An Experimental Survey on Non-Negative Matrix Factorization for Single Channel Blind Source Separation

Mona Nandakumar M
Department of Electronics and Communication Engineering
MES College of Engineering
Kuttippuram, Kerala, India

Edet Bijoy K
Department of Electronics and Communication Engineering
MES College of Engineering
Kuttippuram, Kerala, India

ABSTRACT

In applications such as speech and audio denoising, music transcription, music and audio based forensics, it is desirable to decompose a single-channel recording into its respective sources, commonly referred to as blind source separation (BSS). One of the techniques used in BSS is non-negative matrix factorization (NMF). In NMF both supervised and unsupervised mode of operations is used. Among them supervised mode outperforms well due to the use of pre-learned basis vectors corresponding to each underlying sources. In this paper NMF algorithms such as Lee Seung algorithms (Regularized Expectation Minimization Maximum Likelihood Algorithm (EMML) and Regularized Image Space Reconstruction Algorithm (ISRA)), Bregman Divergence algorithm (Itakura Saito NMF algorithm (IS-NMF)) and an extension to NMF, by incorporating sparsity, Sparse Non-Negative Matrix Factorization(SNMF) algorithm are used to evaluate the performance of BSS in which supervised mode is used. Here signal to distortion ratio (SDR), signal to interference ratio (SIR) and signal to artifact ratio (SAR) are measured for different speech and/or music mixtures and performance is evaluated for each combination.

General Terms:

Source Separation, Blind Source Separation, BSS Evaluation

Keywords:

NMF, EMML, ISRA, IS-NMF, SNMF, Bregman Divergence

1. INTRODUCTION

Separation of mixed signals has long been considered as an important and fundamental issue in signal processing, with a wide variety of applications in telecommunications, audio and speech signal processing, and biomedical signal processing etc. Audio and speech separation systems find a variety of potential applications including automatic speech recognition (ASR) under adverse noise conditions, and multimedia or music analysis where signals are mixed purposefully from multiple sources.

Source separation methods can usually be classified as blind and non-blind methods based on the characteristics of underlying mixtures. In blind source separation (BSS), the completely unknown sources are separated without the use of any other information besides the mixture. These methods typically rely on the assumption that the sources are non-redundant, and the methods are based on, decorrelation, statistical independence, or the minimum description length principle. In Non-blind methods, the separation is based on the availability of further information such as prior distribution, about the mixture. The NMF based algorithms are used here for blind source separation scenario [1].

Like NMF the most commonly used method in BSS is Independent Component Analysis (ICA). In ICA, the linear representation of nongaussian data is calculated so as to make the components statistically independent, or as independent as possible. Such a representation seems to capture the essential structure of the data in many applications, including feature extraction and signal separation etc. But when both sources and mixing matrices are unknown; ICA cannot determine the variances (energies) of the independent components as well as the order of the independents sources because of the basis functions are ranked by non-gaussianities [2]. Lee and Seung [3] have suggested a solution for BSS problem with non negativity constraints. In NMF, the non negativity constraint leads to the parts based representation of the input mixture which helps to develop structural constraints on the source signals. NMF does not require the independence assumption, and is not restricted to data lengths. It yields more importants to the basis vectors for reconstructing the underlying signal than the activation vectors. In NMF the basis functions are not ranked; the order of the underlying sources does not change in ICA. As from [1], [2], it is found that NMF is attractive and it out performs ICA in BSS environment. The spatial and temporal correlations between variables are more accurately taken into account by NMF which helps to makes NMF a useful tool for decomposition of multivariate data. This paper focused on BSS using NMF with decorrelation as a method for updating the activation vectors.

Most NMF algorithms focus on minimizing the cost function such as Kullback-Leibler divergence or squared Frobenius norm or Itakura Saito Divergence etc using multiplicative or additive updates. This paper compares the performances of four multiplicative algorithms Regularized EMML and Regularized ISRA proposed by Lee and Seung [3], [4], IS-NMF proposed by Bregman [5], and SNMF proposed by Hoyer [6] for NMF.

