21}$$

The equality of Eq. (20) holds if and only if

$$\gamma^l_{k,\omega,t} = \frac{w^l_{\omega,k} h^l_{k,t}}{\sum_{k'=1}^{K^l} w^l_{\omega,k'} h^l_{k',t}}. \tag{22}$$

The second term $\mathsf{s}^l_{\omega,t} \log g_{\omega,t}$ is concave in $g_{\omega,t}$. Hence, we can use the fact that a tangent line to the graph of a differentiable concave function lies entirely above the graph:

$$\log g_{\omega,t} \leq \sum_k \frac{w_{\omega,k} h_{k,t}}{\eta_{\omega,t}} + \log \eta_{\omega,t} - 1, \tag{23}$$

where $\eta_{\omega,t}$ is an arbitrary positive number. The equality of this inequality holds if and only if

$$\eta_{\omega,t} = g_{\omega,t}. \tag{24}$$

Since a quadratic function is convex, we can apply Jensen's inequality to the third term, which yields

$$g_{\omega,t}^{l\ 2} \leq \sum_{k=1}^{K^l} \frac{w_{\omega,k}^{l\ 2} h_{k,t}^{l\ 2}}{\beta_{k,\omega,t}^l}, \tag{25}$$

where $\beta_{k,\omega,t}^l > 0$ is also a positive number that sums to unity:

$$\sum_{k=1}^{K^l} \beta_{k,\omega,t}^l = 1. \tag{26}$$

The equality of Eq. (25) holds if and only if

$$\beta_{k,\omega,t}^l = \frac{w_{\omega,k}^l h_{k,t}^l}{\sum_{k'=1}^{K^l} w_{\omega,k'}^l h_{k',t}^l}. \tag{27}$$

As regards the fourth term, we can use the fact that the function $1/x^2$ is convex in the first quadrant and use Jensen's inequality to obtain a majorizer:

$$\frac{1}{g_{\omega,t}^2} \leq \sum_k \frac{\theta_{k,\omega,t}^3}{w_{\omega,k}^2 h_{k,t}^2}, \tag{28}$$

where $\theta_{k,\omega,t}$ is a positive number that sums to unity:

$$\sum_k \theta_{k,\omega,t} = 1. \tag{29}$$

We can confirm that the equality of this inequality holds if and only if

$$\theta_{k,\omega,t} = \frac{w_{\omega,k} h_{k,t}}{\sum_{k'} w_{\omega,k'} h_{k',t}}. \tag{30}$$

From Eqs. (18), (20), (25), and (28), we can construct a majorizer for the objective function with KL divergence as

$$\begin{aligned} &f_{KL}(\mathbf{W}) \\ &\leq \sum_l \alpha_l \sum_{\omega,t,k} \left(\frac{\mathsf{s}_{\omega,t}^l w_{\omega,k} h_{k,t}}{\eta_{\omega,t}} - \mathsf{s}_{\omega,t}^l \gamma_{k,\omega,t}^l \log \frac{w_{\omega,k}^l h_{k,t}^l}{\gamma_{k,\omega,t}^l} \right. \\ &\quad \left. + \frac{\lambda_{\omega,t}^l \mathsf{m}_{\omega,t}}{2\beta_{k,\omega,t}^l} w_{\omega,k}^{l\ 2} h_{k,t}^{l\ 2} + \frac{\mathsf{m}_{\omega,t} \theta_{k,\omega,t}^3}{2\lambda_{\omega,t}^l w_{\omega,k}^2 h_{k,t}^2} \right) + d \\ &=: f_{KL}^+(\mathbf{W}, \boldsymbol{\Gamma}), \end{aligned} \tag{31}$$

where $\boldsymbol{\Gamma}$ denotes a set of all the auxiliary variables, $\{\lambda_{\omega,t}^l\}$, $\{\gamma_{k,\omega,t}^l\}$, $\{\eta_{\omega,t}\}$, $\{\beta_{k,\omega,t}^l\}$ and $\{\theta_{k,\omega,t}\}$, and d denotes a term that does not depend on $\mathbf{W}$.

By using Lemma 2, Jensen's inequality and the concave inequality, we can also derive a majorizer for the case of the IS divergence in a similar manner:

$$\begin{aligned} &f_{IS}(\mathbf{W}) \\ &= \sum_l \alpha_l \mathcal{D}_{IS}\left(\mathbf{S}^l \middle| \frac{\mathbf{W}^l \mathbf{H}^l}{\mathbf{W}\mathbf{H}} \odot \mathbf{M} \right) \\ &= \sum_l \alpha_l \sum_{\omega,t} \left(\frac{\mathsf{s}_{\omega,t}^l g_{\omega,t}}{\mathsf{m}_{\omega,t} g_{\omega,t}^l} - \log g_{\omega,t} + \log g_{\omega,t}^l \right) + d' \\ &\leq \sum_l \alpha_l \sum_{k,\omega,t} \left(\frac{\lambda_{\omega,t}^l \mathsf{s}_{\omega,t}^l w_{\omega,k}^2 h_{k,t}^2}{2\mathsf{m}_{\omega,t} \beta_{k,\omega,t}} + \frac{\mathsf{s}_{\omega,t}^l \theta_{k,\omega,t}^{l\ 3}}{2\lambda_{\omega,t}^l \mathsf{m}_{\omega,t} w_{\omega,k}^{l\ 2} h_{k,t}^{l\ 2}} \right. \\ &\quad \left. - \gamma_{k,\omega,t} \log \frac{w_{\omega,k} h_{k,t}}{\gamma_{k,\omega,t}} + \frac{w_{\omega,k}^l h_{k,t}^l}{\eta_{\omega,t}^l} \right) + d'' \\ &=: f_{IS}^+(\mathbf{W}, \boldsymbol{\Gamma}), \end{aligned} \tag{33}$$

where d' and d'' denote terms that do not depend on $\mathbf{W}$.

