Estimating Number of Columns in Mixing Matrix for Under-Determined ICA Using Observed Signal Clustering and Exponential Filtering

Charuwan Saengpratch and Chidchanok Lursinsap

Abstract Under-determined Independent Component Analysis arises in a variety of signal processing applications, including speech processing. In this paper, we proposed a new approach focusing on the estimation of the number of columns and their values of the mixing matrix. The method is based on the observation that the observed vectors must be clustered along the direction of the column vectors of the mixing matrix. A new clustering measure and cluster direction finding are introduced. The propose algorithms are tested with real speech signals and compared with both AICA method and Information Index Removal, Perturbed Mean Shift Algorithm. Our result gives the correct number of columns with higher accuracy under the performance measure of algebraic matrix distance index.

Keywords Independent component analysis · Under-determined ICA · Blind source separation · Linear transformation · Sparse representation

1 Introduction

Knowledge of Independent Component Analysis (ICA) has a great importance for potential speech signal processing, brain signal processing, feature extraction, and acoustic processing. Defined ICA is very closely related to the method called Blind Source Separation (BSS). ICA problem arises in many application domains, such as speech separation, array antenna processing, multi-sensor biomedical records and financial data analysis [1]. Blind Source Separation was using a two stage sparse representation approach and presented an algorithm for estimating the mixing

C. Saengpratch (✉) · C. Lursinsap
Faculty of Science, Chulalongkorn University, Bangkok, Thailand
e-mail: c.saengpratch@hotmail.com

S.-S. Yeo et al. (eds.), *Computer Science and its Applications*, Lecture Notes in Electrical Engineering 203, DOI: 10.1007/978-94-007-5699-1_26,

matrix [2]. The over-complete were introduced that the source separation problem consist of estimating the original sources, the observed signals, then were introduced the over-complete source separation case where there are the mixture less than source signals [3]. A novel BSS algorithm for de-mixing under-determined presented anechoic mixtures [4].

In this paper, we proposed a new method to find the number of columns in the mixing matrix and estimated the value of each element. The remainder of this paper has the following structure. Section 2 describes the problem with constraints. Section 3 focuses on the main idea and the detail of the proposed algorithms. The experimental result is presented in Sect. 4. Section 5 concludes the paper.

2 Problem Statement

Under-determined ICA is considered in this paper. Let $\mathbf{S} = [\mathbf{s_1}, s_2, \ldots, s_N]^T \in R^{m \times T}$ and $\mathbf{X} = [x_1, x_2, \ldots, x_M]^T \in \mathbf{R}^{n \times M}$ be the source signals of n dimensions and observed signals of m dimensions, respectively. Each observed signal x_i is computed by using a mixing matrix $\mathbf{A} = [a_1, a_2, \ldots, a_m] \in \mathbf{R}^{n \times m}$ for n < m as follows.

$$X = AS \tag{1}$$

The source signals are based on the sparseness assumption [5]. The problem of how to recover the source signals after estimating the mixing matrix will not be considered here. However, we adopted the recovering method from [6].

3 Main Concept and Algorithms

Our method to find the columns of mixing matrix A is based on this simple observation. Let $X = [x_{i,1}, x_{i,2}, \ldots, x_{i,3}]^T$ be an observed signal i for 1 <= i <= n in a p-dimensional space. The mixing matrix A maps each incoming signal s_i to a new location considered as an observed signal x_i in lower dimensions. An observed signal can be viewed as a vector. These observed vectors are clustered along each new independent basis represented by each column vector of mixing matrix A. Therefore, to compute each column vector of A, a new similarity measure for all observed vectors clustered at any column must be introduced. Our method consists of the following main steps.

1. Computing the direction of each observed vector x_i in terms of radian degrees.
2. Grouping all observed vectors based on their directions computed from Step 1.
3. Finding the direction of each group and use it as a column vector of the mixing matrix A.

In this paper, we consider only the case where the observed vectors are in two dimensions. Hence, $X_i = [x_{i,1}, x_{i,2}]^T$. The detail of the first two steps is given in the following Algorithm.

