Block-GMCA

Christophe Kervazo and Jérôme Bobin
CEA, IRFU, SEDI/Service d'Astrophysique
91191 Gif-sur-Yvette Cedex, France
Email: christophe.kervazo@cea.fr

***Abstract*—Blind Source Separation (BSS) is a powerful method to analyze multichannel data in fields that involve processing large-scale data (*e.g.* astrophysical data, spectroscopic data in medicine and nuclear physics, etc.). However, standard methods fail at correctly tackling BSS problems when the number of sources becomes large, especially when the number of available samples is low. Building upon a standard BSS algorithm, namely GMCA (Generalized Morphological Component Analysis - [1] [2]), we propose investigating the performances of block-coordinate optimization strategies to tackle sparse BSS problems in the large-scale regime. Preliminary results reveal that the proposed approach, the block-GMCA algorithm, significantly improves the performances of the standard GMCA algorithm.**

I. The proposed algorithm

In the framework of BSS, the multichannel data $\mathbf{Y}$ are composed of m row observations and are assumed to be the linear combination of n unknown elementary sources $\mathbf{S}$ of t samples such that $\mathbf{Y} = \mathbf{AS}+\mathbf{N}$. The matrix $\mathbf{N}$ represents the noise and model imperfections. Neither the mixing matrix $\mathbf{A}$ nor the sources $\mathbf{S}$ are known, which makes this problem an ill-posed unsupervised matrix factorization problem [3]. So far, several approaches have been introduced to make BSS a better-posed problem. In this study, we will focus on the case of sparse BSS problems [4]. The proposed approach builds upon the GMCA algorithm [1], which seeks a critical point of :

$$\underset{\mathbf{A},\mathbf{S}}{\text{minimize}}\ \frac{1}{2}\left\|(\mathbf{Y}-\mathbf{AS})\right\|_2^2 + \left\|\Lambda \odot \mathbf{S}\right\|_1 \quad (1)$$

where Λ includes regularization terms. The first term of (1) enforces the proximity between observations and factorization, while the ℓ_1 norm enforces the sparsity of the estimated sources. Problem (1) is non-convex, which entails that the minimization strategy plays a key role to provide robustness with respect to spurious local critical points. The GMCA algorithm is built upon an iterative procedure alternately minimizing the problem (1) with respect to the sources $\mathbf{S}$ and the mixing matrix $\mathbf{A}$. This procedure is reminiscent of the popular algorithm ALS (Alternate Least-Square [5]) in NMF (Non-negative Matrix Factorization) and dictionary learning methods [6].

For a large number of sources n, numerical experiments show that this approach fails (Figure 1) because the GMCA algorithm is very likely prone to be trapped in local critical points. To alleviate this problem, we propose investigating the use of block-coordinate minimization schemes [7], [8]. Unlike GMCA, the proposed block-GMCA algorithm performs by minimizing (1) using *blocks* I of $\mathbf{A}$ and $\mathbf{S}$ of size r (the whole $\mathbf{Y}$ being used): $\text{minimize}_{\mathbf{A}_I,\mathbf{S}_I}\ \frac{1}{2}\left\|(\mathbf{Y}-\mathbf{A}_I\mathbf{S}_I-\mathbf{A}_{I^c}\mathbf{S}_{I^c})\right\|_2^2 + \left\|\Lambda \odot \mathbf{S}_I\right\|_1$. Each iteration (k) can then be described as follows:

1 - Randomly select a subset I of size r and update the submatrix $\mathbf{S}_I$ assuming $\mathbf{A}$ is fixed. This is performed by computing a projected least-square solution: $\hat{\mathbf{S}}_I = \mathcal{S}_{\Lambda^{(k)}}({\mathbf{A}_I^{(k)}}^\dagger \mathbf{R}_I)$, where the residual term is defined by $\mathbf{R}_I = \mathbf{Y} - \mathbf{A}_{I^c}^{(k)}\mathbf{S}_{I^c}^{(k)}$, the set I^c is the complement of I in $[1, n]$ and ${\mathbf{A}_I^{(k)}}^\dagger$ is the pseudo-inverse of $\mathbf{A}_I^{(k)}$. The operator $\mathcal{S}_{\Lambda^{(k)}}(\,.\,)$ is the standard soft-thresholding operator with thresholds $\Lambda^{(k)}$.

2 - Update the submatrix $\mathbf{A}_I$: $\hat{\mathbf{A}}_I = \mathbf{R}_I {\mathbf{S}_I^{(k)}}^\dagger$.

3 - Update the threshold matrix $\Lambda^{(k)}$. Following [1], the use of a decreasing thresholding strategy significantly improves the robustness of the GMCA algorithm with respect to spurious critical points and noise. This is carried out by retaining at each iteration an increasing number of samples with amplitudes larger than 3σ, σ being the noise standard deviation.

Note that a new subset I is randomly chosen at each iteration, ensuring some robustness due to several sources combinations tests.

