International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:13, No:6, 2019

A Mixing Matrix Estimation Algorithm for Speech Signals under the Under-Determined Blind Source Separation Model

Jing Wu, Wei Lv, Yibing Li, Yuanfan You

Abstract—The separation of speech signals has become a research hotspot in the field of signal processing in recent years. It has many applications and influences in teleconferencing, hearing aids, speech recognition of machines and so on. The sounds received are usually noisy. The issue of identifying the sounds of interest and obtaining clear sounds in such an environment becomes a problem worth exploring, that is, the problem of blind source separation. This paper focuses on the under-determined blind source separation (UBSS). Sparse component analysis is generally used for the problem of under-determined blind source separation. The method is mainly divided into two parts. Firstly, the clustering algorithm is used to estimate the mixing matrix according to the observed signals. Then the signal is separated based on the known mixing matrix. In this paper, the problem of mixing matrix estimation is studied. This paper proposes an improved algorithm to estimate the mixing matrix for speech signals in the UBSS model. The traditional potential algorithm is not accurate for the mixing matrix estimation, especially for low signal-to noise ratio (SNR).In response to this problem, this paper considers the idea of an improved potential function method to estimate the mixing matrix. The algorithm not only avoids the inuence of insufficient prior information in traditional clustering algorithm, but also improves the estimation accuracy of mixing matrix. This paper takes the mixing of four speech signals into two channels as an example. The results of simulations show that the approach in this paper not only improves the accuracy of estimation, but also applies to any mixing matrix.

Keywords—Clustering algorithm, potential function, speech signal, the UBSS model.

I. Introduction

WITH the development of science and the technology, blind source separation is widely used in many fields, such as image processing [1], biomedicine [2], speech signal processing [3], communication [4], industrial machinery [5] and so on. According to the number of observed signals and the number of source signals, blind source separation can be divided into three categories: normal blind source separation (NBSS), overdetermined blind source separation (OBSS) and under-determined blind source separation (UBSS) [6]. When the number of observed signals and source signals is equal, it is NBSS. When the number of the observed signal is more than the source signal, it is OBSS. When the number of the observed signal is less than the source signal, it is UBSS. Methods such as information maximization and independent component analysis (ICA) can be used to solve blind separation problems[7]. The method of ICA has been widely used. This method solves the problems of NBSS and OBSS. But this method does not solve the UBSS problem. Michael and Terrence in [8] proposed the sparsity of the signal for the first time and the blind signal is separated by the sparse characteristic of the signal, which is a pioneering algorithm in the history of under-determined blind separation. This method consists of two steps. The first step is estimating the mixing matrix according to observed signals. The next step is recovering the source signals based on the known mixing matrix. In this paper, the problem of mixing matrix estimation is studied. Mixing matrix estimation problem is divided into two classes. One is that the source signal is completely sparse, the other is the source signal is not completely sparse[9]. This paper discusses the case where the source signal is sufficiently sparse. First, a simple single source point detection method is used to improve the sparsity of signals[10]. Then an improved method is proposed to estimate the number of source signals and the mixing matrix. Many scholars have studied the mixing matrix estimation algorithm. The k-means algorithm is proposed by Li et al in [11] to estimate the mixing matrix. In order to set the number of initial clustering centers, the algorithm needs to know the number of source signals in advance. At the same time, the algorithm is greatly affected by the initial clustering center. Due to these shortcomings of k-means, some scholars have proposed DBSCAN algorithm to estimate the number of source signals and the mixing matrix[12]. However, the algorithm is greatly affected by the clustering radius and the minimum number of clusters. Dong in [13] proposed an improved potential function to estimate the mixing matrix. However, the estimation accuracy of the mixing matrix is not good, especially in the low signal-to-noise ratio. In order to solve the problem, this paper proposes an improved potential function to estimate the mixing matrix. This method improves the estimation accuracy at low SNR and is applicable to any mixing matrix at the same time. The rest of the paper is arranged as follows. Section II introduces the basic model of our method. In Section III, our algorithm is derived. Followed by Section IV, it describes the simulation results and analysis. Finally, conclusions are drawn in Section V.

Jing Wu, Yibing Li, and Yuanfan You are with Information and Communication Engineering Department, Harbin Engineering University, Harbin, China(e-mail: wj9557@ 126.com, liyibing@126.com, China2016080326@hrbeu.edu.cn).

Wei Lv is with China United Network Communication limited, Heilongjiang branch company, China (e-mail: lvweivip@126.com).