3.1 Grouping Observed Vectors Algorithm

Let W be an empty set.
For each x_i, $1 \le i \le n-1$ do

$$\theta_i = \frac{180}{\pi} x \left(\arctan\left(\frac{x_{i,2}}{x_{i,1}} \right) \right)$$

$$\theta_{i+1} = \frac{180}{\pi} x \left(\arctan\left(\frac{x_{i+1,2}}{x_{i+1,1}} \right) \right)$$

If $\left| \frac{\theta_i}{\theta_{i+1}} \right| \le |\theta_i - \theta_{i+1}|$ then

$W = W \cup \{x_i, x_{i+1}\}$.

End
Let k = 1.
Let group = 0.
While W $\neq \varnothing$ do

Let $\mathbf{x}_f^{(W)}$ be the first element in W.

Let θ_{old} be the radian degree of $\mathbf{x}_f^{(W)}$.

Let B_k be a new empty set.

$B_k = B_k \cup \{\mathbf{x}_f^{(W)}\}$.

$W = W - \{\mathbf{x}_f^{(W)}\}$.

For each $x_j \in W$ do

Let θ_{new} be the radian degree of x_j .

If $\left| \frac{\theta_{new}}{\theta_{old}} \right| \ge |\theta_{new} - \theta_{old}|$ then

$B_k = B_k \cup \{x_j\}$.

$W = W - \{x_j\}$.

EndIf
EndFor
k = k + 1.
group = group + 1.
EndWhile

Variable group in the algorithm is for counting the total number of groups. After grouping the observed vectors according to their radian degrees, the actual direction of each group will be derived and used as the column vector in mixing matrix A. Not every group generated by Grouping Observed Vectors Algorithm is feasible enough to derive a column vector in mixing matrix A. Only the group with high density of clustering will be selected. The detail of how to find the actual column of mixing matrix A is the following.

3.2 Finding Actual Column Algorithm

Let Q be an empty set.
For $1 \le i \le group$ do
$t = |B_i|$.
$q = 1 - e^{-\frac{1}{t}}$.
$q' = 1 - q$.
If $q \ge q'$ then
$Q = Q \cup \{B_i\}$.
EndFor
For each element $B_j \in Q$ do
Compute eigenvector of B_j.
EndFor

Steps 4 and 5 are used to measure the density of clustering of each group. The eigenvectors extracted from steps 10 in Finding Actual Column Algorithm above are used as column vectors in mixing matrix A. Figure 1 shows an example as the result of the above algorithms. Figure 1a is the observed vectors. After being grouped by Grouping Observed Vectors Algorithm, Fig. 1b is obtained. The actual column vectors of the mixing matrix are shown in Fig. 1c.

4 The Experimental Results

The algorithms are tested with a real speech as shown in Fig. 2a. There are three sources of signals and two observed signals. The source signals are mixed by using the following mixing matrix.

$$\mathbf{A} = \begin{bmatrix} 0.7071 & -0.4472 & -0.9487 \\ 0.7071 & 0.8944 & 0.3162 \end{bmatrix} \tag{2}$$

The wave form of each observed signal is shown in Fig. 2b. The distribution of both observed vectors are plotted as shown in Fig. 2c. After applying Grouping Observed Vectors Algorithm to all observed vectors, three groups are obtained as shown in Fig. 3a. The direction of each eigenvector in each group is computed and illustrated in Fig. 3b. Our result was compared with the results produced by AICA method [7] and the Information Index Removal and Perturbed Mean Shift Algorithm [8]. To measure the accuracy of the estimated mixing matrix, Algebraic

