A sparsity-based method to solve the permutation indeterminacy in frequency domain convolutive blind source separation

Prasad Sudhakar, Rémi Gribonval

A sparsity-based method to solve permutation indeterminacy in frequency-domain convolutive blind source separation

Prasad Sudhakar and Rémi Gribonval

METISS Team,
Centre Recherche INRIA Rennes - Bretagne Atlantique, Campus de Beaulieu, 35042 Rennes cedex, France
{firstname.lastname}@irisa.fr
http://www.irisa.fr/metiss

Abstract. Existing methods for frequency-domain estimation of mixing filters in convolutive blind source separation (BSS) suffer from permutation and scaling indeterminacies in sub-bands. However, if the filters are assumed to be sparse in the time domain, it is shown in this paper that the ℓ_1-norm of the filter matrix increases as the sub-band coefficients are permuted. With this motivation, an algorithm is then presented which solves the source permutation indeterminacy, provided there is no scaling indeterminacy in sub-bands. The robustness of the algorithm to noise is also presented.

Key words: convolutive BSS, permutation ambiguity, sparsity, ℓ_1-minimization

1 Introduction

The need to separate source signals from a given set of convolutive mixtures arises in various contexts. The underlying model of having M mixtures $x_m(t), m = 1 \dots M$ from N source signals $s_n(t), n = 1 \dots N$, given a discrete time index t, is given by

$$x_m(t) = \sum_{n=1}^{N} \sum_{k=0}^{K-1} a_{mnk} s_n(t-k) + v_m(t) \tag{1}$$

with $v_m(t)$ the noise term. Concisely, it can be written in the matrix notation as

$$\mathbf{x}(t) = \sum_{k=0}^{K-1} \mathbf{A}_k \mathbf{s}(t-k) + \mathbf{v}(t) \tag{2}$$

where $\mathbf{x}(t), \mathbf{v}(t)$ are $m \times 1$ vectors, $\mathbf{A}_k$ is an $M \times N$ matrix which contains the filter coefficients at k^{th} index. The notation $\mathbf{A}_{mn}(t) = a_{mnt}$ will also be used for each mixing filter, which is of length K. The ultimate objective of a BSS system is to recover back the original source signals $s_n, n = 1 \dots N$ given only the mixtures $x_m(t), m = 1 \dots M$.

A standard approach is to first estimate the mixing matrix $\mathbf{A}(t) = (\mathbf{A}_{mn}(t))$ $m = 1 \dots M, n = 1 \dots N$ and then recover the sources $s_n(t)$. This paper focusses on the estimation of the mixing matrix.

1.1 Permutation problem description

Several methods have been proposed by the signal processing community to estimate the mixing matrices in convolutive BSS. Pedersen et. al. present an excellent survey of the existing methods [1]. Broadly, the techniques can be classified into time-domain and frequency-domain techniques. Both approaches have their own advantages and disadvantages. They are summarized in table 3 of [1].

The context of our problem arises in the frequency-domain approach. A survey of these techniques is provided in [2]. The main advantage of frequency-domain techniques is that the convolutive mixture case is transformed (under the narrowband assumption) into complex-valued instantaneous mixture case for each frequency bin:

$$\mathbf{x}(f,t) = \mathbf{A}(f)\mathbf{s}(f,t) + \mathbf{v}(f,t) \tag{3}$$

where $f = 1 \dots F$ are the sub-band frequencies.

The central task of frequency-domain mixing matrix estimation techniques is to provide an estimate $\hat{\mathbf{A}}(f)$ of $\mathbf{A}(f)$. However, the frequency-domain approach suffers from permutation and of scaling indeterminacies *in each sub-band* f. Specifically, the estimated $\hat{\mathbf{A}}(f)$ is related to the true filter matrix $\mathbf{A}(f)$, for each f in the following form

$$\hat{\mathbf{A}}(f) = \mathbf{A}(f)\mathbf{\Lambda}(f)\mathbf{P}(f) \tag{4}$$

where $\mathbf{P}(f)$ is the frequency-dependent permutation matrix, $\mathbf{\Lambda}(f)$ is a diagonal matrix containing the arbitrary scaling factors.

The frequency-domain methods have to invariably solve the permutation and scaling indeterminacy to eventually estimate $\mathbf{A}(t)$ up to a unique global permutation and scaling $\hat{\mathbf{A}}(t) = \mathbf{A}(t)\mathbf{\Lambda}\mathbf{P}$.

