Journal club presentation

Imane El hamzaoui
imane.el-hamzaoui@cea.fr

Séparation non-supervisée de composantes multivaluées parcimonieuses et applications en astrophysique

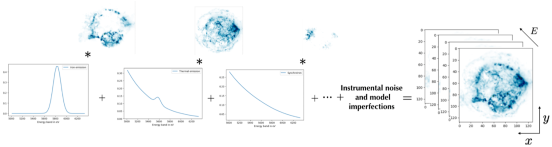
20/05/2019

Outline

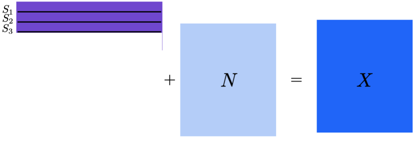
Multivalued data analysis with BSS

Introduction to BSS

- Modelization of the data through the Linear Mixture Model (LMM):



$$\mathbf{A}^1\mathbf{S}_1 + \mathbf{A}^2\mathbf{S}_2 + \mathbf{A}^3\mathbf{S}_3 + \ldots + \mathbf{N} = \mathbf{X}$$



- Blind Source Separation (BSS) aims at **disentangling** mixed components to retrieve spectral and spatial information.

Multivalued data analysis with BSS

Introduction to BSS

$$min_{\mathbf{A},\mathbf{S}} \|\mathbf{X} - \mathbf{AS}\|_F^2. \quad (1)$$

Ill-posed problem requiring further assumptions:

- Statistical independence of the sources: **I**ndependent **C**omponent **A**nalysis [1],
- Non-negativity of the components: **N**on-negative **M**atrix **F**actorization [2],
- Sparsity of the sources (possibly in a transformed domain): sparse BSS *e.g.* Generalized Morphological Component Analysis algorithm (GMCA [3]).

[1] *Comon et al 2010*

[2] *Lee et al 1999*

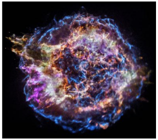
[3] *Bobin et al 2007*

Multivalued data analysis with BSS

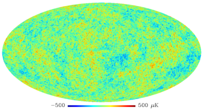
Goal and organization of the PhD

- BSS has proven to be efficient in the detection of microwave and infra-red rays, including the detection of the oldest observable electromagnetic radiation of the Universe [4].

This PhD is aimed at extending BSS methods to high energy data (detection of supernova remnants, blackholes...).



Supernova remnant Cassiopeia A seen by X-ray telescope Chandra



CMB reconstruction with sparse BSS

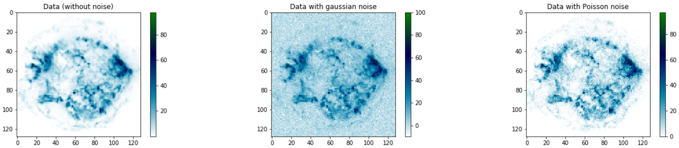
[4] *Bobin et al 2013*

Multivalued data analysis with BSS

Goal and organization of the PhD

Challenges raised by high-energy imaging

1. **Poisson noise:** High-energy photon count is so low that we cannot consider the noise gaussian. The modelisation $\mathbf{X} = \mathbf{AS} + \mathbf{N}$ is <u>no more valid</u>.



Observation derived from Chandra simultations with gaussian noise and poissonian noise. Both of the noises have the same level in terms of mean square error.

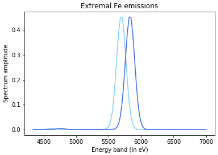
Poisson noise, unlike gaussian noise is <u>data dependent</u>.

Multivalued data analysis with BSS

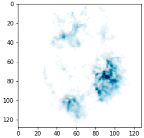
Goal and organization of the PhD

Challenges raised by high-energy imaging

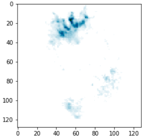
2. **Spectral variabilities:**
   - spatially variant spectra are ubiquitious to X-ray imaging.
   - Their estimation is of great astrophysical interest
   - *e.g.* Fe line shifting allows to estimate the speed of supernovae remnants for example.



Spectral distribution



Spatial distribution of Fe1



Spatial distribution of Fe2

Necessity of a method fully accounting for spectral variabilities.