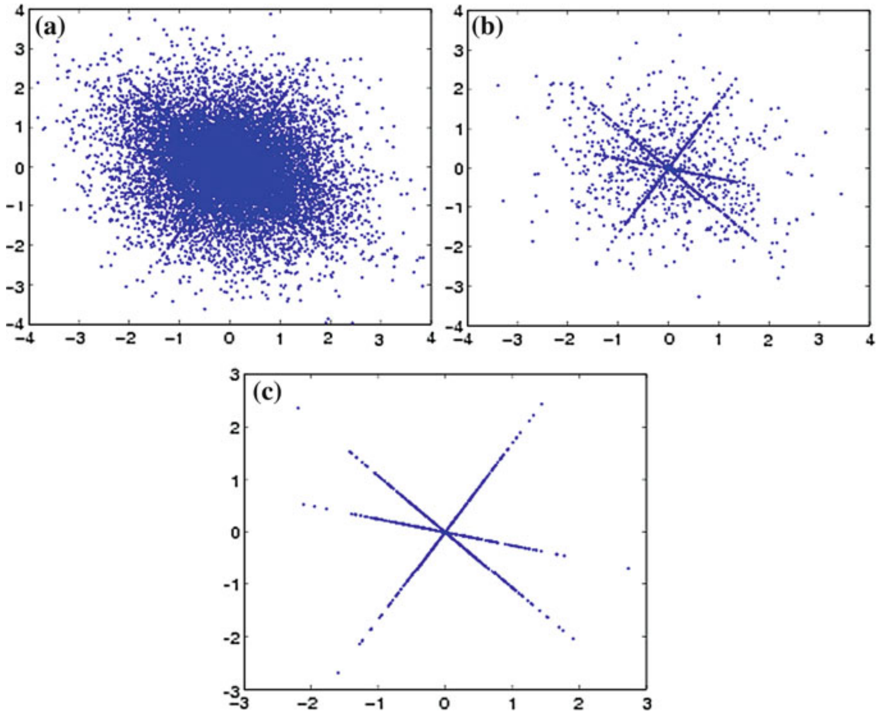

Fig. 1 An example of the result from the proposed algorithm. **a** The given synthetic observed vectors. **b** The result after grouping. **c** The computed eigenvectors which are used as column vectors in the mixing matrix

Matrix-Distance Index (AMDI) [7] was used. The following matrix is the result from our algorithms. Table 1 summarizes the accuracy of each method.

$$\hat{\mathbf{A}} = \begin{bmatrix} -0.9487 & 0.7071 & -0.4472 \\ 0.3162 & 0.7071 & 0.8944 \end{bmatrix} \tag{3}$$

5 Conclusions

This paper presents a new approach for de-mixing mixtures under-determined Independent Component Analysis. This technique is based on the observation that the observed vectors usually clustered along the direction of column vectors in the mixing matrix. Our algorithms can find the correct number of column vectors in the mixing matrix with the lowest AMDI value. However, a further modification must be studied to scope with higher dimensional signals.

