SHIFTED NMF WITH GROUP SPARSITY FOR CLUSTERING NMF BASIS FUNCTIONS

Rajesh Jaiswal,

Audio Research Group
Department of Electrical Engineering
Dublin Institute of Technology
Dublin, Ireland
rajesh.enc@gmail.com

Derry Fitzgerald,

Audio Research Group
Department of Electrical Engineering
Dublin Institute of Technology
Dublin, Ireland
derry.fitzgerald@dit.ie

Eugene Coyle,

Audio Research Group
Department of Electrical Engineering
Dublin Institute of Technology
Dublin, Ireland

Scott Rickard,

Department of Electronic Engineering
University College Dublin
Dublin, Ireland

ABSTRACT

Recently, Non-negative Matrix Factorisation (NMF) has found application in separation of individual sound sources. NMF decomposes the spectrogram of an audio mixture into an additive parts based representation where the parts typically correspond to individual notes or chords. However, there is a need to cluster the NMF basis functions to their sources. Although, many attempts have been made to improve the clustering of the basis functions to sources, much research is still required in this area. Recently, Shifted Non-negative Matrix Factorisation (SNMF) was used to cluster these basis functions. To this end, we propose that the incorporation of group sparsity to the Shifted NMF based methods may benefit the clustering algorithms. We have tested this on SNMF algorithms with improved separation quality. Results show that this gives improved clustering of pitched basis functions over previous methods.

1. INTRODUCTION

The process of estimation of individual sound sources from a mixture of single channel audio mixture is known as Monaural sound source separation (SSS). This is a difficult problem due to the complex overlapping of audio signals, produced by sources, in time and frequency. SSS would help in many audio applications which requires analysis, manipulation and re-localisation of audio data like music transcription, pitch modification and conversion of monophonic sound to 5.1 surround system.

Recently, NMF [1] based algorithms have been widely used in separating individual sound sources from a single channel audio mixture. NMF decomposes time-frequency representations such as the magnitude spectrogram of an audio signal into additive parts-based basis functions. NMF approximately decomposes the magnitude spectrogram $\mathbf{X}$ of size $n \times m$ into multiplicative factors $\mathbf{A}$ and $\mathbf{B}$ such that

$$\mathbf{X} \approx \hat{\mathbf{X}} = \mathbf{AB} \tag{1}$$

where $\mathbf{A}$ is $n \times r$ matrix and $\mathbf{B}$ is $r \times m$ matrix. The number of basis functions i.e. r is chosen such as $r < m, n$. Matrix $\mathbf{A}$ contains r frequency basis functions and the corresponding time activation functions are stored in matrix $\mathbf{B}$. The cost function of the form $\mathbf{D}(\mathbf{X}||\hat{\mathbf{X}})$ is minimised to obtain frequency basis functions.

A commonly used cost function for NMF is the generalised Kullback-Leibler (KL) divergence $\mathbf{D}_{KL}(\mathbf{X}||\hat{\mathbf{X}})$ is used as shown in equation 2:

$$\mathbf{D}_{KL}(\mathbf{X}||\hat{\mathbf{X}}) = \sum_{i,j}(\mathbf{X}_{ij} log\frac{\mathbf{X}_{ij}}{\hat{\mathbf{X}}_{ij}} - \mathbf{X}_{ij} + \hat{\mathbf{X}}_{ij}) \tag{2}$$

NMF basis functions typically corresponds to individual notes or chords played by the instruments in the music mixture. Thus, the spectral envelop of each NMF basis function along with the time activation function can be used to re-synthesis the original note. However, there will usually be many more notes than sources. Thus, clustering of these NMF basis functions to their respective sources is required to determine the individual sound sources in a given piece of music. Much research have been done to cluster these basis functions into active sources. Supervised clustering methods have been discussed in [2] to map the separated signals into sources. Unsupervised clustering of separated basis functions by mapping the basis functions to the Mel frequency cepstral domain has been implemented in [3]. Recently, we have proposed a clustering method [4] which uses Shifted Non-negative Matrix Factorisation (SNMF). The basic principle used in this paper is covered in [5], we will go through the details in section 4.1.

A property of NMF is that it typically generates a sparse representation of the given data. This makes the frequency basis function sparse in nature. However, the sparse is random and does not give any spatial or temporal information of the data. Additional constraints on NMF basis functions can be imposed to control the degree of sparseness to improve clustering. One such constraint is group sparsity (GS). GS assumes that each instrument is turned *on* (played) for as little a time as possible and that an individual instrument activation is much sparser than that of a mixture of instru-

ments. Such a constraint has been proposed by [6] that generates the NMF basis functions which benefits sparsity at group level. Thus, each basis function in $\mathbf{A}$ belongs to a group corresponding to an instrument in the mixture. Hence, GS reduces the overlapping of basis functions in time. In [6], GS is incorporated in NMF with the hypothesis that the local amplitudes of the sources are independent and may be derived as a marginal distribution for the activation function $\mathbf{B}$. Further, they used Itakura-Saito (IS) divergence as the cost function. This is done to exploit the equivalence between IS-NMF method and maximum-likelihood estimation of $(\mathbf{A}, \mathbf{B})$ when power spectrum density (PSD) of the input signal is used to calculate frequency basis functions. However, many recent works in audio have used the NMF of magnitude spectra instead of power spectra with better sound separation quality.

