Wavelet Packet Transform-Based Algorithm for Mixing Matrix Estimation

Yujie Zhang, Huiming Peng, Hongwei Li*
School of Mathematics and Physics, China University of Geosciences, Wuhan 430074, China
Email: zhangyujie@cug.edu.cn, phm821@cug.edu.cn, hwli@cug.edu.cn

***Abstract*—The sparsity of signals in their transformed domain is widely used for under-determined blind source separation. The most challenging task of under-determined BSS is to estimate the mixing matrix. In this paper, a new cost function is proposed to detect the sparsest sub-band. Samples in the sub-band can be used to estimate the mixing matrix. Finally some numerical experiments are performed to evaluate the effectiveness of the proposed algorithm.**

***Index Terms*—under-determined blind source separation, sparse component separation, wavelet packet transform, k-means clustering**

I. Introduction

Blind source separation (BSS) is the process of separating a set of unknown original signals from their mixtures without any knowledge about the mixing process or the signals [1-2]. The linear BSS problem can be described as:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t), \quad t = 1, 2, \cdots, N, \tag{1}$$

where $\mathbf{x}(t) = [x_1(t), x_2(t), \cdots, x_m(t)]^T$ are the observed signals and $\mathbf{s}(t) = [s_1(t), s_2(t), \cdots, s_n(t)]^T$ are the unknown original source signals. t is the discrete time sequence and T is the transpose operator. $\mathbf{A} = (a_{ij})_{m\times n}$ is the unknown mixture matrix with full row rank. The main objective of BSS is to estimate the matrix $\mathbf{A}$ and the source vectors $\mathbf{s}$. In the last 20 years, a lot of algorithms were proposed for this purpose. In the cases of over-determined and determined, i.e. the number of sources is less than or equal to the number of mixtures, independent component analysis (ICA) method was widely used [1-2]. However, in practical application the number of sources is usually more than the number of observed sources, which is called under-determined BSS. When mixing is under-determined, the performance of the ICA method was poor and instead sparse component analysis (SCA) was used [3-6]. Sparsity arises in many practical problems, e.g. unmixing spectral data [7], feature extraction and image processing [8], music transcription [9], etc. A signal is said to be sparse in the temporal domain only when a few number of sources have significant values and most of them are almost zero at each time point [10]. Even if the sources are non-sparse in time domain, they may be sparse in another linear transformed domain. For example, acoustic signals may not be sparse enough in time domain, but they are sparse in time-frequency domains [11], and image signals have strong low frequency components and few edges, but their wavelet coefficients are sparse [11]. Hence if we transform the time domain signal into the transformed domains, the sparsity can be utilized to separate the signals from their mixtures.

By making additional assumptions on the source signals, several time-frequency (TF) algorithms are proposed to solve the SCA problems [12-14]. In [13], it divides the TF methods into three sets. And all the approaches presented in the papers which involve in [8] use the same TF transforms i.e. the Short-Time Fourier Transform (STFT). However, the STFT includes the choice of the window functions and overlap length. In order to overcome these restrictions, wavelet packet transform has been proposed for BSS to instead of STFT, which allows isolation of the fine details within each decomposition level and enables adaptive subband decomposition [14]. In [14], it acquires the mixed matrix by searching the independent sub-components based on approximation of the mutual information (MI) with the assumption of wide-band source signals are dependent and some of their sub-components are independent. This assumption may help to solve some practical BSS problems, but it does not necessarily hold. In this paper, we assume that the mixing model is expressed as model (1) in which the number of sources is equal to or larger than the number of observations (i.e. $m \le n$), in addition, all of the source signals are sparse in at least one frequency sub-band. Without loss of generality, it is further assumed that the columns of $\mathbf{A}$ are normalized [3-4], i.e. $\sum_{i=1}^{m} a_{ij}^2 = 1$, $j = 1, \cdots, n$.

