IEICE TRANSACTIONS

on Fundamentals of Electronics, Communications and Computer Sciences

VOL. E97-A NO. 5
MAY 2014

A PUBLICATION OF THE ENGINEERING SCIENCES SOCIETY
The Institute of Electronics, Information and Communication Engineers
Kikai-Shinko-Kaikan Bldg., 5-8, Shibakoen 3 chome, Minato-ku, TOKYO, 105-0011 JAPAN

LETTER

Music Signal Separation Based on Supervised Nonnegative Matrix Factorization with Orthogonality and Maximum-Divergence Penalties

Daichi KITAMURA[†a)], ***Student Member***, **Hiroshi SARUWATARI**[†], ***Member***, **Kosuke YAGI**[†], ***Nonmember***, **Kiyohiro SHIKANO**[†], ***Fellow***, **Yu TAKAHASHI**[††], ***Nonmember***, ***and*** **Kazunobu KONDO**[††], ***Member***

SUMMARY In this letter, we address monaural source separation based on supervised nonnegative matrix factorization (SNMF) and propose a new penalized SNMF. Conventional SNMF often degrades the separation performance owing to the basis-sharing problem. Our penalized SNMF forces nontarget bases to become different from the target bases, which increases the separated sound quality.

key words: *music signal separation, nonnegative matrix factorization, supervised method*

1. Introduction

In this study, we address monaural music signal separation using nonnegative matrix factorization (NMF) [1]–[3], which decomposes an observed spectrogram into a basis matrix and an activation matrix. The basis matrix involves frequently appearing spectral patterns in the observed spectrogram as the column vectors, and the activation matrix involves the time-varying gain corresponding to each basis. In particular, supervised NMF (SNMF) [4] has been proposed, which utilizes some sound samples of a target signal as supervision for a priori training. In SNMF, a spectrogram of the supervision sound is decomposed by conventional NMF in a training process to generate the supervised basis matrix. In the separation process, an observed spectrogram that consists of target and interference sources is decomposed using the supervised bases and other bases. Finally, the separated target signal can be reconstructed from the supervised bases and their activation.

Conventional SNMF incurs a risk of degrading the separation performance owing to the simultaneous generation of similar spectral patterns in the supervised bases and other bases. Here, we explain this phenomenon via a simplified example. Let $\boldsymbol{X}$ be an observed nonnegative matrix whose rank is one. We can represent $\boldsymbol{X}$ using a (nonnegative column) supervised basis vector $\boldsymbol{a}_1$ and the corresponding (nonnegative column) activation vector $\boldsymbol{b}$ as $\boldsymbol{X}=\boldsymbol{a}_1\boldsymbol{b}^{\mathrm{T}}+\boldsymbol{a}_2\boldsymbol{0}^{\mathrm{T}}$, where $\boldsymbol{a}_2$ is an arbitrary basis vector and $\boldsymbol{0}$ is a zero vector. If $\boldsymbol{a}_2$ is identical to $\boldsymbol{a}_1$, we can also represent $\boldsymbol{X}$ using other activation vectors $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ as $\boldsymbol{X}=\boldsymbol{a}_1\boldsymbol{b}_1^{\mathrm{T}}+\boldsymbol{a}_2\boldsymbol{b}_2^{\mathrm{T}}$, where $\boldsymbol{b}=\boldsymbol{b}_1+\boldsymbol{b}_2$; the original activation vector $\boldsymbol{b}$ is split into $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$, and then the separated signal $\boldsymbol{a}_1\boldsymbol{b}_1^{\mathrm{T}}$ is distorted. In SNMF, such basis sharing between the supervised bases and other bases often occurs. This is because the cost function in NMF is defined as the divergence $\mathcal{D}(\cdot\|\cdot)$ between the observed and reconstructed matrices, and unique decomposition is not guaranteed ($\mathcal{D}(\boldsymbol{X}\|\boldsymbol{a}_1\boldsymbol{b}^{\mathrm{T}})=\mathcal{D}(\boldsymbol{X}\|\boldsymbol{a}_1\boldsymbol{b}_1^{\mathrm{T}}+\boldsymbol{a}_2\boldsymbol{b}_2^{\mathrm{T}})$ in the above case).

To solve this problem, we propose a new penalized SNMF (PSNMF), which employs a penalty term in the cost function to force the other bases to become as different as possible from the supervised bases. In this study, we introduce two types of penalty term based on orthogonality and divergence maximization, and we confirm their efficacy via experimental evaluations.

2. Proposed PSNMF

2.1 Decomposition Model

The following equation represents the decomposition model of PSNMF:

$$\boldsymbol{Y} \simeq \boldsymbol{F}\boldsymbol{G} + \boldsymbol{H}\boldsymbol{U}, \tag{1}$$

where $\boldsymbol{Y}(\in \mathbb{R}_{\geq 0}^{\Omega\times T})$ is an observed spectrogram, $\boldsymbol{F}(\in \mathbb{R}_{\geq 0}^{\Omega\times K})$ is a matrix that involves supervised spectral bases (dictionary) of the target source as column vectors, $\boldsymbol{G}(\in \mathbb{R}_{\geq 0}^{K\times T})$ is the activation matrix that corresponds to $\boldsymbol{F}$, $\boldsymbol{H}(\in \mathbb{R}_{\geq 0}^{\Omega\times L})$ is another basis matrix that ideally involves residual spectral patterns, and $\boldsymbol{U}(\in \mathbb{R}_{\geq 0}^{L\times T})$ is the activation matrix that corresponds to $\boldsymbol{H}$. Moreover, Ω is the number of frequency bins, T is the number of frames of the observed signal, K is the number of supervised bases, and L is the number of the other bases. In PSNMF, the supervised basis matrix $\boldsymbol{F}$ is trained in advance via a target sound sample. After fixing $\boldsymbol{F}$, the matrices $\boldsymbol{G}$, $\boldsymbol{H}$, and $\boldsymbol{U}$ are optimized. Hence, $\boldsymbol{F}\boldsymbol{G}$ ideally represents the target source components and $\boldsymbol{H}\boldsymbol{U}$ represents the other different components after decomposition.

