A NEW TWO-STAGE APPROACH TO UNDERDETERMINED BLIND SOURCE SEPARATION USING SPARSE REPRESENTATION

Wei Zhang, Ju Liu, Jiande Sun and Shuzhong Bai*

School of Information Science and Engineering, Shandong University, Jinan, 250100, China
(*Email: juliu@sdu.edu.cn)

ABSTRACT

In this paper we focus on the two-stage underdetermined blind source separation (BSS), which consists of the mixing matrix estimation stage, the first stage, and the source estimation stage, the second stage. In the first stage, both the mixing matrix and the number of sources are estimated by a new potential-function-based clustering method using a new potential function constructed by Laplacian-like window function. In the second stage, in order to overcome the disadvantage of l^1-norm solution, a new sparse representation based on high-order statistics in transformed domain, which is called statistically sparse component analysis (SSCA), is proposed to recover the sources. Compared with the existing two-stage methods, the proposed approach can achieve higher reconstructed signal-to-noise ratios (SNRs).

Index Terms**—**Underdetermined, sparse representation, two-stage, blind source separation

1. INTRODUCTION

Blind source separation (BSS), which has been widely studied, consists in estimating *N* original sources only from their *M* observed mixtures. In this paper we focus on the underdetermined case, i.e., *M*<*N,* which is a challenging problem. In the case of underdetermined BSS, estimating the mixing system is not sufficient for reconstructing the sources because the mixing matrix is not invertible. So it requires important prior information on the sources, e.g. sparsity, to resolve the underdetermined BSS problem.

Sparse representation of signals has received a great deal of attention in recent years [1-3]. When the sources are sparse, small values are more likely and thus, for a given data point *k*, one of the sources is significantly larger, the remaining ones are likely to be close to zero. The sources can be sparse in the Time-Frequency (TF) domain if a suitable linear transformation is performed, e.g. discrete wavelet packets transformation [4]. Sparse representation has been widely applied in underdetermined BSS.

Several two-stage methods were proposed in underdetermined BSS [4-7]. In the mixing matrix estimation stage, time-frequency-transform-based clustering method [4], k-means clustering method [5], absolute winner-takes-all learning method [6] and potential-function-based method [7] have been proposed. In the source estimation stage, there are l^1-norm solution method [5][8], shortest path decomposition method [6][7], statistically sparse decomposition principle (SSDP) [9], etc. Bofill and Zibulevsky [7] first proposed the shortest path decomposition algorithm which obtained the minimal l^1-norm representation of each data point by a linear combination of the pair of basis vectors that enclose it. In shortest path decomposition algorithm, it is not reasonable that the linear combinations of two basis vectors can't occur when these two vectors are not adjoining. The sources must be sufficiently sparse in l^1-norm solution. In [9], a sparse representation based on minimizing correlation coefficient in a fixed time interval was proposed to solve this problem, but it only estimated the sources in time domain where the sources are not sufficiently sparse.

In this paper, a new two-stage approach to underdetermined BSS based on sparse representation is proposed. Firstly, both the mixing matrix and the number of sources are estimated by a new potential-function-based clustering method using a new potential function constructed by Laplacian-like window function. Secondly, in order to overcome the disadvantage of l^1-norm solution, a new sparse representation based on high-order statistics in transformed domain, which is called statistically sparse component analysis (SSCA), is proposed to estimate the sources. Compared with the existing two-stage methods, better separation performance is obtained in the proposed approach.

2. SYSTEM MODELS AND ASSUMPTIONS

The following noise-free model and noisy model of BSS are considered:

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

$$\mathbf{X}=\mathbf{AS}+\mathbf{N} \tag{2}$$

where $\mathbf{X}=\left[\mathbf{x}(1),\cdots\mathbf{x}(K)\right]\in\Re^{M\times K}$ is the matrix whose rows denote *M* mixtures of sources, $\mathbf{S}=\left[\mathbf{s}(1),\cdots,\mathbf{s}(K)\right]\in\Re^{N\times K}$ is the matrix whose rows denote *N* sources, and

$\mathbf{A} = [\mathbf{a}_1, \cdots, \mathbf{a}_N] \in \Re^{M \times N}$ is the mixing matrix, $\mathbf{N} \in \Re^{M \times K}$ is the noise matrix.