2. BSS USING NMF

BSS is a method to separate the independent sources from mixed observations, where mixing process is unknown. It may lead to determined (no: of sensors = no: of sources), overdetermined (no: of sensors $>$ no: of sources) or underdetermined (no: of sensors $<$ no: of sources) cases when number of sources and number of sensors varies. When single-channel source separation problem is taken as underdetermined one, it cannot in general be solved without the prior knowledge of underlying sources within the mixture. Due to this, the problem of estimating several overlapping sources from one input mixture is ill-defined and complex in BSS environment. But NMF gives a solution to this single channel source separation problem by utilizing its non negativity constraint as well as supervised mode of operation for source separation [7].
NMF is defined as

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

Where $V \in R_+^{F\times T}$ is the speech spectrogram, $W \in R_+^{F\times K}$ is the matrix of basis vectors (columns) and $H \in R_+^{K\times T}$ is the matrix of activations (rows) of the input mixture.
In NMF when the spectrogram of mixture V is given, the matrices W and H can be computed via an optimization problem by

$$\underset{W,H\geq 0}{\text{Min}}\ D(V\|W,H) \tag{2}$$

where D denotes the divergence.
The complex sound needs more than one basis vector for separation in unsupervised mode of operation, it is difficult to control which basis vector explains which source within the mixture. The 'right' value of K avoids the factorization errors makes BSS accurate. One way to control the factorization problem is by modifying F, T and K values which defines dimensionality of factorized matrices. But operation in supervised mode is much simpler than modifying dimensionalities, uses an isolated training data of each source within a mixture to pre-learn individual models of underlying source [8]. The speech and/or music data base for source separation are taken from [9] and is used as input to evaluate the performance of Regularized EMML, Regularized ISRA, IS-NMF and SNMF algorithms by varying K from 5 to 100 with constant number of iteration 100. The performance evaluation measures of SDR, SIR and SAR determines the quality of the underlying algorithms [10].

3. NMF ALGORITHMS

EMML and ISRA algorithms are two among the Regularized Lee-Seung algorithms group, usually uses an alternating minimization of a cost function $D(V\|WH)$ which subject to the non negativity constraints $(W,H \geq 0)$. In Regularized EMML algorithm Kullback-Leibler cost function is minimized where as in Regularized ISRA algorithm it minimizes the squared Frobenius norm. In NMF algorithms any suitably designed cost function has two sets of parameters (W and H), it usually employ constrained alternating minimization, i.e., in one step W is estimated and H fixed, and in the other step fix H and estimate W.

3.1 Regularized EMML Algorithm

Kullback-Leibler cost function is given by

$$D(V\|\hat{V}) = \sum_{i,j}(V_{ij} log\frac{V_{ij}}{\hat{V}_{ij}} - V_{ij} + \hat{V}_{ij}) \tag{3}$$

To minimize

$$D(V\|W,H) = \sum_{i,j}(V_{ij} log\frac{V_{ij}}{(WH)_{ij}} - V_{ij} + (WH)_{ij}) \tag{4}$$

block coordinate descent technique is used.

$$D(V\|W,H) = \sum_{i,j} -V_{ij}\log\sum_k W_{ik}H_{kj} + \sum_{i,j}\sum_k W_{ik}H_{kj} \tag{5}$$

Since both W and H are not convex together, closed form optimization is not possible. So to minimize divergence, the auxiliary function taken in Expectation-Maximization algorithm [4] is also used here. To maximize the function, a useful tool is Jensen's inequality, which says that for convex functions f: f(average) $\leq$ average of f.
To apply Jensen's inequality, weights introduced as $\sum_k \pi_{ijk} = 1$

$$D(V\|W,H) \leq \sum_{i,j}(-V_{ij}\log\sum_k \pi_{ijk}\frac{W_{ik}H_{kj}}{\pi_{ijk}} + \sum_{i,j}\sum_k W_{ik}H_{kj}) \tag{6}$$

