Sparse BSS in the presence of outliers

Cécile Chenot
CEA Saclay
DSM/Irfu/SEDI/LCS
91191 Gif-sur-Yvette, France
Email: cecile.chenot@cea.fr

Jérôme Bobin
CEA Saclay
DSM/Irfu/SEDI/LCS
91191 Gif-sur-Yvette, France
Email: jerome.bobin@cea.fr

Jérémy Rapin
CEA Saclay
DSM/Irfu/SAP/LCS
DSM/LIST/LOAD
91191 Gif-sur-Yvette, France

***Abstract*—While real-world data are often grossly corrupted, most techniques of blind source separation (BSS) give erroneous results in the presence of outliers. We propose a robust algorithm that jointly estimates the sparse sources and outliers without requiring any prior knowledge on the outliers. More precisely, it uses an alternative weighted scheme to weaken the influence of the estimated outliers. A preliminary experiment is presented and demonstrates the advantage of the proposed algorithm in comparison with state-of-the-art BSS methods.**

I. Problem Formulation

Suppose we are given m noisy observations $\{\mathbf{X}_i\}_{i=1..m}$ of unknown linear mixtures of $n \leq m$ sparse sources $\{\mathbf{S}_j\}_{j=1..n}$ with $t > m$ samples. It is generally assumed that these data are corrupted by a Gaussian noise, accounting for instrumental or model imperfections. However in many applications, some entries are additionally corrupted by outliers, leading to the following model: $\mathbf{X} = \mathbf{AS} + \mathbf{O} + \mathbf{N}$, with $\mathbf{X}$ the observations, $\mathbf{A}$ the mixing matrix, $\mathbf{S}$ the sources, $\mathbf{O}$ the outliers, and $\mathbf{N}$ the Gaussian noise.

The presence of outliers in real-world data have been taken into account in related domains such as compressive-sensing [9], dictionary learning [8] and denoising of hyperspectral images [4]. Contrary to these problems, both the mixing matrix and the sources should be individually retrieved in our problem, bringing further difficulties. In this case, the key issue lies in separating the components $\mathbf{O}$ from $\mathbf{AS}$. To this end, assuming that the term $\mathbf{AS}$ has low-rank, some strategies [4] suggest to pre-process the data to estimate and remove the outliers with RPCA [3]. However, besides the fact that low-rankness is generally restrictive for most BSS problems, the source separation is severely hampered if the outliers are not well estimated. Therefore, we introduce a method that estimates the sources in the presence of the outliers without pre-processing. For the best of our knowledge, it has only been studied in [5] by using the β-divergence. Unlike [5], we propose to estimate jointly the outliers and the sources by exploiting their sparsity.

II. Algorithm

Building upon the GMCA algorithm which has been shown to be an efficient BSS method for sparse sources [1], our approach estimates iteratively $\mathbf{O}$, $\mathbf{S}$ and $\mathbf{A}$. However, the straightforward extension of the GMCA algorithm struggles to distinguish between the sources and the outliers. Indeed, we can face important leakages from the outliers towards the estimated sources, misleading the estimation of $\mathbf{A}$. In the spirit of [2], we propose to further penalize the columns with detected outliers according to their amplitudes to limit their influence. Accordingly, the proposed robust-GMCA (rGMCA) algorithm aims at solving:

$$\underset{\mathbf{O},\mathbf{A},\mathbf{S}}{\text{minimize}}\ \frac{1}{2}\left\|(\mathbf{X}-\mathbf{AS}-\mathbf{O})\mathbf{W}\right\|_2^2 + \lambda\left\|\mathbf{S}\right\|_1 + \alpha\left\|\mathbf{O}\right\|_1,$$

This work was (partially) funded by the PHySIS project, contract no. 640174, within the H2020 Framework Program of the European Commission.

Algorithm 1: rGMCA-WO

Initialize $\tilde{\mathbf{O}}$, $\tilde{\mathbf{S}}$, $\tilde{\mathbf{A}}$, $\tilde{\mathbf{W}}$, $\tilde{\alpha}$, and $\tilde{\lambda}$.
while $k < K$ **do**
  $\tilde{\mathbf{O}} = \text{argmin}_{\mathbf{O}} \frac{1}{2}\left\|(\mathbf{X}-\tilde{\mathbf{A}}\tilde{\mathbf{S}}-\mathbf{O})\tilde{\mathbf{W}}\right\|_2^2 + \tilde{\alpha}\left\|\mathbf{O}\right\|_1$
  Update $\tilde{\mathbf{W}}$
  while $j < J$ **do**
    $\tilde{\mathbf{S}} = \text{argmin}_{\mathbf{S}} \frac{1}{2}\left\|(\mathbf{X}-\tilde{\mathbf{A}}\mathbf{S}-\tilde{\mathbf{O}})\tilde{\mathbf{W}}\right\|_2^2 + \tilde{\lambda}\left\|\mathbf{S}\right\|_1$
    $\tilde{\mathbf{A}} = \text{argmin}_{\mathbf{A}} \frac{1}{2}\left\|(\mathbf{X}-\mathbf{A}\tilde{\mathbf{S}}-\tilde{\mathbf{O}})\tilde{\mathbf{W}}\right\|_2^2$
    Choose $\tilde{\lambda}$
  end while
  Choose $\tilde{\alpha}$
end while