To this end, we propose that this incorporation of GS in NMF of magnitude spectra can improve the clustering in recently proposed SNMF-based algorithm [4]. Here, we use the relation between Kl-NMF and ML problem of estimating $\mathbf{A}$ and $\mathbf{B}$ using Poisson distribution [7] as explained in section 3. We also propose that GS constraint can further be integrated in SNMF algorithm for better separation of the individual sources. This can be explained as follows. Let the number of groups to be defined in SNMF algorithm be equal to the number of sources. Again, the sparseness in NMF basis functions can be controlled by giving information about the groups upon activation of SNMF. Then, GS in SNMF would enable the given frequency basis functions to iterate towards the source it belongs and thus improve the quality of separation.

The structure of the paper is as follows: Section 2 outlines the working of statistical model and signal flow of the proposed algorithm. Section 3 illustrates the penalized ML estimation method for GS in KL-NMF. Section 4 gives a overview of the SNMF algorithm and gives the update equations for the proposed SNMF algorithm. A comparison of various SNMF algorithms is done in section 5. Finally, the results of the proposed SNMF algorithm are compared against a previously proposed algorithm in section 6.

2. OVERVIEW OF STATISTICAL MODEL

Figure 1 shows the statistical model for the algorithm proposed. The spectrogram of the input signal is obtained by using the short-time Fourier transform (STFT). Then, the NMF basis functions are obtained from the magnitude spectrogram of a given mixture. Thereafter, the NMF basis functions are then converted into constant Q domain using CQT [8] to exploit the shift-invariant property of the SNMF algorithm. Then, the activation of the SNMF model results in determining the instrument basis functions $\mathbf{A}_r$ for the respective sources. The individual source spectrograms are obtained from $\mathbf{A}_r$ using SNMF masking as explain in section 4.4.

We have incorporated group sparsity at two stages of the proposed algorithm. First of the two stages is calculating the NMF basis functions and second stage is the activation SNMF algorithm to determine the instrument basis functions. However, it can be noted that no knowledge of GS at first stage is used to model the SNMF at second stage and vice versa. We have also tested the performance of IS divergence with and without group sparsity in SNMF algorithms. The prefix g has been used in subscript of SNMF to indicate the use of GS in the given SNMF algorithm. In other words, gkl represents KL divergence with GS and kl refers to KL divergence without GS. For example, SNMF$_{gis-gis}$ represents two stages of SNMF algorithm with group sparsity at both the stages with IS divergence. - in the subscript divides the two

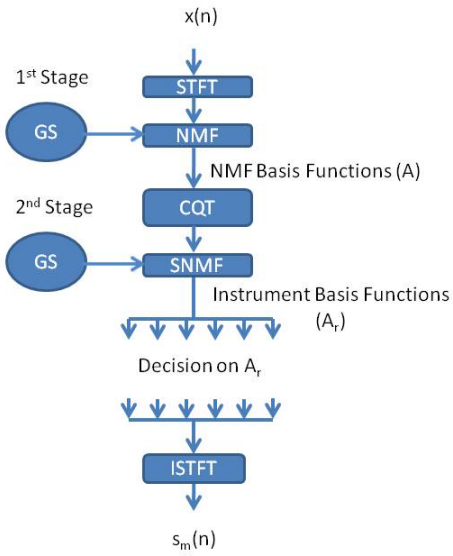


Figure 1: Signal flowchart of the System model

stages where the left side refers to the first stage and the right side represents the second stage. However, NMF$_{kl}$ denotes standard NMF method with KL divergence for calculating frequency basis functions and SNMF$_{gkl}$ represents the 2^{nd} stage of the SNMF algorithm with KL divergence incorporated with group sparsity. We will use these notations for rest of the paper. In the following section we will explain the NMF method with GS to determine NMF basis functions.

3. GROUP SPARSITY WITH KL-NMF

3.1. Equivalence between KL-NMF and ML estimation

The minimising of the cost function in equation 2 to determine $\mathbf{A}$ and $\mathbf{B}$ can be derived from a probabilistic model described in [7]. This can be illustrated as follows. Given the magnitude spectrogram $\mathbf{X}$ of the input signal $\mathbf{x}$, we assume that at every time-frequency interval, the sum of the magnitude of individual source signals $x^r_{m,n}$ is the total magnitude of the observed signal $x_{m,n}$, such that:

$$x_{m,n} = \sum_{r}^{R} x^r_{m,n} \tag{3}$$

where $x^r_{m,n}$ represents the time-frequency atom in the instrument spectrogram x^r produced by the r^{th} source. R is the number sources in the mixture. Also, we make the hypothesis that signals in $x^r_{m,n}$ follows the Poisson distribution. Thus, the magnitude of each $x^r_{m,n}$ can be represented as:

$$x^r_{m,n} \sim \mathcal{P}l(x^r_{m,n}; A_{m,r}B_{r,n}) \tag{4}$$

$$\mathcal{PO}(k;\lambda) = \frac{\lambda^k e^{-\lambda}}{\Gamma(k+1)!} \tag{5}$$

where $B_{r,n}$ is the activation gain for the basis function $A_{m,r}$. Equation 5 defines the Poisson distribution. It can be noted that the summation of the statistically independent Poisson random variable is also a Poisson random variable. Further, as mentioned in [7] the determination of basis functions can be modelled as

$$p(X|A,B) = \mathcal{PO}(X; AB) \tag{6}$$

Alternatively, it can be written as:

$$p(X|A,B) = \prod \frac{e^{-AB}[AB]^X}{\Gamma(X+1)!} \tag{7}$$

The ML solution can be given by taking log and solving which is as follows:

$$\begin{aligned}(A,B) &= arg\max_{A,B} \log p(X|A,B) \\ &= \sum -[AB] + X\log([AB]) - \log(\Gamma(X+1) \\ &\equiv -\mathbf{D}_{KL}(\mathbf{X}||\mathbf{AB})\end{aligned} \tag{8}$$

Thus, we derived a ML estimation of the basis vectors using the probability model in equation 8. We find that this objective is same as minimising the cost function $\mathbf{D}_{KL}(\mathbf{X}||\hat{\mathbf{X}})$ defined in equation 2. In the next section we will incorporate group sparsity with the ML estimation that would favour NMF using KL divergence.