So the function can be minimized exactly as

$$H^*_{kj} = \frac{\sum_i V_{ij}\pi_{ijk}}{\sum_i W_{ik}} \tag{7}$$

where $\pi_{ijk} = \frac{W_{ik}H^{(l)}_{kj}}{\sum_k W_{ik}H^{(l)}_{kj}}$ so

$$H^{(l+1)}_{kj} \leftarrow H^{(l)}_{kj}.\frac{\sum_i (\frac{V}{WH^{(l)}})_{ij}W_{ik}}{\sum_i W_{ik}} \tag{8}$$

In matrix form it can be represented as

$$H_{(l+1)} \leftarrow H_{(l)}.\frac{W^T\frac{V}{WH^{(l)}}}{W^T1} \tag{9}$$

In similar manner W can be calculated.

Algorithm 1 Regularized EMML Algorithm

Using $D(V\|WH) = D(V^T\|W^TH^T)$, it obtain a similar update for W
Now just iterate between

(1) Updating W.
(2) Updating H.
(3) Checking $D(V\|WH)$. If the change since the last iteration is small, then declare convergence.

1: Initialize W, H
2: **repeat**

$$H \leftarrow H\cdot\frac{W^T\frac{V}{WH}}{W^T1} \qquad W \leftarrow W\cdot\frac{\frac{V}{WH}H^T}{1H^T}$$

3: **until** convergence return W, H.

3.2 Regularized ISRA Algorithm

The squared Euclidean distance (squared Frobenius norm) cost function is given by,

$$D_F(V\|WH) = \frac{1}{2}\|V - WH\|_F^2 \tag{10}$$

Applying the standard gradient descent technique to the cost function, it get

$$W_{ij} \leftarrow W_{ij}\frac{[VH^T]_{ij}}{[WHH^T]_{ij}} \tag{11}$$

$$H_{jk} \leftarrow H_{kj}\frac{[W^TV]_{jk}}{[W^TWH]_{jk}} \tag{12}$$

In matrix notation, the Lee-Seung Euclidean multiplicative updates become

$$W \leftarrow W \odot \frac{VH^T}{WHH^T} \qquad H \leftarrow H \odot \frac{W^TV}{W^TWH} \tag{13}$$

Algorithm 2 Regularized ISRA Algorithm

Using $D_F(V\|WH) = D_F(V^T\|W^TH^T)$, it obtain a similar update for W
Now just iterate between.

(1) Updating W.
(2) Updating H.
(3) Checking $\|V - WH\|$. If the change since the last iteration is small, then declare convergence.

1: Initialize W, H
2: **repeat**

$$H \leftarrow H \cdot \frac{W^TV}{WHH^T} \qquad W \leftarrow W \cdot \frac{VH^T}{W^TWH}$$

3: **until** convergence return W, H.

3.3 Itakura Saito Divergence Algorithm

Itakura Saito NMF is belongs to the class of Bregmans divergence where the underlying function is strictly considered as convex in real space. NMF with Itakura Saito Divergence is given by

$$d_{IS}(\frac{V}{WH}) = \frac{V}{WH} - log\frac{V}{WH} - 1 \tag{14}$$

It is obtained from the maximum likelihood (ML) estimation[4] of short-time speech spectra under autoregressive modeling. The IS divergence belongs to the class of Bregman divergences and is a limit case of the $\beta-$ divergence. Thus, the gradient descent multiplicative rules are applied here. The gradients of criterion $D\beta(V|WH)$ wrt W and H is represented as

$$\nabla_H D_\beta(V|WH) = W^T((WH)^{\cdot\beta-2} \cdot (WH - V)) \tag{15}$$

$$\nabla_W D_\beta(V|WH) = ((WH)^{\cdot\beta-2} \cdot (WH - V))H^T \tag{16}$$

where $\cdot$ denotes Hadamard entry wise product and A^n denotes the matrix with entries $[A]_{ij}^n$. The multiplicative gradient descent approach taken is equivalent to updating each parameter by multiplying its value at previous iteration by the ratio of the negative and positive parts of the derivative of the criterion wrt this parameter, namely$\theta \leftarrow \theta \cdot [\nabla f(\theta)]_- / [\nabla f(\theta)]_+$, where $\nabla f(\theta) = [\nabla f(\theta)]_+ - [\nabla f(\theta)]_-$ and the summands are both non negative. This ensures non negativity of the parameter updates, provided initialization with a nonnegative value. A fixed point θ^* of the algorithm implies either $\nabla f(\theta^*) = 0$ or $\theta^* = 0$. This leads to the updates for W and H,

