DISCRIMINATIVE NON-NEGATIVE MATRIX FACTORIZATION WITH MAJORIZATION-MINIMIZATION

Li Li[1], *Hirokazu Kameoka*[2] *and Shoji Makino*[1]

[1] University of Tsukuba, Japan
[2] NTT Communication Science Laboratories, NTT Corporation, Japan
lili@mmlab.cs.tsukuba.ac.jp, kameoka.hirokazu@lab.ntt.co.jp, maki@tara.tsukuba.ac.jp

ABSTRACT

Non-negative matrix factorization (NMF) is a powerful approach to single channel audio source separation. In a supervised setting, NMF is first applied to train the basis spectra of each sound source. At test time, NMF is applied to the spectrogram of a mixture signal using the pretrained spectra. The source signals can then be separated out using a Wiener filter. A typical way to train the basis spectra of each source is to minimize the objective function of NMF. However, the basis spectra obtained in this way do 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, a framework called discriminative NMF (DNMF) has recently been proposed. In in this work a multiplicative update algorithm was proposed for the basis training, however one drawback is that the convergence is not guaranteed. To overcome this drawback, this paper proposes using a majorization-minimization principle to develop a convergence-guaranteed algorithm for DNMF. Experimental results showed that the proposed algorithm outperformed standard NMF and DNMF using a multiplicative update algorithm as regards both the signal-to-distortion and signal-to-interference ratios.

Index Terms— Discriminative non-negative matrix factorization, majorization-minimization, single channel, speech enhancement

1. INTRODUCTION

Single channel audio source separation is the challenging task of extracting individual source signals from a monaural recording of a mixture signal. Non-negative matrix factorization (NMF) [1, 2] has attracted a lot of attention in recent years after being proposed as a powerful approach for audio source separation.

Factorizing an observed magnitude (or power) spectrogram of a mixture signal, interpreted as a non-negative matrix, into the product of two non-negative matrices amounts to approximating the observed spectra by a linear sum of basis spectra scaled by time-varying amplitudes. In a supervised/semi-supervised 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 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. To address this, a framework called [1]discriminative NMF (DNMF) has recently been proposed [3]. 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, as shown later, 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 [3] Weninger proposed a multiplicative update algorithm for the basis training, however one drawback is that the convergence is not guaranteed. To overcome this drawback, this paper proposes using a majorization-minimization principle to derive a convergence-guaranteed algorithm for DNMF.

2. DISCRIMINATIVE NON-NEGATIVE MATRIX FACTORIZATION

2.1. Standard NMF

We start by reviewing standard NMF for single channel source separation. Let us denote the number of sources by L, and an observed power spectrogram by $\boldsymbol{Y} = (Y_{\omega,t})_{\Omega\times T} \in \mathbb{R}^{\geq 0,\Omega\times T}$, where ω and t are frequency and time indices. Supervised NMF factorizes an observed spectrogram $\boldsymbol{Y}$ of a mixture signal into the product of a non-negative basis matrix $\boldsymbol{W} = [\boldsymbol{W}^1, \boldsymbol{W}^2, \ldots .\boldsymbol{W}^L]$ and a non-negative coefficient (activation) matrix $\boldsymbol{H} = [\boldsymbol{H}^1; \boldsymbol{H}^2; \ldots; \boldsymbol{H}^l]$, where

This work was supported by JSPS KAKENHI under Grants 26730100 and 26280060, and SECOM Science and Technology Foundation.

[1] While many methods called "discriminative NMF" have been proposed with the aim of enhancing the discriminative power of the basis spectra [3]–[10], here we use this term in relation to the work done by Weninger [3].

978-1-5090-5925-6/17/$31.00 ©2017 IEEE

$\boldsymbol{W}^l = (W^l_{\omega,k})_{\Omega \times K^l} \in \mathbb{R}^{\geq 0, \Omega \times K^l}$ with $l = 1, 2, \ldots, L$ is pretrained using the spectrograms of training samples $\boldsymbol{S}^l = (S^l_{\omega,t})_{\Omega \times T}$. A typical criterion for this is

$$\boldsymbol{W}^l = \underset{\boldsymbol{W}^l}{\operatorname{argmin}}\, \mathcal{D}(\boldsymbol{S}^l | \boldsymbol{W}^l \boldsymbol{H}^l), \tag{1}$$