Outline

Overview of the problem

Model $\mathbf{X} = \mathbf{AS} + \mathbf{N}$ not valid.

The noise corrupting high energy data follows shot noise statistics.

Poisson law

$$\mathcal{P}(\mathbf{X}_j[t] \mid [\mathbf{AS}]_j[t]) = \frac{e^{-[\mathbf{AS}]_j[t]}[\mathbf{AS}]_j[t]^{\mathbf{X}_j[t]}}{\mathbf{X}_j[t]!}, \quad (2)$$

where $[\mathbf{AS}]_j[t]$ is the sample of the pure mixture $\mathbf{AS}$ located at the j-th row and t-th column.

Overview of the problem

Sparse BSS with Poisson measurements

$$\mathcal{P}(\mathbf{X}_j[t] \mid [\mathbf{AS}]_j[t]) = \frac{e^{-[\mathbf{AS}]_j[t]}[\mathbf{AS}]_j[t]^{\mathbf{X}_j[t]}}{\mathbf{X}_j[t]!}$$

Optimization function

$$\min_{\mathbf{A},\mathbf{S}} \underbrace{\mathcal{L}(\mathbf{X} \mid \mathbf{AS})}_{\text{antiloglikelihood}} + \underbrace{\| \Lambda \odot \mathbf{S}\Phi^T \|_1 + i_{\cdot \geq 0}(\mathbf{S})}_{\text{constraints on } \mathbf{S}} + \underbrace{i_C(\mathbf{A})}_{\text{constraints on } \mathbf{A}}, \quad (3)$$

where $\mathcal{L}$ the antiloglikelihood of Poisson noise:

$$\begin{aligned} \mathcal{L}(\mathbf{X} \mid \mathbf{AS}) &= \sum_{j,t} [\mathbf{AS}]_j[t] - \mathbf{X}_j[t] \log([\mathbf{AS}]_j[t]) \\ &= \mathbf{AS} - \mathbf{X} \odot \log(\mathbf{AS}), \end{aligned} \quad (4)$$

Λ contains the regularization parameters;
$C = \mathrm{OB}(m) \cap K^+$.

Overview of the problem

Sparse BSS with Poisson measurements

Optimization function

$$\min_{\mathbf{A},\mathbf{S}} \underbrace{\mathcal{L}(\mathbf{X} \mid \mathbf{AS})}_{\text{antiloglikelihood}} + \underbrace{\| \Lambda \odot \mathbf{S}\Phi^T \|_1 + i_{\cdot \geq 0}(\mathbf{S})}_{\text{constraints on } \mathbf{S}} + \underbrace{i_C(\mathbf{A})}_{\text{constraints on } \mathbf{A}} , \quad (5)$$

- Multiconvex problem $\Longrightarrow$ Block Coordinate Descent (BCD) algorithm [5];
- Smooth and differentiable likelihood required $\Longrightarrow$ replace it by a smooth approximate.

$$\mathcal{L}(\mathbf{X} \mid \mathbf{AS}) = \mathbf{AS} - \mathbf{X} \odot \log(\mathbf{AS}).$$

[5] *Tseng 2001*

Smooth approximation of the data-fidelity term

We propose an approximation of $\mathcal{L}$ based on Nesterov's technique

Nesterov, 2005

$$\mathcal{L}_\mu(\mathbf{X} \mid \mathbf{Y}) = \inf_{\mathbf{U}} < \mathbf{Y}, \mathbf{U} > - \underbrace{\mathcal{L}^*(\mathbf{X} \mid \mathbf{U})}_{\text{Fenchel dual of } \mathcal{L}} - \underbrace{\mu \|\mathbf{U}\|_F^2}_{\text{regularization function}}, \quad (6)$$

where $\mu \in \mathbb{R}^+$ is the **smoothing parameter**.

$\mathcal{L}_\mu$ is **differentiable** and admits a $\frac{1}{\mu}$-**Lipschitzian gradient**.