1.2 Existing approaches to solve the described problem

There are mainly two categories of approaches to solve the permutation indeterminacy in the sub-bands of the estimated mixing filters [1].

The first set of techniques use consistency measures across the frequency sub-bands of the filters to recover the correct permutations, such as inter-frequency smoothness, etc. This category also includes the beamforming approach to identify the direction of arrival of sources and then adjust the permutations [3].

The second set of techniques use the consistency of the spectrum of the recovered signals to achieve the same. The consistency across the spectrum of the recovered signals is applicable for only those signals which have strong correlation across sub-bands, such as speech [4].

Different definitions of consistency have been used to propose different methods. Table 4 in [1] contains a categorization of these approaches.

1.3 Proposed approach: Sparse filter models

Here we propose to use a special type of consistency that can be assumed on the mixing filters: sparsity. That is, the number S of non-negligible coefficients in each filter $\mathbf{A}_{mn}(t)$ is significantly less than its length K. Acoustic impulse responses have few reflection paths relative to its duration, and hence the sparsity of mixing filters is a realistic approximation.

The idea behind our approach is that the permutations in the sub-bands decreases the sparsity of the reconstructed filter matrix $\hat{\mathbf{A}}(t)$. So, one can solve the permutations by maximizing the sparsity.

The sparseness of the filter matrix is measured by its ℓ_0-norm, defined as $||\mathbf{A}(t)||_0 = \sum_{m=1}^{M} \sum_{n=1}^{N} \sum_{t=1}^{K} |\mathbf{A}_{mn}(t)|^0$. Lesser the norm, sparser is the filter matrix. However, the ℓ_1-norm of the filter matrix defined by

$$||\mathbf{A}(t)||_1 = \sum_{m=1}^{M} \sum_{n=1}^{N} \sum_{t=1}^{K} |\mathbf{A}_{mn}(t)| \tag{5}$$

is also a sparsity promoting norm. It is shown by Donoho that working with ℓ_1-norm is as effective as working with ℓ_0-norm while looking for sparse solutions to linear systems [5]. Furthermore ℓ_1-norm being convex, has certain computational advantages over the former and is robust to noise.

We concentrate on the permutation problem and will assume that the estimated filter sub-bands are free from scaling indeterminacy. That is

$$\hat{\mathbf{A}}(f) = \mathbf{A}(f)\mathbf{P}(f), f = 1 \dots F. \tag{6}$$

Hence, when $\hat{\mathbf{A}}(t)$ is reconstructed using $\hat{\mathbf{A}}(f)$, given by Eqn. (6) one can expect an increase in $||\hat{\mathbf{A}}(t)||_1$. In Sec. 2 we experimentally show that this is indeed the case.

Then in Sec. 3 we show that if there is no scaling indeterminacy, a simple algorithm which minimizes the ℓ_1-norm of the filter matrix solves the permutation indeterminacy, even under noisy conditions.

2 Source permutations and ℓ_1-norm

This section presents some preliminary experimental study on how the ℓ_1-norm of the filter matrix $\mathbf{A}(t)$ is affected by permutation in the sub-bands.

For experimental purposes, 50 different filter matrices were synthetically created with $N = 5$ sources and $M = 3$ channels, and each filter having a length of $K = 1024$ with $S = 10$ non-zero coefficients. The non-zero coefficients were i.i.d. Gaussian with mean zero and variance one. The locations of non-zero coefficients were selected uniformly at random. Then for each such instance of filter matrix $\mathbf{A}(t)$ the discrete Fourier transform (DFT) $\hat{\mathbf{A}}_{mn}(f)$ was computed for each filter $\mathbf{A}_{mn}(t)$, to obtain the frequency-domain representation $\hat{\mathbf{A}}(f)$ of $\mathbf{A}(t)$. These filters were used in the following two kinds of experiments.

