MultiCraft

International Journal of Engineering, Science and Technology
Vol. 3, No. 6, 2011, pp. 273-282

INTERNATIONAL JOURNAL OF ENGINEERING, SCIENCE AND TECHNOLOGY

www.ijest-ng.com

Performance analysis of alpha divergence in nonnegative matrix factorization of monaural musical sounds

Bhuwan Mehta*, Nishchal K. Verma and Pradip Sircar

Department of Electrical Engineering, Indian Institute of Technology Kanpur, INDIA
**Corresponding Author: e-mail: bhuwan@iitk.ac.in*

Abstract

Estimation of original instrument sound signals from complex music signals without any prior information is one of the most challenging problems under the framework of Blind Source Separation (BSS). Due to their effectiveness in other applications of BSS, Nonnegative Matrix Factorization (NMF) based methods have particularly gained attention in the context of musical sound source separation. These techniques are based on decomposing the magnitude or power spectrum of an input signal into a sum of components with time varying gains. This is achieved by using a suitable cost function to determine the optimal factorization. Most work in this field has focused on the use of Euclidean and Kullback-Liebler (KL) divergence. This study looks into the use of α-divergence based non negative factorization in the context of single channel musical sound separation. Simulation experiments were carried on single channel mixtures of randomly mixed pitched musical instrument samples to determine optimal α values for this problem. The paper also looks into the performance of the algorithm as important system parameters are varied.

Keywords: Nonnegative Matrix Factorization, α-divergence, musical sound source separation, Blind Source Separation

DOI: http://dx.doi.org/10.4314/ijest.v3i6.22

1. Introduction

Blind Source Separation represents a larger framework of the class of unsupervised learning algorithms used in estimation of source signals from a mixture signal with no or very little information. Lately, the technique of representation of an observation signal into non negative factors or Nonnegative Matrix Factorization (NMF) is emerging as a popular technique in blind source separation. Besides BSS, NMF has been increasingly used in other related areas such as data mining, pattern recognition, object detection and dimensionality reduction. Paatero and Trapper (1997) first introduced it and since then many variants have been proposed by researchers. Its wide spectrum of applications includes face recognition (Li et al., 2001), medical imaging (Lee et al., 2006), polyphonic music transcription (Smaragdis and Brown, 2003), portfolio diversification (Drakakis et al., 2008), document clustering (Xu et al., 2003), and Scotch whiskies clustering (Young et al., 2006).

NMF factorizes a given non negative data matrix $\mathrm{X} = [\mathrm{X}_1, \mathrm{X}_2, \ldots \ldots, \mathrm{X}_\mathrm{T}] \in \Re^{\geq 0, F \times T}$ as a product of two non negative matrices $\mathrm{W} \in \Re^{\geq 0, F \times K}$ and $\mathrm{H} \in \Re^{\geq 0, K \times T}$ such as:

$$X \approx WH \tag{1}$$

where *W* contains the basis vectors in its columns and *H* is the associated variable gain matrix.

In literature, the factorization (1) is usually sought through a minimization problem

$$\min_{W,H \geq 0} D(X \mid WH) \tag{2}$$

where $D(X \mid WH)$ is a cost function or simply put, an error measure.