II. The Basic Model

We use a simple linear instantaneous mixing model. For convenience, we ignore the effects of noise. Then the data model can be represented as follows.

$$\mathrm{x}(t)=\mathrm{As}(t)=\sum_{i=1}^{M}\mathrm{a}_i\mathrm{s}_i(t) \tag{1}$$

where $\mathrm{x}(t)=[x_1(t),x_2(t),\cdots,x_N(t)]^T$ is a N-dimensional observation signal vector, $\mathrm{A}=[a_1,a_2,\cdots,a_M]$ is a $N\times M$ dimension mixing matrix, $\mathrm{s}(t)=[s_1(t),s_2(t),\cdots,s_M(t)]^T$ is a M-dimensional source signal vector, t is the time sampling point and represents the M-th column vector of the mixing matrix.

III. The Proposed Algorithm

A. Single Source Points Detection

In general, the signal sparsity is not good in time domain, but it is good in time frequency domain, so the signal is usually converted to time frequency domain for processing. Applying STFT on both sides of (1), we can get the expression of time and frequency domain.

$$\mathrm{X}(t,f)=\mathrm{AS}(t,f)=\sum_{i=1}^{M}\mathrm{a}_iS_i(t,f) \tag{2}$$

where $\mathrm{X}(t,f)=[X_1(t,f),\cdots,X_N(t,f)]^T$ and $\mathrm{S}(t,f)=[S_1(t,f),\cdots,S_M(t,f)]^T$ are the STFT coefficients of the observed signal and the source signal at the time frequency point (t,f). In this paper, a two-channel mixing signal is taken as an example. Then (2) can be written.

$$\begin{bmatrix}X_1(t,f)\\X_2(t,f)\end{bmatrix}=\begin{bmatrix}a_{11}\ a_{12}\cdots a_{1M}\\a_{21}\ a_{22}\cdots a_{2M}\end{bmatrix}\begin{bmatrix}S_1(t,f)\\S_2(t,f)\\\vdots\\S_M(t,f)\end{bmatrix} \tag{3}$$

The estimation of mixing matrix mainly depends on the sparsity of signals. For the method of single source point detection, any two column vectors are irrelevant. Because the signal is completely sparse in this paper, there is only one source signal at any TF point. So (3) can be written as:

$$\begin{bmatrix}X_1(t,f)\\X_2(t,f)\end{bmatrix}=\begin{bmatrix}a_{1i}\\a_{2i}\end{bmatrix}S_i(t,f) \tag{4}$$

If there is only one source signal at a time frequency point, there are the following expressions.

$$\begin{aligned}X_1(t,f)&=a_{11}S_1(t,f)\\&=a_{11}\left[\mathrm{Re}\left(S_1(t,f)\right)+j\mathrm{Im}\left(S_1(t,f)\right)\right]\end{aligned} \tag{5}$$

$$\begin{aligned}X_2(t,f)&=a_{21}S_1(t,f)\\&=a_{21}\left[\mathrm{Re}\left(S_1(t,f)\right)+j\mathrm{Im}\left(S_1(t,f)\right)\right]\end{aligned} \tag{6}$$

where Re() and Im() denote the real part and the imaginary part of the complex number, respectively. From (5) and (6), we can get:

$$\frac{\mathrm{Re}\left[X_1(t,f)\right]}{\mathrm{Re}\left[X_2(t,f)\right]}=\frac{\mathrm{Im}\left[X_1(t,f)\right]}{\mathrm{Im}\left[X_2(t,f)\right]} \tag{7}$$

If there are two source signals at a time frequency point, the following expression exists:

$$X_1(t,f)=a_{11}S_1(t,f)+a_{12}S_2(t,f) \tag{8}$$

$$X_2(t,f)=a_{21}S_1(t,f)+a_{22}S_2(t,f) \tag{9}$$

If this situation is still satisfied with (7), it will need to:

$$\frac{a_{11}}{a_{21}}=\frac{a_{12}}{a_{22}} \tag{10}$$

or

$$\frac{\mathrm{Re}\left(S_1\right)}{\mathrm{Re}\left(S_2\right)}=\frac{\mathrm{Im}\left(S_1\right)}{\mathrm{Im}\left(S_2\right)} \tag{11}$$

Since any two columns of the mixing matrix are irrelevant, (10) is not valid. And the probability of (11) is very small. So you can use (7) to detect a single source point. In fact, this condition is very harsh, so we need to relax the condition.

$$\left|\frac{\mathrm{Re}\left[X_1(t,f)\right]}{\mathrm{Re}\left[X_2(t,f)\right]}-\frac{\mathrm{Im}\left[X_1(t,f)\right]}{\mathrm{Im}\left[X_2(t,f)\right]}\right|<\varepsilon_1 \tag{12}$$

where ε_1 is a positive number close to zero.