2.1 Random source permutations

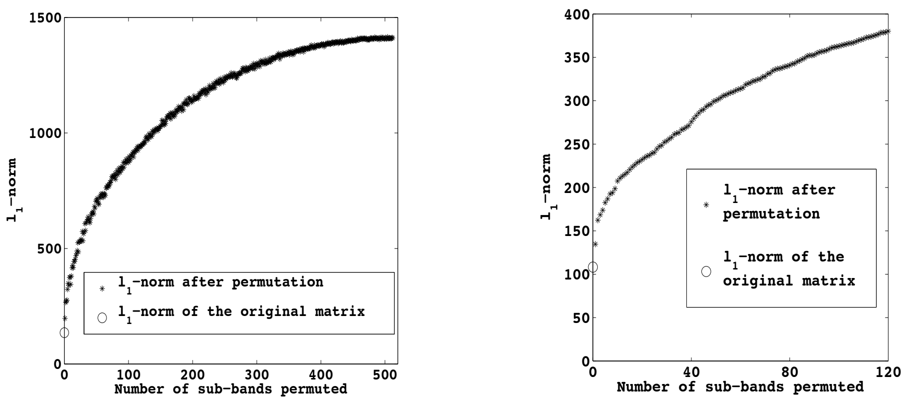
In a practical scenario, each sub-band can have a random permutation. So, for each filter matrix in the frequency domain, the sources were permuted randomly in an increasing of number of sub-bands (chosen at random), and their ℓ_1-norms were computed. The positive and negative frequency sub-bands were permuted identically.

For one such experimental instance, Fig. 1(a) shows the variation of ℓ_1-norm against increasing number of randomly permuted sub-bands. The circle shows the norm of the true filter matrix. Each star represents the norm after randomly permuting the sub-bands at random locations. Note the gradual increase in the norm as the number of sub-bands being permuted increases. Similar experiments were conducted with combinations of $M = 3, N = 4$ and $M = 2, N = 3$ and $S = 10, 15, 20$, leading to similar observations.

2.2 Sensitivity of ℓ_1-norm to permutations

In order to show that even a *single* permutation in only one sub-band can increase the norm, only two sources, chosen at random were permuted in increasing number of sub-bands.

For one such instance, Fig. 1(b) shows a plot of the variation in ℓ_1-norm with the number of sub-bands permuted.The circle in the plot shows the ℓ_1-norm of the true filter matrix. Each star shows the norm after permuting the sources 2 and 3.



(a) (b)

Fig. 1. The variation in the ℓ_1-norm of the filter matrix against the number of sub-bands permuted. (a) All the sources are permuted randomly (b) Only sources 2 and 3 are permuted.

3 Proposed algorithm

With the inspiration from the previous section, we present an algorithm in this section to solve the permutation indeterminacy.

Assumption: The estimate $\hat{\mathbf{A}}(f)$ of the sub-band coefficients $\mathbf{A}(f)$ are provided by some other independent technique and $\hat{\mathbf{A}}(f) = \mathbf{A}(f)\mathbf{P}(f), f = 1 \dots F$.

The absence of scaling indeterminacy is not totally a realistic assumption but we feel that this is the first step towards solving the bigger problem.

3.1 Description

We denote the set of all the possible source permutations by $\mathcal{P}$ ($|\mathcal{P}| = N!$) and the inverse discrete Fourier transform by $IDFT$. At each sub-band, every $\mathbf{P} \in \mathcal{P}$ is explored, keeping the other sub-bands fixed, and that permutation is retained which minimizes the ℓ_1-norm. This ensures that ℓ_1-norm of the filter matrix is lowered to the minimum possible extent by aligning that particular sub-band. At the end of one such iteration through all the sub-bands, the norm of the filter matrix would have significantly reduced. However, as the sub-bands were locally examined, the resulting norm may not be the global minimum. Hence, the entire process is iterated until the difference in the norm does not reduce significantly between iterations.

Input: $\hat{\mathbf{A}}, \theta$: The estimated sub-band coefficients of $\mathbf{A}(t)$ and a threshold
Output: $\tilde{\mathbf{A}}$: The sub-band coefficient matrix after solving for the permutations

(1) Initialize $\tilde{\mathbf{A}} \leftarrow \hat{\mathbf{A}}$;
(2) Update all sub-bands;
  foreach $f = 1 : F$ **do**
    $old\tilde{\mathbf{A}} \leftarrow \tilde{\mathbf{A}}$;
    foreach $\mathbf{P} \in \mathcal{P}$ **do**
      $\tilde{\mathbf{A}}(f) \leftarrow \hat{\mathbf{A}}(f)\mathbf{P}$;
      $val(\mathbf{P}) \leftarrow ||IDFT(\tilde{\mathbf{A}}(f'))||_1, f' = 1 \dots F$;
    end
    $\mathbf{P}(f) \leftarrow_{\mathbf{P} \in \mathcal{P}}^{\arg\min} val(\mathbf{P})$;
    $\tilde{\mathbf{A}}(f) \leftarrow \hat{\mathbf{A}}(f)\mathbf{P}(f)$;
  end
(3) Test if the algorithm should stop;
  if $||\hat{\mathbf{A}}(t)||_1 \geq ||old\hat{\mathbf{A}}(t)||_1 - \theta$ **then** Output $\tilde{\mathbf{A}}$ **else** Go to step (2)