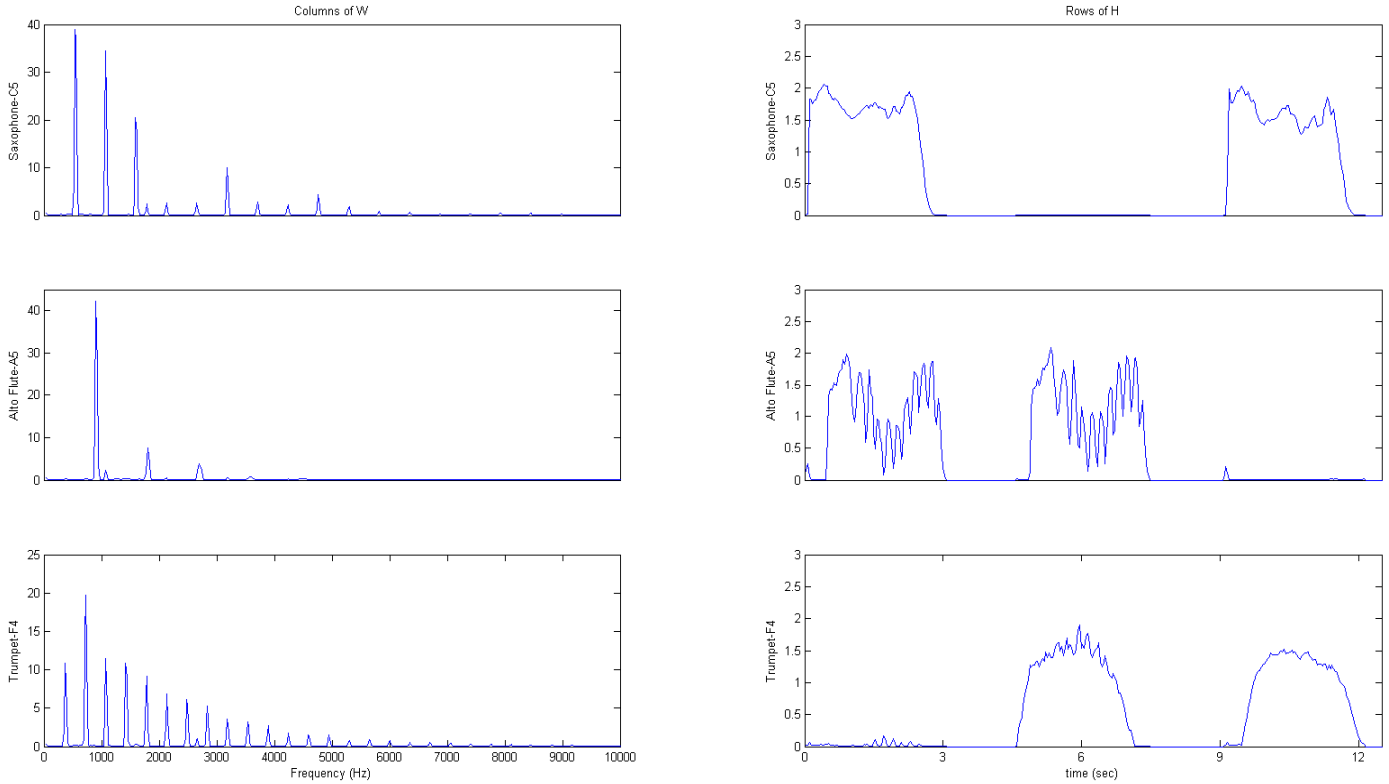

Figure 1. Columns of W and rows of H obtained by applying KL-NMF on a mixture signal consisting of notes: C5 of Saxophone, A5 of Alto Flute and F4 of Trumpet played in pairs one at a time.

Various error measures for the minimization problem (2) have been proposed such as Euclidean distance measure, Kullback Liebler (KL) divergence, Csiszár's divergences (Cichocki et al., 2006), β divergence (Dhillon and Sra, 2005), with several other cost functions considered in Cichocki et al. (2006). Popular choices for the cost functions are Euclidean distance measure that we here define as:

$$D_{euc}(X \mid WH) = \| X - WH \|^2 \tag{3}$$

and the generalized KL divergence which is defined as

$$D_{KL}(X \mid WH) = \sum_{i,j} X_{i,j} \log \frac{X_{i.j}}{[WH]_{i,j}} - X_{i,j} + [WH]_{i,j} \tag{4}$$

It was Lee and Seung (1999) who proposed the multiplicative update rules for these two measures, under which the cost function has been shown to be non-increasing in subsequent iterations. The simplicity of these update rules has greatly contributed to the popularity of these measures. Multiplicative update rules for these two NMF algorithms are as follows:

1) Euclidean Update: A local minimum of the cost function (3) is reached by an update algorithm that is of form:

$$W_{i,j} \leftarrow W_{i,j} \frac{[XH^T]_{i,j}}{[WHH^T]_{i,j}}$$

$$H_{i,j} \leftarrow H_{i,j} \frac{[W^T X]_{i,j}}{[W^T WH]_{i,j}} \tag{5}$$

2) KL-divergence Update: A local minimum of the error measure (4) is found by the following update algorithm:

$$W_{i,j} \leftarrow W_{i,j}\left[\frac{\sum_k [H_{j,k}(\frac{X_{i,k}}{[WH]_{i,k}})]}{\sum_l H_{j,l}}\right]$$
$$H_{i,j} \leftarrow H_{i,j}\left[\frac{\sum_k [W_{k,i}(\frac{X_{k,j}}{[WH]_{k,j}})]}{\sum_l W_{l,i}}\right] \quad (6)$$

They have been found to work reliably for sound source separation and majority of the work has focused on these cost functions. The choice of the cost function is dictated by the type of the data to analyze. However, little literature is available exploring performance improvement by using other cost functions on a certain data. In this paper, we use a more general parameterisable divergence, known as α-divergence and study its performance in the context of musical sound source separation. The divergence was proposed in Havrda and Charvat (1967) and contains various well known divergence measures for various values of α. The outline of the paper is as follows: Section 2 describes the Amari's Alpha divergence, its related properties and update algorithms. The application of NMF in musical sound source separation is explained in Section 3, and simulation experiments and results are discussed in Section 4.

2. Amari's Alpha Divergence

The α-divergence is a parametric family of divergence functional. It includes several well known divergences as its special cases. The objective function of α-NMF which is based on the divergence between *X* and *WH* can be defined as

$$D_\alpha(X \mid WH) = \frac{1}{\alpha(1-\alpha)}\sum_{i,j} X_{i,j}^{\alpha}[WH]_{i,j}^{1-\alpha} - \alpha X_{i,j} + (\alpha - 1)[WH]_{i,j} \quad (7)$$

where $\alpha \in (-\infty, \infty)$. As in KL-divergence and Euclidean distance, α-divergence is zero if $X=WH$. It is a convex function. Alpha divergence can also be expressed as

$$D_\beta(X \mid WH) = \frac{4}{(1-\beta^2)}\sum_{i,j} X_{i,j}^{\frac{1-\beta}{2}}[WH]_{i,j}^{\frac{1+\beta}{2}} - \frac{1-\beta}{2}X_{i,j} - \frac{1+\beta}{2}[WH]_{i,j} \quad (8)$$

which is arrived at by setting α = (1 - β)/2 and 1 - α = (1 + β)/2 in (7). From (8), one can notice one of the important properties of duality of the α-divergence i.e.

$$D_{-\beta}(X \mid WH) = D_\beta(WH \mid X) \quad (9)$$

KL-divergence, Hellinger's divergence, chi-squared divergence corresponding to α=1, 0.5, 2 respectively are some of the special cases of the α-divergence. The α-divergence belongs to a family of convex divergence measure which is known as Csiszár's f-divergence (Cichocki et al., 2006, 2008).