These majorizers are particularly noteworthy in that they can be minimized analytically with respect to $w_{\omega,k}^l$ since they are given as the sum of the reciprocal, logarithmic, first-order, and second-order functions.

C. UPDATE RULES

We can obtain the update rules for $w_{\omega,k}^l$ by setting the partial derivatives of the above majorizers with respect to $w_{\omega,k}^l$ at zeros. Thus, the optimal update of $w_{\omega,k}^l$ is given by the positive solution of

$$\begin{aligned} &\alpha_l \left(\sum_t \frac{\lambda_{\omega,t}^l \mathsf{m}_{\omega,t}}{\beta_{k,\omega,t}^l} h_{k,t}^{l\ 2} \right) w_{\omega,k}^{l\ 4} - \alpha_l \left(\sum_t \mathsf{s}_{\omega,t}^l \gamma_{k,\omega,t}^l \right) w_{\omega,k}^{l\ 2} \\ &\quad + \left(\alpha_l \sum_t \frac{\mathsf{s}_{\omega,t}^l h_{k,t}^l}{\eta_{\omega,t}} + \sum_{l':l' \neq l} \alpha_{l'} \sum_t \frac{\mathsf{s}_{\omega,t}^{l'} h_{k,t}^l}{\eta_{\omega,t}} \right) w_{\omega,k}^{l\ 3} \\ &\quad - \left(\alpha_l \sum_t \frac{\mathsf{m}_{\omega,t} \theta_{k,\omega,t}^3}{\lambda_{\omega,t}^l h_{k,t}^{l\ 2}} + \sum_{l':l' \neq l} \alpha_{l'} \sum_t \frac{\mathsf{m}_{\omega,t} \theta_{k,\omega,t}^3}{\lambda_{\omega,t}^{l'} h_{k,t}^{l\ 2}} \right) = 0 \end{aligned} \tag{34}$$

for the KL divergence case and

$$\begin{aligned} &\left(\alpha_l \sum_t \frac{\lambda_{\omega,t}^l \mathsf{s}_{\omega,t}^l h_{k,t}^{l\ 2}}{\mathsf{m}_{\omega,t} \beta_{k,\omega,t}} + \sum_{l':l' \neq l} \alpha_{l'} \sum_t \frac{\lambda_{\omega,t}^{l'} \mathsf{s}_{\omega,t}^{l'} h_{k,t}^{l\ 2}}{\mathsf{m}_{\omega,t} \beta_{k,\omega,t}} \right) w_{\omega,k}^{l\ 4} \\ &\quad - \left(\alpha_l \sum_t \gamma_{k,\omega,t} + \sum_{l':l' \neq l} \alpha_{l'} \sum_t \gamma_{k,\omega,t} \right) w_{\omega,k}^{l\ 2} \\ &\quad + \alpha_l \sum_t \frac{h_{k,t}^l}{\eta_{\omega,t}^l} w_{\omega,k}^{l\ 3} - \alpha_l \sum_t \frac{\mathsf{s}_{\omega,t}^l \theta_{k,\omega,t}^{l\ 3}}{\lambda_{\omega,t} \mathsf{m}_{\omega,t} h_{k,t}^{l\ 2}} = 0 \end{aligned} \tag{35}$$

for the IS divergence case. It is worth noting that since each element of $\mathbf{W}$ is isolated in a separate term in $f_{KL}^+(\mathbf{W}, \boldsymbol{\Gamma})$ and $f_{IS}^+(\mathbf{W}, \boldsymbol{\Gamma})$, we can update each of the elements in parallel. Thus, this algorithm is well suited to parallel implementations. Furthermore, since each of the update rules consists of a negative zeroth-order term and a negative second-order term, it turns out that there is only one positive solution, implying that there is no need to solve a solution selection problem.

$f_{KL}^+(\mathbf{W}, \boldsymbol{\Gamma})$ is minimized with respect to the auxliary variables when the exact bounds of Eqs. (18), (20), (23), (25) and (28) are achieved, namely when Eqs. (19), (22), (24), (27), and (30) are achieved. The proposed basis training algorithm with the KL divergence can therefore be summarized as *Algorithm 1*. The algorithm with the IS divergence can be developed in the same way.