The SCA problem is usually solved in two steps. The first step is the estimation of the mixing matrix, and the second step is the recovery of the source signals by knowing the mixing matrix. Note that in the under-determined case, knowing the mixing matrix does not directly result in the recovery of the sources [11]. In this paper, we only consider the problem of the estimation of the mixing matrix. The main contribution of this paper is

* Corresponding author

doi:10.4304/jcp.7.11.2605-2611

the efficient algorithm proposed in section 3 for the detection of sparse sub-band which can be used for the mixing matrix estimation. We propose an approach to get the sparse sub-band, which is based on multiresolution decomposition using wavelet packet (WP) based iterative filter banks [14]. In order to enable the filter bank to adaptively select the sub-band with the sparsest components of the source signals, we have introduced a criterion based on the minimum sums of within-cluster average distance (*WCAD*). To state the idea more precisely, firstly, we introduce a WP based approach to BSS obtaining adaptive sub-band decomposition of the observed sources. Secondly, we introduce the minimum sums of *WCAD* in each sub-band as a criterion for the selection of the sub-band with the sparsest components. Then, the sub-band so obtained is clustered using the k-means clustering algorithm and principal component analysis (PCA) for the mixing matrix.

The rest of the paper is organized as follows. In section 2, we describe the problem formulation and compare several time-frequency translations. The algorithm for estimating the mixing matrix is proposed in section 3. In section 4, we present some simulation results and finally the conclusion is drawn in Section 5.

II. PROBLEM FORMULATION

A. SCA Model

For the SCA, model (1) can be written as

$$\mathbf{x}(t)=\begin{pmatrix}a_{11}\\ \vdots\\ a_{m1}\end{pmatrix}s_1(t)+\cdots+\begin{pmatrix}a_{1n}\\ \vdots\\ a_{mn}\end{pmatrix}s_n(t),\quad t=1,2,\cdots,N. \qquad (2)$$

For a fixed time point *t*, if only the *i*-th source is active, i.e. $s_i(t)\neq 0$ and $s_j(t)=0,\ j\neq i,\ j=1,2,\cdots,N$. Then model (2) can be reduced to the following formula,

$$\begin{pmatrix}x_1(t)\\ \vdots\\ x_m(t)\end{pmatrix}=\begin{pmatrix}a_{1i}\\ \vdots\\ a_{mi}\end{pmatrix}s_i(t). \qquad (3)$$

Obviously, formula (3) is a linear equation, which means that all the columns of mixing matrix $\mathbf{A}$ are the hyperline directions in the scatter plot of the observed data $\mathbf{x}$. In other words, for the estimation of mixing matrix $\mathbf{A}$, we have to find these directions by solving the adaptive clustering problem.

The essence of the sparse approach is the identification of line orientations vectors from the observed signals. And the higher sparsity is a requirement for good estimation of mixing matrix. In this situation a possible solution is to look for a linear transform T such that the new representation of the data is sparser [15]. The transformation being linear, the mixing matrix is preserved and model (1) can be rewritten as

$$T(\mathbf{x})=\mathbf{A}T(\mathbf{s}). \qquad (4)$$

The key is how to choose the linear transform T . For this purpose, STFT has been used in [3-4, 15-16]. Although the signals can be sparser by using STFT, it involves the choice of the window function and overlap length. In [11], it was said that the Hanning window performed better than the Hamming window or the square window. As in [16], with the same source signals and the same mixing matrix $\mathbf{A}$, the difference in overlap length may lead to the error in mixing matrix estimation.

As a consequence, we choose wavelet packet transform (WPT) instead of STFT. The main reason is due to the existence of WPT in a form of iterative filter bank and the multiresolution property of WPT which allows isolation of the fine details within each decomposition level and enables the adaptive sub-band decomposition [14]. It means that a sparse sub-band which is arbitrarily narrow can be isolated by progressing to the higher decomposition levels.

B. Multiresolution Wavelet Packet Decomposition

We express each source signal in terms of its decomposition coefficients [14]:

$$s_{ki}^{j}(t)=\sum_{\alpha}c_{ki\alpha}^{j}\varphi_{j\alpha}(t), \qquad (5)$$

where the indexes j,k,i,α represent the scale level, the sub-band index, the source index and the shift index, respectively, herein $k=1,2,\cdots,2^j$. $\varphi_j(t)$ is the chosen wavelet and $c_{ki\alpha}^{j}$ are the corresponding decomposition coefficients. If we choose the same representation space as for the source signals, each component of the observed data $\mathbf{x}$ can be written as

$$x_{kp}^{j}(t)=\sum_{\alpha}d_{kp\alpha}^{j}\varphi_{j\alpha}(t), \qquad (6)$$

where p is the observed signal index. Let vectors $\mathbf{c}_{k\alpha}^{j}=[c_{k1\alpha}^{j},c_{k2\alpha}^{j},\cdots,c_{kn\alpha}^{j}]^T$ and $\mathbf{d}_{k\alpha}^{j}=[d_{k1\alpha}^{j},d_{k2\alpha}^{j},\cdots,d_{km\alpha}^{j}]^T$ be constructed from the *l*-th coefficients of the sources and mixtures, respectively. From (1) and (6) using the orthogonally property of the functions $\varphi_j(t)$, we obtain

