Generalization of an Active Set Newton Algorithm with Alpha-Beta divergences for audio separation

1st Auxiliadora Sarmiento Vega
Department of Teoría de la Señal y Comunicaciones
Escuela Técnica Superior de Ingeniería
Universidad de Sevilla, Seville, Spain
0000-0003-2587-1382

2nd Iván Durán Díaz
Department of Teoría de la Señal y Comunicaciones
Escuela Técnica Superior de Ingeniería
Universidad de Sevilla, Seville, Spain
0000-0001-6097-0402

3rd Irene Fondón
Department of Teoría de la Señal y Comunicaciones
Escuela Técnica Superior de Ingeniería
Universidad de Sevilla, Seville, Spain
0000-0002-8955-7109

4th Sergio Cruces
Department of Teoría de la Señal y Comunicaciones
Escuela Técnica Superior de Ingeniería
Universidad de Sevilla, Seville, Spain
0000-0003-4121-7137

***Abstract*—This article considers the decomposition of a non-negative signal into a non-negative linear combination of the contributions of pre-specified atomic units, which are also non-negative. This model, referred as compositional model, is evident in the time-frequency characterisations of audio signals, where the sound can be viewed as a blending of spectral patterns of the component sounds that are present simultaneously. The algorithm proposed in this article obtains the activation vector of the atoms through an Active-Set Newton algorithm that employ the Alpha-Beta-divergence between the observed signal and the decomposition. This divergence family has been proved to be more efficient than other more common divergences, such as the generic Kullback-Leibler divergence in various audio signal processing applications. We have evaluated the proposed algorithm in a signal separation application of polyphonic music.**

***Index Terms*—Alpha-Beta divergence, Active Set-Newton algorithm, signal separation, dictionary learning, compositional model**

I. Introduction

The problem of signal separation is one of the most popular in the field of signal processing. In its broadest definition, it consists of the recovery of a series of source signals that are mixed in a set of observations. When the source signals are audio signals, the compositional model [1] has been effective in solving the problem of signal separation, both blind [2] and supervised [3], [4]. The compositional model of sound considers that any audio signal can be represented through an additive combination of elementary sound units or atoms, a dictionary being a collection of atoms representative of a certain type of sound.

Mathematically, the compositional model can be expressed as follows. Given a non-negative column vector $\mathbf{x}$ of length F, a set of N atoms $\mathbf{b}_n \in \mathbb{R}_+^{F\times 1}$ and N non-negative weights w_n, with $n = 1 \cdots N$, the compositional model of $\mathbf{x}$ can be written as follows:

$$\mathbf{x} \approx \hat{\mathbf{x}} = \sum_{n=1}^{N} w_n \mathbf{b}_n, \quad w_n \geq 0, \quad \forall n \tag{1}$$

and in matrix form as:

$$\mathbf{x} \approx \hat{\mathbf{x}} = \mathbf{B}\mathbf{w}, \quad \mathbf{w} \geq 0 \tag{2}$$

where $\mathbf{w} = [w_1, \cdots, w_N]^T$ is the vector of weights and $\mathbf{B} = [\mathbf{b}_1, \cdots, \mathbf{b}_N]$ is the dictionary, a matrix in whose columns the atoms are arranged.

The compositional model is most evident in the time-frequency transformed domain, e.g. through the Short-Time Fourier Transform (STFT). In this transformed domain, each atom corresponds to a spectral pattern of the audio signal (in magnitude or power). The dictionaries of each type of sound would therefore be composed of a set of spectral patterns characteristic of that type of sound. In the vast majority of cases, a given type of sound has a large number of spectral patterns associated with it. For example, the sound coming from a certain musical instrument must contemplate the different spectral patterns that are obtained depending on the musical note, the attack or stabilisation phase of the sound as well as the different materials and/or manufacturers of the instrument. For all these reasons, audio signal dictionaries are usually composed of a high number of atoms, so that the compositional model of the audio signal usually has a sparse or low-density characteristic, i.e. with a small number of non-zero weights in the decomposition.