The cost function we aim at minimizing becomes:

$$\min_{\mathbf{A},\mathbf{S}} \mathcal{L}_\mu(\mathbf{X}|\mathbf{AS}) + \| \Lambda \odot \mathbf{S}\Phi^T \|_1 + i_C(\mathbf{A}) + i_{\cdot \geq 0}(\mathbf{S}), \quad (7)$$

Multiconvex problem

Block Coordinate Descent algorithm (BCD)

$$\min_{\mathbf{A},\mathbf{S}} \mathcal{L}_\mu(\mathbf{X}|\mathbf{AS}) + \| \Lambda \odot \mathbf{S}\Phi^T \|_1 + i_C(\mathbf{A}) + i_{.\geq 0}(\mathbf{S}), \tag{8}$$

BCD: alternative estimation of the convex subproblems.

Structure of pGMCA algorithm

- **Initialization:** $(\mathbf{A}^{(0)}, \mathbf{S}^{(0)})$ obtained with robust (to initialization) sparse BSS algorithm GMCA[a].
- **Iteration** k:
  - Update of $\mathbf{A}$ assuming $\mathbf{S}$ fixed,
  - Update of $\mathbf{S}$ assuming $\mathbf{A}$ fixed.

[a] *Bobin et al, 2007*

Update of **A** assuming **S** fixed

$$\min_{\mathbf{A}} \mathcal{L}_\mu(\mathbf{X} \mid \mathbf{AS}) + i_C(\mathbf{A}). \quad (9)$$

- $\mathcal{L}_\mu(\mathbf{X} \mid \mathbf{AS})$ is a differentiable function whose gradient is $\frac{\|\mathbf{S}^T\mathbf{S}\|_2}{\mu}$,
- $i_C(.)$ is an indicator function of a convex set, it is proximable.

Update of **A** at iteration (k)

$$\mathbf{A}^{(k+1)} \Leftarrow \text{FISTA}^1(\mathbf{S}^{(k)}, \mathbf{A}^{(k)})$$

[1] *Fast Iterative Shrinkage-Thresholding Algorithm, Beck et al 2009*

Update of **S** assuming **A** fixed

$$\min_{\mathbf{S}} \mathcal{L}_\mu(\mathbf{X} \mid \mathbf{AS}) + \| \Lambda \odot \mathbf{S}\Phi^T \|_1 + i_{\cdot \geq 0}(\mathbf{S}). \quad (10)$$

- $\mathcal{L}_\mu(\mathbf{X} \mid \mathbf{AS})$ is a differentiable function whose gradient is $\frac{\|\mathbf{S}^T\mathbf{S}\|_2}{\mu}$,
- $\| \Lambda \odot \mathbf{S}\Phi^T \|_{\ell_1}$ is the proximable ℓ_1 norm.
- $i_C(.)$ is an indicator function of a convex set, it is proximable.

Update of **S** at iteration (k)

$$\mathbf{S}^{(k+1)} \Leftarrow \text{Generalized Forward Backward}^1(\mathbf{S}^{(k)}, \mathbf{A}^{(k)})$$

Question: how to set the regularization parameter Λ?

[1] *Raguet et al, 2013*

Set-up of the threshold

Gaussian noise

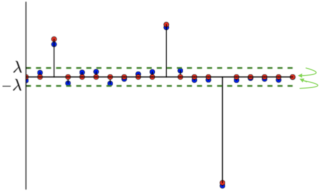
Let's consider the generic problem

$$min_X \frac{1}{2}\|Y - AX\|_F^2 + \|\lambda \odot X\|_1,$$

with gaussian noise and X sparse in direct domain.

•At iteration (k) of a proximal algorithm (*e.g* GFB), we have

$$X^{(k+1)} = \mathcal{S}_\lambda(\underbrace{X^{(k)}}_{\text{sparse contribution}} + \underbrace{\nabla_X \|Y - AX\|_F^2)}_{\mathcal{G}:\ \text{gaussian noise contribution}}.$$



$X + \mathcal{G}$ (blue) and X (red).

Set-up of the threshold

Gaussian noise

We want to **remove** gaussian noise and **keep** sparse signal coefficients (of X) with **their amplitudes**.

$\mathcal{S}_\lambda(X+\mathcal{G}) = Z \; s.t. \; \forall j \; Z_j = (|X_j + \mathcal{G}_j| - \lambda)_+$.