Algorithm 1 Proposed Basis Training Algorithm With KL Divergence

Require: $\mathbf{S}_1, \ldots, \mathbf{S}_L, \mathbf{M}$
  Compute $\tilde{\mathbf{B}}$ and $\tilde{\mathbf{H}}$ using NMF to solve Eq. (7) for all l.
  Compute $\hat{\mathbf{H}}$ using NMF to solve Eq. (8).
  Initialize $\mathbf{W}$ by, for example, $\mathbf{W} \leftarrow \tilde{\mathbf{B}}$.
  Fix $\mathbf{H}$ at $\mathbf{H} \leftarrow \hat{\mathbf{H}}$.
  while not converged **do**
    Update $\boldsymbol{\Gamma}$ via Eqs. (19), (22), (24), (27), and (30).
    Update $\mathbf{W}$ by solving Eq. (34).
  end while
  return $\tilde{\mathbf{B}}$, $\mathbf{W}$

D. TEST INFERENCE ALGORITHM

Let $\mathbf{Y}$ and $\boldsymbol{\Phi}$ be the power and phase spectrograms of a test mixture signal and let $\tilde{\mathbf{B}}$ and $\tilde{\mathbf{W}}$ be the pretrained basis matrices. The test inference algorithm for the DNMF approach consists of computing $\hat{\mathbf{H}}$ by solving

$$\hat{\mathbf{H}} = \underset{\mathbf{H}}{\operatorname{argmin}}\, \mathcal{D}(\mathbf{Y}|\tilde{\mathbf{B}}\mathbf{H}) + \mu||\mathbf{H}||_1, \tag{36}$$

computing $\mathbf{C}^1, \ldots, \mathbf{C}^L$ using

$$\mathbf{C}^l = \frac{\tilde{\mathbf{W}}^l\hat{\mathbf{H}}^l}{\tilde{\mathbf{W}}\hat{\mathbf{H}}} \odot \mathbf{Y}, \tag{37}$$

and performing the inverse STFT on $\mathbf{C}^l \odot \boldsymbol{\Phi}$ for all l. Note that the test inference algorithm for the standard NMF approach corresponds to a special case where $\tilde{\mathbf{B}} = \tilde{\mathbf{W}}$.

IV. EXPERIMENTAL EVALUATIONS

A. SPEECH ENHANCEMENT TASK

First, we evaluated the effect of the proposed algorithm in a speech enhancement task, namely $l \in \{s, n\}$. For comparison, we tested (i) the standard supervised NMF method [21] with Euclidean distance (SNMF_EU), KL divergence (SNMF_KL), and IS divergence (SNMF_IS); (ii) DNMF using the MU-based basis training algorithm [11] with KL divergence (DNMF_MU_KL) and Euclidean distance (DNMF_MU_EU); and (iii) DNMF using the proposed basis training algorithm with KL divergence (DNMF_MM_KL) and IS divergence (DNMF_MM_IS). Note that we have excluded DNMF_MU_IS from the baselines since it was not studied in [11]. Also note that the results for DNMF_MM_EU are not provided. This is because we have yet to come up with an auxiliary function with a tractable form for the Euclidean distance case.

1) DATASET AND EXPERIMENTAL SETTINGS

We constructed the training and test datasets using speech signals excerpted from the Wall Street Journal (WSJ-0) corpus [22] and noise signals excerpted from the CHiME4 background noise database [23], which includes four types of noise recorded in a bus, cafe, pedestrian area, and street, respectively. The training dataset consisted of 600 utterances, each of which was created by mixing randomly selected utterances from `si_tr_s` and noise signals with signal-to-noise ratios (SNRs) set at $\{-5, 0, 5\}$dB. In the same way, we also created a validation dataset consisting of 90 utterances. Each of the four test datasets consisted of 100 utterances, half of which we created using speech signals in `si_tr_s` and the other half using speech signals of different speakers in `si_dt_05`. The SNRs for three of the four test datasets were set at $\{-5, 0, 5\}$ dB. and those for the remaining dataset were randomly set between $[-10, 10]$ dB.

All the audio signals were monaural and downsampled to 16 kHz. The STFT was computed using a Hanning window that was 32-ms long with a 16-ms overlap. We used the same basis number k for speech and noise, i.e., $K^s = K^n = K$. In this task, we tested $K = \{25, 50, 100\}$. For $K = 100$, we evaluated the effectiveness of sparse regularization in the case of a large number of basis numbers by setting $\mu = \{0, 0.5, 1, 5, 10\}$. SNMF_KL was run for 100 iterations. For the DNMF algorithms, SNMF_KL was used for initialization. For the separation, the Wiener filter was constructed using the trained basis and activation matrices obtained using the standard NMF run for 100 iterations.

2) CONVERGENCE BEHAVIOR AND COMPUTATIONAL COST

We compared the convergence behaviors of the proposed algorithms, DNMF_MU_EU and DNMF_MU_KL, within the first 500 iterations. For all the algorithms, we used the same initialization and evaluated the signal-to-distortion ratio (SDR) [24] improvements. Two examples are shown in Fig. 2. As can be seen from the example when tested on bus noise with $k = 100$, DNMF_MU_EU and DNMF_MU_KL did not decrease the objective functions monotonically. This indeed shows that each update in the MU algorithms does not guarantee a decrease in the objective functions. It is also worth noting that the objective function value does not directly reflect the speech enhancement performance, as shown in the experimental results when tested on street noise with $k = 50$. According to the SDR results obtained with the validation dataset as well as the setting in [11], in the following experiments, we set the iteration number at 150 for the proposed algorithms and 25 for the MU algorithms.

