Divergence optimization based on trade-off between separation and extrapolation abilities in superresolution-based nonnegative matrix factorization *

☆ Daichi Kitamura, Hiroshi Saruwatari, Satoshi Nakamura (Nara Institute of Science and Technology)
Kazunobu Kondo, Yu Takahashi (Yamaha Corporation)

1 Introduction

In recent years, music and acoustic signal separation based on nonnegative matrix factorization (NMF) [1] has been a very active area of signal processing research. NMF for acoustical signals decomposes an input spectrogram into the product of a spectral basis matrix and its activation matrix. In particular, supervised NMF (SNMF) [2, 3], which includes a priori training with some sample sounds of a target instrument, can extract the target signal to some extent. However, for the case of a mixture consisting of many sources, the source extraction performance is markedly degraded when only single-channel observation is available.

Multichannel NMF, which is a natural extension for multichannel signal, has been proposed as an unsupervised method [4]. However, such an unsupervised separation is a difficult problem because the decomposition is underspecified. Hence, these algorithms involve strong dependence on initial values and lack robustness.

As another means for addressing multichannel signal separation, a hybrid method, which concatenates superresolution-based SNMF after directional clustering, has been proposed by the authors [5]. This method uses index information generated by binary masking of directional clustering to regard the spectral chasms as *unseen* observations, and finally reconstructs the target source components via spectrogram extrapolation using the supervised bases. In this paper, we extend the method [5] to be a more generalized form introducing new parametric cost functions for both separation and regularization based on β-divergence criterion. Also, we discuss the optimal divergence for this method.

2 Conventional Method

2.1 SNMF

The unsupervised NMF approaches have difficulty in clustering the decomposed spectral patterns into a specific target instrumental sound. To solve this problem, SNMF has been proposed [2, 3]. In particular, a penalized SNMF (PSNMF), which imposes a penalty term to force supervised bases and other bases to become uncorrelated with each other, achieves good performance [3]. SNMF consists of two processes, a priori training and observed signal separation, as described below in detail.

2.1.1 Training Process of Supervision

In SNMF, a priori spectral patterns (bases) should be trained in advance as a basis *dictionary*. Hereafter, we assume that we can obtain a specific solo-played instrumental sound, which is the target of the separation task. The trained bases are constructed by simple NMF as

$$\boldsymbol{Y}_{\text{target}} \simeq \boldsymbol{F}\boldsymbol{Q}, \tag{1}$$

where $\boldsymbol{Y}_{\text{target}}(\in \mathbb{R}_{\geq 0}^{\Omega \times T_s})$ is an amplitude (or a power) spectrogram of the specific signal for training, $\boldsymbol{F}(\in \mathbb{R}_{\geq 0}^{\Omega \times K})$ is a nonnegative matrix that involves bases of the target signal as column vectors, and $\boldsymbol{Q}(\in \mathbb{R}_{\geq 0}^{K \times T_s})$ is a nonnegative matrix that corresponds to the activation of each basis of $\boldsymbol{F}$. In addition, Ω is the number of frequency bins, T_s is the number of frames of the training signal, and K is the number of bases. Therefore, the basis matrix $\boldsymbol{F}$ constructed by (1) is used as the supervision (dictionary) of the target instrumental signal.

To construct the basis matrix $\boldsymbol{F}$ and the activation matrix $\boldsymbol{Q}$, a cost function is given by

$$\mathcal{J}_{\text{NMF}} = \mathcal{D}\left(\boldsymbol{Y}_{\text{target}} \| \boldsymbol{F}\boldsymbol{Q}\right), \tag{2}$$

where $\mathcal{D}(\cdot\|\cdot)$ is an arbitrary distance function.

