Towards complex nonnegative matrix factorization with the beta-divergence

Paul Magron, Tuomas Virtanen

TOWARDS COMPLEX NONNEGATIVE MATRIX FACTORIZATION WITH THE BETA-DIVERGENCE

Paul Magron, Tuomas Virtanen

Laboratory of Signal Processing, Tampere University of Technology, Finland
{firstname.lastname}@tut.fi

ABSTRACT

Complex nonnegative matrix factorization (NMF) is a powerful tool for decomposing audio spectrograms while accounting for some phase information in the time-frequency domain. While its estimation was originally based on the Euclidean distance, in this paper we propose to extend it to any beta-divergence, a family of functions widely used in audio to estimate NMF. To this end, we introduce the beta-divergence in a heuristic fashion within a phase-aware probabilistic model. Estimating this model results in performing an NMF with Itakura-Saito (IS) divergence on a quantity called the phase-corrected posterior power of the sources, which is both phase-dependent and nonnegative-valued. Therefore, we replace IS with the beta-divergence, so that the factorization uses an optimal distortion metric and remains phase-aware. Even though by doing so we loose theoretical convergence guarantees, the resulting algorithm demonstrates its potential for an audio source separation task, where it outperforms previous complex NMFs approaches.

***Index Terms*—** Nonnegative matrix factorization (NMF), complex NMF, beta-divergence, anisotropic Gaussian model, audio source separation

1. INTRODUCTION

Many audio signal processing techniques act on a time-frequency (TF) representation of the data, such as the short-time Fourier transform (STFT), since the structure of audio signals is more prominent in that domain. For instance, in audio source separation [1], it is common to decompose the magnitude or power spectrogram of the mixture in order to further recover the constitutive sources [2].

To tackle this issue, many approaches rely on nonnegative matrix factorization (NMF) [3]. NMF consists in expressing a nonnegative data matrix as the product of two nonnegative matrices, respectively representing a dictionary of spectral templates and a matrix of time-varying activations. This factorization is usually obtained by minimizing a function that measures the mismatch between the data and the model. In particular, the beta-divergences [4], a family of functions which includes the Euclidean distance [3], Kulback-Leibler (KL) [5] and Itakura-Saito (IS) [6] divergences, are widely used in audio. Indeed, tuning their shape parameter β results in functions that hold interesting properties for audio [7, 8].

However, the main drawback of NMF is that it assumes the additivity of the spectrograms, which does not hold when the sources overlap in time and frequency, and may results in artifacts in the estimated signals [9]. To alleviate this issue, complex NMF [10] has been proposed. This model decomposes the complex-valued STFT of the mixture into a sum of components whose magnitudes are structured by means of an NMF. This model enables one to jointly estimate the magnitude and the phase of each source. Estimating complex NMF was originally based on the Euclidean distance, which however does not properly characterize the properties of audio (e.g., its large dynamic range).

Extending complex NMF to other divergences is cumbersome since, for instance, KL and IS only accept nonnegative-valued data as inputs. However, it has been extended to the KL divergence by using a primal-dual formulation of the optimization problem [11]. It has also recently been extended to the IS divergence [12] by considering a probabilistic framework based on anisotropic Gaussian (AG) distributions [13]. This family of distributions permits us to model the sources with non-uniform phase parameters and to structure the variance parameters by means of an NMF model. This results in an extension of the ISNMF to the case of complex-valued sources, hence its name of complex ISNMF.

In this paper, we propose to extend complex NMF to other beta-divergences. To achieve this goal, we introduce the beta-divergence in a heuristic fashion within a phase-aware probabilistic framework. We introduce an AG source model that is a simplified version of the original one [12]. Inference of this model results in minimizing the IS divergence between the NMF model and a novel quantity which is the *phase-corrected posterior power* of the latent sources. This quantity appears as particularly interesting because it is nonnegative-valued and it accounts for the phase of the sources. Therefore, performing an NMF on this quantity leads to a phase-aware decomposition of the data. We further propose to replace IS with the beta-divergence, so that in addition to be phase-aware, the factorization uses an optimal distortion metric. Minimizing this divergence with multiplicative update rules [6] results in an algorithm called complex βNMF. Experiments conducted on real audio data for a harmonic/percussive source separation task show that this approach outperforms previous complex NMF methods.

