Sparse BSS in the large-scale regime

Christophe Kervazo, Jérôme Bobin, Cécile Chenot

CEA Saclay, Irfu/SEDI, France | christophe.kervazo@cea.fr

Abstract

Standard Blind Source Separation (BSS) methods allow to efficiently analyze large-scale multichannel data. They nevertheless fail at dealing with problems involving a large number of sources. Building upon standard BSS algorithms, we propose investigating the performances of block-coordinate optimization strategies to handle this case. The use of blocks of intermediate sizes will be shown to yield both a dramatic improvement in the separation quality, which can potentially become perfect up to the numerical errors, and a gain in computation time.

Context

- In BSS, some multichannel data $\mathbf{X}$ composed of m row observations are assumed to be the linear combination, entached of noise $\mathbf{N}$, of n unknown elementary sources $\mathbf{S}$ of t samples: $\mathbf{X} = \mathbf{AS} + \mathbf{N}$ [1].
- Goal: estimate $\mathbf{A}$ and $\mathbf{S}$ from the sole $\mathbf{X}$, which is an ill-posed unsupervised matrix separation problem.
- Classical methods promote some discriminant information. ICA (or its efficient implementation FastICA) assumes statistically independent sources [1]. Non-Negative Matrix Factorization (NMF) assumes that the coefficients of $\mathbf{A}$ and $\mathbf{S}$ are non-negative. We will focus on sparse sources (EFICA, RNA, GMCA [2]...).
- With most algorithms, the separation tends however to deteriorate for large values of n:

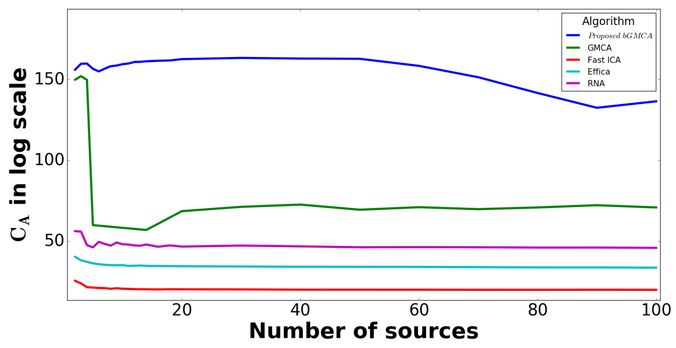


Illustration of the results of four classical BSS algorithms, as well as our proposed *bGMCA* algorithm.

- We aim at finding a way to efficiently extend the framework of BSS to a high number of sources.

Multi-convex optimization and proposed method

- A way to perform sparse BSS is to look for a minimizer of:

$$\min_{\mathbf{A},\mathbf{S}} \frac{1}{2} \|\mathbf{X} - \mathbf{AS}\|_F^2 + \mathcal{J}(\mathbf{A}) + \mathcal{G}(\mathbf{S})$$

- $\mathcal{J}(.)$ and $\mathcal{G}(.)$ are used to enforce some constraints on $\mathbf{A}$ and $\mathbf{S}$. We studied sevaral cases:
  - For $\mathbf{S}$:
    - ℓ_1 sparsity constraint: $\mathcal{G}(\mathbf{S}) = \|\Lambda_\mathbf{S} \odot \mathbf{S}\|_1$
    - ℓ_1 sparsity constraint in a transformed domain and non-negativity in the direct domain: $\mathcal{G}(\mathbf{S}) = \|\Lambda_\mathbf{S} \odot \left(\mathbf{S}\mathbf{\Phi}_\mathbf{S}^\mathbf{T}\right)\|_{\ell_1} + \iota_{\{\forall j,k; \mathbf{S}[j,k] \geq 0\}}(\mathbf{S})$
  - For $\mathbf{A}$:
    - Oblique constraint: $\mathcal{J}(\mathbf{A}) = \iota_{\{\forall i; \|\mathbf{A}^i\|_2^2 \leq 1\}}(\mathbf{A})$
    - Non-negativity and oblique constraint: $\mathcal{J}(\mathbf{A}) = \iota_{\forall i; \|\mathbf{A}^i\|_2^2 \leq 1}(\mathbf{A}) + \iota_{\forall i,j; \mathbf{A}[i,j] \geq 0}(\mathbf{A})$
- Multi-convex problem, yielding a determining role to the optimization strategy. So far, blocks [3] of size of 1 (*deflation* or *hierarchical* methods, e.g. HALS) or n (e.g. GMCA) were used, yielding a tradeoff between a less computationally difficult and simpler problem, and error propagations between the estimated sources. This is why we tried an in-between approach: we propose investigating the performances of block-coordinate optimization strategies using blocks I of sizes $r \in [1, n]$.

A two step minimization approach

- Our algorithm uses *blocks* indexed by a set of indices I of size r, $1 \leqslant r \leqslant n$.

