Underdetermined Blind Separation of an Unknown Number of Sources Based on Fourier Transform and Matrix Factorization

Ossama S. Alshabrawy*, Mohamed E. Ghoneim†
Damietta University, faculty of Science, Department of Mathematics & Computer Science, 34517, Damietta, Egypt
A. A. Salama
Port Said University, faculty of Science, Department of Mathematics & Computer Science, Port Said, Egypt
Aboul Ella Hassanien*
Cairo University, faculty of Computers and Information Technology, Cairo, Egypt

* Scientific Research Group in Egypt (SRGE), http://www.egyptscience.net

Abstract—This paper presents an approach for underdetermined blind source separation that can be applied even if the number of sources is unknown. Moreover, the proposed approach is applicable in the case of separating 1×3 sources from 1 mixtures without additive noise. This situation is more challenging and suitable to practical real world problems. Also, the sparsity conditions are not imposed unlike to those employed by some conventional approaches. Firstly, the number of source signals are estimated followed by the estimation of the mixing matrix based on the use of short time Fourier transform and rough-fuzzy clustering. Then, source signals are normalized and recovered using modified Lin’s projected gradient algorithm with modified Armijo rule. The simulation results show that the proposed approach can separate 1×3 source signals from 1 mixed signals, and it has superior evaluation performance compared to conventional approaches.

Keywords—Underdetermined Blind Source Separation; Rough Fuzzy clustering; Short Time Fourier transform; Lin’s Projected Gradient; Armijo rule

I. INTRODUCTION

BLIND Signal Separation (or Blind Source Separation, BSS) has received a great deal of attention in the fields of digital communication systems, speech processing, medical imaging, water marking, biomedical engineering, and data mining [3]-[7] in recent years in combination with artificial neural networks, information theory, and computer science applications. Blindness or blind separation means that no or very little information is known about the source signals or the mixing system [1].

The objective of BSS is to extract original source signals using only the information gathered from observed signals with no or very limited knowledge about the source signals or the mixing system. The approaches developed by researchers in the last few years can be classified into two methodologies, namely over-determined BSS and underdetermined BSS, according to the number of source signals and observable mixed signals [20]. BSS that has fewer sensors or observable mixed signals than source signals is called underdetermined BSS while a BSS that has more sensors than sources is called over-determined BSS. Underdetermined BSS is challenging and is more realistic in practical situations. However, most approaches for BSS rarely involve underdetermined BSS cases. The classical independent component analysis (ICA) approach fails to solve underdetermined BSS problems [10]. Moreover in many practical problems, there are a large number of source signals but a few numbers of sensors that means the underdetermined case. Another major difficulty of ICA is that the mixing matrix and the magnitude of original source signals cannot be estimated due to its ambiguities and that the order, sign, and the variances of the independent components cannot be determined [2].

Most of the current traditional BSS methods assume that the source signals are as statistically independent as possible given the observed data and that the mixing matrix is of full column rank. In many real-world situations, however, this hypothesis is not valid. Consequently, recovering the source signals by multiplying the observable data matrices by the pseudo inverse of the mixing matrix cannot be used. This makes recovering the source signals a very challenging task [8]. In practical terms, the over-determined mixture assumption does not always hold (e.g., in radio communications the probability of receiving more sources than sensors increases with increase of reception bandwidth), thus it is necessary to solve the problem of underdetermined blind source separation (UBSS) [9].

Nonnegative Matrix Factorization (NMF) has been widely applied to BSS problems. However, the separation results are sensitive to the initialization of parameters, also the additive parts by NMF are not necessarily localized, and consequently the solution is not unique. Avoiding the subjectivity of choosing parameters, we use general matrix factorization (GMF), which completely relaxes the non-negativity constraints from its factors with the Alternative Least Squares (ALS) method as an initialization to the
source signals instead of random initial values. GMF is a
generalization of the well-known NMF where the NMF is
constrained by non-negativity on all its factors, is not
necessarily localize, has low convergence and, does not
provide a unique solution in some cases without additive
constraints and parameters. However, GMF has no
constraints of non-negativity and is fast convergent with the
ALS method used for initialization and improvement.

The motivation of this research is to separate sparse, and
super and sub-Gaussian signals in the underdetermined case
with an unknown number of source signals without resorting
to any sparsity conditions, and to increase the performance
of the separation.