The rest of this paper is organized as follows. Section 2 presents the necessary background on NMF and complex NMF, and Section 3 introduces complex βNMF. Section 4 experimentally validates its potential for a source separation task. Finally, Section 5 draws some concluding remarks.

2. BACKGROUND

2.1. NMF with the beta-divergence

Let us consider a nonnegative-valued matrix $\mathbf{V} \in \mathbb{R}_+^{F\times T}$, which in audio is usually a magnitude or power spectrogram with F frequency channels and T time frames. NMF consists in finding a factorization $\mathbf{V} \approx \mathbf{WH}$ with $\mathbf{W} \in \mathbb{R}_+^{F\times K}$ and $\mathbf{H} \in \mathbb{R}_+^{K\times T}$, where K is the rank of the factorization. This factorization is usually obtained by minimizing an error between the data $\mathbf{V}$ and the model $\mathbf{WH}$. An important

P. Magron is supported by the Academy of Finland, project no. 290190.

class of functions in audio is the family of beta-divergences [14]. The beta-divergence between two matrices $\mathbf{A}$ and $\mathbf{B}$ with entries a_{ft} and b_{ft} is $D_\beta(\mathbf{A},\mathbf{B}) = \sum_{f,t} d_\beta(a_{ft}, b_{ft})$, with:

$$d_\beta(a,b) = \begin{cases} \dfrac{a^\beta + (\beta-1)b^\beta - \beta a b^{\beta-1}}{\beta(\beta-1)} & \beta \in \mathbb{R}\backslash\{0,1\} \\ a\log\dfrac{a}{b} + b - a & \beta = 1 \\ \dfrac{a}{b} - \log\dfrac{a}{b} - 1 & \beta = 0. \end{cases} \tag{1}$$

Special cases are the Euclidean distance ($\beta = 2$), and KL ($\beta = 1$) and IS ($\beta = 0$) divergences. This divergence is usually minimized by applying the following multiplicative updates [4]:

$$\mathbf{W} \leftarrow \mathbf{W} \odot \frac{([\mathbf{WH}]^{\odot\beta-2} \odot \mathbf{V})\mathbf{H}^T}{[\mathbf{WH}]^{\odot\beta-1}\mathbf{H}^T}, \tag{2}$$

and

$$\mathbf{H} \leftarrow \mathbf{H} \odot \frac{\mathbf{W}^T([\mathbf{WH}]^{\odot\beta-2} \odot \mathbf{V})}{\mathbf{W}^T[\mathbf{WH}]^{\odot\beta-1}}, \tag{3}$$

where $\odot$, $.^{\odot}$ and $\div$ respectively denote the element-wise matrix multiplication, power and division, and ^{T} is the matrix transposition.

2.2. Statistical interpretation of ISNMF

NMF with beta-divergence can be framed in a general probabilistic framework using Tweedie distributions [15], but a particularly interesting case is ISNMF [6], which we detail hereafter.

Let $\mathbf{X} \in \mathbb{C}^{F\times T}$ be the STFT of a single-channel audio signal. $\mathbf{X}$ is the instantaneous mixture of J sources $\mathbf{S}_j \in \mathbb{C}^{F\times T}$:

$$\mathbf{X} = \sum_j \mathbf{S}_j. \tag{4}$$

The TF coefficients of all sources are modeled with independent circularly-symmetric Gaussian random variables: $s_{j,ft} \sim \mathcal{N}(0, v_{j,ft} I)$, where I is the identity matrix, and the variances follow an NMF model: $\mathbf{V}_j = \mathbf{W}_j\mathbf{H}_j$, where $\mathbf{W}_j \in \mathbb{R}_+^{F\times K_j}$ and $\mathbf{H}_j \in \mathbb{R}_+^{K_j\times T}$. Then, thanks to the additivity of the Gaussian distributions, the mixture given by (4) is also Gaussian so $x_{ft} \sim \mathcal{N}(0, v_{x,ft} I)$ with $\mathbf{V}_x = \mathbf{WH} = \sum_j \mathbf{W}_j\mathbf{H}_j$.

To estimate $\mathbf{W}$ and $\mathbf{H}$, an intuitive approach in such a probabilistic framework consists in maximizing the likelihood of the data. It can be shown [6] that the maximum likelihood estimation of this model is equivalent to performing an NMF with IS divergence on $\mathbf{V} = |\mathbf{X}|^{\odot 2}$, hence the name of ISNMF model. This IS divergence is usually minimized with the multiplicative updates (2) and (3), or with variants of the expectation-maximization (EM) algorithm [6].

