A PROBABILISTIC APPROACH FOR BLIND SOURCE SEPARATION OF UNDERDETERMINED CONVOLUTIVE MIXTURES

J. Michael Peterson

SIPI, University of Southern California
PHE 304, Los Angeles, CA 90089
johnmpet@usc.edu

Shubha Kadambe

HRL Laboratories, LLC,
3011 Malibu Canyon Rd Malibu, CA 90265
skadambe@hrl.com

ABSTRACT

There are very few techniques that can separate signals from the convolutive mixture in the underdetermined case. We have developed a method that uses overcomplete expansion of the signal created with a time-frequency transform and that also uses the property of sparseness and a Laplacian source density model to obtain the source signals from the instantaneously mixed signals in the underderdetermined case. This technique has been extended here to separate signals (a) in the case of underdetermined convolutive mixtures, and (b) in the general case of more than 2 mixtures. Here, we also propose a geometric constrained based search approach to significantly reduce the computational time of our original ”dual update” algorithm. Several examples are provided. The results of signal separation from the convolutive mixtures indicate that an average signal to noise ratio improvement of 5.3 dB can be obtained.

1. INTRODUCTION

The method of blind source separation (BSS) attempts to estimate the sources or inputs of a system by observing the outputs of the system, without knowing how the sources were mixed together. There are two cases of problems - the instantaneous mixture (IM) where the system has no memory and the convolutive mixture where the length of the filters in the system is greater than one. The IM case can be written in matrix form as

$$\mathbf{x}[n] = \mathbf{a}\mathbf{s}[n] \tag{1}$$

where $\mathbf{x}[n]$ is the output, $\mathbf{s}[n]$ is the source and $\mathbf{a}$ is the mixing matrix of size N by M, where N is the number of sensors and M is the number of sources.

Recently, several authors have shown the feasibility of BSS when N is less than M[1, 2, 3]. This can be achieved by transforming the sensed signals to the time-frequency domain and using the properity of sparseness to help in the estimation of the mixing matrix. After the mixing matrix has been estimated, the resulting estimate is used to estimate the sources, where the sources are assumed to be independent and have a Laplacian density function. However [3] uses what the authors call a ”dual update” approach that iteratively and jointly estimates the source and the mixing matrix by minimizing L1 and L2 norms. This approach is extended to the convolutive mixture in this paper.

So far there has been little done to solve the convolutive mixture separation especially, for the undermined case. However, one author has developed a method to perform underdetermined BSS when the system's transfer functions contain only delay elements that are attenuated by a factor [4]. Here, we attempt to fill that void. The next section discusses how to perform BSS using an overcomplete expansion. This is followed by challenges encountered in the convolutive case and the proposed extensions to handle those challenges. Finally, the results of several experiments showing the effectiveness of this method is discussed.

2. PROBABILISTIC BASED BSS FOR UNDERDETERMINED IM

This section summarizes our previous algorithm described in [3] and generalizes it by proposing some modifications (a) to handle more than 2 mixtures, (b) to robustly estimate the initial mixing matrix and (c) to speed up the iterative ”dual update” algorithm.

In order to separate sources that have been combined instantaneously, first, the sensed signals are transformed into a time-frequency domain using the short-time Fourier Transform (STFT), thus resulting in

$$\mathbf{X}(k,m) = \mathbf{A}(k,m)\mathbf{S}(k,m) \tag{2}$$

where $\mathbf{X}(k,m)$, $(A)(k,m)$ and $\mathbf{S}(k,m)$ are the time-frequency distributions of $\mathbf{x}[n]$, (a) and $\mathbf{s}[n]$, respectively. Since $\mathbf{a}$ represents an instantaneous system, $\mathbf{A}(k,m) = \mathbf{a}$ and is constant for all frequency bands, k and time m.

In order to find an initial estimate of $\mathbf{a}$, the mutual information between sensors is computed for each subband.