The rest of the paper is organized as follows. Section II
formulates the problem. In Section III, we introduce an
overview, background, and the basic concepts of Projected
Gradient and GMF, alternative least square, and rough fuzzy
clustering. In Section IV, we present the details of the
proposed approach. In section V, we show the analysis of
typical experiments and the results obtained by different
BSS methods, where the simulation results show the
effectiveness and high performance of the proposed
algorithm. Finally, a short conclusion and future work are
presented in Section 6.

II. PROBLEM STATEMENT

The problem considered in this paper is an
underdetermined instantaneous BSS with an unknown
number of source signals but without background noise,
which can be mathematically formulated as follows:

Assume that for I unobservable components \( X(t) = tr[X_1(t), X_2(t), \ldots, X_J(t)] \), where \( J \) is
the number of source signals, and \( X(t) \) is a zero-mean vector.
The available sensor vector \( Y(t) = tr[Y_1(t), Y_2(t), \ldots, Y_J(t)] \), where \( I \) is the number of sensors and \( tr \) is the transpose of
the vector, is given by

\[
Y(t) = AX(t).
\]

Here \( A \in \mathbb{R}^{J \times I} \) is a non-singular and unobservable matrix
and has a non-zero determinant, and the rank of \( A \) is \( I \).
\( X \in \mathbb{R}^{J \times T} \), \( Y \in \mathbb{R}^{R \times T} \). \( t=0, \ldots, T-1 \) are the sampling instant
time points.

III. PRELIMINARY TOPICS

This section provides a brief explanation of the basic
technologies used in this paper including projected gradient
and GMF, alternative least squares, and rough fuzzy
clustering.

A. Projected Gradient and General Matrix Factorization

GMF is a generalization of NMF where there are no
nonnegative constraints on all of the factors [12] and is the
focus of a great deal of attention in Mathematics and
Computer Science. NMF has been widely used in many
areas including BSS [11], [13], [14]. However, the solution
is not unique since NMF is non-convex programming, and
in most algorithms it frequently converges to local optima.
Unlike NMF, GMF is convergent and has good local optima
avoidance when initialized with ALS. In this paper, GMF is
regarded as a good tool for solving the problem of UBSS.
The novelty in this paper is that GMF is to solve the UBSS
problem for the first time.

The basic GMF decomposition model for BSS is as follows:

\[
Y = AX
\]

where, \( Y \in \mathbb{R}^{J \times T} \) represents the observable mixtures,
\( A \in \mathbb{R}^{I \times J} \) is the mixing matrix, and \( X \in \mathbb{R}^{J \times T} \) is the source
signals matrix. Hence, \( Y \), \( A \), and \( X \) have both signs unlike
NMF where \( Y \), \( A \), and \( X \) are non-negative. For BSS, \( I \) is the
number of mixtures or sensors, \( T \) is the number of sample
time points, and \( J \) is the number of sources. With only the
data observable mixtures \( Y \) as the only known variable, the
mixing matrix \( A \) and the source signals \( X \) are estimated
using Equation (2).

We will use the projected gradient based update rules in
GMF. These updates take the following generalized form of
iterative rules [11]:

\[
X^{(n+1)} = X^{(n)} - \alpha_X P_X \tag{2}
\]

\[
A^{(n+1)} = A^{(n)} - \alpha_A P_A \tag{3}
\]

where \( P_A \) and \( P_X \) are the descent directions, and \( \alpha_X \) and \( \alpha_A \)
are the learning rates, of \( A \) and \( X \) respectively.

The projected gradient algorithms for GMF are based on
the alternating minimization technique which can be written
in the matrix form as follows:

\[
\min_{X} \text{Cost} (Y \parallel AX) = \frac{1}{2} \| Y - AX \|^2_F \tag{4}
\]

\[
\min_{X} \text{Cost} (Y^T \parallel X^T A^T) = \frac{1}{2} \| Y^T - X^T A^T \|^2_F \tag{5}
\]

Basically, the matrix \( A \) is assumed to be full rank.
Consequently, this provides the existence of a unique
solution \( X^* \in \mathbb{R}^{J \times T} \). The gradient matrix for \( A \) and \( X \) is
given by the following equations:

\[
\text{Grad}_A (X) = \nabla_X \text{Cost} (Y \parallel AX) = A^T (AX - Y) \tag{6}
\]

\[
\text{Grad}_A (A) = \nabla_A \text{Cost} (Y^T \parallel X^T A^T) = (AX - Y) X^T \tag{7}
\]