$$H \leftarrow H \cdot \frac{W^T((WH)^{\cdot\beta-2}V)}{W^T(WH^{\cdot\beta-1})} \tag{17}$$

$$W \leftarrow W \cdot \frac{((WH)^{\cdot\beta-2}V)H^T}{(WH^{\cdot\beta-1})H^T} \tag{18}$$

where β=0.

Algorithm 3 Itakura Saito NMF Algorithm

Using $D_F(V\|WH) = D_F(V^T\|W^TH^T)$, it obtain a similar update for W
Now just iterate between.

(1) Updating W.
(2) Updating H.
(3) Checking $\|V - WH\|$. If the change since the last iteration is small, then declare convergence.

1: Initialize W, H
2: **repeat**

$$H \leftarrow H \cdot \frac{W^T\frac{V}{(WH)^2}}{W^T\frac{1}{WH}} \qquad W \leftarrow W \cdot \frac{\frac{V}{(WH)^2}H^T}{\frac{1}{WH}H^T}$$

3: **until** convergence return W, H.

3.4 SNMF Algorithm

One of the most useful properties of NMF is that it usually produces a sparse representation of the data. Such a representation encodes much of the data using few active components, which makes the encoding easy to interpret. So Sparse NMF is an extension of NMF, in which an additional sparsity constraint is enforced on either the matrix **H** or **W**, i.e., a solution is sought where only a few basis vectors are active simultaneously. The sparse NMF problem can be formulated as

$$\underset{W,H\geq 0}{\text{Min}}\ D(V, WH) + \beta(H) \tag{19}$$

where β is a penalty term that enforces the sparsity. This penalty could be selected as the 0-norm, i.e., the count of non-zero elements in **H**, but this leads to a very rough cost function that is hard to minimize because of its many local minima. A penalty function that leads to a smoother regularization while still inducing sparsity is the the 1-norm, which, in Bayesian terms, corresponds to assuming an exponential prior over **H**. In practice $\beta(H = \lambda\sum_{i,j} H_{i,j})$, where λ is a parameter which controls the tradeoff between sparsity and accuracy of the approximation. To use this penalty function a normalization constraint on either **W** or **H** is introduced, since trivial solutions minimizing β can be found by letting **H** decrease and **W** increase accordingly. With the sparseness penalty, the data is modeled not only as a non-negative linear combination of a set

of basis vectors, but as linear combinations using only a few basis vectors at a time. This allows us to compute an over complete factorization, i.e., a factorization with more basis vectors than the dimensionality of the data. Without the sparsity constraint, any basis spanning the entire positive orthant would be a solution.
Patrik O. Hoyer [6] has developed a projected gradient descent algorithm for NMF with sparseness constraints. This algorithm essentially takes a step in the direction of the negative gradient, and subsequently projects onto the constraint space, making sure that the taken step is small enough that the objective function $E(W, H) = \|V - WH\|^2$ is reduced at every step. The projection operator, which enforces the desired degree of sparseness in the algorithm.

Algorithm 4 NMF with Sparseness Constraint on W

(1) Initialize **W** and **H** to random positive matrices
(2) If sparseness constraints on **W** apply, then project each column of **W** to be non-negative, have unchanged L_2 norm, but L_1 norm set to achieve desired sparseness

1: Iterate
2: **if** Sparseness constraints on **W** apply **then**
3: Set $W := W - \mu_W (WH - V)H^T$
4: Project each column of **W** to be non-negative, have unchanged L_2 norm, but L_1 norm set to achieve desired sparseness
5: **else** {Take standard multiplicative step $W := W \otimes (VH^T) \oslash (WHH^T)$.}
6: **until** convergence return W, H.