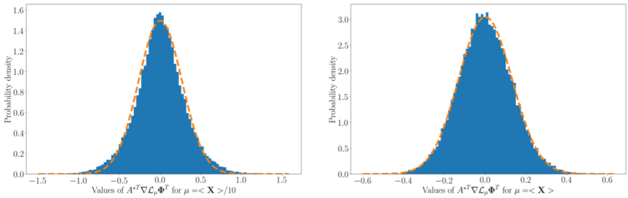
- "3-σ rule": the probability that an amplitude higher than 3-σ corresponds only to Gaussian noise is 0.4%
- σ unknow but we have $\sigma_{\mathcal{G}} = 1.48 MAD(\mathcal{G})$
- $MAD(X+\mathcal{G}) = MAD(\mathcal{G})$

λ set up as a denoising threshold

$$\lambda = 1.48k\mathrm{MAD}$$

where $k = 3$ for Gaussian noise

Set-up of the threshold
From Gaussian noise to Poisson noise



Histogram of the gradient of $\mathcal{L}_\mu$ with respect to **S** at the true input and their Gaussian best fit

$$\lambda = 1.48k\mathrm{MAD}$$

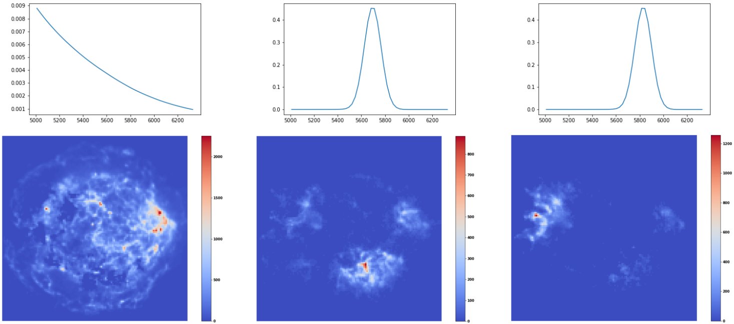
where $k = 1$ for Poisson noise; and to limit biaises [a]:

$$\lambda_f = \frac{w}{w+\epsilon}\lambda$$

[a] *Candes et al, 2008*

Dataset

Realistic Chandra simulations



1.Synchrotron 2. Fe 1 3. Fe 2

Figure: Spectral (*top line*) and spatial (*bottom line*) distribution of the three components

State-of-the-Art algorithms compared to pGMCA

- **G**eneralized **M**orphological **C**omponent **A**nalysis: standard sparse BSS algorithm;
- **H**ierarchial **Al**ternating **L**east **S**quare algorithm: NMF algorithm with sequential updates [6];
- β **NMF**: NMF with Kullback-Leibler divergence (β=1) [7];
- sparse **NMF** algorithm [8].

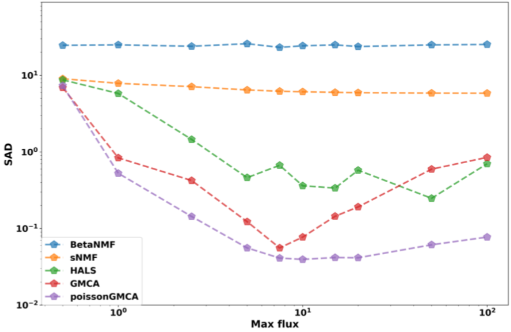
[6] *Gillis et al, 2012*

[7] *Mihoko et al, 2002*

[8] *Le Roux et al, 2015*

State-of-the-Art algorithms compared to pGMCA

Metric: $\mathrm{SAD} = \frac{1}{n}\sum_{i=1}^{n} \arccos(< \mathbf{A}^i \mid {\mathbf{A_g}}^i >)$ with $\mathbf{A}$ the mixing matrix recovered with the proposed approach and $\mathbf{A_g}$ the ground-truth mixing matrix.