$$\mathbf{d}_{k\alpha}^{j}=\mathbf{A}\mathbf{c}_{k\alpha}^{j}. \qquad (7)$$

Note that the relation between decomposition coefficients of the mixtures and the sources is exactly the same as in the original domain of signals. The estimation of the mixing matrix is performed using the decomposition coefficients $\mathbf{y}_k^j(t)=[d_{k1}^j,d_{k2}^j\cdots,d_{km}^j]^T$ of the mixtures.

III. PROPOSED APPROACH

At *j*-th scale level, in each sub-band k , to reduce the effect of outliers, we first remove columns $\mathbf{y}_k^j(t)$ from the matrix $\mathbf{y}_k^j$ that are close to origin (i.e., their norms are below a small threshold). Most of them are generated from the source signal whose components are close to zero. Then, we normalize $\mathbf{y}_k^j(t)$ to unit length in the preprocessing stage by the following formula:

$$\mathbf{y}_k^j(t)=\mathbf{y}_k^j(t)/\left\|\mathbf{y}_k^j(t)\right\|. \qquad (8)$$

Next, mirror the directions. That is, the points lying in the under side of the horizontal axis in the scatter diagram are mapped to the upper side by changing their sign:

$$\mathbf{y}_k^j(t) = \mathbf{y}_k^j(t) \times sign(\mathbf{y}_k^j(t)). \quad (9)$$

Through processing the sub-band sources, which is said standardization, all of the data are projected to the upper unit semicircle or unit sphere or hyper unit semisphere. Each line corresponds to one category on the unit semicircle.

Then, we divide $\mathbf{y}_k^j(t)$ into n categories using k-means clustering, where $\mathbf{y}_{ki}^j(t)$ is the i-th cluster and $\bar{\mathbf{y}}_{ki}^j$ is the mean of the corresponding cluster. The within-cluster sum of point-to-centroid distance can be written as

$$D_k^j(i) = \frac{\sum d(\mathbf{y}_{ki}^j(t), \bar{\mathbf{y}}_{ki}^j)}{l_{ki}^j}, \ i = 1, 2 \cdots, n. \quad (10)$$

where l_{ki}^j is sample size of the i-th cluster. The distance values can be calculated using the l^2-norm. After that calculate the sum of within-cluster average distance ($WCAD$):

$$WCAD_k^j = \sum_{i=1}^{n} \frac{D_k^j(i)}{n}. \quad (11)$$

Note that, the more obvious the straight line is, the smaller the distance $WCAD_k^j$ will be.

Once we succeed in obtaining the most sparse sub-band q, the n cluster centers in this sub-band correspond to the column vectors of the mixing matrix $\mathbf{A}$ up to permutation and scaling. In order to accurately estimate the mixing matrix $\mathbf{A}$, calculate the first eigenvector and its corresponding largest eigenvalue of the matrix $\mathbf{y}_{kq}^j \cdot \left(\mathbf{y}_{kq}^j\right)^T / l_{kq}^j$. This eigenvector is chosen as one column of the estimation of the mixing matrix.

The error in mixing matrix estimation can be further decreased by removing the points which are away from the center of the cluster. Here we remove those points by $c\sigma_{kqi}^j$, where c is a constant and σ_{kqi}^j is the standard deviation of the samples in the i-th cluster. In other words, the h-th sample in the i-th cluster is removed if $\left|\mathbf{y}_{kqi}^j(h) - \mu_{kqi}^j\right| > c\sigma_{kqi}^j$, where μ_{kqi}^j is the center of the i-th cluster.

WP algorithm for estimating A can be outlined as follows:

Step 1. Perform multiresolution wavelet packet decomposition of the mixed signals $\mathbf{x}$.

Step 2. Remove the $\mathbf{y}_k^j(t)$ from the matrix $\mathbf{y}_k^j$ when $\left\|\mathbf{y}_k^j(t)\right\| < \rho$.

Step 3. Standardize each sub-band signals.

Step 4. Split each sub-band signals to n clusters by k-means clustering.

Step 5. Calculate $WCAD_k^j$, find the minimum $WCAD$ corresponding to the sub-band $\mathbf{y}_{kq}^j$.