0-7803-7663-3/03/$17.00 ©2003 IEEE

The subband with the highest mutual information is chosen since it represents the subband that exhibits the most separation and hence, can best estimate $\mathbf{a}$. For the estimation of initial $\mathbf{a}$, the following equation is used

$$\theta_{i,j}(k,m) = \arctan(\frac{X_j(k,m)}{X_i(k,m)}) \tag{3}$$

where k is the chosen subband and i and j represent the signal received at the i^{th} and j^{th} sensors. If $S_l(k,m)$ is much larger than all the other sources at a particular k and m, then (3) simplifies to

$$\begin{aligned} \theta_{i,j}(k,m) &\approx \arctan(\frac{A_{j,l}(k)S_l(k,m)}{A_{i,l}(k)S_l(k,m)}) \\ &= \arctan(\frac{A_{j,l}(k)}{A_{i,l}(k)}) \end{aligned} \tag{4}$$

This results in clusters of measurements that correspond to the arctangent of the ratios of rows $\mathbf{a}_j$ and $\mathbf{a}_i$. Several methods could be used to find these clusters. Some authors use a peak picking of the histogram [3]. Instead a hierarchical cluster is computed in the proposed approach, where each observation is taken in succession and merged with its nearest neighbor. This is computationally less intensive than finding the two observations that are closest together and then merging them. The result is used as an initial guess for a k-means clustering algorithm. The cluster means are then used for the initial estimate of the mixing matrix. It was found experimentally that this method works better than using either hierarchical clustering or k-means separately and also is more robust than peak picking method.

The next step is to jointly minimize the cost function (”dual update”). In this case the sources are modeled as a Laplacian and the noise is assumed to be white Gaussian, so the resulting cost function is

$$-L(\mathbf{S}|\mathbf{X},\mathbf{A}) = (\mathbf{X}-\mathbf{AS})^t(\mathbf{X}-\mathbf{AS}) + \lambda\mathbf{c}^t|\mathbf{S}| \tag{5}$$

where $\mathbf{c}^t = [1\ 1\ \ldots\ 1]$, t denotes the Hermitian transpose of a matrix and k and m are assumed in order to simplify the expression. This expression is optimized by first finding $\mathbf{S}(k,m)$ that minimizes $\lambda\mathbf{c}^t|\mathbf{S}(k,m)|$ under the constraint that $\mathbf{X}(k,m) = \mathbf{A}(k)\mathbf{S}(k,m)$. The second part of the procedure re-estimates $\mathbf{A}(k)$ so that the sources will be more independent.

The easiest way to perform the first part of the “dual-update” method is to recognize that there is a local minimum whenever there are $M-N$ zeros in the $\mathbf{S}(k,m)$ vector. This can be shown by using a geometrical argument. First draw the shape formed by all points at a certain cost, ϵ. The resulting shape is an M-dimensional cube with vertices located on the axes at a distance of ϵ away from the origin. Now the constraint has dimension $M-N$, so when there is one more source than sensor the constraint is a line. If the line goes through the cube, then the portion inside the cube is at a lower cost and the portion outside the cube is at a higher cost. If the cube is shrunk until the constraint only touches the edge of the cube, the point of intersection is the lowest cost. If the line is parrellel to one of the sides of the box, then there are an infinite number of solutions. This case corresponds to the $\mathbf{A}(k)$ matrix having at least two identical column vectors. The other case requires that the line intersect a vertex of the cube. Of course this occurs when the line passes through a plane created by all combinations of N axes or in other words there are $M-N$ zeros in $\mathbf{S}(k,m)$. This gives a finite number of points to check. The point with the lowest cost is the global minimum. Inclusion of this geometric constrained based search approach has speeded up our original Armijo line constrained search of the ”dual update” algorithm significantly. A similar method is used in [1].