Outline

Overview of the problem

The mixing matrix is pixelwise dependent:

$$\forall \textit{ sample } k \; \mathbf{X}[k] = \sum_{i=1}^{n} \mathbf{A}^i[k]\mathbf{S}_i[k] + \mathbf{N}[k], \qquad (11)$$

State-of-the-Art model: Perturbed Linear Mixture Model (PLMM) [a]

[a] *Thouvenin et al, 2016*

$$\mathbf{X} = \bar{\mathbf{A}}\mathbf{S} + \Delta\mathbf{A}\mathbf{S}$$

- Applied to hyperspectral terrestrial images.
- sum-to-one assumption $\sum_i \mathbf{S}_i = 1$: the sources are not independent,
- pure pixel assumption $\forall i, \exists k' / \mathbf{X}[k'] = A^i$.

Overview of the problem

Linearization and angular variability

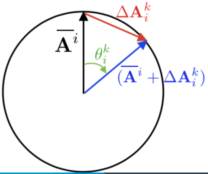
Inspired by Perturbated Linear Mixture Model (PLMM), we linearise **A**

$$\forall i, j, k \; \mathbf{A}_j^i[k] = \bar{\mathbf{A}}_j^i + \Delta_j^i[k]$$

Angular variability:

$$\|\Delta^i[k]\|_2 = 2\sin\left(\frac{\theta^i[k]}{2}\right) \simeq \theta^i[k] \ll \|\bar{\mathbf{A}}^i\|_2;$$

since $\mathbf{A} \in \mathrm{OB}(m)$ in the sparsity framework.



Overview of the problem

Linearization and spatial regularization

$\forall i, j, k \ \mathbf{A}_j^i[k] = \bar{\mathbf{A}}_j^i + \Delta_j^i[k]$ with $\|\Delta^i[k]\|_2 = \theta^i[k]$.
Optimization function:

$$\min_{\mathbf{A},\mathbf{S}} \frac{1}{2} \left\| (\mathbf{X} - \sum_i \mathbf{A}^i \mathbf{S}_i) \right\|_F^2 + \|\Lambda \odot \mathbf{S}\|_1$$

Underdetermined and (very) ill-posed problem $\rightarrow$ constraint on spectral variabilities (SV) required.

Spatial regularization of the SV:

$$\forall i \|\gamma \odot \theta^i \ \Psi^T\|_1 = \|\gamma \odot \|\mathbf{A}^i - \bar{\mathbf{A}}^i\|_2 \ \Psi^T\|_1 \simeq \|\gamma \odot (\mathbf{A}^i - \bar{\mathbf{A}}^i)\Psi^T\|_{2,1}. \quad (12)$$

Optimization function

$$\min_{\mathbf{A},\mathbf{S}} \frac{1}{2}\left\|(\mathbf{X} - \sum_i \mathbf{A}^i \mathbf{S}_i)\right\|_{\mathrm{F}}^2 + \|\Lambda \odot \mathbf{S}\|_1 + \sum_{i=0}^{n}\left\|\gamma \odot (\mathbf{A}^i - \bar{\mathbf{A}}^i)\Psi^T\right\|_{2,1} + i_C(\mathbf{A}) \quad (13)$$

where

- Λ (*resp* γ) contains the regularization parameters and weights for the source matrix and *resp.* the spectral variabilities,
- $C = \mathrm{OB}(m) \cap K^+$.

For these preliminary tests, we do not enforce sparsity of the sources.

Multiconvex problem --→ BCD algorithm.

Initialization and update of **S**

Initialization

GMCA per patch algorithm: GMCA on blocks + filtering. [a]

[a] *Bobin et al, 2013*

Update of **S** assuming **A** is fixed

$$\min_{\mathbf{S}} \frac{1}{2} \left\| (\mathbf{X} - \sum_i \mathbf{A}^i \mathbf{S}_i) \right\|_{\mathrm{F}}^2 .$$

Moore-Penrose pseudo-inverse : $\mathbf{S}^{(l)} = \mathbf{A}^{(l)\dagger} X$.

Update of **A**

Update of **A** assuming **S** is fixed

$$\min_{\mathbf{A}} \frac{1}{2} \left\| (\mathbf{X} - \sum_i \mathbf{A}^i \mathbf{S}_i) \right\|_{\mathrm{F}}^2 + \sum_{i=0}^{n} \left\| \gamma \odot (\mathbf{A}^i - \bar{\mathbf{A}}^i) \Psi^T \right\|_{2,1} + \iota_C(\mathbf{A}).$$

- The proximal operator of the positivity and the oblique constraint is their composition,
- The proximal operator of the $\ell_{2,1}$ norm in transformed domain is analytical...
- ... but there is no proximal operator for all three constraint.

