Stage M2, 2019

Large-scale sparse blind decomposition

Sparse signal modelling, blind component separation, proximal algorithms, large-scale data, multi-convex optimization

Context

In the context of astrophysics, the data collected are often very large. This is for instance the case with the Chandra X-ray observatory[1] (collected images with up to 10^9 pixels), the Planck mission[2] (450×10^6 pixels) or continental-size interferometers such as the SKA[3] (typically producing terabytes per second). Such huge dimensions mandates the development of computationally efficient algorithms enabling to extract the core information. In this context, blind component separation (or blind source separation – BSS [1]) is one of the key tools enabling the study of multi-valued data such as multi-wavelength ones. The goal of BSS is to decompose the observed data $\mathbf{X}$ into a linear decomposition of some elementary signals, called sources. Mathematically speaking, BSS aims at finding two matrices $\mathbf{A}$ and $\mathbf{S}$, called respectively the sources and the mixing matrix, such that :

$$\mathbf{X} = \mathbf{AS} + \mathbf{N} \tag{1}$$

Where $\mathbf{N}$ is an unknown noise perturbating the mixing. Since problem (1) is an ill-posed problem, additional assumptions are needed; during this internship, $\mathbf{S}$ will be assumed to have a large number of zero coefficients, that is to be sparse [4]. An example of BSS problem in astrophysical imaging (study of a supernovae remnant) is given Fig. 1.

In the large-scale context, classical BSS methods fail due to two major limitations: i) memory and computational burden (it is sometimes not even possible to store the full matrices into memory); ii) deteriorated separation performances (in particular, when the number of sources is high). Several works have already attempted to tackle large-scale BSS problems. Among them, two can be highlighted: *block-coordinate* methods [2,3] that split the $\mathbf{A}$ and $\mathbf{S}$ matrices; and *mini-batch* methods [5] that split the $\mathbf{X}$ matrix. While these works have shown good results both in terms of computational time and accuracy, several paths are promising to still enhance them.

Goal

The goal of this internship is to work on large scale BSS and more specifically to contribute to the development of block-coordinate and mini-batch methods, while merging these approaches into a general framework. Several paths are to be studied :

[1] http://chandra.harvard.edu/
[2] https://www.cosmos.esa.int/web/planck/
[3] https://skatelescope.org/

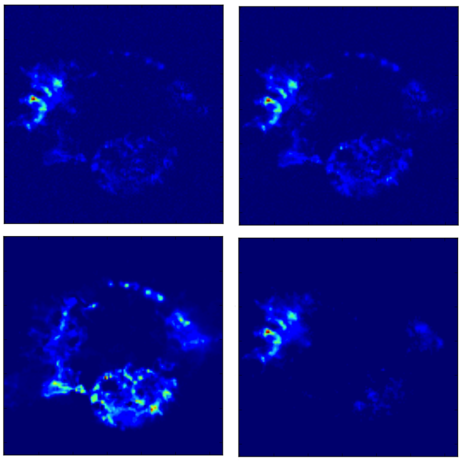

Figure 1: BSS problem on data captured by the Chandra X-ray laboratory. *Up*: exemple of 2 mixings in $\mathbf{X}$. *Down*: the 2 sources at the origin the mixings (which correspond to the thermal and iron emission respectively).

- Combine both block-coordinate and mini-batch methods into a single algorithm and study the resulting performances.
- Work on a distributed implementation of block-coordinate methods using an asynchronous strategy (contrary to the current sequential methods).

Depending on remaining time, the candidate will also have the opportunity to test the developed algorithms on the difficult under-determined case, in which the number of sources to find is larger than the number of observations.

The candidate

The candidate must have a Master 2 degree (or equivalent) and should have a good knowledge in signal/image processing. Knowledge in convex optimization is a plus, as well as Python or C++ coding skills.
The candidate will acquire an expertise in sparse image processing, multichannel data processing, machine learning and modern large-scale non-convex optimization techniques. The knowledge acquired during the internship are applicable to a wide range of applications in various fields such as biomedical imaging, astrophysics, hyper-spectral imaging...

Contact

The internship will take place in the CosmoStat lab (CEA Saclay), which is a joint laboratory at the interface in signal processing and astrophysics. The student will take an active participation in the European project LENA (http://lena.cosmostat.org).

- Contact: christophe.kervazo@cea.fr
- Lab: CEA/IRFU in Saclay
- Applications are expected before the 28th of February 2019.

References

[1] Comon, Pierre, and Christian Jutten, eds. Handbook of Blind Source Separation: Independent component analysis and applications. Academic press, 2010.
[2] Kervazo, C., Jerome Bobin, and Cecile Chenot. "Blind separation of a large number of sparse sources." Signal Processing 150 (2018): 157-165.
[3] Tseng, Paul. "Convergence of a block coordinate descent method for nondifferentiable minimization." Journal of optimization theory and applications 109.3 (2001): 475-494.
[4] Starck, Jean-Luc, Fionn Murtagh, and Jalal M. Fadili. Sparse image and signal processing: wavelets, curvelets, morphological diversity. Cambridge university press, 2010.
[5] Mairal, Julien, et al. "Online learning for matrix factorization and sparse coding." Journal of Machine Learning Research 11.Jan (2010): 19-60.