The multiplicative update rules for α-divergence as derived in Cichocki et al. (2008) and are as follows:

$$W_{i,j} \leftarrow W_{i,j}\left[\frac{\sum_k [H_{j,k}(\frac{X_{i,k}}{[WH]_{i,k}})^{\alpha}]}{\sum_l H_{j,l}}\right]^{\frac{1}{\alpha}}$$
$$H_{i,j} \leftarrow H_{i,j}\left[\frac{\sum_k [W_{k,i}(\frac{X_{k,j}}{[WH]_{k,j}})^{\alpha}]}{\sum_l W_{l,i}}\right]^{\frac{1}{\alpha}} \quad (10)$$

The update rule for KL-divergence as discussed in (6) can be arrived at by putting α=1 in (10), which verifies KL-divergence as being a special case of α-divergence for α=1.

3. Application of NMF in Musical Source Separation

Sound source separation from music signals without any prior information about the original sources is one of the most challenging problems in the field of blind source estimation. The basic nature of most of the music styles results in significant overlap in both time and frequency domains, thus making separation harder. Moreover, absence of spatial information in monaural sounds further increases the complexity of the problem. Nonnegative Factorization (NMF) techniques are

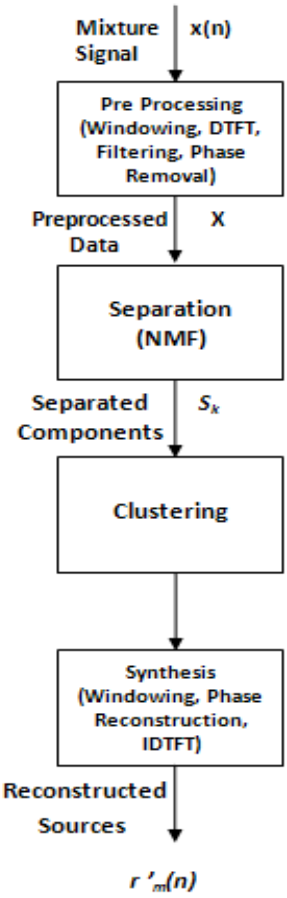

Figure 2. Pictorial Scheme of NMF based source separation

increasingly being applied in musical sound source separation due to their success in other applications of blind source separation. Smaragdis and Brown (2003) introduced the use of NMF for musical separation tasks. Figure 2 shows the basic pictorial scheme of the process of musical sound source separation, where $\boldsymbol{r'_m(n)}$ is defined as the reconstructed source signals. In this paper, since we are only studying the performance of the NMF algorithm in musical sound factorization, we have not dealt with the clustering and synthesis part in this paper. We will now discuss the application of NMF for source separation in detail in the remaining part of the section.

A) Nonnegative Matrix Factorization: The NMF algorithm is applied in musical source separation based on a signal model where the magnitude or power spectrum $\boldsymbol{X_n}$ of an input mixture signal in frame $\boldsymbol{n}$ containing $\boldsymbol{F}$ frequency points**,** is modeled as a linear combination of time independent basis functions $\boldsymbol{W_k}$ which can be written as

Table 1. Generation of Test Signals

Parameter	Interval
Number of pitched musical instruments	[5,9]
Length of each note (sec.)	[0.7, 1.4]
Onset time for each note (sec.)	[0,3]

$$X_n = \sum_{k=1}^{K} H_{k,n} W_k \tag{11}$$

where $\boldsymbol{K}$ is the number of basis functions, and $\boldsymbol{H_{k,n}}$ is a scalar value representing time varying gain of the $\boldsymbol{k^{th}}$ basis function at time frame $\boldsymbol{n}$. The basis function $\boldsymbol{W_k}$ along with its time varying gain $\boldsymbol{H_{k,n}}$ represents a single component *k* of the input signal. Each source can be modelled as a sum of one or more components. A component can be interpreted as a meaningful entity or parts of a musical sound source signal. This parts based representation is facilitated by a rather static spectral structure of individual musical notes over time, as compared to speech signals, thus accounting for fixed basis function $\boldsymbol{W_k}$ over time with time varying gains $\boldsymbol{H_{k,n}}$.

The model in (11) can be represented in matrix format as

$$X = WH \tag{12}$$

where $\mathrm{X} = [\mathrm{X}_1, \mathrm{X}_2, \ldots \ldots, \mathrm{X}_\mathrm{T}] \in \Re^{\geq 0, F \times T}$ is the input signal magnitude or power spectrogram in $\boldsymbol{T}$ time frames, $\mathrm{W} = [\mathrm{W}_1, \mathrm{W}_2, \ldots \ldots, \mathrm{W}_\mathrm{K}] \in \Re^{\geq 0, F \times K}$ is the matrix containing the spectral structure of $\boldsymbol{K}$ bases and $\mathrm{H} \in \Re^{\geq 0, K \times T}$ is the time varying gain matrix for $\boldsymbol{K}$ bases at $\boldsymbol{T}$ time frames. In this paper, we have used magnitude spectrogram as it performs better than the power spectrogram as noted in Virtanen (2007). This may be due to the fact that power spectrogram suppresses the lower intensity spectra as compared to the higher intensity ones.

The decomposition $\boldsymbol{X=WH}$ can be carried on by using the minimization problem (2) on various error measures such as Euclidean distance measure (3), KL-divergence (4), α-divergence (7) and others. The estimation of W and H is done as shown in the following steps:

1) Initialize the entries in $\boldsymbol{W}$ and $\boldsymbol{H}$ to random non-negative values.
2) Update W using α-divergence update (10), KL-update (6) or Euclidean update (5).
3) Update H using α-divergence update (10), KL-update (6) or Euclidean update (5).
4) Iterate until the value of cost function converges.

One of the key issues with NMF is the estimation of the number of components or $\boldsymbol{K}$. Selecting $\boldsymbol{K}$ can be a complex procedure which requires estimating the dimensionality of the input matrix. Typically, $\boldsymbol{K}$ is chosen such that it is larger than the estimated number of sources, and follows the constraint $\boldsymbol{(F+T)K < < FT}$.