Signal separation using the compositional model of sound considers that the magnitude of the STFT of each source signal $\mathbf{S}_i(f,t)$ can be approximated by an additive combination of atoms that are in known and pre-trained dictionaries, which

This research was funded by Ministerio de Economía, Industria y Competitividad (MINECO) of the Government of Spain, grant number TEC2017-82807-P, by the European Regional Development Fund (ERDF) of the European Commission and by FEDER/Junta de Andalucía-Consejería de Economía y Conocimiento/ Project (US-1264994).

have been previously estimated for that type of audio signal [5]. Therefore, the magnitude of the STFT of a mixture of M audio signals $\mathbf{X}(f,t)$, can be expressed as an additive sum of atoms stored in M pre-trained dictionaries $\mathbf{B}_i(f)$ where $i = 1, \cdots, M$. The separation of the source signals is achieved by estimating the matrix of weights $\mathbf{w}$ that approximates the observation to the compositional model at each time index t of the STFT:

$$\hat{\mathbf{X}}(f,t) = \mathbf{B}(f)\mathbf{w}(f,t) \tag{3}$$

where $\mathbf{B}(f,t)$ is the matrix of dictionaries composed of the concatenation of the M pre-trained dictionaries $\mathbf{B}_i(f)$:

$$\mathbf{B}(f) = [\mathbf{B}_1(f)\mathbf{B}_2(f)\cdots\mathbf{B}_M(f)] \tag{4}$$

and $\mathbf{w} = \left[\mathbf{w}_1(f,t)^T\mathbf{w}_2(f,t)^T\cdots\mathbf{w}_M(f,t)^T\right]^T$.

Once the weights are estimated, the magnitude of the STFT of each source signal $\hat{\mathbf{S}}_i(f,t)$ is obtained by "Wiener-style" reconstruction as:

$$\hat{\mathbf{X}}_i(f,t) = \mathbf{X}_i(f,t) \circ \frac{\mathbf{B}_i(f)\mathbf{w}_i(f,t)}{\mathbf{B}(f)\mathbf{w}(f,t)} \tag{5}$$

and then reconstructed to the time domain through the inverse STFT copying the phases from the mixture spectrum.

The decomposition of a signal following the compositional model can be interpreted as a non-negative matrix factorisation (NMF) problem in which the decomposition is performed by minimising a distance or divergence measure between the observation $\mathbf{X}(f,t)$ and the linear model $\hat{\mathbf{X}}(f,t)$, where it is necessary to introduce sparse constraints to take into account the sparsity characteristic of the compositional model of the audio signal. However, these methods are computationally very expensive when the dictionaries are very large and when it is necessary to add sparse constraint regularisation terms.

Another family of methods that are computationally less expensive than the NMF methods are the active set methods that are part of the large set of iterative methods for the optimisation of non-linear functions with linear constraints. In our problem to be solved, the linear constraints consist in the non-negativity of the weights to be estimated. This type of iterative strategy has been used for the separation of audio signals, resulting in the ASNA algorithm (acronym for Active-Set Newton Algorithm) proposed in [5] and in [6] with the addition of sparse constraints. The original ASNA algorithm minimised the Kullback-Leibler (KL) divergence between the observation $\mathbf{x}$ and the linear model $\mathbf{x}$ by employing Newton's algorithm and the full Hessian matrix for optimisation. In [4] the ASNA-AB algorithm was proposed by replacing the KL divergence in the cost function by the family of Alpha-Beta divergences. This family of divergences, governed by two parameters α and β, has a huge potential since it integrates and connects a large number of already known divergences, such as the Kullback-Leibler, Itakura-Saito, Alpha-divergences, etc. Moreover, it has been shown to be useful in several applications related to audio signal processing, such as for the separation of convolutional mixtures of speech [7], the recognition of speech in noise [8] and the classification of audio signals into musical genres [9].

