ICIT 2015 The 7th International Conference on Information Technology
doi:10.15849/icit.2015.0073

Unsupervised Single Channel Source Separation with Nonnegative Matrix Factorization

A.M. Darsono, Shakir Saat, N.M. Z. Hashim, A.A.M ISA
Faculty of Electronics & Computer Engineering, Universiti Teknikal Malaysia Melaka,
Melaka, Malaysia
{abdmajid, shakir, nikzarifie, azmiawang}@utem.edu.my

***Abstract*— In this paper, a novel single channel source separation using two-dimensional nonnegative matrix factorization (NMF2D) is proposed. In NMF2D, the time-frequency (TF) profile of each source is modeled as two-dimensional convolution of the temporal code and the spectral basis. The proposed model used Beta-divergence as a cost function and updated by maximizing the joint probability of the mixing spectral basis and temporal codes using the multiplicative update rules. Results have concretely shown the effectiveness of the algorithm in blindly separating the audio sources from single channel mixture.**

Keywords- Blind Source Separation; Nonnegative Matrix Factorization ; Machine Learning; Beta Divergence.

I. INTRODUCTION

Blind source separation (BSS) refers to the statistical technique of separating a mixture of underlying source signals. BSS has become one of the promising and exciting topics with solid theoretical foundations and potential applications in the fields of signal processing, neural computation and advanced statistics. Single channel source separation (SCSS) is a branch of BSS family where the blind signal separation is achieved when only one single recording is available. For many practical applications such as audio scenarios, generally only one channel recording is available in the hardware and in such cases conventional source separation techniques are not appropriate. Several approaches have been developed to solve the MSS problem such as the computer auditory scene analysis (CASA) [1] and underdetermined BSS [2, 3]. However both techniques are supervised technique which relies on a priori knowledge of sources obtained during the training phase to perform the separation. To overcome this, nonnegative matrix factorization (NMF) [4] approach is introduced where separation is performing without using any prior knowledge about the corresponding source signal. In NMF, given the matrix, **Y** of a dimension of *F* x *N* with nonnegative elements, nonnegative matrix factorization (NMF) is the problem of approximate the factorization

$$\mathbf{Y} \approx \mathbf{WH} \tag{1}$$

where $\mathbf{W} \in \Re^{F \times C}$ and $\mathbf{H} \in \Re^{C \times N}$ are a non-negative matrices. *F* represents the frequency bins while *N* represents the time slot in the TF domain. **W** contains the spectral basis vectors while **H** represents the amplitude of each basis vector at each time point. *C* is the numbers of component from data sources being used and it is determine such that *FC+CN<<FN* so that the data can be compressed to its integral component. This problem can be formulated as the minimization of an objective function.

$$D\left(\mathbf{Y}\middle|\mathbf{WH}\right) = \sum_{f,n} d\left(Y_{f,n}\middle|\sum_{c} W_{f,c}H_{c,n}\right) \tag{2}$$

where *d* is a scalar divergence. common way to measure how close **Y** and **WH** are to use a so-called Beta divergence [5], defined by

$$d_{\beta}(y|x) = \begin{cases} \dfrac{y^{\beta}}{\beta(\beta-1)} + \dfrac{x^{\beta}}{\beta} - \dfrac{yx^{\beta-1}}{\beta-1} & \beta \in \Re \setminus \{0,1\} \\ y(\log y - \log x) + (x-y) & \beta = 1 \\ \dfrac{y}{x} - \log\dfrac{y}{x} - 1 & \beta = 0 \end{cases} \tag{3}$$

The limiting cases β =0 and β =1 correspond to the Itakura-Saito (IS) and Kullback-Leibler (KL) divergences, respectively. Another case of note is β =2 which corresponds to the Least Square (LS) distance. The Beta divergence offers a continuum of noise statistics that interpolates between these three specific cases. This paper proposed a new model of monaural source separation based on two-dimensional NMF (NMF2D) model [6] with the Beta-divergence as an objective function. We develop a novel solution that efficiently performs source separation to be used in audio source separation. The proposed solution operated in time-frequency domain and the objective function was minimized using multiplicative update rules.