We compared the computational times of all the algorithms with $k = 50$ using the training data with a length of about one hour. The algorithms were implemented using MATLAB and run on an Intel Xeon Gold 5120 @2.2GHz processor. Table 1 shows the average computational time for updating $\mathbf{B}$ or $\mathbf{W}$ at each iteration and that of the entire process. Note that the total time of DNMF includes the time of computing $\tilde{\mathbf{B}}$ for initialization and $\hat{\mathbf{H}}$. Note that the time complexity of the proposed algorithm is $O(\Omega KTL^2)$, whereas that of the standard NMF and DNMF algorithms with multiplicative update rules is $O(\Omega KTL)$. Since L was 2 in the speech enhancement task, it did not have a significant impact on the computation time. Rather, the increase in the number of iterations in the proposed algorithm led to an increase in the total computation time.

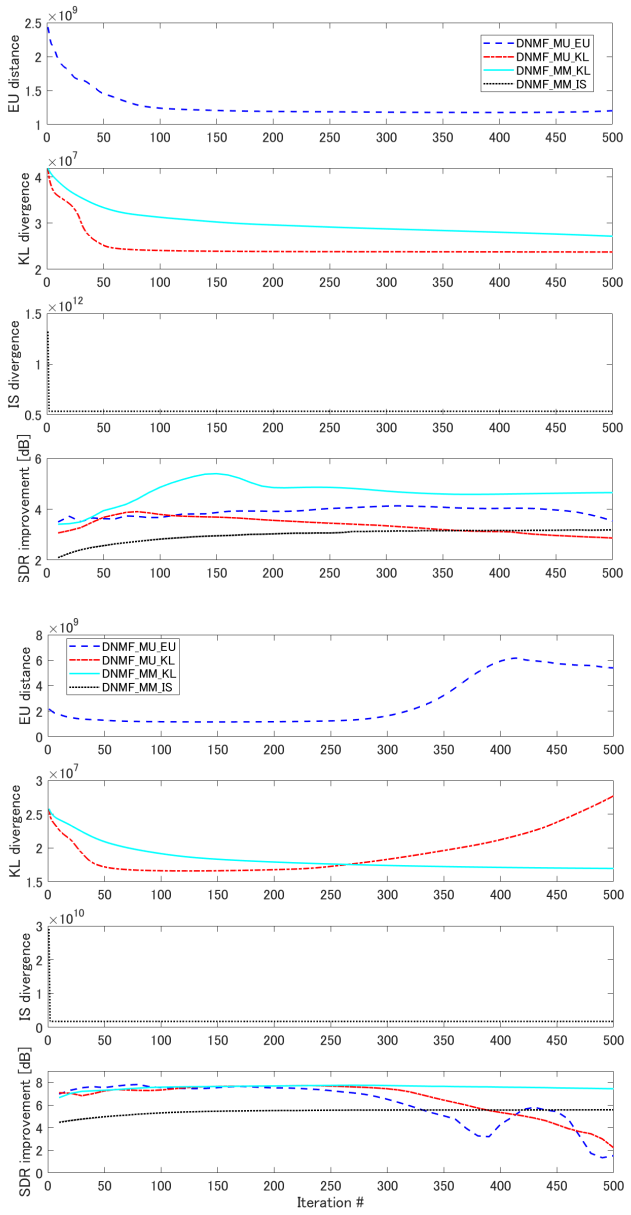

FIGURE 2. Convergence behavior and corresponding SDR improvements obtained with each method in street noise with the $K = 50$ case (top) and bus noise with basis number $K = 100$ case (bottom).

TABLE 1. Comparison of computational times [sec] with basis number $K = 50$.

Method	Time / Iteration	Total time
SNMF_EU	0.1468	62.6800
SNMF_KL	0.4687	192.3910
SNMF_IS	0.2820	121.4615
DNMF_MU_EU	1.3460	256.3109
DNMF_MU_KL	0.6234	236.6248
DNMF_MM_KL	1.4947	434.0287
DNMF_MM_IS	1.5184	437.2275

3) SPEECH ENHANCEMENT PERFORMANCE

The speech enhancement performance was numerically evaluated in terms of SDRs, signal-to-interference ratios (SIRs), and signal-to-artificial ratios (SARs) [24]. Table 2 shows the average SDRs taken over all the test data with basis number $K = \{25, 50, 100\}$. For each noise type with different k, we conducted 5 trials with different initializations. The average input SDR of the test data was about 0.063 dB. As Table 2 shows, increasing the bases did not always lead to an improvement in speech enhancement performance. Comparing the results of the standard NMF and DNMF algorithms, we found that the latter outperformed the former. This indicates the effectiveness of the ability to learn discriminative bases. Furthermore, the proposed algorithm performed best among all the algorithms based on the same divergence measure. Table 3 shows the average SDRs, SIRs, and SARs evaluated using $K = 25$ with various input SNRs. These results were averaged over the four noise types. As the results show, DNMF_MM_KL performed best among all the algorithms in terms of the SDR and SIR. Specifically, it achieved about 1.2-dB improvements over DNMF_MU_EU and DNMF_MU_KL, and about 1.7-dB improvements over SNMF_KL. This shows that the proposed algorithm with the KL divergence criterion had a better ability

TABLE 2. SDR [dB] obtained with $K = \{25, 50, 100\}$, average over all the test datasets (four types noise) with five random initializations. The average input SDR was about 0.063 dB.