2.2 Cost Functions

In this section, we propose two types of PSNMF algorithm.

Manuscript received September 3, 2013.
Manuscript revised January 11, 2014.

[†]The authors are with Nara Institute of Science and Technology, Ikoma-shi, 630-0192 Japan.
[††]The authors are with Research & Development, Yamaha Corporation, Iwata-shi, 438-0192 Japan.
a) E-mail: daichi-k@is.naist.jp

DOI: 10.1587/transfun.E97.A.1113

Hereafter, we denote the entries of the nonnegative matrices $\boldsymbol{Y}$, $\boldsymbol{F}$, $\boldsymbol{G}$, $\boldsymbol{H}$, and $\boldsymbol{U}$ as $y_{\omega,t}$, $f_{\omega,k}$, $g_{k,t}$, $h_{\omega,l}$, and $u_{l,t}$, respectively. The cost function in NMF is defined as the arbitrary divergence between $\boldsymbol{Y}$ and $\boldsymbol{FG} + \boldsymbol{HU}$. In this study, we propose the use of the following generalized cost function:

$$\mathcal{J}_{\mathrm{NMF}} = \mathcal{D}_{\beta}\left(\boldsymbol{Y} \| \boldsymbol{FG} + \boldsymbol{HU}\right), \tag{2}$$

where $\mathcal{D}_{\beta}(\cdot\|\cdot)$ indicates β-divergence [5], defined as

$$\mathcal{D}_{\beta}(\boldsymbol{B}\|\boldsymbol{A}) = \sum_{m,n}\left\{\frac{b_{m,n}^{\beta}}{\beta(\beta-1)} + \frac{a_{m,n}^{\beta}}{\beta} - \frac{b_{m,n}a_{m,n}^{\beta-1}}{\beta-1}\right\}, \tag{3}$$

where $\boldsymbol{A}(\in \mathbb{R}^{M\times N})$ and $\boldsymbol{B}(\in \mathbb{R}^{M\times N})$ are matrices whose entries are $a_{m,n}$ and $b_{m,n}$, respectively. This generalized divergence is a family of cost functions parameterized by a single shape parameter β that takes Itakura-Saito divergence (*IS-divergence*), generalized Kullback-Leibler divergence (*KL-divergence*), and Euclidean distance (*EUC-distance*) as special cases ($\beta = 0$, 1, and 2, respectively).

In PSNMF, to avoid the sharing of bases, we make $\boldsymbol{H}$ as different as possible from $\boldsymbol{F}$. We impose the following minimization in addition to the cost function:

$$\arg\min_{\boldsymbol{H}} \|\boldsymbol{F}^{\mathrm{T}}\boldsymbol{H}\|_{\mathrm{Fr}}^2, \tag{4}$$

where the conditions $\sum_{\omega} f_{\omega,k} = 1$ and $\sum_{\omega} h_{\omega,l} = 1$ are applied, and $\|\cdot\|_{\mathrm{Fr}}$ indicates the Frobenius norm. This minimization corresponds to the maximization of orthogonality between $\boldsymbol{F}$ and $\boldsymbol{H}$. The cost function with the orthogonality penalty is given by

$$\begin{aligned}\mathcal{J}_1 &= \mathcal{J}_{\mathrm{NMF}} + \mu_1\|\boldsymbol{F}^{\mathrm{T}}\boldsymbol{H}\|_{\mathrm{Fr}}^2 \\ &= \mathcal{J}_{\mathrm{NMF}} + \mu_1\sum_{k,l}\left(\sum_{\omega} f_{\omega,k}h_{\omega,l}\right)^2,\end{aligned} \tag{5}$$

where μ_1 is the weighting parameter for the penalty term.

As another means for making $\boldsymbol{H}$ different from $\boldsymbol{F}$, the maximization of all divergence combinations between the supervised bases in $\boldsymbol{F}$ and the other bases in $\boldsymbol{H}$ can be used, which is given by

$$\arg\max_{\boldsymbol{H}} \sum_{k,l,\omega} \mathcal{D}_{\beta_{\mathrm{m}}}\left(f_{\omega,k}\|h_{\omega,l}\right), \tag{6}$$

where β_{m} is the shape parameter of the divergence for this penalty, and the conditions $\sum_{\omega} f_{\omega,k} = 1$ and $\sum_{\omega} h_{\omega,l} = 1$ are applied. This maximization forces the other bases in $\boldsymbol{H}$ to become different from the target spectral patterns. The cost function with the maximum-divergence penalty is given by

$$\mathcal{J}_2 = \mathcal{J}_{\mathrm{NMF}} + \mu_2\exp\left(-\frac{1}{\lambda}\sum_{k,l,\omega}\mathcal{D}_{\beta_{\mathrm{m}}}\left(f_{\omega,k}\|h_{\omega,l}\right)\right), \tag{7}$$

where μ_2 and λ are the weighting and sensitivity parameters, respectively. Here, exponentiation is applied to make the penalty term nonnegative.