2.1.2 Signal Separation Process

The following equation represents the decomposition of SNMF using the trained supervision matrix $\boldsymbol{F}$:

$$\boldsymbol{Y} \simeq \boldsymbol{F}\boldsymbol{G} + \boldsymbol{H}\boldsymbol{U}, \tag{3}$$

where $\boldsymbol{Y}(\in \mathbb{R}_{\geq 0}^{\Omega \times T})$ is an observed spectrogram, $\boldsymbol{G}(\in \mathbb{R}_{\geq 0}^{K \times T})$ is an activation matrix that corresponds to $\boldsymbol{F}$, $\boldsymbol{H}(\in \mathbb{R}_{\geq 0}^{\Omega \times L})$ represents the residual spectral patterns that cannot be expressed by $\boldsymbol{F}\boldsymbol{G}$, and $\boldsymbol{U}(\in \mathbb{R}_{\geq 0}^{L \times T})$ is an activation matrix that corresponds to $\boldsymbol{H}$. Moreover, T is the number of frames of the observed signal and L is the number of bases of $\boldsymbol{H}$. In SNMF, the matrices $\boldsymbol{G}$, $\boldsymbol{H}$, and $\boldsymbol{U}$ are optimized under the condition that $\boldsymbol{F}$ is known in advance. Hence, ideally, $\boldsymbol{F}\boldsymbol{G}$ represents the target instrumental components and $\boldsymbol{H}\boldsymbol{U}$ represents other components different from the target sounds after the decomposition.

2.1.3 Problem of PSNMF

PSNMF can extract the target signal particularly in the case of a small number of sources. However, for the case of a mixture consisting of many sources, the source extraction performance is markedly degraded because of the existence of instruments with similar timbre.

2.2 Directional Clustering and Its Hybrid Method with Superresolution-Based SNMF

Decomposition methods employing directional (spatial) information for the multichannel signal have also been proposed as unsupervised techniques [6]. These methods quantize the direction via time-frequency binary masking. Such directional clustering works well, even in an underdetermined situation. However, there is an inherent problem that the sources located in the same direction cannot be separated only using the directional information. Furthermore, the extracted signal is likely to be distorted because the signal has many spectral chasms resulting of the binary masking procedure as shown in Fig. 1.

To solve this problem, a hybrid method that concatenates superresolution-based SNMF after the directional clustering has been proposed [5]. This SNMF algorithm explicitly utilizes index information determined by time-frequency binary masking in directional clustering. For

* "超解像型非負値行列因子分解における分離性能と外挿能力のトレードオフに基づく最適なダイバージェンスの検討," by 北村大地, 猿渡洋, 中村哲 (奈良先端大), 近藤多伸, 高橋祐 (ヤマハ株式会社).

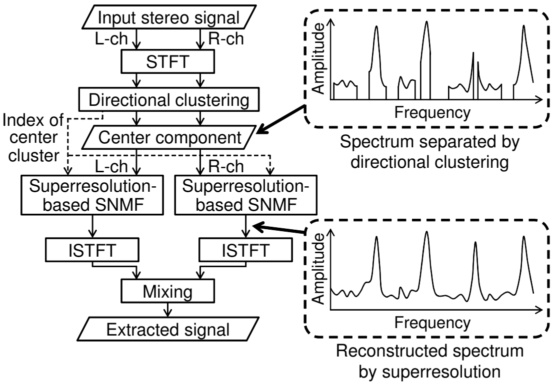

Fig. 1 Signal flow of proposed hybrid method.

example, if the target instrument is localized in the center cluster along with the interference, superresolution-based SNMF is only applied to the existing center components using index information (see Fig. 1). Therefore, the spectrogram of the target instrument is reconstructed using more matched bases because spectral chasms are treated as *unseen*, and these chasms have no impact on the cost function in SNMF. In addition, the components of the target instrument lost after directional clustering can be extrapolated using the supervised bases. In other words, the resolution of the target spectrogram is recovered with the superresolution by the supervised basis extrapolation. Furthermore, a regularization term is added in the cost function to this SNMF for avoiding the basis extrapolation error.