The remainder of the paper is organized as follows: Single channel mixture model in the TF domain is introduced in Section II. The derivation of proposed separation technique of Beta-divergence two dimensional NMF is detailed in Section III. Section IV presents the results of experimental tests as well as the analysis. Finally, Section V concludes the paper.

doi:10.15849/icit.2015.0073

II. TWO-DIMENSIONAL NMF WITH BETA-DIVERGENCE

A. Source Model

In this section, the proposed nonnegative matrix factorization framework is derived. Firstly, we considered a source model of $\mathbf{Y}$ which is defined as a follows:

$$\mathbf{Y} \approx \sum_{\tau=0}^{\tau_{\max}} \sum_{\phi=0}^{\phi_{\max}} \overset{\downarrow\phi}{\mathbf{W}}^{\tau} \overset{\rightarrow\tau}{\mathbf{H}}^{\phi} \approx \sum_{\tau=0}^{\tau_{\max}} \sum_{\phi=0}^{\phi_{\max}} \left(\sum_{j=1}^{J} \overset{\downarrow\phi}{\mathbf{W}}_{j}^{\tau} \overset{\rightarrow\tau}{\mathbf{H}}_{j}^{\phi} \right) \tag{4}$$

where J is the number of sources. The matrix $\mathbf{W}^{\tau}$ represents the τ^{th} slice spectral basis and $\mathbf{H}^{\phi}$ represents the ϕ^{th} slice of temporal code for each spectral basis element. The vertical arrow in $\overset{\downarrow\phi}{\mathbf{W}}^{\tau}$ denotes downward shift operator which moves each element in the matrix by ϕ row down. By the same token, the horizontal arrow in $\overset{\rightarrow\tau}{\mathbf{H}}^{\phi}$ denotes the right shift operator which moves each element in the matrix by τ column to the right. This can be interpreted as follows, i.e:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \\ 1 & 2 & 3 \end{bmatrix}, \quad \overset{\rightarrow 1}{\mathbf{A}} = \begin{bmatrix} 0 & 1 & 2 \\ 0 & 1 & 2 \\ 0 & 1 & 2 \end{bmatrix}, \quad \overset{\downarrow 2}{\mathbf{A}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 2 & 3 \end{bmatrix}.$$

The factorization for NMF2D source model in (4) is based on a model that represents temporal structure and pitch change. In audio processing, the model represents each instrument by a single time-frequency profile convolved in both time and frequency by a time-pitch weight matrix. This model thoroughly decreases the number components need to model various instruments and efficiently solves the monaural source separation problem. In the following, novel algorithm of sparse NMF2D with Beta-divergence is proposed to estimate the parameter of $\overset{\downarrow\phi}{\mathbf{W}}_{j}^{\tau}$ and $\overset{\rightarrow\tau}{\mathbf{H}}_{j}^{\phi}$ from the mixture.

B. Cost Function with Multiplicative Update Rules.

Now, we incorporated the Beta-divergence as defined in (3) such that it will minimize the cost function as follow:

$$C_{\beta}\left(\mathbf{Y}\middle|\hat{\mathbf{Y}}\right) = \sum_{f,n} \left(\frac{\left(\mathbf{Y}_{f,n}\right)^{\beta}}{\beta(\beta-1)} + \frac{\left(\hat{\mathbf{Y}}_{f,n}\right)^{\beta}}{\beta} - \frac{\mathbf{Y}_{f,n}\left(\hat{\mathbf{Y}}_{f,n}\right)^{\beta-1}}{\beta-1} \right) \tag{5}$$

for $f = 1,...,F$, $n = 1,...,N$ where $\hat{\mathbf{Y}} = \sum_{j,\tau,\phi} \overset{\downarrow\phi}{\mathbf{W}}_{j}^{\tau} \overset{\rightarrow\tau}{\mathbf{H}}_{j}^{\phi}$. In this paper, we employed the multiplicative update rules which consist in updating each parameter by multiplying its value at the previous iteration by a certain coefficient. The derivatives of (5) corresponding to $\mathbf{W}^{\tau}$ and $\mathbf{H}^{\phi}$ of Beta-NMF2D are given by:

$$\begin{aligned} \frac{\partial C_{\beta}}{\partial \mathbf{W}_{f',j'}^{\tau'}} &= \frac{\partial}{\partial \mathbf{W}_{f',j'}^{\tau'}} \left(\sum_{f,n} \left(\frac{\left(\mathbf{Y}_{f,n}\right)^{\beta}}{\beta(\beta-1)} + \frac{\left(\hat{\mathbf{Y}}_{f,n}\right)^{\beta}}{\beta} - \frac{\mathbf{Y}_{f,n}\left(\hat{\mathbf{Y}}_{f,n}\right)^{\beta-1}}{\beta-1} \right) \right) \\ &= \sum_{\phi,n} \left(\left(\hat{\mathbf{Y}}_{f'+\phi,n}\right)^{\beta-1} - \left|\mathbf{Y}\right|_{f'+\phi,n}^{2} \left(\hat{\mathbf{Y}}_{f'+\phi,n}\right)^{\beta-2} \right) \mathbf{H}_{j',n-\tau'}^{\phi} \end{aligned} \tag{6}$$

and

$$\begin{aligned} \frac{\partial C_{\beta}}{\partial \mathbf{H}_{j',n'}^{\phi'}} &= \frac{\partial}{\partial \mathbf{H}_{j',n'}^{\phi'}} \left(\sum_{f,n} \left(\frac{\left(\mathbf{Y}_{f,n}\right)^{\beta}}{\beta(\beta-1)} + \frac{\left(\hat{\mathbf{Y}}_{f,n}\right)^{\beta}}{\beta} - \frac{\mathbf{Y}_{f,n}\left(\hat{\mathbf{Y}}_{f,n}\right)^{\beta-1}}{\beta-1} \right) \right) \\ &= \sum_{\tau,f} \mathbf{W}_{f-\phi',j'}^{\tau} \left(\left(\hat{\mathbf{Y}}_{f,n'+\tau}\right)^{\beta-1} - \left|\mathbf{Y}\right|_{f,n'+\tau}^{2} \left(\hat{\mathbf{Y}}_{f,n'+\tau}\right)^{\beta-2} \right) \end{aligned} \tag{7}$$

Thus, by applying the standard multiplicative update rule:

$$\mathbf{W}_{f',j'}^{\tau'} \leftarrow \mathbf{W}_{f',j'}^{\tau'} - \eta_{W} \frac{\partial C_{\beta}}{\partial \mathbf{W}_{f',j'}^{\tau'}} \quad \text{and} \quad \mathbf{H}_{j',n'}^{\phi'} \leftarrow \mathbf{H}_{j',n'}^{\phi'} - \eta_{H} \frac{\partial C_{\beta}}{\partial \mathbf{H}_{j',n'}^{\phi'}} \tag{8}$$

where η_{W} and η_{H} are positive learning rates which can be obtained by following [7], namely:

$$\eta_{W} = \frac{\mathbf{W}_{f',j'}^{\tau'}}{\sum_{\phi,n} \left(\hat{\mathbf{Y}}_{f'+\phi,n}\right)^{\beta-1} \mathbf{H}_{j',n-\tau'}^{\phi}} \quad \text{and} \quad \eta_{H} = \frac{\mathbf{H}_{j',n'}^{\phi'}}{\sum_{\tau,f} \mathbf{W}_{f-\phi',j'}^{\tau} \left(\hat{\mathbf{Y}}_{f,n'+\tau}\right)^{\beta-1}} \tag{9}$$

Thus, the multiplicative update rules for $\mathbf{W}^{\tau}$ and $\mathbf{H}^{\phi}$ become:

$$\mathbf{H}^{\phi} \leftarrow \mathbf{H}^{\phi} . \frac{\sum_{\tau} \overset{\downarrow\phi}{\mathbf{W}}^{\tau\,\mathrm{T}} \left(\left(\overset{\leftarrow\tau}{\hat{\mathbf{Y}}} \right)^{(\beta-2)} . \overset{\leftarrow\tau}{\mathbf{Y}} \right)}{\sum_{\tau} \overset{\downarrow\phi}{\mathbf{W}}^{\tau\,\mathrm{T}} \left(\overset{\leftarrow\tau}{\hat{\mathbf{Y}}} \right)^{(\beta-1)}} \tag{10}$$

