UNDERDETERMINED SOURCE SEPARATION IN THE TIME-FREQUENCY DOMAIN

Zeyong Shan, Jacob Swary, and Selin Aviyente

Department of Electrical and Computer Engineering
Michigan State University, East Lansing, MI 48824, USA
E-mail: {shanzeyo, swaryjac, aviyente@egr.msu.edu}

ABSTRACT

Underdetermined blind source separation (UBSS) is a challenging problem that has recently been formulated in the time-frequency domain. Previous work in the area of UBSS problem focuses on using sparse representations of signals, such as matching pursuit and wavelet packet decomposition, for identifying the sources. However, these methods are in general computationally expensive and rely on the choice of an appropriate basis function for obtaining a sparse representation. In this paper, we propose a new approach based on Cohen's class of distributions. The new approach takes advantage of the high resolution of time-frequency distributions for obtaining a sparse representation, and separates the sources by a simple clustering algorithm followed by a convex optimization problem. Compared to other time-frequency based separation methods, the presented approach is characterized by its simplicity and ease of implementation. Experimental results indicate the effectiveness of the proposed approach at separating the sparse signals in the time-frequency domain.

Index Terms— Time-frequency distribution, blind source separation, sparsity

1. INTRODUCTION

Underdetermined blind source separation (UBSS) considers the recovery of the underlying source signals from mixtures when there are more sources than sensors. It has applications in different areas, such as communications, speech processing, image processing, and biomedical signal processing [1]. UBSS is a more challenging problem compared with the (over)determined source separation. Contrary to the (over)-determined case, estimating the mixing system is not sufficient for reconstruction of the sources since the mixing matrix is not invertible. Therefore, additional *a priori* information about the sources is needed to allow for reconstruction. One increasingly popular and powerful assumption is the sparsity of the sources for a given basis [2]. Sparse signal representations lend themselves to good separability of the sources, because most of the energy of a basis coefficient at any time instant belongs to a single source.

Since most real life signals are non-stationary and not sufficiently sparse in the time domain, the time-frequency representations of underlying signals are used for source separation. One advantage of using time-frequency analysis is that the sources are much sparser in the time-frequency domain compared to only in the time or frequency domains due to the high resolution of time-frequency representations. Several time-frequency based UBSS algorithms have been proposed to achieve source separation using time-frequency distributions (TFDs) [3, 4, 5, 6]. In [3], the binary time-frequency masks are constructed using the sparsity of speech signals in the short time Fourier transform (STFT) domain to extract sources from only two mixtures. However, the algorithm assumes a specific signal model and is applicable to only two mixtures at a time. The method presented in [4] is an extension of that in [3], with increased implementation complexity. The mixing matrix is estimated using vector clustering based on the co-linearity of linear time-frequency representations of mixtures in [5]. Recently, a two-stage cluster-then-l^1-optimization approach for UBSS problem in the wavelet packet domain has been proposed where the mixing matrix and the sources are estimated separately [6]. In this paper, we extend this two-stage sparse representation approach to the time-frequency domain, and compare its performance with that of wavelet packets.

2. BACKGROUND ON TIME-FREQUENCY ANALYSIS

A time-frequency distribution (TFD), $X(t,\omega)$, from Cohen's class can be expressed as [1] [7]:

$$X(t,\omega) = \int\int\int \phi(\theta,\tau) s(u+\frac{\tau}{2}) s^*(u-\frac{\tau}{2}) e^{j(\theta u - \theta t - \omega\tau)} du\, d\theta\, d\tau, \quad (1)$$

where $\phi(\theta,\tau)$ is called the kernel function and $s(t)$ is the signal. Some of the most desired properties of TFDs are the energy preservation and the marginals. They are satisfied when

[1] All integrals are from $-\infty$ to ∞ unless otherwise stated.

$\phi(\theta,0)=\phi(0,\tau)=1 \ \ \forall\tau,\theta$ and are given as follows:

$$\begin{aligned}\int\int X(t,\omega)\,dt\,d\omega &= \int |s(t)|^2\,dt = \int |S(\omega)|^2\,d\omega,\\ \int X(t,\omega)\,d\omega &= |s(t)|^2 \ , \int X(t,\omega)\,dt = |S(\omega)|^2.\end{aligned} \tag{2}$$

Since severe cross-terms exist in different time-frequency regions for some TFDs particularly when the signal is multicomponent, kernel functions that minimize the interference are developed. For cross-term minimization, $\phi(\theta,\tau)$ should satisfy

$$\phi(\theta,\tau) << 1 \quad \text{for } \theta\tau >> 0. \tag{3}$$

These kernel functions produce reduced interference distributions (RIDs).

3. TIME-FREQUENCY BASED SPARSE REPRESENTATION APPROACH FOR UBSS

In this section, a two-stage approach for the UBSS problem in the time-frequency domain is presented, in which the first stage is for determining the mixing matrix, and the second stage is for estimating the source signals.