One of the projected gradients based approaches, and will
be applied in this paper in a modified version, is Lin's
projected gradient algorithm [15]. Lin's projected gradient
(LPG) algorithm can be induced by the iterative formulas
(2) and (3) with \( P_A \) and \( P_X \) expressed by the equations (6)
and (7). Moreover, the projection on the subspace of non-
negative real numbers is not considered.
B. Alternative least squares (ALS)

The minimization of cost function in equations (4) and (5) which represent the standard squared Euclidean distance can be formulated as follows:

$$\text{Cost}(Y \parallel AX) = \frac{1}{2} ||Y - AX||_F^2$$

$$= \frac{1}{2} tr(Y - AX)(Y - AX)$$

where $tr$ stands for the transpose of the matrix. The above cost function can be alternately minimized with respect to the two factors A and X [11]. Moreover, each time during the optimization process of one factor while keeping the other one fixed, and finding the stationary or critical points, which are obtained by equating the gradients to zero. This corresponds to the following two minimization problems:

$$A^{(k+1)} = \min_A ||Y - AX^{(k)}||_F^2$$

$$X^{(k+1)} = \min_X ||Y^T - X^T(A^{(k+1)})||_F^2$$

The gradients after equating them by zero according to the Karush-Kuhn-Tucker (KKT) optimality conditions are:

$$\frac{\partial D_F(Y \parallel AX)}{\partial a_{ij}} = [YX^T + AXX^T]_{ij} = 0,$$

$$\frac{\partial D_F(Y \parallel AX)}{\partial x_{ij}} = [-AY + A^TAX]_{ij} \forall ij.$$ (10)

Consequently,

$$A = YX^T (XX^T)^{-1} \text{ and, } X = (A^T A)^{-1} A^T Y.$$ (11)

This method will be used as an initialization in our proposed system.

C. Rough fuzzy clustering

In fuzzy c-means (FCM) algorithm developed by Dunn in 1973, improved by Bezdek in 1981, and is the best known method for fuzzy clustering, based on optimizing objective function, the concept of traditional k-means clustering algorithm is extended which for each data point a degree of membership or membership function $\zeta_{ik} \in [0,1]$ of clusters is calculated.

$$\zeta_{ik} = \frac{1}{\sum_{j=1}^c \left( \frac{d_{ik}}{d_{jk}} \right)^{2/\delta -1}}$$

where $\delta$ is the degree of fuzziness.

In contrast to fuzzy clustering, in rough c-means (RCM), the concept of k-means is extended by considering each cluster as an interval or rough set $Y$ [16]. It is characterized by the lower approximation $BY$ and the upper approximations $BY$ with the following properties: (i) an object or a sample $y_k$ can be part of at most one lower approximation; (ii) if $y_k \in BY$ of cluster $X$, then simultaneously $y_k \in BY$; and (iii) if $y_k \notin BY$ is not a part of any lower approximation, then it belongs to two or more upper approximations. This permits overlaps between clusters.

A rough–fuzzy c-means algorithm which involves the integration of fuzzy and rough sets has been developed [17]. This allows incorporating the fuzzy membership value $\zeta_{ik}$ of a sample $y_k$ to cluster center $\beta_j$. Moreover, instead of absolute individual distance $d_{ik}$ from the centroid, the membership to the cluster center of each $\beta_j$ is relative to the other centers $\beta_i \forall i \neq j$. Consequently, the robustness of the clustering will be enhanced with respect to different choices of the parameters. The centroid $\beta_i$ of cluster $U_i$ can be determined by the following equation:

$$\beta_i = \left\{ \begin{array}{ll}
Z & \text{if } BU_i \cup BU_i \neq \emptyset, BU_i \neq \emptyset,
\sum_{y_k \in BU_i \cup BU_i} y_k \quad & \text{if } BU_i \neq \emptyset, BU_i = \emptyset,
\sum_{y_k \in BU_i} y_k \quad & \text{otherwise.}
\end{array} \right.$$

where,

$$Z = W_{\text{upper}} \sum_{y_k \in BU_i} y_k + W_{\text{lower}} \sum_{y_k \in BU_i} y_k$$

The algorithm of rough fuzzy c-means is stated below in Algorithm 1.