The sparsity of sources plays a key role in the two-stage method. To obtain a sparse representation, discrete wavelet packets transformation is applied to (1), and a transformed model is obtained:

$$\tilde{\mathbf{X}} = \mathbf{A}\tilde{\mathbf{S}} \tag{3}$$

or

$$\tilde{\mathbf{x}}(k) = \mathbf{A}\tilde{\mathbf{s}}(k), \quad k = 1, \cdots, K \tag{4}$$

where each row of $\tilde{\mathbf{X}}$ is composed of the discrete wavelet packets transformation coefficients of a corresponding row of $\mathbf{X}$, each row of $\tilde{\mathbf{S}}$ is the TF representation of the corresponding source in $\mathbf{S}$, $\tilde{\mathbf{x}}(k)$ is the kth column of $\tilde{\mathbf{X}}$, $\tilde{\mathbf{s}}(k)$ is the kth column of $\tilde{\mathbf{S}}$.

The purpose of BSS is to find the solution to the above equations when $\mathbf{A}$ and $\mathbf{S}$ are unknown. It is very difficult to resolve the problem only by the mixture $\mathbf{X}$ without any hypothesis. So the following assumptions should be satisfied in the proposed method:

A1) $M<N$, for convenience, we assume $M=2$, $N>2$;

A2) The sources are statistically mutual independent and sparse in certain degree in transformed domain (the column vector $\tilde{\mathbf{s}}(k)$ has at most M nonzero elements);

A3) The mixing matrix $\mathbf{A}$ is of full row rank, that is, its any $M \times M$ square submatrix is nonsingular.

A4) The noise is additive and independent to the sources.

3. MIXING MATRIX ESTIMATION

In this section, the potential-function-based method [7] is improved.

As shown in Fig.1, because of the sparsity of sources, the scatter plot of $\tilde{\mathbf{X}}$ shows a tendency to cluster along the directions of the basis vectors $\mathbf{a}_j$, which are columns of $\mathbf{A}$. So estimating the cluster directions is equal to estimating that of $\mathbf{a}_j$. Sparsity of sources is often modeled by Laplacian distribution [6]. So a new Laplacian-like window basis function is defined instead of that was used in [7]:

$$\varphi(\alpha) = \exp(-\lambda \,|\, \alpha \,|) \tag{5}$$

So the weighted potential function (WPF) becomes:

$$\Phi(\theta) = \sum_i w_i \varphi(\eta(\theta_i - \theta)) \tag{6}$$

where θ is the assumed cluster center. θ_i is the angle of the scatter plot point i. η is the parameter of scale. w_i is the weight of each scatter plot point, and generally is the square of the modulus of the data.

The number of sources can be determined by the number of local maxima of the resulting function $\Phi(\theta)$, since $\theta \in [0, \pi)$, . The mixing matrix is estimated by the localizations of the local maxima of the WPF. In practice, the range of θ can be divided into P equal spaces. Discrete data of WPF is obtained and WPF obtain its local maxima. Estimating the cluster directions is to localize the peaks of WPF, as shown in Fig.2.

In the proposed algorithm, a threshold of w_i is set: $w_i = 0$ if $w_i < \ell$, which can not only reduce the computational complexity but also improve the accuracy of estimation.

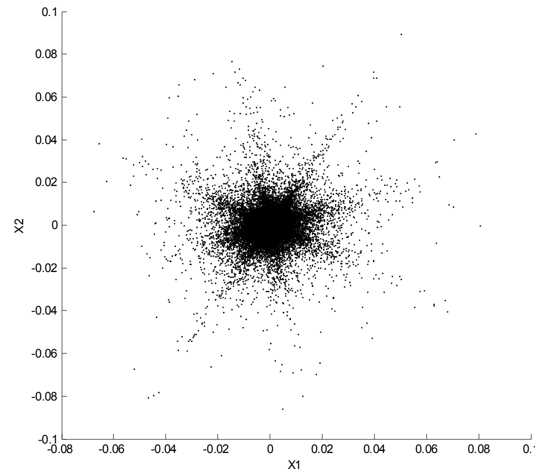


Fig.1 Scatter plot of $\tilde{\mathbf{X}}$, a mixture of four voices.

