BLIND SOURCE SEPARATION USING MONOCHANNEL OVERCOMPLETE DICTIONARIES

B. Vikrham Gowreesunker and Ahmed H. Tewfik

University of Minnesota
Dept. of Electrical and Computer Engineering
200 Union Street SE. Minneapolis, MN 55455

ABSTRACT

We propose a new approach to underdetermined Blind Source Separation (BSS) using sparse decomposition over monochannel dictionary atoms and compare it to multichannel dictionary approaches. We show that the new approach is easily extended to any single channel decomposition method and allows for faster computation of algorithms such as the Bounded Error Subset Selection (BESS) because of the reduced dimension of the search space. Experimental results on Matching Pursuit (MP) and BESS algorithms show that our method can give better Signal to Interference Ratio performance than pursuit methods based on multichannel dictionary atoms.

Index Terms— Underdetermined Blind Source Separation, Sparse Decomposition, Bounded Error Subset Selection

1. INTRODUCTION

In the blind source separation (BSS) problem, we have mixtures of several source signals and the goal is to separate them with as little prior information as possible, hence the term blind. In this work, we study the instantaneous underdetermined BSS case, where we have more sources than mixtures. We are concerned with separating mixtures of speech signals when the mixing matrix and number of underlying sources are unknown. This problem is ill-defined and its solution requires some additional assumptions compared to its overdetermined counterpart. The difficulty of the underdetermined setup can be somewhat alleviated if there exists a representation wherein all the sources are rarely simultaneously active, which entails finding a representation where the sources are sparse. Some authors have shown that speech signals are sparser in the time-frequency than in the time domain [1], and that there exists several other representations such as wavelets packets, where different degrees of sparsity can be obtained. It has been shown that better separation can indeed be achieved by exploiting such sparsity [2].

In this paper, we investigate methods for performing BSS using overcomplete dictionaries. The fundamental success of the separation depends on two factors, namely the type of dictionary used and the type of decomposition method employed. In [3], we studied dictionary design methods, and showed the performance improvement good dictionaries offer. In this paper, we introduce a new approach to sparse decomposition (SD) for multichannel BSS. Commonly known SD algorithms like Matching Pursuit (MP) [4], Basis Pursuit (BP) [5], or Bounded Error Subset Selection (BESS) [6] were designed for single channel problems and cannot always be directly applied to multichannel data. One popular approach to the multichannel BSS is to use MP over so called multichannel dictionaries[7], a method hard to extend to BP or BESS because of the prohibitive cost of the expanded search space. In this work, we propose an algorithm that can easily be combined with most single channel decompositions available in the literature.

Our proposed method searches the best dictionary atoms that can represent a linear combination of the mixtures, and use these atoms to find a representation of each mixture. Because of the flexibility of our approach, we can also investigate how well different SD algorithms and their corresponding formulations recover the underlying sparsity of the sources. We combine our method with a single channel Matching Pursuit and BESS and compare their performance with the Matching Pursuit with multichannel dictionary for a 2 channel case[7]. In section 2, we give a mathematical description of the problem and an explanation of how sparsity is used in source separation, in section 3 we describe the motivation behind our approach and give a description of our proposed algorithm. In section 4, we compare the performance improvement of our approach with a class of multichannel decomposition methods.

2. MIXING MODEL AND SPARSITY

In this section, we formulate the instantaneous mixing model for underdetermined Blind Source Separation. Given M mixtures of N sound sources such that $M \leq N$, our goal is to recover the underlying sources up to a scale factor. For a data frame of T samples, we can represent the mixture X, as an $M \times T$ matrix which is the result of the product of an unknown

mixing matrix, A and the $N \times T$ source matrix S. Matrix A is an $M \times N$ matrix with each column corresponding to the direction of arrival of one source. Without loss of generality, we assume that 2 mixtures and 3 sources are available. The m^{th} mixture can be represented as follows,

$$x_m = \sum_{n=1}^{N} a_{m,n} s_n, \tag{1}$$

where s_n is vector corresponding to the n^{th} source, and $a_{m,n}$ corresponds to the $(m,n)^{th}$ entry of matrix A. We can see that for $M < N$, we have fewer equations than unknowns.

If we assume that only the n^{th} source is active, we find that the ratio x_1 to x_2 is the ratio of the mixing matrix column, $a_{1,n}/a_{2,n}$. In fact, if the sources are sufficiently sparse such that they never overlap, a scatter plot of x_1 against x_2 will reveal 3 clear lines, whose gradients corresponds to $a_{1,n}/a_{2,n}$. The points on each line belong to only one source, and can be separated. Evidently, in the time domain, the sources are not sparse enough for such separation to be done. However, several authors have found that sound signals exhibits very high sparsity in alternative representations such as time-frequency [1], and have successfully exploited this in algorithms such as DUET [8]. We now extend the above model for overcomplete dictionaries.