Hence, we have now built upon a framework for using NMF on audio signals. An important thing to note here is that NMF does not allow for the usage of phase spectra in the estimation phase. During the synthesis of the extracted components, as explained later, phases are extracted from the original mixture signal. The perceptual quality of reconstructed signals is not greatly affected in this case as the human ear is less sensitive to phase distortions (Goldstein, 1967).

Figure 1 presents an example of implementation of NMF on mixture signals. The mixture signal is composed of note A5 (880Hz) of Alto Flute, note F4 (350Hz) of Trumpet and note C5 (524Hz) of Saxophone played in pairs at different time intervals. NMF using KL-divergence was applied on the mixture signal and number of components was set to be 3. We noticed that the frequency of the lowest partial of the extracted components matches with the fundamental frequency of the constituent notes. Also, the gain information in the rows of H corresponds to the placement of particular notes in the mixture signal. Thus, we can see that the source signals have been successfully separated from the mixture X by decomposing it in the form of a product of matrices $\boldsymbol{W}$ and $\boldsymbol{H}$.

In this paper, we have used NMF for blind estimation of source signals. In this method, NMF learns from the mixture without any prior knowledge of the source signals and tries to capture the repeating spectral patterns like note spectra from the available mixture. However, the simple model in (11) and (12) has also been used in supervised learning framework, in which basis are learned prior to estimation (Paulus and Virtanen, 2005; Wilson et al., 2008). Also, many other extensions to the model in (11) and (12) have been proposed (Blumensath and Davies, 2006; Virtanen et al., 2008) to improve the performance. We will be using the simple unsupervised NMF model in this paper as the aim is to evaluate the performance of α-divergence.

Table 2. Performance of Euclidean Distance and KL-Divergence as window length is varied

Window Length	Euclidean Distance			KL-Divergence		
	Detection Fraction	SDR	Multi-SDR	Detection Fraction	SDR	Multi-SDR
128	0.5782	2.7562	4.9021	0.6375	3.2076	5.4646
256	0.6189	3.6782	6.5741	0.7233	4.4549	7.3018
512	0.655	4.8155	8.3911	0.7674	6.0052	9.385
1024	0.6669	5.5863	9.5076	0.7973	7.0252	10.6164
2048	0.6776	5.8595	9.6733	0.799	7.0497	10.6655
4096	0.6533	5.3716	9.1803	0.7578	6.0557	9.4212
8192	0.6364	4.5866	8.173	0.7171	4.9809	8.1638

B) Separation and Quality Measures: The NMF separates the mixture spectrogram $\boldsymbol{X}$ into $\boldsymbol{K}$ components. The $\boldsymbol{k^{th}}$ column of W can be multiplied with the $\boldsymbol{k^{th}}$ row of H to obtain the component $\boldsymbol{S_k}$. However, clustering of the components into sources is a difficult issue in unsupervised domain. Hence, we have to use source signals before mixing as reference to cluster the components.

Signal to Distortion Ratio (SDR) between each component and source is used as a measure to cluster the components. The SDR between the magnitude spectrograms of $\boldsymbol{m^{th}}$ reference signal $\boldsymbol{R_m}$ and $\boldsymbol{k^{th}}$ component $\boldsymbol{S_k}$ can be defined as:

$$SDR_{m,k} = \frac{\sum_{f,t}[R_m]_{f,t}^2}{\sum_{f,t}[[R_m]_{f,t} - [S_k]_{f,t}]^2} \tag{13}$$

Each component $\boldsymbol{k}$ is assigned to a reference source $\boldsymbol{m}$ corresponding to which it produces the highest SDR. Multiple components might be mapped to a reference source under this scheme. However, it might lead to non salient components like noise and transients mapped to a single source. To overcome this, we will only select the component with the highest SDR among all the components mapped to a source for that particular reference source.

The quality evaluation of the separated components is a tricky issue. Ultimately, human perception of the separated sounds is the best measure of the quality of separated sounds. However, listening tests for each mixture signal is very time consuming and subjective. This is where arises the need for an object quality measure for studying the performance of the algorithm. In this paper, we have used the SDR between the magnitude spectrum of the separated components and corresponding reference sources as described in (13) for quality evaluation. The average SDR (in decibels) is finally reported as a performance measure. Often, source signals of weaker intensity can be overshadowed by signals of stronger intensity and could not be extracted. Such source signals that could not be associated with any component are listed as undetected. Increasing the number of extracted components can be helpful to extract the undetected source signal in such cases. The detection fraction defined as the ratio of the total number of detected sources and the total number of sources was also used as a performance measure. Also, SDR using multiple components mapped to a single reference source, known as multi-SDR, was also calculated to find an upper limit on the performance of such algorithms. These performance evaluation measures were used and described in Virtanen (2007). The components with maximum SDR less than 0 dB were rejected as non-salient components.

Example- Given a mixture signal made of three reference signals (R_1, R_2 and R_3). NMF is applied on the mixture signal resulting in 4 components (S_1, S_2, S_3 and S_4). The SDR values of components w.r.t reference signals is as given in Table 3.