Algorithm 5 NMF with Sparseness Constraint on H

(1) Initialize **W** and **H** to random positive matrices
(2) If sparseness constraints on **H** apply, then project each row of **H** to be non-negative, have unit L_2 norm, and L_1 norm set to achieve desired sparseness

1: Iterate
2: **if** Sparseness constraints on **H** apply, **then**
3: Set $H := H - \mu_H W_T (WH - V)$
4: Project each row of **H** to be non-negative, have unit L_2 norm, and L_1 norm set to achieve desired sparseness
5: **else** {Take standard multiplicative step $H := H \otimes (W^T V) \oslash (W^T WH)$.}
6: **until** convergence return W, H.

where, $\otimes$ and $\oslash$ denote element wise multiplication and division, respectively. Moreover, μ_W and μ_H are small positive constants (step sizes) which must be set appropriately for the algorithm to work. Many of the steps in the Algorithm 5 and Algorithm 6 require a projection operator which enforces sparseness by explicitly setting both L_1 and L_2 norms (and enforcing non-negativity). This operator is defined as, for any vector **x**, the closest non-negative vector **s** with a given L_1 norm and a given L_2 norm can be obtained as

Algorithm 6 Projection Operaror Calculation

1: Set $s_i := x_i + (L_1 \sum x_i)/dim(x), \forall i$
2: Set Z := {}
3: Iterate
4: Set m$_i$:= $\begin{cases} L_1/(dim(x)size(Z)) & \text{if } i \notin Z \\ 0 & \text{if } i \in Z \end{cases}$
5: Set $s := m + \alpha(s - m)$, where $a \geq 0$ is selected such that the resulting **s** satisfies the L_2 norm constraint. This requires solving a quadratic equation.
6: **if** all components of **s** are non-negative **then**
7: **return s**
8: **else** {Set $Z := Z \cup i; s_i < 0$}
9: Set $s_i := 0, \forall i \in Z$
10: Calculate $c := (\sum s_i - L_1)/(dim(x) - size(Z))$
11: Set si := $s_i - c, \forall i \notin Z$
12: go to (4)
13: **until** s become non-negative.

3.5 Procedure for Complete Supervised Process

In this paper the supervised procedure found in [8] is incorporated in NMF algorithms to make the reconstruction effective in BSS methods. The complete procedure for supervised source separation process is as follows

Algorithm 7 Procedure for Complete Supervised Process

1: Use isolated training data to learn a factorization $(W_s H_s)$ for each source s
2: Throw away activationsH_s for each source s.
3: Concatenate basis vectors of each source $(W_1; W_2, ...)$ for complete dictionary W
4: Hold W fixed, and factorize unknown mixture of sources V (only estimate H)
5: Once complete, use W and H as before to filter and separate each source.

4. PERFORMANCE EVALUATION OF NMF ALGORITHMS

The principle of the performance measures, SDR, SIR, SAR described in [10] is to decompose a given estimate $\hat{s}(t)$ of a source $s_i(t)$ as a sum

$$\hat{s}(t) = s_{target}(t) + e_{interf}(t) + e_{noise}(t) + e_{artif}(t). \quad (20)$$

where $s_{target}(t)$ is an allowed deformation of the target source $s_i(t)$, $e_{interf}(t)$ is an allowed deformation of the sources which accounts for the interferences of the unwanted sources, $e_{noise}(t)$ is an allowed deformation of the perturbating noise , and $e_{artif}(t)$ is correspond to artifacts of the separation algorithm used in sepration process etc. SDR, SIR, SAR can be computed as

(1) ***Source to Distortion Ratio***

$$SDR = 10log_{10} \frac{\|s_{target}\|^2}{\| + e_{interf} + e_{noise} + e_{artif}\|^2} \quad (21)$$