3. CHALLENGES FOR THE CONVOLUTIVE CASE

This section focuses on the problems to be addressed when using the above mentioned method (”basic method”) to separate signals in the convolutive case where the mixing matrix is not constant as in the case of IM but is a function of time. The logical extension of the ”basic method” would be to take the FFT of the signal with a length of FFT that is long enough to ensure that the convolution can be approximated as multiplication in the Fourier domain. Then the ”dual update” algorithm can be applied in each subband independently. This approach does have several drawbacks. First of all, the algorithm finds the signal separation within an arbitrary scale factor and arbitrary permutation. This means that the scale factors and permutations will need to be consistent between different subbands. Incorrect scale factors cause spectral distortion. This scale factor problem is solved here by constraining the filter structure such that $\|\mathbf{A}_j(k)\|^2 = 1$, where $\mathbf{A}_j(k)$ is the j^{th} column vector of $\mathbf{A}(k)$. This was also used in [5]. The method that is used to find permutation is based on the inter-frequency correlation[6]. This relies on the fact that the adjacent subbands of non-stationary signals are correlated. This can be used in the following equation

$$\hat{\mathbf{P}}(k) = \arg\max_{\mathbf{P}(k)} \sum_{m=1}^{L}\sum_{j=1}^{k-1}(\mathbf{P}(k)\bar{\mathbf{S}}(k,m))^t\bar{\mathbf{S}}(j,m) \tag{6}$$

where $\mathbf{P}(k)$ is the permutation matrix and $\bar{\mathbf{S}}(k,m)$ is the envelope signal of $\mathbf{S}(k,m)$ which is created by passing the absolute values of the source signals through a low pass filter. The permutation of the first subband is designated as the correct permutation. The permutation of the next subband is found by using (6). The source signals at that subband are permuted according to the resulting $\mathbf{P}(k)$ and this is continued for all subbands.

Lastly, the "basic" algorithm assumes both sensed signals and the mixing matrix are real. However, they can be complex. Threfore, the algorithm needs to be modified to handle complex data. This is beneficial, not only for the convolutive case, but the IM case where $\mathbf{a}$ is complex.

4. INITIAL SYSTEM (MIXING MATRIX) ESTIMATE

In order for convolutive BSS to be possible, the method finding the initial estimate must be changed to allow for a complex mixing matrix. Assume that $\mathbf{A}(k)$ is complex. The ratio in polar form of the i^{th} and j^{th} rows for the n^{th} column is

$$\frac{A_{j,n}}{A_{i,n}} = \frac{|A_{j,n}|}{|A_{i,n}|} e^{\sqrt{-1}(\angle A_{j,n} - \angle A_{i,n})} \tag{7}$$

Using an argument similar to (4) that $S_l(k,m)$ is larger than all the other sources results in

$$\begin{aligned} \phi_{i,j} &= \angle X_j(k,m) - \angle X_i(k,m) \\ &= (\angle S_l(k,m) + \angle A_{j,l}(k)) \\ &\quad - (\angle S_l(k,m) + \angle A_{i,l}(k)) \\ &= \angle A_{j,l}(k) - \angle A_{i,l}(k) \end{aligned} \tag{8}$$

This shows that the estimation of $\mathbf{A}(k)$ requires two components - the ratios of the magnitudes of the $\mathbf{A}(k)$ elements and the difference in phase between the elements.

The remaining procedure is similar to the IM case, in that clustering is used to determine the initial estimate of $\mathbf{A}(k)$. Since ϕ is between 0 and 2π and θ is between 0 and $\pi/2$, ϕ is appropriately weighted in the clustering, so that the same amount of weight will be placed on the ϕ components as the θ components.

A value of ϕ that is slightly larger than 0 should be considered close to a value that is slightly less than 2π. If the phase difference is close to zero or 2π then the clustering algorithm could see two clusters. In order to avoid this possibility the histogram of ϕ is computed and the values are shifted so that the discontinuity will occur at a point that would not divide a cluster. An example scatter plot is shown in figure 1. Unfortunately, due to the ambiguity of scale factor, the actual phase values of the mixing matrix cannot be determined. However the use of the phase difference, ϕ greatly improves the robustness of the separation for convolutive mixtures and complex IM.