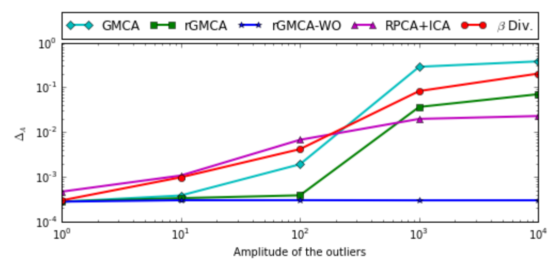

Fig. 1. Δ_A versus the amplitudes of the outliers (median of 80 simulations).

where $\mathbf{W}$ is the penalizing diagonal matrix of size $t \times t$ such that $w_{i,i} = \frac{1}{1+\left\|\mathbf{O}^i\right\|_1}$ with $\mathbf{O}^i$ the ith column of $\mathbf{O}$.

The problem is non-convex but can be efficiently tackled using block coordinate descent [7], see Algorithm 1. The parameters λ and α are set automatically based on the noise level, similarly to the thresholding strategy in [2].

III. Experiment

We compare our algorithm with rGMCA (the unweighted version of our algorithm) and also with standard BSS techniques: GMCA [1], RPCA+ICA [3], β Divergence [5] (implementation from [6]). We create 16 measurements of 8 sources of size 1024 generated from a Bernoulli-Gaussian law with parameter of activation 0.12 and standard deviation 100, to which a Gaussian noise is added. The support of the outliers is created such that: 160 are drawn uniformly at random, and 10 columns are entirely corrupted. Their amplitudes are Gaussian, with a standard deviation chosen according to fig.1. The variation of the global performance criterion $\Delta_A = \frac{\left\|(\tilde{\mathbf{A}}^T\tilde{\mathbf{A}})^{-1}\tilde{\mathbf{A}}^T.\mathbf{A}-\mathbf{I}\right\|_1}{n^2}$, with $\tilde{\mathbf{A}}$ the estimated matrix [2], versus the amplitude of the outliers is represented in fig.1. Our weighting procedure outperforms the standard algorithms and is almost not influenced by the amplitude of the outliers, fig.1. In the second

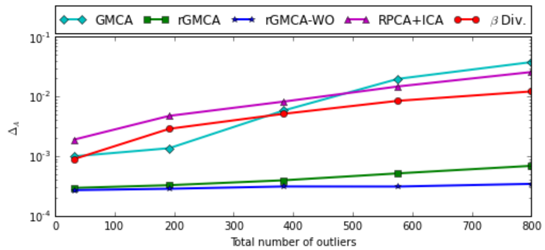

Fig. 2. Δ_A versus the total number of outliers (median of 80 simulations).

experiment, the amplitude of the outliers is set to 10^2. The total number of outliers is chosen according to fig.2. One half of the outliers is drawn uniformly at random and the other one corresponds to the entirely corrupted columns of $\mathbf{X}$. Our algorithm gets again better results and the estimation of $\mathbf{A}$ is nearly not affected by the increasing number of outliers. Furthermore, whereas the problem is not convex and has no free parameter, only 8 of the 800 simulations return an error larger than 10^{-3}.

IV. Conclusion

We propose a new robust BSS algorithm which estimates jointly outliers and sources and exploits their sparsity. These explicit estimations make any addition of a weighting scheme or of further prior easy. Numerical experiments show that the estimation of the mixing matrix is clearly enhanced with our weighting procedure.

References

[1] J.Bobin, J.L.Starck, Y.Moudden and M.J.Fadili, "Blind Source Separation: The Sparsity Revolution", *Advances in Imaging and Electron Physics*, vol.152, 2008.
[2] J.Bobin, J.Rapin, A.Larue, J.L.Starck, "Sparsity and Adaptivity for the Blind Separation of Partially Correlated Sources", *arXiv:1412.4005*, in press.
[3] E.J.Candès, X.Li, Y.Ma and J.Wright, "Robust Principal Component Analysis?", *Journal of ACM*, vol.58, 2011.
[4] Q.Li, H.Li, Z.Lu, Q.Lu and W.Li, "Denoising of Hyperspectral Images Employing Two-Phase Matrix Decomposition", *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* , vol.7, 2014.
[5] M.Mihoko and S.Eguchi, "Robust blind source separation by beta divergence", *Neural Computation*, vol.14, 2002.
[6] N.Gadhok and W.Kinsner, "An Implementation of β- Divergence For Blind Source Separation", *IEEE CCGE/CCGEI, Ottawa*, 2006.
[7] P.Tseng, "Convergence of a Block Coordinate Descent Method for Nondifferentiable Minimization", *Optimization*, 2001.
[8] Z.Chen and W.Ying, "Robust dictionary learning by error source decomposition", *Computer Vision (ICCV), 2013 IEEE International Conference on*, 2013.
[9] C.Studer, P.Kuppinger, G.Pope and H.Bölcskei, "Recovery of sparsely corrupted signals", *Information Theory, IEEE Transactions on*, vol.58, 2013.