Sparse BSS with spectral variabilities

Imane El Hamzaoui and Jérôme Bobin
CEA, IRFU, DEDIP/DAP
91191 Gif-sur-Yvette Cedex, France
Email: imane.el-hamzaoui@cea.fr

Abstract—Blind Source Separation has proven to be an efficient tool to recover meaningful information from multivalued data. However, most methods do not take into account spectral variabilities, therefore failing at modeling many real-world data (remote sensing, astrophysics to cite only a few) and leading to estimation errors. To overcome this pitfall, we propose a method that fully accounts for spectral variabilities by using the Perturbed Linear Mixture Model [1] and we exploit sparse modeling to recover spatially variant spectra. Preliminary results show a significant improvement on the detection of spectral variabilities in comparison with State-of-the-Art algorithms.

I. PROPOSED ALGORITHM

In Blind Source Separation (BSS), we assume that the \( m \) observations, each composed of \( t \) samples, denoted as \( \mathbf{X} \) follow the so-called Linear Mixture Model. Each observation corresponds to a weighted linear combination of \( n \) elementary sources such that

\[
\mathbf{X} = \sum_{i=1}^{n} \mathbf{A}^i \mathbf{S}_i + \mathbf{N},
\]

\( \mathbf{N} \) denotes both the noise and model imperfections. This framework, well-suited for many applications (astrophysics [3], spectrometry [4], ...) still fails at modeling real-world data in the presence of spectral variabilities (e.g. in Astrophysics, most components of the Cosmic Microwave Background have pixel-dependent emission spectra [5]).

Inspired by the Pertubated Linear Mixture Model proposed in [1], we model the mixture weights by a tensor \( \mathbf{A}^k \) model the mixture weights by a tensor

\[
\mathbf{A}^k = \mathbf{A}^0 \mathbf{S}_0 + \mathbf{N}^k,
\]

with \( \mathbf{A} \) spatially variant. This is an ill-posed, underdetermined problem which requires regularization.

In this study, we take advantage of spatial information and particularly sparse modeling of the variabilities in a transformed domain to discriminate between the solutions. The idea is to linearize each local mixing matrix \( \mathbf{A}^i[k] \) as the sum of a reference matrix \( \mathbf{A}^i \) and a local spectral variability \( \Delta \mathbf{A}^i[k] \) spatially sparse in a transformed domain, with \( \forall (i, j, k) \Delta \mathbf{A}^i[k] \ll \mathbf{A}^i_j[k] \). The minimization problem we tackle can be recast as follows:

\[
\min_{\mathbf{A}, \mathbf{S}} \frac{1}{2} \| (\mathbf{X} - \mathbf{A} \mathbf{S}) \|^2_F + \| \mathbf{A} \otimes \mathbf{S} \|_1 + \sum_{i=0}^{n} \| \Delta \mathbf{A}^i \|_F + \epsilon \| \mathbf{A} \|_2 + \epsilon \| \mathbf{S} \|_1.
\]

\( \otimes \) stands for the Hadamard product. The first term is the classical data-fidelity term well suited for additive gaussian noise, the second one refers to the classical sparse BSS constraint on the sources. For the sake of simplicity and since the recovery of \( \mathbf{A} \) is our focus, we will consider \( \Lambda = 0 \). Our applications being in the low-noise regime, this will have little impact. The third term accounts for the spatial sparsity of the variabilities given by the \( \ell_{2,1} \) norm in the transformed domain. The two last terms are, respectively, the positive and oblique constraints on the mixing matrix, standard constraints in BSS.

Equation (1) is a matrix factorization problem for which we have implemented the Block Coordinate Descent algorithm [6]. The initialization is given by the GMCA per patch algorithm [5]. This is an extension of the sparse BSS algorithm GMCA [12]. It consists on applying the GMCA to patches and filtering the results on the basis of a Frechet mean. Each iteration \( l \) can be described as follows:

1. Updating \( \mathbf{S} \) assuming \( \mathbf{A} \) is fixed: the source matrix is obtained with a classical pseudo-inverse problem: \( \mathbf{S}^{(l)} = \mathbf{A}^{(l)} \mathbf{X} \).
2. Updating \( \mathbf{A} \) assuming \( \mathbf{S} \) is fixed: Since there is an analytical proximal operator of each constraint \([7][8]\) but not of their sum, we use a Generalized Forward Backward algorithm [9] to recover \( \mathbf{A} \).