where $\mathcal{D}$ is a cost function that measures the difference between $\boldsymbol{S}^l$ and $\boldsymbol{W}^l\boldsymbol{H}^l$. At test time, $\boldsymbol{W}$ is fixed at the pretrained basis spectra and the activation matrix $\boldsymbol{H}$ is estimated so that the objective funtion

$$\boldsymbol{H} = \underset{\boldsymbol{H}}{\operatorname{argmin}}\, \mathcal{D}(\boldsymbol{Y} | \boldsymbol{W}\boldsymbol{H}), \tag{2}$$

is minimized subject to non-negativity. When $\mathcal{D}$ is a generalized Kullback-Leibler (KL) divergence, the objective function can be written as

$$\begin{aligned} &\mathcal{D}_{\mathrm{KL}}(\boldsymbol{Y}|\boldsymbol{W}\boldsymbol{H}) \\ &= \sum_{\omega,t} \left(Y_{\omega,t} \log \frac{Y_{\omega,t}}{[\boldsymbol{W}\boldsymbol{H}]_{\omega,t}} - Y_{\omega,t} + [\boldsymbol{W}\boldsymbol{H}]_{\omega,t} \right), \end{aligned} \tag{3}$$

where $[\cdot]_{i,j}$ denotes the $\{i,j\}$-th element of a matrix. Once $\boldsymbol{W}$ and $\boldsymbol{H}$ are obtained, the signals can be separated by a Wiener filter constructed using the estimated power spectrogram as follows

$$\hat{S}^l = \frac{\boldsymbol{W}^l \boldsymbol{H}^l}{\boldsymbol{W}\boldsymbol{H}} \otimes \boldsymbol{Y}, \tag{4}$$

where $\otimes$ and $\div$ denote elementwise multiplication and division.

2.2. DNMF and multiplicative update algorithm

Instead of using (1), Weninger [3] proposed directly using the reconstruction error of the separated signals as an objective function for the basis training

$$\mathcal{J} = \sum_l \alpha_l \mathcal{D}_{\mathrm{KL}}(\boldsymbol{S}^l | \hat{\boldsymbol{S}}^l), \tag{5}$$

where $\alpha_l \geq 0$ weighs the importance of source signal l. This framework is called discriminative NMF by analogy with the discriminative models for classification or regression.

For convenience of explanation, here we consider a speech enhancement problem where the sources are speech and noise, $\boldsymbol{S}^{\mathrm{s}} = (S^{\mathrm{s}}_{\omega,t})_{\Omega \times T} \in \mathbb{R}^{\geq 0, \Omega \times T}$ denotes training samples of clean speech and $\boldsymbol{S}^{\mathrm{n}} = (S^{\mathrm{n}}_{\omega,t})_{\Omega \times T} \in \mathbb{R}^{\geq 0, \Omega \times T}$ denotes training samples of noise. Since we are concerned with reconstructing clean speech as well as possible, we set α at 1 for speech and 0 for noise. In the same way as in 2.1, we use a Wiener filter to separate the spectrogram of speech $\hat{\boldsymbol{S}}^{\mathrm{s}}$ from the spectrogram of a mixture signal $\boldsymbol{M} = (M_{\omega,t})_{\Omega \times T} \in \mathbb{R}^{\geq 0, \Omega \times T}$. The discriminative training problem can be cast as an optimization problem

$$\begin{aligned} &\text{minimize} \quad f(\boldsymbol{W}, \boldsymbol{H}) = \mathcal{D}_{\mathrm{KL}} \left(\boldsymbol{S}^{\mathrm{s}} \,\middle|\, \frac{\boldsymbol{W}^{\mathrm{s}} \boldsymbol{H}^{\mathrm{s}}}{\boldsymbol{W}\boldsymbol{H}} \otimes \boldsymbol{M} \right), \\ &\text{subject to} \quad \forall k,\ \sum_{\omega} W_{\omega,k} = 1, \end{aligned} \tag{6}$$

where the concatenated basis matrix $\boldsymbol{W} = [\boldsymbol{W}^s, \boldsymbol{W}^n]$ consists of a total of K basis vectors with K_{s} for speech and K_{n} for noise. It should be noted that in [3] a sparse regularization term [9] is used to promote the sparsity of $\boldsymbol{H}$.