Step 6. Calculate μ_{kqi}^j, σ_{kqi}^j in the q-th sub-band, then remove columns $\mathbf{y}_{kqi}^j(h)$ from i-th cluster if $\left|\mathbf{y}_{kqi}^j(h) - \mu_{kqi}^j\right| > c\sigma_{kqi}^j$. After that, we obtain a set of sub-matrices $\tilde{\mathbf{y}}_{kqi}^j$.

Step 7. In each cluster, calculate the first eigenvector corresponding to the maximum eigenvalue of the correlation matrix $\tilde{\mathbf{y}}_{kqi}^j \cdot \left(\tilde{\mathbf{y}}_{kqi}^j\right)^T / \tilde{l}_i$, where $\tilde{l}_i$ is the sample size of i-th cluster. Finally this eigenvector is taken as the corresponding column of the estimation of mixing matrix $\mathbf{A}$ up to permutation and scaling.

Note that, in order to select adaptive decomposition level, one way to automate algorithm completely is to monitor the rate of change of the $WCAD$, as decomposition level is increasing and stop when rate of the change is small or there is a $WCAD_k^j < 0.05$. That is, we can compare $WCAD$ in the parent and child nodes and decide whether the further splitting was fruitful. If the child nodes do not differ significantly among each other in the $WCAD$ or there is $WCAD_k^j < 0.05$, we can stop the further decomposing in that direction.

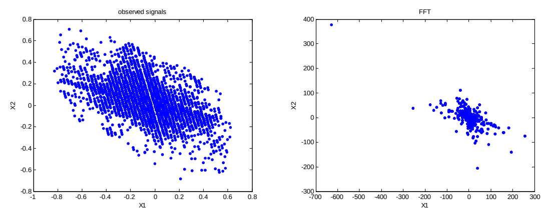
IV. SIMULATION

A. Comparison with Different linear transformations

We begin with an example of two speech signals in different linear transformations. The considered two sources are from the experiment "FourVoices_src" in [16], each source has 65536 samples. The mixing matrix is followed by

$$\mathbf{A} = \begin{bmatrix} 0.2588 & 0.9659 \\ -0.9659 & -0.2588 \end{bmatrix},$$

two mixtures were obtained by $\mathbf{x} = \mathbf{As}$. Fig. 1 gives the scatter plot $\mathbf{x}_1$ against $\mathbf{x}_2$ in different linear transformations. Fig. 1(a) presents a scatter plot of the observed signals showing a single big cloud. As can be seen, the different sources are indistinguishable. Then each mixture is Fast Fourier transform (FFT) transformed and the scatter plot of the frequency domain data is shown in Fig. 1(b) from which we cannot see any noticeable line directions. So STFT with Hanning window is used (see [15] for details). The



(a) (b)

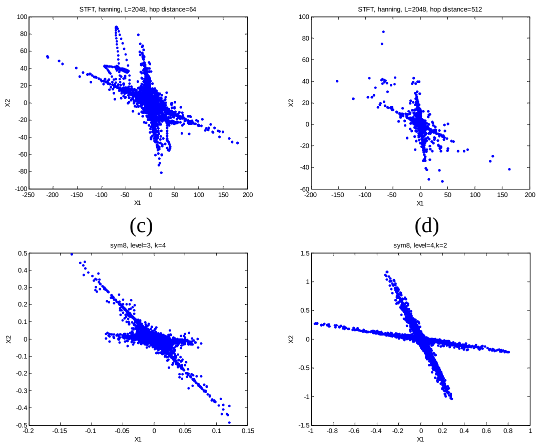


Figure 1. Scatter plot $\mathbf{x}_1$ vs. $\mathbf{x}_2$ which mixed by two speech signals.

(a) the observed signals $\mathbf{x}_1$ and $\mathbf{x}_2$; (b) the transform domains of FFT; (c) the transform domains consists of the coefficients of 2048-point windowed FFT, the hop distance is 64; (d) the hop distance is 512; (e) the wavelet packet is sym8 with 3 scale level and the sub-band is 4; (f) the scale level is 4 and the sub-band is 9.

scatter plot of $\mathbf{x}$ with $L = 2048$ and $d = 64$ is shown in Fig. 1(c). And the scatter plot of $\mathbf{x}$ with $L = 2048$ and $d = 512$ is shown in Fig. 1(d). Further, wavelet packet is applied on observed data， in which the wavelet filter is chosen by the sym8 from SymletsA family [17]. The scatter plots of the frequency domain data $\mathbf{y}_k^j$ for $j = 3,\ k = 4$ and $j = 4$, $k = 9$ are depicted in Fig. 1(e) and Fig. 1(f), respectively. In contrast, WPT is obviously better then FFT and STFT in this example. Almost all significant data are clustered along the two directions of the basis vectors.

