Processing MUSE hyperspectral data: denoising, deconvolution and detection of astrophysical sources.

Sébastien Bourguignon, David Mary and Éric Slezak

Observatoire de la Côte d’Azur, CNRS, University of Nice

Astronomical Data Analysis 6th conference, Monastir, 3-7 May 2010
Outline

1. Astrophysical context: MUSE instrumental project
   - Expected data: images and spectra
   - Expected noise characteristics

2. Sparse representations for "denoising"
   - General framework
   - Observational model: accounting for instrumental specificities

3. Results on realistic simulations
   - Restoration of spectra and images
   - Towards detection: how to exploit extracted information?

4. Conclusions and further work
Astrophysical context: MUSE instrumental project
- Expected data: images and spectra
- Expected noise characteristics

Sparse representations for "denoising"
- General framework
- Observational model: accounting for instrumental specificities

Results on realistic simulations
- Restoration of spectra and images
- Towards detection: how to exploit extracted information?

Conclusions and further work
MUSE instrument: Multi Unit Spectroscopic Explorer

© Very Large Telescope (Chile) in 2012
Integral-field spectrograph:
- 4000 wavelengths (450 – 900 nm)
- 300 × 300 pixels

→ exploring deep field space: 1 arcmin²
http://www-obs.univ-lyon1.fr/muse

What will data look like (simulations):
- spatially: deep field → galaxies = a few pixels
MUSE instrument: Multi Unit Spectroscopic Explorer

@ Very Large Telescope (Chile) in 2012
Integral-field spectrograph:
- 4000 wavelengths (450 – 900 nm)
- 300 × 300 pixels
→ exploring deep field space: 1 arcmin²
http://www-obs.univ-lyon1.fr/muse

What will data look like (simulations):
- spatially: deep field → galaxies = a few pixels
- spectrally: lines
MUSE instrument: Multi Unit Spectroscopic Explorer

© Very Large Telescope (Chile) in 2012
Integral-field spectrograph:
- 4000 wavelengths (450 – 900 nm)
- 300 × 300 pixels

→ exploring deep field space: 1arcmin²
http://www-obs.univ-lyon1.fr/muse

What will data look like (simulations):
- spatially: deep field → galaxies = a few pixels
- spectrally: lines, continuum
MUSE instrument: Multi Unit Spectroscopic Explorer

© Very Large Telescope (Chile) in 2012
Integral-field spectrograph:
- 4000 wavelengths (450 – 900 nm)
- 300 × 300 pixels
→ exploring deep field space: 1 arcmin$^2$
http://www-obs.univ-lyon1.fr/muse

What will data look like (simulations):
- spatially: deep field → galaxies = a few pixels
- spectrally: lines, continuum, both...
MUSE instrument: Multi Unit Spectroscopic Explorer

© Very Large Telescope (Chile) in 2012
Integral-field spectrograph:
- 4000 wavelengths (450 – 900 nm)
- 300 × 300 pixels

→ exploring deep field space: 1arcmin²
http://www-obs.univ-lyon1.fr/muse

What will data look like (simulations):
- spatially: deep field → galaxies = a few pixels
- spectrally: lines, continuum, both...
- challenge: detect and characterise faint sources (galaxies) with small spatial extensions and variable spectral profiles
1. **Astrophysical context: MUSE instrumental project**
   - Expected data: images and spectra
   - Expected noise characteristics

2. **Sparse representations for ”denoising”**
   - General framework
   - Observational model: accounting for instrumental specificities

3. **Results on realistic simulations**
   - Restoration of spectra and images
   - Towards detection: how to exploit extracted information?