An inspection of (1) and (6) shows that the training criterion for DNMF is more analytically complex than the objective function of standard NMF. In [3], Weninger proposed a multiplicative update algorithm for solving the above optimization problem. The algorithm consists of two stages: First, the activation matrix $\boldsymbol{H}$ is obtained by solving (2) using NMF. The basis matrix $\boldsymbol{W}$ is then iteratively updated according to the following rules

$$\boldsymbol{W}^{\mathrm{s}} \leftarrow \boldsymbol{W}^{\mathrm{s}} \otimes \frac{\frac{\boldsymbol{S}^{\mathrm{s}} \otimes \boldsymbol{W}^{\mathrm{s}} \boldsymbol{H}^{\mathrm{s}}}{\boldsymbol{W}\boldsymbol{H} \otimes \boldsymbol{W}^{\mathrm{s}} \boldsymbol{H}^{\mathrm{s}}} \boldsymbol{H}^{\mathrm{s}\mathsf{T}}}{\frac{\boldsymbol{M} \otimes \boldsymbol{W}^{\mathrm{n}} \boldsymbol{H}^{\mathrm{n}}}{(\boldsymbol{W}\boldsymbol{H})^2} \boldsymbol{H}^{\mathrm{s}\mathsf{T}}},$$

$$\boldsymbol{W}^{\mathrm{n}} \leftarrow \boldsymbol{W}^{\mathrm{n}} \otimes \frac{\frac{\boldsymbol{M} \otimes \boldsymbol{W}^{\mathrm{s}} \boldsymbol{H}^{\mathrm{s}}}{(\boldsymbol{W}\boldsymbol{H})^2} \boldsymbol{H}^{\mathrm{n}\mathsf{T}}}{\frac{\boldsymbol{S}^{\mathrm{s}}}{\boldsymbol{W}\boldsymbol{H}} \boldsymbol{H}^{\mathrm{n}\mathsf{T}}}.$$

Here, the multiplicative factors are given by dividing the negative parts by the positive parts of the partial derivative of f with respect to $\boldsymbol{W}^{\mathrm{s}}$ and $\boldsymbol{W}^{\mathrm{n}}$ in the same way as in [11]. Although this algorithm is easy to implement and works reasonably well in practice, one drawback is that convergence to a stationary point of f is not guaranteed.

3. DNMF WITH MAJORIZATION-MINIMIZATION

3.1. Majorization-minimization principle

To address the problem of the multiplicative update algorithm shown above, we derive a novel convergence-guaranteed algorithm for DNMF based on a majorization-minimization (MM) principle. 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. Suppose $F(\Theta)$ is an objective function that we wish to minimize with respect to Θ. A majorizer $F^+(\Theta, \alpha)$ is then defined as a function satisfying $F(\Theta) = \min_\alpha F^+(\Theta, \alpha)$, where α is called an auxiliary parameter. An algorithm that consists of iteratively minimizing $F^+(\Theta, \alpha)$ with respect to Θ and α is guaranteed to converge to a stationary point of the objective function. It should be noted that this concept is adopted in many existing algorithms. For example, the expectation-maximization (EM) algorithm [13] builds a surrogate for a likelihood function of latent variable models by using Jensen's inequality. It is also well known for its use in devising an algorithm for NMF [1, 12]. In general, if we can build a tight majorizer that is easy to optimize for the objective function of some optimization problem, we can expect to obtain a fast-converging algorithm.

3.2. Majorizer for objective function

In this section, we derive a novel majorizer for the objective function (6). First, let us focus on the term

$$\frac{\sum_{k=1}^{K^{\mathrm{s}}} W^{\mathrm{s}}_{\omega,k} H^{\mathrm{s}}_{k,t}}{\sum_{k=1}^{K} W_{\omega,k} H_{k,t}}. \tag{7}$$

To construct a majorizer for this term, we can use the following inequality:

Lemma 1. *For $a \in \mathbb{R}^{>0}$ and $b \in \mathbb{R}^{>0}$,*

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

The equality holds if and only if $\lambda = 1/(ab)$.