3 Superresolution-Based SNMF with Generalized Cost Function

3.1 Generalized Cost Function Using β-Divergence

In this section, we derive the update rules of the superresolution-based SNMF with generalized cost functions. Here, the index matrix $\boldsymbol{I}(\in \mathbb{R}_{\{0,1\}}^{\Omega\times T})$ is obtained from the binary masking preceding the directional clustering. This index matrix has specific entries of unity or zero, which indicate whether or not each grid of the spectrogram belongs to the target directional cluster. The generalized form of the cost function in superresolution-based SNMF is defined using the index matrix as

$$\mathcal{J}(\theta)=\mathcal{J}_{\mathrm{NMF}}(\theta|\,\boldsymbol{Y},\boldsymbol{I},\boldsymbol{F})+\lambda\mathcal{J}_{\mathrm{reg}}(\boldsymbol{G}|\,\boldsymbol{I},\boldsymbol{F})+\mu\|\boldsymbol{F}^{\mathrm{T}}\boldsymbol{H}\|_{\mathrm{F}}^{2}, \tag{4}$$

where $\theta=\{\boldsymbol{G},\boldsymbol{H},\boldsymbol{U}\}$ is the set of objective variables, λ and μ are the weighting parameters for each cost, and $\|\cdot\|_{\mathrm{F}}$ represents the Frobenius norm. The first term $\mathcal{J}_{\mathrm{NMF}}$ represents the main cost for the separation of superresolution-based SNMF, the second term $\mathcal{J}_{\mathrm{reg}}$ represents the cost for the regularization, and the third penalty term $\|\boldsymbol{F}^{\mathrm{T}}\boldsymbol{H}\|_{\mathrm{F}}^{2}$ indicates that $\boldsymbol{F}$ and $\boldsymbol{H}$ are forced to become uncorrelated with each other to avoid sharing the same basis [3].

In this paper, we propose to use β-divergence as the separation and regularization costs to generalize the criteria of these costs and find optimal divergences fit for superresolution-based SNMF. The β-divergence is defined as

$$\mathcal{D}_{\beta}(y\|x)=\begin{cases}\dfrac{y^{\beta}}{\beta(\beta-1)}+\dfrac{x^{\beta}}{\beta}-\dfrac{yx^{\beta-1}}{\beta-1} & (\beta\in\mathbb{R}_{\setminus\{0,1\}})\\ y(\log y-\log x)+x-y & (\beta\rightarrow 1)\\ \dfrac{y}{x}-\log\dfrac{y}{x}-1 & (\beta\rightarrow 0)\end{cases}. \tag{5}$$

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). Using (5), we can define the separation and regularization costs as follows:

$$\mathcal{J}_{\mathrm{NMF}}(\theta|\,\boldsymbol{Y},\boldsymbol{I},\boldsymbol{F})=\textstyle\sum_{\omega,t}i_{\omega,t}\left\{\dfrac{z_{\omega,t}^{\beta_{\mathrm{NMF}}}}{\beta_{\mathrm{NMF}}}-\dfrac{y_{\omega,t}z_{\omega,t}^{\beta_{\mathrm{NMF}}-1}}{\beta_{\mathrm{NMF}}-1}\right\}, \tag{6}$$

$$\mathcal{J}_{\mathrm{reg}}(\boldsymbol{G}|\,\boldsymbol{I},\boldsymbol{F})=\textstyle\sum_{\omega,t}\overline{i_{\omega,t}}\left(\sum_{k}f_{\omega,k}g_{k,t}\right)^{\beta_{\mathrm{reg}}}/\beta_{\mathrm{reg}}, \tag{7}$$

$$z_{\omega,t}=\textstyle\sum_{k}f_{\omega,k}g_{k,t}+\sum_{l}h_{\omega,l}u_{l,t}, \tag{8}$$