2.3 Auxiliary Functions

In this section, we derive the update rules based on the cost functions Eqs. (5) and (7), similarly to [6]. Since it is difficult to analytically derive the optimal $\boldsymbol{G}$, $\boldsymbol{H}$, and $\boldsymbol{U}$, we define auxiliary functions $\mathcal{J}_1^{+}$ and $\mathcal{J}_2^{+}$ that represent the upper bounds of $\mathcal{J}_1$ and $\mathcal{J}_2$, respectively. Here, we can rewrite Eq. (2) as

$$\mathcal{J}_{\mathrm{NMF}} = \sum_{\omega,t}\left\{\frac{y_{\omega,t}^{\beta}}{\beta(\beta-1)} + \frac{z_{\omega,t}^{\beta}}{\beta} - \frac{y_{\omega,t}z_{\omega,t}^{\beta-1}}{\beta-1}\right\}, \tag{8}$$

where $z_{\omega,t}$ is defined as

$$z_{\omega,t} = \sum_{k} f_{\omega,k}g_{k,t} + \sum_{l} h_{\omega,l}u_{l,t}. \tag{9}$$

First, we define the upper bound for the second term on the right-hand side of Eq. (8). This term is convex for $\beta \geq 1$ and concave for $\beta < 1$. If β satisfies $\beta \geq 1$, the upper bound function $Q_{\omega,t}^{(\beta)}$ is defined using auxiliary variables $\alpha_{\omega,t,k} \geq 0$, $\gamma_{\omega,t,l} \geq 0$, $\eta_1 \geq 0$, and $\eta_2 \geq 0$ that satisfy $\sum_k \alpha_{\omega,t,k} = 1$, $\sum_l \gamma_{\omega,t,l} = 1$, and $\eta_1 + \eta_2 = 1$. Applying Jensen's inequality to this, we have

$$\begin{aligned}\frac{z_{\omega,t}^{\beta}}{\beta} &\leq \frac{1}{\beta}\left\{\sum_{k}\alpha_{\omega,t,k}\eta_1\left(\frac{f_{\omega,k}g_{k,t}}{\alpha_{\omega,t,k}\eta_1}\right)^{\beta} + \sum_{l}\gamma_{\omega,t,l}\eta_2\left(\frac{h_{\omega,l}u_{l,t}}{\gamma_{\omega,t,l}\eta_2}\right)^{\beta}\right\} \\ &\equiv Q_{\omega,t}^{(\beta)}.\end{aligned} \tag{10}$$

The equality in Eq. (10) holds if and only if the auxiliary variables are set as follows:

$$\alpha_{\omega,t,k} = \left(f_{\omega,k}g_{k,t}\right) / \left(\textstyle\sum_{k'} f_{\omega,k'}g_{k',t}\right), \tag{11}$$

$$\gamma_{\omega,t,l} = \left(h_{\omega,l}u_{l,t}\right) / \left(\textstyle\sum_{l'} h_{\omega,l'}u_{l',t}\right), \tag{12}$$

$$\eta_1 = \left(\textstyle\sum_{k} f_{\omega,k}g_{k,t}\right) / \left(\textstyle\sum_{k'} f_{\omega,k'}g_{k',t} + \sum_{l'} h_{\omega,l'}u_{l',t}\right), \tag{13}$$

$$\eta_2 = \left(\textstyle\sum_{l} h_{\omega,l}u_{l,t}\right) / \left(\textstyle\sum_{k'} f_{\omega,k'}g_{k',t} + \sum_{l'} h_{\omega,l'}u_{l',t}\right). \tag{14}$$

If β satisfies $\beta < 1$, the upper bound function $\mathcal{R}_{\omega,t}^{(\beta)}$ is defined using the auxiliary variable $\sigma_{\omega,t} \geq 0$. Applying the tangent line inequality to this, we have

$$\frac{z_{\omega,t}^{\beta}}{\beta} \leq \sigma_{\omega,t}^{\beta-1}\left(z_{\omega,t} - \sigma_{\omega,t}\right) + \frac{\sigma_{\omega,t}^{\beta}}{\beta} \equiv \mathcal{R}_{\omega,t}^{(\beta)}. \tag{15}$$

The equality in Eq. (15) holds if and only if the auxiliary variable is set to

$$\sigma_{\omega,t} = \sum_{k'} f_{\omega,k'}g_{k',t} + \sum_{l'} h_{\omega,l'}u_{l',t}. \tag{16}$$

Second, we define the upper bound function for the third term on the right-hand side of Eq. (8). This term is convex for $\beta \geq 2$ and concave for $\beta < 2$. Similarly to Eqs. (10) and (15), we can derive the auxiliary function for the third term of Eq. (8) as

$$-\frac{z_{\omega,t}^{\beta-1}}{\beta-1} \leq \begin{cases} -Q_{\omega,t}^{(\beta-1)} & (\beta \geq 2) \\ -\mathcal{R}_{\omega,t}^{(\beta-1)} & (\beta < 2) \end{cases}. \tag{17}$$