Proof of Lemma 1. For $a, b, \lambda \in \mathbb{R}^{>0}$,

$$\begin{aligned} \lambda \left(a - \frac{1}{\lambda b} \right)^2 &= \lambda \left(a^2 - 2\frac{a}{\lambda b} + \frac{1}{\lambda^2 b^2} \right) \geq 0 \\ \Rightarrow \frac{a}{b} &\leq \frac{\lambda a^2}{2} + \frac{1}{2\lambda b^2}. \end{aligned} \tag{8}$$

The equality holds if and only if $a - 1/(\lambda b) = 0$. □

Since $M_{\omega,t}$ is non-negative, we can construct an upper bound for $\frac{\boldsymbol{W}^s \boldsymbol{H}^s}{\boldsymbol{W}\boldsymbol{H}} \otimes \boldsymbol{M}$ according to the above lemma,

$$\begin{aligned} &\mathcal{D}_{\mathrm{KL}} \left(\boldsymbol{S}^{\mathrm{s}} \middle| \frac{\boldsymbol{W}^{\mathrm{s}} \boldsymbol{H}^{\mathrm{s}}}{\boldsymbol{W}\boldsymbol{H}} \otimes \boldsymbol{M} \right) \\ &\overset{c}{=} \sum_{\omega,t} \left(- S^{\mathrm{s}}_{\omega,t} \log G^{\mathrm{s}}_{\omega,t} + S^{\mathrm{s}}_{\omega,t} \log G_{\omega,t} + \frac{G^{\mathrm{s}}_{\omega,t}}{G_{\omega,t}} M_{\omega,t} \right) \\ &\leq \sum_{\omega,t} \Bigg(- S^{\mathrm{s}}_{\omega,t} \log G^{\mathrm{s}}_{\omega,t} + S^{\mathrm{s}}_{\omega,t} \log G_{\omega,t} \\ &\qquad + \frac{\lambda_{\omega,t} M_{\omega,t} {G^{\mathrm{s}}_{\omega,t}}^2}{2} + \frac{M_{\omega,t}}{2\lambda_{\omega,t} G^2_{\omega,t}} \Bigg), \end{aligned} \tag{9}$$

where $=^c$ denotes equality up to a constant term, $G^{\mathrm{s}}_{\omega,t} = \sum_{k=1}^{K^{\mathrm{s}}} W^{\mathrm{s}}_{\omega,k} H^{\mathrm{s}}_{k,t}$ and $G_{\omega,t} = \sum_{k=1}^{K} W_{\omega,k} H_{k,t}$. The equality of (9) holds if and only if

$$\lambda_{\omega,t} = \frac{1}{G^{\mathrm{s}}_{\omega,t} G_{\omega,t}}. \tag{10}$$

In the following, we construct a majorizer for each of the terms in the right-hand side of (9).

Since $S^{\mathrm{s}}_{\omega,t}$ is positive, $-S^{\mathrm{s}}_{\omega,t} \log G^{\mathrm{s}}_{\omega,t}$ is convex in $G^{\mathrm{s}}_{\omega,t}$. Hence, we can use Jensen's inequality to obtain a majorizer for this term as

$$-\log G^{\mathrm{s}}_{\omega,t} \leq -\sum_{k=1}^{K^{\mathrm{s}}} \gamma_{k,\omega,t} \log \frac{W^{\mathrm{s}}_{\omega,k} H^{\mathrm{s}}_{k,t}}{\gamma_{k,\omega,t}}, \tag{11}$$

where $\gamma_{k,\omega,t}$ is a positive weight that sums to unity. The equality of (11) holds if and only if

$$\gamma_{k,\omega,t} = \frac{W^{\mathrm{s}}_{\omega,k} H^{\mathrm{s}}_{k,t}}{\sum_{k'} W^{\mathrm{s}}_{\omega,k'} H^{\mathrm{s}}_{k',t}}. \tag{12}$$

The second term $S^{\mathrm{s}}_{\omega,t} \log G_{\omega,t}$ 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:

$$\log G_{\omega,t} \leq \sum_k \frac{W_{\omega,k} H_{k,t}}{\eta_{\omega,t}} + \log \eta_{\omega,t} - 1, \tag{13}$$

where $\eta_{\omega,t}$ is an arbitrary positive number. The equality of this inequality holds if and only if $\eta_{\omega,t} = G_{\omega,t}$. Since a quadratic function is convex, we can apply Jensen's inequality to the third term, which yields

$${G^{\mathrm{s}}_{\omega,t}}^2 \leq \sum_{k=1}^{K^{\mathrm{s}}} \frac{{W^{\mathrm{s}}_{\omega,k}}^2 {H^{\mathrm{s}}_{k,t}}^2}{\beta_{k,\omega,t}}, \tag{14}$$

where $\beta_{k,\omega,t} > 0$ is also a positive number that sums to unity, i.e., $\sum_k \beta_{k,\omega,t} = 1$. The equality of (14) holds if and only if

$$\beta_{k,\omega,t} = \frac{W^{\mathrm{s}}_{\omega,k} H^{\mathrm{s}}_{k,t}}{\sum_{k'=1}^{K^{\mathrm{s}}} W^{\mathrm{s}}_{\omega,k'} H^{\mathrm{s}}_{k',t}}. \tag{15}$$