**Algorithm 1** Rough fuzzy c-means clustering

Step 1: Assign initial means $\beta_i$ for c clusters.

Step 2: Compute the fuzzy membership $\zeta_{ik}$ for c clusters and N data objects according to equation (12) and Normalize the distances used for fuzzy membership in [0,1].

Step 3: Assign each data object $y_k$ to the lower or upper approximation of cluster pair $U_i$ and $U_j$.

Step 4: Compute the difference $\zeta_{ik} - \zeta_{jk}$ to cluster centroids $\beta_i$ and $\beta_j$.

Step 5: Let $\zeta_{ik}$ be maximum and $\zeta_{jk}$ be the next to maximum.

Step 6: If $abs(\zeta_{ik} - \zeta_{jk})$ is less than some threshold

Then,

$y_k \in \overline{BU_i}$ and $y_k \in \overline{BU_j}$ and $y_k$ cannot be a member of any lower approximation.

Else,
where, $\beta$ is the cluster centers, $c$ is the number of clusters, and $C_{\text{max}}$ is the chosen maximum number of clusters. Here,

$$Scat(c) = \frac{1}{c} \sum_{i=1}^{c} \sigma(\beta_i) \| \| \sigma(Y) \|$$

(17)

Also, the value of $Scat(c)$ varies from 0 to 1. The term that represents the separation between clusters is defined by

$$sep(c) = \frac{D_{\text{min}}^{2} \sum_{i=1}^{c} (\sum_{j=1}^{c} \| \beta_i - \beta_j \|^{-1})}{D_{\text{max}} \max_{i,j} \| \beta_i - \beta_j \|^{-1}}$$

(18)

where,

$$D_{\text{min}} = \min_{i,j} \| \beta_i - \beta_j \|, \quad D_{\text{max}} = \max_{i,j} \| \beta_i - \beta_j \|$$

After clustering, and determining the number of source signals, the $i$th column vector of $A$, denoted as $\tilde{a}_i$, is estimated as

$$\tilde{a}_i = \frac{1}{|\chi_i|} \sum_{(t,r) \in \chi_i} \text{Re} \{ Y^F(t,r) \}$$

(19)

Here $|\chi_i|$ represents the number of TF points in cluster $C_i$ for $i=1,2,...,J$.

Algorithm 2 Mixing estimation and determining the number of source signals

**Input:** the observable mixtures $Y = [Y_1, Y_2, \ldots, Y_J]$

**Output:** number of source signals, the mixing matrix $A$

Step 1: Calculate STFT $Y$ using equations (14)

Step 2: Calculate $\chi_F$ using equation (15)

Step 3: Cluster $\chi_F$ using rough fuzzy $c$-means clustering stated in Algorithm 1 for different number of clusters by choosing $C_{\text{min}}, C_{\text{max}}$ (i.e. min and max chosen number of clusters, respectively) using equations (16)-(18) and the cluster number that minimizes $V$ is considered to be the optimal value for number of source signals.

Step 4: Determine the TF points and their quantity in each cluster.

Step 5: Calculate the columns of the mixing matrix $A$ using equation (19)

**B. Lin's Projected Gradient (LPG) with Armijo rule based GMF**

In Lin’s projected gradient algorithm the learning rates $\alpha_k$ and $\alpha_s$ are not fixed diagonal matrices in the inner iterations but are scalars. These learning rates are computing by inexact estimation techniques. Lin considered two options to estimate the learning rates. The first option is the Armijo rule along the projective arc of the algorithm proposed by Bertsekas [23]. The value of the learning rate $\alpha_s$, for every iterative step of the algorithm, is given by:

$$\alpha_s^{(k)} = \rho^{m^k}$$

(20)

where $m^k$ is the first non-negative integer $m$ for which

$$\text{Cost}(Y \| AX^{(k-1)}) - \text{Cost}(Y \| AX^{(k)}) \leq \nu tr \left\{ \nabla_X \text{Cost}(Y \| AX^{(k)}) (X^{(k-1)} - X^{(k)}) \right\}$$

(21)
The parameter inputs of the modified LPG algorithm are the observable mixtures matrix \( Y \), and the mixing matrix \( A \) obtained from algorithm 2. We choose the maximum number of iterations to be only 25 iterations. We investigate the performance of the proposed UBSS approach in the above mentioned cases by comparing its results with the results of approaches in Khor (2006) [22], Kim and Yoo (2009) [20], Xiang and Peng (2010) [8]. Here, the simulation of the separation of sparse and Gaussian signals is provided followed by some discussion. Then, the cases of a variety of sparse, non-sparse, and super- and sub-Gaussian signals are stated.