(2) ***Source to Interference Ratio***

$$SIR = 10log_{10} \frac{\|s_{target}\|^2}{\|e_{interf}\|^2} \quad (22)$$

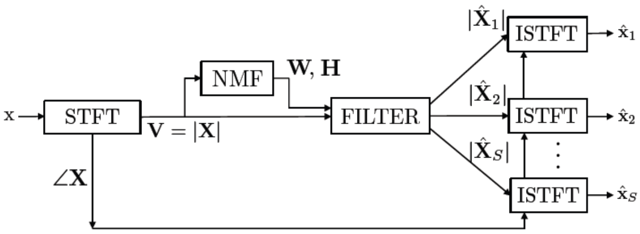

Fig. 1. Basic Source Separation Pipeline

Table 1. Maximum Time Elapsed for Speech+Speech Mixture in sec

K	5	25	50	75	100
REMML	29.8897	38.3651	55.7259	68.1227	87.6488
RISRA	11.2503	17.4523	28.6843	36.9398	56.7109
IS NMF	75.6281	107.2751	112.0757	139.0591	163.2998
SNMF	8.0893	8.9456	9.5673	11.7803	12.1244

(3) ***Source to Artifact Ratio***

$$SAR = 10log_{10}\frac{\|s_{target} + e_{interf} + e_{noise}\|^2}{\|e_{artif}\|^2} \quad (23)$$

5. EXPERIMENTAL RESULTS

In basic source separation pipeline for a complex input mixture the Short Time Fourier Transformed of input mixture is taken place first and then magnitude and phase components are evaluated. After that, NMF decomposition is performed in magnitude spectrogram of the input mixture to split the mixture into its basis and activation vectors. At source synthesis, filtering followed by Inverse Short Time Fourier Transform is performed and the mixture is separated into its individual sources by multiplying the complete basis with each column of activation matrix. Figure 1 shows the general source separation pipeline [7].

From the performance evaluation of each source, it is found that when mixture contain speech or music as underlying source, the SDR, SIR and SAR values are obtained high for $K = 25$. But if the number of underlying sources increases from 2 to 5, the maximum separation is obtained for $K = 50$. Figure1, Figure 2 and Figure 3 gives the performance evaluation values obtained for Speech+Speech, Music+Music and Speech+Music mixture, by varying K from 5 to100, maximum number of iteration=100 . In all cases, as K value varies from 5 to 100, from 5 to 50 an increasing behavior in performance evaluation is obtained and after $K = 50$ the SDR, SIR, SAR values get saturated and then decreased accordingly.

From the evaluation it is found that for mixture containing only two underlying sources, Regularized EMML algorithm performs well for Speech+Speech as well as Music+Speech mixtures. But for Music+Music mixture IS-NMF algorithm is the best. The maximum time elapsed for each of the NMF algorithm can be represented in Table 1, Table II and Table III for Speech+Speech, Music+Music and Music+Speech mixture respectively. From that evaluation SNMF algorithm outperforms the other three algorithms.

Table II and Table III shows the performance evaluation of mixture containing 3 and 5 underlying sources respectively for $K = 50$, maximum number of iteration = 100. When number of underlying sources in the mixture increases from 3 to 5, minimum K value required for accurate source separation in ISRA algorithm (i.e. without the loss of any underlying source from the mixture) is 25 and for Regularized EMML algorithm, is 10. IS-NMF and SNMF can't perform BSS when the mixture contain more than 2 underlying sources. As K value decreases, computation time also get decreases and SNMF algorithm gives more accurate output within the minimum computation time and also with source separation from minimum number of active components. But Regularized EMML algorithm gives more accurate outputs than the other three when number of underlying sources increases in the mixture. So Regularized EMML algorithm is much suitable for BSS in complex

