A sparsity-based method to restore MUSE hyperspectral data

Dedicated Algorithms for Hyperspectral Imaging in Astronomy

url: dahlia.oca.eu

E. Slezak – S. Bourguignon – D. Mary
(UNS / CNRS / Observatoire de la Côte d'Azur)

- References -
Bourguignon, S., Mary, D., Slezak, E.: 2011, Statistical Methodology 9(1)
The era of integral-field spectroscopy

- giraffe/flames@vlt
- sinfoni@vlt
- kmos@vlt2
- osiris@keck
- miri@jwst

sauron@wht

crism@mro

- nirspec@jwst

muse@vlt2

XShooter@vlt2

sphere@vlt2
Astronomical hyperspectral data: specificities

• **Low signal to noise ratio** (S/N < 1 dB):
  → the high noise level and its detailed statistics have to be taken into account
  → **fusion/combination** of exposures (deep field, large field: mosaicing)

• **Ground-based observations** mostly (cf. MUSE @ VLT2):

  – perturbations from: atmosphere, AO, telescope, instrument, ...
  → **PSF determination** (spatial and wavelength dependencies) + deconvolution
  – objects are superimposed onto a (varying) background
  → **background estimation** and subtraction

• **Astrophysical sources are diverse**: 
  – spatial and spectral shapes may be very different → **segmentation / detection**
  – overlaps and crowding → **source separation**

• **Massive and complex data**:
  → data **visualization** and simulation
  300x300 x 4000 elements
Interest of hyperspectral data in astrophysics

- various physical processes with different spectral behaviours
- discovery of distant (emission-line) objects whatever their redshifts are
MUSE hyperspectral data Segmentation
astronomical objects vs. astrophysical sources vs. homogeneous zones

• reminder :
  – PSF variations + expected DRS residuals
  – very low S/N; a spectrally varying noise
  – lack of spectral continuum may occur → LAEs
  – dimension heterogeneity: images / spectra

• Two strategies:
  – a 2+1 D cube: a set of shapes with a spectral information (Marked Point Process)
  – a 1+2 D cube: a set of spectra (cf. MUSE instrumental characteristics) to be aggregated
    1) use and restore the spectral information
       (noise reduction, deconvolution, characterization)
    2) segmentation

  chosen approach
  huge data set → dimension reduction is welcomed
  very low S/N → a robust method is mandatory
  an inverse problem with parcimony constraints
MUSE instrument and noise specificities

- MUSE response → spatial and spectral spreading variable with the wavelength
  \[ \text{PSF : Field Spread Function (FSF) + Line Spread Function (LSF)} \]

- a spectrally variable noise: (strong) atmospheric emission lines, q. efficiency, laser star (AO)
Restoration of MUSE-like data

Inverse problem: observations: \( Y = H \, X + N \) → estimate \( X \)?

Direct inversion: \( H^{-1} \, X + H^{-1} \, N \) → noise amplification
> constrain restoration with additional prior assumptions <

Prior information: sparsity in the spectral domain where the information is located
> galaxy spectrum: a set of elementary features (continuum, lines, discontinuities, etc.) <

→ dictionary \( W \) of possible spectral features

\[
W = [W_{\text{lines}}, W_{\text{steps}}, W_{\text{cos}}]
\]
\( N \approx 3500 \text{ données} \times M \approx 26000 \text{ atomes} \)

\[
x = W \, u \text{ with } u \text{ sparse}
\]
Estimation setting

Considering \( Y = HX + N \) the estimate of \( X \) is given by \( \min_X \| Y - HX \|_2^2 \)

so that, for each pixel \( k \ (1,...,K) \), \( x_k = W x_k \) with sparse \( x_k \)

s.t. \( X = W^{(K)} U \) with \( W^{(K)} \) made of \( K \) blocks \( W \)

\[ \hat{U} = \arg \min_U \| Y - HW^{(K)} U \|_2^2 + \alpha \| U \|_1, \alpha > 0 \]

→ estimation of active coefficients in each spectrum : \( \hat{U}_* = \{ \hat{U}_j \neq 0 \} \)

i.e. detection of significant spectral features

\[ \| . \|_1 \rightarrow \text{amplitudes are biased} \rightarrow \]

\[ \hat{U}' = \arg \min_{U_*} \| Y - HW^{(K)} U_* \|_2^2 \]

→ spectra restoration : \( \hat{X} = W^{(K)} \hat{U}' \)