4. **Conclusions and further work**
Sky Background Emission (data collected through the atmosphere)
\[ \text{data}(r, \lambda) = \text{object}(r, \lambda) + \text{SB}, \; \text{SB} \gg \text{object} \]

- small field of view: spatially constant but spectrally variable: \( \overline{\text{SB}}(\lambda) \)
- sky-subtracted data: \( \text{object}(r, \lambda) + \text{SB}(r, \lambda) - \overline{\text{SB}}(\lambda) \)

MUSE efficiency is spectrally variable:

\[ e(\lambda) = \frac{\# \text{detected photons}}{\# \text{received photons}} \]
\[ \Rightarrow \sigma(\text{noise}) \propto \frac{1}{e(\lambda)} \]

Dark-current and read-out white noise
Finally:

\[ SB(\lambda) + \sigma_{\text{eff}}(\lambda) + \sigma_{\text{elec}}(\lambda) \]

⇒ highly \textit{coloured} noise with known \( \sigma^2(\lambda) \):

true spectrum \hspace{1cm} MUSE-like spectrum (SNR 14 dB)
Finally:

\[ SB(\lambda) + \sigma_{\text{eff}}(\lambda) + \sigma_{\text{elec}} \]

\[ \Rightarrow \text{highly} \ \textit{coloured} \ \text{noise with known} \ \sigma^2(\lambda): \]

\[ \sigma(\lambda) \]

true spectrum \hspace{1cm} MUSE-like spectrum (SNR -3 dB)
Finally:

\[ SB(\lambda) + \sigma_{\text{eff}}(\lambda) + \sigma_{\text{elec}} \]

⇒ highly \textit{coloured} noise with known \( \sigma^2(\lambda) \):

true spectrum \hspace{2cm} MUSE-like spectrum (SNR -3 dB)

- high noise level with high spectral variability
- spectra: strong parasite lines
  ⇒ denoising is mandatory

S. Bourguignon (OCA - University of Nice)  Processing MUSE hyperspectral data  ADA VI, Monastir, May 2010
1. Astrophysical context: MUSE instrumental project
   - Expected data: images and spectra
   - Expected noise characteristics

2. Sparse representations for "denoising"
   - General framework
   - Observational model: accounting for instrumental specificities

3. Results on realistic simulations
   - Restoration of spectra and images
   - Towards detection: how to exploit extracted information?

4. Conclusions and further work
Denoising: from observed spectrum $y = s + \epsilon$, we want to restore $s$.

Suppose transform $W$ concentrates most information of $s$ in few points...

Denoising: apply $W$ to $y$ and only keep highest coefficients in $Wy = z$ (hard-thresholding):

$$\hat{z}_n = z_n \text{ if } |z_n| > \tau, \quad \hat{z}_n = 0 \text{ if } |z_n| \leq \tau,$$

then $s_{\text{denoised}} = W^{-1}\hat{z}$

Threshold at $\tau = 4\sigma$

21 coefficients

SNR $+11$ dB
Sparsity-based denoising: a simple example

[Donoho and Johnstone, 1994]

- **Denoising**: from *observed* spectrum $y = s + \epsilon$, we want to restore $s$.
- **Suppose transform $\mathbf{W}$ concentrates most information of $s$ in few points...**

![Graph showing signal energy concentrated in a few large coefficients]

- **Denoising**: apply $\mathbf{W}$ to $y$ and only keep highest coefficients in $\mathbf{Wy} = z$ (*hard-thresholding*):

  $$\hat{z}_n = z_n \text{ if } |z_n| > \tau, \quad \hat{z}_n = 0 \text{ if } |z_n| \leq \tau,$$

  then $\hat{s}_{\text{denoised}} = \mathbf{W}^{-1} \hat{z}$

- **Threshold at $\tau = 3\sigma$**
  - 37 coefficients
  - SNR +12 dB
Denoising: from *observed* spectrum $y = s + \epsilon$, we want to restore $s$.

Suppose transform $\mathbf{W}$ *concentrates most information of* $s$ *in few points*.

![Graph showing signal energy concentrated in a few large coefficients](image)

Denoising: apply $\mathbf{W}$ to $y$ and only keep highest coefficients in $\mathbf{Wy} = \mathbf{z}$ (*hard-thresholding*):

$$\hat{z}_n = z_n \text{ if } |z_n| > \tau, \quad \hat{z}_n = 0 \text{ if } |z_n| \leq \tau,$$

then $\hat{s}_{\text{denoised}} = \mathbf{W}^{-1}\hat{z}$

Threshold at $\tau 2\sigma$

195 coefficients

SNR + 6 dB
Sparsity-based denoising: a simple example

[Donoho and Johnstone, 1994]

- Denoising: from observed spectrum \( y = s + \epsilon \), we want to restore \( s \).