Method	Basis number k		
	25	50	100
SNMF_EU	2.55	2.52	2.53
SNMF_KL	2.42	2.38	2.44
SNMF_IS	2.11	1.83	1.68
DNMF_MU_EU	2.87	2.69	2.71
DNMF_MU_KL	2.52	2.62	2.63
DNMF_MM_KL	**3.49**	**3.39**	**3.36**
DNMF_MM_IS	2.10	2.26	2.34

TABLE 3. From top to bottom: average SDRs, SIRs, SARs [dB] over four types of noise with basis number $K = 25$.

	Method	Input SNR [dB]				
		-5	0	5	[-10,10]	Avg
SDR	unprocessed	-4.92	0.05	5.03	0.09	0.06
	SNMF_EU	-2.01	2.69	7.07	2.48	2.55
	SNMF_KL	-2.04	2.63	6.79	2.29	2.42
	SNMF_IS	-2.35	2.35	6.45	2.00	2.11
	DNMF_MU_EU	-1.61	3.02	7.30	2.77	2.87
	DNMF_MU_KL	-1.99	2.73	6.94	2.41	2.52
	DNMF_MM_KL	**-0.92**	**3.78**	**7.77**	**3.34**	**3.49**
	DNMF_MM_IS	-2.35	2.19	6.52	2.04	2.10
SIR	SNMF_EU	-1.18	3.73	8.75	3.87	3.79
	SNMF_KL	-1.04	3.94	8.99	4.05	3.94
	SNMF_IS	-0.87	4.22	9.22	4.26	4.21
	DNMF_MU_EU	-0.41	4.49	9.51	4.61	4.55
	DNMF_MU_KL	-1.01	3.94	8.74	3.91	3.90
	DNMF_MM_KL	**0.75**	**5.77**	**10.53**	**5.73**	**5.70**
	DNMF_MM_IS	-1.29	3.44	8.22	3.51	3.47
SAR	SNMF_EU	**10.04**	**11.62**	**13.06**	**11.62**	**11.58**
	SNMF_KL	8.90	10.33	11.56	10.24	10.26
	SNMF_IS	7.11	8.77	10.48	8.83	8.80
	DNMF_MU_EU	8.85	10.60	12.35	10.64	10.61
	DNMF_MU_KL	9.26	10.97	12.56	10.91	10.93
	DNMF_MM_KL	7.88	10.00	11.99	9.94	9.95
	DNMF_MM_IS	8.65	10.40	12.43	10.61	10.52

TABLE 4. SDR (top) and SIR (bottom) improvement [dB] achieved for the four-source separation task averaged over five random initializations. Bold font shows the highest average score for each song.

	ID	SNMF_KL					DNMF_MM_KL				
		Bass	Drums	Other	Vocals	Avg	Bass	Drums	Other	Vocals	Avg
SDRi [dB]	060	8.09	8.06	7.52	12.03	8.92	8.86	7.72	8.02	12.23	**9.21**
	070	5.62	10.06	5.17	7.69	7.13	6.00	9.83	6.24	6.60	**7.17**
	080	10.37	11.56	4.55	7.53	8.50	11.70	11.87	4.80	7.93	**9.07**
	090	7.42	8.63	6.44	10.03	8.13	8.02	8.31	6.78	10.15	**8.32**
	100	8.07	9.57	7.60	10.19	8.86	8.33	10.07	8.92	10.00	**9.33**
	shared	5.96	8.08	4.49	7.17	6.42	6.64	8.35	6.01	6.98	**6.99**
	ID	SNMF_KL					DNMF_MM_KL				
		Bass	Drums	Other	Vocals	Avg	Bass	Drums	Other	Vocals	Avg
SIRi [dB]	060	10.82	12.06	9.94	14.52	11.83	12.09	11.29	10.40	14.88	**12.16**
	070	9.40	15.53	6.73	10.76	10.61	10.00	15.33	8.36	8.66	10.59
	080	12.86	15.26	6.38	10.85	11.34	14.87	15.73	6.52	11.43	**12.14**
	090	9.90	11.55	8.89	12.96	10.82	11.25	10.63	9.39	12.86	**11.03**
	100	11.67	13.51	9.89	13.89	12.24	11.20	14.39	11.38	13.16	**12.53**
	shared	8.67	10.77	6.40	9.56	8.85	8.93	11.50	8.89	8.72	**9.51**
	ID	SNMF_IS					DNMF_MM_IS				
		Bass	Drums	Other	Vocals	Avg	Bass	Drums	Other	Vocals	Avg
SDRi [dB]	060	7.41	7.87	7.31	11.42	8.50	5.48	4.36	3.49	8.25	5.39
	070	5.30	9.91	6.98	7.13	7.10	5.83	6.48	4.47	6.44	5.80
	080	9.06	11.32	3.83	5.96	7.54	1.97	5.68	0.17	3.82	2.91
	090	5.98	8.43	5.78	9.90	7.52	0.64	3.25	1.50	-1.88	0.89
	100	7.74	10.97	7.34	8.43	8.62	4.33	5.41	3.88	7.80	5.35
	shared	5.16	8.23	4.28	6.37	6.01	3.30	4.54	1.63	4.55	3.50
	ID	SNMF_IS					DNMF_MM_IS				
		Bass	Drums	Other	Vocals	Avg	Bass	Drums	Other	Vocals	Avg
SIRi [dB]	060	10.26	12.54	9.78	14.07	11.66	8.10	6.42	5.15	11.67	7.83
	070	8.63	16.83	8.21	10.74	**11.10**	9.84	9.65	6.02	9.64	8.79
	080	12.03	15.24	5.61	8.72	10.40	3.79	9.20	3.75	9.16	6.47
	090	8.46	11.35	7.89	14.62	10.58	2.60	5.22	5.97	7.07	5.22
	100	10.29	16.19	9.76	12.20	12.11	6.74	7.96	5.91	11.60	8.05
	shared	7.87	12.68	6.36	9.13	9.01	5.04	6.59	3.33	7.16	5.53