and

$$\mathbf{W}^{\tau} \leftarrow \mathbf{W}^{\tau} . \frac{\sum_{\phi} \left(\left(\overset{\uparrow\phi}{\hat{\mathbf{Y}}} \right)^{(\beta-2)} . \overset{\uparrow\phi}{\mathbf{Y}} \right) \overset{\rightarrow\tau\,\mathrm{T}}{\mathbf{H}^{\phi}}}{\sum_{\phi} \left(\overset{\uparrow\phi}{\hat{\mathbf{Y}}} \right)^{(\beta-1)} \overset{\rightarrow\tau\,\mathrm{T}}{\mathbf{H}^{\phi}}} \tag{11}$$

In equations (10) and (11), *A.B* denotes element wise multiplication and $\frac{A}{B}$ denotes the element wise division.

C. Reconstruction of the Separated Sources

From mixture $\mathbf{Y}$, we seek the two estimated sources which are $\hat{\mathbf{X}}_{1} = \sum_{\tau=0}^{\tau_{\max}} \sum_{\phi=0}^{\phi_{\max}} \overset{\downarrow\phi}{\mathbf{W}}_{1}^{\tau} \overset{\rightarrow\tau}{\mathbf{H}}_{1}^{\phi}$ and $\hat{\mathbf{X}}_{2} = \sum_{\tau=0}^{\tau_{\max}} \sum_{\phi=0}^{\phi_{\max}} \overset{\downarrow\phi}{\mathbf{W}}_{2}^{\tau} \overset{\rightarrow\tau}{\mathbf{H}}_{2}^{\phi}$. Then, by

doi:10.15849/icit.2015.0073

using binary masking technique [8], we obtained mask, $\mathbf{M}_j$ as follows:

$$\mathbf{M}_j = \begin{cases} 1, & if \quad \hat{\mathbf{X}}_j > \hat{\mathbf{X}}_k \\ 0, & Otherwise \end{cases} \tag{12}$$

Then, the time domain estimated signal $\hat{\mathbf{x}}_j$ is obtained by resynthesizing $\mathbf{M}_j$ with the mixture $\mathbf{Y}$ i.e. $\hat{\mathbf{x}}_j = \text{resynthesize}\left(\mathbf{M}_j \bullet \mathbf{Y}\right)$. Here, 'resynthesize' signifies the inverse mapping of log-frequency axis to the original frequency axis and then followed by inverse short-time Fourier transform (STFT) back to the time domain.

III. Experiments & Analysis

A. Experiment Setup

The proposed algorithm is tested on audio signals containing female speech and jazz music. The mixture is approximately 6s long and sampled at 16 kHz. For audio separation, after conducting the Monte-Carlo experiments over 50 independent realizations of the mixture, the parameters of the convolutive factors of τ and ϕ shifts are set to be $\tau_{\max} = 8$ and $\phi_{\max} = 32$. This is the best realistic parameter setting to represent the temporal code and spectral basis in the factorization for most of music signals.To evaluate this, the performances of the algorithm have been measured using signal to distortion ratio (SDR) [9] which measures an overall sound quality of the source separation. SDR value which is higher than 7dB can be considered as good because it shows that there is less distortion in the recovered signal and represents an acceptable perceptual measure.

B. Audio Source Separation Results

Figure 1 show the average SDR values obtained from various values of Beta using multiplicative update NMF2D algorithm. The value of β tested was varied from 0 to 2 in steps of 0.1. It ought to cover Least Square (LS) distant, the Kullback-Leibler (KL) divergence and the Itakura-Saito (IS) divergence of NMF2D. The average separation performance was obtained from the estimated SDR value for each source in a speech-music mixture, thereby providing a measure of overall separation for each signal. From Figure 1, as we increase the value of β, the performance also increase and it reach its peak value when β=0.8 with average SDR value of 8.5dB is obtained for each source. A tail-off in performance occurs as the value of β increases from 0.8 goes up to 2. From this experiment, it suggests that beta around 0.8 is an optimal value for audio separation. Figure 2 shows the audio separation results in time domain. From Figure 2, the proposed algorithm shows the capability to separate the single mixture and recover the female speech and jazz music very well in single channel mixture. It can be seen that the separated signals almost replicate the original sources.