B. Removal of Low Energy Points and Normalization Treatment

In addition to some noise points, the detected single source points have low energy around the origin, which affect the estimation accuracy of the mixing matrix. In order to improve the estimation accuracy, these low energy points are generally removed. It can be achieved by the following formula:

$$\|X(t,f)\|>\lambda\cdot\max\|X(t,f)\| \tag{13}$$

where the parameter $\lambda\in(0,1)$.

The scatter diagram before removing the low energy points is shown in Fig. 1. Four straight lines can be roughly seen from the diagram. After removing the low energy points, the scatter plot becomes as shown in Fig. 2. The four straight lines can be clearly seen from the diagram. Then, we can estimate the mixing matrix more accurately. It can be seen from Fig. 1 that

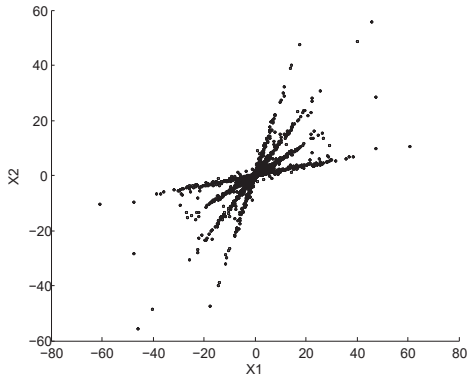

Fig. 1 The scatter diagram before removing the low energy points

the observed signal shows obvious clustering characteristics

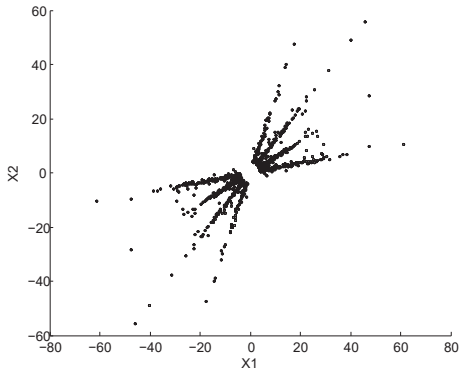

Fig. 2 The scatter diagram after removing the low energy points

after normalized processing, which is convenient for the estimation of mixing matrix.

In order to estimating clustering centers, this paper introduces potential functions.

$$J\left(\mathrm{z_k}\right)=\sum_{i=1}^{T}\left\|\hat{\mathrm{x}}_i\right\|\left[\exp\left(-\frac{\left\|\mathrm{x}_i-\mathrm{z}_k\right\|^2}{b}\right)\right]^{\gamma}\quad k=1,2,\cdots,K \tag{14}$$

In (14), $\mathrm{z_k}$ is the k-th clustering center vector on the hypersphere. K is the number of cluster centers and b is the scale parameter. $\hat{\mathrm{x}}_i$ is a normalized term of x_i. Parameter γ can be obtained by correlation comparison method. The estimation of parameter b is as follows:

$$\hat{b}=\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{m}\left|\tilde{x}_i\left(t\right)-\mu_{\tilde{x}_i}\right|\right) \tag{15}$$

where $\mu_{\tilde{x}_i}$ is the mean value of a mixing signal. Potential function obtains local maximum value at cluster centers. The number of local maximum values is the number of source signals. The vector corresponding to the local maximum value is the column vector of the mixing matrix. In order to obtain good estimation accuracy at low signal-to-noise ratio, a method to improve the potential function is used in this paper. The specific steps are as follows.

- Conduct STFT for observed signals $\mathrm{x}\left(t\right)$ and convert it to $\mathrm{X}\left(t,f\right)$.
- Delete the points in $\mathrm{X}\left(t,f\right)$ according to the single source point detection condition.
- Normalize the remaining points in $\mathrm{X}\left(t,f\right)$.
- Find the points corresponding to local maxima for potential function.
- Use fixed point iteration method to find the clustering center.
- Obtain the distance between the cluster centers and the points detected by single source points.
- Classify each single source point into the nearest class and get the new cluster center by averaging the single source points in each class.
- If the cluster centers obtained by the two iterations are the same, the cluster centers are the estimate of the mixing matrix, otherwise repeat the above steps.

IV. Simulation Results and Analysis

In order to determine the performance of the algorithm, the normalized mean square error (NMSE) is used to evaluate the mixing matrix estimation. The expression is as follows:

$$NMSE=10\log\left(\frac{\sum_{ij}\left(\tilde{a}_{ij}-a_{ij}\right)^2}{\sum_{ij}a_{ij}^2}\right) \tag{16}$$

where $\tilde{a}_{ij}$ is the (i,j) element of the estimated mixing matrix and a_{ij} is the (i,j) element in the original mixing matrix.