2.3. Complex NMF

Complex NMF [10] consists in directly modeling the complex-valued STFT of the mixture as a sum of components whose magnitudes are structured by means of an NMF model:

$$x_{ft} \approx \sum_{k=1}^K w_{fk} h_{kt} e^{ij\phi_{k,ft}}. \tag{5}$$

We rather consider here a model known as complex NMF with intra-source additivity [16]: it consists in modeling the phase of each source instead of the phase of each rank-1 component. This significantly reduces the number of parameters of the model, thus it lowers both the memory and computation time required for the estimation of the model, at the cost of a moderate drop in terms of separation quality [16]. Grouping the components into J sources then leads to:

$$x_{ft} \approx \sum_{j=1}^J [\mathbf{W}_j\mathbf{H}_j]_{ft} e^{j\phi_{j,ft}}, \tag{6}$$

It is usually estimated by minimizing the Euclidean distance between the data and the model (6), and its main advantage is that it enables us to promote phase constraints [17, 18]. However, this distortion metric is not well adapted to audio [7]. To alleviate this issue, complex NMF has recently been extended to KL [11] and IS [12] divergences.

3. COMPLEX NMF WITH THE BETA-DIVERGENCE

Here, we extend complex NMF to any beta-divergence. This is done in a probabilistic framework in which inference of the parameters results in applying NMF to a quantity that is both phase-aware and nonnegative. It then becomes possible to perform this factorization with the beta-divergence.

3.1. Anisotropic Gaussian model

First, we introduce a probabilistic source model based on the AG distribution, which is a simplified version of the model in [12]. Indeed, we assume here that the sources are centered (i.e., their mean is null), which was not the case in the original model [12]. This novel model is easier to manipulate and it preserves the possibility to account for a phase value.

Our approach consists in generalizing the Gaussian model presented in Section 2.2, by considering sources $s_{j,ft} \sim \mathcal{N}(0, \Gamma_{j,ft})$ whose covariance matrices $\Gamma_{j,ft}$ are no longer diagonal:

$$\Gamma_{j,ft} = \begin{pmatrix} \gamma_{j,ft} & c_{j,ft} \\ \bar{c}_{j,ft} & \gamma_{j,ft} \end{pmatrix}, \tag{7}$$

where $\gamma_{j,ft} = \mathbb{E}(|s_{j,ft}|^2) \in \mathbb{R}_+$ and $c_{j,ft} = \mathbb{E}(s_{j,ft}^2) \in \mathbb{C}$ are the variance and relation term of $s_{j,ft}$, and $\bar{z}$ denotes the complex conjugate of z. This approach allows us to account for a phase value through non-null relation parameters. Drawing on [12], we define the variance and relation terms as follows:

$$\begin{aligned} \gamma_{j,ft} &= (1-\lambda^2)[\mathbf{W}_j\mathbf{H}_j]_{ft}, \\ c_{j,ft} &= \rho[\mathbf{W}_j\mathbf{H}_j]_{ft} e^{i2\mu_{j,ft}}. \end{aligned} \tag{8}$$

where $\mu_{j,ft}$ is a phase location parameter, and

$$\lambda = \frac{\sqrt{\pi}}{2}\frac{I_1(\kappa)}{I_0(\kappa)} \text{ and } \rho = \frac{I_2(\kappa)}{I_0(\kappa)} - \lambda^2. \tag{9}$$

where I_q is the modified Bessel function of the first kind of order q, and $\kappa \geq 0$ is an *anisotropy* parameter which quantifies how important the phase location parameter is. In particular, if $\kappa = 0$, then $\lambda = \rho = 0$ and $\Gamma_{j,ft} = v_{j,ft} I$. Consequently, the distribution becomes circularly-symmetric and this model becomes equivalent to ISNMF, as presented in Section 2.2.

