Joint Source Localization and Separation in Spherical Harmonic Domain Using a Sparsity Based Method

Sachin N Kalkur, Sandeep Reddy C, and Rajesh M Hegde

Indian Institute of Technology Kanpur
{kalkur, csreddy, rhegde}@iitk.ac.in

Abstract

In this paper, we address the problem of source localization and separation using sparse methods over a spherical microphone array. A sparsity based method is developed from the observed data in spherical harmonic domain. A solution to the sparse model formulated herein is obtained by imposing orthonormal constraint on the sparsity matrix. Subsequently, a splitting method based on bregman iteration is used to jointly localize and separate the sources from the mixtures of sources. A joint estimate of location and the separated sources is finally obtained after fixed number of iterations. Experiments on source localization and separation are conducted at different SNRs on the grid database. Experimental results based on RMSE analysis and objective evaluation indicate a reasonable performance improvement when compared to other methods in literature.

Index Terms: Source separation, Spherical harmonics, Bregman iteration.

1. Introduction

Multi-channel source separation attempts to separate $L$ unknown sources from mixtures obtained from an array of $I$ sensors. Source separation plays an important role in applications such as Automatic Speech Recognition (ASR) and Musical Information Retrieval (MIR) [1] and hence has been an active area of research.

Linear and circular arrays are commonly used geometric configuration of sensors in source separation. In recent decade, spherical microphone array have gained a lot of interest in study of acoustic signal processing, especially because of its geometry. 3-D symmetry of spherical microphone array is able to capture the spatial structure of acoustic field without any distortion [2] [3]. In this work, we address the problem of source separation using a spherical microphone array.

The problem of source separation has been addressed in various domains such as spatial [4], time [5] [6], frequency [7], time-frequency [8] [9]. The time-domain based separation relies on modelling the mixture as an instantaneous model and has been extensively studied. This formulation does not take into account reverberation. Consequently, frequency domain modeling employs a convolutive model and often tackles the problem of separation using Independent Component Analysis (ICA). Probabilistic methods [10] involve an assumption that the mixture coefficients are non-negative and use an iterative algorithm such as Non-negative Matrix Factorization (NMF). Time-frequency masking employs separation of sources based on their sparsity in the time-frequency domain. The efficiency of this technique depends on the design of mask. Spatial domain source separation utilizes the knowledge of locations of the sources to separate individual sources from the mixture of sources. Beamforming and adaptive beamforming are commonly used techniques. Generally, source separation methods in spatial domain is a two step process. Estimate source locations and subsequently compute the informations of the sources [11] by beamforming. In this work, we utilize both the simplicity of sparsity based methods and the accuracy of spatial domain method to provide an efficient solution to the problem of source localization and separation.

The pressure observed at the microphones of a spherical microphone array is expressed as a product of steering matrix and the strength of the sources in frequency domain [12] [13]. In general, the steering vector matrix is in spherical harmonic domain. The source locations are considered to be sparse and can be modelled using an overcomplete dictionary. The sparsity modelled presented in this work is developed on this principle. Subsequently to estimate both source locations and strengths, we follow an iterative approach i.e, a splitting method based on bregman iteration [14] [15]. Finally, to the proposed sparsity based method, performance evaluation is done as in [16] and the results are compared with [17].

The rest of the paper is organized as follows, Section 2 introduces the data model in spherical harmonic domain and describes the proposed method. Section 3 evaluates and compares the proposed method with existing methods and Section 4 provides conclusion and remarks on the proposed method.

2. Sparsity based method for joint source localization and separation in spherical harmonic domain

Sparsity based methods have been used extensively for blind source separation using linear and planar arrays [18]. Development of a sparse framework, in spherical harmonic domain has hitherto not been studied extensively. In this work, we develop a sparsity based method for joint source localization and separation in spherical harmonic domain. An introduction to the data model is first provided. Subsequently, development of sparse framework is described. A solution to the joint source localization and separation problem based on sparse model is then presented using a splitting method on bregman iteration.