TABLE 5. SDR [dB] obtained with $\mu = \{0, 0.5, 1, 5, 10\}$ and $K = 100$ average over all the test datasets with 5 random initializations. Bold font shows the highest score for each method.

Method	μ				
	0	0.5	1	5	10
SNMF_EU	2.53	**2.66**	2.64	2.37	2.14
SNMF_KL	2.44	**2.48**	2.41	2.40	2.40
SNMF_IS	1.68	**1.98**	1.96	1.80	1.78
DNMF_MU_EU	2.71	**3.62**	3.52	3.12	2.88
DNMF_MU_KL	2.63	**3.78**	3.77	3.77	3.77
DNMF_MM_KL	3.36	**3.88**	3.87	3.87	3.87
DNMF_MM_IS	**2.34**	1.99	1.93	1.71	1.65

to learn discriminative bases than the baseline algorithms did. However, the SARs obtained with the proposed algorithms tended to be lower than those obtained with the baseline algorithms.

We also evaluated the effectiveness of sparse regularization. The results are shown in Table 5. We found that $\mu = 0.5$ achieved the best score for each method except for DNMF_MM_IS, where the best performance was obtained without sparse regularization. DNMF_MM_KL outperformed the other methods regardless of the sparse regularization.

B. SINGLE-CHANNEL SOURCE SEPARATION

We also evaluated the performance of the proposed algorithms in source separation tasks.

1) DATASET AND EXPERIMENTAL CONDITIONS

We excerpted five recordings from Demixing Secrets Dataset 100 (DSD100) [25], which was used in the SiSEC 2016 MUS task. Each of the recordings consisted of four sources, namely bass, drums, vocals, and the other. The task was thus a four-source separation problem, namely $\alpha_l = 1, l = \{1, 2, 3, 4\}$. Each of the recordings was about four to five minutes long. We divided each recording into two segments, namely a training data segment and a test data segment.

Here, we conducted two experiments. In the first experiment, we trained the basis matrix separately using the training data segment of each recording and tested on the test data segment. In the second experiment, we trained a shared basis matrix using the collection of the training data segments of all the recordings and tested on the test data segment of each recording. As in the speech enhancement task, we used monaural audio signals and downsampled them to 16 kHz. The STFT was computed using a 256-ms long Hanning window with 1/2 window overlap. Considering the characteristics of the four sources, we set the basis number at [10, 10, 15, 15] for bass, drums, vocals, and other, respectively, for the first experiment and [20, 20, 50, 50] for the second experiment. For each experiment, we also ran five trails with random initialization and evaluated the average SDR and SIR scores. SNMF_KL was run for 100 iterations and was used as the initialization for the DNMF algorithms. In the source separation

experiments, we set the number of iterations for training $\tilde{\mathbf{W}}$, at 25.

2) EXPERIMENTAL RESULTS

Fig. 3 shows an example of the convergence behavior of the proposed algorithms. Table 4 shows the SDR and SIR improvements [dB]. As the results show, the proposed algorithm with KL divergence outperformed SNMF_KL for most of the test data. It is interesting to note that even though in the first experiment the standard NMF was relatively advantageous as regards the training condition, DNMF_MM_KL still obtained higher SDR and SIR scores.

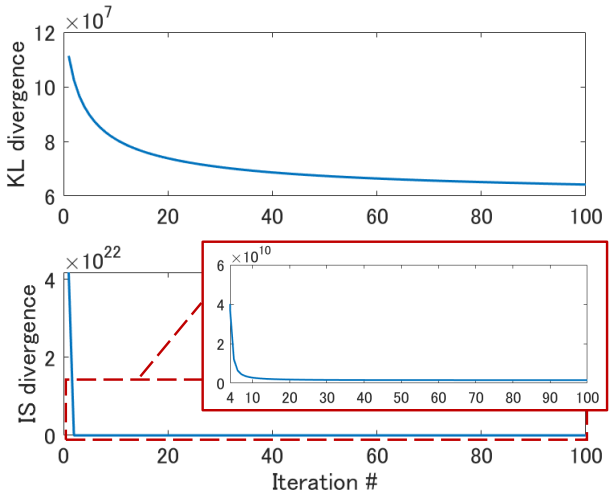

FIGURE 3. Example of convergence behavior of proposed algorithms for source separation task with shared basis matrix.