Third, for the orthogonality penalty term in Eq. (5), the upper bound function $\mathcal{P}^+$ is defined using an auxiliary variable $\delta_{k,l,\omega} \geq 0$ that satisfies $\sum_\omega \delta_{k,l,\omega} = 1$. Similarly to Eq. (10), we obtain

$$\mu_1 \sum_{k,l} \left(\sum_\omega f_{\omega,k} h_{\omega,l} \right)^2 \leq \mu_1 \sum_{k,l,\omega} \frac{f_{\omega,k}^2 h_{\omega,l}^2}{\delta_{k,l,\omega}} \equiv \mathcal{P}^+, \tag{18}$$

where the equality in Eq. (18) holds if and only if the auxiliary variable is set to

$$\delta_{k,l,\omega} = \left(f_{\omega,k} h_{\omega,l} \right) / \left(\textstyle\sum_{\omega'} f_{\omega',k} h_{\omega',l} \right). \tag{19}$$

Finally, using Eqs. (10), (15), (17), and (18), we can define the upper bound functions $\mathcal{J}_1^+$ and $\mathcal{J}_2^+$ as

$$\mathcal{J}_1^+ = \mathcal{J}_{\mathrm{NMF}}^+ + \mathcal{P}^+, \tag{20}$$

$$\mathcal{J}_2^+ = \mathcal{J}_{\mathrm{NMF}}^+ + \mu_2 \exp\left(-\frac{1}{\lambda} \sum_{k,l,\omega} \mathcal{D}_{\beta_{\mathrm{m}}} \left(f_{\omega,k} \| h_{\omega,l} \right) \right), \tag{21}$$

where

$$\mathcal{J}_{\mathrm{NMF}}^+ = \sum_{\omega,t} \frac{y_{\omega,t}^\beta}{\beta(\beta-1)} + \sum_{\omega,t} \mathcal{S}_{\omega,t}^{(\beta)}, \tag{22}$$

$$\mathcal{S}_{\omega,t}^{(\beta)} = \begin{cases} \mathcal{R}_{\omega,t}^{(\beta)} - y_{\omega,t} Q_{\omega,t}^{(\beta-1)} & (\beta < 1) \\ Q_{\omega,t}^{(\beta)} - y_{\omega,t} Q_{\omega,t}^{(\beta-1)} & (1 \leq \beta \leq 2) \\ Q_{\omega,t}^{(\beta)} - y_{\omega,t} \mathcal{R}_{\omega,t}^{(\beta-1)} & (\beta > 2) \end{cases}. \tag{23}$$

2.4 Update Rules for PSNMF

The update rules with respect to each variable are determined by setting the gradient of the cost function to zero. From $\partial \mathcal{J}_1^+ / \partial h_{\omega,l} = 0$, we obtain

$$\sum_t \left(\mathcal{V}_{\omega,t,l}^{(\beta)} - \mathcal{W}_{\omega,t,l}^{(\beta)} \right) + 2\mu_1 \sum_k \frac{f_{\omega,k}^2 h_{\omega,l}}{\delta_{k,l,\omega}} = 0, \tag{24}$$

where

$$\mathcal{V}_{\omega,t,l}^{(\beta)} = \begin{cases} \sigma_{\omega,t}^{\beta-1} u_{l,t} & (\beta < 1) \\ h_{\omega,l}^{\beta-1} \left(\gamma_{\omega,t,l} \eta_2 \right)^{1-\beta} u_{l,t}^\beta & (\beta \geq 1) \end{cases}, \tag{25}$$

$$\mathcal{W}_{\omega,t,l}^{(\beta)} = \begin{cases} y_{\omega,t} h_{\omega,l}^{\beta-2} \left(\gamma_{\omega,t,l} \eta_2 \right)^{2-\beta} u_{l,t}^{\beta-1} & (\beta \leq 2) \\ y_{\omega,t} \sigma_{\omega,t}^{\beta-2} u_{l,t} & (\beta > 2) \end{cases}. \tag{26}$$

By solving Eq. (24) for $h_{\omega,l}$ assuming nonnegativity, and substituting Eqs. (11)–(14), (16), and (19) into the solution, we can obtain the update rule of $h_{\omega,l}$ with the orthogonality penalty as

$$h_{\omega,l} \leftarrow h_{\omega,l} \left(\frac{\sum_t y_{\omega,t} u_{l,t} z_{\omega,t}^{\beta-2}}{\sum_t u_{l,t} z_{\omega,t}^{\beta-1} + 2\mu_1 \sum_k f_{\omega,k} \sum_{\omega'} f_{\omega',k} h_{\omega',l}} \right)^{\varphi(\beta)}, \tag{27}$$

where $\varphi(\beta)$ is given by

$$\varphi_{(\beta)} = \begin{cases} 1/(2-\beta) & (\beta < 1) \\ 1 & (1 \leq \beta \leq 2) \\ 1/(\beta-1) & (\beta > 2) \end{cases}. \tag{28}$$

