Introduction

Sparse BSS from Poisson measurements

Sparse BSS with spectral variabilities

Conclusion

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

1. Introduction

2. Sparse BSS from Poisson measurements
   - Overview of the problem
   - Full description of the pGMCA algorithm
   - Applications in astrophysics

3. Sparse BSS with spectral variabilities
   - Overview of the problem
   - Proposed algorithm
   - Preliminary results

4. Conclusion
Multivalued data analysis with BSS

Introduction to BSS

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

\[ A^1S_1 + A^2S_2 + A^3S_3 + \ldots + N = X \]

- Blind Source Separation (BSS) aims at disentangling mixed components to retrieve spectral and spatial information.
Multivalued data analysis with BSS

**Introduction**

**Introduction to BSS**

\[
\min_{\mathbf{A}, \mathbf{S}} \| \mathbf{X} - \mathbf{A} \mathbf{S} \|^2_F. \tag{1}
\]

Ill-posed problem requiring further assumptions:

- Statistical independence of the sources: **Independent Component Analysis** \(^1\),
- Non-negativity of the components: **Non-negative Matrix Factorization** \(^2\),
- 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

Université Paris-Saclay

Laboratoire Cosmostat, CEA Paris-Saclay
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
**Introduction**

Sparse BSS from Poisson measurements

Sparse BSS with spectral variabilities

**Conclusion**

**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 \( X = AS + N \) is no more valid.

Observation derived from Chandra simulations 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 data dependent.
Challenges raised by high-energy imaging

2 Spectral variabilities:
- spatially variant spectra are ubiquitous 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.

Necessity of a method fully accounting for spectral variabilities.
Outline

1. Introduction

2. Sparse BSS from Poisson measurements
   - Overview of the problem
   - Full description of the pGMCA algorithm
   - Applications in astrophysics

3. Sparse BSS with spectral variabilities
   - Overview of the problem
   - Proposed algorithm
   - Preliminary results

4. Conclusion
Model \( \mathbf{X} = \mathbf{AS} + \mathbf{N} \) not valid.

The noise corrupting high energy data follows shot noise statistics.

Poisson law

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

(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

\[ P(X_j[t] | [AS]_j[t]) = \frac{e^{-[AS]_j[t][AS]_j[t]} X_j[t]}{X_j[t]!} \]

Optimization function

\[
\begin{align*}
\min_{A,S} \quad & \mathcal{L}(X | AS) + \| \Lambda \odot S \Phi^T \|_1 + i_{\geq 0}(S) + i_C(A), \\
\text{antiloglikelihood} \quad & \text{constraints on } S \quad \text{constraints on } A
\end{align*}
\]

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

\[
\mathcal{L}(X | AS) = \sum_{j,t} [AS]_j[t] - X_j[t] \log([AS]_j[t])
\]

\[
= AS \odot \log(AS)
\]

\( \Lambda \) contains the regularization parameters;
\( C = OB(m) \cap K^+ \).
Overview of the problem
Sparse BSS with Poisson measurements

Optimization function

\[
\min_{A,S} \mathcal{L}(X \mid AS) + \left\| \Lambda \odot S\Phi^T \right\|_1 + i_{\geq 0}(S) + i_C(A), \quad (5)
\]

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

\[
\mathcal{L}(X \mid AS) = AS - X \odot \log(AS).
\]

\(^5\) Tseng 2001
Smooth approximation of the data-fidelity term

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

\[ \mathcal{L}_\mu(X \mid Y) = \inf_U \langle Y, U \rangle - \mathcal{L}^*(X \mid U) - \mu \|U\|_F^2, \]

Fenchel dual of $\mathcal{L}$

regularization function

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:

\[ \min_{A,S} \mathcal{L}_\mu(X \mid AS) + \|\Lambda \odot S\Phi^T\|_1 + i_C(A) + i_{\geq 0}(S), \]
Multiconvex problem
Block Coordinate Descent algorithm (BCD)

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

BCD: alternative estimation of the convex subproblems.

**Structure of pGMCA algorithm**

- **Initialization**: \((A^{(0)}, S^{(0)})\) obtained with robust (to initialization) sparse BSS algorithm GMCA\(^a\).

- **Iteration** \(k\):
  - Update of \(A\) assuming \(S\) fixed,
  - Update of \(S\) assuming \(A\) fixed.

\(^a\) Bobin et al, 2007
Update of $\mathbf{A}$ assuming $\mathbf{S}$ 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 $\frac{\mu}{\|\mathbf{S}^T\mathbf{S}\|_2}$,
- $i_C(.)$ is an indicator function of a convex set, it is proximable.