Algorithm 1: Algorithm to solve the permutation indeterminacy by minimizing the ℓ_1-norm of the time domain filter matrix

3.2 Objective of the algorithm

The aim of the algorithm is to obtain the sub-band matrix which would give the minimum ℓ_1-norm. However, currently we do not have analytical proof about the convergence of the algorithm to the global minimum. Also, the sources will be globally permuted at the end.

3.3 Complexity

A brute force approach to solve the ideal ℓ_1-minimization problem would need $N!^K K \log K$ operations. In our case, each outer iteration needs to inspect $N! \times K$ permutations and at each step, an inverse FFT has a complexity of $K \log K$. Hence, the complexity in each outer iteration is $N! K^2 \log K$. This is still costly because it grows in factorial with the number of sources, but it is tractable for small problem sizes.

4 Experimental results

In this section we present an illustration of the algorithm presented above.

4.1 The no noise, no scaling case

Firstly, we consider the case where the sub-band coefficients are assumed to be estimated without noise and scaling ambiguity. 20 filter matrices with $N = 3, M = 2, K = 1024$ and $S = 10$ were created, transformed and sub-bands permuted in a similar way as explained in Sec. 2.1 and were the input to the algorithm. The value of θ was set to 0.0001 in all the experiments.

The output was transformed back to time domain to compute the reconstruction error. In all the experiments, the output filters were identical to the true filters up to a global permutation and within a numerical precision in Matlab.

4.2 Effect of noise

The estimation of $\hat{\mathbf{A}}(f)$ by an actual BSS algorithm invariably involves some level of noise (as well as scaling, which we do not deal with here). Hence, the permutation solving algorithm needs to be robust to certain level of noise. Experiments were conducted by permuting the sub-bands and adding noise to the coefficients:

$$\hat{\mathbf{A}}(f) = \mathbf{A}(f)\mathbf{P}(f) + \mathbf{N}(f) \tag{7}$$

where $\mathbf{N}(f)$ is i.i.d. complex Gaussian with mean zero and variance σ^2.

For illustration, Fig. 2 shows an instance of the reconstructed filter matrix using Algorithm 1 with the input corrupted by additive complex Gaussian noise with $\sigma^2 = 0.2$. Each filter had 10 significant coefficients which have been faithfully recovered, along with some amount of noise.

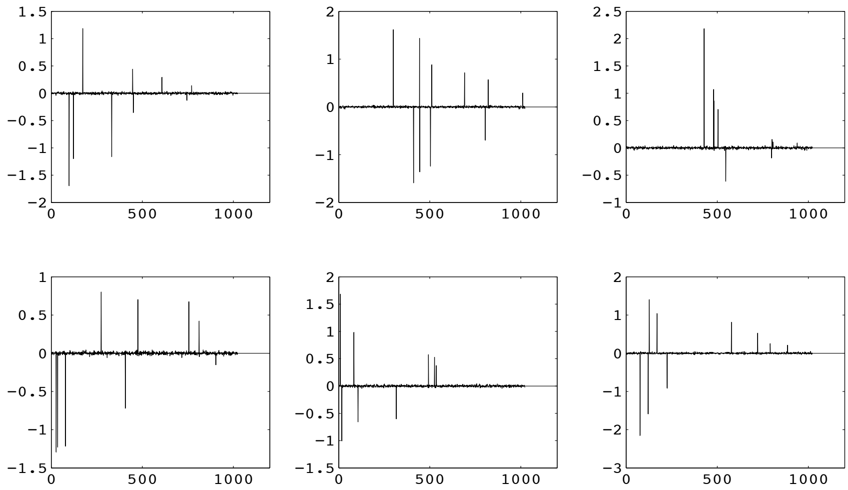


Fig. 2. Reconstructed filter matrix with additive noise in the sub-bands

For a quantitative analysis of the effect of noise, the input SNR was varied between -10 dB and 40 dB in steps of 5 dB and the output SNR was computed. The SNR definitions are given in Eqn. (8). The problem size was $N = 3, M = 2, K = 1024$, and 20 experiments were conducted to obtain each data point.

$$SNR_{in} = 20\log_{10}\left(\frac{||\mathbf{A}(t)||_2}{||\mathbf{N}(t)||_2}\right),\ SNR_{out} = 20\log_{10}\left(\frac{||\mathbf{A}(t)||_2}{||\mathbf{A}(t) - \tilde{\mathbf{A}}(t)||_2}\right) \quad (8)$$

The experiments were repeated for $S = 10$ and $S = 25$.