3.2. ML with Group Sparsity

Given r basis functions, we need to group them into g groups, where each of these non-overlapping groups contains all the basis functions corresponds to a particular source. The sparsity constraint have been previously applied on both $\mathbf{A}$ and $\mathbf{B}$ or either $\mathbf{A}$ or $\mathbf{B}$ for many SSS algorithms but until the introduction of group sparsity, this was done on individual basis functions. However, we want to make a given source active for as little time as possible. Therefore, following the principle used in [6], for a given time-frequency frame n, if a source (group) is not *on*, then the corresponding activation gain $B_{g,n}$ should be set to zero. Here, $B_{g,n}$ is a vector of basis functions r_i such that r_i is a member of a given group g ($r_i \in g$ where $1 \leq i \leq m$). Let $\mathbf{B}_n^g$ be defined as a time envelop of the given source for a given time frame n such as

$$\mathbf{B}_n^g = ||B_{g,n}||_1 \tag{9}$$

where $||.||_1$ is L1 norm function. Furthermore, it is assumed that the activation gain $\mathbf{B}_n^g$ for all the individual sources are statistically independent inverse gamma random variables. Thereafter, by using the conditional probability on the activation function B at frame n for r basis functions, the activation gains $B_{r,n}$ can be factorized into groups to determine respective sources. This can be denoted as:

$$p(\mathbf{B}_n|B_n^g) = \prod_g \prod_{r\in g} p(B_{r,n}|B_n^g) \tag{10}$$

The prior of the activation functions $\mathbf{B}_n$ can be calculated using the marginal distribution as follows:

$$p(\mathbf{B}_n) = \prod \frac{\Gamma(g+\beta)}{\Gamma(\beta)} \frac{\alpha^\beta}{(\alpha + B_n^g)^{(\beta+g)}} \tag{11}$$

where α is the scaling factor and the parameter β defines the shape of the gamma distribution. The ML estimation of basis functions $\mathbf{A}$ and gains $\mathbf{B}$ is done using the prior and the term defined in equation 8. This introduction of the penalized term, i.e. the prior information, for the ML estimation is known as MAP (maximum a posterior) estimation. Therefore, the MAP estimation technique can be formulated as:

$$(A,B) = \min_{A,B\geq 0} \mathbf{D}_{KL}(\mathbf{X}||\mathbf{AB}) + \lambda\Phi(B) \tag{12}$$

where the 2^{nd} term $\Phi(B)$ is an optimisation term and is used to uniquely define the grouping pattern. The regularisaition term $\lambda \in [0,1)$ tunes the quality of factorisation obtained and can be set to zero to obtain standard KL-NMF solution.

The update equation for the activation function $\mathbf{A}$ and $\mathbf{B}$ are follows:

$$B \leftarrow B \circledast \left(\frac{A^T(X \circledast \hat{X}^{-\delta})}{A^T(\hat{X}^{-(\delta-1)}) + \lambda\Phi'(B_n^g)} \right) \tag{13}$$

where

$$\Phi(z) = \log(\alpha + z) \tag{14}$$

$$A \leftarrow A \circledast \left(\frac{(X \circledast \hat{X}^{-\delta})B^T}{(\hat{X}^{-(\delta-1)})B^T + \lambda\sum_n B_{r,n}\Phi'(B_n^g))} \right) \tag{15}$$

were δ is set to 1 for KL divergence. $\circledast$ indicates elementwise matrix multiplication. The derivation of update equations can be found in [6] where δ was set to 2 for IS divergence. All operations in equations 13 and 15 are done elementwise. Using these equations the basis functions with GS constraints can be obtained. The obtained frequency basis functions need to be clustered to respective sources for SSS. Recently, an SNMF based algorithm (SNMF$_{kl-kl}$) was proposed to segregate these basis functions to their sources [4]. We argue that further incorporating GS in SNMF will better the quality of separated sources as it would guide the basis function obtained using NMF$_{gkl}$ towards the sources. In section 5, it is shown that for a choice of NMF basis functions, SNMF$_{gkl}$ gives a better clustering than SNMF$_{kl}$. Also, we mention that we are not using GS grouping at first stage to guide SNMF algorithm.

Here, we will explain the significance of the two stage process. In [6], it is mentioned that, in general, clustering of the basis functions using group sparsity close to that of ideal can be achieved for temporal overlapping of sources up to 66%. Therefore, it can be concluded that the GS in the first stage alone did not give good clustering for the basis functions due to 100% overlapping of sources in time, as in the case of the test set used in this paper. Recently, the second stage alone has been implemented [9], i.e. SNMF decomposition, to segregate the frequency basis functions obtained by CQT of the magnitude spectrogram. After testing, we did not get any improvement on the application of GS on the SNMF algorithm discussed in [9]. However, GS did appear to reduce the amount of temporal overlapping in the separated basis functions. The SNMF clustering stage was designed to remove the overlapping in basis functions and group them into sources. Hence, GS assists the SNMF clustering algorithms discussed in this paper and the two stage process was necessary for improving the quality of separation. Next, we will discuss the implementation of GS in KL-SNMF.