Similarly to Eq. (27), we can obtain the update rule of $h_{\omega,l}$ with the maximum-divergence penalty from $\partial \mathcal{J}_2^+ / \partial h_{\omega,l} = 0$ as

$$h_{\omega,l} \leftarrow h_{\omega,l} \left(\frac{\lambda \sum_t y_{\omega,t} u_{l,t} z_{\omega,t}^{\beta-2} + \mu_2 h_{\omega,l}^{\beta_{\mathrm{m}}-1} C_{\beta_{\mathrm{m}}}}{\lambda \sum_t u_{l,t} z_{\omega,t}^{\beta-1} + \mu_2 h_{\omega,l}^{\beta_{\mathrm{m}}-2} C_{\beta_{\mathrm{m}}} \sum_k f_{\omega,k}} \right)^{\varphi(\beta)}, \tag{29}$$

where

$$C_{\beta_{\mathrm{m}}} = \exp\left(-\frac{1}{\lambda} \sum_{k,l,\omega} \mathcal{D}_{\beta_{\mathrm{m}}} \left(f_{\omega,k} \| h_{\omega,l} \right) \right). \tag{30}$$

The update rules of the activation matrices are obtained as follows:

$$g_{k,t} \leftarrow g_{k,t} \left(\frac{\sum_\omega f_{\omega,k} y_{\omega,t} z_{\omega,t}^{\beta-2}}{\sum_\omega f_{\omega,k} z_{\omega,t}^{\beta-1}} \right)^{\varphi(\beta)}, \tag{31}$$

$$u_{l,t} \leftarrow u_{l,t} \left(\frac{\sum_\omega h_{\omega,l} y_{\omega,t} z_{\omega,t}^{\beta-2}}{\sum_\omega h_{\omega,l} z_{\omega,t}^{\beta-1}} \right)^{\varphi(\beta)}. \tag{32}$$

3. Experiment for Artificial Signals

3.1 Experimental Conditions

To confirm the efficacy of PSNMF, we compared the applicability of PSNMF and conventional SNMF ($\mu_1 = \mu_2 = 0$ in Eqs. (27) and (29)) to a separation task involving monaural multiple instrumental sources. We produced the four melodies depicted in Fig. 1. These instrumental signals were artificially generated by a MIDI synthesizer. As the supervision sound for a priori training, we used the same MIDI sounds of the target instruments containing two octave notes that cover all the notes of the target signal in the observed signal. The spectrograms were computed using a 92-ms-long rectangular window with a half-size shift. The number of iterations for the training and separation processes was 500. Moreover, the number of supervised bases in $\boldsymbol{F}$ was 100 and the number of the other bases in $\boldsymbol{H}$ was 50. In this experiment, the parameters μ_1, μ_2, and λ were changed to

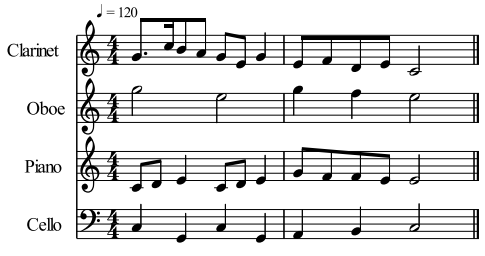

Fig. 1 Scores of each instrument.

evaluate their dependence on the separation performance.

We conducted two experiments to consider two-source and four-source cases. In the two-source case, the observed signal $\boldsymbol{Y}$ was produced by mixing two sources selected from four instruments with the same power. Therefore we prepared 12 combinations of the observed signals. In the four-source case, we produced an observed signal that consisted of four instruments with the same power. Then we calculated the average evaluation scores for each combination of the instruments.

In NMF decomposition, the parameter β of the divergence directly affects the separation accuracy. Hence, we used three values, namely, $\beta = 0$ (IS-divergence), $\beta = 1$ (KL-divergence), and $\beta = 2$ (EUC-distance). Also, we used $\beta_{\mathrm{m}} = 0$, 1, and 2 for the maximum-divergence penalty.

3.2 Experimental Results

We used the signal-to-distortion ratio (SDR), source-to-interference ratio (SIR), and sources-to-artifacts ratio (SAR) defined in [7] as the evaluation scores. SDR indicates the quality of the separated target sound, SIR indicates the degree of separation between the target and other sounds, and SAR indicates the absence of artificial distortion. Therefore, SDR indicates the total evaluation score that involves SIR and SAR.

First, we depict the dependence of SDR values on the parameters μ_1, μ_2, and λ in the two-source case, where only $\beta = 1$ is selected owing to the limit of the space because KL-divergence-based NMF is often used for many signal separation tasks. Figure 2 shows the variation of SDR values of PSNMF with the orthogonality penalty. We plot the average SDR values of 12 combinations of the observed signals and its deviation in the error bar. From this result, we can confirm that the separation performance improves with increasing the value of μ_1 because of the prevention of basis sharing, and high SDR values can be retained when μ_1 is set to be large enough. Figures 3–5 show the variation of SDR values of PSNMF with the maximum-divergence penalty. From these results, we can also prevent the basis-sharing problem by setting the parameter μ_2 to be large. However, the fluctuation exists in the average SDR along with μ_2; it indicates a slight difficulty in optimizing the parameters in the maximum-divergence penalty. In all cases regardless of the type of penalties, the deviation in 12 combinations of the observed signals is within 4 dB. The other results of $\beta = 0$ and 2 showed the similar tendency to that of $\beta = 1$.