In the speech enhancement task, we confirmed that the proposed algorithm performed slightly better than the standard NMF method under the IS divergence criterion. However, this was found not to be the case for the source separation task. This implies that the discriminative basis training and/or MM strategies were less effective for the IS divergence than for the KL divergence. The reason for this will be examined more closely in our future work.

V. CONCLUSION

DNMF is noteworthy in that it directly uses the reconstruction errors of separated signals as the training criteria, which eliminates the inconsistency between the objctive functions for training and separation in the conventional NMF method and can increase the discriminative power of the trained basis. However, such training criteria cause difficulty in optimization. This paper derived a novel majorizer for the objective function of DNMF and successfully developed an MM algorithm that is guaranteed to converge to a stationary point. Experimental results showed that the proposed algorithm with the KL divergence criterion achieved significant improvements in terms of the SDR and SIR over standard NMF and DNMF using the multiplicative update algorithm.

ACKNOWLEDGMENT

This article was presented in part at HSCMA 2017 as a conference paper [1].

REFERENCES

[1] L. Li, H. Kameoka, and S. Makino, "Discriminative non-negative matrix factorization with majorization-minimization," in *Proc. Hands-Free Speech Commun. Microphone Arrays (HSCMA)*, 2017, pp. 141–145.
[2] D. Wang and G. J. Brown, *Computational Auditory Scene Analysis: Principles, Algorithms, and Applications*. Hoboken, NJ, USA: Wiley, 2006.
[3] S. T. Roweis, "One microphone source separation," in *Proc. NIPS*, 2001, pp. 793–799.
[4] G. Cauwenberghs, "Monaural separation of independent acoustical components," in *Proc. ISCAS*, vol. 5, May 1999, pp. 6–65.
[5] F. R. Bach and M. I. Jordan, "Blind one-microphone speech separation: A spectral learning approach," in *Proc. NIPS*, 2005, pp. 65–72.
[6] P. Smaragdis, B. Raj, and M. Shashanka, "Supervised and semisupervised separation of sounds from single-channel mixtures," in *Proc. Integr. Comput.-Aided Eng.*, 2007, pp. 414–421.
[7] P.-S. Huang, M. Kim, M. Hasegawa-Johnson, and P. Smaragdis, "Deep learning for monaural speech separation," in *Proc. ICASSP*, 2014, pp. 1562–1566.
[8] J. R. Hershey, Z. Chen, J. Le Roux, and S. Watanabe, "Deep clustering: Discriminative embeddings for segmentation and separation," in *Proc. ICASSP*, Aug. 2016, pp. 31–35.
[9] D. Yu, M. Kolbæk, Z.-H. Tan, and J. Jensen, "Permutation invariant training of deep models for speaker-independent multi-talker speech separation," in *Proc. ICASSP*, 2017, pp. 241–245.
[10] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. NIPS*, 2001, pp. 556–562.
[11] F. Weninger, J. Le Roux, J. R. Hershey, and S. Watanabe, "Discriminative NMF and its application to single-channel source separation," in *Proc. Interspeech*, 2014, pp. 865–869.
[12] E. M. Grais and H. Erdogan, "Discriminative nonnegative dictionary learning using cross-coherence penalties for single channel source separation," in *Proc. Interspeech*, 2013, pp. 808–812.
[13] G. Bao, Y. Xu, and Z. Ye, "Learning a discriminative dictionary for single-channel speech separation," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 7, pp. 1130–1138, Jul. 2014.
[14] N. Boulanger-Lewandowski, Y. Bengio, and P. Vincent, "Discriminative non-negative matrix factorization for multiple pitch estimation," in *Proc. ISMIR*, 2012, pp. 205–210.
[15] Z. Wang and F. Sha, "Discriminative non-negative matrix factorization for single-channel speech separation," in *Proc. ICASSP*, 2014, pp. 3749–3753.
[16] K. Kwon, J. W. Shin, and N. S. Kim, "Target source separation based on discriminative nonnegative matrix factorization incorporating cross-reconstruction error," *Proc. IEICE Trans. Inf. Syst.*, vol. 98, no. 11, pp. 2017–2020, 2015.
[17] P. Sprechmann, A. M. Bronstein, and G. Sapiro, "Supervised non-Euclidean sparse nmf via bilevel optimization with applications to speech enhancement," in *Proc. HSCMA*, 2014, pp. 11–15.
[18] C. Févotte, N. Bertin, and J.-L. Durrieu, "Nonnegative matrix factorization with the itakura-saito divergence: With application to music analysis," *Neural Comput.*, vol. 21, no. 3, pp. 793–830, 2009.
[19] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc. Ser. B Methodol.*, vol. 39, pp. 1–38, Sep. 1977.
[20] M. Nakano, H. Kameoka, J. Le Roux, Y. Kitano, N. Ono, and S. Sagayama, "Convergence-guaranteed multiplicative algorithms for nonnegative matrix factorization with beta-divergence," in *Proc. MLSP*, 2010, pp. 283–288.
[21] H. Kameoka, "Non-negative matrix factorization and its variants for audio signal processing," in *Applied Matrix and Tensor Variate Data Analysis*, T. Sakata, Ed. Tokyo, Japan: Springer, 2016.
[22] S. J. Garofolo, *CSR-I (WSJ0) Complete LDC93S6A. Web Download*. Philadelphia, PA, USA: Linguistic Data Consortium, 1993.
[23] E. Vincent, S. Watanabe, A. A. Nugraha, J. Barker, and R. Marxer. *The 4th CHiME Speech Separation and Recognition Challenge*. Accessed: Dec. 19, 2020. [Online]. Available: http://spandh.dcs.shef.ac.uk/chime_challenge/chime2016/

