SPARSE COMPONENT ANALYSIS FOR BLIND SOURCE SEPARATION WITH LESS SENSORS THAN SOURCES

*Yuanqing Li, Andrzej Cichocki and Shun-ichi Amari**

Laboratory for Advanced Brain Signal Processing
Laboratory for Mathematical Neuroscience*
RIKEN Brain Science Institute
Wako shi, Saitama, 3510198, Japan

ABSTRACT

A sparse decomposition approach of observed data matrix is presented in this paper and the approach is then used in blind source separation with less sensors than sources. First, sparse representation (factorization) of a data matrix is discussed. For a given basis matrix, there exist infinite coefficient matrices (solutions) generally such that the data matrix can be represented by the product of the basis matrix and coefficient matrices. However, the sparse solution with minimum 1-norm is unique with probability one, and can be obtained by using linear programming algorithm. The basis matrix can be estimated using gradient type algorithm or *K*-means clustering algorithm. Next, blind source separation is discussed based on sparse factorization approach. The blind separation technique includes two steps, one is to estimate a mixing matrix (basis matrix in the sparse representation), the second is to estimate sources (coefficient matrix). If the sources are sufficiently sparse, blind separation can be carried out directly in the time domain. Otherwise, blind separation can be implemented in time-frequency domain after applying wavelet packet transformation preprocessing to the observed mixtures. Three simulation examples are presented to illustrate the proposed algorithms and reveal algorithms performance. Finally, concluding remarks review the developed approach and state the open problems for further studying.

1. INTRODUCTION

Sparse coding or sparse representation of signals, which is modelled by matrix factorization, has been receiving a great deal of interest in recent years [1]-[13]. In [1] was discussed sparse representation of signals by using large scale linear programming under given over-complete basis (e.g., wavelets). In [2], sparse coding of images was discussed based on maximum posterior approach, and the corresponding algorithm was easy to implement, however, no convergence of algorithm was given. In [4, 5, 12], several improved FOCUSS-based algorithms were designed to solve under-determined linear inverse problem when both the dictionary and the sources where unknown. However, only locally optimal solution can be obtained by using this approach and solutions are not unique.

Sparse representation can be used in blind source separation. In several references, the mixing matrix and sources were estimated by using maximum posterior approach and maximum likelihood approach [7, 8, 9]. However, this kinds of algorithms may stick in local minima and they have poor convergence property. Another approach is two steps procedure in which the mixing matrix and the sources are estimated separately [10]. In [10], the blind source separation was performed in frequency domain. A potential-function-based method was presented for estimating the mixing matrix, which was very effective in two dimensional data space. The sources were then estimated by minimizing 1-norm (shortest path separation criterion). In [11] was developed a blind source separation via multi-node sparse representation. Based on several subsets of wavelet packet coefficients, the mixing matrix was estimated by using Fuzzy C-means clustering algorithm, the sources were recovered using the inverse of the estimated mixing matrix. Experiments with two sources and two sensors demonstrated that multi-node sparse representations improved the efficiency of source separation. However, the case of less sensors than sources (including the case in which the number of source was unknown) was not discussed.

This paper first considers sparse representation (factorization) of the observed data matrix based on the following model

$$\mathbf{X} = \mathbf{B}\mathbf{S}, \tag{1}$$

where the $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_N \in R^{n\times N}$ ($N \gg 1$) is a known data matrix, $\mathbf{B} = [\mathbf{b}_1 \cdots \mathbf{b}_m]$ is a $n \times m$ basis matrix, $\mathbf{S} = [s_{ij}]_{m\times N}$ is a coefficient matrix also called a solution corresponding to basis matrix $\mathbf{B}$. Generally, we assume

Correspondence to: Dr. Yuanqing Li, E-mail: liyuan@bsp.brain.riken.go.jp

$m > n$, which implies that the basis is over-complete.

It is well known that there exist many possible solutions of the factorization (1) generally. For a given basis matrix, under the sparsity measure of 1-norm, the uniqueness result of sparse solution is obtained. And number of nonzero entries of the sparse solution can not be reduced generally. We find that the basis matrix of which the column vectors are composed by cluster centers of $\mathbf{X}$ is a sub-optimal basis matrix, which can be obtained using gradient type algorithm or K-Means clustering algorithm.