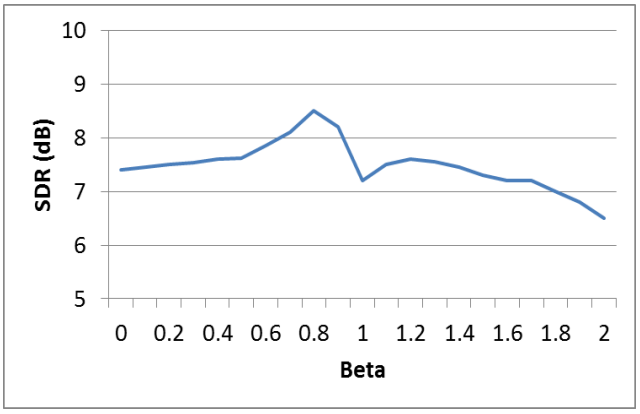

Figure 1 Separation results for various values of β using Beta-divergence NMF2D

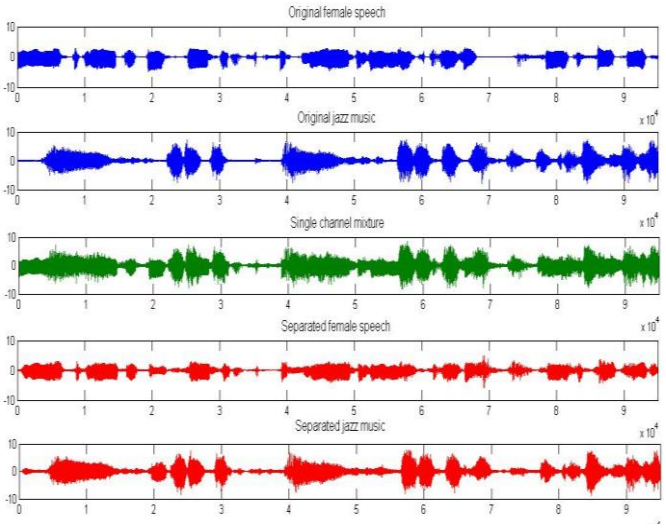

Figure 2 Audio separation results using Beta-divergence NMF2D

IV. Conclusion

The use of the Beta-divergence for audio source separation using NMF2D model has been investigated. The value of Beta-divergence with β=0.8 was found to produce an optimal result. The method proposed are computationally efficient where it avoids strong constrains of separating sources without prior knowledge of the original sources. We confirmed through an experiment that the proposed algorithm performs very well in separation of an audio mixture.

Acknowledgment

The authors would like to thank Universiti Teknikal Malaysia Melaka (UTeM) and Ministry of Education Malaysia for the

doi:10.15849/icit.2015.0073

research grant funding RAGS/2013/FKEKK/TK02/04/B0033 that makes this research work possible.

REFERENCES

[1] Li. Y, Woodruff J. and Wang D.L., 2009, "Monaural musical sound separation based on pitch and common amplitude modulation", IEEE Transaction on Audio, Speech and Language Processing, 17, 1361-1371.

[2] Gao Bin, Woo W.L., and Dlay S.S., 2008 , "Single Channel Blind Source Separation using best characteristic basis," Proceeding of 2008 3rd International Conference of Information and Communication Technologies: From Theory to Applications, 1-5.

[3] Darsono A.M, Gao Bin, Woo W.L, Dlay S.S, 2010, "Nonlinear single channel source separation", International Symposium On Communications Systems, Networks And Digital Signal Processing (CSNDSP), 507-511.

[4] Kompass R., 2005 "A generalized divergence measure for non-negative matrix factorization", Neuroinformatics workshop, Torun, Poland.

[5] Fevotte C., Bertin N., and Durrieu J.L., 2009, "Nonnegative matrix factorization with the Itakura-Saito divergence. with application to music analysis," Neural Computation, 21, 793–830.

[6] Morup M. and Schmidt M.N., 2006 "Sparse nonnegative matrix factor 2-D deconvolution," Technical Report, Technical University of Denmark, Copenhagen, Denmark..