2.1. Data model in spherical harmonic domain

Consider a spherical array with $I$ similar omni-directional microphones, radius $r$ and order $N$. The position of the $i^{th}$ microphone in spherical co-ordinates can be represented as $r_i=(r,\Omega_i)$, where $\Omega_i=(\theta_i,\phi_i)$. The elevation angle $\theta$ is mea-
sured downwards from the z-axis, the azimuth angle $\phi$ is measured counter clockwise from the z-axis. In this work, a sound field consisting of L sound fields in far-field with no multipath is considered. The position of $l^{th}$ source can be represented as $\Psi_l = (\theta_l, \phi_l)$ and corresponding wave vector is $k_l = (k_l \cos(\phi_l) \sin(\theta_l), k_l \sin(\phi_l) \sin(\theta_l), k_l \cos(\phi_l))^T$. Where $k = ||k_l|| = \frac{2\pi f}{c}$ and $f$ is the frequency associated to wave number. The pressure $p(k)$ on the spherical microphone array in frequency domain [12] can be expressed as:

$$p(k) = V(k, \Psi) s(k) + n(k), \quad \forall k$$

(1)

where $s(k) = [s_1(k), \ldots, s_L(k)]^T$ are source strengths in frequency domain. Each element of $s(k)$ is the zero-mean additive white gaussian noise. The noise $n(k)$ is uncorrelated with source strengths $s(k)$. The $I \times L$ steering matrix $V(k, \Psi)$ employs spatial characteristics of the array and represents the random impulse response, it can be expressed as shown below:

$$V(k, \Psi) = [v(k, \Psi_1), \ldots, v(k, \Psi_L)],$$

(2)

where $v(k, \Psi_l) = [e^{i k_l r_1}, \ldots, e^{i k_l r_I}]^T$ is the $1 \times I$ steering vector corresponding to $l^{th}$ source with each element in the steering vector corresponds to an unit plane wave. Applying spherical harmonic expansion to the steering vector matrix, the total pressure on the spherical microphone array [12] can be expressed as

$$p(k) = Y^H(\Omega) B(kr) Y^H(\Psi) s(k) + n(k), \quad \forall k$$

where $Y^H(\Psi) \in \mathbb{C}^{(N+1)^2 \times L}$ is the spherical harmonics matrix with angular positions corresponding to source locations and it is given by

$$Y^H(\Psi) = \left[ y_1^H, y_2^H, \ldots, y_L^H \right],$$

$$y_l = \left[ Y_{n_1}^0(\Psi_l), Y_{n_1}^{-1}(\Psi_l), \ldots, Y_{n_m}^0(\Psi_l) \right],$$

where $Y_{nm}^0$ is spherical harmonics function of order $n$ and degree $m$ is given by

$$Y_n^m(\theta, \phi) = \sqrt{\frac{(2n+1)(n-m)!}{4\pi(n+m)!}} P_n^m(\cos \theta) e^{im\phi},$$

where $P_n^m$ is the associated legendre polynomial. The matrix $Y(\Omega)$ is defined in a similar fashion as $Y^H(\Psi)$. $B(kr)$ is a $(N+1)^2 \times (N+1)^2$ diagonal matrix with each element corresponding to the mode strength $b_n(kr)$ for a rigid spherical microphone array and the matrix is defined as

$$B(kr) = \text{diag}(b_0(kr), b_1(kr), b_2(kr), \ldots, b_N(kr)),$$

$$b_n(kr) = 4\pi n! \left[ j_n(kr) - j_n^*(kr) \frac{h_n(kr)}{h_n^*(kr)} \right],$$

where $j_n$ and $h_n$ denote the spherical Bessel and Hankel functions respectively, $j_n$ and $h_n$ are their corresponding derivatives. The total pressure in spherical harmonic domain is

$$p_{nm} = B(kr) Y^H(\Psi) s(k) + Y^H(\Omega) \Gamma n(k), \quad \forall k$$

where $\Gamma$ are the sampling weights. Now, upon left multiplying $B^{-1}(kr)$ to the above equation we get,