As we can see in Table 3, components S_1, S_2 and S_3 produce largest SDR with reference source R_1, and S_4 with R_2. The components are accordingly mapped to the respective sources. However, since no component can be matched to R_3, it remains undetected. Among the three components mapped to R_1, S_1 has the largest SDR value, and so it is used in SDR evaluation of system. However, all the three components of R_1 will be used in the calculation of multi SDR. Hence, average SDR of system = 0.5(6.78+7.22) = 7, and detection fraction =2/3.*

Table 3. Sample SDR Values Between Extracted Components and Reference Sources

SDR	*R_1*	*R_2*	*R_3*
S_1	***6.78***	*1.21*	*2.15*
S_2	***4.11***	*0.38*	*3.18*
S_3	***1.54***	*0.63*	*0.28*
S_4	*0.26*	***7.22***	*0.85*

C) Clustering and Synthesis: Since we are only concerned with quality evaluation in this study, reconstruction of the signals is not important from our point of view. However, in practical applications and in case of listening tests, reconstruction by clustering and synthesis is a very important aspect. For reconstruction, first the extracted components need to be mapped or clustered to individual sound sources. Various clustering methods, both unsupervised and supervised are available in literature (Dubnov, 2002; Vembu and Baumann, 2005). However, the performance of clustering methods is still limited and need to be improved upon. Once clustered, the components are then synthesized to obtain signals in time domain. For this, first the magnitude/power spectra of all the components corresponding to a particular reference source are added. To get the complex spectra, the phases of the original mixture signal are used. This simple method stems from the sparseness of the audio spectra. Then, by taking an inverse Fourier transform of the complex spectra, one can obtain the synthesized separated signal in the time domain. To account for the discontinuities at frame boundaries, one can use overlap add method where each synthesized time frame is windowed before combining adjacent frames in the time domain. This synthesis method was found to produce good results in our informal listening tests.

4. Experiments and Results

We carried experiments on a large number of signals for reliable performance evaluation. Ideally, the experiments should be performed on real life musical samples for a realistic evaluation. However, for evaluating the quality of the extracted components, as explained earlier, reference signals before mixing are needed. Since, this scenario was not possible with commercially available music samples; we had to generate test signals for our experiments. These test signals were generated on the lines of Virtanen (2007) from a comprehensive database of real instrument recordings, which is available for research purposes at TUIMISD (2011). The instrument sounds in the database were recorded at a sampling frequency of 44.1 kHz and 16 bit/sample. A single note is played at a time, and notes from the complete range of an instrument were available. We took about 25 notes per instrument and a total of 18 pitched musical instruments were used for evaluation. Random number of pitched musical instruments was chosen from within the limits given in Table 1. A random instrument and a random note were then selected from the available samples. Each instrument sample was only used once in the mixture and was truncated to random length within the limits given in Table 1. Each note started randomly at a time interval between 0 and 3 s. The length of each mixture signal was set to 4 s. Majority of frames in the mixture signals produced were overlapping, thus making them ideal for evaluation purposes of the algorithm.

The window length for spectrogram representation of signals greatly affects the performance of the algorithm. The window length chosen determines whether there is a good frequency resolution or good time resolution. A wider window leads to better frequency quantization, which aids the separation of frequency components closely spaced together, but provides poorer time resolution leading to poorer magnitude/power gain change in time. The opposite is true for a narrower window which provides better time resolution but poorer frequency resolution. Table 2 illustrates the effect of window length on the performance of Euclidean Divergence and KL-Divergence functions when applied on randomly generated mixture signals as explained above. Increasing the window length improves the performance, and the best performance is attained for window length of 2048, after which it deteriorates. We can see that the window length of 2048 provides a frequency resolution of about 21.5 Hz and a time window of about 46 ms at a sampling frequency of 44.1 kHz is the most suitable for NMF factorization on this data. Thus, in our experiments ahead, a window length of 2048 samples was used with an overlap of 0.5.

As already presented in Mehta et al. (2011), the performance of α-NMF was evaluated for various values of α. The value of α was varied from -1 to 2 with a resolution of 0.2. The purpose was to study the performance of α-divergence across the whole range of α. The range of α was such chosen because it covered all the popular divergence measures. The experiment was carried on about 450 generated mixture samples. Also, performance was evaluated for Euclidean distance to provide a performance comparison. The number of sources was taken to be 12, a number always larger than the number of sources and an intuitive estimate based on factorization experiments in F´evotte et al. (2008), in which a comparative experimental study of Euclidean-NMF and KL-NMF applied to a short duration piano sequence has been reported.

Figure 3 shows the average SDR obtained for magnitude spectrogram for various values of α. It can be seen that the optimal values for α occur in the region from α= -0.4 to 0.2 with maximum SDR occurring at α = -0.2. SDR continually decreases the further we move away from this region. However, at the limit point α=1, rapid increase in the SDR is observed. This corresponds to KL-divergence. The SDR reported at α=1 is just below the maximum at α= -0.2, thus justifying its use as the commonly used cost function. It can also be observed that the α-divergence performs better than the Euclidean distance measure for which SDR value is 6.02, at all values of α. Similar patterns are observed for other performance measures. Thus, α-divergence is definitely more effective at sound source separation than the Euclidean distance measure.