Given an overcomplete dictionary, $D = \{d_k\}_{k=1}^{K}$, such that each atom d_k is a $T \times 1$ vector and $K \gg T$, we can represent the n^{th} source vector as,

$$s_n = \sum_{k=1}^{K} c_{n,k} d_k, \tag{2}$$

where $c_{n,k}$ is the coefficient associated to the k^{th} dictionary atom. Subsequently, we can represent the mixture signals in terms of the source signals, dictionary atoms, mixing matrix columns and associated coefficients. The resulting representation is,

$$x_m = \sum_{n=1}^{N} \sum_{k=1}^{K} a_{m,n} c_{n,k} d_k. \tag{3}$$

The representation in eq. (2) is not unique and some sparse decomposition algorithm can be used to find a representation with few non-zero coefficients.

3. MULTICHANNEL DECOMPOSITIONS

3.1. Sparse Decomposition for Multichannel data

For the purposes of multichannel BSS, we seek a representation for each mixture such that the sources are as sparse as possible. However, the authors of [3] have found that the independent decomposition of the mixtures result in representations where the mixtures do not always share the same dictionary elements. Under these conditions, it is not possible to separate the sources. The authors of [7] have proposed the use of multichannel dictionary atoms in conjuction with a Matching Pursuit type of algorithm. At each iteration, this method simultaneously decomposes all the mixtures over the same dictionary atom and finds the atom that is best correlated to all the mixtures. For the two channel case, we refer to this as the MP with stereo dictionary. This idea was also extended to directional dictionaries where $\mathbf{a}_n d_k$ is the dictionary atom in the direction of the n^{th} column of the mixing matrix. For both stereo dictionary and the directional extension, the mixtures will share the same dictionary atoms and standard separation methods can be applied to the coefficients. However, the search space for multichannel dictionaries is larger than for single channel dictionaries. We find that these methods are not always easily extended to all sparse decomposition methods, in part due to the higher complexity of a larger search space, which gets worse as the dimension of the problem increases.

3.2. Proposed Method Using Monochannel Dictionaries

One can think of the problem of finding the best representation for source separation as identifying a subset of dictionary atoms common to both mixtures where the underlying sources live disjointly. If such a subset was known, the mixtures' representation could be easily found by least square approximation on this reduced dictionary, and the coefficient could be separated by some standard separation method. Of course, such a subset of dictionary atoms is not known in practice and we seek the next best option, which is the subset of dictionary atoms that can best represent both mixtures. As mentioned earlier, the independent decomposition of the mixtures give representations that are not well suited for both mixtures. We now illustrate how to use a monochannel dictionary to find the good set of dictionary atoms to represent the mixtures.

Let M_λ be the linear combination of the two mixtures such that,

$$M_\lambda = x_1 + \lambda x_2, \tag{4}$$

where λ is a predetermined constant. We can express M_λ as,

$$M_\lambda = \sum_{n=1}^{N} \sum_{k=1}^{K} (a_{1,n} + \lambda a_{2,n}) c_k d_k. \tag{5}$$

The parameter λ affects the weighing of the sources in M_λ. For example, if we compare eq. (1) and (4), we find that by setting $\lambda = -a_{1,n}/a_{2,n}$, we can cancel the n^{th} source from M_λ. By looking at the sparse representation of different linear combination of the mixtures, we can identify the most frequently occurring dictionary atoms as those most likely to represent the underlying source. However, for computational reasons, we would like to limit the number of realizations of λ to a minimum. Given one instance of λ, we can find the subset of dictionary atoms to represent M_λ using some single channel decomposition method such as BESS, MP, or BP. If multiple instances of M_λ are available, we can take the most

frequently occurring atoms for the sparse representations of all the values of λ. Once the atoms are determined, a least square approximation can be used to represent each mixture. The final representation of the mixtures will share the same dictionary atoms and if the sources were sparse enough, the mixture coefficients will be clustered along the mixing matrix columns, and can be separated. In section 4, we show results for this approach and compare it with stereo dictionary decomposition. A summary of the algorithm is given in table 1.