The ASNA-AB algorithm proposed in [4] considered the parameterisation of the most general Alpha-Beta divergence, obtained when $\alpha, \beta, \alpha+\beta \neq 0$. In this article the method has been extended to all possible values of $\alpha, \beta \in \mathbb{R}$.

The organisation of this article is as follows: Section II presents the complete formulation of the Alpha-Beta divergence and some of its most interesting properties for audio signal processing; Section III details the ASNA-AB algorithm for all possible values of $\alpha, \beta \in \mathbb{R}$; Section IV illustrates the performance of the algorithm for audio signal separation while Section V presents the conclusions.

II. The family of Alpha-Beta divergences

The Alpha-Beta divergences, introduced in [12], are a measure of dissimilarity between positive data, governed by two parameters α and β. Given two non-negative matrices $\mathbf{P} \in \mathbb{R}_+^{I\times T}$ and $\mathbf{Q} \in \mathbb{R}_+^{I\times T}$ with entries $p_{it} = [\mathbf{P}]_{it}$ and $q_{it} = [\mathbf{Q}]_{it}$, the Alpha-Beta divergence is given by:

$$D_{AB}^{\alpha,\beta}(\mathbf{P}\|\mathbf{Q}) = \sum_{it} d_{AB}^{\alpha,\beta}(p_{it}, q_{it}) \tag{6}$$

where $d_{AB}^{\alpha,\beta}(p_{it}, q_{it})$ is defined as (7).

Several divergences and distances employed in the field of audio signal processing belong to the family of Alpha-Beta divergences. For example, the Kullback-Leibler divergence is obtained for $(\alpha,\beta) = (1,0)$, the Itakura-Saito divergence for $(\alpha,\beta) = (1,-1)$, the Alpha-divergences for $\alpha+\beta = 1$ and the Beta-divergences for $\alpha = 1$.

In the context of audio signal decomposition, the scaling property of the Alpha-Beta divergence is of particular interest. This property states that we can control the scaling of the divergence arguments by adjusting the α and β parameters even obtaining scaling invariant divergences when $\alpha+\beta = 0$:

$$D_{AB}^{\alpha,\beta}(c\mathbf{P}\|c\mathbf{Q}) = c^{\alpha+\beta} D_{AB}^{\alpha,\beta}(\mathbf{P}\|\mathbf{Q}) \tag{8}$$

The dynamic range of audio signals can be high and the energy of the high-frequency components are, in general, lower than the low-frequency components. If the divergence used in the decomposition of an audio signal is not scaling invariant, the magnitude of the errors made in the decompositions can be much larger for low frequency components than for high frequencies [1].

Another relevant feature of the Alpha-Beta divergences is that they are convex with respect to the second argument q_{it} for values of (α,β) that lie within the convex cone bounded by $\alpha+\beta = 1$ and $\beta = 1$ [12]. Nevertheless, this convex region widens for sufficiently small relative errors between p_{it} and q_{it}, as we will see in the simulation results in Section IV.

III. The ASNA-AB algorithm

The ASNA-AB algorithm we propose in this work, as does the ASNA algorithm in [5] and the algorithm proposed in [4], iteratively update the active set $\mathcal{A}$ to find the optimal

$$d_{AB}^{\alpha,\beta}(p_t,q_t)=\begin{cases}-\frac{1}{\alpha\beta}\left(p_t^{\alpha}q_t^{\beta}-\frac{\alpha}{\alpha+\beta}p_t^{\alpha+\beta}-\frac{\beta}{\alpha+\beta}q_t^{\alpha+\beta}\right), & \text{for } \alpha,\beta,\alpha+\beta\neq 0\\ \frac{1}{\alpha^2}\left(p_t^{\alpha}\ln\frac{p_t^{\alpha}}{q_t^{\alpha}}-p_t^{\alpha}+q_t^{\alpha}\right), & \text{for } \alpha\neq 0,\beta=0\\ \frac{1}{\alpha^2}\left(\ln\frac{p_t^{\alpha}}{q_t^{\alpha}}+\left(\frac{p_t^{\alpha}}{q_t^{\alpha}}\right)^{-1}-1\right), & \text{for } \alpha=-\beta\neq 0\\ \frac{1}{\beta^2}\left(q_t^{\beta}\ln\frac{q_t^{\beta}}{p_t^{\beta}}-q_t^{\beta}+p_t^{\beta}\right), & \text{for } \alpha=0,\beta\neq 0\\ \frac{1}{2}(\ln p_t-\ln q_t)^2, & \text{for } \alpha,\beta=0\end{cases} \tag{7}$$