B. Results with Determined and Underdetermined Case

In order to show the efficiency of the proposed algorithm in identifying the sparsest sub-band and estimating the mixing matrix, the following numerical experiments are carried out in determined and underdetermined cases. To check how well the mixing matrix is estimated, we introduce the following normalized mean square error (NMSE) in dB as a performance index [4]:

$$NMSE = 10\log_{10}\left(\frac{\sum_{\alpha,\beta}(a_{\alpha\beta})^2}{\sum_{\alpha,\beta}(\hat{a}_{\alpha\beta} - a_{\alpha\beta})^2}\right), \qquad (12)$$

where $\hat{a}_{\alpha\beta}$ is the (α,β) th element of the estimated matrix $\hat{\mathbf{A}}$. In all the tests, we use WPT with 4 decomposition levels. Regarding the type of the wavelet, our choice is sym4, from SymletsA family [17]. And let $\rho = 0.1$ and $c = 0.05$. The experiments are conducted on the following sets of signals: a SixFluteMelodies data which are not sparse in the time domain from www.ac.upe.es /homes/pau. Each source has 58492 samples.

Determined Case:

Experiment 1: $m = n = 2$

Two flute signals are mixed into two mixtures using

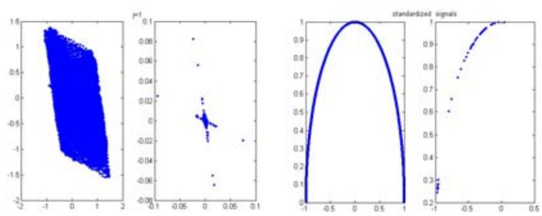
$$\mathbf{A} = \begin{bmatrix} 0.2588 & 0.9659 \\ -0.9659 & -0.2588 \end{bmatrix}.$$

We compute the $WCAD_k^j$ from (9) and (10). Table 1 shows the *WCAD* among the corresponding sub-bands of the wavelet trees obtained after WP based decomposition of the observed data. From table 1, the 3*th* sub-band at the level 2 corresponds to the minimum $WCAD_3^2 = 0.0016 < 0.05$. So we can choose the 3*th* sub-band at the level 2 as the sparsest sub-band for the mixing matrix estimation. Fig. 2 presents the observed signals (left) and the corresponding standardized signals (right) at different decomposition level j, $j = 1,\cdots,4$.

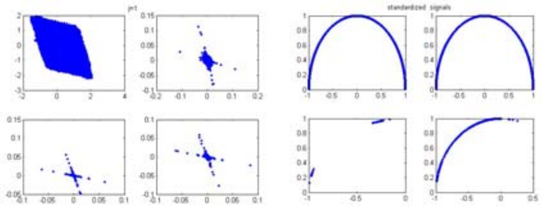
As can be seen, the first subplot in Fig. 2(a) (from left to right) showing a single big cloud, it corresponds to the third subplot in Fig. 2(a) showing a semicircumference. Obviously, the different sources are indistinguishable. Similarly, almost all data are clustered along the two directions of the basis vectors at the fifth subplot in Fig. 2(b), and it corresponds to the seventh subplot in Fig. 2(b) showing two distinct clusters in upper semicircle. From Table 1 and Fig. 2, we observe that the smaller the *WCAD* value, the obvious the clusters, and the sparser the sources.

TABLE I.
THE *WCAD* BETWEEN THE SAME NODES IN THE WAVELET PACKET TREES OF THE OBSERVED SIGNALS IN EXPERIMENT 1

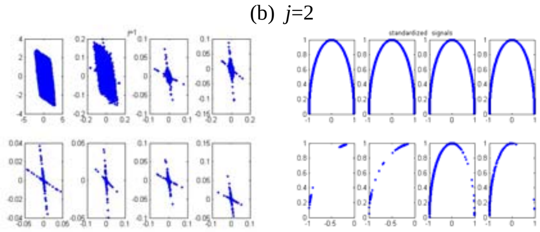
Level 1	Level 2	Level 3	Level 4
0.3621	0.4159	0.3639	0.4228
			0.4012
		0.3771	0.4293
			0.3397
	0.3549	0.3775	0.3873
			0.3592
		0.3773	0.3805
			0.4300
0.0342	**0.0016***	0.0017	0.0025
			0.0046
		0.0153	0.0559
			0.0071
	0.0687	0.0910	0.2120
			0.1085
		0.0716	0.1116
			0.1117