II. Numerical results

For the sake of simplicity, the sources are assumed to be sparse in the sample domain. These results would be identical for any sources that are sparse in an orthogonal representation. For that purpose, the entries of the sources are drawn randomly according to a Generalized Gaussian distribution with a profile parameter $\alpha = 0.3$, which is a reasonable proxy for approximately sparse sources. The number of samples t is 1000. The mixing matrix $\mathbf{A}$ is drawn randomly (with $m = n$), such that its condition number is 1. The initialization consists in a random $\mathbf{A}$ and a zero matrix for $\mathbf{S}$. 2000 iterations are performed so that all sources are updated a sufficient amount of times. Figure 2 displays the performances of the block-GMCA algorithm for different number of sources when the block size r evolves from 1 (the sources are updated individually) to n (the standard GMCA algorithm). These performances are evaluated using a mixing matrix criterion C_A defined in a similar way as in [2]. C_A is the median of the matrix $\mathbf{PA}^\dagger\mathbf{A}^* - \mathbf{I}$, with $\mathbf{A}^*$ the true mixing matrix and $\mathbf{PA}^\dagger$ the pseudo-inverse of the solution estimated by the algorithm, corrected through $\mathbf{P}$ for the scale/permutation indeterminacies.

For low block sizes (*i.e. typically* $r < 5$), the loss of separation quality is likely explained by increased error propagations. Indeed, in block-GMCA, the submatrices $\mathbf{A}_I$ and $\mathbf{S}_I$ are updated from a residual, which might be prone to error propagations. For large batch sizes ($r > 40$ for $n = 50$ and $r > 70$ for $n = 100$), the loss is more likely the consequence of a lack of robustness to local critical points. It is very interesting to see that there is a regime, with potentially small block sizes (*e.g.* $5 < r < 70$), where the separation is of improved quality. While the convergence of the block-coordinate minimization scheme is guaranteed [7], updating random subsets of the sources yields randomness that likely improves the robustness of the minimization procedure with respect to local critical points.

Conclusion

We introduce the block-GMCA algorithm to improve the performances of sparse BSS in the large-scale regime. Preliminary experiments show that the proposed block-coordinate minimization approach significantly improves the performances when the number of sources to be estimated is large. Moreover, the block-GMCA paves the way for computationally effective implementations of sparse BSS methods due to potential parallel implementations. More details about the algorithm and the results will be presented at the conference.

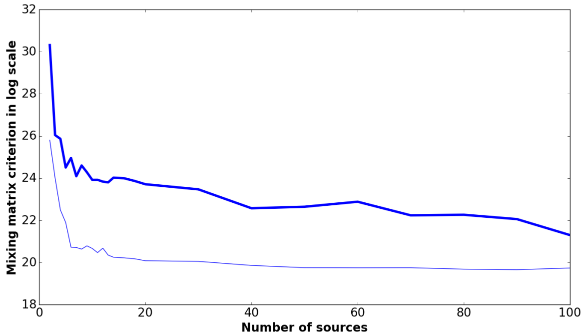

Fig. 1. Results of two BSS algorithms as a function of the number of sources. The values are computed as $-10 \log C_A$, with C_A the mixing matrix criterion, and are high for high quality separation. Thick line: GMCA, thin line: fast ICA, showing that the deterioration of the separation quality for larger n is not limited to GMCA and thus the interest of developping efficient large-scale methods.

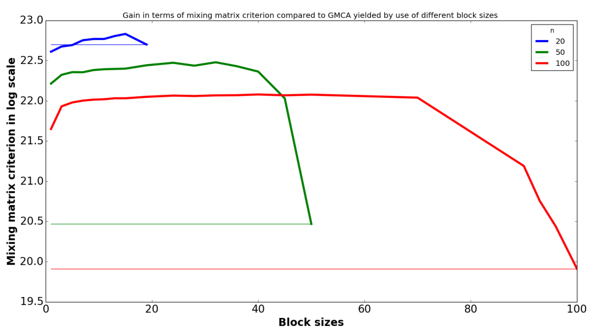

Fig. 2. $-10 \log C_A$ as a function of r and for several numbers of sources. Thick line: block GMCA, thin line: GMCA.

References

[1] J. Bobin, J.-L. Starck, J. Fadili, and Y. Moudden, "Sparsity and morphological diversity in blind source separation," *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2662–2674, 2007.

[2] J. Bobin, J. Rapin, A. Larue, and J.-L. Starck, "Sparsity and adaptivity for the blind separation of partially correlated sources," *IEEE Transactions on Signal Processing*, vol. 63, no. 5, pp. 1199–1213, 2015.

[3] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent component analysis and applications*. Academic press, 2010.

[4] M. Zibulevsky and B. A. Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary," *Neural computation*, vol. 13, no. 4, pp. 863–882, 2001.

[5] A. Cichocki and R. Zdunek, "Regularized alternating least squares algorithms for non-negative matrix/tensor factorization," in *International Symposium on Neural Networks*. Springer, 2007, pp. 793–802.

[6] R. Rubinstein, M. Zibulevsky, and M. Elad, "Efficient implementation of the k-svd algorithm using batch orthogonal matching pursuit," *CS Technion*, vol. 40, no. 8, pp. 1–15, 2008.

[7] P. Tseng, "Convergence of a block coordinate descent method for nondifferentiable minimization," *Journal of optimization theory and applications*, vol. 109, no. 3, pp. 475–494, 2001.

[8] Y. Xu and W. Yin, "A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion," *SIAM Journal on imaging sciences*, vol. 6, no. 3, pp. 1758–1789, 2013.