Next, Tables 1, 2, and 3 show the average scores of SDR, SIR, and SAR in the two-source case for each value of β. Here, the SDR values are the average of 12 combinations of the observed signals with optimization on the parameters μ_1, μ_2, and λ, and the SIR and SAR values are the corresponding ones. Also, Tables 4, 5, and 6 show the average scores in the four-source case. From the SDR results, we can confirm that conventional SNMF cannot achieve a high separation accuracy owing to basis sharing between the supervised bases and other bases, whereas the proposed methods can avoid this basis-sharing problem by using the penalty terms. In addition, the performances of the orthog-

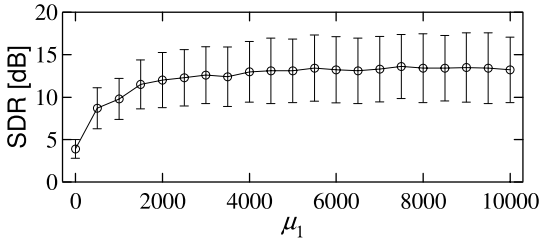

Fig. 2 Variation of SDR values of PSNMF with orthogonality penalty when $\beta = 1$ in two-source case. Circle and error bar represent average SDR of 12 combinations of observed signals and its deviation, respectively.

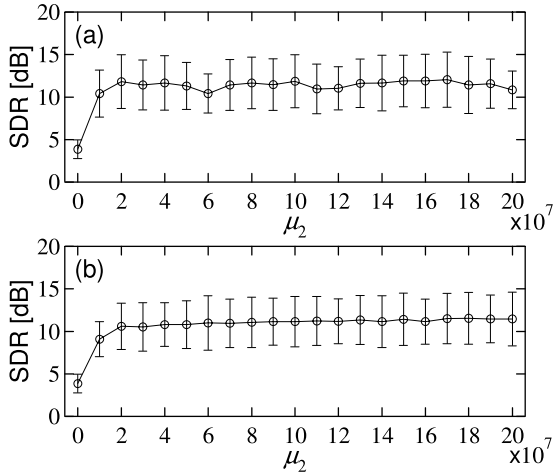

Fig. 3 Variation of SDR values of PSNMF with maximum-divergence penalty when $\beta = 1$ and $\beta_{\mathrm{m}} = 0$ in two-source case: (a) $\lambda = 10^5$ and (b) $\lambda = 10^6$. Circle and error bar represent average SDR of 12 combinations of observed signals and its deviation, respectively.

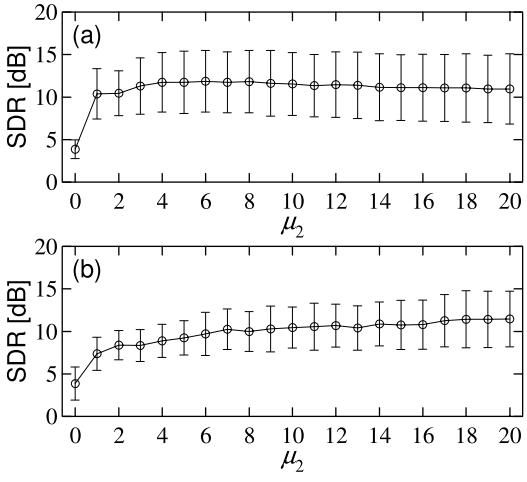

Fig. 4 Variation of SDR values of PSNMF with maximum-divergence penalty when $\beta = 1$ and $\beta_{\mathrm{m}} = 1$ in two-source case: (a) $\lambda = 1$ and (b) $\lambda = 10$. Circle and error bar represent average SDR of 12 combinations of observed signals and its deviation, respectively.

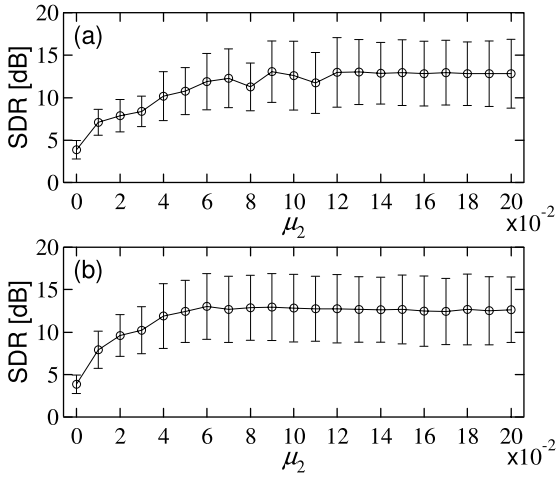

Fig. 5 Variation of SDR values of PSNMF with maximum-divergence penalty when $\beta = 1$ and $\beta_{\mathrm{m}} = 2$ in two-source case: (a) $\lambda = 10^{-4}$ and (b) $\lambda = 10^{-3}$. Circle and error bar represent average SDR of 12 combinations of observed signals and its deviation, respectively.

Table 1 Average scores in two-source case of artificial signals ($\beta = 0$).

Method	SDR	SIR	SAR
Conventional SNMF	6.1	18.5	1.9
PSNMF with orthogonality penalty	9.6	17.8	6.7
PSNMF with maximum-divergence penalty ($\beta_{\mathrm{m}} = 0$)	9.3	16.9	6.3
PSNMF with maximum-divergence penalty ($\beta_{\mathrm{m}} = 1$)	8.5	14.5	6.2
PSNMF with maximum-divergence penalty ($\beta_{\mathrm{m}} = 2$)	8.9	18.1	5.1

Table 2 Average scores in two-source case of artificial signals ($\beta = 1$).

