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 non-gaussian 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 important 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

\[ V \approx WH \]  

(1)

Where \( V \in R_{N \times T}^F \) is the speech spectrogram, \( W \in R_{N \times K}^F \) is the matrix of basis vectors (columns) and \( H \in R_{T \times K}^K \) 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

\[ \min_{W,H} D(V||WH) \]  

(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

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

(3)

To minimize

\[ D(V||WH) = \sum_{i,j} (V_{ij} \log \frac{V_{ij}}{(WH)_{ij}} - V_{ij} + (WH)_{ij}) \]  

(4)

block coordinate descent technique is used.

\[ D(V||WH) = \sum_{i,j} -V_{ij} \log \sum_k W_{ik} H_{kj} + \sum_{i,j,k} W_{ik} H_{kj} \]  

(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 \( f: f(\text{average}) \leq \text{average of } f \).

To apply Jensen’s inequality, weights introduced as \( \pi_{ijk} = 1 \)

\[ D(V||WH) \leq \sum_{i,j,k} (-V_{ij} \log \frac{W_{ik} H_{kj}}{\pi_{ijk}} + \sum_{i,j,k} W_{ik} H_{kj}) \]  

(6)

So the function can be minimized exactly as

\[ H_{kj} = \frac{\sum_i V_{ij} \pi_{ijk}}{\sum_i W_{ik}} \]  

(7)

where \( \pi_{ijk} = \frac{W_{ik} H_{kj}^i}{\sum_k W_{ik} H_{kj}^i} \) so

\[ H_{kj}^{(l+1)} \leftarrow H_{kj}^{(l)} \frac{\sum_i (\frac{V_{ij}}{W_{ij} H_{kj}^i}) w_{ik}}{\sum_i w_{ik}} \]  

(8)

In matrix form it can be represented as

\[ H_{kj}^{(l+1)} \leftarrow H_{kj}^{(l)} W^T \frac{V_{ij}}{W_{ij} H_{kj}^i} \]  

(9)

In similar manner \( W \) can be calculated.

Algorithm 1 Regularized EMML Algorithm

Using \( D(V||WH) = D(V^T||WH^T) \), it obtain a similar update for \( W \).

Now just iterate between

1: Initialize \( W, H \)
2: repeat

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

3: until convergence return \( W, H \).
3.2 Regularized ISRA Algorithm

The squared Euclidean distance (squared Frobenius norm) cost function is given by,
\[ D_F(V \| W H) = \frac{1}{2} \| V - W H \|_F^2 \] (10)

Applying the standard gradient descent technique to the cost function, it gets:
\[ W_{ij} \leftarrow W_{ij} \frac{[V H^T]_{ij}}{[W H H^T]_{ij}} \] (11)
\[ H_{jk} \leftarrow H_{jk} \frac{[W^T V]_{jk}}{[W^T W H]_{jk}} \] (12)

In matrix notation, the Lee-Seung Euclidean multiplicative updates become:
\[ W \leftarrow W \odot \frac{V H^T}{W H H^T} \quad H \leftarrow H \odot \frac{W^T V}{W^T W H} \] (13)

**Algorithm 2 Regularized ISRA Algorithm**

Using \( D_F(V \| W H) = D_F(V^T \| W^T H^T) \), it obtains a similar update for \( W \).

Now just iterate between:
1: Initialize \( W, H \)
2: repeat

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

1: Initialize \( W, H \)
2: repeat

\[ H \leftarrow H \cdot \frac{W^T V}{W^T W H} \quad W \leftarrow W \cdot \frac{V H^T}{W H H^T} \]
3: until convergence return \( W, H \).