- **Suppose transform \( W \) concentrates most information of \( s \) in few points.**

![Graphs showing signal energy concentration](image)

- Denoising: apply \( W \) to \( y \) and only keep highest coefficients in \( Wy = z \) (**hard-thresholding**):

\[
\hat{z}_n = z_n \text{ if } |z_n| > \tau, \quad \hat{z}_n = 0 \text{ if } |z_n| \leq \tau, \text{ then } \hat{s}_{\text{denoised}} = W^{-1}\hat{z}
\]

Denoising efficiency \( \iff \) find transform \( W \) adapted to data
Sparse modelling for galaxies’ spectra

\[ s = s^\ell \text{ (lines)} + s^c \text{ (continuum)} \]

- \( s^\ell \) line spectrum: sparse in the canonical basis
- \( s^c \) smooth: \( s^c \) can be approximated by few coefficients in DCT space:
  \[ x^c = \text{DCT}(s^c) \text{ has only few } x^c_n \neq 0 \]
  \[ \iff s^c = \text{DCT}^{-1}(x^c) \text{ with sparse } x^c \]

\[ s = s^\ell + s^c = I_N x^\ell + W^c x^c = \underbrace{[I_N \ W^c]}_{W} \begin{bmatrix} x^\ell \\ x^c \end{bmatrix} = W x, \ x \text{ sparse} \]

- Complementary properties: \( s \) not sparse in canonical (or DCT) basis alone (cf. [Donoho & Huo, 2001])
Denoising with overcomplete dictionary \( \mathbf{W} \)

**Observed data:** \( \mathbf{y} = \mathbf{Wx} + \epsilon, \quad \mathbf{W} = [\mathbf{I} \ \mathbf{W}^c] \) *dictionary* of \( \delta \) and cosine *atoms*

- \( \mathbf{W} \) overcomplete \((N \times 2N)\): infinity of \( \hat{x} \) with \( \mathbf{y} = \mathbf{W}\hat{x} \)
  - but \( \mathbf{y} \) noisy \( \Rightarrow \mathbf{y} = \mathbf{W}\hat{x} \) not satisfactory (\( \mathbf{y} \) overfitted)
- Better find a *sparse* solution that *approximately* fits the data
  - \( \mathbf{W} \) not invertible: algorithms are required

\[
\hat{x} = \arg \min_{\mathbf{x}} \frac{1}{2} \| \mathbf{y} - \mathbf{Wx} \|^2 + \lambda \| \mathbf{x} \|_1 \quad , \quad \| \mathbf{x} \|_1 = \sum_k |x_k| \]

- For adequate \( \lambda \), \( \hat{x} \) is sparse
- Many optimisation algorithms available to compute \( \hat{x} \)

- Restored spectrum \( \hat{s} = \mathbf{W}\hat{x} = [\hat{x}^\ell + \mathbf{W}^c\hat{x}^c] \), \( \hat{x}^\ell \) lines, \( \hat{x}^c \) continuum
1. Astrophysical context: MUSE instrumental project
   - Expected data: images and spectra
   - Expected noise characteristics

2. Sparse representations for "denoising"
   - General framework
   - Observational model: accounting for instrumental specificities

3. Results on realistic simulations
   - Restoration of spectra and images
   - Towards detection: how to exploit extracted information?

4. Conclusions and further work
Accounting for the Line Spread Function: "deconvolution"

**LSF**$_\lambda$: spectral response to a spectral line at wavelength $\lambda$

\[ s_{\text{LSF}}(\lambda_n) = \sum_k s(\lambda_{n-k}) \cdot \text{LSF}_{\lambda_n}(\lambda_k) \]

→ model $y = Hs + \epsilon$ with $n^{th}$ line of $H = \text{LSF}_{\lambda_n}$

Recall that $s = Wx$: $\hat{x} = \arg\min_x \frac{1}{2} \|y - HWx\|^2 + \lambda \|x\|_1$

**LSF variability:** \[ \Rightarrow \text{H is not a Toeplitz matrix} \]
Accounting for noise statistics

Noise is colored: \( y = HWx + \epsilon, \) \( \text{Var} \ \epsilon_n = \sigma^2(\lambda_n): \)

\[
\chi^2 = \sum_n \frac{(y_n - (HWx)_n)^2}{2\sigma_n^2}
\]