Table 1. Algorithm Overview

1. Choose λ.
2. Create linear combination of mixtures, $M_\lambda = x_1 + \lambda x_2$.
3. Apply sparse decomposition algorithm on (2) such that $M_\lambda = \sum_{i \in A} c_i d_i$, where $A = \{i : c_i \neq 0\}$.
4. Defined reduced dictionary, $D_r = \{d_i\}_{i \in A}$.
5. Find representation of each mixture, x_m over reduced dictionary by least square approximation. $x_1 = \sum_{i \in A} c_{1,i} d_i$, and $x_2 = \sum_{i \in A} c_{2,i} d_i$
6. Separate coefficients by projecting coefficient pairs of $(c_{1,i}, c_{2,i})$ onto closest mixing matrix column.[3, 7]

3.3. Bounded Error Subset Selection for Source Separation

In [6], BESS has been shown to have an enumerated solution with polynomial complexity and better rate distortion than most commonly known decomposition methods. However, for multichannel dictionaries, as the number of channel increases this search space becomes more prohibitive. In this work, we have successfully applied the monochannel dictionary method for BESS and the results are presented in section 4. Computational savings over an augmented dictionary should be evident for an enumerated approach such BESS.

4. EXPERIMENTS AND RESULTS

4.1. Dictionary Atom Selection of Monochannel Method

We devised the following experiment to find out how our method picks the dictionary atoms out of a mixture. We created 3 random artificial signals from a linear combination of C_x dictionary atoms, and mixed them using some known mixing matrix. We found the dictionary atoms used to represent 3 instances of M_λ with $\lambda = -a_{1,n}/a_{2,n}$, and $n = 1, 2, 3$ and took the union of the dictionary atoms for these representations. We then computed the percentage of correct atoms recovered by our method, relative to the total known atoms used by the sources. We also found the number of wrong atoms picked by our methods and computed its percentage relative to the total known source atoms. Results for one such experiment is plotted in the top and bottom portion of fig.1 respectively. It has to be noted that the sum of correct and wrong atoms can be larger than the original number of atoms used

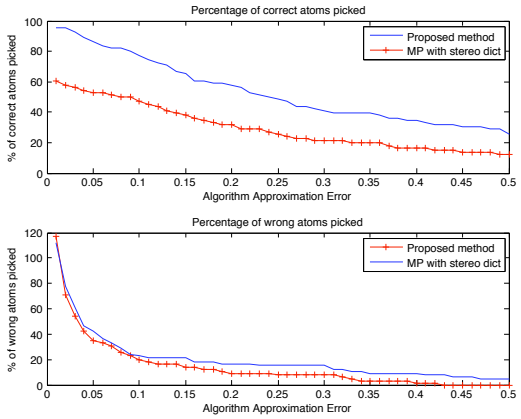

Fig. 1. Plot of % of correct source atom (top fig) and % of incorrect source atoms (bottom fig) that appear in mixture representation. % is relative to total atoms that is known(apriori) to represent sources.

to represent the sources. We compared these features against the Matching Pursuit using stereo dictionary, and found that the monochannel case consistently picks a larger percentage of original dictionary atoms. Not shown here for space consideration is that the sparser the original signal, the better our method is at picking the sources' atoms. This supports our earlier claim that this method does a better job at picking atoms to represent the underlying sources. Furthermore, we find that as the approximation error of the algorithm decreases below 0.1, we get an exponential increase in "wrong" dictionary atoms being picked. We find that source separation performance improves when we pick a higher percentage of correct atoms and a lower percentage of "wrong atoms". Below, we compare separation performance of our method against the MP with stereo dictionaries, for speech mixtures.

4.2. Performance Evaluation

We evaluated the performance of our proposed approach on the blind separation of 3 second tracks of 2 mixtures of 3 male speakers sampled at 16 Khz. These were artifically mixed with an instantaneous mixing matrix. The dictionary used was a KSVD trained dictionary[9], with 4096 atoms of size 512×1. We evaluated three variations of our algorithm. The first case, which we refer to as "Type-1 MP with mono dictionary," is a straightforward implementation of the algorithm in table 1, with $\lambda = 1$, and Matching Pursuit sparse decomposition method used. The second case, which we refer to as "Type-2 MP with mono dictionary" is a modified version of the algorithm in table 1. In this case, we pick 3 values of $\lambda_n = -a_{1,n}/a_{2,n}$, with $n = 1, 2, 3$, and for each case we repeat steps 2-4 from table 1. We take the union of the bases found in step 4 for all 3 instances of λ, and define this as

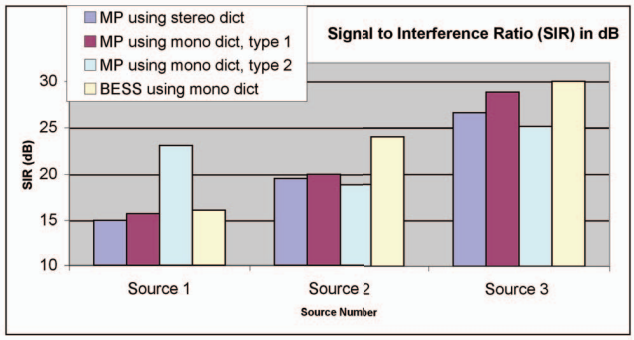

Fig. 2. SIR comparison between 3 mono dictionary methods and a stereo method. We see that type-2 MP and BESS perform the best.