Update of $\mathbf{A}$ at iteration $(k)$

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

$^1$ Fast Iterative Shrinkage-Thresholding Algorithm, Beck et al 2009
Update of $S$ assuming $A$ fixed

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

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

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

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

**Question:** how to set the regularization parameter $\Lambda$?

---

1. Raguet et al, 2013
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

\[ X^{(k+1)} = S_\lambda \left( X^{(k)} + \nabla_X \| Y - AX \|_F^2 \right). \]

\( S_\lambda \): sparse contribution

\( \nabla_X \| Y - AX \|_F^2 \): gaussian noise contribution

\( X + \mathcal{G} \) (blue) and \( X \) (red).
We want to **remove** gaussian noise and **keep** sparse signal coefficients (of $X$) with **their amplitudes**.

$$S_\lambda(X + G) = Z \text{ s.t. } \forall j \ Z_j = (|X_j + G_j| - \lambda)_+.$$

- "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 $\sigma_G = 1.48 \text{MAD}(G)$
- $\text{MAD}(X + G) = \text{MAD}(G)$

\[ \lambda \text{ set up as a denoising threshold} \]

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

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

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

where $k = 1$ for Poisson noise; and to limit biaisès $^a$:

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

$^a$Candes et al., 2008
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

- **Generalized Morphological Component Analysis:** standard sparse BSS algorithm;
- **Hierarchial Alternating Least Square** algorithm: NMF algorithm with sequential updates \(^6\);
- **\(\beta\) NMF:** NMF with Kullback-Leibler divergence \((\beta=1)\) \(^7\);
- **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: \( \text{SAD} = \frac{1}{n} \sum_{i=1}^{n} \arccos(\langle A^i | A_g^i \rangle) \) with \( A \) the mixing matrix recovered with the proposed approach and \( A_g \) the ground-truth mixing matrix.

![Graph showing the performance of different algorithms with SAD metric]
Overview of the problem

The mixing matrix is pixelwise dependent:

\[ \forall \text{ sample } k \quad X[k] = \sum_{i=1}^{n} A^i[k]S_i[k] + N[k], \quad (11) \]

State-of-the-Art model: Perturbed Linear Mixture Model (PLMM) \(^a\)

\(^a\) Thouvenin et al, 2016

\[ X = \bar{A}S + \Delta AS \]

- Applied to hyperspectral terrestrial images.
- Sum-to-one assumption \( \sum_i S_i = 1 \): the sources are not independent.
- Pure pixel assumption \( \forall i, \exists k' / X[k'] = A^i \).
Overview of the problem
Linearization and angular variability

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

$$\forall i, j, k \ A^j_i[k] = \bar{A}^j_i + \Delta^j_i[k]$$

Angular variability:

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

since $A \in \text{OB}(m)$ in the sparsity framework.
Overview of the problem
Linearization and spatial regularization

$\forall i, j, k \ A_{ij}^k = \bar{A}_{ij} + \Delta_{ij}^k$ with $\|\Delta_{ij}^k\|_2 = \theta_{ij}^k$.

Optimization function:

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

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

Spatial regularization of the SV:

$$\forall i \| \gamma \odot \theta^i \Psi^T \|_1 \approx \| \gamma \odot (A^i - \bar{A}^i) \Psi^T \|_{2,1}. \quad (12)$$
Optimization function

\[
\min_{A,S} \frac{1}{2} \left\| (X - \sum_i A^i S_i) \right\|_F^2 + \|\Lambda \odot S\|_1 + \sum_{i=0}^n \left\| \gamma \odot (A^i - \bar{A}^i) \psi^T \right\|_{2,1} + i_C(A)
\]

where

- \( \Lambda \) (resp \( \gamma \)) contains the regularization parameters and weights for the source matrix and resp. the spectral variabilities,
- \( C = OB(m) \cap K^+ \).

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

Multiconvex problem $\rightarrow$ 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

$$\min_S \frac{1}{2} \left\| \left( X - \sum_i A^i S_i \right) \right\|_F^2 .$$

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

Université Paris-Saclay

Laboratoire Cosmostat, CEA Paris-Saclay
Update of $\mathbf{A}$ assuming $\mathbf{S}$ is fixed

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

- 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 $A$
prox of $\ell_{2,1}$

$$\min_A \frac{1}{2} \| (X - AS) \|_F^2 + \sum_{i=0}^n \left\| \gamma \odot (A^i - \bar{A}^i) \psi^T \right\|_{2,1} + \iota \geq 0(A) + \iota \| \iota = 1(A).$$

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

$$\forall i, \lambda_i \approx 1.5 \, k \, \text{MAD}(\| (A^{(k)} - \bar{A} + \frac{1}{L_A} (X - A^{(k)} \odot S)S^T)_i \|_2)$$

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 $S^{m-1}$
Numerical experiments

Figure: Spectral variabilities in the sample domain

<table>
<thead>
<tr>
<th></th>
<th>$\theta = \frac{\pi}{6}$</th>
<th>$\theta = \frac{\pi}{8}$</th>
<th>$\theta = \frac{\pi}{12}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>svGMCA</td>
<td>9.97 $\times 10^{-5}$</td>
<td>1.11 $\times 10^{-4}$</td>
<td>7.13 $\times 10^{-5}$</td>
</tr>
<tr>
<td>GMCA per patch</td>
<td>8.69 $\times 10^{-3}$</td>
<td>3.26 $\times 10^{-4}$</td>
<td>1.57 $\times 10^{-4}$</td>
</tr>
<tr>
<td>GMCA</td>
<td>9.63 $\times 10^{-3}$</td>
<td>4.07 $\times 10^{-4}$</td>
<td>9.31 $\times 10^{-3}$</td>
</tr>
</tbody>
</table>

Table: GMSE for various angles
Outline

1. Introduction

2. Sparse BSS from Poisson measurements
   - Overview of the problem
   - Full description of the pGMCA algorithm
   - Applications in astrophysics

3. Sparse BSS with spectral variabilities
   - Overview of the problem
   - Proposed algorithm
   - Preliminary results

4. Conclusion
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).