4. GROUP SPARSITY WITH KL-SNMF

Having obtained the basis functions using group sparsity in KL-NMF, a knowledge of groups and there sparseness can be intro-

duced in SNMF when clustering these basis functions. This enforcing of basis function towards their respective groups will further improve the clustering and hence improving the separation quality of the individual sources. This can be done in the same way as explained in section 3. Here we will first explain the principle and technique used in SNMF.

4.1. SNMF algorithm

The SNMF algorithm assumes that the timbre of a note produced by a particular instrument does not change for entire range of pitches present in music. Therefore, an instrument in a music mixture can be uniquely defined by the timbre of the note (frequency basis function) play by the particular instrument. Also, a single basis function can be translated to approximate the spectra of all notes played by the instrument in consideration. This can be explained as follows. According to the even tempered chromatic [10] scale, the fundamental frequency of the adjacent notes are geometrically spaced by a constant factor of $\sqrt[12]{2}$. As a result, by translating the frequency basis function up or down in frequency as required, the frequency basis function of one note can be used to approximate that of another note for a particular instrument. A log-frequency spectrogram is required to exploit this shift-invariant property. The log frequency resolution of the frequency basis functions is obtained by the constant Q transform. As SNMF makes use of tensors, we now define the notation used for the tensor parameters in the SNMF model.

We will follow the notations and parameter definitions described in [11] for the SNMF model [5]. Calligraphic upper-case letters ($\mathcal{R}$) denotes tensors of any given dimension. The contracted product of the two tensors of finite dimension results in a tensor. This can be explained as follows.Let a tensor $\mathcal{R}$ be of dimension $I_1 \times \cdots \times I_S \times L_1 \times \cdots \times L_P$ and a tensor $\mathcal{D}$ be of dimension $I_1 \times \cdots \times I_S \times J_1 \times \cdots \times J_N$. Then, the contracted tensor multiplication along the first S modes of $\mathcal{R}$ and $\mathcal{D}$ can be denoted as:

$$\langle \mathcal{RD} \rangle_{\{1,\ldots,S;1,\ldots,S\}} = \sum_{i_1=1}^{I_1} \cdots \sum_{i_1=1}^{I_1} \mathcal{R} \times \mathcal{D} = \mathcal{Z} \tag{16}$$

The dimensions along which the tensors $\mathcal{R}$ and $\mathcal{D}$ are to be multiplied are specified in curly brackets. The resultant tensor $\mathcal{Z}$ will be of dimension $L_1 \times \cdots \times L_P \times J_1 \times \cdots \times J_N$. Indexing of a given tensor is done using lower case letters, such as i and is denoted by $\mathcal{R}(i, j)$.

4.2. Shifted NMF

To incorporate shift-invariant property, the Constant Q spectrogram $\mathcal{C}$ is obtained by multiplying a transform matrix $\mathbf{T}$ with matrix $\mathbf{A}$. Here matrix $\mathbf{A}$ that contains the frequency basis function is considered as a spectrogram and transform matrix $\mathbf{T}$ acts as a warping function which translates linear frequency in $\mathbf{A}$ into Constant Q domain.

$$\mathcal{C} = \mathbf{TA} \tag{17}$$

The spectrogram $\mathcal{C}$ is then factorised using SNMF model to approximately determine the instrument basis function as shown in equation 18 :

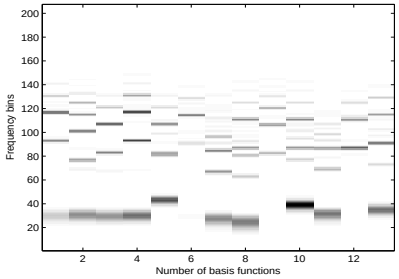


Figure 2: NMF basis function of input mixture in constant Q domain.

$$\mathcal{C} \approx \langle \langle \mathcal{RD} \rangle_{\{3,1\}} \mathcal{H} \rangle_{\{2:3,1:2\}} \tag{18}$$

Here, $\mathcal{R}$ is a translation tensor of dimension $n \times k \times n$ for k possible translations. $\mathcal{R}$ translates the instrument basis functions in $\mathcal{D}$ up or down to approximately cover all the notes played by the required instrument. The tensor $\mathcal{D}$ of size $n \times r$ contains instrument basis functions for each source.Tensor $\mathcal{H}$ of size $k \times r \times m$ is a time activation function such that $\mathcal{H}(i\,,\,j\,,\,:)$ represents the time envelope for the i^{th} translation of the j^{th} source, which indicates when a given note is played by a particular instrument. The cost function used to obtain tensors $\mathcal{D}$ and $\mathcal{H}$ is same as used for NMF.

Therefore, the equivalence between ML estimation of tensors $\mathcal{D}$ and $\mathcal{H}$ and minimising the KL divergence between tensors $\mathcal{C}$ and $\mathcal{D}$ $\mathcal{H}$ can be exploited. The cost function for the decomposition described in equation 18 can be defined as:

$$\mathbf{D}_{KL}(\mathcal{C} || \langle \mathcal{PH} \rangle_{\{2:3,1:2\}})$$

$$\approx \sum_{i,j} (\mathcal{C}_{ij} log \frac{\mathcal{C}_{ij}}{\langle \mathcal{PH} \rangle_{\{2:3,1:2\}}} - \mathcal{C}_{ij} + \langle \mathcal{PH} \rangle_{\{2:3,1:2\}}) \tag{19}$$

where

$$\mathcal{P} = \langle \mathcal{RD} \rangle_{\{3,1\}} \tag{20}$$

where tensor $\mathcal{P}$ contains the translated instrument basis functions.

