Statistical Methodology 9 (2012) 32–43

Contents lists available at SciVerse ScienceDirect

Statistical Methodology

journal homepage: www.elsevier.com/locate/stamet

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

Sébastien Bourguignon [a,*], David Mary [b], Éric Slezak [a]

[a] *Laboratoire Cassiopée UMR 6202, University of Nice Sophia Antipolis, CNRS, Observatoire de la Côte d'Azur, B.P. 4229, 06304 Nice Cedex 4, France*

[b] *Laboratoire A.H. Fizeau UMR 6525, University of Nice Sophia Antipolis, CNRS, Observatoire de la Côte d'Azur, B.P. 4229, 06304 Nice Cedex 4, France*

ARTICLE INFO

Keywords:
Noise reduction
Deconvolution
Sparse decompositions
Detection
Hyperspectral imaging data

ABSTRACT

We address the problem of processing astronomical hyperspectral data cubes in the context of the forthcoming MUSE instrument. MUSE, which is under construction, will provide massive hyperspectral data with about 300×300 pixels at approximately 4000 wavelengths. One of its main astrophysical objectives concerns the observation of extragalactic deep fields, where MUSE should be able to detect and characterize galaxies much fainter than the ones currently observed by other ground-based instruments. The data will suffer, however, from very powerful and spectrally variable perturbations.

In this paper, MUSE data cubes are first considered as a collection of spectra, which are processed independently. A restoration method is proposed, based on the hypothesis that data can be approximated by appropriate sparse representations. Sparsity can be naturally expressed in the spectral domain, where a galaxy spectrum is mainly the superposition of an emission and absorption line spectrum, which is naturally sparse, on a continuum, which is supposed to have a sparse discrete cosine transform. The problem is addressed within the ℓ^1-norm penalization setting. The original features of the model consist, first, in taking into account observational specificities such as the spectrally variable instrumental response and non-identically distributed noise, and, second, in tuning regularization parameters in this setting, which are fixed in order to obtain uniform false alarm rates for decomposition coefficients. In a second step, such

* Corresponding author. Tel.: +33 492 003 029.
E-mail addresses: Sebastien.Bourguignon@oca.eu (S. Bourguignon), David.Mary@unice.fr (D. Mary), Eric.Slezak@oca.eu (É. Slezak).

1572-3127/$ – see front matter
doi:10.1016/j.stamet.2011.04.010

sparse decompositions are used as an input to an object detection and characterization method. The decomposed spectra are first used for spatial segmentation. Then, once a group of pixels has been identified as belonging to the same object, the corresponding spectrum and amplitude map are jointly estimated under the former sparsity assumption. Applications to object identification, the amplitude map and spectrum estimation are presented for realistic deep field simulated data cubes provided by the MUSE consortium.

1. Introduction

The ESO (European Southern Observatory) second-generation VLT (Very Large Telescope) instrument MUSE (Multi-Unit Spectroscopic Explorer) is an extremely powerful integral field spectrograph, which will provide hyperspectral data cubes with 300×300 spatial elements and up to 4000 spectral channels covering the essential part of the visible spectrum [2]. In its wide field mode configuration (covering a field of view of 1 arcmin^2), the expected performance of MUSE should allow the detection of galaxies which appear a hundred million times fainter than the faintest stars observable with the naked eye. A challenging scientific objective is then to detect and characterize highly redshifted astrophysical sources.

Many methods in the field of hyperspectral imaging have been proposed for object detection and segmentation (see, *e.g.,* [5]), that are however not easily transposable to MUSE data cubes. An important characteristic of data that will be provided by MUSE is a very low signal-to-noise ratio (SNR) together with a highly spectrally variable noise distribution, caused by the powerful parasite emission of the atmosphere at specific wavelengths and by instrumental limitations. Hence, taking into account observational (instrument and noise) specificities is a crucial point for achieving the ambitious scientific goals of the instrument. Moreover, usual segmentation methods are not adapted to the high density and diversity of objects in expected deep field observations: there are as many segments as objects in the field – typically, several hundreds or even thousands – whereas usual segmentation methods are efficient for a relatively small number of regions.