We use a **Generalized Forward Backward (GFB) algorithm**.

Update of $\mathbf{A}$

prox of $\ell_{2,1}$

$$\min_{\mathbf{A}} \frac{1}{2}\|(\mathbf{X}-\mathbf{AS})\|_{\mathrm{F}}^2+\sum_{i=0}^{n}\left\|\gamma \odot (\mathbf{A}^i-\bar{\mathbf{A}}^i)\Psi^T\right\|_{2,1}+\iota_{\cdot\geq 0}(\mathbf{A})+\iota_{\|\cdot\|_2=1}(\mathbf{A}).$$

At iteration $(k+1)$,γ acts as a threshold applied to $\|\mathbf{A}^{(k)}-\bar{\mathbf{A}}+\frac{1}{L_{\mathbf{A}}}(\mathbf{X}-\mathbf{A}^{(k)}\odot\mathbf{S})\mathbf{S}^T\|_2$ (χ distribution). Following the previous reasonning (applied to a χ distribution), we have:

$$\forall i,\;\; \lambda_i \simeq 1.5\; k\; \mathrm{MAD}(\|(\mathbf{A}^{(k)}-\bar{\mathbf{A}}+\frac{1}{L_{\mathbf{A}}}(\mathbf{X}-\mathbf{A}^{(k)}\odot\mathbf{S})\mathbf{S}^T)_i\|_2)$$

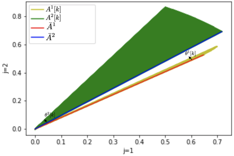
To allievates biases errors:

$$\forall i, \forall k, \gamma_i[k]=\lambda_i\frac{w_i[k]}{w_i[k]+\epsilon}$$

Numerical experiments

Dataset

- 2 sources, 1500 samples and 5 observations,
- sources generated from Generalized Gaussian distribution with $\rho = 0.3$,
- Spectral variabilities exactly sparse in DCT domain (2 activated coefficients for each source) and low frequency,
- Maximal amplitude of $\frac{\theta}{6}$,
- comparison with GMCA and GMCA per patch.



Projection of the two first sources onto the slice defined by the two first observations of the hypersphere $\mathcal{S}^{m-1}$

Numerical experiments

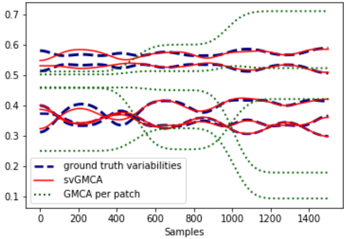


Figure: Spectral variabilities in the sample domain

	$\theta = \frac{\pi}{6}$	$\theta = \frac{\pi}{8}$	$\theta = \frac{\pi}{12}$
svGMCA	$\mathbf{9.97 \times 10^{-5}}$	$\mathbf{1.11 \times 10^{-4}}$	$\mathbf{7.13 \times 10^{-5}}$
GMCA per patch	8.69×10^{-3}	3.26×10^{-4}	1.57×10^{-4}
GMCA	9.63×10^{-3}	4.07×10^{-4}	9.31×10^{-3}

Table: GMSE for various angles

Outline

Conclusion

First half of the PhD

- **pGMCA**
  - pGMCA has been presented in iTwist'19 conference
  - Journal papier on pGMCA submitted on IEEE journal
  - pGMCA is being currently applied to Chandra data by F.Acero[9].
- **svGMCA**
  - Preliminary results
  - svGMCA submitted to SPARS conference

[9] Département d' Astrophysique, CEA

Conclusion

Second half of the PhD

- **svGMCA**
  - Encouraging preliminary results on SVs.
  - Work on progress: leverage morphological diversity between the sources and the spectral variabilities
  - Preparation of a journal article with focus on the introduced methodology
  - Application to high-energy astronomical data
- **spectral variabilities with shape information**
  - Methodology and applications to come.

Acknowledgement

This work is supported by the European Community through the grant LENA (ERC StG - contract no. 678282).