3.1. Linear Mixture Model and Assumptions

In this paper, we consider the problem of determining the source signals when the number of observed mixtures is less than the number of source signals. Assume that the M mixtures, $\mathbf{z}(t) = [z_1(t), z_2(t), \ldots, z_M(t)]^T$, of the N non-stationary complex source signals, $\mathbf{s}(t) = [s_1(t), s_2(t), \ldots, s_N(t)]^T$, are given with

$$\mathbf{z}(t) = \mathbf{B}\mathbf{s}(t), \tag{4}$$

where $\mathbf{B}$ is the $M \times N$ instantaneous mixing matrix ($M < N$). We want to extract the underlying sources $\mathbf{s}(t)$.

Each mixture, $z_i(t)$, is first transformed to the time-frequency plane, and then the corresponding time-frequency distribution is vectorized to form a matrix of time-frequency distributions, $\mathbf{X}$. In our source separation problem, the observed time-frequency distributions, $\mathbf{X}$, can be written as a linear combination of the original sources' TFDs, $\mathbf{S}$, assuming the cross-terms between the sources are negligible by using a RID:

$$\mathbf{X} \approx \mathbf{B}^2\mathbf{S} = \mathbf{A}\mathbf{S}, \tag{5}$$

where $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_P] \in R^{M\times P}$ is the mixtures of the sources, P is the total number of time and frequency points, $\mathbf{A} = \mathbf{B}^2 = [\mathbf{a}_1, \cdots, \mathbf{a}_N] \in R^{M\times N}$ is an unknown mixing matrix, $\mathbf{B}^2$ is the element-by-element square of the mixing matrix in the time domain, and $\mathbf{S} = [\mathbf{s}_1, \cdots, \mathbf{s}_P] \in R^{N\times P}$ is the time-frequency representations of the N unknown source signals. Note that equation (5) is applicable to any signal. It is also assumed that the source signals are sparse in the time-frequency domain.

3.2. Determination of the Mixing Matrix

Due to the sparsity of the source signals in the time-frequency domain, there exists many columns of $\mathbf{S}$ with only one nonzero entry. For instance, suppose that $\mathbf{s}_{i_1}, \cdots, \mathbf{s}_{i_K}$ are K columns of $\mathbf{S}$, where only the first entry of each of these columns is nonzero, then we have

$$\mathbf{A}\mathbf{s}_{i_j} = \mathbf{a}_1 s_{1i_j} \quad j = 1, \cdots, K, \tag{6}$$

and

$$[\mathbf{x}_{i_1}, \cdots, \mathbf{x}_{i_K}] = \mathbf{A}[\mathbf{s}_{i_1}, \cdots, \mathbf{s}_{i_K}] = [\mathbf{a}_1 s_{1i_1}, \cdots, \mathbf{a}_1 s_{1i_K}], \tag{7}$$

where, $\mathbf{x}_{i_j}$ is the i_jth column of $\mathbf{X}$, $\mathbf{a}_1$ is the first column of $\mathbf{A}$, and s_{1i_j} is the first entry of $\mathbf{s}_{i_j}$. From equation (7), we see that each $\mathbf{x}_{i_j}$ is equal to $\mathbf{a}_1$ multiplied by a scalar s_{1i_j}, which means that these K column vectors of $\mathbf{X}$, $\mathbf{x}_{i_1}, \cdots, \mathbf{x}_{i_K}$, are distributed along the direction of $\mathbf{a}_1$. Thus, ideally after normalization, $\mathbf{x}_{i_1}, \cdots, \mathbf{x}_{i_K}$ are mapped to a unique vector on the multidimensional unit circle which is equal to $\mathbf{a}_1$. However, in practice, since the mixture matrix $\mathbf{X}$ is not completely sparse in the time-frequency domain, $\mathbf{x}_{i_1}, \cdots, \mathbf{x}_{i_K}$ are not exactly in the same direction as $\mathbf{a}_1$, but rather in the neighborhood of $\mathbf{a}_1$. This means that $\mathbf{a}_1$ lies at the center of $\mathbf{x}_{i_1}, \cdots, \mathbf{x}_{i_K}$.

Therefore, we use the K-means clustering method to cluster the column vectors of the mixture matrix $\mathbf{X}$ into different clusters, where the center of each cluster corresponds to one column vector of the mixing matrix $\mathbf{A}$. By doing so, we can obtain an estimate of the mixing matrix $\mathbf{A}$.