Usually, \( \hat{x} = \arg \min_x \frac{1}{2} \left( \| y - WX \|^2 + \lambda \| x \|_1 \right) \)

but \( \epsilon_n \sim \mathcal{N}(0, \sigma_n^2) \)

With \( \Sigma \overset{\Delta}{=} \text{Cov} \ \epsilon = \text{diag}\{\sigma_n^2\}_n, \) one has equivalently:

\[
\chi^2 = \frac{1}{2} \left\| \Sigma^{-1/2}y - \Sigma^{-1/2}HWx \right\|^2 = \frac{1}{2} \left\| z - Bx \right\|^2 \]

"whitened" data \( z_n = \frac{y_n}{\sigma_n} \)

weighted dictionary \( B = \Sigma^{-1/2}HW \)
Hyperparameter tuning

\[ \hat{x} = \arg \min_x \frac{1}{2} \left\| \sum^{-1/2} y - \sum^{-1/2} HW z B x \right\|^2 + \lambda \|x\|_1 \]

Let \( \hat{e} = z - B \hat{x} \), one has (KKT conditions):

\[ |b_k^T \hat{e}| = \lambda \text{ for } \hat{x}_k \neq 0 \text{ and } |b_k^T \hat{e}| < \lambda \text{ for } \hat{x}_k = 0 \]

- **\( \lambda \): detection threshold on \( b_k^T \hat{e} \)**
- if \( \hat{e} = \) noise, then \( b_k^T \hat{e} \sim \mathcal{N}(0, \|b_k\|^2) \) \( \Rightarrow \lambda = q \|b_k\| \) (\( q = 2, 3 \ldots \))

and \( \|b_k\| \) depends on \( k \) because of LSF (\( H \)) and noise (\( \sum^{-1/2} \))

\[ \rightarrow \text{ instead of } \lambda \|x_1\|, \text{ use } \sum_k \lambda_k |x_k|, \text{ with } \lambda_k = q \|b_k\| \]
Optimisation

\[
\min_{x} J(x) = \frac{1}{2} \| z - \Sigma^{-1/2} H W x \|^2 + \sum_{k} \lambda_k |x_k|, \text{ with } \begin{cases} \\ \Sigma^{-1/2}, \ N \text{ diagonal} \\ H \text{ convolution} \\ W = [I, \text{DCT}^{-1}] \end{cases}
\]

- Gradient-based strategies (Iterative Thresholding and accelerations...) [Daubechies et al., 2004]
  - computes fast transforms \((W \cdot, W^T \cdot)\) on sparse vectors in \(\mathbb{R}^N\)
  - but \(H \cdot, H^T \cdot\) slow (\(H\) is not a Toeplitz matrix)
  - convergence issues...

- Coordinatewise optimization ICD [Sardy et al., 2000]
  - successive updates \(x_k^{(t)} = \arg \min_{x_k} J(x)\) (soft thresholding)
  - updates only non-zero coefficients ⇒ efficient for sparse \(\hat{x}\)
  - convergence ensured

- Other alternatives are possible...
Outline

1. Astrophysical context: MUSE instrumental project
   - Expected data: images and spectra
   - Expected noise characteristics

2. Sparse representations for "denoising"
   - General framework
   - Observational model: accounting for instrumental specificities

3. Results on realistic simulations
   - Restoration of spectra and images
   - Towards detection: how to exploit extracted information?

4. Conclusions and further work
Restoration of spectra and images \((\lambda_k = q \|b_k\|)\)

- De-noising and deconvolution of spectra

\[ |\hat{x}^c| \text{ (5 coefs)} \quad (q_c = 3.5) \]

\[ \hat{s}^c = \text{DCT}^{-1}(x^c) \]

\[ \hat{s}^l = \hat{x}^l \text{ (0 line)} \quad (q_l = 4) \]

\[ \text{restored } \hat{s}^l + \hat{s}^c \]

S. Bourguignon (OCA - University of Nice)  Processing MUSE hyperspectral data  ADA VI, Monastir, May 2010
Restoration of spectra and images \( (\lambda_k = q \| b_k \| ) \)