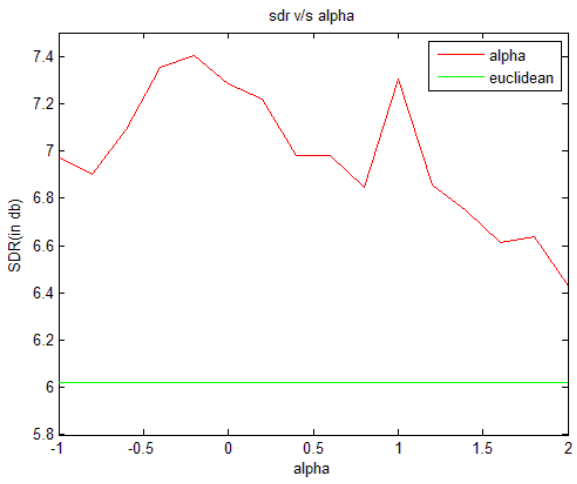

Figure 3. SDR v/s alpha

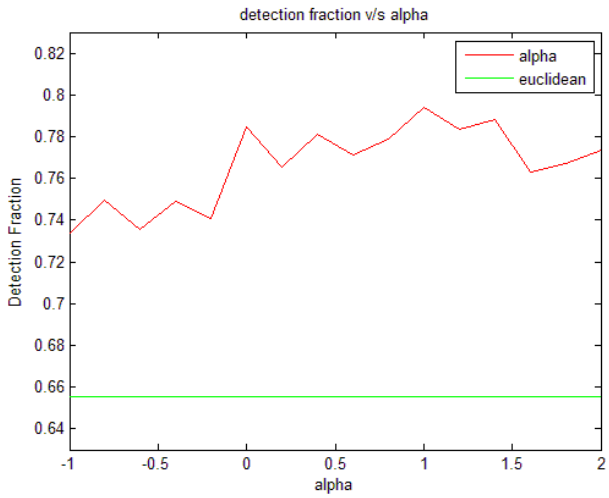

Figure 4. Detection Fraction v/s alpha

In Figure 4, we can see that Detection fraction is the maximum at α=1, i.e. KL-divergence. The general tendency is to decrease as we move away from α=1. The detection fraction for Euclidean Distance measure is 0.655, which is well below detection fraction values for all α. More or less, same pattern can be observed in the case of multi-SDR. In the case of multi-SDR, α=1 easily outperforms other α values as can be seen in Figure 5. The region α= -0.4 to α= 0.8 has almost comparable multi-SDR values, beyond which a rapid decrease in multi-SDR is observed barring the limit point maximum at α=1. In fact, it even decreases below the Euclidean distance multi-SDR value=10.1, for α beyond -0.5. The most suitable conclusion that one can make from these observations is that values of α in the region beyond α= -0.4 and α= 1 carry less significance in terms of musical sound source separation. The number of iterations measure in Figure 6 also seems to suggest the same. The number of iterations involved increase rapidly in this region thus suggesting that the convergence takes more time as magnitude of α increases. Another important observation is that KL-divergence i.e. α=1 is the only value that performs well with respect to all performance measures.

For the purpose of analysis, we can also dissect the region of interest from α= -0.4 to 0.9 into two regions, one with higher SDR and lower Detection Fraction (α= -0.4,-0.2,..), and the other with lower SDR and higher. In the former region, one can interpret that the undetected components mainly contain residual noise and transients, thus lowering the Detection Fraction. This is why despite of high SDR, multi-SDR obtained are not much higher because of insignificant components in terms of noise and transients. Likewise, in the latter region, one can interpret that the detected components corresponding to a note might contain portions of notes from other sources, resulting in artifacts in separations thus accounting for low SDR values and also multi-SDR values observed are not much higher.

Intuitively, it can be suggested that the detection factor can be improved by increasing the number of components. However, it might also affect the quality of components separated i.e. SDR. Thus, to gain a better understanding of component factorization, we tried to look into the performance of alpha divergence as we vary the number of components. For this, we conducted a different set

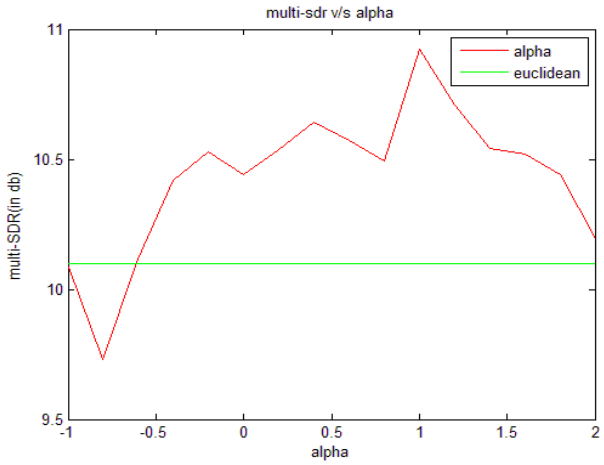

Figure 5. Multi-SDR v/s alpha

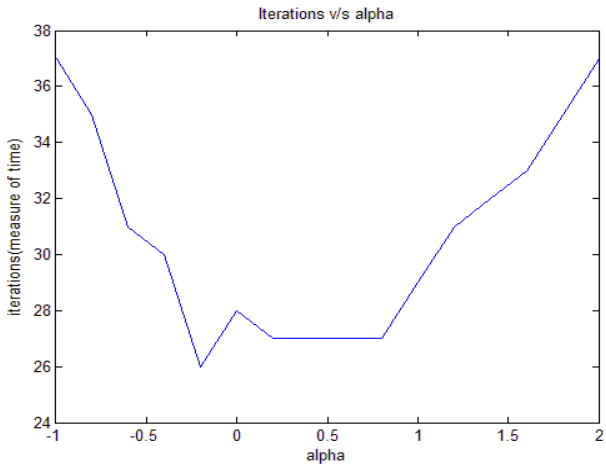

Figure 6. Iterations v/s alpha

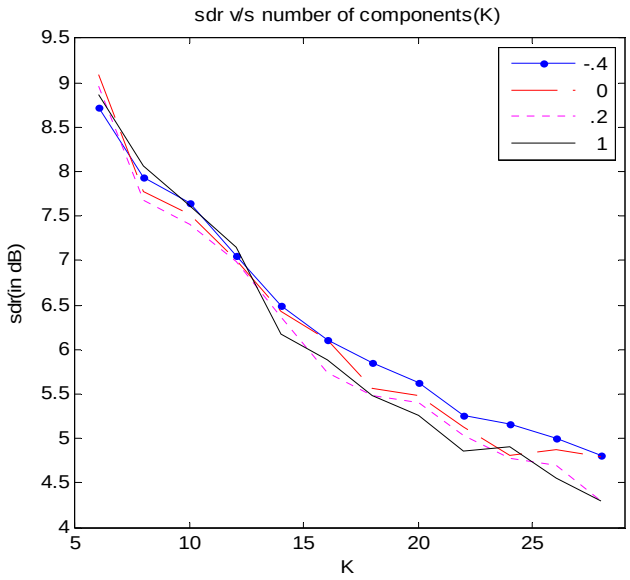

Figure 7 a) SDR(dB) v/s Number of components (K)