set of active atoms. The active set $\mathcal{A}$ is therefore composed of the indices corresponding to the atoms of the dictionary matrix with non-zero weights. At each iteration the weights of the active atoms are updated using Newton's algorithm. For simplicity of notation we will describe the algorithm for a generic observation $\mathbf{x}$ and later, in Section IV, we will describe the use of the algorithm for the decomposition of an audio mixture $\mathbf{X}(f,t)$.

The proposed algorithm consists of the following steps:

Step 1. Inicialization

First, the dictionary atoms are normalised to have Euclidean unit norm. Then the optimal weights of the atoms w_n are obtained by minimising the Alpha-Beta divergence between the observation vector and the corresponding atom $\mathbf{b}_n$:

$$w_n=\arg\min_{w_n} D_{AB}\left(\mathbf{x}\|w_n\mathbf{b}_n\right) \tag{9}$$

The solution to this optimisation problem is obtained by deriving the above cost function with respect to the weight w_n and equalling zero, yielding:

$$w_n=\begin{cases}\left(\frac{\mathbf{x}^{\alpha T}\mathbf{b}_n^{\beta}}{\mathbf{1}^T\mathbf{b}_n^{\alpha+\beta}}\right)^{\frac{1}{\alpha}}, & \text{for } \alpha\neq 0\\ \exp\left(\frac{{\mathbf{b}_n}^{\beta T}\ln\left(\frac{\mathbf{x}}{\mathbf{b}_n}\right)}{\mathbf{1}^T\mathbf{b}_n^{\beta}}\right), & \text{for } \alpha=0\end{cases} \tag{10}$$

where $\mathbf{1}$ represents a vector of ones of length F. Once the optimal weights have been obtained, the active set is initialised with the atom and its corresponding weight that provides the minimum cost in terms of Alpha-Beta divergence $D_{AB}\left(\mathbf{x}\|w_n\mathbf{b}_n\right)$:

$$\mathcal{A}\leftarrow\arg\min_{n} D_{AB}\left(\mathbf{x}\|w_n\mathbf{b}_n\right). \tag{11}$$

Step 2. Updating the active set

The active set is iteratively updated so that a new atom is added every $K=2$ iterations. Let $AB(\mathbf{w})$ be the cost function defined as the Alpha-Beta divergence between the observation vector $\mathbf{x}$ and the model $\hat{\mathbf{x}}=\mathbf{B}\mathbf{w}$:

$$AB(\mathbf{w})=D_{AB}\left(\mathbf{x}\|\mathbf{B}\mathbf{w}\right). \tag{12}$$

The atom added to the active set is the atom with the largest negative value of the partial derivative of the cost function $AB(\mathbf{w})$ with respect to the weight w_n:

$$\mathcal{A}\leftarrow\mathcal{A}\cup\left\{\arg\min_{n\notin\mathcal{A}}\frac{\partial}{\partial w_n}AB(\mathbf{w})\right\} \tag{13}$$