Method	SDR	SIR	SAR
Conventional SNMF	3.9	18.1	−0.2
PSNMF with orthogonality penalty	13.6	19.0	11.3
PSNMF with maximum-divergence penalty ($\beta_{\mathrm{m}} = 0$)	12.0	18.6	9.0
PSNMF with maximum-divergence penalty ($\beta_{\mathrm{m}} = 1$)	11.8	15.9	10.5
PSNMF with maximum-divergence penalty ($\beta_{\mathrm{m}} = 2$)	13.0	17.5	11.6

Table 3 Average scores in two-source case of artificial signals ($\beta = 2$).

Method	SDR	SIR	SAR
Conventional SNMF	6.7	18.4	3.4
PSNMF with orthogonality penalty	11.9	18.1	9.8
PSNMF with maximum-divergence penalty ($\beta_{\mathrm{m}} = 0$)	13.0	17.8	11.4
PSNMF with maximum-divergence penalty ($\beta_{\mathrm{m}} = 1$)	10.8	17.0	9.5
PSNMF with maximum-divergence penalty ($\beta_{\mathrm{m}} = 2$)	12.1	15.8	12.3

Table 4 Average scores in four-source case of artificial signals ($\beta = 0$).

Method	SDR	SIR	SAR
Conventional SNMF	6.9	14.9	3.6
PSNMF with orthogonality penalty	8.6	12.6	6.8
PSNMF with maximum-divergence penalty ($\beta_{\mathrm{m}} = 0$)	8.6	13.8	6.0
PSNMF with maximum-divergence penalty ($\beta_{\mathrm{m}} = 1$)	8.8	14.3	6.3
PSNMF with maximum-divergence penalty ($\beta_{\mathrm{m}} = 2$)	8.6	13.8	6.0

onality and maximum-divergence penalties are roughly the same.

Figure 6 shows sample spectrograms after the separation, which extracts the cello signal from the observed signal

Table 5 Average scores in four-source case of artificial signals ($\beta = 1$).

Method	SDR	SIR	SAR
Conventional SNMF	7.1	14.2	4.3
PSNMF with orthogonality penalty	10.8	14.1	10.6
PSNMF with maximum-divergence penalty ($\beta_{\mathrm{m}} = 0$)	11.1	13.8	10.9
PSNMF with maximum-divergence penalty ($\beta_{\mathrm{m}} = 1$)	9.7	12.2	9.9
PSNMF with maximum-divergence penalty ($\beta_{\mathrm{m}} = 2$)	10.1	12.0	11.8

Table 6 Average scores in four-source case of artificial signals ($\beta = 2$).

Method	SDR	SIR	SAR
Conventional SNMF	7.3	13.0	6.4
PSNMF with orthogonality penalty	9.6	11.8	11.4
PSNMF with maximum-divergence penalty ($\beta_{\mathrm{m}} = 0$)	10.3	12.1	12.6
PSNMF with maximum-divergence penalty ($\beta_{\mathrm{m}} = 1$)	9.2	12.1	11.3
PSNMF with maximum-divergence penalty ($\beta_{\mathrm{m}} = 2$)	8.6	11.1	11.3

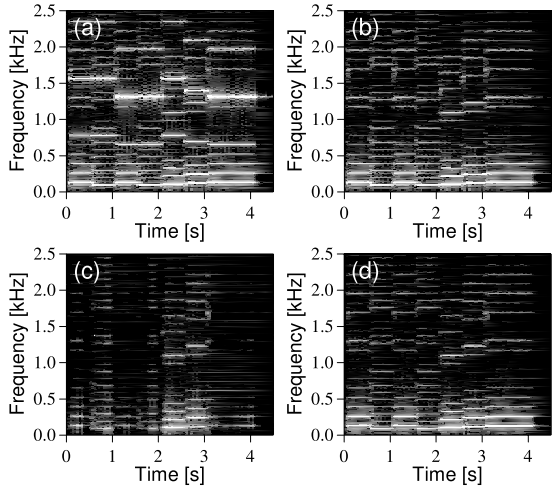

Fig. 6 Spectrograms of (a) observed signal consisting of cello and oboe, (b) oracle signal of target cello signal, (c) cello signal extracted by conventional SNMF, and (d) cello signal extracted by proposed PSNMF with orthogonality penalty.

that consists of cello and oboe signals. The signal extracted by conventional SNMF loses some of the target spectra (see Fig. 6(c)) because of basis sharing, but the proposed method extracts the target source with high accuracy (see Fig. 6(d)).

Finally, as another means of preventing the basis sharing problem, some people may guess that the orthogonality penalty on *activations*, $\|\boldsymbol{G}\boldsymbol{U}^{\mathrm{T}}\|_{\mathrm{Fr}}^2$, can also be introduced instead of the proposed penalty, $\|\boldsymbol{F}^{\mathrm{T}}\boldsymbol{H}\|_{\mathrm{Fr}}^2$. However, this penalty term has a risk to force $\boldsymbol{G}$ to become $\boldsymbol{0}$, which yields that the input matrix $\boldsymbol{Y}$ is represented using only the other matrix $\boldsymbol{HU}$. Indeed, for instance, the average SDR value of KL-divergence-based PSNMF with this penalty term was −8.8 dB in two source case, and the output signal did not contain the sufficient target components.

4. Experiment for Real-Recorded Signals

4.1 Experimental Conditions

We also conducted another experiment using real-recorded

Table 7 Average scores in two-source case of recorded signals ($\beta = 0$).