3.2. Inference

We propose to estimate the model parameters $\Theta = \{\mathbf{W}, \mathbf{H}, \boldsymbol{\mu}\}$ with the EM algorithm [19], which consists in maximizing the following lower bound of the data log-likelihood:

$$Q(\Theta, \Theta') = \int p(\mathbf{Z}|\mathbf{X};\Theta') \log p(\mathbf{X}, \mathbf{Z}; \Theta) d\mathbf{Z}, \tag{10}$$

where Θ' contains the current set of estimated parameters, and $\mathbf{Z}$ denotes a set of latent (hidden) variables. The EM algorithm consists in alternatively computing the functional Q given the current set of parameters Θ' (E-step) and maximizing it with respect to Θ (M-step). This is proven [19] to increase the value of the criterion (10). Due to the mixing constraint (4), we consider, as in [12], a reduced set of $J-1$ free variables $\mathbf{Z} = \{\{s_{j,ft}\}_{j=1}^{J-1}\}_{ft}$.

Due to space constraints, we do not detail the full derivation of the E-step, but it can be found in [12]. It results in maximizing the following functional:

$$Q(\Theta, \Theta') \stackrel{c}{=} -\sum_{f,t}\sum_{j=1}^{J} \log(\sqrt{|\Gamma_{j,ft}|}) + \frac{1}{|\Gamma_{j,ft}|}\left(\gamma_{j,ft}(|m'_{j,ft}|^2 + \gamma'_{j,ft}) - \Re(\bar{c}_{j,ft}(m'^2_{j,ft} + c'_{j,ft}))\right), \tag{11}$$

where $|A|$ is the determinant of a matrix A, $\stackrel{c}{=}$ denotes equality up to an additive constant, and the posterior mean and covariance matrix are given by:

$$m'_{j,ft} = \begin{pmatrix}\gamma_{j,ft} & c_{j,ft}\end{pmatrix}\Gamma^{-1}_{x,ft}\begin{pmatrix}x_{ft}\\ \bar{x}_{ft}\end{pmatrix}, \tag{12}$$

$$\Gamma'_{j,ft} = \begin{pmatrix}\gamma'_{j,ft} & c'_{j,ft}\\ \bar{c}'_{j,ft} & \gamma'_{j,ft}\end{pmatrix} = \Gamma_{j,ft} - \Gamma_{j,ft}\Gamma^{-1}_{x,ft}\Gamma_{j,ft}. \tag{13}$$

We can rewrite the functional (11) by isolating the phase parameters. This leads to minimizing:

$$\Re\left(e^{-2i\mu_{j,ft}}(c'_{j,ft} + m'^2_{j,ft})\right), \tag{14}$$

which is easily achieved with the following update:

$$\mu_{j,ft} \leftarrow \frac{1}{2}\angle(m'^2_{j,ft} + c'_{j,ft}) - \frac{\pi}{2}. \tag{15}$$

Finally, we rewrite (11) to put the emphasis on the NMF parameters:

$$Q(\Theta|\Theta') \stackrel{c}{=} -\sum_{j=1}^{J}\sum_{f,t} \log([\mathbf{W}_j\mathbf{H}_j]_{ft}) + \frac{p_{j,ft}}{[\mathbf{W}_j\mathbf{H}_j]_{ft}}, \tag{16}$$

where $p_{j,ft}$ is the *phase-corrected posterior power* of $s_{j,ft}$:

$$p = \frac{(1-\lambda^2)\left(\gamma' + |m'|^2\right) - \rho\Re\left(e^{-2i\mu}(c' + m'^2)\right)}{(1-\lambda^2)^2 - \rho^2}, \tag{17}$$

where we removed the indices j, ft for clarity. We recognize in (16) the IS divergence between $\mathbf{P}_j$ and $\mathbf{W}_j\mathbf{H}_j$. In the original complex ISNMF model [12], this divergence was slightly different because it included an extra-term that was due to the non-zero means in the AG model (*cf.* Section 3.1). Therefore, we proposed in [12] to minimize the corresponding function with the auxiliary function method. However, we can here use the multiplicative updates (*cf.* (2) and (3)) where $\mathbf{V}$, $\mathbf{W}$ and $\mathbf{H}$ are respectively replaced by $\mathbf{P}_j$, $\mathbf{W}_j$ and $\mathbf{H}_j$.

3.3. Introducing the β-divergence

The key aspect of this derivation was to highlight a novel quantity $p_{j,ft}$ which is both nonnegative and phase-aware. It was therefore possible to extend complex NMF to the IS divergence by performing the factorization on this quantity rather than on the magnitude spectrogram of the mixture $\mathbf{V}$, which does not account for the phase.