This partial derivative can be expressed in matrix form as:

$$\frac{\partial}{\partial w_n}AB(\mathbf{w})=\begin{cases}\frac{1}{\alpha}\mathbf{b}_n^T\left(\hat{\mathbf{x}}^{\alpha+\beta-1}\circ\left(\mathbf{1}-\left(\frac{\mathbf{x}}{\hat{\mathbf{x}}}\right)^{\alpha}\right)\right), & \text{for } \alpha\neq 0\\ \mathbf{b}_n^T\left(\hat{\mathbf{x}}^{\beta-1}\circ\ln\left(\frac{\hat{\mathbf{x}}}{\mathbf{x}}\right)\right), & \text{for } \alpha=0\end{cases} \tag{14}$$

where $\circ$ denotes the Hadamard product and the vector division is done point by point. The weight assigned to the atom to be added to the active set is initialised to a small positive value, in particular 10^{-15} . If all partial derivatives calculated following (14) are positive, no atom is added to the active set.

Step 3. Updating weights

Let $\mathbf{B}_A$ be the dictionary composed of the bases that are in the active set $\mathcal{A}$ and $\mathbf{w}_A$ be the weights of these bases. The observation vector model can then be expressed as $\hat{\mathbf{x}}=\mathbf{B}_A$. The update of the weights of the atoms in the active set is done following Newton's method:

$$\mathbf{w}_A\leftarrow\mathbf{w}_A-\mu{\mathbf{H}_{\mathbf{w}_A}}^{-1}\nabla_{\mathbf{w}_A} \tag{15}$$

where μ is the step size, $\mathbf{x}$ is the gradient of the cost function $D_{AB}(\mathbf{x}\|\mathbf{B}_A\mathbf{w}_A)$ with respect to the vector of weights of the

active set and $\mathbf{H}_{\mathbf{w}_A}$ is the Hessian matrix. The resulting matrix expressions for the gradient and the Hessian are as follows:

$$\nabla_{\mathbf{w}_\mathcal{A}} = \begin{cases} \frac{1}{\alpha}\mathbf{B}_\mathcal{A}^T\left(\hat{\mathbf{x}}^{\alpha+\beta-1} \circ \left(\mathbf{1} - \left(\frac{\mathbf{x}}{\hat{\mathbf{x}}}\right)^\alpha\right)\right), \\ \qquad\qquad \text{for} \alpha \neq 0 \\ \mathbf{B}_\mathcal{A}^T\left(\hat{\mathbf{x}}^{\beta-1} \circ \ln\left(\frac{\hat{\mathbf{x}}}{\mathbf{x}}\right)\right), \\ \qquad\qquad \text{for } \alpha = 0 \end{cases} \tag{16}$$

and

$$\mathbf{H}_{\mathbf{w}_A} = \mathbf{B}_A^T \text{diag}\left(\mathbf{v}\right)\mathbf{B}_A \tag{17}$$

where $\mathbf{v}$ is a vector defined as:

$$\mathbf{v} = \begin{cases} \hat{\mathbf{x}}^{\alpha+\beta-2} \circ \left(\left(\frac{\alpha+\beta-1}{\alpha}\right)\mathbf{1} + \left(\frac{1-\beta}{\alpha}\right)\left(\frac{\mathbf{x}}{\hat{\mathbf{x}}}\right)^\alpha\right), \\ \qquad\qquad \text{for } \alpha \neq 0 \\ \hat{\mathbf{x}}^{\beta-2} \circ \left(\mathbf{1} + (1-\beta)\ln\left(\frac{\mathbf{x}}{\hat{\mathbf{x}}}\right)\right), \\ \qquad\qquad \text{for } \alpha = 0 \end{cases} \tag{18}$$

To ensure numerical stability in the inversion of the Hessian matrix in (15), it is necessary to add to the Hessian matrix an identity matrix multiplied by a small positive constant, 10^{-10} in our implementation, before performing the inversion. Finally, given the vector $\mathbf{r}$ defined as:

$$\mathbf{r} = \frac{\mathbf{w}_A}{\mathbf{H}_{\mathbf{w}_A}^{-1}\nabla_{\mathbf{w}_A}} \tag{19}$$

the step size μ that guarantees that the resulting weights are non-negative is given by:

$$\mu = \min(\min_{r_i>0} r_i, 1) \tag{20}$$

If, as a result of the above update, the weight obtained is zero, the corresponding atom is removed from the active set.