This kind of sparse representation approach is then used in blind sparse source separation with less sensors than sources and with unknown source number. If the sources are sufficiently sparse, blind separation can be carried out directly in the time domain. Otherwise, blind separation can be implemented in time-frequency domain after wavelet package transformation preprocessing of the mixture signals.

The rest of this paper is organized as follows. Section 2 presents the analysis of data matrix sparse representation based on 1-norm sparsity measure. Section 3 discusses blind source separation with less sensors than sources via sparse representation. Simulation results are shown in Section 4. The concluding remarks in Section 5 summarizes the advantages of the proposed algorithm and states the remaining studying tasks.

2. SPARSE REPRESENTATION OF OBSERVED DATA MATRIX

In this section, sparse representation of data matrix is discussed. Two algorithms are presented, one is for estimating coefficient matrix, another is for estimating basis matrix.

2.1. Linear programming algorithm for estimating coefficient matrix for a given basis matrix

For a given basis matrix $\mathbf{B}$ in (1), the coefficient matrix can be found by maximizing posterior distribution $P(\mathbf{S}|\mathbf{X},\mathbf{B})$ of $\mathbf{S}$ [6]. Under the assumption that the prior is Laplacian, maximizing posterior distribution can be implemented by solving the following optimization problem [1],

$$\min \sum_{i=1}^{m}\sum_{j=1}^{N}|s_{ij}|, \text{ subject to } \mathbf{BS}=\mathbf{X}. \quad (2)$$

Thus, the 1-norm

$$J_1(\mathbf{S})=\sum_{i=1}^{m}\sum_{j=1}^{N}|s_{ij}|, \quad (3)$$

is used as the sparsity measure in this paper.

For sake of convenient discussion, we present a definition.

Definition 1: For a given basis matrix $\mathbf{B}$, denote the set of all solutions of (1) as D. The solution $\mathbf{S}_B = \arg\min_{\mathbf{S}\in D} J_1(\mathbf{S})$ is called a sparse solution with respect to the basis matrix $\mathbf{B}$. The corresponding factorization (1) is said to be a sparse factorization or sparse representation.

For a given basis matrix $\mathbf{B}$, the sparse solution of (1) can be obtained by solving the optimization problem (2). It is not difficult to show that the optimization problem (2) is equivalent to the following N (smaller scale) simultaneous LP problems:

$$\min \sum_{i=1}^{m}|s_{ij}|, \text{ subject to } \mathbf{B}\mathbf{s}_j=\mathbf{x}_j, \quad (4)$$

for $j=1,\cdots,N$.

By setting $\mathbf{S}=\mathbf{U}-\mathbf{V}$, where $\mathbf{U}=[u_{ij}]_{m\times N}\geq 0$, $\mathbf{V}=[v_{ij}]_{m\times N}\geq 0$, (4) can be converted to the following standard LP problems with non-negative constraints,

$$\min \sum_{i=1}^{m}(u_{ij}+v_{ij}),$$
$$\text{subject to } [\mathbf{B},-\mathbf{B}][\mathbf{u}_j^T,\mathbf{v}_j^T]^T=\mathbf{x}_j,\ \mathbf{u}_j\geq 0,\ \mathbf{v}_j\geq 0, \quad (5)$$

for $j=1,\cdots,N$.

Theorem 1 *For almost all basis $\mathbf{B}\in R^{n\times m}$, the sparse solution of (1) is unique. That is, the set of bases matrix $\mathbf{B}$, under which the sparse solution of (1) is not unique, has measure zero in $R^{n\times m}$.*

Due to limit of space we omit all profs in this paper.

Remark 1: Fig. 1 shows in the two dimensional data space, the data vector $\mathbf{x}_i$ can be represented by the basis vectors $\mathbf{b}_2$ and $\mathbf{b}_3$, or $\mathbf{b}_1$, $\mathbf{b}_3$ and $\mathbf{b}_4$, etc. Obvious only the unique sparse representation is based on $\mathbf{b}_2$ and $\mathbf{b}_3$.