where $y_{\omega,t}$, $f_{\omega,k}$, $g_{k,t}$, $h_{\omega,l}$, and $u_{l,t}$ are the nonnegative entries of matrices $\boldsymbol{Y}$, $\boldsymbol{F}$, $\boldsymbol{G}$, $\boldsymbol{H}$, and $\boldsymbol{U}$, respectively, $i_{\omega,t}$ is the entry of index matrix $\boldsymbol{I}$, which maps the values of one and zero onto the time-frequency (ω-t) region, $\bar{\cdot}$ represents the binary complement of each entry in the index matrix, and β_{NMF} and β_{reg} are the parameters that define the shape of divergences of separation and regularization. Since the divergence $\mathcal{J}_{\mathrm{NMF}}$ is only defined in grids whose index is one, SNMF treats only the valid components except for the spectral chasms. In addition, the regularization cost $\mathcal{J}_{\mathrm{reg}}$, which corresponds to the grids of spectral chasms, forces to minimize the target component $\boldsymbol{FG}$ in proportion to the number of zeros in index matrix $\boldsymbol{I}$ in each frame. Hence, the supervised bases are chosen so as to minimize $\boldsymbol{FG}$ to avoid the extrapolation error.

3.2 Auxiliary Function and Update Rules

The update rules based on (4), (9), and (13) are obtained by the auxiliary function approach, similarly to [7]. First, we define the upper bound function for $\mathcal{J}_{\mathrm{NMF}}$. The first term of $\mathcal{J}_{\mathrm{NMF}}$ is convex for $\beta_{\mathrm{NMF}}\geq 1$ and concave for $\beta_{\mathrm{NMF}}<1$, and the second term is convex for $\beta_{\mathrm{NMF}}\geq 2$ and concave for $\beta_{\mathrm{NMF}}<2$. Applying Jensen's inequality to the convex function and tangent line inequality to the concave function, we can define the upper bound function $\mathcal{J}'_{\mathrm{NMF}}$ using auxiliary variables $\alpha_{\omega,t,k}\geq 0$, $\gamma_{\omega,t,l}\geq 0$, $\eta_1\geq 0$, $\eta_2\geq 0$, and $\sigma_{\omega,t}$ that satisfy $\sum_k\alpha_{\omega,t,k}=1$, $\sum_l\gamma_{\omega,t,l}=1$, and $\eta_1+\eta_2=1$ as

$$\mathcal{J}_{\mathrm{NMF}}\leq\mathcal{J}'_{\mathrm{NMF}}=\textstyle\sum_{\omega,t}i_{\omega,t}\mathcal{P}_{\omega,t}^{(\beta_{\mathrm{NMF}})}, \tag{9}$$

where

$$\mathcal{P}_{\omega,t}^{(\beta_{\mathrm{NMF}})}=\begin{cases}\mathcal{N}_{\omega,t}^{(\beta_{\mathrm{NMF}})}-y_{\omega,t}\mathcal{M}_{\omega,t}^{(\beta_{\mathrm{NMF}}-1)} & (\beta_{\mathrm{NMF}}<1)\\ \mathcal{M}_{\omega,t}^{(\beta_{\mathrm{NMF}})}-y_{\omega,t}\mathcal{M}_{\omega,t}^{(\beta_{\mathrm{NMF}}-1)} & (1\leq\beta_{\mathrm{NMF}}\leq 2)\\ \mathcal{M}_{\omega,t}^{(\beta_{\mathrm{NMF}})}-y_{\omega,t}\mathcal{N}_{\omega,t}^{(\beta_{\mathrm{NMF}}-1)} & (\beta_{\mathrm{NMF}}>2)\end{cases}, \tag{10}$$

$$\mathcal{M}_{\omega,t}^{(\beta_{\mathrm{NMF}})}=\frac{1}{\beta_{\mathrm{NMF}}}\Big\{\textstyle\sum_k\alpha_{\omega,t,k}\eta_1\left(f_{\omega,k}g_{k,t}/\alpha_{\omega,t,k}\eta_1\right)^{\beta_{\mathrm{NMF}}}$$