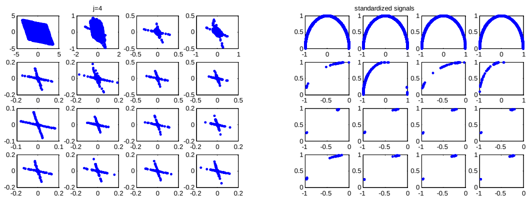
(a) j=1



(b) j=2



(c) j=3



(d) j=4

Figure 2. the WPT of the observed signals (left) and the corresponding standardized signals (right) at different decomposition level j, $j=1,\cdots,4$.

Using the proposed algorithm, we obtain the estimated mixing matrix as follows:

$$\hat{\mathbf{A}}=\begin{bmatrix}-0.2607 & -0.9658\\ 0.9654 & 0.2593\end{bmatrix},$$

the estimation error obtained is $NMSE=56.934\text{db}$. In this case, because $\mathbf{A}$ is invertible, we can recover the sources by $\tilde{\mathbf{s}}(t)=\hat{\mathbf{A}}^{-1}\mathbf{x}(t)=\hat{\mathbf{A}}^{-1}\mathbf{A}\mathbf{s}(t)$.

Underdetermined Case

Experiment 2: $m=2,\ n=3$

Three flute signals are mixed into two mixtures using

$$\mathbf{A}=\begin{bmatrix}0.7071 & 0.2588 & 0.9659\\ 0.7071 & -0.9659 & -0.2588\end{bmatrix}.$$

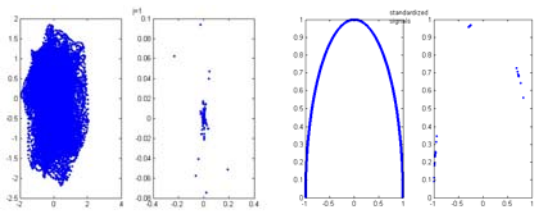
We compute the $WCAD_k^j$ from (9) and (10). The *WCAD* among the corresponding sub-bands of the wavelet trees obtained after WP based decomposition of the observed data is shown in Table 2. From table 2, the 2*th* sub-band at the level 1 corresponds to the $WCAD_2^1=0.0073$. Although the $WCAD_2^1$ is not the minimum $(WCAD_9^4=0.0013)$, $WCAD_2^1<0.05$ has satisfied the stopping condition for decomposition. So we can choose the 2*th* sub-band at the level 1 as the sparse sub-band for the mixing matrix estimation.

TABLE II.
THE *WCAD* BETWEEN THE SAME NODES IN THE WAVELET PACKET TREES OF THE OBSERVED SIGNALS IN EXPERIMENT 2

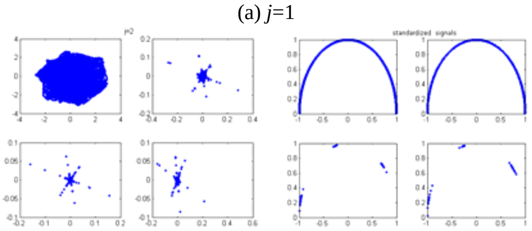
Level 1	Level 2	Level 3	Level 4
0.2261	0.2258	0.2260	0.2232
			0.2232
		0.2309	0.2321
			0.2260
	0.2279	0.2480	0.2503
			0.2434
		0.2140	0.2135
			0.2353
0.0073*	0.2463	0.0015	0.0013
			0.0035
		0.0066	0.0219
			0.2320
	0.0091	0.2563	0.1577
			0.0362
		0.0077	0.0146
			0.0097

Like *Experiment 1* analyses, the first subplot in Fig. 3(a) (from left to right) showing a single big cloud, it corresponds to the third subplot in Fig. 3(a) showing a semicircumference. Obviously, the different sources are indistinguishable. Similarly, almost all data are clustered along the three directions of the basis vectors at the second subplot in Fig. 3(a), and it corresponds to the fourth subplot in Fig. 3(a) showing three distinct clusters in upper semicircle. From table 2 and Fig. 3, we find that the smaller the value *WCAD* , the obvious the clusters, and the sparser the sources.