- De-noising and deconvolution of spectra

\[
\hat{x}^c \quad (1 \text{ coef}) \quad (q_c = 3.5)
\]

\[
\hat{s}^c = \text{DCT}^{-1}(x^c)
\]

\[
\hat{s}^\ell = \hat{x}^\ell \quad (65 \text{ lines}) \quad (q_\ell = 3)
\]

\[
\text{restored} \quad \hat{s}^\ell + \hat{s}^c
\]
Restoration of spectra and images \( (\lambda_k = q \| b_k \|) \)

- De-noising and deconvolution of spectra

\[ \hat{x}^c \] (1 coefs) 
\[ (q_c = 3.5) \]

\[ \hat{s}^c = \text{DCT}^{-1}(x^c) \]

\[ \hat{s}^\ell = \hat{x}^\ell \] (65 lines) 
\[ (q_\ell = 3) \]

\[ \hat{s}^\ell + \hat{s}^c \] restored

\[ \text{noise-free} \]

\[ \text{noisy} \]
Restoration of spectra and images \((\lambda_k = q \parallel b_k \parallel)\)

- De-noising and deconvolution of spectra

\[
\hat{x}^c \quad (2 \text{ coefs}) \\
(q_c = 3.5)
\]

\[
\hat{s}^c = \text{DCT}^{-1}(x^c)
\]

\[
\hat{s}^\ell = \hat{x}^\ell \quad (6 \text{ lines}) \\
(q_\ell = 3)
\]

\[
\text{restored } \hat{s}^\ell + \hat{s}^c
\]

- Noise-free

- Noisy

S. Bourguignon (OCA - University of Nice)  Processing MUSE hyperspectral data  ADA VI, Monastir, May 2010  18 / 24
Restoration of spectra and images ($\lambda_k = q \|b_k\|$)

- De-noising and deconvolution of spectra

- (Indirect) image restoration: example at $\lambda = 600$ nm
Outline

1. Astrophysical context: MUSE instrumental project
   - Expected data: images and spectra
   - Expected noise characteristics

2. Sparse representations for "denoising"
   - General framework
   - Observational model: accounting for instrumental specificities

3. Results on realistic simulations
   - Restoration of spectra and images
   - Towards detection: how to exploit extracted information?

4. Conclusions and further work
Towards detection: how to exploit extracted information?

Group together neighbour pixels with close estimated parameters:

- **lines:** if \( \hat{s}_1^\ell \) and \( \hat{s}_2^\ell \) share at least one detected line
- **continuum:** if \( \theta(\hat{s}_1^c, \hat{s}_2^c) = \arccos \left( \frac{\langle \hat{s}_1^c, \hat{s}_2^c \rangle}{\| \hat{s}_1^c \| \| \hat{s}_2^c \|} \right) < \theta_{\text{lim}} \)

Mean image (log scale)

Detection based on continuous part

\[ \theta_{\text{lim}} = 35^\circ \]

Detection based on lines

colour = \( \lambda \)
1. Astrophysical context: MUSE instrumental project
   - Expected data: images and spectra
   - Expected noise characteristics

2. Sparse representations for "denoising"
   - General framework
   - Observational model: accounting for instrumental specificities

3. Results on realistic simulations
   - Restoration of spectra and images
   - Towards detection: how to exploit extracted information?

4. Conclusions and further work
Conclusions

- Sparse modelling in two bases (Identity and DCT)
  - adaptation of classic Basis Pursuit De-Noising problem to MUSE "observations": variable LSF, noise statistics
  - hyperparameter tuning: weighted $\ell_1$-norm
  - efficient optimisation (Iterative Coordinate Descent)
  - satisfactory denoising and deconvolution for individual spectra

- Extraction of relevant information
  - location of spectral lines: physically meaningful
  - input to object detection and classification methods
Search for **better bases** than DCT (wavelets, XXlets)

- Dictionaries with **more than 2 bases**
- Include spatial information for **joint spatial-spectral denoising**
  - data-fit term: accounting for **spatial and spectral convolution**
  - **joint sparsity** on neighbour pixels (eg. mixed norms)
  - use 3D atoms

- Final objective for MUSE science = object **detection**
  - \( \forall \) object, spatial extension, estimated spectrum, uncertainties . . .
That’s all!

Thank you for your interest.