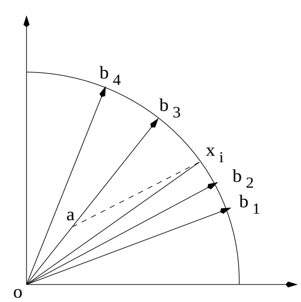

Fig. 1. In two dimensional data space, the representation of $\mathbf{x}_i$ based on unit length vectors $\mathbf{b}_2$ and $\mathbf{b}_3$ is a unique sparse representation

It is interesting to note that in some references, $J_0(\mathbf{S})=\sum_{i=1}^{n}\sum_{j=1}^{N} sign(|s_{ij}|)$ is used as a sparsity measure of $\mathbf{S}$ which is the number of nonzero entries of $\mathbf{S}$. Under this measure, the sparse solution is obtain by solving the problem [12]

$$\min \sum_{i=1}^{m} \sum_{j=1}^{N} sign(|s_{ij}|), \text{ subject to } \mathbf{BS} = \mathbf{X}. \quad (6)$$

Obviously, the solution of (6) is not unique generally, which can be seen easily in Fig. 1. In Fig. 1, every representation of $\mathbf{x}_i$ based on 2 basis vectors is a solution of (6) and $J_0 = 2$.

From Theorem 1, we have the following corollary.

Corollary 1 *Suppose that the number of nonzero entries of the sparse solution* $\mathbf{S}$ *obtained in (2) is* k_0. *If there is a solution* $\mathbf{S}'$ *of (6), such that* $\sum_{i=1}^{m}\sum_{j=1}^{N} |s'_{ij}| \leq \sum_{i=1}^{m}\sum_{j=1}^{N} |s_{ij}|$, *then* $J_0 = k_0$ *with probability one.*

2.2. The algorithm for estimating a sub-optimal basis matrix

To find reasonable over-completed basis of (1) such that the coefficients are as sparse as possible, the following two trivial cases should be removed in advance: 1. the number of basis vector is arbitrary; 2. the norms of basis vectors is unbounded. In the Case 1, we can set the number of basis vector to be that of data vectors, and the basis are composed of data vectors themselves. In the Case 2, if the norms of basis vectors tend to infinity, the coefficients will tend to zeros. Thus we have the two assumptions in this paper beforehand:

Assumption 1: The number of basis vectors m is assumed to be fixed in advance and satisfy that $n \leq m \ll N$.

Assumption 2: All basis vectors are normalized with their $2-$norms being one.

It follows from Theorem 1 that for any given basis, there exists a unique sparse solution of (2) with probability one. Among all basis matrices which satisfy the two assumptions above, there exists at least a basis matrix such that the corresponding solution of (2) is the sparsest. However, it is very difficult to find the best basis. In the following, we will minimize the following objective function to find a sub-optimal basis matrix,

$$\begin{aligned} J'_1 &= \sum_{j=1}^{m} \sum_{\mathbf{x}_i \in \theta'_j} d(\mathbf{x}_i, \mathbf{b}_j) \\ &= \sum_{j=1}^{m} \sum_{\mathbf{x}_i \in \theta'_j} \sqrt{(x_{1i} - b_{1j})^2 + \cdots + (x_{ni} - b_{nj})^2}, \end{aligned} \quad (7)$$

where $\mathbf{x}_i = [x_{1i}, \cdots, x_{ni}]^T$ is normalized data vector, the set θ'_j is composed by all these data vectors $\mathbf{x}_i$'s which satisfy that $d(\mathbf{x}_i, \mathbf{b}_j) = \min\{d(\mathbf{x}_i, \mathbf{b}_k), k = 1, \cdots, m\}$.

That is, by solving the following optimization problem, the sub-optimal basis matrix can be estimated,

$$\min J'_1, \;\; s.t. \; b_{1j}^2 + \cdots + b_{nj}^2 = 1, \quad j = 1, \cdots, m. \quad (8)$$

In fact, the columns of this basis matrix are the cluster centers of the normalized known data vectors.