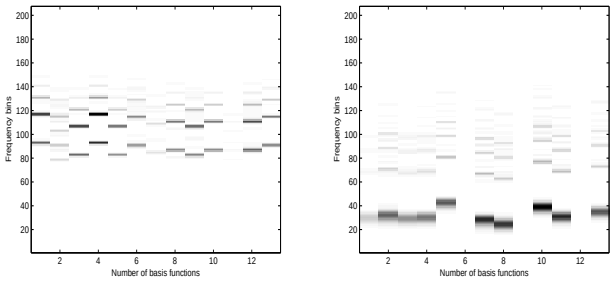
The basis functions in $\mathcal{D}$ are translated using the translation tensor $\mathcal{P}$ as shown in equation 20.

4.3. Update equations for $\mathcal{H}$ and $\mathcal{D}$ with Group Sparsity

Assuming that the number of groups is equal to the number of sources, we can get the required clustering of frequency basis functions. The GS in SNMF can be incorporated by applying the group sparsity constraint on $\mathcal{H}$ and determining the priors using gamma distribution as done in equation 10 and 11. For a given time-frequency frame, let the activation gain $\mathcal{H}_{g,k}$ in SNMF model be the summation of all the components defined by $\mathcal{H}(k, :, :)$ for a particular g. This can be expressed as:

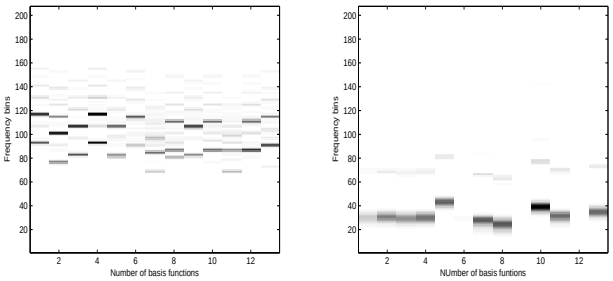
$$\mathcal{H}_{g,k} = \sum_{k} \mathcal{H}(k, g, :) \tag{21}$$

where k is the number of frequency shifts. Further, with the knowledge of priors of the activation function $\mathcal{H}$, the SNMF problem can be reduced to the ML estimation of the tensors $\mathcal{D}$ and $\mathcal{H}$. The penalised ML solution for the KL-SNMF problem can be defined as:



(a) Source 1 (b) Source 2

Figure 3: Clustering of NMF_{kl} basis function using SNMF_{gkl}



(a) Source 1 (b) Source 2

Figure 4: Clustering of NMF_{kl} basis function using SNMF_{kl} .

$$\langle \mathcal{P}, \mathcal{H} \rangle = \min_{\mathcal{P},\mathcal{H} \geq 0} \mathbf{D}_{KL}(\mathcal{C} || \langle \mathcal{P}\mathcal{H} \rangle_{\{2:3,1:2\}}) + \lambda \Phi(\mathcal{H}) \tag{22}$$

The optimisation term $\Phi(\mathcal{H})$ is again used to define the group sparsity constraint. The interactive multiplicative update equations for $\mathcal{P}$ and $\mathcal{H}$ can be derived in a manner similar to [5]. This can be formulated as follows:

$$\mathcal{H} \leftarrow \mathcal{H} \circledast \left(\frac{\langle \langle \mathcal{R}\mathcal{D} \rangle_{\{3,1\}} \mathcal{Y} \rangle_{\{3,1\}}}{\langle \langle \mathcal{R}\mathcal{D} \rangle_{\{3,1\}} \mathcal{O} \rangle_{\{1,1\}} + \lambda \Phi'(\mathcal{H}_{g,k})} \right) \tag{23}$$

where

$$\mathcal{Y} = \frac{\mathcal{C}}{\langle \mathcal{P}\mathcal{H} \rangle_{\{2:3,1:2\}}} \tag{24}$$

and $\mathcal{O}$ is a tensor of all ones. The multiplicative updates for the translated basis functions in $\mathcal{D}$ can be by using following equations:

$$\mathcal{W} = \langle \mathcal{R}\mathcal{O} \rangle_{\{1,1\}} \tag{25}$$

$$\mathcal{D} \leftarrow \mathcal{D} \circledast \left(\frac{\langle \mathcal{Z}\mathcal{H} \rangle_{\{1:3,1:3\}}}{\langle \mathcal{W}\mathcal{H} \rangle_{\{1:3,1:3\}} + \lambda \sum_n \mathcal{H}_{r,n} \Phi'(\mathcal{H}_{g,k})} \right) \tag{26}$$

where

$$\mathcal{Z} = \langle \mathcal{R}\mathcal{Y} \rangle_{\{1,1\}}$$

where function $\Phi(z)$ is same as stated in equation 14. The multiplicative updates and the positive (random numbers) initialization for $\mathcal{D}$ and $\mathcal{H}$ ensures the positive tensor factorisation. The number of translations k in $\mathcal{R}$ is chosen such that the translated (frequency-shifted) instrument basis functions cover all the notes or chords corresponding to basis functions in the mixture. Thus, by using Constant Q spectrogram $\mathcal{C}$ as an input, the SNMF with group sparsity helps in separating the instrument basis functions.