[24] E. Vincent, R. Gribonval, and C. Fevotte, "Performance measurement in blind audio source separation," *IEEE Trans. Audio, Speech, Language Process.*, vol. 14, no. 4, pp. 1462–1469, Jul. 2006.
[25] A. Liutkus, F.-R. Stöter, Z. Rafii, D. Kitamura, B. Rivet, N. Ito, N. Ono, and J. and Fontecave, "The 2016 signal separation evaluation campaign," in *Proc. LVA/ICA*, 2017, pp. 323–332.

LI LI (Student Member, IEEE) received the B.E. degree from the Shanghai University of Finance and Economics, China, in 2014, and the M.S. degree from the University of Tsukuba, Japan, in 2018, where she is currently pursuing the Ph.D. degree with the Graduate School. Since 2018, she has been a Research Fellow of the Japan Society of Promotion of Science. Her research interests include audio and speech signal processing, source separation, and machine learning. She received the 13th Student Presentation Award from the Acoustical Society of Japan and the second IEEE Signal Processing Society Tokyo Joint Chapter Student Award.

HIROKAZU KAMEOKA (Senior Member, IEEE) received the B.E., M.S., and Ph.D. degrees from The University of Tokyo, Japan, in 2002, 2004, and 2007, respectively. From 2011 to 2016, he was an Adjunct Associate Professor with The University of Tokyo. He is currently a Senior Distinguished Researcher with NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation. He is also an Adjunct Associate Professor with the National Institute of Informatics. He is the author or coauthor of about 150 articles in journal articles and peer-reviewed conference proceedings. His research interests include audio, speech, and music signal processing, and machine learning. He has been a member of the IEEE Audio and Acoustic Signal Processing Technical Committee, since 2017, and a member of the IEEE Machine Learning for Signal Processing Technical Committee, since 2019. He has received 17 awards, including the IEEE Signal Processing Society 2008 SPS Young Author Best Paper Award. Since 2015, he has been an Associate Editor of the IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING.

SHOJI MAKINO (Fellow, IEEE) received the B.E., M.E., and Ph.D. degrees from Tohoku University, Japan, in 1979, 1981, and 1993, respectively.

He joined NTT, in 1981. He is currently a Professor with the University of Tsukuba. He is the author or coauthor of more than 200 papers in journals and conference proceedings. He is responsible for more than 150 patents. His research interests include adaptive filtering technologies, realization of acoustic echo cancellation, blind source separation of convolutive mixtures of speech, and acoustic signal processing for speech and audio applications.

Dr. Makino was a member of the IEEE Jack S. Kilby Signal Processing Medal Committee, from 2015 to 2018, and the James L. Flanagan Speech & Audio Processing Award Committee, from 2008 to 2011. He is an IEICE Fellow, a Board Member of the ASJ, and a Member of EURASIP. He received the ICA Unsupervised Learning Pioneer Award, in 2006, the IEEE MLSP Competition Award, in 2007, the IEEE SPS Best Paper Award, in 2014, the Achievement Award for Science and Technology by the Minister of Education, Culture, Sports, Science and Technology, in 2015, the Hoko Award of the Hattori Hokokai Foundation, in 2018, the Outstanding Contribution Award of the IEICE, in 2018, the Technical Achievement Award of the IEICE, in 2017 and 1997, the Outstanding Technological Development Award of the ASJ, in 1995, and eight Best Paper Awards. He was the Chair of SPS Audio and Acoustic Signal Processing Technical Committee, from 2013 to 2014 and the Blind Signal Processing Technical Committee of the IEEE Circuits and Systems Society, from 2009 to 2010. He was the General Chair of IWAENC 2018, WASPAA2007, IWAENC2003, and the Organizing Chair of ICA2003. He is the designated Plenary Chair of ICASSP2012. He has served on IEEE SPS Board of Governors, from 2018 to 2020, Technical Directions Board, from 2013 to 2014, Awards Board, from 2006 to 2008, Conference Board, from 2002 to 2004, and Fellow Evaluation Committee, from 2018 to 2020. He was an Associate Editor of the IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, from 2002 to 2005, and the *EURASIP Journal on Advances in Signal Processing*, from 2005 to 2012. He was a Guest Editor of the Special Issue of the *IEEE Signal Processing Magazine*, from 2013 to 2014. He was a Keynote Speaker at ICA2007, a Tutorial Speaker at EMBC2013, Interspeech2011, and ICASSP2007. From 2009 to 2010, he was an IEEE SPS Distinguished Lecturer.