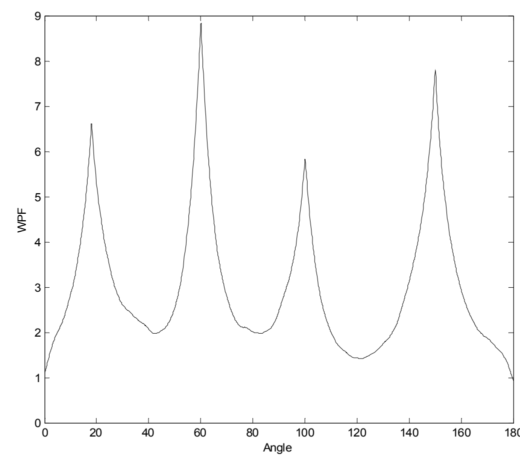


Fig.2 Plot of the WPF

4. SOURCE RECOVERY

Assuming the sources are sparse and $\mathbf{A}$ is given, maximizing a posterior (MAP) method, which is usually used in source estimation in underdetermined BSS, can be solved by the following linear programming problem [1-2]:

$$\min \sum_{k=1}^{K} \sum_{j=1}^{N} |s_j(k)|, \qquad \textit{subject to} \quad \mathbf{AS} = \mathbf{X} \tag{7}$$

Then, l^1-norm $|\mathbf{S}|_1 = \sum_{k=1}^{K} \sum_{j=1}^{N} |s_j(k)|$ can be used to measure the sparsity. The source recovery stage is to minimize the l^1-norm $|\mathbf{S}|_1$ under the constraint $\mathbf{AS} = \mathbf{X}$. If the sources are not sparse enough in time domain, the following linear programming problem in TF domain will be considered:

$$\min \sum_{k=1}^{K} \sum_{j=1}^{N} \left| \tilde{s}_j(k) \right|, \qquad \textit{subject to} \quad \mathbf{A}\tilde{\mathbf{S}} = \tilde{\mathbf{X}} \tag{8}$$

After $\tilde{\mathbf{S}}$ is estimated, $\mathbf{S}$ can be obtained by inverse wavelet packets transformation [4].

In shortest path decomposition algorithm[7], the sources must be sparse enough. And the approach proposed in [9] just estimated the sources in the time domain where the sources are not sufficiently sparse.

To resolve the above problems, a new sparse representation based on high-order statistics in transformed domain, which is called statistically sparse component analysis (SSCA), is proposed to recover the sources in this paper.

A sparse model can be obtained from Eq. (4) and the assumptions in section 2. Given a data point *k*, if the number of the non-zero valued elements in source vector $\tilde{\mathbf{s}}(k)$ is no more than *M*, the model is given:

$$\begin{cases} \tilde{\mathbf{x}}(k) = \mathbf{A}_J \widetilde{\mathbf{s}_J}(k) \\ \tilde{s}_j(k) = 0, \quad j \notin J \end{cases} \tag{9}$$

where $J = \{j_1, j_2, \cdots, j_M\} \subset \{1, 2, \cdots, N\}$; $\mathbf{A}_J = (\mathbf{a}_{j_1}, \cdots, \mathbf{a}_{j_M})$;

$\widetilde{\mathbf{s}_J}(k) = [\tilde{s}_{j_1}(k), \tilde{s}_{j_2}(k), \cdots, \tilde{s}_{j_M}(k)]^{\mathrm{T}}$

Under the assumptions in Section 2, the matrix $\mathbf{A}_J$ is $M\times M$ nonsingular matrix. The following purpose is to obtain the matrix $\mathbf{A}_J$.

First, the data is centered [10]:

$$\tilde{\mathbf{x}}(k) \leftarrow \tilde{\mathbf{x}}(k) - E_{\tilde{\mathbf{x}}}, \tag{10}$$

where $E_{\tilde{\mathbf{x}}}$ is the mean of the data.

Then $\tilde{\mathbf{x}}(k)$ is whitened to obtain the vector $\mathbf{z}$ whose correlation matrix $E\{\mathbf{z}\mathbf{z}^{\mathrm{T}}\}$ is equal to unity [10], as follows:

$$\begin{cases} \mathbf{z} = \mathbf{V}\tilde{\mathbf{x}}(k) \\ \mathbf{V} = \mathbf{E}\mathbf{D}^{1/2}\mathbf{E}^{\mathrm{T}} \\ E\left\{\tilde{\mathbf{x}}(k)(\tilde{\mathbf{x}}(k))^{\mathrm{T}}\right\} = \mathbf{E}\mathbf{D}\mathbf{E}^{\mathrm{T}} \end{cases}, \tag{11}$$

where $\mathbf{V}$ is *M*-by-*M* matrix, $\mathbf{E}$ is the orthogonal matrix of eigenvectors of $E\left\{\tilde{\mathbf{x}}(k)(\tilde{\mathbf{x}}(k))^{\mathrm{T}}\right\}$ and $\mathbf{D}$ is the diagonal matrix of its eigenvalues.