4.4. Signal reconstruction

Individual source spectrograms $\mathbf{C_r}$ are reconstructed by using the slices of tensors, $\mathcal{D}(:, r)$ and $\mathcal{H}(:, r, :)$, corresponding to the r^{th} source.

$$\mathbf{C}_r = \mathcal{C}(:, :, r) = \langle \langle \mathcal{R}\mathcal{D}(:, r) \rangle_{\{3,1\}} \mathcal{H}(:, r, :) \rangle_{\{2:3,1:2\}} \tag{27}$$

A limitation of using SNMF algorithm is that there is no true inverse of CQT which results in lower separation quality of separated sources. Therefore, in the absence of inverse CQT, the recovered individual source spectrograms $\mathbf{C_r}$ are mapped back to the linear domain to obtain $\mathbf{A_r}$. This can be done as follows. An approximate inverse of $\mathbf{T}$ (see equation 17) is multiplied with individual source spectrograms $\mathbf{C}_r$ to recover corresponding $\mathbf{A}_r$ as shown in equation:

$$\mathbf{A}_r = \mathbf{T}^{'} \mathbf{C}_r \tag{28}$$

Having obtained source frequency basis functions $\mathbf{A}_r$, individual sound sources can be reconstructed, thus achieving the source separation. The details of the synthesis of the sources can be found in [4].

Figure 2 shows the log-frequency spectra of the NMF_{gkl} basis functions of a test mixture of two sources. The x-axis shows the frequency basis functions for all the notes played by the instruments present in mixture .The application of SNMF algorithm separates the basis functions into two groups corresponding to the individual sources. The separated basis functions of source 1 and source 2 respectively can be seen in figure 4. SNMF_{kl} was used to generate the figure 4. The separated source spectrograms using SNMF_{gkl} can also be seen in figure 3. The separated basis functions are more visible for respective sources for SNMF_{gkl} as compared against SNMF_{kl}. Thus, by inspecting the figures 4 and 3, we can conclude that SNMF_{gkl} works better than SNMF_{kl} to separate basis functions. The results show distinct groupings of basis functions and can further be used to separate sources in the mixture. Hence, SNMF with GS constraint can be used to cluster basis functions in monaural mixtures.

5. EXPERIMENTS

A dataset of 25 monaural test mixtures were used to test the performance of all the SNMF algorithms discussed in this paper. The 25 test signals were the mixtures of 2 instruments and were generated by using a huge library of orchestral samples of notes and chords produced by a total of 15 different orchestral instruments [12]. The sampling rate of the input mixtures were 44.1 kHz and were of 4 to 8 seconds in length. The test mixtures contains overlapping harmonics and notes played by different instruments in the mixture. This ensures the capability of SNMF-based algorithm to separate notes played simultaneously by harmonic instruments. The details of the dataset can be found in [13]

The widely used quality measures [14] signal-to-distortion ratio (SDR), the signal-to-interference ratio (SIR), and the signal-to-artefacts ratio (SAR) were used for evaluation of audio outputs from various algorithms. SDR calculates the amount of distortion