• spatial and spectral deconvolution with only spectral prior information

• 15 x 15 pixels, 4 000 wavelengths → 9 \( 10^5 \) data points, 7 \( 10^6 \) unknowns
A two-step (sub-optimal) procedure

\[ \hat{U} = \arg \min_U \| Y - HW^{(K)}U \|^2 + \alpha \| U \|_1, \alpha > 0 \]  
(eq.1)

i) Decompose all spectra \( y_k \) in \( Y \) independently:

→ sparse approximation of a spatially spread spectral information

\[ \hat{u}_k = \arg \min_u \| y_k - L_k W u_k \|^2 + \beta \| u_k \|_1, \beta > 0 \]  

(\( L_k \): LSF at pixel \( k \))

→ detection of significant atoms in \( y_k \) : \( W \Omega_k, \Omega_k = \{ j | u_{kj} \neq 0 \} \)

in practice: **low \( \beta \) values** → detection of **faint features**, but more false alarms

ii) Consider eq.1 with \( W \) being restricted to the atoms selected at step i)

• \( x_k = W \Omega_k u_{\Omega_k} \)

• \( x_k = W \Omega'_k u_{\Omega'_k} \) with \( \Omega'_k = \cup_{j \in V(k)} \Omega_j \) and \( V(k) \) a local neighbourhood < FSF

→ optimization in a parameter space \( \Omega \) of lower dimensionality

\[ \hat{U} = \arg \min_U \| Y - HW_{\Omega} U_{\Omega} \|^2 + \gamma \| U_{\Omega} \|_1, \gamma > 0 \]

in practice: the 2\textsuperscript{nd} sparsity constraint allows one to remove false alarms from i)
i) Dictionary normalization:
Let us consider a Gaussian noise with covariance matrix $\Sigma$. The data misfit term is:

$$\|Y - HWU\|_\Sigma^2 = (Y - HWU)^T \Sigma^{-1} (Y - HWU) = \|\Sigma^{-1/2} Y - \Sigma^{-1/2} HWU\|^2$$

- the equivalent dictionary $\Sigma^{-1/2} H W$ is not normalized
- column normalization is necessary for coherent detection statistics

ii) Optimization specificities:
- $W$ structure: no fast transform is available
- involv. sizes: matrix storage is impossible
→ Iterative Coordinate Descent algorithm with specific accelerations
Experimental results: a single spectrum

- Initial spectrum
- Convolved and noisy spectrum
- Restored: $\beta = 4.2$
- Restored: $\beta = 3.5$
- Restored: $\beta = 2.5$
Experimental results: two point sources

- Original data (X)
- Convolved data (HX)
- Convolved and noisy data (Y)
- Individual spectral restoration
- Spatial-spectral restoration
Computational aspects

Image $9 \times 9$ ; FSF $5 \times 5$ $\rightarrow$ restored image : $13 \times 13$ pixels

3 463 wavelengths $\rightarrow 280\,000$ data

- whole dictionary : $\sim 4 \times 10^6$ unknowns, computation time $= ???$
- restoration of $x_k$ using only atoms selected on pixel $k$
  $\sim 10^3$ unknowns, 82 active coefficients, execution time $= 6$ mn
- restoration of $x_k$ using selected atoms within a $5 \times 5$ neighbourhood
  $\sim 10^4$ unknowns, 89 active coefficients, execution time $= 45$ mn

initial data                       blurred and noisy data averaged images                       restored data
Experimental results: spectral unmixing

\[ X_{\text{true}} = a_1 \times s_1 + a_2 \times s_2 \]

original X  convolved HX  convolved and noisy Y

\[ \hat{X} = \hat{a}_1 \times \hat{s}_1 + \hat{a}_2 \times \hat{s}_2 \]
Summary and Perspectives

- A restoration scheme which accounts for spectrally varying PSF and noise
- Spatial and spectral deconvolution with spectral prior information
- Sparsity: physically motivated (vs. generic) set of atoms → robustness
- Efficiency in separating point sources and in “simple” unmixing cases

> Topics in progress <

→ Dictionaries
→ Automatic tuning of hyperparameters
→ More complex priors for specific problems (cf. extended sources)
→ Source detection and characterization:
  - Group pixels with the same spectral signature
    (distance measure: KL; statistical decision)
    A first attempt with a basic criterion ... →
  - No constraint on morphologies
  - Vs. input for a MPP approach?
→ Very faint but extended structures