Warm-up stage: GMCA with blocks [2]

- The results are a first guess used as an input of the PALM stage, which is more sensitive to the initialization and threshold choice.
- Iterative. Each update is performed using a projected least-square solution.
- Heuristic approach: decreasing threshold based on GMCA.

PALM algorithm with blocks [4]

- Convergence properties [5].
- Iterative. For each block: a gradient step on the smooth part and a proximal step on the nonsmooth part.
- Use the proximal operators of $\mathcal{J}$ and $\mathcal{G}$: $\texttt{prox}_{\mathcal{J}(.)}$ and $\texttt{prox}_{\mathcal{G}(.)}$.

Summary of the algorithm

- Warm-up step
  For (k) from 0 to nmax:
  - Choose a set of indices I
  - Estimation of $\mathbf{S}$ with a fixed $\mathbf{A}$: $\mathbf{S}_I^{(k)} = \texttt{prox}_{\mathcal{G}(.)}(\mathbf{A}_I^{(k-1)\dagger}\mathbf{R}_I)$
  - Estimation of $\mathbf{A}$ with a fixed $\mathbf{A}_I^{(k)} = \texttt{prox}_{\mathcal{J}(.)}(\mathbf{R}_I\mathbf{S}_I^{(k)\dagger})$
  - Choice of a new threshold $\Lambda^{(k)}$ (heuristic)
- Refinement step: initialize with the previous $\mathbf{A}$ and $\mathbf{S}$ and the corresponding thresholds.
  While not converged:
  - Choose a set of indices I
  - $\mathbf{S}_I^{(k)} = \texttt{prox}_{\frac{\gamma\mathcal{G}(.)}{\left\|\mathbf{A}_I^{(k-1)^T}\mathbf{A}_I^{(k-1)}\right\|_2}}\left(\mathbf{S}_I^{(k-1)} - \frac{\gamma}{\left\|\mathbf{A}_I^{(k-1)^T}\mathbf{A}_I^{(k-1)}\right\|_2}\mathbf{A}_I^{(k-1)^T}(\mathbf{A}^{(k-1)}\mathbf{S}^{(k-1)} - \mathbf{X})\right)$
  - $\mathbf{A}_I^{(k)} = \texttt{prox}_{\frac{\delta\mathcal{J}(.)}{\left\|\mathbf{S}_I^{(k)}\mathbf{S}_I^{(k)^T}\right\|_2}}\left(\mathbf{A}_I^{(k-1)} - \frac{\delta}{\left\|\mathbf{S}_I^{(k)}\mathbf{S}_I^{(k)^T}\right\|_2}(\mathbf{A}^{(k-1)}\mathbf{S}^{(k)} - \mathbf{X})\mathbf{S}_I^{(k)^T}\right)$

Numerical experiments: simulated matrices

Setting and metric

- $\mathbf{S}$ is taken exactly sparse in the direct domain following a Bernouilli-Gaussian distribution.
- $\mathbf{A}$ follows a standard normal distribution, has unit columns and a given condition number.
- The metric is computed using the median of $\mathbf{D} = |\mathbf{PA}^\dagger\mathbf{A}^*| - \mathbf{Id}$, where $\mathbf{A}^*$ is the true mixing matrix.

Modeling block minimization

- Modelization of the residual as the desired term plus errors:

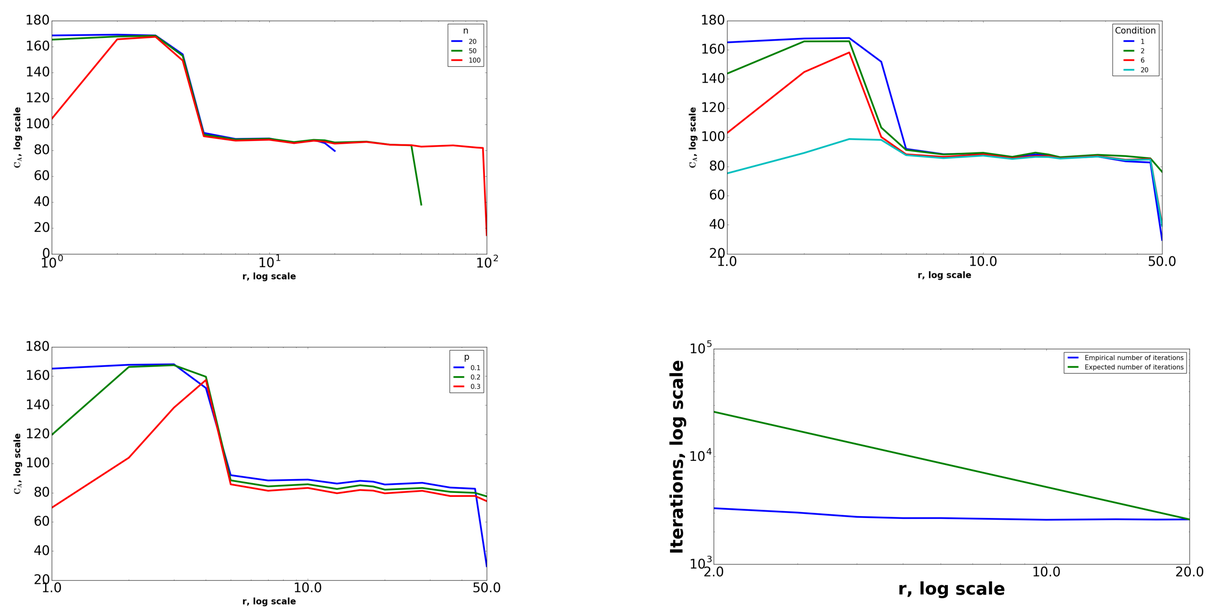
$$\mathbf{R}_I = \mathbf{X} - \mathbf{A}_{IC}\mathbf{S}_{IC} = \mathbf{A}_I^*\mathbf{S}_I^* + \mathcal{E} + \mathbf{N} \quad (1)$$