Algorithm 1: Complex βNMF

Inputs: Mixture $\mathbf{X} \in \mathbb{C}^{F\times T}$
Anisotropy $\kappa \in \mathbb{R}_+$ and divergence $\beta \in \mathbb{R}_+$
Initial NMF matrices $\forall j$, $\mathbf{W}_j \in \mathbb{R}_+^{F\times K_j}$, $\mathbf{H}_j \in \mathbb{R}_+^{K_j \times T}$
Initial phases $\forall j$, $\mu_j \in [0, 2\pi[^{F\times T}$
Initialization: compute λ and ρ with (9)
while stopping criterion not reached **do**
    % E-step
    Update γ and c with (8)
    $\Gamma_x = \sum_{j=1}^{J}\Gamma_j$
    Update m' with (12) and Γ' with (13)
    Update p with (17)
    % M-step
    $\forall j$ Update $\mathbf{W}_j$ with (19)
    Update $\mathbf{H}_j$ with (20)
    Normalize $\mathbf{W}_j$ and $\mathbf{H}_j$
    Update $\boldsymbol{\mu}_j$ with (15)
end
Update γ and c with (8)
$\Gamma_x = \sum_{j=1}^{J}\Gamma_j$
Update m' with (12),
Outputs: $m' \in \mathbb{C}^{J\times F\times T}$.

Consequently, we propose here to further extend complex NMF to any beta-divergence. Replacing the IS cost function in (16) by the beta-divergence leads to maximizing the following functional:

$$Q_\beta(\Theta|\Theta') \stackrel{c}{=} -\sum_{j=1}^{J}\sum_{f,t} d_\beta(p_{j,ft}, [\mathbf{W}_j\mathbf{H}_j]_{ft}). \tag{18}$$

for which we apply the multiplicative updates (2) and (3):

$$\mathbf{W}_j \leftarrow \mathbf{W}_j \odot \frac{(\mathbf{P}_j \odot \mathbf{V}_j^{\odot\beta-2})\mathbf{H}_j^T}{(\mathbf{V}_j^{\odot\beta-1})\mathbf{H}_j^T}, \tag{19}$$

and

$$\mathbf{H}_j \leftarrow \mathbf{H}_j \odot \frac{\mathbf{W}_j^T(\mathbf{P}_j \odot \mathbf{V}_j^{\odot\beta-2})}{\mathbf{W}_j^T(\mathbf{V}_j^{\odot\beta-1})}. \tag{20}$$

The phase update (15) is left unchanged, though in future work we will investigate on alternative phase update schemes. This procedure is called complex βNMF and it is summarized in Algorithm 1. This algorithm includes a normalization step after updating $\mathbf{W}_j$ and $\mathbf{H}_j$, which eliminates trivial scale indeterminacies and avoids numerical instabilities. We impose a unitary ℓ_2-norm on each column of $\mathbf{W}_j$ and scale $\mathbf{H}_j$ accordingly, so that the cost function is not affected. One final E-step is performed after looping in order to estimate the sources with the most up-to-date set of parameters.

We derived here a procedure that extends complex NMF to any beta-divergence in a heuristic way. Indeed, replacing the IS divergence in (16) by the beta-divergence results in loosing the convergence properties of the algorithm. However, we observed in our experiments that the objective function (i.e., the log-likelihood of the data in the AG model) was still non-decreasing under those updates. Note that this algorithm is also inconsistent regarding the underlying statistical model. We leave to future work the design of alternative phase-aware probabilistic models in which the beta-divergence (18) naturally arises as the likelihood function.