Step 4. Finalisation

The algorithm iteratively repeats steps 2 and 3 until all derivatives in (14) take non-negative values, although in practice a maximum number of iterations is fixed. As in the ASNA algorithm, two Newton updates (step 3) are performed before each update of the active set (step 2).

IV. Simmulations

The ASNA-AB algorithm proposed in this article has been implemented in MATLAB using a columnar scheme following the recommendations in [13]. The resulting implementation can work simultaneously with multiple observations with $\mathbf{X} \approx \hat{\mathbf{X}} = \mathbf{WB}$ subject to $\mathbf{W}, \geq 0$. With this implementation, the O observations are in the rows of $\mathbf{X} \in \mathbb{R}_+^{O\times F}$, and the non-negative weights corresponding to each observation are found in the rows of $\mathbf{W} \in \mathbb{R}_+^{O\times NA}$, where NA is the total number of atoms in the dictionary matrix $\mathbf{B}$.

Simulations in the field of audio signal separation have been performed using the same methodology as in [5], but using the Bach10 music database [14]. The Bach10 database

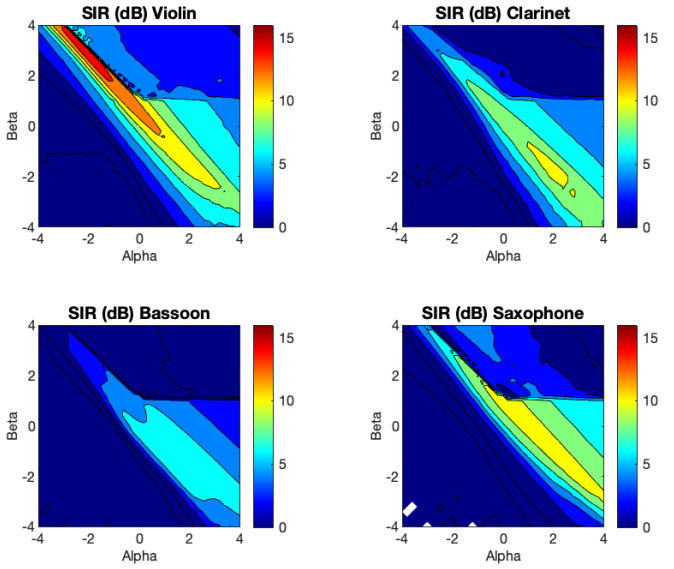

Fig. 1. Performance of the ASNA-AB algorithm for the separation of four instruments in the $\alpha\beta$-plane in terms of SIR (dB).

is composed of 10 polyphonic pieces of music (soprano, alto, tenor and bass) performed by the instruments violin, clarinet, saxophone and bassoon. The database provides the .wav files of the four instruments as well as the mix of them. The sampling frequency is 44100 Hz and the duration of the pieces varies between 25 and 42 seconds. The first 9 pieces have been used to obtain the dictionaries of each of the instruments and the last piece has been used to separate the four instruments from the mix using the proposed ASNA-AB algorithm.

To obtain the observation vector in the time-frequency domain, we have used the magnitude of the short Fourier transform (STFT) calculated with a Hanning-type windowing of 60 ms and a displacement size of 15 ms. The dictionaries of the four instruments, consisting of 1000 atoms each, have been obtained by grouping the different spectral components obtained by the STFT with the k-means algorithm. The separation results obtained for the four instruments in terms of SIR (dB) measured with the tool BSSEVAL [15] are presented in Fig. 1. Simulations have been performed for values of α and β within the ranges $-4 \leq \alpha \leq 4$ and $-4 \leq \beta \leq 4$ with step size equal to 0.2.