We can use gradient-type algorithm or $K-$means clustering algorithm to solve the problem (8) and estimate the sub-optimal basis matrix. Thus, we have the following algorithm steps:

Algorithm 1 outline: Estimation of basis matrix

Step 1. Normalize the data vectors,

$$\mathbf{X}' = [\frac{\mathbf{x}_1}{||\mathbf{x}_1||}, \cdots, \frac{\mathbf{x}_N}{||\mathbf{x}_N||}] = [\mathbf{x}'_1, \cdots, \mathbf{x}'_N],$$

where $\mathbf{x}_j, j = 1, \cdots, N$ are column vectors of the data matrix $\mathbf{X}$;

Step 2. Determine starting points of iteration. For the $i = 1, \cdots, n$, find out its maximum and minimum of the i-th row of $\mathbf{X}'$ denoted as M_i and m_i. Take a sufficient large positive integer K, and divide the set $[m_1, M_1] \times \cdots \times [m_n, M_n]$ into K subsets. The centers of the K subsets are used as the initial values;

Step 3. Start $K-$means clustering iteration gradient-type iteration followed by normalization to estimate the sub-optimal basis matrix. Noting that if two basis vectors have opposite directions, only one is taken.

End

3. BLIND SOURCE SEPARATION BASED ON SPARSE REPRESENTATION

In this section, we discuss blind source separation based on sparse representation of observed data matrix $\mathbf{X}$. The proposed approach is also suitable for the case in which the number of sensors is less than the number of sources, as well as the case in which the source number is equal to sensor number.

In this section, the following noise free mixing model is considered,

$$\mathbf{X} = \mathbf{AS}, \quad (9)$$

where the mixing matrix $\mathbf{A} \in R^{n\times m}$ is unknown, the matrix $\mathbf{S} \in R^{m\times N}$ is composed of the m unknown sources, the only observable $\mathbf{X} \in R^{n\times N}$ is a data matrix with its rows being linear mixtures of sources, $n \leq m$. The task of blind source separation is to recover the sources only using the observed data matrix $\mathbf{X}$.

3.1. Blind sparse source separation

If the sources are sufficiently sparse in time domain, then sparse representation approach in this paper can be used directly in blind separation. The process is divided into two steps: the first step is to estimate the mixing matrix using Algorithm 1 presented in the previous section. When the iteration is terminated, the obtained matrix is denoted as $\hat{\mathbf{A}}$ which is the estimation of the mixing matrix $\mathbf{A}$. Although

$\hat{\mathbf{A}}$ has more columns than $\mathbf{A}$ generally, we can see in the simulation examples that the sources can be obtained provided that $\hat{\mathbf{A}}$ contains of all columns of $\mathbf{A}$. After the mixing matrix is estimated, the next is to estimate the source matrix by solving linear programming problem (5).

There is a recoverability problem occurred here, that is, whether the estimated sources are equal to the true sources even if the mixing matrix is estimated correctly.

At first, we have an assumption on the mixing matrix:

Assumption 3: Each group of n column vectors of the mixing matrix $\mathbf{A}$ are linearly independent.

Noting that the large scale linear programming problem (2) is equivalent to the N smaller linear programming problems in (4), we consider the following problem,

$$\min \sum_{i=1}^{m} |s_i|, \text{ subject to } \mathbf{A}\mathbf{s} = \mathbf{x}_0, \tag{10}$$

where the mixture vector $\mathbf{x}_0 = \mathbf{A}\ \mathbf{s}_0$, $\mathbf{A}$ is a known mixing matrix, $\mathbf{s}_0$ is a source vector.

Without loss of generality, suppose that $\mathbf{s}_0$ has p nonzero entries, and denote the estimated source vector in (10) as $\tilde{\mathbf{s}}$. We have the following recoverability result:

Theorem 2 *Under the Assumption 3, if* $\tilde{\mathbf{s}}$ *has minimum number of nonzero entries, and* $p < \frac{n+1}{2}$*, then* $\tilde{\mathbf{s}} = \mathbf{s}_0$ *with probability one.*

Remarks 2: 1. For a fixed n, if p is smaller, then there is more possibility that the estimated source vector is equal to the true source vector. So as for the case of fixed p and larger n. This can be seen in the simulation of Example 1. 2. For the source matrix $\mathbf{S}$ in (9), if the sources are sparse, the column vectors $\mathbf{S}$ have much less nonzero entries than zero entries generally, then the sources can be estimated ideally by using linear programming provided that the mixing matrix can be estimated correctly. At the same time, if the sources are more sparse, it is easier to estimate the mixing matrix according to the analysis of the previous section (the cluster centers of data column vectors are more obvious). Thus the sparseness of sources plays an important role.