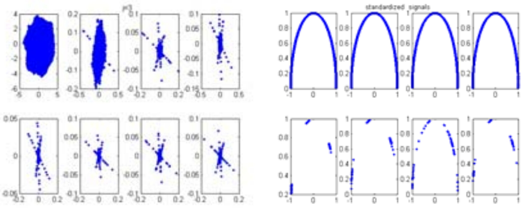
Using the proposed algorithm, at the 2*th* sub-band at the level 1, we obtain the estimated mixing matrix as follows:



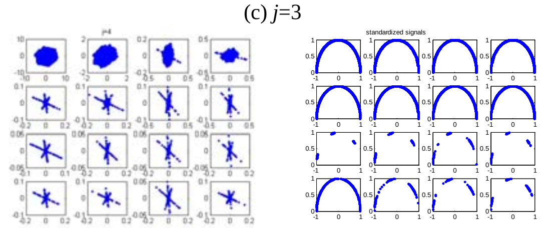
(a) j=1



(b) j=2



(c) j=3



(d) j=4

Figure. 3 the WPT of the observed signals (left) and the corresponding standardized signals(right) at different decomposition level j, $j=1,\cdots,4$.

$$\hat{\mathbf{A}}_1 = \begin{bmatrix} -0.2521 & -0.7163 & -0.9652 \\ 0.9677 & -0.6978 & 0.2614 \end{bmatrix},$$

the estimation error obtained is $NMSE = 41.2203\text{db}$. At the same time, we calculate the $\hat{\mathbf{A}}$ at the 9*th* sub-band at the level 4. It is

$$\hat{\mathbf{A}}_2 = \begin{bmatrix} -0.2611 & -0.7128 & -0.966 \\ 0.9653 & -0.7013 & 0.2585 \end{bmatrix},$$

the estimation error is $NMSE = 46.2051\text{db}$. Compare with these two *NMSE*s, there is only a little difference, so we choose the 2*th* sub-band at the level 1 as the sparse sub-band for the mixing matrix estimation is reasonable.

V. CONCLUSION

In this paper, we proposed a novel WP-based approach to estimate the mixing matrix for linear instantaneous mixtures. Adaptive sub-band decomposition was realized through WPT implemented in a form of iterative filter bank. The sub-band with the sparsest component is selected by measuring *WCAD* between the same nodes of the wavelet tree in the multiresolution decomposition of the observed signals. Then we can estimate the mixing matrix $\mathbf{A}$ in this sub-band through clustering algorithm. The proposed algorithm does not have any restriction on the number of sources and mixtures. Our method provided some major advantages over classical ICA methods. That is:

(1) We can accurately estimate $\mathbf{A}$ as long as the source signals satisfy sparse at one sub-band regardless of under-determined case or over-determined case.

(2) If $\mathbf{A}$ is invertible, we can recover the sources by $\tilde{\mathbf{s}}(t) = \hat{\mathbf{A}}^{-1}\mathbf{x}(t) = \hat{\mathbf{A}}^{-1}\mathbf{A}\mathbf{s}(t)$. (Ex.1)

(3) If the source signals are enough sparse, we can get the estimation of the source signals through the FOCUSS algorithms [18-19], linear programming (LP) [20-21] or the short path decomposition [15,21] after knowing $\hat{\mathbf{A}}$.

(4) This method can be extended to higher-dimensional situation. (Ex.2)

Our further investigations will focus on the extension of the proposed method to the dependent sources separation, which is a more important but complicated problem.

ACKNOWLEDGMENT

This work is supported by the Natural Science Foundation of China (No. 61071188), the Natural Science Foundation of Hubei Province of China (No. 2009CDB077) and the Special Fund for Basic Scientific Research of Central Colleges, China University of Geosciences Wuhan (No. CUGL090252).

REFERENCES

[1] P. Common, C. Jutten, *Handbook of blind source separation*, Academic Press, New York, 2010.

[2] A. Cichocki, S. Amari, *Adaptive blind signal and image processing*, J. Wiley, New York, 2002.

[3] Y. Li, S. Amari, A. Cichocki, D.W.C. Ho, and S. Xie, Underdetermined blind source separation based on sparse representation, *IEEE Transactions on Signal Processing*, 54(2), pp. 423-437, 2006.