### A. Sparse and Gaussian signals

The separation of \( J \) synthetic Gaussian and sparse signals from \( I=3 \) mixtures was performed in the time domain for \( J=4, 5, \) and \( 6 \) source signals. In this simulation, the mixing matrix was estimated using algorithm 2. The proposed approach was compared to the abovementioned algorithms. The simulation settings were as follows. Synthetic sparse signals were generated by generating 5000 Gaussian samples using the \texttt{randn} command of Matlab and substituting 80% of the samples chosen randomly by zeros for each source. The results show that the proposed approach can separate \( I+3 \) source signals from \( I \) mixtures, unlike the previous approaches. This is confirmed in the next simulations. The analysis aims at comparing mainly the reconstruction index Signal-to-Interference Ratio (SIR) to evaluate the performance of the proposed approach. Given original source signals \( X \) and its estimations \( \hat{X} \) obtained by the proposed approach, SIR in decibels is defined as

\[
SIR = -10 \log_{10}(\frac{\|\hat{X}_i - X_i\|^2}{\|X_i\|^2}), \quad i = 1, 2, \ldots, J
\]  

(15)

Fig. 1 illustrates the averaged SIRs when the number of the sources increases from 4 to 6 signals.
B. Synthetic signals

We investigated the effectiveness of the proposed UBSS approach by comparing with the methods mentioned above. We chose the number of mixtures to be only 2 and the number of sources to be 5 in order to prove that the proposed algorithm can separate 1+3 source signals from 1 mixture. The mixture signals that we perform our experiments on are mixed by the following randomly generated mixing matrix:

\[
\begin{bmatrix}
0.5377 & -2.2588 & 0.3188 & -0.4336 & 3.5784 \\
1.8339 & 0.8622 & -1.3077 & 0.3426 & 2.7694
\end{bmatrix}
\]

The mixing matrix was once again estimated using algorithm 2. The true and estimated values of A are shown in Table 1. The six source signals, two observable mixtures, and estimated source signals are plotted in Fig. 2. The number of sampling time points is 10,000. The simulation results of the proposed approach in addition to those of the five different UBSS methods are shown in Fig. 4. The performance of the source recovery method can be evaluated by Eqs. (15) and (16).

\[
SNR = \frac{1}{J} \sum_{j=1}^{J} 10 \log \left( \frac{\| X_j \|_F^2}{\| \hat{X}_j - X_j \|_F^2} \right)
\]  

Where J is the number of sources and \( \| \cdot \|_F \) is the Frobenius norm. The efficiency of the separation results is good when \( SNR \geq 25 \) [10].

We note from Fig. 3 that the proposed approach achieves about 5.7 dB higher SIR for J=5 sources with only two mixtures than the highest performance algorithm among the other three approaches. Another comparison of the proposed approach with the other five approaches is presented using...
the following examples with 2 observable mixtures where $X$ is the chosen source signal shown in Fig. 4.

**Example 1.** $X = \{X_1, X_2, X_3\}$

**Example 2.** $X = \{X_1, X_2, X_3, X_4\}$

**Example 3.** $X = \{X_1, X_2, X_3, X_4, X_5\}$

![Fig. 4. Performance estimation of Examples 1–3 from 2 observable mixtures.](image)

From the results in Figs. 3 and 4, we can conclude that the separation performance of the proposed approach is very high, has faster convergence, and can separate 1+3 source signals from 1 mixed signals.

II. CONCLUSION

In this paper, we addressed the problem of underdetermined blind source separation with the challenging case that the true number of source signals is unknown. A new two-step approach for optimum estimation of the source signals without additive noise. In this approach, STFT is combined with rough fuzzy $c$-means clustering to estimate the mixing matrix and determine the number of source signals. Then the source signals are estimated by a modified LPG algorithm with Armijo rule based general matrix factorization. Simulation experiments demonstrated the validity and superior performance of the proposed approach.

REFERENCES