3.3. Estimation of the Source Signals for a Given Mixing Matrix

After obtaining the estimated mixing matrix, the next stage is to estimate the source signals. For a given mixing matrix $\mathbf{A}$ in equation (5), the source signals can be estimated by maximizing the posterior distribution $P(\mathbf{S}|\mathbf{X}, \mathbf{A})$ of $\mathbf{S}$. Under the assumption that the prior is Laplacian, maximizing posterior distribution can be implemented by solving the following optimization problem [8]:

$$\min \sum_{i=1}^{N}\sum_{j=1}^{P} |s_{ij}|, \quad \text{subject to } \mathbf{A}\mathbf{S} = \mathbf{X}. \tag{8}$$

Hence, the l^1-norm

$$J_1(\mathbf{S}) = \sum_{i=1}^{N}\sum_{j=1}^{P} |s_{ij}| \tag{9}$$

can be used as the sparsity measure.

It is not difficult to prove that the optimization problem (8) is equivalent to the following set of P smaller scale linear programming (LP) problems:

$$\min \sum_{i=1}^{N} |s_{ij}|, \quad \text{subject to } \mathbf{A}\mathbf{s}_j = \mathbf{x}_j \quad \text{for } j = 1, \cdots, P. \tag{10}$$

Finally, we propose the following algorithm for estimating the source signals:

Algorithm:

1. Pre-threshold the mixture matrix $\mathbf{X}$ to obtain a sparser data matrix $\widehat{\mathbf{X}}$.
2. Normalize the column vectors of the data matrix $\widehat{\mathbf{X}}$.
3. Take a sufficiently large positive integer K as the number of clusters. Choose the initial points of iteration and the distance measure criterion. In this paper, we choose the squared Euclidean distance as the criterion.
4. Do K-means clustering on $\widehat{\mathbf{X}}$ followed by normalization to estimate the sub-optimal mixing matrix $\mathbf{A}$.
5. Using the estimated mixing matrix $\mathbf{A}$ and the mixtures $\mathbf{X}$, estimate the time-frequency representations $\mathbf{S}$ by solving the set of LP problems in equation (10).

4. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, several examples will be used to illustrate the effectiveness of the proposed approach to separate the sparse source signals from their fewer mixtures in the time-frequency domain. The binomial kernel [7] is used for computing the TFD since it belongs to the class of reduced interference distributions (RIDs).

Example 1: The set of observed signals are two linear combinations of four Gabor logons. These four Gabor logons are centered at the time sample point and the normalized frequency (30,0.7), (160,-0.7), (70,-0.4), and (120,0.1), respectively. Each observed signal is first transformed to the time-frequency domain with $I = 50$ time samples and $L = 64$ frequency samples. Each TFD is then vectorized to form a TFD mixture matrix $\mathbf{X} = [\mathbf{X}_1; \mathbf{X}_2]$ of size 2×3200.

Fig. 1 presents a scatter plot of the mixtures $\mathbf{X}$ ($\mathbf{X}_2$ vs. $\mathbf{X}_1$) in the time-frequency domain. It can be seen from this plot that almost all significant data points are distributed along four different directions, thus providing very good separability. The separation results using the proposed approach are illustrated in Fig. 2. Fig. 2(a) and (b) represent the two mixtures. The four extracted Gabor logon signals are shown in Fig. 2(c), (d), (e), and (f). The results indicate that these four Gabor logons can be successfully separated from their two mixtures using the proposed approach based on their sparsity

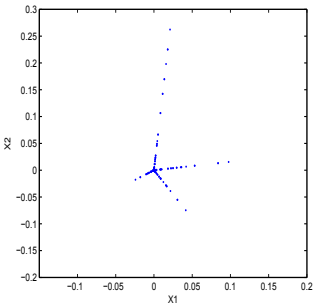

Fig. 1. Scatter plot of two mixtures of four Gabor logons in the time-frequency domain

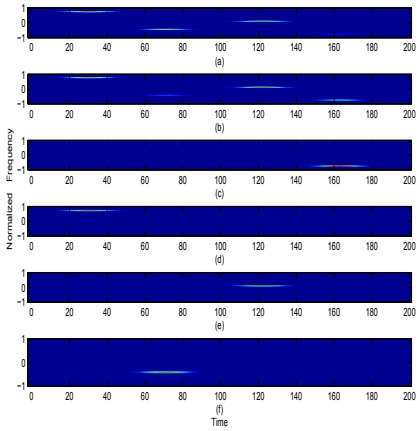

Fig. 2. The mixtures and the separation of four Gabor logons: (a) and (b) two mixtures; (c), (d), (e), and (f) four extracted Gabor logons

with an average signal to interference ratio (SIR) of 36.1251 dB.

Example 2: Two mixtures of a chirp signal and two Gabor logons are given. The chirp signal has a linear frequency increasing from an initial normalized frequency of -0.2 to a normalized frequency of 0.2. The Gabor logons are the first two Gabor logons given in Example 1. A scatter plot of the two mixtures in Fig. 3 shows that it is similar to the first example in that the distributions of data points belonging to different sources are along three different directions. Since the chirp signal overlaps with the two Gabor logons in the time domain, typical time domain separation methods can not be used to perfectly recover them. However, it is illustrated in Fig. 4 that these three signals can be effectively extracted in the time-frequency domain using the proposed method with an average SIR of 32.7634 dB.