[4] V.G. R, S.N. Koh, I.Y. Soon, An algorithm for mixing matrix estimation in instantaneous blind source separation, *Signal Processing*, 89(9), pp.1762-1773, 2009.

[5] Z. He, S. Xie, S. Ding, and A. Cichocki, Convolutive blind source separation in the frequency domain based on sparse representation, *IEEE Transactions on Audio, Speech and Language Processing*, 15(5), pp.1551-1563, 2007.

[6] Z. He, S. Xie, L. Zhang, and A. Cichocki, A note on Lewicki-Sejnowski gradient for learning overcomplete representation, *Neural Computation*, 20(3) , pp.636-643, 2008.

[7] P. Pauca, R. Plemmons, M. Giffin, K.Hamada, Unmixing spectral data for space objects using independent component component analysis and nonnegative matrix factorization, *in: Proceedings Smos Technical Conference*, 2004.

[8] A. Hyvarinen, P.O. Hoyer, and E. Oja, *Image denoising by sparse code shrinkage*. In S. Hakin and B. Kosko, editors, Intelligent Signal Processing, IEEE Press, 2001.

[9] S. Abdallah, M. Plumbley, Polyphonic transcription by non-negative sparse coding of power spectra, *in: Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Barcelona Spain, pp.318-325, 2004.

[10] R. Gribonval and S, Lesage, A survey of sparse component analysis for blind source separation: principles, perspectives, and new challenges, *in: Proceedings of 14th European Symposium on Artificial Neural Networks (ESANN'06)*, Bruges, Belgium, pp.323-330, 2006.

[11] A.A. Amini, M. Babaie-Zadeh, and C. Jutten, A fast

method for sparse component analysis based on iterative detection-estimation, in Proceedings of 26th International Workshop on Bayesian Inference and Maximum Entropy Methods *in Science and Engineering (MaxEnt'06)* , pp.123-130, 2006.
[12] F. Abrard, Y. Deville, A time-frequency blind signal separation method applicable to underdetermined mixtures of dependent sources, *Signal Processing85,* pp.1389-1403, 2005.
[13] Y. Deville, M. Puigt, Temporal and time-frequency correlation-based blind source separation methods, Part I: Determined and underdetermined linear instantaneous mixtures, *Signal Processing 87,* pp.374-407, 2007.
[14] I. Koprova and D. Sersic, Wavelet packets approach to blind separation of statistically dependent sources, *Neurocomputing,* 71(7-9) , pp.1642-1655, 2008.
[15] P. Bofill and M. Zibulevsky, Underdetermined blind source separation using sparse representations, *Signal Processing,* 81(11), pp.2353-2362, 2001.
[16] Z.S. He, S.L. Xie, and Y.L. Fu, sparse representation and blind source separation of ill-posed mixtures, *Science In China Ser. E Information Sciences,* 36(8), pp.864-879, 2006.
[17] S. Mallat, *A wavelet tour of signal processing*, Academic Press, New York, 1999.
[18] A. Cichocki and S. Amari, *Adaptive blind signal and image processing: learning algorithms and applications,* Wiley, New York, 2003.
[19] S.F. Cotter, B.D. Rao, K. Engan, and K. Kreutz-Delgado, Sparse solutions to linear inverse problems with multiple measurement vectors, *IEEE Transactions on Signal Processing,* 53(7), pp.2477-2488, 2005.
[20] Y.Q. Li, A. Cichocki, and S. Amari, Analysis of sparse representation and blind source separation, *Neural Computation* 16(6) , pp.1193-1234, 2004.
[21] I. Takigawa, M. Kudo, and J. Toyama, Performance analysis of minimum l_1 -norm solutions for underdetermined source separation, *IEEE Transactions on Signal Processing*, 52(3) , pp.582-591, 2004.

Yujie Zhang was born in Hubei, China. She received her M.S. degree in Applied Mathematics in 2006 from China University of Geosciences, China.

She is currently a lecturer at the China University of Geosciences, China. Her research interests include Blind Signal Processing, Time-frequency Analysis and their applications.

Huiming Peng was born in Hunan, China. He received his M.S. degree in Probability & Statistics in 2005 from Wuhan University. His research interests included Statistical Signal Processing and Bioinformatics.

Hongwei Li was born in Hunan, China. He received his Ph.D. degree in Applied Mathematics in 1996 from Beijing University. Currently, he is a professor at China University of Geosciences, China. His research interests include Statistical Signal Processing, Non-stationary Signal Processing and Blind Signal Processing.