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

$$\frac{1}{G^2_{\omega,t}} \leq \sum_k \frac{\theta^3_{k,\omega,t}}{W^2_{\omega,k} H^2_{k,t}}, \tag{16}$$

where $\theta_{k,\omega,t} > 0$ and $\sum_k \theta_{k,\omega,t} = 1$. We can confirm that the equality of this inequality holds if and only if

$$\theta_{k,\omega,t} = \frac{W_{\omega,k} H_{k,t}}{\sum_{k'} W_{\omega,k'} H_{k,t'}}. \tag{17}$$

From (9), (11), (14) and (16), we can construct a majorizer for the objective function as

$$\begin{aligned} &f(\boldsymbol{W}, \boldsymbol{H}) \leq f^{+}(\boldsymbol{W}, \boldsymbol{H}, \boldsymbol{\Gamma}) \\ &= -\sum_{\omega,t,k} S^{\mathrm{s}}_{\omega,t} \gamma_{k,\omega,t} \log \frac{W^{\mathrm{s}}_{\omega,k} H^{\mathrm{s}}_{k,t}}{\gamma_{k,\omega,t}} + \sum_{\omega,t,k} \frac{S^{\mathrm{s}}_{\omega,t} W_{\omega,k} H_{k,t}}{\eta_{\omega,t}} \\ &+ \sum_{\omega,t,k} \frac{\lambda_{\omega,t} M_{\omega,t}}{2\beta_{k,\omega,t}} {W^{\mathrm{s}}_{\omega,k}}^2 {H^{\mathrm{s}}_{k,t}}^2 + \sum_{\omega,t,k} \frac{M_{\omega,t} \theta^3_{k,\omega,t}}{2\lambda_{\omega,t} W^2_{\omega,k} H^2_{k,t}} + d, \end{aligned}$$

where $\boldsymbol{\Gamma}$ denotes a set of all the auxiliary variables, $\lambda_{\omega,t}$, $\gamma_{k,\omega,t}$, $\eta_{\omega,t}$, $\beta_{k,\omega,t}$ and $\theta_{k,\omega,t}$, and d denotes a constant term. This majorizer is particularly noteworthy in that it can be minimized analytically with respect to $W_{\omega,k}$ and $H_{k,t}$ since it is given as the sum of the reciprocal, logarithmic, first-order and second-order functions.

3.3. Update rules

We can obtain the update rules for $W_{\omega,k}$ and $H_{k,t}$ by setting the partial derivatives of the proposed majorizer with respect to $\boldsymbol{W}^{\mathrm{s}}$, $\boldsymbol{H}^{\mathrm{s}}$, $\boldsymbol{W}^{\mathrm{n}}$ and $\boldsymbol{H}^{\mathrm{n}}$ at zero. Thus, the update rules can be obtained as the positive solution of the following quartic

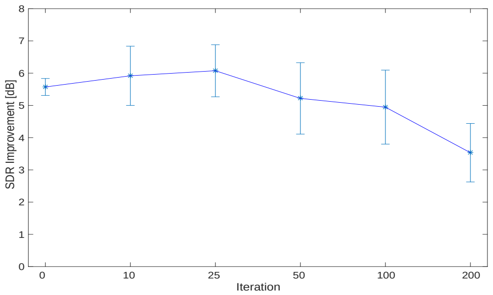

Fig. 1. Means and standard deviations of the SDR improvements [dB] obtained with the proposed algorithm with various iteration numbers (0, 10, 25, 50, 100, 200) initialized at 5 random values.