$$+\textstyle\sum_l\gamma_{\omega,t,l}\eta_2\left(h_{\omega,l}u_{l,t}/\gamma_{\omega,t,l}\eta_2\right)^{\beta_{\mathrm{NMF}}}\Big\}, \tag{11}$$

$$\mathcal{N}_{\omega,t}^{(\beta_{\mathrm{NMF}})}=\sigma_{\omega,t}^{\beta_{\mathrm{NMF}}-1}\left(z_{\omega,t}-\sigma_{\omega,t}\right)+\sigma_{\omega,t}^{\beta_{\mathrm{NMF}}}/\beta_{\mathrm{NMF}}. \tag{12}$$

Second, we define the upper bound function for $\mathcal{J}_{\mathrm{reg}}$. This term is convex for $\beta_{\mathrm{reg}}\geq 1$ and concave for $\beta_{\mathrm{reg}}<1$. Similarly to (9)-(12), we can define the upper bound function $\mathcal{J}'_{\mathrm{reg}}$ using auxiliary variables $\alpha_{\omega,t,k}$ and $\rho_{\omega,t}$ as

$$\mathcal{J}_{\mathrm{reg}}\leq\mathcal{J}'_{\mathrm{reg}}=\textstyle\sum_{\omega,t}\overline{i_{\omega,t}}\mathcal{S}_{\omega,t}^{(\beta_{\mathrm{reg}})}, \tag{13}$$

Fig. 2 Scores of each part.

where

$$\mathcal{S}^{(\beta_{\mathrm{reg}})}_{\omega,t} = \begin{cases} \mathcal{R}^{(\beta_{\mathrm{reg}})}_{\omega,t} & (\beta_{\mathrm{reg}} < 1) \\ \mathcal{Q}^{(\beta_{\mathrm{reg}})}_{\omega,t} & (1 \leq \beta_{\mathrm{reg}}) \end{cases}, \tag{14}$$

$$\mathcal{Q}^{(\beta_{\mathrm{reg}})}_{\omega,t} = \textstyle\sum_k \alpha_{\omega,t,k} \left(f_{\omega,k} g_{k,t} / \alpha_{\omega,t,k} \right)^{\beta_{\mathrm{reg}}} / \beta_{\mathrm{reg}}, \tag{15}$$

$$\mathcal{R}^{(\beta_{\mathrm{reg}})}_{\omega,t} = \rho^{\beta_{\mathrm{reg}}-1}_{\omega,t} \left(\textstyle\sum_k f_{\omega,k} g_{k,t} - \rho_{\omega,t} \right) + \rho^{\beta_{\mathrm{reg}}}_{\omega,t} / \beta_{\mathrm{reg}}. \tag{16}$$

Third, we define the upper bound function for $\|\boldsymbol{F}^{\mathrm{T}}\boldsymbol{H}\|^2_{\mathrm{F}}$ using auxiliary variables $\delta_{k,l,\omega} \geq 0$ that satisfy $\sum_\omega \delta_{k,l,\omega} = 1$ as

$$\|\boldsymbol{F}^{\mathrm{T}}\boldsymbol{H}\|^2_{\mathrm{F}} \leq \textstyle\sum_{k,l,\omega} \left(f^2_{\omega,k} h^2_{\omega,l} / \delta_{k,l,\omega} \right). \tag{17}$$

Finally, using (9), (13), and (17), we can define the upper bound function $\mathcal{J}'(\theta, \hat{\theta})$ as

$$\mathcal{J}'(\theta, \hat{\theta}) = \mathcal{J}'_{\mathrm{NMF}} + \mathcal{J}'_{\mathrm{reg}} + \textstyle\sum_{k,l,\omega} \left(f^2_{\omega,k} h^2_{\omega,l} / \delta_{k,l,\omega} \right), \tag{18}$$

where $\hat{\theta}$ is the set of auxiliary variables. The update rules are obtained from the derivative of $\mathcal{J}'$ w.r.t. each objective variable and substitution of equality condition w.r.t. each auxiliary variable as follows:

$$g_{k,t} \leftarrow g_{k,t} \left(\frac{\sum_\omega i_{\omega,t} y_{\omega,t} f_{\omega,k} r^{\beta_{\mathrm{NMF}}-2}_{\omega,t}}{\sum_\omega i_{\omega,t} f_{\omega,k} z^{\beta_{\mathrm{NMF}}-1}_{\omega,t} + \lambda v_{k,t}} \right)^{\varphi(\beta_{\mathrm{NMF}})}, \tag{19}$$

$$h_{\omega,l} \leftarrow h_{\omega,l} \left(\frac{\sum_t i_{\omega,t} y_{\omega,t} u_{l,t} r^{\beta_{\mathrm{NMF}}-2}_{\omega,t}}{\sum_t i_{\omega,t} u_{l,t} z^{\beta_{\mathrm{NMF}}-1}_{\omega,t} + \mu w_{\omega,l}} \right)^{\varphi(\beta_{\mathrm{NMF}})}, \tag{20}$$

$$u_{l,t} \leftarrow u_{l,t} \left(\frac{\sum_\omega i_{\omega,t} y_{\omega,t} h_{\omega,l} r^{\beta_{\mathrm{NMF}}-2}_{\omega,t}}{\sum_\omega i_{\omega,t} h_{\omega,l} z^{\beta_{\mathrm{NMF}}-1}_{\omega,t}} \right)^{\varphi(\beta_{\mathrm{NMF}})}, \tag{21}$$

where $r_{\omega,t}$, $v_{k,t}$, $w_{\omega,l}$, and $\varphi(\beta)$ are given by

$$r_{\omega,t} = \textstyle\sum_{k'} f_{\omega,k'} g_{k',t} + \sum_{l'} h_{\omega,l'} u_{l',t}, \tag{22}$$

$$v_{k,t} = \textstyle\sum_\omega \overline{i_{\omega,t}} f_{\omega,k} \left(\sum_{k'} f_{\omega,k'} g_{k',t} \right)^{\beta_{\mathrm{reg}}}, \tag{23}$$

$$w_{\omega,l} = \textstyle\sum_k f_{\omega,k} \sum_{\omega'} f_{\omega',k} h_{\omega',l}, \tag{24}$$

$$\varphi(\beta) = \begin{cases} 1/(2-\beta) & (\beta < 1) \\ 1 & (1 \leq \beta \leq 2) \\ 1/(\beta - 1) & (2 < \beta) \end{cases}. \tag{25}$$

4 Experiments

4.1 Optimal Divergence and Regularization for Superresolution-Based SNMF

4.1.1 Experimental Conditions

In this experiment, we compared some evaluation scores with various β_{NMF} and β_{reg} to find the optimal divergence and regularization fit for superresolution-based SNMF. We compared the simple PSNMF [3] and the hybrid method with superresolution-based SNMF by setting four divergences and regularizations, namely, $\beta = 0, 1, 2$, and 3. We used the same divergence (β_{NMF}) in the training and separation processes. In this experiment, we used stereo signals containing four melody parts (depicted in Fig. 2) with three compositions of instruments shown in Table 1. These signals were artificially generated by a MIDI synthesizer, and the observed signals $\boldsymbol{Y}$ were produced by mixing four sources with the same power. The target source is always located in the center direction along with another interfering source, and there are two sources in the left- and right-hand sides, which are located at $\theta = \pm 15^\circ$ based on a sine law. In addition, we used the same MIDI sounds of the target instruments as supervision for a priori training. The training sounds contained two octave notes that cover all notes of the target signal in the observed signal. The spectrograms were computed using a 92-ms-long rectangular window with a 46-ms overlap shift. The number of iterations for the training was 500 and that for the separation was 400. Moreover, the number of clusters of used in directional clustering was 3, the number of a priori bases was 100, and the number of bases for matrix $\boldsymbol{H}$ was 30. In this experiment, the weighting parameters λ and μ were experimentally determined.