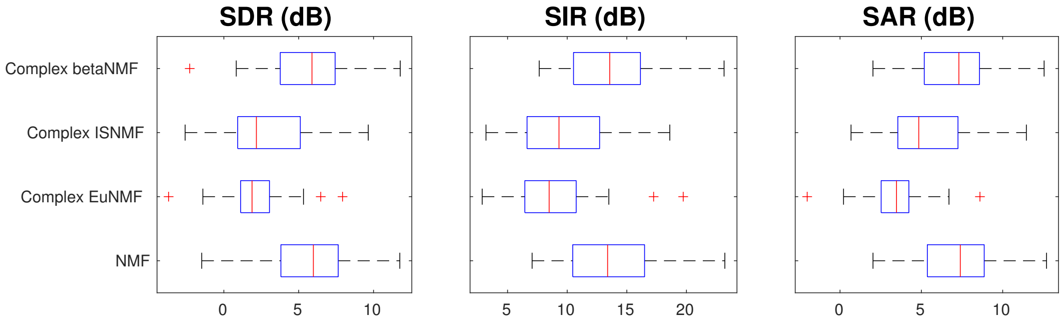

Fig. 1. Source separation performance on the DSD100 test dataset. Each box-plot is made up of a central line indicating the median, box edges indicating the 1^{st} and 3^{rd} quartiles, whiskers indicating the minimum and maximum values, and crosses representing the outliers.

4. EXPERIMENTAL EVALUATION

4.1. Setup

We propose to asses the potential of complex βNMF for a harmonic/percussive source separation task. As audio data, we use music song excerpts from the DSD100 database [20]. Each excerpt is 20 seconds long and is made up of $J = 2$ sources: a percussive source (the `drums` track) and a harmonic source (the sum of the other tracks). The database consists of two subsets of 50 songs (learning and test sets). The signals are sampled at 44100 Hz and the STFT is computed with a 92 ms Hann window and 75 % overlap.

We consider a supervised separation scenario. Each excerpt is split into two signals of 10 seconds. The first segment is used for learning the dictionaries $\mathbf{W}_j$ on the power spectrogram of each isolated track, by means of a k-means clustering algorithm with $K = 50$ basis per dictionary. The second segment is used for performing the separation, so only the activation matrices $\mathbf{H}_j$ and phase parameters $\boldsymbol{\mu}_j$ are computed. For a fair comparison, all the algorithms use the same random initial matrices $\mathbf{H}_j$ and 100 iterations, and the complex NMFs are initialized with the mixture's phase.

Source separation quality is measured with the signal-to-distortion, signal-to-interference, and signal-to-artifact ratios (SDR, SIR, and SAR) [21] expressed in dB.

The code of this experimental study is available online[1].

4.2. Learning β

We first consider the 50 songs that form the learning subset in order to learn the optimal β parameter for complex βNMF. We use the value $\kappa = 1$ for the anisotropy parameter, which yielded good results in previous experiments [12, 13]. Results are presented in Fig. 2, and suggest that the value $\beta = 0.5$ leads to the best trade-off between the different indicators. We also learn the optimal β for the classical NMF presented in Section 2.1. Previous works have been conducted on the optimal beta-divergence [14] for NMF, so we simply perform here a basic learning for this specific task. Similarly to complex βNMF, we obtain the best results for $\beta = 0.5$.

4.3. Comparison with other methods

We now compare complex βNMF with other approaches. We test the phase-unaware NMF, complex NMF with Euclidean distance (complex EuNMF) as presented in Section 2.3 and complex ISNMF as introduced in [12]. NMF and complex βNMF use the value $\beta = 0.5$ in conformity with the results of the previous experiment. Finally, even though the complex NMF with KL divergence introduced in [11] would have been an appropriate comparison reference, we were unfortunately not able to re-implement it in this framework.

We run the algorithms on the test set and present the results in Fig 1. Firstly, we remark that the previous complex NMFs are outperformed by NMF: even if it is phase-unaware, it uses a distortion metric that is optimally chosen for this task. We observe that the proposed complex βNMF outperforms the other complex NMFs for all indicators, which confirms the interest of extending complex NMF to any beta-divergence. It performs slightly better than NMF (an improvement of 0.1, 0.3 and 0.1 dB in median SDR, SIR and SAR respectively), though the difference is not statistically significant. However, it should be noted that the full potential of complex NMFs relies on the possibility of incorporating phase constraints [18], which is not possible in the classical NMF model. This complex βNMF technique is therefore a promising tool for phase-aware source separation.

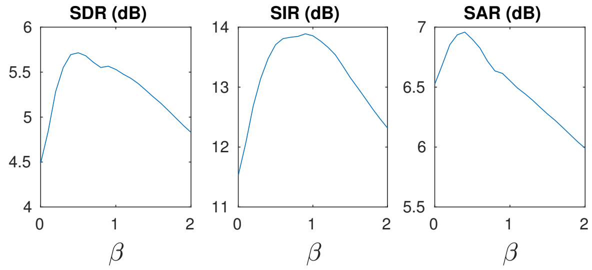

Fig. 2. Influence of β on the source separation quality for the complex βNMF algorithm on the DSD100 learning set.