5. EXPERIMENTAL RESULTS

In order to test this method, the experimental setup shown in Figure 2 was used. This algorithm was tested for several cases involving different numbers of sensors and sources. This include two sensors tested with three and four sources,

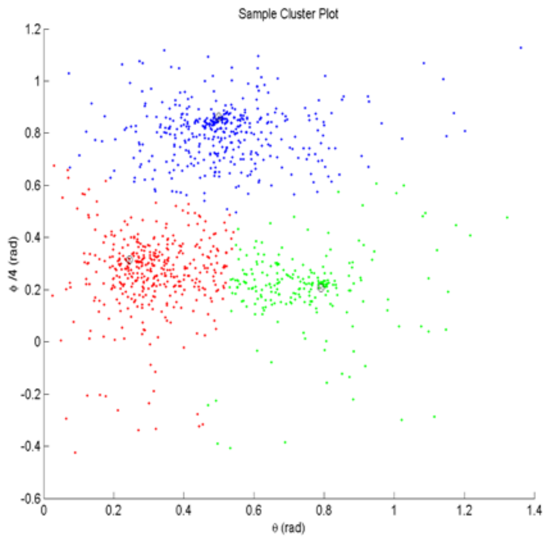

Fig. 1. Scatter plot of data for BSS of 3 sources using 2 sensors. The circled x shows the true value computed from the mixing matrix.

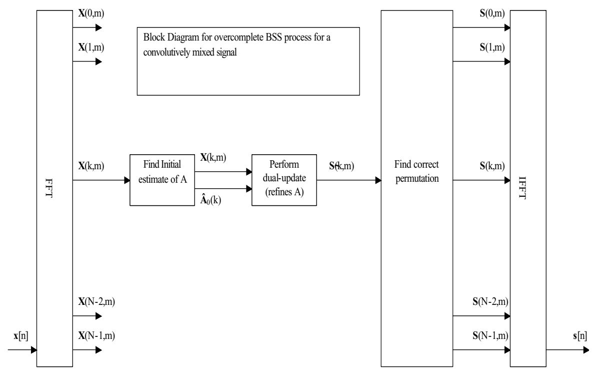

Fig. 2. A block diagram of the complete BSS system

and three sensors tested with four and five sources. The length of the filters used was five taps long and the filter taps were randomly distributed between -0.5 and 0.5 using a uniform distribution. The signals were mixed together with additive white Guassian noise with a variance of 0.01. The input signals had a variance of 1.0. The sources were short speech clips, including a child saying the alphabet, a woman counting in Spanish, a man counting in Ukrainian, and two women speaking in English. Each experiment was repeated 10 times each with randomly determined filter taps. In each experiment, the signals were reconstructed without using any knowledge of how they were mixed together and for comparison the signals were reconstructed using the known values for $\mathbf{a}$.

In order to evaluate the performance, a measure of the average SNR improvement was used. First the SNR of each source in each mixture was computed using the following equation.

$$SNR_{s_i,x_k} = 10\log_{10}(\frac{var(a_{k,j} * s_i)}{var(x_k - a_{k,j} * s_i)}) \qquad (9)$$

The average result is then subtracted from

$$SNR_{s_i} = 10\log_{10}(\frac{var(s_i)}{var(\hat{s}_i/c - s_i)}) \qquad (10)$$

where $\hat{s}_i = c(s_i + n_i)$ is the estimated source and c is a constant due to the ambiguity of scale. Of course, c is unknown and therefore it must be estimated. One method uses the ratio of the cross-correlation of s_i with $\hat{s}_i$ and the correlation of s_i. If the noise and source are independent and zero mean then the ratio is

$$\frac{R_{s_i,\hat{s}_i}}{R_{s_i,s_i}} = \frac{cR_{s_i,s_i}}{R_{s_i,s_i}} = c \qquad (11)$$

After c has been estimated, then the SNR of the estimated sources can be compared with the SNR of the sensors to show the average SNR improvement.