Table 1 Compositions of musical instruments

Dataset	Melody 1	Melody 2	Midrange	Bass
No. 1	Oboe	Flute	Piano	Trombone
No. 2	Trumpet	Violin	Harpsichord	Fagotto
No. 3	Clarinet	Horn	Piano	Cello

4.1.2 Experimental Results and Discussion

We used the signal-to-distortion ratio (SDR), source-to-interference ratio (SIR), and sources-to-artifacts ratio (SAR) defined in [8] 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.

Figure 3 shows the average SDR, SIR, and SAR for each divergence and each regularization, where four instruments are shuffled with 12 combinations in each composition. Therefore, these results are the averages of 36 input signal patterns. From the SDR in Fig. 3, we can confirm that the EUC-distance-based cost function (β_{NMF}) is an optimal divergence for the hybrid method with superresolution-based SNMF. This is of great interest because EUC-distance-based NMF cannot achieve the best performance for common music source separation, but instead KL-divergence is often used (as also noted in many papers, e.g., [9]). In fact, SDR in Fig. 3 indicates that KL-divergence is the best divergence for PSNMF. In addition, we can confirm that the regularization with KL-divergence is slightly better than the other divergences but the difference is not obvious.

In summary, we can assert as follows.

- The optimal divergence of separation (β_{NMF}) differs between simple SNMF and superresolution-based SNMF, regardless of the type of regularization.
- We can speculate that this marked shift of optimal divergence in the SNMF methods is due to balance of separation and extrapolation abilities; this is because superresolution-based SNMF should achieves both separation and extrapolation simultaneously, as described in Sect. 2.2.

The latter issue will be addressed in the next subsection.

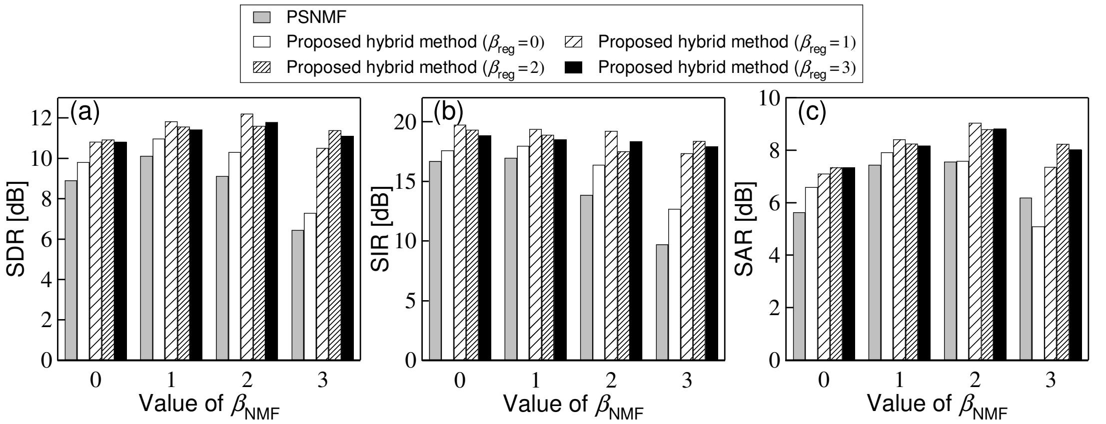

Fig. 3 Average scores: (a) shows SDR, (b) shows SIR, and (c) shows SAR for each divergence and regularization.

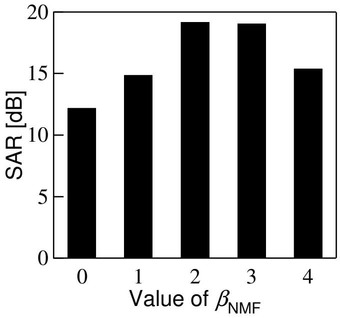

Fig. 4 Average SAR for each divergence. This score shows that extrapolation ability decreases when β_{NMF} is close to zero.