According to (9) and (11), we can obtain:

$$\widetilde{\mathbf{s}_J}(k) = (\mathbf{V}\mathbf{A}_J)^{-1}\mathbf{z}, \tag{12}$$

From ICA algorithms [10], estimating the mutual independent sources is based on the maximization of nongaussianity. So nongaussianity can be used to estimate the components of $\widetilde{\mathbf{s}_J}(k)$. The approximating negentropy is used as the measure of nongaussianity in this paper, and is defined as following:

$$C(y) = [E\{G(y)\} - E\{G(v)\}]^2 \tag{13}$$

where $G(y) = -\exp(-y^2/2)$, and v is a standardized Gaussian variable.

Given a point k and a not-too-short interval Δk , $C(\widetilde{\mathbf{s}_J}(k))$ can be written as:

$$C(\widetilde{\mathbf{s}_J}(k)) = \left\| [\frac{1}{\Delta k} \sum_{k}^{k+\Delta k-1} G((\mathbf{V}\mathbf{A}_J)^{-1}\mathbf{z}) - E\{G(\mathbf{v})\}]^2 \right\| \tag{14}$$

where $\mathbf{v}$ is a vector whose components are standardized Gaussian variables, $\|\cdot\|$ is l^2-norm.

$\hat{J}$, which maximizes $C(\widetilde{\mathbf{s}_J}(k))$, is selected as the estimation of J . Thus, the estimation $\hat{J}$ can be obtained:

$$\hat{J} = \underset{\substack{j_1, j_2, \cdots, j_M \\ =1, 2, \cdots, N}}{\arg} \max C(\widetilde{\mathbf{s}_J}(k)), \tag{15}$$

Finally, the sources in transformed domain can be recovered by:

$$\hat{\tilde{\mathbf{s}}}(k) = \begin{cases} \widetilde{\mathbf{s}_{\hat{J}}}(k) = \mathbf{A}_{\hat{J}}^{-1}\tilde{\mathbf{x}}(k) \\ \tilde{s}_j(k) = 0, \quad j \notin \hat{J} \end{cases} \tag{16}$$

For $k \in [1, K]$, the inverse discrete wavelet packets transformation is applied to obtain the estimation of sources in time domain.

In the proposed approach, that the sources are unnecessary to be the sparsest approximates the truth.

5. SIMULATION AND RESULTS

Two male speeches and two female speeches from TIMIT speech database sampled at 8 kHz with 16 bits resolution were used in our experiments, as shown in Fig.5 (a). The duration of each source is 4 seconds.

5.1. The first stage----mixing matrix recovery

To construct the two mixtures which are shown in Fig.4, a 2×4 mixing matrix is generated by four fixed angles, and every columns is normalized:

$$\mathbf{A} = \begin{bmatrix} \cos 18^\circ & \cos 60^\circ & \cos 100^\circ & \cos 150^\circ \\ \sin 18^\circ & \sin 60^\circ & \sin 100^\circ & \sin 150^\circ \end{bmatrix}$$

$\hat{\mathbf{A}}$ is the estimation of $\mathbf{A}$ estimated by the proposed approach and $\tilde{\mathbf{A}}$ is the one obtained by the method in [7]. Their results in the experiment are:

$$\hat{\mathbf{A}} = \begin{bmatrix} 0.9511 & 0.5000 & -0.1719 & -0.8660 \\ 0.3088 & 0.8660 & 0.9851 & 0.5001 \end{bmatrix}$$

$$\tilde{\mathbf{A}} = \begin{bmatrix} 0.9527 & 0.4969 & -0.1702 & -0.8642 \\ 0.3038 & 0.8678 & 0.9854 & 0.5031 \end{bmatrix}$$

The differences between the mixing matrix and its estimation are presented:

$$\hat{\mathbf{A}} - \mathbf{A} = \begin{bmatrix} 0.0001 & -0.0000 & -0.0017 & 0.0001 \\ -0.0002 & 0.0000 & 0.0003 & 0.0001 \end{bmatrix}$$

$$\tilde{\mathbf{A}} - \mathbf{A} = \begin{bmatrix} 0.0017 & -0.0031 & 0.0035 & 0.0018 \\ -0.0052 & 0.0018 & 0.0006 & 0.0031 \end{bmatrix}$$

In the experiment, the parameters are set as: λ =360, η =0.2, ℓ =0.2× (mean of the data), P=1440.

Compared with the mixing matrix identification method used in [7], the basis function $\varphi(\alpha)$ defined in this paper is better, and the results of estimation are more accurate.