3.2. Wavelet Packets transformation preprocessing and blind separation for general sources

Generally, sources can not be recovered directly using sparse factorization if the sources are not sufficiently sparse. In this section, blind separation algorithm based on wavelet packets transformation of mixtures is presented for non-sparse sources.

Algorithm 2 outline: Blind separation for non-sparse sources

Step 1. Apply wavelet packets transformation to mixture signals $\mathbf{x}_i$, where $\mathbf{x}_i$ is the $i-$th row of $\mathbf{X}$, $i = 1, \cdots, n$;

Step 2. Select several wavelet packets tree nodes with coefficients being as sparse as possible. Based on these coefficient vectors, estimate the mixing matrix $\hat{\mathbf{A}}$ using the Algorithm 1 presented in Subsection 2.2. If there exists noise in the mixing model, we can estimate mixing matrix for several times, and then take the mean of these estimated matrixes;

Step 3. Based on the estimated mixing matrix $\hat{\mathbf{A}}$ and the coefficients of all nodes obtained in step 1, estimate the coefficients of all nodes of wavelet packets tree of sources by solving linear programming problems (5);

Step 4. Reconstruct sources using inverse wavelet packets transformation;

Step 5. Perform de-noising of the estimated sources using wavelet transformation if there exists noise in the mixtures.

End.

Remarks 3: 1. We found by simulations experiments that the coefficients of wavelet transformation or Fourier transformation are often not sufficiently sparse to estimate the mixing matrix and sources. Therefore wavelet packets transformation is used in this paper; 2. Sometimes, Steps 3 and 4 can be implemented only based on wavelet transformation coefficients to reduce computation burden.

4. SIMULATION EXAMPLES

Simulation results presented in this section are divided into three categories. Example 1 is concerned with the recoverability of sparse sources by using linear programming method. Example 2 concerned with blind sparse source separation. In Example 3, based on wavelet packets preprocessing and sparse representation, blind source separation of 8 speech sources only using 5 observable mixtures is presented.

Example 1: In this example, consider the mixing model

$$\mathbf{x}_0 = \mathbf{A}\mathbf{s}_0, \tag{11}$$

where $\mathbf{A} \in R^{n\times 15}$ is a known mixing matrix, $\mathbf{s}_0 \in R^{15}$ is a source vector, $\mathbf{x}_0 \in R^n$ is the mixture vector.

Based on the known mixture vector and the mixing matrix, the source vector is estimated by solving the following linear programming problem

$$\min \sum_{i=1}^{15} |s_i|, \ s.t. \ \mathbf{A}\mathbf{s} = \mathbf{x}_0. \tag{12}$$

The task of this example is to check whether the estimated source vector is equal to the true source vector.

There are two simulations in this example. In the first simulation, n is fixed to be 10. 15 loop experiments are completed each of which contain 1000 optimization experiments. In each of the 1000 experiments of the first loop experiment, the mixing matrix $\mathbf{A}$ and the source vector $\mathbf{s}_0$ vector are chosen randomly, however, $\mathbf{s}_0$ has only one nonzero component. After the 1000 optimization experiments, the ratio that the source vectors are estimated correctly is obtained. The k-th loop experiment is carried out similarly but then the source vector has only k nonzero entries.

The first subplot of Fig. 2 shows the curve of the ratios that the source vectors are estimated correctly obtained in 15 different loop experiments. We can see that the source can be estimated correctly when $k = 1, 2, 3$, and the ratio is larger than 0.96 when $k \leq 5$.

Noting that when $k = 5$, the ratio should be closed to 1 from Theorem 2. However it is about 0.96, the reason for this is that the condition in Theorem 2 is not satisfied. Since the mixing matrix $\mathbf{A}$ is chosen randomly in each experiment, so there maybe exist several singular 5×5 submatrices.