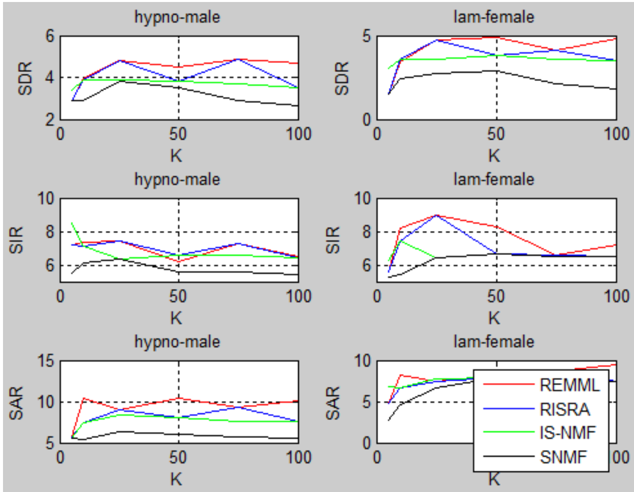

Fig. 2. Speech+Speech mixture

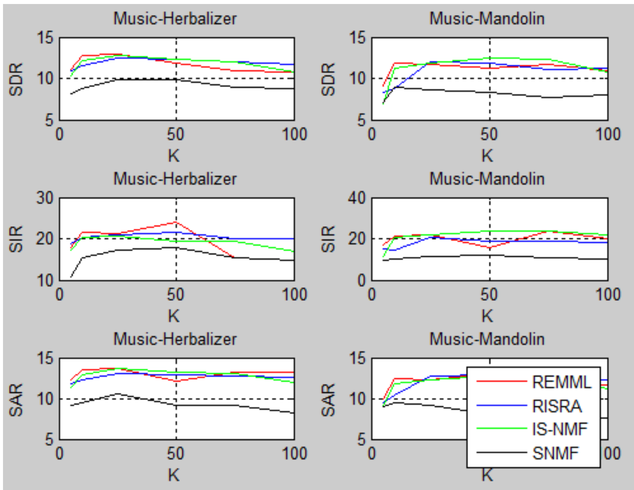

Fig. 3. Music+Music mixture

Table 2. Maximum Time Elapsed for Music+Music Mixture in sec

K	5	25	50	75	100
REMML	30.4729	39.5176	50.3255	65.8112	78.9805
RISRA	11.2231	17.4465	25.5172	40.3055	47.4318
IS NMF	77.8978	73.8568	111.0574	114.1779	157.3814
SNMF	6.5647	7.8456	9.1256	10.1263	12.1123

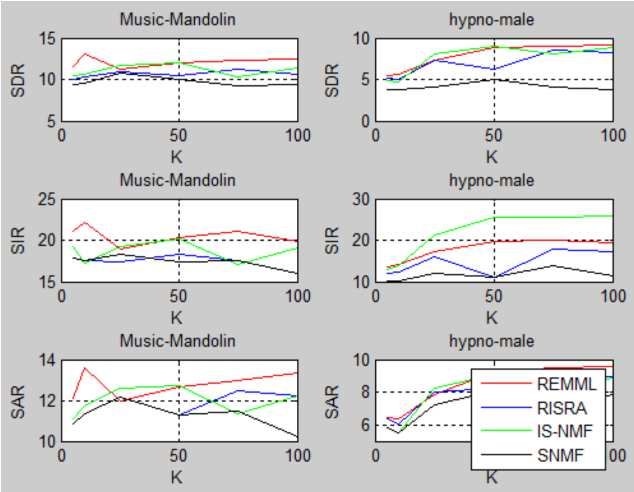

Fig. 4. Music+Speech mixture

Table 3. Maximum Time Elapsed for Music+Speech Mixture in sec

K	5	25	50	75	100
REMML	30.2477	39.3374	46.8076	66.5748	81.5750
RISRA	11.3183	17.0710	36.0914	47.1217	50.7862
IS NMF	75.1075	83.5273	106.4168	144.5353	157.5029
SNMF	5.0895	8.1237	8.9893	10.0456	12.1235

Table 4. Mixture with 3 underlying sources

$K = 50$ $Iter = 100$	EMML			ISRA		
	SDR (dB)	SIR (dB)	SAR (dB)	SDR (dB)	SIR (dB)	SAR (dB)
Music_ Herbalizer	5.1747	9.9017	7.8322	5.7434	12.0216	7.1746
hypno_ male	4.3267	8.0095	7.3924	4.5321	8.4603	7.3639
Music_ Mandolin	7.8322	16.3467	8.5907	7.7669	13.9651	9.1296

mixtures with more than 2 underlying sources than ISRA,IS-NMF and SNMF algorithms. But when number of underlying sources in the mixture increases rapidly there is the possibility of complete loss of underlying sources due to high distortion and high interference from other underlying signals within the mixture, make source separation inaccurate.