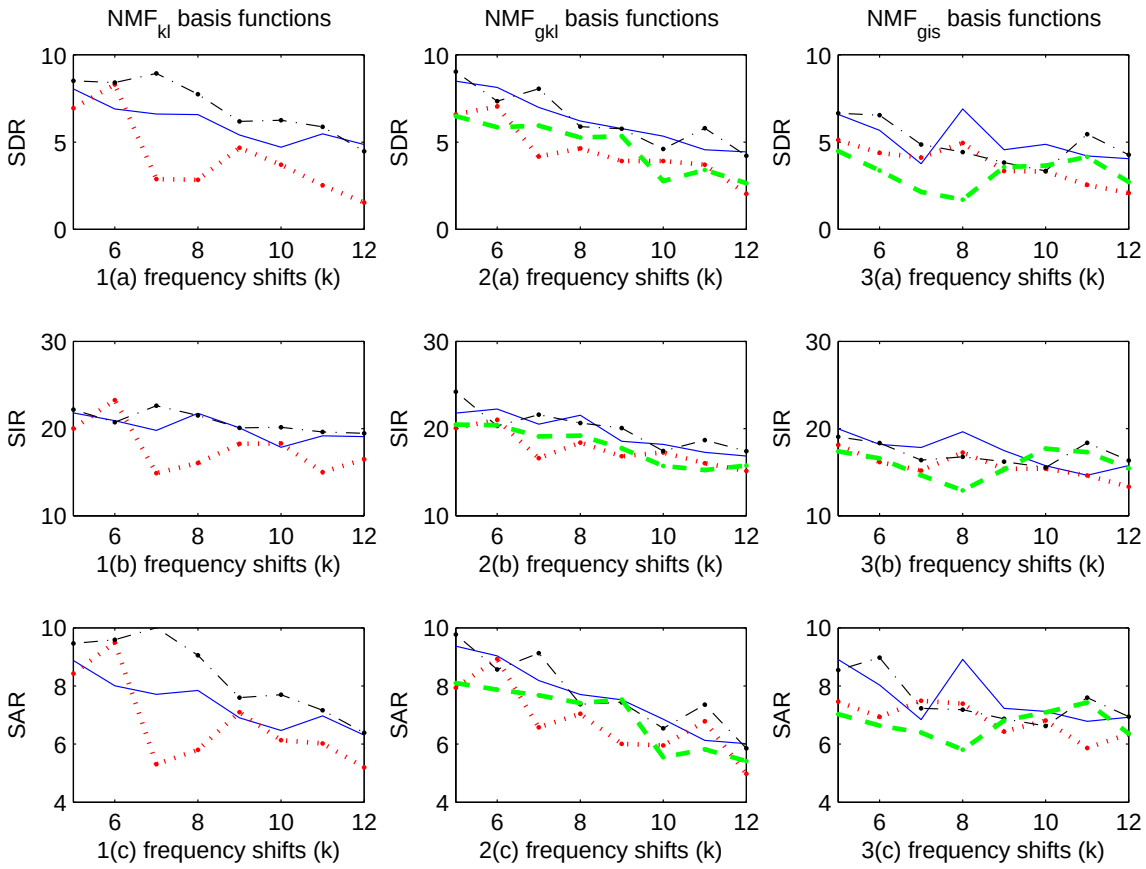


Figure 5: *Performance evaluation of SNMF$_{kl}$(blue solid line), SNMF$_{is}$(red dotted line), SNMF$_{gkl}$(black dash-dot line) and SNMF$_{gis}$(green dashed line) to group basis functions generated by NMF$_{kl}$(1st column), NMF$_{gkl}$(2nd column) and NMF$_{gis}$(3rd column) for different number of frequency shifts*

present in the reconstructed signal, SIR determines the interference of other sound sources in the separated signal and SAR measures the artefacts present in the separated signal as a result of data processing and reconstruction. The details of definition of the quality measures can be found in [14]. These algorithms were tested for the same set of input mixtures of 2 instruments. The magnitude spectra and not the power spectra of the input signal were used to calculate frequency basis functions as it gave better results for the test mixtures. The magnitude spectrogram of the time-domain signal were obtained using the STFT with a 75% overlapping Hann window, 4096 samples in length. The number of basis functions were set to 13 to cover all the notes played in the mixture.

Matrices $\mathbf{A}$ and $\mathbf{B}$ in equation 1 were initialised with random positive numbers and NMF was run for 300 iterations. 24 frequency bins per octave ranging from $55Hz$ to $22.05kHz$ were used for CQT. The number of sources in the SNMF algorithm was set to 2. The SNMF algorithm ran for 50 iterations. A number of different tests were conducted to efficiently determine the frequency basis functions using NMF and to determine the effect of the number of different translations 'k' in frequency on various SNMF algorithms. We were hoping that the frequency shifts would give some insights to the clustering obtained and would give a clear comparison of the various SNMF methods. Number of frequency shifts were ranged from 5 to 12. The number of groups in GS was limited to 2 for the given tests.

A summary of the results for all the SNMF algorithms are shown in figure 5. The scores for all the quality measures were calculated and graphed against the allowable frequency shifts k. The results were determined by finding the average of the quality measures obtained for each separated source for each input mixture. Each set of quality measure, say SDR in figures (a), (d) and (g), illustrates the comparison of all the listed SNMF algorithms for NMF$_{kl}$, NMF$_{gkl}$ and NMF$_{gis}$ basis functions respectively. Although, the GS constraint in SNMF$_{gis}$ helps in enhance the clustering of NMF$_{gkl}$ basis functions as compared against SNMF$_{is}$ but it fails to improve the grouping of for NMF$_{gis}$ basis functions than that of SNMF$_{is}$. Also, the results of SNMF$_{kl-gis}$ were not good and were not included. Through visual inspection it can be seen that SNMF algorithms with KL divergence (SNMF$_{kl}$ and SNMF$_{gkl}$) completely outperforms SNMF model with IS divergence (SNMF$_{is}$ and SNMF$_{gis}$). As a result, we will elaborate more on SNMF algorithms based on KL divergence in section 6.

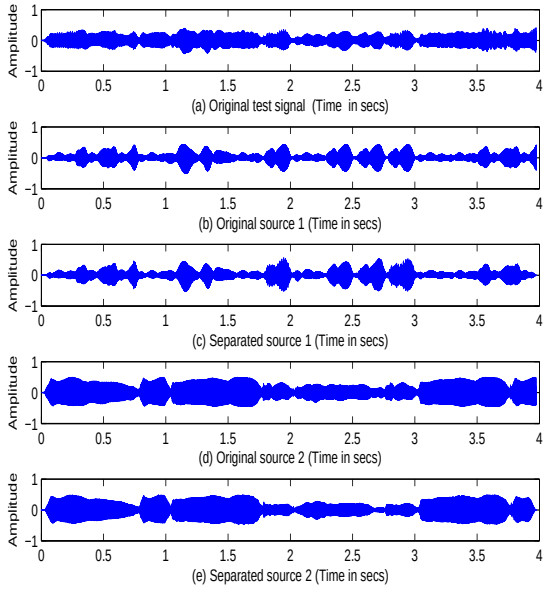


Figure 6: The figure shows (a) the original test signal, (b) the original source 1, (c) the separated source 1, (d) the original source 2, (e) the separated source 2 in time domain.

Figure 6 shows the audio waveforms in the time domain of an input mixture signal of two sources. The corresponding original separated sources and synthetic sources are also shown. This waveforms are obtained by using SNMF$_{gkl-gkl}$ method. It can be seen from the various waveforms that the original sound source and reconstructed sound signals closely match with each other and after listening to the separated sources it was found that the notes and the melodies played by the sources in mixture have separated well. Thus, the proposed SNMF algorithm with GS can be used to separate sound source signals in a monaural mixture.

6. RESULTS

In this section, we will compare the result of the proposed SNMF model with GS constraint against a recently proposed SNMF clustering algorithm (SNMF$_{mask}$) [4]. It is important to note that the SNMF$_{mask}$ algorithm is same as SNMF$_{kl-kl}$ as denoted in this paper. As discussed in section 5 that the SNMF algorithms with KL divergence works better for clustering the basis functions as compared against the SNMF algorithms with IS divergence. Therefore, we will discuss more on SNMF algorithms with KL divergence. It can be concluded from the figure 5 that for NMF$_{kl}$ basis functions, SNMF$_{gkl}$ improves the grouping of basis functions as compared to SNMF$_{kl}$. Also, SNMF$_{gkl}$ is marginally better than SNMF$_{kl}$ to group the NMF$_{kl}$ basis functions. However, both the SNMF algorithms, SNMF$_{gkl-gkl}$ and SNMF$_{gkl-kl}$ scores low as the frequency shifts increases.