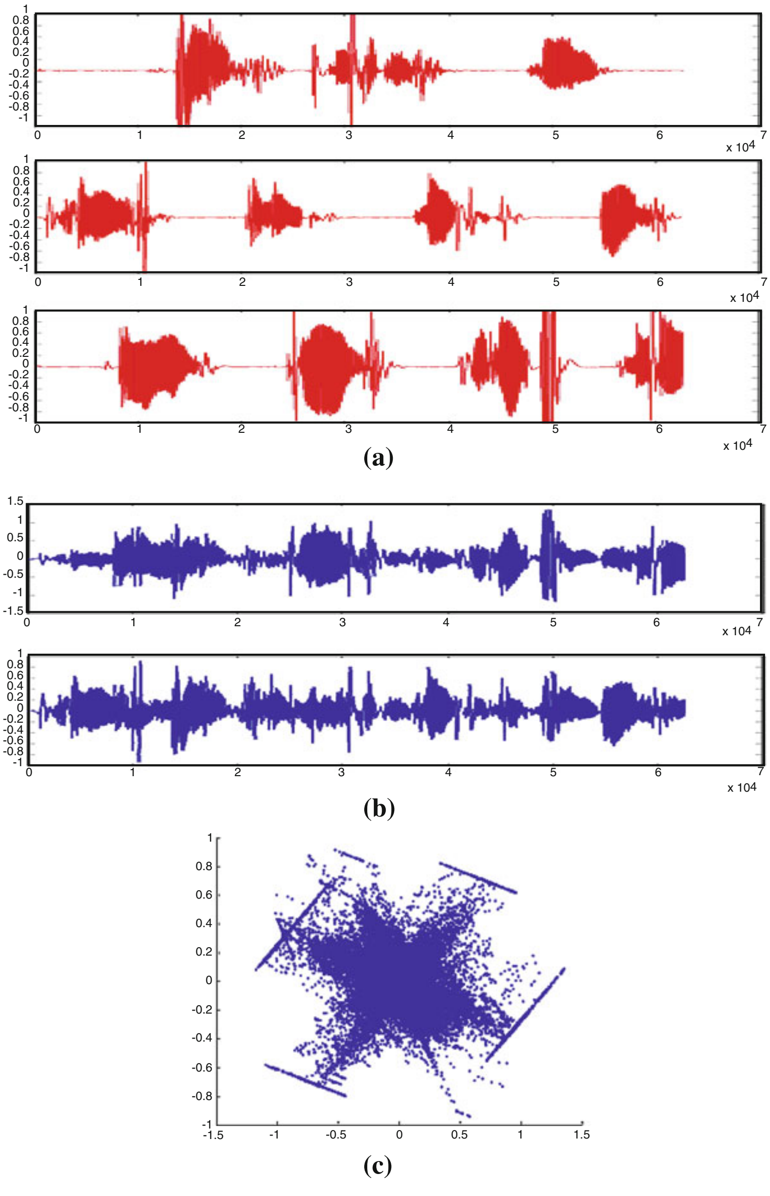

Fig. 2 The tested data. **a** Three given source signals. **b** Two observed signals. **c** The distribution plot of observed vectors

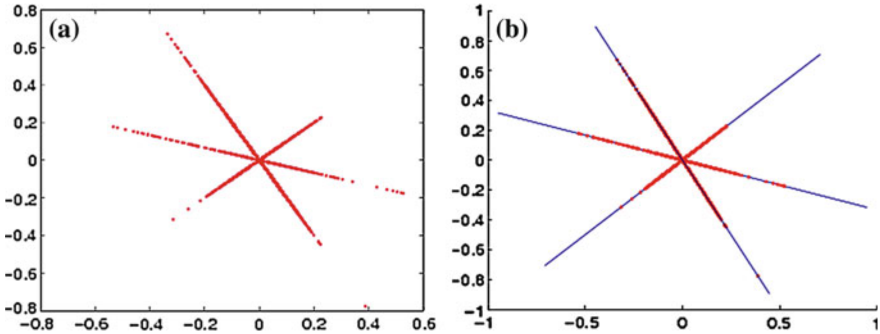

Fig. 3 The result from our algorithm. **a** Groups of observed vectors in the experiment obtained after applying **Grouping Observed Vectors Algorithm**. **b** The direction of each eigenvector in each group

Table 1 The comparison of the estimated mixing matrix by different methods

Comparative methods	Data 2D (n = 3)
Our proposed method	2.1950e-05
The perturbed mean shift	2.6000e-05
The AICA method	<0.001

Acknowledgments The Commission on Higher Education, Thailand for supporting by grant fund under the program Strategic Scholarships for Frontier Research.

References

1. Chinnasarn, K., Lursinsap, C., Palade, V.: Blind separation of mixed kurtosis signed signals using partial observations and low complexity activation functions. Int. J. Comput. Intell. Appl. **4**, 207–223 (2004)
2. Yuanqing, L., Amari, S., Cichocki, A., Daniel, W.C.H., Shengli, X.: Under-determined blind source separation based on sparse representation. IEEE Trans. Signal Process. **54**, 423–437 (2006)
3. Mitianoudis, N., Stathaki, T.: Over-complete source separation using laplacian mixture models. IEEE Signal Process. Lett. **12**, 277–280 (2005)
4. Rayan, S., Yilmaz, O., Martin, J.M., Rafeef, A.: Blind separation of anechoic under-determined speech mixtures using multiple sensors. In: IEEE International Symposium on Signal Processing and Information Technology (2006)
5. Paul, D.O., Barak, A.P.: The LOST algorithm: finding lines and separating speech mixtures. EURASIP J. Adv. Signal Process. **2008**,784296 Hindawi Publishing Corporation (2008)
6. Steven, V.V., Ignacio, S.: A spectral clustering approach to under-determine postnonlinear blind source separation of sparse sources. IEEE Trans. Neural Netw. **17**, 811–814 (2006)

7. Khurram, W., Fathi, M.S.: Algebraic independent component analysis: an approach for separation of over-complete speech mixtures. In: IEEE-INNS Joint International Conference on Neural Networks (2003)
8. Panyangam, B., Chinnasarn, K., Lursinsap, C.: Estimating columns of under-determined mixing matrix by information index removal and perturbed mean shift algorithm. In: IEEE Third International Conference on Natural Computation (2007)