As it can be observed, the best results for all instruments are reached in a region delimited by the lines $\alpha + \beta = 1$ and $\alpha = -\beta$ which correspond to the family of Alpha-divergences and a Generalized Itakura-Saito (Gen-IS) distance with an α-zoom of its arguments respectively:

$$D_{AB}^{(\alpha,1-\alpha)}(\mathbf{P}\|\mathbf{Q}) = D_A^{(\alpha)}(\mathbf{P}\|\mathbf{Q}) \tag{21}$$

$$D_{AB}^{(\alpha,-\alpha)}(\mathbf{P}\|\mathbf{Q}) = \frac{1}{\alpha^2}D_{IS}\left(\mathbf{P}^{\cdot[\alpha]}\|\mathbf{Q}^{\cdot[\alpha]}\right) \tag{22}$$

where $\mathbf{P}^{\cdot[\alpha]}$ denotes the one to one transformation that raises each element of the vector $\mathbf{P}$ to the power α.

TABLE I
PERFORMANCE OF THE ASNA-AB ALGORITHM FOR SOME SPECIFIC DISTANCES AND DIVERGENCES IN TERMS OF SIR (DB)

Alpha-Beta Parametrization	**SIR (dB)**			
	Violin	***Clarinet***	***Bassoon***	***Saxophone***
Kullback-Leibler (1,0)	10.86	7.88	6.63	11.02
Euclidean (1,1)	7.94	6.22	5.07	9.54
Itakura-Saito (1,-1)	9.93	9.66	7.53	7.53
Alpha-div. (0,1)	11.76	7.38	5.81	10.86
Alpha-div. (2,-1)	10.20	8.23	7.21	10.88
Alpha-div. (-1,2)	13.21	6.75	4.73	10.13
Alpha-div. (-2,3)	15.44	5.51	3.26	8.34
Beta-div. (1,-0.4)	12.06	9.52	6.59	10.13
Gen-IS div. (2,-2)	9.71	10.34	7.39	6.91

Nevertheless, not all instruments achieve their best result with similar parameterisations of the algorithm. The violin performs best for negative values of alpha, while the other instruments perform best for positive values of alpha. This suggests a dependence of the optimal parameters α and β with the timbres of the instruments (which are related to their spectral envelopes) and possibly with their ranges (related to the played notes). Additionally, we observe that for $\alpha+\beta < 0$, the performance of the algorithm deteriorates drastically probably due to the inversion of the arguments of the AB-divergence. This behaviour is also observed in other applications of AB-divergences [9], [12].

Additionally, one can observe that the best performing regions of the algorithm for the violin, bass and saxophone are inside the convex cone in which the convexity of the AB-divergence is guaranteed with respect to the second argument, while for the clarinet the best region is outside the convex cone. As mentioned in Sec. II, the region of convexity is not strictly limited to the convex cone, and therefore it is always interesting to explore what happens outside the convex cone since the algorithm can still converge in nearby regions.

Table I summarises the results for some specific divergences and distances. While Kullback-Leibler divergence $(\alpha, \beta) = (1, 0)$, exploited in [5], can be a good choice, other alternatives can improve the results. The pair $(\alpha, \beta) = (-2, 3)$ could be desirable when we are exclusively interested in the extraction of the violin track, improving the results of Kullback-Leibler divergence by more than 4.5 dB. For the separation of the four instruments the pair $(\alpha, \beta) = (2, -1)$ as well as the pair $(\alpha, \beta) = (1, -0.4)$ can be preferable to Kullback-Leibler divergence. The first one, $(\alpha, \beta) = (2, -1)$, improves the results for the "worst" instruments (the bassoon and the clarinet) while the results for the "best" instruments do not noticeably worsen, providing a global improvement of 0.13 dB. On the other hand, the second pair $(\alpha, \beta) = (1, -0.4)$ provides an apreciable improvement of the SIR for the violin and clarinet (1.2 dB and 1.64 dB respectively), while the behaviour for saxophone and bassoon do not noticeably worsen.