Method	SDR	SIR	SAR
Conventional SNMF	7.5	19.6	3.5
PSNMF with orthogonality penalty	10.6	19.1	11.8
PSNMF with maximum-divergence penalty ($\beta_m = 0$)	10.5	18.2	7.1
PSNMF with maximum-divergence penalty ($\beta_m = 1$)	10.5	17.2	7.3
PSNMF with maximum-divergence penalty ($\beta_m = 2$)	10.5	19.7	6.7

Table 8 Average scores in two-source case of recorded signals ($\beta = 1$).

Method	SDR	SIR	SAR
Conventional SNMF	4.4	20.2	−0.1
PSNMF with orthogonality penalty	14.4	20.0	11.7
PSNMF with maximum-divergence penalty ($\beta_m = 0$)	13.8	20.3	10.4
PSNMF with maximum-divergence penalty ($\beta_m = 1$)	13.6	18.2	11.0
PSNMF with maximum-divergence penalty ($\beta_m = 2$)	14.4	18.5	12.4

Table 9 Average scores in two-source case of recorded signals ($\beta = 2$).

Method	SDR	SIR	SAR
Conventional SNMF	7.5	19.6	3.5
PSNMF with orthogonality penalty	11.9	19.0	9.2
PSNMF with maximum-divergence penalty ($\beta_m = 0$)	14.7	19.5	12.4
PSNMF with maximum-divergence penalty ($\beta_m = 1$)	11.9	18.2	9.9
PSNMF with maximum-divergence penalty ($\beta_m = 2$)	13.2	16.3	12.8

Table 10 Average scores in four-source case of recorded signals ($\beta = 0$).

Method	SDR	SIR	SAR
Conventional SNMF	8.3	17.4	4.6
PSNMF with orthogonality penalty	11.3	16.7	8.8
PSNMF with maximum-divergence penalty ($\beta_m = 0$)	11.4	15.3	9.5
PSNMF with maximum-divergence penalty ($\beta_m = 1$)	10.9	16.2	8.2
PSNMF with maximum-divergence penalty ($\beta_m = 2$)	11.6	16.1	9.2

music signals. We recorded each instrumental solo signal and the supervision sound, which are the same as those in the previous section, in an experimental room whose reverberation time was 200 ms. The distance between a loudspeaker and binaural microphone NEUMANN KU-100 was 1.5 m. The binaurally recorded signals in both ears were mixed down to a monaural signal. The observed signal $\boldsymbol{Y}$ was produced by mixing these recorded signals as the same power. Other conditions were the same as those of the previous experiment, and we prepared the observed signals in two-source and four-source cases.

4.2 Experimental Results

Tables 7, 8, and 9 show the average scores of SDR, SIR, and SAR in the two-source case for each value of β. Here, the SDR values are the average of 12 combinations of the observed signals with optimization on the parameters μ_1, μ_2, and λ, and the SIR and SAR values are the corresponding ones. Also, Tables 10, 11, and 12 show the average scores in the four-source case. From these results, we can confirm

Table 11 Average scores in four-source case of recorded signals ($\beta = 1$).

Method	SDR	SIR	SAR
Conventional SNMF	8.2	16.2	4.6
PSNMF with orthogonality penalty	13.4	16.2	12.4
PSNMF with maximum-divergence penalty ($\beta_m = 0$)	14.0	15.3	12.0
PSNMF with maximum-divergence penalty ($\beta_m = 1$)	12.2	16.1	11.6
PSNMF with maximum-divergence penalty ($\beta_m = 2$)	13.0	16.2	11.6

Table 12 Average scores in four-source case of recorded signals ($\beta = 2$).

Method	SDR	SIR	SAR
Conventional SNMF	9.1	13.9	6.8
PSNMF with orthogonality penalty	12.3	15.1	11.6
PSNMF with maximum-divergence penalty ($\beta_m = 0$)	12.9	15.0	13.2
PSNMF with maximum-divergence penalty ($\beta_m = 1$)	11.9	14.3	12.5
PSNMF with maximum-divergence penalty ($\beta_m = 2$)	10.3	11.8	12.2

that our proposed method can achieve higher separation accuracy compared with the conventional method even in the case of real-recorded signals.

5. Conclusion

In this study, we propose a new penalized SNMF with two types of penalty that force the other bases to become different from the target bases trained in advance. From the experimental results, it can be confirmed that the proposed method prevents the simultaneous generation of similar spectral patterns in the supervised bases and other bases, and increases the separation performance compared with the conventional method.

References

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

[2] T. Virtanen, "Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria," IEEE Trans. Audio Speech Language Process., vol.15, pp.1066–1074, 2007.

[3] H. Kameoka, M. Nakano, K. Ochiai, Y. Imoto, K. Kashino, S. Sagayama, "Constrained and regularized variants of non-negative matrix factorization incorporating music-specific constraints," Proc. ICASSP, pp.5365–5368, 2012.

[4] P. Smaragdis, B. Raj, and M. Shashanka, "Supervised and semi-supervised separation of sounds from single-channel mixtures," Proc. LVA/ICA 2010, LNCS 6365, pp.140–148, 2010.

[5] S. Eguchi and K. Yano, "Robustifying maximum likelihood estimation," Technical Report, Institute of Statistical Mathematics, 2001.

[6] 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," Proc. 2010 IEEE International Workshop on Machine Learning for Signal Processing, pp.283–288, 2010.

[7] 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, 2006.