$$x_{nm} = Y^H(\Psi) s(k) + n_1(k), \quad \forall k$$

where $n_1(k) = B^{-1}(kr) Y^H(\Omega) \Gamma n(k)$. For the sake of ease of notation we rewrite the above equation as shown below,

$$x(k) = A s(k) + n_1(k), \quad \forall k$$

(3)

where $x_{nm}$ and $Y^H$ is rewritten as $x(k)$ and $A$, respectively. Thus (3) is the data model for the joint source localization and separation problem in spherical harmonic domain.

### 2.2. Development of sparse framework for joint source localization and separation

Let us reconsider the data model in Equation (3). $A$ can be assumed to be sparse in spatial domain. $A$ contains spatial information of the sources in spherical harmonic domain. $A$ can be expressed as,

$$A = DR,$$

(4)

where $D \in \mathbb{C}^{(N+1)^2 \times d}$ is the overcomplete dictionary matrix with location uniformly sampled in azimuth $\phi$ and elevation $\theta$ and $d \gg L$ atoms in spherical harmonic domain and it is given by,

$$D = \left[ y_1^H, y_2^H, \ldots, y_L^H \right],$$

(5)

where,

$$y_d = \left[ Y_{n_1}^0(\Psi_d), Y_{n_1}^{-1}(\Psi_d), \ldots, Y_{n_m}^0(\Psi_d) \right],$$

and $R \in \mathbb{R}^{d \times L}$ is the sparse dictionary coefficient matrix. The size of the dictionary (number of atoms in a dictionary) is chosen based on spatial resolution required for the estimation of location of sources. Hence to jointly localize and separate the sources the following problem is formulated as,

$$\arg\min_{A, R} \| x - A s \|_2 + \lambda \| R \|_1$$

(6)

where $s, d A = DR$ and $s = (A^H A)^{-1} A^H x$

The problem is formulated as a standard minimization problem under the sparse framework. The above formulation is a non-convex problem. It is also important to note that the problem involves orthonormal constraints.

#### 2.2.1. Significance of the orthonormal constraint

One important property of spherical harmonic function is orthonormality. Spherical harmonics matrix is rectangular matrix. Hence with sampling weights, it is left orthogonal i.e,

$$Y^H Y = I,$$

(7)

$$YTY^H \neq I.$$

Utilizing this property in a sparse framework, we have $AA^H \approx I$. In order to satisfy this condition accurately, it is required that,

$$R^T R = I.$$  

(8)

It is also implicit in this discussion that $DD^H \approx I$ because $D$ is the dictionary matrix consisting of spherical harmonic functions as the elements. Hence (8) has been imposed as a constraint in the minimization program.
2.3. Solution using splitting method based on bregman iteration

In the previous section, it may be noted that the formula is orthonormally constrained. Thus, minimization problem is non-convex. Instead of relaxing the orthonormal constraints, the non-convex problem is solved by using a splitting method based on bregman iteration as in [15].

In order to provide an insight into the splitting method based on bregman iteration [15], let us consider an optimization problem with orthonormal constraints as

\[ X = \arg\min_X J(X) \quad \text{s.t.} \quad X^T Q X = I, \]

where \( J(X) \) is convex objective function and \( Q \succ 0 \) is a positive definite matrix. Introduce a new variable \( P \) such that \( P = QX \). Further, using bregman iteration, the minimization problem is divided into sub-problems and iteratively solved as given below,

1. \( X^n = \arg\min_{X \in \mathbb{R}^{M \times N}} J(X) + \frac{\alpha}{2} \|LX - P^{n-1} + B^{n-1}\|_F^2, \)
2. \( P^n = \arg\min_{P \in \mathbb{R}^{M \times N}} \frac{\alpha}{2} \|P - (LX^n + B^{n-1})\|_F^2, \) \hspace{1cm} (9)
3. \( B^n = B^{n-1} + LX^n - P^n. \)

The second subproblem (9) has a closed form solution as discussed in [15]. The closed form solution given by

\[ P^n = FE^{-1/2}E^T, \] \hspace{1cm} (10)

where \( F = LX^n + B^{n-1} \). \( E \) is the orthonormal matrix and \( T \) is the diagonal matrix satisfying the SVD factorization of \( F^T F = ETET^T \).