The tables show the resulting SNR improvement for 3 sensors and four sources and 3 sensors and five sources. The average SNR improvement for four sources was 5.3 dB and for five sources, it was 4.0 dB. When the knowledge of mixing system $\mathbf{a}$ was used to estimate the sources, then the improvement was 13.1 dB for four sources and 11.7 dB for five (not shown in table 2 due to lack of space).

Full BSS				**BSS with known a**			
s_1	s_2	s_3	s_4	s_1	s_2	s_3	s_4
1.6	11.3	8.0	10.0	7.7	17.4	13.0	14.9
6.0	2.9	6.3	5.5	13.7	12.9	14.8	9.2
2.6	4.4	6.5	8.5	8.8	8.6	15.0	20.1
4.2	2.4	3.2	2.5	17.3	14.0	9.5	10.7
3.9	6.9	4.3	6.6	12.9	13.6	13.2	13.4
8.0	9.6	0.7	3.9	15.6	15.0	13.7	12.2
7.8	8.2	6.7	5.4	12.3	20.3	13.2	10.8
8.6	1.6	5.5	5.9	15.0	10.6	14.1	12.7
0.3	6.7	2.1	10.0	8.3	16.5	8.2	13.8
6.9	-1.7	5.0	3.9	16.6	7.7	13.9	11.7

Table 1. The SNR improvements over the average SNR in dB of the sensor measurements for the case of three sensors and four sources.

6. CONCLUSION

An extension of our probabilistic approach for the BSS of underdetermined convolutive mixture is presented. This extension includes, capability to handle complex data and more than 2 mixtures, and robust estimation of initial mixing matrix and the number of sources. Further, the geometric constraint based search introduced in this paper speeded up our previous ”dual update” algorithm significantly. The method of inter-frequency correlation is also developed to find the correct permutations between frequencies. The experimental results indicate that this technique can efficiently separate convolutively mixed signals. However, there still exists resolving ambiguity of scale which is the topic of our future research. The good estimation of coherent set of scale factors will further improve performance of our method.

Full BSS				
s_1	s_2	s_3	s_4	s_5
0	3.6	-5.5	7.0	9.1
3.4	-0.7	4.5	1.4	11.3
5.8	3.8	6.0	4.1	5.4
4.2	0.6	7.3	-1.3	4.4
2.4	5.1	0.2	4.3	6.5
8.0	4.8	2.7	8.2	5.8
0.3	-0.6	-7.5	7.3	5.1
5.6	7.2	-1.1	2.2	10.3
3.3	4.2	2.4	6.5	8.2
8.1	8.1	-2.8	1.0	9.2

Table 2. The SNR improvements over the average SNR in dB of the sensor measurements for the case of three sensors and five sources.

7. REFERENCES

[1] Pau Bofill and Michael Zibulevsky, “Underdetermined blind source seperation using sparse representations,” *Signal Processing*, pp. 2353–2362, 2001.

[2] Te-Won Lee, Michael S. Lewicki, Mark Girolami, and Terrence J. Sejnowski, “Blind source seperation of more sources than mixtures using overcomplete representations,” *IEEE Signal Processing Letters*, vol. 6, no. 4, April 1999.

[3] A. Ossadtchi and S. Kadambe, “Over-complete blind source seperation by applying sparse decomposition and information theoretic based probabilistic approach,” in *Proceedings of the ICASSP*, 2000.

[4] Pau Bofill, “Underdetermined blind separation of delayed sound sources in the frequency domain,” Submitted to Neurocomputing, March 2001.

[5] Lucas Parra and Clay Spence, “Convolutive blind seperation of non-stationary sources,” *IEEE Transactions of Speech and Audio Processing*, vol. 8, no. 3, May 2000.

[6] Futoshi Asano and Shiro Ikeda, “Evaluation and realtime implementation of blind source seperation system using time-delayed decorrelation,” in *Proceedings of ICA*, Helsinki, 2000.