In this paper, four speech source signals are mixed into two observed signals as an example. The source signals used in the simulation experiment is from [14]. The original mixing matrix is.

$$\mathrm{A}=\begin{bmatrix}0.3420 & 0.6428 & 0.8660 & 0.9848\\ 0.9397 & 0.7660 & 0.5000 & 0.1736\end{bmatrix} \tag{17}$$

The results of Fig. 3 can be obtained after a single source point detection and normalization process.

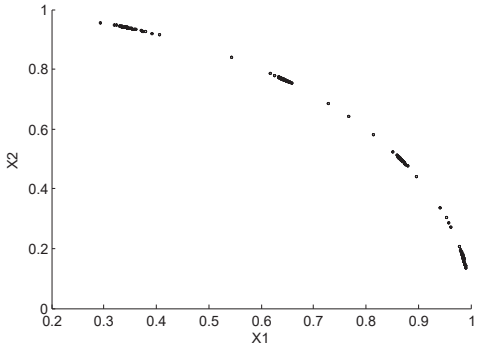

Fig. 3 The scatter diagram normalized

From Fig. 3, we can clearly see that there are four categories, that is , four source signals. After the process of the potential function, we can get the three-dimensional plot of the potential function J(z), as shown in Fig. 4.

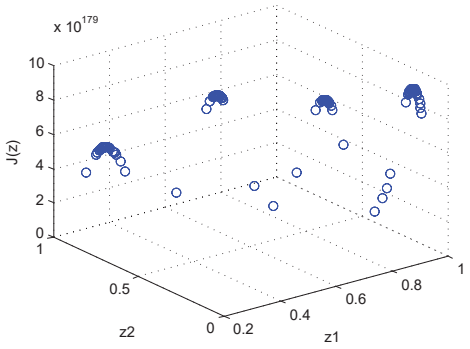

Fig. 4 Three-dimensional plot of J(z)

From Fig. 4, we can clearly see that there are four peaks. We can estimate the number of source signals based on the number of peaks.In order to estimate the mixing matrix, we need to estimate the location of the peak value. Therefore, we transform the three-dimensional plot of J(z) into a two-dimensional plot to better estimate the location of the peak value. The two-dimensional plot of J(z) is shown in Fig. 5.

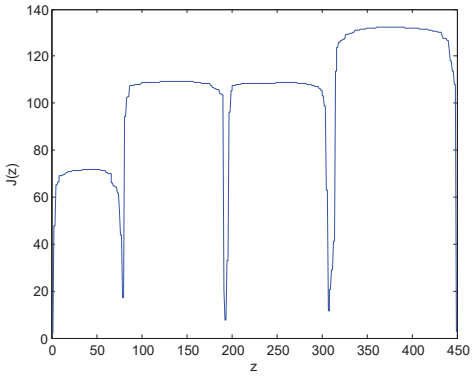

Fig. 5 Two-dimensional plot of J(z)

According to the result of Fig. 5, we can estimate the position of the peak value, but the accuracy of the mixing matrix estimated by this method is not high. In order to improve the estimation accuracy of the mixing matrix, this paper proposes an improved potential function algorithm to estimate the mixing matrix. The mixing matrix $\tilde{A}$ estimated by the algorithm in this paper is as follows:

$$\tilde{A} = \begin{bmatrix} 0.3438 & 0.6415 & 0.8666 & 0.9846 \\ 0.9391 & 0.7671 & 0.4990 & 0.1749 \end{bmatrix} \quad (18)$$

By comparing the original mixing matrix A, it is found that the proposed algorithm is effective. To better illustrate the generality of the algorithm, another set of mixing matrix is used to verify the algorithm. The original mixing matrix A and the estimated mixing matrix $\tilde{A}$ are as follows:

$$A = \begin{bmatrix} 0.1673 & 0.5317 & 0.8499 & 0.9920 \\ 0.9850 & 0.8472 & 0.5177 & 0.1247 \end{bmatrix} \quad (19)$$

$$\tilde{A} = \begin{bmatrix} 0.1716 & 0.5308 & 0.8506 & 0.9915 \\ 0.9847 & 0.8477 & 0.5212 & 0.1289 \end{bmatrix} \quad (20)$$

The experimental results show that the algorithm is effective for other mixing matrices.