The second simulation contains 11 loop experiments. The number k of nonzero entries of the source vectors is fixed to be 5, and the dimension n of mixture vectors changes from 5 to 15. After each loop experiment which also contains 1000 experiments, the ratio that the source vectors are estimated correctly is obtained. The second subplot of Fig. 2 shows the curve of the ratios.

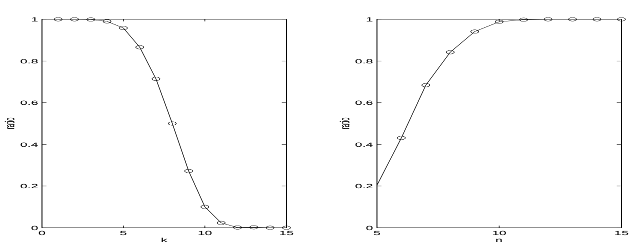

Fig. 2. The first subplot: the curve of ratios that the source vectors are estimated correctly as a function of k obtained in the first simulation of Example 1; the second subplot: the curve of ratios that the source vectors are estimated correctly as a function of n obtained in the second simulation of Example 1.

Example 2: Consider the linear mixing model (9), where source matrix $\mathbf{S}$ is composed by 10 sparse face images, the nonnegative 6×10 mixing matrix is taken randomly of which every column is normalized. In Fig.3, 10 sparse face images are shown in the subplots of the first and second rows, whereas the 6 mixtures are shown in the subplots of the third and fourth rows.

By using the algorithm in Section 3.2, we obtain the separation result shown in Figs. 4. Obviously, the first 10 ten outputs are the recovered sources.

Example 3: Consider the linear mixing model (9), where source matrix $\mathbf{S}$ is composed by 8 speech signals shown in Fig. 5, the 5×8 mixing matrix is taken randomly of which every column is normalized, only 5 mixtures are available

Fig. 3. The ten subplots in the first two rows: the 10 sparse face images; The six subplots in the third and fourth rows: the six mixtures in Example 2

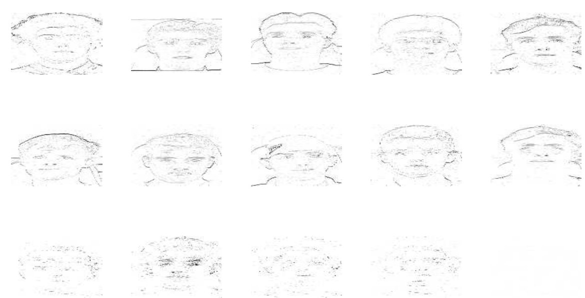

Fig. 4. The fifteen outputs of which the first ten are recovered sources in Example 2

as shown in the subplots of the first row Fig. 5.

By using Algorithm 2, an estimated 5×10 dimensional mixing matrix, and 10 estimated sources are obtained. The subplots in the second and third rows of Fig. 6 show 10 estimated signals, from which we can see that all 8 sources have been recovered quite well, and the other two estimated sources with much smaller amplitudes represent spurious sources, which should be ignored.

5. CONCLUDING REMARKS

Sparse representation of data matrix and blind source separation are discussed with less sensors than sources in this paper. In this paper, 1-norm is used as a sparseness measure. For a given basis matrix, although there exist an infinite number of solutions (factorizations) generally, the sparse solution with minimum 1-norm is proved to be unique with probability one, which can be obtained by using linear programming algorithm. The basis matrix can be estimated by using gradient type algorithm or K-means clustering algorithm. The column vectors of the basis matrix are cluster