6. CONCLUSION

Separation of underdetermined mixtures is an important problem in signal processing that has attracted a great deal of attention over the years. Prior knowledge required to solve such problems is obtained by incorporating complete supervised procedure for source separation using NMF algorithms. From the performance evaluation it is found that as number of underlying sources in the mixture increases possibility of accurate reconstruction get decreases due to the occurrence of traces of underlying sources in separated signal. Even though Regularized EMML algorithm has found higher priority to separate complex mixture than Regularized ISRA, IS-NMF and SNMF algorithms by considering the case of mixture containing 5 underlying sources. When $K = 50$ maximum separation of sources in the mixture take place and minimum value of K required

Table 5. Mixture with 5 underlying sources

$K = 50$ $Iter = 100$	EMML			ISRA		
	SDR (dB)	SIR (dB)	SAR (dB)	SDR (dB)	SIR (dB)	SAR (dB)
Music_ Vivaldi	2.2659	8.9479	3.8364	2.2554	7.1214	4.7398
numbers_ male	2.0547	7.1645	4.4189	2.0151	6.9829	4.4738
Music_ trumpet	0.3497	12.1247	0.9067	0.1768	10.7054	0.9338
Music_ loopbass	6.1852	14.8634	6.9573	6.8925	18.7273	7.2447
hypno_ male	1.4836	5.4004	4.8448	1.4562	5.3789	4.8181

for source separation is found as 2. From these experiments it was shown that Regularized EMML algorithm outperforms the Regularized ISRA,IS-NMF and SNMF algorithms for NMF-based single channel speech and music separation when complexity of the mixture increases. But the computation time of the algorithm is comparatively smaller for SNMF, so mixture with only two underlying sources SNMF outperforms the other three algorithms. Even though all the NMF algorithms are itself easy to implement and compute, makes NMF good for BSS method.

7. REFERENCES

[1] Menaka Rajapakse and Lnnce Wyse, *"NMF vs ICA for Face Recognition"*, Proceedings of the 3rd International Symposium on Image and Signal Processing and Analysis (Proc. ISPAO3), pp 605-610, 2003.

[2] F. Cong, Z. Zhang, I. Kalyakin, T. Huttunen-Scott, H. Lyytinen, and T. Ristaniemi, *"Non-negative Matrix Factorization Vs. FastICA on Mismatch Negativity of Children"*, Proceedings of International Joint Conference on Neural Networks, pp 586-600, June 2009.

[3] Daniel D. Lee and H. Sebastian Seung, *"Algorithms for Non-negative Matrix Factorization"*, Neural Inf. Process. Syst, vol.13, pp 556-562, 2001.

[4] Dempster, AP, Laird, NM Rubin, *"Maximum Likelihood from Incomplete Data via the EM Algorithm"*, J. Royal Stat. Soc, vol.39, 1977.

[5] I. S. Dhillon and S. Sra, *"Generalized nonnegative matrix approximations with Bregman divergences"*, Advances in Neural Information Processing Systems, 19, 2005.

[6] Patrik O. Hoyer, *"Non-negative Matrix Factorization with Sparseness Constraints "*, Journal of Machine Learning Research, vol.5 , pp. 1457-1469, 2004.

[7] Mikkel N. Schmidt, *"Speech Separation Using Non-negative Features and Sparse Non-negative Matrix Factorization"*, Elsevier, 2007.

[8] Dennis L. Sun and Gautham J. Mysore, *"Universal Speech Models for Speaker Independent Single Channel Source Separation"*, ICASSP, 2013.

[9] http://www.telecom.tuc.gr/ ~nikos/BSS_Nikos.html(for obtaining speech/music data base)

[10] Emmanuel Vincent, Rmi Gribonval, and Cdric Fvotte, *"Performance Measurement in Blind Audio Source Separation"*, IEEE Transactions on Audio, Speech, and Language Processing, vol.14, no.4, pp 1462-1469, July 2006.