Gaussian white noise is used in this paper. The contrast algorithm is the algorithm of Dong in [14]. The parameters of the algorithm are derived from [14]. Fig. 6 shows the average NMSE obtained after 100 Monte Carlo experiments. From Fig. 6, we can see that the estimation accuracy of the proposed algorithm is obviously better than that of the other algorithm in 5-30 dB. The experimental results show the effectiveness of an improved potential function algorithm.

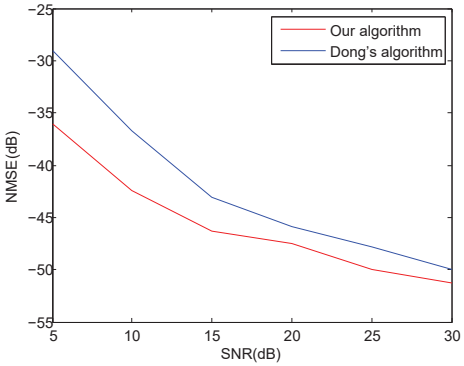

Fig. 6 Comparison between this algorithm and other algorithms in performance

V. Conclusions

This paper presents an improved algorithm to estimate the mixing matrix for speech signals in UBSS model. The estimation accuracy of traditional potential function algorithm is not high, especially in low SNR. Therefore, we consider an improved potential function method to estimate the mixing matrix. The algorithm not only improves the estimation accuracy of mixing matrix, but also applies to other matrices. This paper takes the mixing of four speech signals into two channels as an example. It can be seen from the experimental results that the algorithm can estimate the mixing matrix well without noise. In the case of noise, the algorithm has a good estimation accuracy compared with other algorithms. At the same time, it is found that the algorithm is applicable to other mixing matrices by replacing the mixing matrix.

Acknowledgment

This paper is funded by the National key research and development program of China (Grant No. 2016YFF0102806),the National Natural Science Foundation of China (Grant No. 61701134), the Natural Science Foundation of Heilongjiang Province, China (Grant No. F2017004), the Fundamental Research Funds for the Central Universities of China (HEUCFM180802). And this work is supported by the International Exchange Program of Harbin Engineering University for Innovation-oriented Talents Cultivation.

References

[1] Aziz M A E, Khidr W. Nonnegative matrix factorization based on projected hybrid conjugate gradient algorithm. Signal Image and Video Processing, 2015, 9(8):1825-1831.
[2] Abolghasemi V, Ferdowsi S, Sanei S. Fast and incoherent dictionary learning algorithms with application to fMRI. Signal Image and Video Processing, 2015, 9(1):147-158.
[3] Michael Syskind Pedersen, DeLiang Wang, Jan Larsen, et al. Two-Microphone Separation of Speech Mixtures. IEEE Transactions on Neural Networks, 2008, 19(3):475-492.

[4] Wang X, Huang Z, Zhou Y. Semi-Blind Signal Extraction for Communication Signals by Combining Independent Component Analysis and Spatial Constraints. Sensors, 2012, 12(7):9024-9045.
[5] Wang H, Li R, Tang G, et al. A Compound fault diagnosis for rolling bearings method based on blind source separation and ensemble empirical mode decomposition. Plos One, 2014, 9(10):e109166.
[6] Chen J, Ye F, Jiang T, et al. Conflicting Information Fusion Based on an Improved DS Combination Method. Symmetry, 2017, 9(11):278.
[7] Sun Q, Tian Y, Diao M. Cooperative Localization Algorithm based on Hybrid Topology Architecture for Multiple Mobile Robot System. IEEE Internet of Things Journal, PP(99):1-1
[8] Li Y, Nie W, Ye F, et al. A complex mixing matrix estimation algorithm in under-determined blind source separation problems. Signal Image and Video Processing, 2016:1-8.
[9] Lewicki M S, Sejnowski T J. Learning nonlinear overcomplete representations for efficient coding. Conference on Advances in Neural Information Processing Systems. MIT Press, 1998:556-562.
[10] Wen-Sheng L I, Yi-Bing L I. A new algorithm for spectrum detection in cognitive radio system. Applied Science & Technology, 2011.
[11] Guo Q, Ruan G, Liao Y. A Time-Frequency Domain Underdetermined Blind Source Separation Algorithm for MIMO Radar Signals. Symmetry, 2017, 9(7):104.
[12] Li Y, Cichocki A, Amari S I. Analysis of sparse representation and blind source separation. MIT Press, 2004.
[13] Sun J, Li Y, Wen J, et al. Novel mixing matrix estimation approach in underdetermined blind source separation. Neurocomputing, 2016, 173(P3):623-632.
[14] Dong T, Lei Y, Yang J. An algorithm for underdetermined mixing matrix estimation. Neuro-computing, 2013, 104:26-34.