To compare the results listed in [4], the highest scores of the quality measures for the separated sound sources for each mixture were hand-picked for the given range of frequency shifts such that

$$SDR = \max_k SDR_k,\ k \in K \tag{29}$$

where K is the number of frequency shifts. The results were then calculated by averaging the metrics (SDR, SIR and SAR) over each of the separated sources for all the test mixtures. Thereafter, the mean SDR, SIR and SAR were obtained by finding the average over each of the input mixture. The mean of the quality measures shown in table 1 are in DB. It can be seen from the table each of the SNMF algorithm with group sparsity performs better than SNMF$_{kl-kl}$. We can also see that SNMF$_{gkl}$ performs better clustering for basis functions generated by NMF$_{kl}$ and is marginally better for NMF$_{gkl}$. Hence, the GS in SNMF improves clustering for NMF basis functions.

SNMF algorithm	SDR	SIR	SAR
SNMF$_{kl-kl}$	10.81	26.75	11.50
SNMF$_{kl-gkl}$	11.79	27.09	12.38
SNMF$_{gkl-kl}$	10.83	26.04	11.43
SNMF$_{gkl-gkl}$	10.98	25.81	11.64

Table 1: Mean SDR, SIR and SAR for separated sound sources using SNMF algorithm

7. CONCLUSIONS

We have presented a Shifted NMF based clustering technique to cluster the frequency basis functions. We have incorporated group sparsity at two stages of the SNMF algorithm. We have explained how the incorporation of group sparsity at first stage can improve the clustering of frequency basis functions by reducing the overlapping of basis functions. Subsequently, at the second stage, the group sparsity would guide the basis functions to their respected groups corresponding to sources. A probabilistic model is used to exploit the equivalence between ML problem and minimising KL divergence cost function to estimate of frequency basis functions. Group sparsity was incorporated in the activation gain functions $\mathbf{B}$ and $\mathcal{H}$ respectively for the first and second stages of the SNMF algorithm. An optimisation term was used to tune the grouping criteria. Results show that incorporating GS improves the clustering of frequency basis function in SNMF model, thus improving the separation quality. The presented algorithm can be potentially used to separate multiple sources in a monaural mixture.

8. ACKNOWLEDGMENTS

The authors wish to acknowledge the Dublin Institute of Technology (Dublin, Ireland) for funding under the ABBEST scholarship programme.

9. REFERENCES

[1] D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorisation, " *Advances in Neural Information Processing System*, 2000, pp. 556-562.

[2] T. Virtanen, "Sound Source Separation Using Sparse Coding with Temporal Continuity Objective," *International Computer Music Conference*, 2003.

[3] M. Spiertz and V. Gnann, "Source-Filter based clustering for monaural blind source separation," *Proceedings of the* 12^{th} *International Conference on Digital Audio Effects*, Italy, 2009.

[4] R. Jaiswal, D. FitzGerald, E. Coyle and S. Rickard, "Clustering NMF basis functions using Shifted NMF for monaural sound source separation," *in Proc. IEEE Int. Conference on Acoustic Speech and Signal Processing ICASSP*, May, 2011.

[5] D. FitzGerald, M. Cranitch and E. Coyle, "Shifted Non-negative matrix factorisation for sound source separation," *IEEE Workshop of Statistical Signal Processing, Bordeaux*, France, 2005.

[6] A. Lefevre, F. Bach and C. Fevotte, "Itakura-Saito nonnegative matrix factorization with group sparsity," *in Proc. IEEE Int. Conference on Acoustic Speech and Signal Processing ICASSP*, May, 2011.

[7] T. Virtanen, A. T. Cemgil and S. Godsill, "Bayesian Extensions to Non-Negative Matrix Factorisation for Audio Signal Modelling," *in Proc. IEEE Int. Conference on Acoustic Speech and Signal Processing ICASSP*, April, 2008.

[8] J. C. Brown, "Calculation of a Constant Q spectral transform," *Journal of the Acoustic Society of America*, vol. 89, no.1, pp 425-434, 1991.

[9] R. Jaiswal, D. FitzGerald, E. Coyle and S. Rickard, "Shifted NMF Using an Efficient Constant Q Transform for Monaural Sound Source Separation," *22nd IET Irish Signals and Systems Conference*, 23-24 June, 2011.

[10] E. M. Burns, "Intervals, Scales and Tuning," *The Psychology of Music*, D. Deutsch, Ed. Academic Press, 1999.

[11] B. W. Bader and T. G. Kolda, "Algorithm 862: MATLAB tensor classes for fast algorithm prototyping," *ACM Transactions on Mathematical Software*, vol. 32, no. 4, pp. 635-653, 2006.

[12] P. Siedlaczek, "Advanced Orchestra Library Set," 1997.

[13] D. FitzGerald, M. Cranitch and E. Coyle, "Extended Non-negative Tensor Factorisation Models for Musical Sound Source Separation," *Computational Intelligence and Neuroscience*, Hindawi Publishing Corp., 2008.

[14] E. Vincent, R. Gribonval, and C. Fevotte, "Performance measurement in blind audio source separation," *IEEE Transactions on Audio, Speech and Language Processing*, 2006.