2.3.1. Splitting method based on bregman iteration for source separation

In this section, we utilize the iterative method as in previous subsection to jointly localize and separate the sources. Also it is very important to note that the minimization problem is solved only for the initial frequency bin. Consequently, the results obtained from initial frequency bin is used for other frequency bins iteratively. Consider the problem in sparse framework as in (6). Bregman method, can be utilized to solve the problem (6) in an iterative fashion as shown below:

\[ (A_n, R^n) = \arg\min_A C(A) \quad \text{s.t.} \quad P^n \in DR, \]

\[ P^n = \arg\min_{P \in \mathbb{R}^{M \times N}} \frac{\alpha}{2} \|P - (R^n + B^{n-1})\|_F^2, \]

\[ \text{s.t.} \quad P^n T P = I, \]

\[ B^n = B^{n-1} + R^n - P^n, \]

\[ s^n = (A_n^H A_n)^{-1} A_n^H x, \]

where \( C(A) \) is a convex objective function given by,

\[ C(A) = \|x - As^{-1}\|_2 + \lambda \|R\|_1 + \frac{\alpha}{2} \|R - P^{-1} + B^{n-1}\|_2^2. \]

Also (11) has a closed solution as described in the previous section, \( P^n = FE^{-1/2}E^T \), where \( F = R^n + B^{n-1} \). The orthonormal matrix \( E \) and diagonal matrix \( T \) satisfies the SVD condition i.e., \( F^T F = ETET^T \). Now the obtained \( A_n \) is further used to compute \( s(k) \) for all values of \( k \).

Accuracy of the obtained solution is increased by adding additional linear constraints to the first subproblem in (11). The constraint is an implication of a property of spherical harmonic function \( Y_{\lambda \phi}^{m} \) that the value of \( Y_{\lambda \phi}^{m} \) irrespective of the angles is a constant. Hence the additional constraint that can be applied is formulated as

\[ -\epsilon 1^T < (A_n^T e_1 - 1^T Y_{\lambda \phi}^{m}) < \epsilon 1^T, \]

where \( e_1 = [1, 0, 0, ..., 0]^T \) is a \((N + 1)^2 \times 1 \) vector and \( 1^T \) is a column vector containing all 1’s. The final minimization problem (11) is given by,

\[ (A_n, R^n) = \arg\min_{A_n} C(A) \quad \text{s.t.} \quad A_n = DR, \]

\[ -\epsilon 1^T < (A_n^T e_1 - 1^T Y_{\lambda \phi}^{m}) < \epsilon 1^T. \]

Subsequently, algorithm 1 provides a brief listings of the steps involved in joint source localization and separation. The parameters herein \( \epsilon, \lambda, \alpha \) are chosen based on a line search. Also a fixed number of iteration of the algorithm 1 is carried out to obtain the required results. Hence the algorithm computes the localization parameters \( A_n \) and individual source strengths by estimating \( s(k) \).

Algorithm 1 Algorithm for source localization and separation using splitting method based on bregman iteration