V. CONCLUSSIONS

A generalization of the Active-Set Newton Algorithm has been proposed, exploiting the generalized Alpha-Beta divergence between the observed signal and the decomposition, instead of the Kullback-Leibler divergence. The proposed algorithm decomposes the non-negative observed signal into a non-negative linear combination of the contributions on non-negative pre-specified atoms. This compositional model is exploited in the time-frequency domain for the characterisation of audio signals. A set of numerical experiments have been performed in order to evaluate the use of the proposed method for the separation of polyphonic music. These experiments have shown that there are some pairs of the alpha-beta parameters that improve the results of the separations with respect to the Kullback-Leibler divergence.

REFERENCES

[1] 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.

[2] A. Ozerov and C. Fevotte, "Multichannel Nonnegative Matrix Factorization in Convolutive Mixtures for Audio Source Separation, "*IEEE Trans. on Audio, Speech, and Language Processing*, vol. 18, no. 3, pp. 550-563, March 2010.

[3] P. Smaragdis, "Convolutive Speech Bases and Their Application to Supervised Speech Separation," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 15, no. 1, pp. 1-12, Jan. 2007.

[4] A. Sarmiento, I. Durán-Díaz, I. Fondón and S. Cruces, "Descomposición de señales de audio mediante un algoritmo de conjunto activo con $\alpha\beta$-Divergencias, " *XXXIV Simposium Nacional de la Unión Científica Internacional de Radio (URSI'19)*, Sevilla, Sep. 2019, Art ID. pp. 1-4

[5] T. Virtanen, J. F. Gemmeke and B. Raj, "Active-Set Newton Algorithm for Overcomplete Non-Negative Representations of Audio," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 21, no. 11, pp. 2277-2289, Nov. 2013.

[6] T. Virtanen, B. Raj, J. F. Gemmeke and H. Van-hamme, "Active-set newton algorithm for non-negative sparse coding of audio,"*2014 IEEE Int. Conf.on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, 2014, pp. 3092-3096.

[7] A. Sarmiento, I. Durán-Díaz, A. Cichocki and S. Cruces, "A Contrast Function Based on Generalized Divergences for Solving the Permutation Problem in Convolved Speech Mixtures, " *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, vol. 23, no. 11, pp. 1713-1726, Nov. 2015.

[8] E. Yilmaz, J. F. Gemmeke and H. Van hamme, "Noise-robust speech recognition with exemplar-based sparse representations using Alpha-Beta divergence, " *2014 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, 2014, pp. 5502-5506.

[9] A. Sarmiento, I. Fondón, I. Durán-Díaz and S. Cruces, "Centroid-Based Clustering with $\alpha\beta$-Divergences, " *Entropy*, vol. 21, no. 2, 196, Feb. 2019. doi:10.3390/e21020196.

[10] K. O'Hanlon and M. D. Plumbley, "Automatic Music Transcription using row weighted decompositions, " *2013 IEEE Int. Conf.on Acoustics, Speech and Signal Processing*, Vancouver, BC, 2013, pp. 16-20.

[11] C. O'Brien and M. D. Plumbley, "Automatic music transcription using low rank non-negative matrix decomposition, " *25th European Signal Processing Conf. (EUSIPCO)*, Kos, 2017, pp. 1848-1852.

[12] A. Cichocki, S. Cruces and S.I. Amari, "Generalized Alpha-Beta Divergences and Their Application to Robust Nonnegative Matrix Factorization, " *Entropy*, vol. 13, pp. 134-170, 2011.

[13] P. Sebastián, T. Virtanen, V. García-Molla and A. Vidal, "Analysis of an efficient parallel implementation of active-set Newton algorithm " *The Journal of Supercomputing*, vol. 75,pp. 1298-1309, 2018. doi:10.1007/s11227-018-2423-5

[14] Z. Duan and B. Pardo, "Soundprism: an online system for score-informed source separation of music audio, " *IEEE Journal of Selected Topics in Signal Process.*, vol. 5, no. 6, pp. 1205-1215, 2011.

[15] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 4, pp. 1462–1469, Jul. 2006.