Example 3: In this example, the same two mixtures of four Gabor logons given in Example 1 are used. The effectiveness of the presented approach is compared for TFDs and wavelet packets (WP) in the presence of noise. Haar wavelet with five levels is used for the wavelet packet decomposition.

To show the effect of increased sparsity of TFDs, the mix-

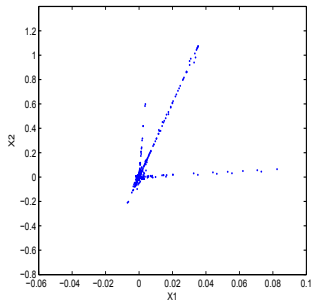

Fig. 3. Scatter plot of two mixtures of a chirp and two Gabor logons in the time-frequency domain

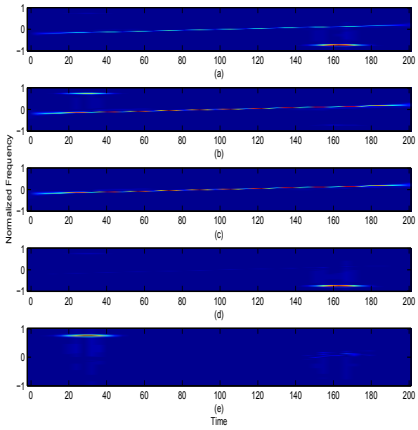

Fig. 4. The mixtures and the separation of a chirp and two Gabor logons: (a) and (b) two mixtures; (c) extracted chirp; (d) and (e) two extracted Gabor logons

tures at different levels of white Gaussian noise are considered. 100 Monte Carlo simulations are used for each noise level. The average mean squared error (MSE) between the extracted sources and the original sources is calculated for both the TFD and WP. The TFD and WP provide similar results when there is no noise. However, as the noise level increases, the performance of the WP rapidly degrades compared to that of the TFD. The MSE versus the signal-to-noise ratio (SNR) is shown in Fig. 5 for both the TFD and WP. This result shows that the RID results in a more sparse time-frequency surface compared to the WP, which improves the robustness of BSS under noise.

5. CONCLUSIONS

This paper introduces a two-stage approach for underdetermined blind separation of sparse and non-stationary sources using TFDs. The mixing matrix is estimated using K-means clustering algorithm based on the sparsity of the sources; for a given mixing matrix, the sources are extracted using a linear programming method. The performance of the proposed

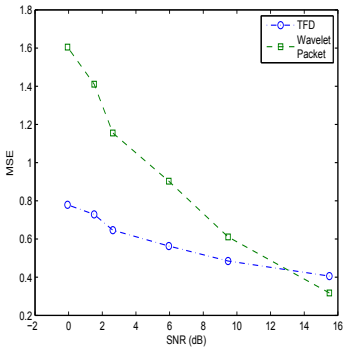

Fig. 5. Comparison of MSE versus SNR for extracted sources with TFD and WP

approach is compared with wavelet packets under different noise levels. The results show that the presented method is simple and effective at separating the sources from their mixtures, and is more robust than wavelet packets under noisy environments. Future work will consider the separation of source signals that are less sparse in the time-frequency domain, including a pre-sparsification stage.

6. REFERENCES

[1] A. K. Nandi (editor), *Blind estimation using higher-order-statistics*, Kluwer Acadimic Publishers, 1999.

[2] S. Mallat and Z. Zhang, "Matching pursuits with time–frequency dictionaries," *IEEE Trans. on Signal Processing*, vol. 41, pp. 3397–3415, 1993.

[3] O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time–frequency masking," *IEEE Trans. on Signal Processing*, vol. 52, pp. 1830–1847, 2004.

[4] Y. Q. Li, S. Amari, A. Cichocki, W. C. Ho, and S. Xie, "Underdetermined blind source separation based on sparse representation," *IEEE Trans. on Signal Processing*, vol. 54, pp. 423–437, 2006.

[5] B. Barkat, F. Sattar, and K. Abed-Meraim, "Source separation of instantaneous mixtures using a linear time-frequency representation and vectors clustering," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, 2006, vol. 3, pp. 460–463.

[6] P. Bofill and M. Zibulevsky, "Underdetermined blind source separation using sparse representations," *Signal Processing*, vol. 81, pp. 2353–2362, 2001.

[7] L. Cohen, *Time–Frequency Analysis*, Prentice Hall, New Jersey, 1995.

[8] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Scientific Comp.*, vol. 20, pp. 33–61, 1999.