4.2 Extrapolation Ability in Superresolution-Based SNMF

4.2.1 Experimental Conditions

To quantify the extrapolation ability, we applied superresolution-based SNMF to the binary-masked signals that contain only the target instrument. The binary masks were made by directional clustering, which was the same as the previous experiments. Therefore, the SAR of this experiment indicates the net extrapolation ability. The parameter β_{NMF} was set to 0, 1, 2, 3, and 4.

4.2.2 Experimental Results and Discussion

Figure 4 shows the average SAR of 36 input signal patterns for each divergence. From this result, superresolution-based SNMF loses the extrapolation ability with decreasing β_{NMF}. If β_{NMF} is close to zero, the spectral bases become more sparsity-aware, and this property is suitable for normal NMF-based music source separation. However, for superresolution-based SNMF, which attempts to fit the trained bases using spectral components except chasms, sparsity-aware bases are not suitable because it is difficult to extrapolate the target signal component using such sparse bases. Therefore, it can be confirmed that the optimal divergence for superresolution-based SNMF is shifted to EUC-distance rather than KL-divergence because of the trade-off between separation and extrapolation abilities, as illustrated in Fig. 5.

5 Conclusions

In this paper, we address a stereo signal separation problem and derive the generalized multiple update rules for superresolution-based SNMF based on β-divergence. Also, we discuss about the optimal divergence for the

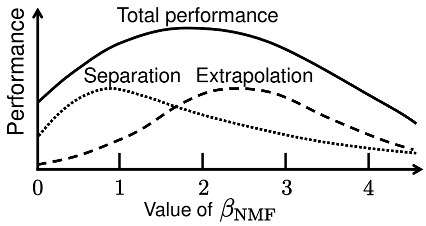

Fig. 5 Trade-off between separation and extrapolation abilities. Total performance gets highest when $\beta_{NMF} = 2$.

hybrid method based on the trade-off between separation and extrapolation abilities. From the experimental results, it can be confirmed that EUC-distance is an optimal divergence for superresolution-based SNMF.

References

[1] D. D. Lee, H. S. Seung, "Algorithms for non-negative matrix factorization," *Neural Info. Process. Syst.*, vol.13, pp.556–562, 2001.

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

[3] K. Yagi, Y. Takahashi, H. Saruwatari, K. Shikano, K. Kondo, "Music signal separation by orthogonality and maximum-distance constrained nonnegative matrix factorization with target signal information," *Proc. AES 45th*, 2012.

[4] A. Ozerov, C. Fevotte, "Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation," *IEEE Trans. Audio, Speech and Language Processing*, vol.18, no.3, pp.550–563, 2010.

[5] D. Kitamura, H. Saruwatari, K. Shikano, K. Kondo, Y. Takahashi, "Superresolution-based stereo signal separation via supervised nonnegative matrix factorization," *Proc. DSP*, no.T3C-2, 2013.

[6] S. Araki, H. Sawada, R. Mukai, S. Makino, "Underdetermined blind sparse source separation for arbitrarily arranged multiple sensors," *Signal Processing*, vol.87, no.8, pp.1833–1847, 2007.

[7] M. Nakano, H. Kameoka, J. Le Roux, Y. Kitano, N. Ono, S. Sagayama, "Convergence-guaranteed multiplicative algorithms for non-negative matrix factorization with beta-divergence," *Proc. MLSP*, 2010.

[8] E. Vincent, R. Gribonval, C. Fevotte, "Performance measurement in blind audio source separation," *IEEE Trans. Audio, Speech and Language Processing*, vol.14, no.4, pp.1462–1469, 2006.

[9] D. FitzGerald, M. Cranitch, E. Coyle, "On the use of the beta divergence for musical source separation," *Proc. Irish Signals and Systems Conference*, 2009.