The sparsity framework is particularly suitable for a thresholding procedure based on detection theory \([10]\): we set the threshold \( \gamma \) such that the entries we keep have a probability of 0.4% to correspond only to noise. Since the contribution of gaussian noise follows a \( \chi \) distribution with \( m \) degrees of freedom, the standard deviation of which is unknown, we will use the median absolute deviation (MAD). The MAD operator is robust to sparse contribution and we can show empirically \( \sigma\_y \approx 1.5 \text{ MAD}(Y) \) for \( Y \) following a \( \chi \) law. Therefore \( \gamma \approx 1.5 \neq \text{ MAD}([\mathbf{A} - \bar{\mathbf{A}}] - \alpha(\mathbf{X} - \mathbf{A} \mathbf{S})\mathbf{S}^T_2 \|_1) \) with \( \alpha \) the stepsize and \( k \) the p-value associated to a probability of 0.4%.

To avoid biases due to soft-thresholding, we impose a reweighting \( \ell_{2,1} \) \([11]\). The weights are obtained from the mixing matrix given by the initialization (i.e GMCA per patch).

II. NUMERICAL EXPERIMENTS AND RESULTS

We use 2 approximately sparse sources, generated according to a Generalized Gaussian distribution of shape parameter \( p = 0.3 \). The number of samples \( t \) is 1500. Since we assume that the columns of the mixing matrix are normalized (oblique constraint on \( \mathbf{A} \)), each \( \mathbf{A}^i \) lives in the hypersphere \( \mathbf{S}^{m-1} \). Two reference vectors are generated to ensure an angle \( \theta_{\text{max}} = \frac{\pi}{2} \) between them. The samples of the local mixing matrix are drawn from a vector (centered around the reference matrix) exactly sparse in the DCT domain (peaks at positions 1 and 5 for the first source; 3 and 5 for the second one). The maximal amplitude of the spectral variabilities is the same for the two sources and is given by the angle \( \theta \). The figure 1 gives an example of the generated spectral variabilities (blue dashed curves). We have \( \text{SNR} = 65\text{dB} \).

We have implemented the GMCA per patch with patchsize of 500 samples, which is the optimal size for our applications. The stopping criterion is given by the relative difference between two successive iterates for \( \mathbf{A} \) and \( \mathbf{S} \) that must be inferior to \( 10^{-14} \) for the algorithm to stop. The performances are compared to State-of-the-Art methods, the GMCA and the GMCA per patch.

Conclusion

We introduced a new unsupervised decomposition algorithm accounting for spectral variabilities. To the best of our knowledge, it is the first method that exploits the spatial sparsity of the variabilities. The sparsity-based framework makes this method well-suited for many applications (astrophysics, spectrometry...). Further work will focus on the detection of the sources and the spectral variabilities by exploiting the morphological diversity between them \([2]\). More details about the algorithm and results will be presented at the conference.
We display generated variabilities of the first source obtained with \( \theta = \frac{\pi}{6} \). We also show the captured variabilities given by our algorithm coined svGMCA and the GMCA per patch. The svGMCA results are very consistent with the true variabilities unlike the GMCA per patch variabilities.

The performances of the algorithms have been described in terms of SAD for each sample \( k \):

\[
SAD[k] = \frac{1}{n} \sum_{i=1}^{n} \arccos\left( \frac{\langle A_i[k] | A_g[k] \rangle}{\| A_i[k] \|_2 \| A_g[k] \|_2} \right)
\]

with \( A \) the mixing matrix recovered with the algorithm implemented and \( A_g \) the ground-truth mixing matrix. We have used the same dataset as in fig 1. svGMCA yields better performances than the other methods (gain on the angular criterion up to 60dB).

Table. 1. This table illustrates the performances of the three algorithms in terms of GMSE for different spectral variabilities amplitude.

<table>
<thead>
<tr>
<th>( \theta = \frac{\pi}{6} )</th>
<th>( \theta = \frac{\pi}{4} )</th>
<th>( \theta = \frac{\pi}{12} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>svGMCA</td>
<td>9.97 \times 10^{-3}</td>
<td>1.11 \times 10^{-3}</td>
</tr>
<tr>
<td>GMCA per patch</td>
<td>8.69 \times 10^{-3}</td>
<td>3.26 \times 10^{-4}</td>
</tr>
<tr>
<td>GMCA</td>
<td>9.63 \times 10^{-3}</td>
<td>4.07 \times 10^{-4}</td>
</tr>
</tbody>
</table>

The svGMCA achieves better reconstruction of the mixing matrix.

REFERENCES