3.3 Itakura Saito Divergence Algorithm

Itakura Saito NMF is belongs to the class of Bregman’s divergence where the underlying function is strictly considered as convex in real space. NMF with Itakura Saito Divergence is given by
\[ d_{IS}(V \| W H) = V W H - \log V W H - 1 \] (14)
It is obtained from the maximum likelihood (ML) estimation of short-time speech spectra under autoregressive modeling. The IS divergence belongs to the class of Bregman’s divergences and is a limit case of the \( \beta \) – divergence. Thus, the gradient descent multiplicative rules are applied here. The gradients of criterion \( D_{\beta}(V \| W H) \) wrt \( W \) and \( H \) is represented as
\[ \nabla_W D_{\beta}(V \| W H) = W H^T ((W H)^{\beta-2} \cdot (W H - V)) \] (15)
\[ \nabla_H D_{\beta}(V \| W H) = (W H)^{\beta-2} \cdot (W H - V) H^T \] (16)

where \( \cdot \) denotes Hadamard entry wise product and \( A^n \) denotes the matrix with entries \( A_{ij} \). 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 \( \theta^* \) of the algorithm implies either \( \nabla f(\theta^*) = 0 \) or \( \theta^* = 0 \). This leads to the updates for \( W \) and \( H \),
\[ H \leftarrow H \cdot \frac{W^T ((W H)^{\beta-2} V)}{(W^T W H^{\beta-1}) H^T} \] (17)
\[ W \leftarrow W \cdot \frac{(W H)^{\beta-2} V H^T}{(W^T W H^{\beta-1})} \] (18)
where \( \beta \geq 0 \).

**Algorithm 3 Itakura Saito NMF Algorithm**

Using \( D_F(V \| W H) = D_F(V^T \| W^T H^T) \), it obtains a similar update for \( W \).

Now just iterate between:
1: Initialize \( W, H \)
2: repeat

\[ H \leftarrow H \cdot \frac{W^T V}{W^T W H} \quad W \leftarrow W \cdot \frac{V H^T}{W H H^T} \]
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. Sparse NMF is an extension of NMF, in which an additional sparsity constraint is enforced on either the encoding easy to interpret. So Sparse NMF is an extension of NMF, much of the data using few active components, which makes the sparse NMF problem can be formulated as
\[ \min_{W,H} D(V,WH) + \beta(H) \] (19)
where \( \beta \) 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 \( \beta(H = \lambda \sum_i H_{i,j}) \), where \( \lambda \) 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 \( \beta \) 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 orthonal 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 Sparserness constraints on $W$ apply then
3. Set $W := W - \mu_W (WH - V)^T H^T$
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 $W := W \otimes (VHT) \otimes (WHHT^T)$},
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 Sparserness 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 $H := H \otimes (WTV) \otimes (W^TH)$},
6. until convergence return $W, H$.

where, $\otimes$ and $\odot$ denote element wise multiplication and division, respectively. Moreover, $\mu_W$ and $\mu_H$ 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 Operator Calculation

1: Set $s_i := x_i + (L_1 \sum x_i)/\text{dim}(x), \forall i$
2: Set $Z := \{\}$
3: Iterate
4: Set $m_i := \begin{cases} L_1/(\text{dim}(x)\text{size}(Z)) & \text{if } i \notin Z \\ 0 & \text{if } i \in Z \end{cases}$
5: Set $s := m + \alpha(s - m)$, where $\alpha \geq 0$ 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 $c := (\sum s_i - L_1)/(\text{dim}(x) - \text{size}(Z))$
11: Set $s_i := c, \forall i \notin Z$
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 $(W_sH_s)$ for each source $s$
2: Throw away activations $H_s$ 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

$$\hat{s}(t) = s_{\text{target}}(t) + e_{\text{interf}}(t) + e_{\text{noise}}(t) + e_{\text{artef}}(t).$$

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

(1) Source to Distortion Ratio

$$SDR = 10\log_{10} \frac{\|s_{\text{target}}\|^2}{\|e_{\text{interf}} + e_{\text{noise}} + e_{\text{artef}}\|^2}$$

(2) Source to Interference Ratio

$$SIR = 10\log_{10} \frac{\|s_{\text{target}}\|^2}{\|e_{\text{interf}}\|^2}$$
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 $K = 50$. Figure 1, 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 to 100, 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 Speech+Music 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
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 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