Figure. 3 shows the variation of output SNR (in dB) with input SNR. Due to the absence of scaling, a perfectly reconstructed filter will have the same SNR as the input. The thick line shows the ideal relationship for reference. In the range of 5 to 10 dB input SNR, the curve for $S = 10$ coincides with the ideal line. For $S = 25$, the curve coincides with the ideal line for range of 5 to 20 dB input SNR. At other places, both the curves closely follow the ideal line suggesting perfect reconstruction in most number of experiments.

5 Conclusion and future work direction

Frequency-domain estimation of mixing matrices in convolutive BSS suffers from the indeterminacies of source permutations and scaling of sub-bands. Hence, solving the permutation indeterminacy is an important aspect of such BSS systems. In this paper, it has been shown that in the absence of scaling, the ℓ_1-norm of the filter matrix is very sensitive to permutations in sub-bands. An algorithm has been presented based on the minimization of ℓ_1-norm of the filter matrix

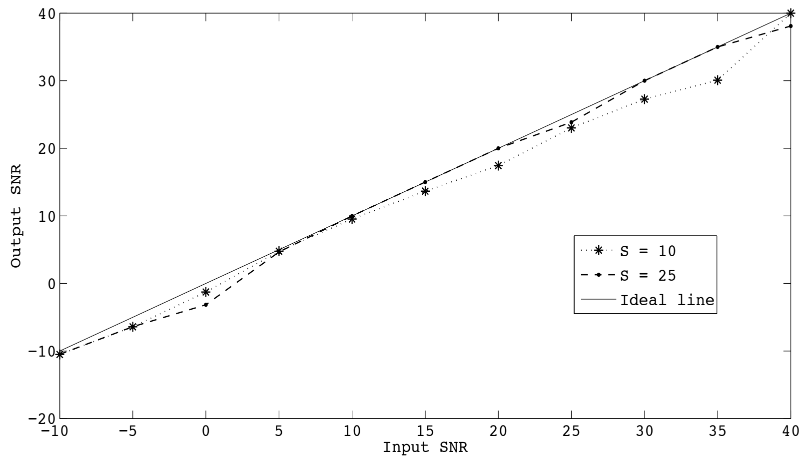


Fig. 3. Output SNR versus input SNR in dB

to solve for the permutations. Experimental results show that in the absence of scaling, the ℓ_1-minimization principle to solve the permutations performs well even in the presence of noise.

Though the absence of scaling is not a realistic assumption, it can be a first step towards sparsity motivated permutation solving methods. Also, the complexity of the algorithm grows with the $N!$ and K^2, which is expensive even for moderate values for the number of sources N and filter length K. Our future work focusses on replacing the combinatorial optimization step by an efficient convex optimization formulation and devising ℓ_1-norm based methods to solve the permutation problem in presence of arbitrary sub-band scaling.

References

1. Pedersen, M. S., Larsen, J., Kjems, U., Parra, L. C.:A Survey of Convolutive Blind Source Separation, Springer Handbook on Speech Processing and Speech Communication (2006)
2. Makino, S., Sawada, H., Mukai, R., Araki, S.: Blind source separation of convolutive mixtures of speech in frequency domain, IEICE Trans. Fundamentals, vol. E88-A, no. 7, pp. 1640-1655, Jul 2005
3. Ikram, M. Z., Morgan, D. R.:A beamforming approach to permutation alignment for multichannel frequency-domain blind source separation, Proc. ICASSP 2002, pp.881-884, May 2002
4. Murata, N., Ikeda, S., Ziehe, A.:An approach to blind source separation based on temporal structure of speech signals, RIKEN Brain Science Institute, BSIS Technical reports 98-2, Apr 1998
5. Donoho, D.:For most large underdetermined systems of linear equations, the minimal ell-1 norm near-solution approximates the sparsest near-solution, Communications on Pure and Applied Mathematics, 59(7), pp. 907-934, July 2006