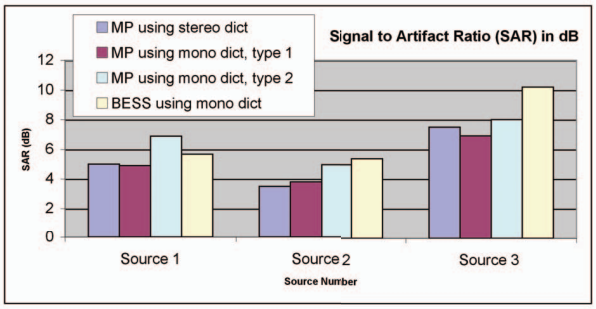

Fig. 3. SAR comparison between 3 mono dictionary methods and a stereo dictionary method. We see that type-2 MP and BESS perform the best.

our reduced dictionary. Steps 5-6 are done over this reduced dictionary. The third case, is called "BESS with mono dictionary", is a direct implementation of the algorithm in table 1 with $\lambda = 1$ and using BESS sparse decomposition method. The three approaches were compared with MP over a stereo dictionary. For fair comparison, we assume the mixing matrix is known for all the methods, and use coefficient space partitioning to separate the coefficients once we have the representation in step 5[7]. The performance of the separated signals were evaluated using the BSS-EVAL toolbox [10] and its two main metrics, Signal to Interference Ratio (SIR) and Signal to Artifact Ratio (SAR) are plotted in fig. 2 and fig. 3.

4.3. Discussion

As can be seen in fig. 2, we get good results for all three monochannel methods for SIR and SAR. Of particular interest is the concurrent improvement in SAR for the BESS algorithm, and the type-2 monochannel MP. Artifacts remain an area where much improvements is still needed for sparse decomposition methods, and it common to have improvement in SIR at the expense of SAR. We find that our approach has indeed opened new avenues for improving performance for both metrics. We find that using this method with only one λ works well with enumerated approaches such as BESS, but does not necessarily provide much improvement for an MP algorithm. However, by using a few strategically picked values of λ, we get a notable improvement in performance even with MP. Not shown in this paper for space consideration is that the type-2 MP method is very robust to mixing matrix estimation errors. We are currently working on computationally efficient ways to take advantage of more combinations of λ.

5. CONCLUSION

We have proposed a new approach to multichannel sparse decomposition using monochannel dictionaries and successfully applied it to underdetermined BSS. The new approach allowed us to overcome some computational hurdles that previously made algorithms such as BESS unattractive for separation. We showed that BESS can offer significant improvement in performance metrics such as interference rejection and level of artifacts. Also, we demonstrated that there is a lot of promise with this type of sparse decomposition method if λ is well chosen. We are currently working on formalizing this approach to find how to optimally pick this parameter.

6. REFERENCES

[1] P. Bofill and M. Zibulevsky, "Underdetermined blind source separation using sparse representation," *Signal Processing*, vol. 81, no. 11, pp. 2353–2362, 2001.

[2] M. Zibulevsky, B.A Pearlmutter, P. Bofill, and P Kisilev, "Blind source separation by sparse decomposition," chapter in the book: S. J. Roberts, and R.M. Everson eds., Independent Component Analysis: Principles and Practice, Cambridge, 2001.

[3] B. Vikrham Gowreesunker and Ahmed H. Tewfik, "Two improved sparse decomposition methods for blind source separation," in *Independent Component Analysis and Signal Separation (ICA)*, London, UK, September 2007, vol. 4666, pp. 365–372.

[4] S. Mallat and Z. Zhang, "Matching pursuit with time-frequency dictionaries," *IEEE Trans. on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[5] S. Chen and D. Donoho, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[6] Masoud Alghoniemy and Ahmed H. Tewfik, "Reduced complexity bounded error subset selection," in *IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, March. 2005, pp. 725–728.

[7] S. Lesage, S. Krstulovic, and R. Gribonval, "Under-determined source separation: Comparison of two approaches based on sparse decomposition," in *Independent Component Analysis and Signal Separation (ICA)*, March 2006, pp. 633–640.

[8] O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," *IEEE Trans. on Signal Processing*, vol. 52, no. 7, pp. 1830–1847, July. 2004.

[9] M. Aharon, M. Elad, and A.M. Bruckstein, "The k-svd: An algorithm for designing of overcomplete dictionaries for sparse representation," *IEEE Trans. on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, November. 2006.

[10] E. Vincent, C. Fevotte, and R. Gribonval, "Performance measurement in blind audio source separation," *IEEE Trans. Audio, Speech and Language Processing*, vol. 14, no. 4, pp. 1462–1469, 2006.