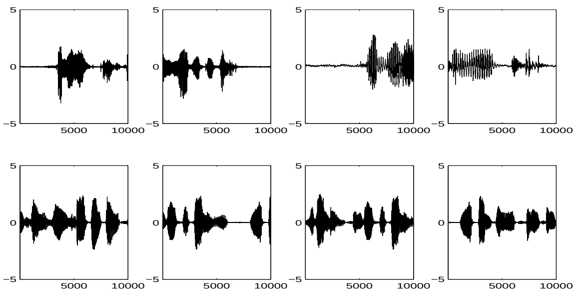

Fig. 5. 8 sources in Example 3

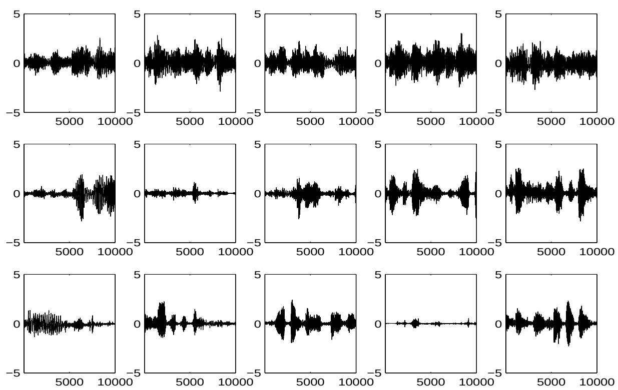

Fig. 6. Blind source separation result in Example 3. The first row: 5 observable mixtures; The second and third rows: estimated signals of which 8 signals are recovered sources, the other two with much smaller amplitudes are spurious sources

centers of normalized data vectors. This sparse representation approach can be used in blind source separation, especially in the case of less sensors than sources. If the sources are sufficiently sparse in the sense of Theorem 2 (i.e., they can be overlapped to some degree), blind separation can be carried out directly in the time domain. Otherwise, wavelet packets transformation preprocessing is necessary, and the blind separation is implemented in time-frequency domain. Two simulation examples illustrate the validity and performance of the proposed algorithms in this paper.

The further work include the algorithm study for estimating the best basis matrix of sparse representation, and application extensions, e.g., applications in image processing and visual computation.

6. REFERENCES

[1] S. Chen, D. L. Donoho, & M. A. Saunders, "Automic decomposition by basis pursuit", *SIAM Journal on Scientific Computing* vol. 20(1), pp. 33-61, 1998.

[2] B. A. Olshausen, "Sparse codes and spikes," In R. P. N. Rao, N. A. Olshausen and M. S. Lewicki (eds.), *Models of Perception and Brain Function*, MIT Press, 2001.

[3] B. A. Olshausen, & D. J. Field, "Sparse coding with an overcomplete basis set: a strategy employed by V1?" *Vision Research*, Vol. 37, pp.3311-3325, 1997.

[4] J. F. Murray, & K. K. Delgado, "An improved FOCUSS-based learning algorithm for solving blind sparse linear inverse problems," *Conference Record of the 35rd Asilomar Conference on Signals, Systems and Computers (IEEE)*,2001.

[5] K. K. Delgado, J. F. Murray, B. D. Rao, K. Engan, T. W. Lee, & T. J. Sejnowski, "Dictionary Learning Algorithms for sparse representation," *Neural Computation* (to appear).

[6] M. S. Lewicki, & T. J. Sejnowski, "Learning overcomplete representations," *Neural Computation* Vol. 12(2), pp. 337-365, 2000.

[7] M. Zibulevsky, & B. A. Pearlmutter, "Blind Source Separation by Sparse Decomposition," *Neural Computations*, vol. 13(4), 2001.

[8] M. Zibulevsky, B. A. Pearlmutter, P. Boll, & P. Kisilev, "Blind source separation by sparse decomposition in a signal dictionary," In Roberts, S. J. and Everson, R. M., editors, *Independent Components Analysis: Princeiples and Practice*, Cambridge University Press, 2000.

[9] T. W. Lee, M. S. Lewicki, & M. Girolami & T. J. Sejnowski, "Blind source separation of more sources than mixtures using overcomplete representations," *IEEE Signal Processing Letter*, vol.6, no. 4, pp. 87-90, 1999

[10] Bofill P., & M. Zibulevsky, "Underdetermined Blind Source Separation using Sparse Representations," *Signal Processing*, Vol.81, No 11, pp.2353-2362, 2001.

[11] M. Zibulevsky, P. Kisilev, Y.Y. Zeevi, & B.A. Pearlmutter, "Blind source separation via multinode sparse representation," *NIPS-2001*.

[12] A. Cichocki & S. Amari, *Adaptive Blind Signal and Image Processing*, Wiley, Chichester, 2002.

[13] P.O. Hoyer, "Non-negative sparse coding", Proc. Workshop on Neural Network for Signal Processing, (NNSP-2002), Martigny, Switzerland, pp.557-565, Sept. 2002