5.2. The second stage----source recovery

Four voices were estimated from two mixtures using the proposed SSCA method presented in Section 4. The fixed data interval Δk is set to 256. $\widehat{\mathbf{A}}$ estimated in the first stage was used. The results estimated by the proposed method are shown in Fig.5 (b).

A reconstruction index is defined as a signal-to-noise ratio [4]:

$$SNR = 10\log\frac{\|\mathbf{S}\|^2}{\|\hat{\mathbf{S}} - \mathbf{S}\|^2}$$

where $\hat{\mathbf{S}}$ is the estimated source, $\mathbf{S}$ is the original one.

The SNR comparison between the shortest path decomposition [7] and the SSDP method proposed [9] are listed in Table 1. It can be seen that the proposed approach is more effective and has better results.

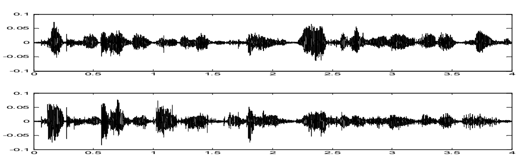
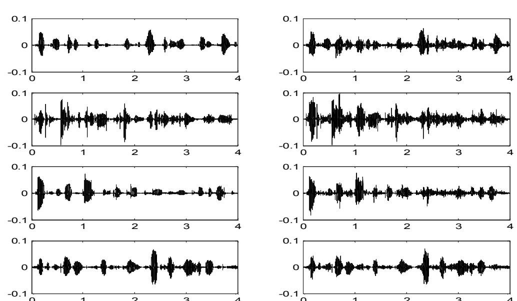


Fig.4 Two mixtures of four sources



(a) (b)

Fig.5 (a) Four speech signals (b) The recovered signals

SNR / Method	S1	S2	S3	S4
l^1-norm	5.4712	6.0278	6.5811	7.0149
SSDP	6.5828	7.5589	8.7182	9.6681
SSCA	**6.9519**	**8.0726**	**9.5967**	**10.6176**

Table 1 The SNRs using different methods

6. CONCLUSION

A new two-stage approach of underdetermined blind source separation is proposed in this paper. In the first stage, a new potential-function-based clustering method is proposed to estimate the mixing matrix. In the second stage, a new sparse representation based on high-order statistics, which is called statistically sparse component analysis (SSCA), is proposed to recover the sources. Simulation results demonstrated that the proposed approach can achieve higher reconstructed signal-to-noise ratios (SNRs).

ACKNOWLEDGMENT

The work is supported by Program for New Century Excellent Talents in University (NCET-05-0582), Specialized Research Fund for the Doctoral Program of Higher Education (Grant No. 20050422017) and the Project sponsored by SRF for ROCS, SEM ([2005]55). The contact author is Ju Liu (juliu@sdu.edu.cn).

REFERENCES

[1] S. Chen, D.L. Donoho, and M.A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput*., vol. 20, no. 1, pp. 33–61, 1998.

[2] M. Zibulevsky and B.A. Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary," *Neural Computation 13*, pp. 863–882, 2001.

[3] M.S. Lewicki and T.J. Sejnowski, "Learning overcomplete representations," *Neural Comput.*, vol. 12, no. 2, pp. 337–365, 2000.

[4] Y. Li, S.I Amari and A. Cichocki, "Underdetermined blind source separation based on sparse representation," *signal Processing*., vol. 54, no. 2, pp. 423–437, 2006.

[5] Y.Q. Li, C. Andrzej and S. Amari. "Analysis of sparse representation and blind source separation," *Neural Computation.* Vol.16, pp. 1193-1234, 2004

[6] F. J. Theis, W. E. Lang and C. G. Puntonet, "A geometric algorithm for overcomplete linear ICA," *Neurocomputing*. Vol.56, pp. 381-398, 2004

[7] P. Bofill and M. Zibulevsky, "Underdetermined blind source separation using sparse representations," *Signal Process*., vol. 81, no. 11, pp. 2353–2362, 2001.

[8] D.L. Donoho and M. Elad, "Maximal sparsity representation via l^1-norm minimization," *Proc. Nat. Aca. Sci*., vol.100, pp. 2197-2202, 2003

[9] M. Xiao, S. Xie and Y. Fu, "A statistically sparse decomposition principle for underdetermined blind source separation." *Intelligent Signal Processing and Communication Systems*, pp. 165-168, 2005

[10] A. Hyvarinen and E. Oja, "Independent component analysis: Algorithms and applications, " *Neural Networks*, vol. 13, pp. 411--430, 2000.