and cubic equations:

$$\sum_t \frac{\lambda_{\omega,t} M_{\omega,t}}{2\beta_{k,\omega,t}} {H^{\mathrm{s}}_{k,t}}^2 {W^{\mathrm{s}}_{\omega,k}}^4 + \sum_t \frac{S^{\mathrm{s}}_{\omega,t} H^{\mathrm{s}}_{k,t}}{\eta_{\omega,t}} {W^{\mathrm{s}}_{\omega,k}}^3 - \sum_t S^{\mathrm{s}}_{\omega,t} \gamma_{k,\omega,t} {W^{\mathrm{s}}_{\omega,k}}^2 - \sum_t \frac{M_{\omega,t} \theta^3_{k,\omega,t}}{2\lambda_{\omega,t} {H^{\mathrm{s}}_{k,t}}^2} = 0, \quad (18)$$

$$\sum_t \frac{S^{\mathrm{s}}_{\omega,t} H^{\mathrm{n}}_{k_n,t}}{\eta_{\omega,t}} {W^{\mathrm{n}}_{\omega,k}}^3 - \sum_t \frac{M_{\omega,t} \theta^3_{k,\omega,t}}{2\lambda_{\omega,t} {H^{\mathrm{n}}_{k,t}}^2} = 0, \quad (19)$$

$$\sum_\omega \frac{\lambda_{\omega,t} M_{\omega,t}}{2\beta_{k,\omega,t}} {W^{\mathrm{s}}_{\omega,k}}^2 {H^{\mathrm{s}}_{k,t}}^4 + \sum_\omega \frac{S^{\mathrm{s}}_{\omega,t} W^{\mathrm{s}}_{\omega,k}}{\eta_{\omega,t}} {H^{\mathrm{s}}_{k,t}}^3 - \sum_\omega S^{\mathrm{s}}_{\omega,t} \gamma_{k,\omega,t} {H^{\mathrm{s}}_{k,t}}^2 - \sum_\omega \frac{M_{\omega,t} \theta^3_{k,\omega,t}}{2\lambda_{\omega,t} {W^{\mathrm{s}}_{\omega,k}}^2} = 0, \quad (20)$$

$$\sum_\omega \frac{S^{\mathrm{s}}_{\omega,t} W^{\mathrm{n}}_{\omega,k}}{\eta_{\omega,t}} {H^{\mathrm{n}}_{k,t}}^3 - \sum_\omega \frac{M_{\omega,t} \theta^3_{k,\omega,t}}{2\lambda_{\omega,t} {W^{\mathrm{n}}_{\omega,k}}^2} = 0. \quad (21)$$

Although there are two quartic equations that must be solved, it is worth noting that the parameters can be updated in parallel using these update rules. This means that this algorithm is well suited to parallel implementations. Furthermore, since each of the update rules consists of a negative 0th-order term and a negative 2nd-order term, it turns out that there is only one positive solution, implying that there is no need to solve a solution selection problem.

4. EXPERIMENTS

To evaluate the effect of the proposed algorithm for speech enhancement tasks, we tested baseline supervised NMF (SNMF), DNMF using the multiplicative update algorithm proposed in [3] (DNMF-MU) and DNMF using the proposed algorithm (DNMF-MM) using speech data excerpted from the ATR503 database [14] and two types of measured noise, namely department store and subway station noise, excerpted from the ATR ambient noise sound database. We used signal-to-distortion ratios (SDRs) and signal-to-interference ratios (SIRs) [15] for the evaluation.

The test data were created by adding noise signals to clean speech signals with signal-to-noise ratios (SNRs) of -6, -3, 0, and 3 dB. All the audio signals were monaural and sampled at 16KHz. The STFT was computed using a Hanning window that was 32ms long with a 16ms overlap.

In the training phase, 200 utterances spoken by 2 male and 2 female speakers were used to train 40 speech basis spectra. For noise we used the same number of basis spectra. Fig. 1 shows the means and standard deviations of the SDR improvements [dB] obtained with the proposed algorithm with various iteration numbers (0, 10, 25, 50, 100, 200) initialized at 5 random values. On the basis of these results, we set the iteration number at 25 in the following experiments. We used 40 utterances selected randomly from the ATR database as the test set.

Tabs. 1 and 2 show the results of the SDR and SIR improvements obtained with the proposed algorithm (DNMF-MM) and the other two algorithms (SNMF, DNMF-MU) using two types of noise. The proposed algorithm (DNMF-MM) yielded an SDR improvement that was 0.86 dB higher than SNMF and 0.71 dB higher than DNMF-MU. It is worth noting that the proposed algorithm obtained an average SIR improvement of more than a 2.9 dB over SNMF, showing the discriminative power of the basis spectra.