- The errors can be described in terms of interferences and artefacts (ϵ_{IC} being the errors on $\mathbf{S}_{IC}^*$):

$$\mathcal{E} = (\mathbf{A}_{IC}^* - \mathbf{A}_{IC})\mathbf{S}_{IC}^* - \mathbf{A}_{IC}\epsilon_{IC} \quad (2)$$

- Trade-off introduced by small blocks: a simpler separation problem vs a larger error.

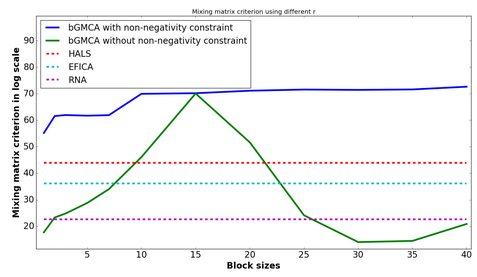
Simulated experiments

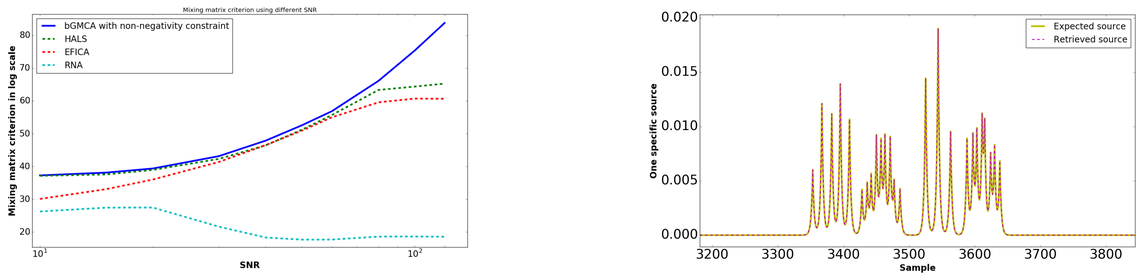


Top left: impact of r and n, top right: impact of the condition number of $\mathbf{A}$, bottom left: impact of the sparsity degree, bottom right: number of iterations

Numerical experiments: realistic sources

- Simulated LC - ^{1}H NMR spectroscopic data of a fluid composed of mixtures of chemical compounds.
- True sources $\mathbf{S}$ [6] but simulated $\mathbf{A}$.
- Sparsity enforced on $\mathbf{S}$ in the wavelet domain and non-negativity for $\mathbf{A}$ and $\mathbf{S}$ in the direct domain.





Top: results with 120dB of noise, bottom left: evolution of the results for $r = 10$ as a function of the SNR, bottom right: exemple of a particular estimated source

Conclusion and perspectives

The proposed *bGMCA* algorithm exhibits good performances in comparison to several standard methods, outperforming them in terms of separation quality on many experiments and even reaching perfect separation up to the numerical errors in some cases. The interest of *bGMCA* is furthermore a dramatic decrease in the computation time, which allows to deal with a large number of sources. To go along with our algorithm, we tried several strategies to choose the blocks.

References

[1] P. Comon, C. Jutten, Handbook of Blind Source Separation: Independent component analysis and applications, Academic press, 2010.
[2] J. Bobin, J.-L. Starck, J. M. Fadili, Y. Moudden, Sparsity and morphological diversity in blind source separation, IEEE Transactions on Image Processing 16 (11) (2007) 2662-2674.
[3] Y. Xu, W. Yin, A globally convergent algorithm for nonconvex optimization based on block coordinate update, arXiv preprint arXiv:1410.1386.
[4] J. Bolte, S. Sabach, M. Teboulle, Proximal alternating linearized minimization for nonconvex and nonsmooth problems, Mathematical Programming 146 (1-2) (2014) 459-494.
[5] E. Chouzenoux, J.-C. Pesquet, A. Repetti, A block coordinate variable metric forward-backward algorithm, Journal of Global Optimization 66 (3) (2016) 457-485.
[6] N. I. of Advanced Industrial Science, T. (AIST), Spectral database for organic compounds.