5. CONCLUSION

In this paper, we extended complex NMF to any beta-divergence. This novel algorithm jointly estimates the magnitudes and phases of the sources by using a distortion metric well adapted to audio. We proposed to perform the factorization on the phase-corrected posterior power of the sources, a phase-dependent and nonnegative quantity. This technique outperformed previous complex NMF approaches in source separation experiments. Future work will focus on obtaining convergence guarantees for complex βNMF by formulating it in an end-to-end probabilistic framework rather than in a heuristic fashion. Besides, we will incorporate phase constraints in this framework in order to fully exploit its potential [12, 18, 22].

[1] `https://github.com/magronp/complex-beta-nmf`

6. REFERENCES

[1] P. Comon and C. Jutten, *Handbook of blind source separation: independent component analysis and applications*, Academic press, 2010.

[2] T. Virtanen, J. F. Gemmeke, B. Raj, and P. Smaragdis, "Compositional models for audio processing: Uncovering the structure of sound mixtures," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 125–144, March 2015.

[3] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[4] C. Févotte and J. Idier, "Algorithms for nonnegative matrix factorization with the beta-divergence," *Neural Computation*, vol. 23, no. 9, pp. 2421–2456, Sept 2011.

[5] T. Virtanen, "Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 3, pp. 1066–1074, March 2007.

[6] C. Févotte, 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, March 2009.

[7] R. Gray, A. Buzo, A. Gray, and Y. Matsuyama, "Distortion measures for speech processing," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 4, pp. 367–376, August 1980.

[8] B. King, C. Févotte, and P. Smaragdis, "Optimal cost function and magnitude power for NMF-based speech separation and music interpolation," in *Proc. of IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, September 2012.

[9] P. Magron, R. Badeau, and B. David, "Model-based STFT phase recovery for audio source separation," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 26, no. 6, pp. 1095–1105, June 2018.

[10] H. Kameoka, N. Ono, K. Kashino, and S. Sagayama, "Complex NMF: A new sparse representation for acoustic signals," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2009.

[11] H. Kameoka, H. Kagami, and M. Yukawa, "Complex NMF with the generalized Kullback-Leibler divergence," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2017.

[12] P. Magron and T. Virtanen, "Complex ISNMF: a phase-aware model for monaural audio source separation," 2018, https://arxiv.org/abs/1802.03156.

[13] P. Magron, R. Badeau, and B. David, "Phase-dependent anisotropic Gaussian model for audio source separation," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2017.

[14] D. Fitzgerald, M. Cranitch, and E. Coyle, "On the use of the beta divergence for musical source separation," in *Proc. of IET Irish Signals and Systems Conference (ISSC)*, June 2008.

[15] U. Simsekli, A. T. Cemgil, and Y. K. Yilmaz, "Learning the beta-divergence in Tweedie compound Poisson matrix factorization models," in *Proc. of International Conference on Machine Learning (ICML)*, June 2013.

[16] B. King, *New Methods of Complex Matrix Factorization for Single-Channel Source Separation and Analysis*, Ph.D. thesis, University of Washington, 2012.

[17] J. Le Roux, H. Kameoka, E. Vincent, N. Ono, K. Kashino, and S. Sagayama, "Complex NMF under spectrogram consistency constraints," in *Proc. of Acoustical Society of Japan Autumn Meeting*, September 2009.

[18] P. Magron, R. Badeau, and B. David, "Complex NMF under phase constraints based on signal modeling: application to audio source separation," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016.

[19] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the royal statistical society. Series B (methodological)*, vol. 39, no. 1, pp. 1–38, 1977.

[20] A. Liutkus, F-R. Stöter, Z. Rafii, D. Kitamura, B. Rivet, N. Ito, N. Ono, and J. Fontecave, "The 2016 signal separation evaluation campaign," in *Proc. of International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA)*, 2017.

[21] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," *IEEE Transactions on Speech and Audio Processing*, vol. 14, no. 4, pp. 1462–1469, July 2006.

[22] J. Bronson and P. Depalle, "Phase constrained complex NMF: Separating overlapping partials in mixtures of harmonic musical sources," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014.