1. Obtain data at each microphones of a spherical microphone array as in equation (4).
2. Create an overcomplete dictionary \( D \) with required resolution in \( \theta \) and \( \phi \).
3. Consider the model data corresponding to the frequency bin with wavenumber \( k \), satisfying \( kr > 1 \).
4. Choose the optimization parameters \( \lambda, \alpha \) and \( \epsilon \).
5. Initialize prior \( s^0 = x(0), x(0) - x^*(0)^T \).
6. Initialize \( P^0 \) to a random matrix obtained from a normal distribution and \( B^0 \) to all zero matrix.
7. for each \( n \in N \) do
8. Solve the minimization problem (12) to get \( A_n \).
9. Assign \( F = R^n + B^{n-1} \).
10. Compute SVD of \( F^{-1} P^0 = ETET^T \).
11. Compute \( P^n = F^{-1} ET^{-1/2}E^T \).
12. Compute \( B^n = B^{n-1} + R^n - P^n \).
13. Source strength at \( k^0 \) wavenumber is given by, \( s^n = (A_n^H A_n)^{-1} A_n^H x \).
14. end for
15. for each vector \( r_i \in R_L \) and \( i \in L \) do
16. Compute the maxima, which corresponds to respective source location.
17. end for
18. Multiply, the pseudo inverse of \( A_n \) to the model data to all frequency to obtain STFT of the separated sources.
19. Compute the inverse STFT to obtain individual sources in the mixture.

3. Performance Evaluation

In this section, the proposed method for source localization and separation is evaluated using datasets obtained from a spherical microphone array. Spectrographic analysis and localization error analysis are also presented.
3.1. Experimental conditions

In this work, we utilize speech data from GRID corpus [19] in the experiments. A rigid spherical microphone array as used in [20] is utilized to simulate the pressure at spherical array. The array consists of 50 microphones with radius of the sphere is 9.75 cm and order of the microphone array is $N = 3$. The dictionaries used for the experiments are created using a resolution of 2 degree in azimuth and 2 degree in elevation. Equation (12) is solved using a numerical method [21][22]. The experiments are conducted for 2 speaker scenario with two uncorrelated speech sources at locations $\Psi_1 = (20^\circ,60^\circ)$ and $\Psi_2 = (30^\circ,45^\circ)$ respectively. The sensor noise is considered to be white and uncorrelated Gaussian random variable.

3.2. Experiments on joint source localization and separation

In this section source localization and separation results are discussed. The proposed method is called as SB (Splitting method on Bregman iteration) method. Spectrograms of true signals and estimated signals using the proposed method are given in Figure 1. From the results, we observe that artifacts in the reconstructed signals are minimal. A satisfactory separation of signals is produced. Subsequently, the source localization is illustrated in figure 2 by plotting the magnitude of the sparse matrix with elevation angle. The columns of the sparsity matrix $R$ correspond to the locations of sources. The maxima in each column vector indicate individual source locations, which is clearly noted in Figure 2.

The efficiency of the SB method is evaluated by computing RMSE in localization and objective evaluations like logspectraldistance (LSD) and Perceptual evaluation score (PESQ) for separation of sources. RMSE analysis for source elevation estimation is illustrated in Figure 3. In this analysis, we observe that error in localization using the SB method is reasonably better than the standard methods of source localization. Quality of separation of sources by using SB method is evaluated using LSD and PESQ. LSD and PESQ scores are obtained by comparison of the clean speech and reconstructed speech. The scores provided in Table 1 are the average of the scores obtained for each speaker. An improvement in separation of sources compared to the proposed method [17] is observed.

In this work, a novel method for joint source localization and separation over a spherical microphone array using sparsity based methods is proposed. Then a solution to the problem is provided using a splitting method based on bregman iteration. Spectrographic analysis and objective evaluations indicate the efficiency of the proposed method. In the future work, the problems of permutation and scaling in source separation can be addressed. Incporation of the time-frequency sparsity in the model can also be investigated as a part of future work.

<table>
<thead>
<tr>
<th>SNR (dB)</th>
<th>LSD PESQ</th>
<th>LSD PESQ</th>
<th>LSD PESQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1.34 2.47</td>
<td>1.19 2.62</td>
<td>1.009 2.89</td>
</tr>
<tr>
<td>5</td>
<td>1.39 1.89</td>
<td>1.33 2.17</td>
<td>1.21 2.32</td>
</tr>
</tbody>
</table>

Table 1: Objective evaluation of source separation using LSD and PESQ.
5. References