Table 1. SDR improvement [dB] evaluated under department store noise (top row) and subway station (bottom row) conditions.

Method	Input SNR				
	-6 dB	-3 dB	0 dB	3 dB	Avg
SNMF	5.58	5.53	5.18	4.64	5.23
DNMF_MU	5.88	5.68	5.11	**4.70**	5.34
DNMF_MM	**6.41**	**6.29**	**5.72**	**4.70**	**5.78**
SNMF	5.79	5.65	5.19	4.06	5.17
DNMF_MU	5.51	5.86	5.22	4.80	5.35
DNMF_MM	**6.82**	**7.20**	**6.50**	**4.89**	**6.35**

Table 2. SIR improvement [dB] evaluated under department store noise (top row) and subway station conditions (bottom row).

Method	Input SNR				
	-6 dB	-3 dB	0 dB	3 dB	Avg
SNMF	7.23	7.44	7.44	7.31	7.36
DNMF_MU	8.07	7.87	7.44	7.34	7.68
DNMF_MM	**9.76**	**9.66**	**10.16**	**9.74**	**9.83**
SNMF	7.78	8.04	8.10	8.16	8.02
DNMF_MU	8.04	8.67	7.95	8.29	8.24
DNMF_MM	**10.77**	**11.58**	**11.89**	**11.28**	**11.38**

5. CONCLUSION

While DNMF is noteworthy in that it directly uses the reconstruction errors of the separated signals as the training criteria, it causes difficulty as regards 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 achieved significant improvements in terms of the SDR and SIR criteria over standard NMF and DNMF using the multiplicative update algorithm.

6. REFERENCES

[1] D. D. Lee, and H. S. Sung, “Algorithms for non-negative matrix factorization.”, In Advances in neural information processing systems, pp. 556–562, 2001.

[2] P. Smaragdis, R. Bhiksha, and S. Madhusudana, “Supervised and semi-supervised separation of sounds from single-channel mixtures.”, In Proc. ICA, pp. 414–421, 2007.

[3] F. Weninger, J. L. Roux, J. R. Hershey, and S. Watanabe, “Discriminative NMF and its application to single-channel source separation.”, In Proc. INTERSPEECH, pp. 865–869, 2014.

[4] E. M. Grais, and H. Erdogan, “Discriminative nonnegative dictionary learning using cross-coherence penalties for single channel source separation.”, in Proc. INTERSPEECH, pp. 808–812, 2013.

[5] G. Bao, Y. Xu, and Z. Ye, “Learning a discriminative dictionary for single-channel speech separation.”, IEEE/ACM Transactions on Audio, Speech, and Language Processing 22(7), pp. 1130–1138, 2014.

[6] N. Boulanger-Lewandowski, Y. Bengio, and P. Vincent, “Discriminative non-negative matrix factorization for multiple pitch estimation.” In Proc. ISMIR, pp. 205–210, 2012.

[7] Z. Wang, and F. Sha, “Discriminative non-negative matrix factorization for single-channel speech separation.” In Proc. ICASSP, pp. 3749–3753, 2014.

[8] K. Kwon, J. W. Shin, and N. S. Kim, “Target source separation based on discriminative nonnegative matrix factorization incorporating cross-reconstruction error.”, IEICE Transactions on Information and Systems 98(11), pp. 2017–2020, 2015.

[9] J. Eggert, and E. Korner, “Sparse coding and NMF.”, in Proc. of Neural Networks, vol 4, pp. 2529–2533, 2004.

[10] P. Sprechmann, A. M. Bronstein, and G. Sapiro, “Supervised non-Euclidean sparse NMF via bilevel optimization with applications to speech enhancement.”, In Proc. HSCMA, pp. 11–15, 2014.

[11] C. Fevotte, 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.

[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 Proc. MLSP, pp. 283–288, 2010.

[13] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” Journal of Royal Statistical Society Series B, vol. 39, pp. 1–38, 1977.

[14] A. Kurematsu, K. Takeda, Y. Sagisaka, S. Katagiri, H. Kuwabara, and K. Shikano, “ATR Japanese speech database as a tool of speech recognition and synthesis,” Speech Communication, vol. 9, pp. 357–363, 1990.

[15] E. Vincent, R. Gribonval, and C. Fevotte, “Performance measurement in blind audio source separation.”, IEEE transactions on audio, speech, and language processing, vol. 14, no. 4, pp. 1462–1469, 2016.