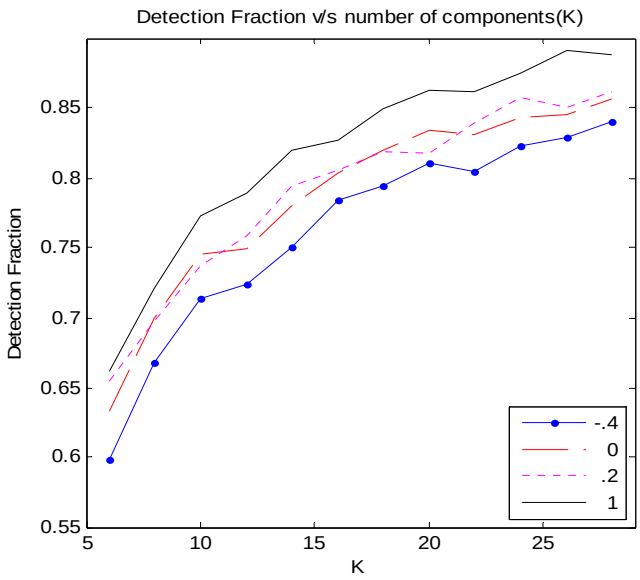

Figure 7 b) Detection Fraction v/s Number of components (K)

of experiments by carrying out the performance evaluation on certain values of α, as the number of components is varied. The number of components was varied from 6 to 28 in steps of 2. We used α= -0.4, 0, 0.2 and 1 for this evaluation, the idea being choosing the better performing values of α.

Figure 7 a) and Figure7 b), illustrate the performance of the above mentioned α values in terms of SDR and detection fraction as function of the number of components. The SDR decreases, whereas the detection error rate increases as the number of components increases. Interestingly, for K=6 and K=8, where the number of components was lesser than the number of sources, SDR values were higher and expectedly detection factor was low. This implies that NMF favours the factorization of dominant sources in the mixture signal. In our experiments, we also observed that multi-SDR values dropped a bit initially and did not vary much thereafter as the number of components increased. The decrease in SDR values can thus be primarily attributed to breaking of a single note into multiple components as the components increase. The increase in detection fraction is expected as the result of separation of more number of sources with the increase in number of components.

5. Conclusion

The use of α-divergence in the context of NMF based musical sound source separation was explored. The performance evaluation was done in an unsupervised framework where little prior information is available about the music signals. The performance was evaluated in terms of SDR and detection fraction. Also, multi-SDR values were measured to observe the upper limit on the performance of such algorithms. It was observed that factorization based on α-divergence works reasonably well in case of sound source separation. It outperformed the commonly used Euclidean distance measure for almost all values of α. However, it was observed that KL-divergence corresponding to α=1 performs best in terms of sound source separation, thus justifying its use as a popular measure for the same. The results suggest that there is no significant advantage of using other alpha values over using KL-divergence in a broader framework of musical sound factorization. Further, experiments were also carried on to observe the change in performance as system parameters such as window length and the number of components were varied.

However, the performance of the algorithm was found to be limited. The performance of the algorithm can be improved by incorporating various constraints such as temporal continuity, sparseness, harmonicity and others, typically associated with audio signals. The results also indicate research possibilities for further research on the use of alpha divergence, and their usability in some specific types of musical sounds can be studied in future research. Also, a proper mathematical study of α-divergence needs to be done. Noise modelling of the divergence can provide useful insights in terms of understanding the results and performance improvement.

References

Paatero P. and Tapper U., 1997. Least squares formulation of robust non- negative factor analysis, *Chemometr. Intell. Lab. Systems* Vol. 37, pp.23–35.

Li S.Z., Hou X.W., Zhang H.J. and Chang Q.S., 2001. Learning spatially localized parts based representation, *In Proc. IEEE International Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii*, pp. 207–212.

Lee H., Cichocki A., Choi S., 2006. Nonnegative matrix factorization for motor imagery EEG classification, *In Proc. International Conference on Artificial Neural Networks, Springer, Athens, Greece*.

Smaragdis P. and Brown J. C., 2003. Non-negative matrix factorization for polyphonic music transcription, *In IEEE Workshop on Signal Processing to Audio and Acoustics (WASPAA).*

Drakakis K., Rickard S., Frein R. de, and Cichocki A., 2008. Analysis of financial data using non-negative matrix factorization, *In International Mathematical Forum 3*, pp.1853-1870.

Xu W., Liu X. and Gong Y., 2003. Document clustering based on non-negative matrix factorization, *In Proc. of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, *Toronto, Canada*, pp. 267–273.

Young S. S., Fogel P., and Hawkins D., 2006. Clustering Scotch whiskies using non-negative matrix factorization, *Joint Newsletter for the Section on Physical and Engineering Sciences and the Quality and Productivity Section of the American Statistical Association*, Vol. 14, No. 1, pp. 11–13.

Cichocki A., Zdunek R., and Amari S., 2006. Csiszar's divergences for non- negative matrix factorization: Family of new algorithms, In *Proc. International Conf. Independent Component Analysis and Blind Signal Separation*, Charleston, South Carolina.

Dhillon I. and Sra S., 2005. Generalized nonnegative matrix approximations with Bregman divergences, *In Advances in Neural Information Processing Systems 17. Cambridge, MA: MIT Press.*

Cichocki A., Amari S., Zdunek R., Kompass R., Hori G., and He Z., 2006. Extended SMART algorithms for non-negative matrix factorization, *In Proc. International Conference on Artificial Intelligence and Soft Computing , Zakopane, Poland*, pp. 548–562.

Lee D.D. and Seung H.S., 1999."Learning the parts of objects with nonnegative matrix factorization," *In Nature*, Vol. 401, pp.788-791.

Havrda J. and Charvat F., 1967. Quantification method of classification process. Concept of structural α- entropy, Kybernetika, Vol. 3, pp. 30–35.

Cichocki A., Lee H., Kim Y.D., & Choi S., 2008. Nonnegative matrix factorization with alpha-divergence, In *Pattern Recognition Letters*, Vol. 29, pp.1433–1440.

Goldstein J.L., 1967. Audiotry spectral filtering and monaural phase perception, In Journal of Acoustic Society of America, Vol. 41, pp. 458-479.

Paulus J. and Virtanen T., 2005. Drum transcription with non-negative spectrogram factorization, *In Proc. of European Signal Processing Conference, Turkey.*

Wilson K. W., Raj B., Smaragdis P., and Divakaran A., 2008. Speech denoising using nonnegative matrix factorization with priors, *In Proc. of ICASSP,* Las Vegas, USA.

Blumensath T. and Davies M., 2006. Sparse and shift-invariant representations of music, *IEEE Trans. Audio, Speech, Lang. Process.*, Vol. 14, No. 1, pp. 50–57.

Virtanen T., Cemgil A. T., and Godsill S., 2008. Bayesian extensions to non-negative matrix factorisation for audio signal modelling*," In Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, Nev, USA*, pp. 1825–1828.

Virtanen T., 2007. Monaural sound source separation by non-negative matrix factorization with temporal continuity and sparseness criteria, *In IEEE Trans. Audio, Speech, Lang. Process.*, Vol. 15, No. 3, pp. 1066–1074.

Dubnov S., 2002. Extracting sound objects by independent subspace analysis," *In Proc. of 22nd International Audio Engineering Society Conference*, Espoo, Finland.

Vembu S. and Baumann S., 2005. Separation of vocals from polyphonic audio recordings, *In Proc. Of International Conference on Music Information Retrieval*, London, U.K., pp. 337-344

The University of Iowa Musical Instrument Samples Database (TUIMISD)[Online], 2011. Available for Research Purposes: http://theremin.music.uiowa.edu, accessed September 30.

Mehta B., Verma N., and Sircar P., 2011.Single channel musical source separation by nonnegative matrix factorization using alpha divergence, *In Proc. Of International Conference on Computer Application and Network Security(ICCANS-2011), Male, Maldives,* pp. 682-687.

F´evotte C., Bertin N., and Durrieu J.L., 2008. Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis, *In Neural Computation*.

Biographical notes

Bhuwan Mehta was born in India in June, 1988. He is currently pursuing his B.Tech-M.Tech Dual Degree from the Department of Electrical Engineering, IIT Kanpur. Currently, he is working on the problem of source separation in Music signals. His major research interests are Signal Processing, Machine Learning and Auditory Scene Analysis.

Nishchal K. Verma was born in India on September 9, 1973. He received B.E. from the Faculty of Engineering, Dayalbagh Educational Institute, Agra, India, in 1996, M.Tech. from the Indian Institute of Technology (IIT) Roorkee, India in 2003, and Ph.D. from IIT Delhi, New Delhi, India, in 2007, all in Electrical Engineering. From 1996 to 2000, he was with the Central India Machinery Manufacturing Company Birla Limited, Rajasthan, India, as an Engineer, where he was responsible for the maintenance and design of industrial equipments. During 2001–02, he was an Engineering Consultant to industries. He later joined as a Post Doctoral Research Associate in Dept. of Computer Science, Louisiana Tech University Ruston LA, USA for half of 2008 and then Post Doctoral Research Fellow in Center for Integrative and Translational Genomics, University of Tennessee Health Science Center, Memphis, TN, USA from Sep'08-Mar'09, where he later acted as a visiting professor for summer of 2010. Since Mar'09, he has been an Assistant Professor with Dept. of Electrical Engineering, IIT Kanpur and has more than 50 publications to his credit. His research interests include fuzzy systems, neural networks, data mining, fault diagnosis, bioinformatics, video clip or image sequence modeling, machine learning and computational intelligence. Dr. Verma is a reviewer for several reputed national and international journals and conferences, including the IEEE Transactions on Fuzzy systems, SMC A,B and C, Pattern analysis and Machine Intelligence and Pattern Recognition

Pradip Sircar received the B.Sc. degree in Physics from Calcutta University in 1974, the B.Tech. degree in Instrumentation and Electronics Engineering from Jadavpur University, Calcutta in 1978, and the M.S. and Ph.D. degrees both in Electrical Engineering from Syracuse Univ., NY, in 1983 and 1987, respectively. He was appointed an Assistant Professor in the Department of Electrical and Computer Engineering of Syracuse University in 1987. He joined the Department of Electrical Engineering, Indian Institute of Technology Kanpur in 1988, where presently he is a Professor. He was a Visiting Professor at Ecole Nationale Superieure des Telecommunications, Paris for one year(1998–1999).. His research interests are in the areas of signal processing, computations and communications. He is a Fellow of the Institution of Electronics and Telecommunication Engineers, a Senior Member of the IEEE, and a Member of the European Association for Signal and Image Processing. He is an Associate Editor of the Journal of The Franklin Institute.

Received September 2011
Accepted December 2011
Final acceptance in revised form February 2012