Because of the heterogeneity of such three-dimensional data (two spatial dimensions plus one spectral dimension), it seems natural to decompose any processing method into successive steps, operating either in the spectral or in the spatial domain. In this paper, data cubes are first considered as a collection of spectra. Indeed, except at pixels where no source is emitting, every spectrum in the MUSE data mainly corresponds to the spectral signature of a galaxy located at the corresponding spatial coordinates. Consequently, a coherent structure in the data is expected along the spectral axis – but this is not the case spatially, where the field of view is expected to be composed of thousands of objects with different shapes. We propose a restoration method for MUSE-like spectra, based on sparse representations, which has become very much a classical approach to denoising [9,12]. Restoration is formulated as a linear inverse problem, which takes into account the line spread function (LSF) of MUSE – that is, the impulse response in the spectral domain – and the statistical distribution of the noise, both of which are spectrally variable. Estimates are then searched for under additional prior information, which is formulated through sparsity constraints. Indeed, the spectra of galaxies are mainly the superposition of a *line spectrum* (composed of both emission and absorption lines), which is naturally sparse – that is, only a few lines appear in the spectrum – upon a continuous spectrum. Such a continuum is supposed to have smooth variations, and it is modeled here with a sparse representation in the discrete cosine transform (DCT) domain. Estimation is performed by minimizing a quadratic data misfit functional, penalized by the – now standard – ℓ^1-norm of the decomposition coefficients [6]. Such a principle of decomposition into both spiky and continuously varying components was already proposed by Ciuciu et al. [7] with a different penalization approach and by Donoho and Huo [8] with the sparsity-based model used here. In our case, specificities include the LSF and the unusual noise structure, which are shown to affect both the equivalent *dictionary* of

the sparse estimation problem and the tuning of the corresponding weights of the ℓ^1-norm. Section 2 first formulates the inverse problem and details the structure of both the LSF and the noise, and then introduces the sparsity-based spectral prior model. Estimation is addressed in Section 3 within the ℓ^1-penalization framework. We focus on the modifications imposed by the observational model in terms of dictionary and hyperparameter tuning, for which a statistical interpretation is proposed in terms of false alarm rates. Amplitude debiasing is considered, and Monte Carlo (MC) simulations are used to study the validity of this model by assessing line detection rates on a simulated example.

A second part of this paper concerns the exploitation of the former methodology for astrophysical object detection and characterization. We first consider the aggregation of neighbor pixels which have *similar* spectra, in order to define spatial regions with homogeneous spectral properties, which are associated with an astrophysical source. A binary spectral similarity is used, where two spectra are supposed similar if they share at least one detected line at a common wavelength. With such an approach, the former statistical interpretation used in the spectral decomposition in terms of false alarms is propagated, allowing one to derive false alarm probabilities in the object detection procedure. Once a list of pixels has been associated with an object, estimation is addressed again for each object. In this step, we consider a separable model where the corresponding part of the cube is composed of an amplitude map weighting a unique spectrum. Here again, the noise spectral signatures and the LSF are taken into account in a joint data misfit measurement. Estimation is performed on both the amplitude map and the spectrum, with the formerly used spectral sparsity constraints. Pixel aggregation and corresponding object estimation are described in Section 4, together with MC simulations revealing the efficiency of such a joint estimation approach. Finally, Section 5 presents results on realistic deep field simulated data[1] provided by the MUSE consortium.

2. The problem formulation and the main assumptions

2.1. The observational model: the specific line spread function and noise statistics

Let $\boldsymbol{s} = [s_1 \dots s_N]^T$ denote the spectrum of a galaxy, discretized at wavelengths $\lambda_1 \dots \lambda_N$. MUSE observations are formalized by the model

$$\boldsymbol{y} = \mathbf{H}\boldsymbol{s} + \boldsymbol{\epsilon} \tag{1}$$

where $\boldsymbol{y} = [y_1 \dots y_N]^T$ collects the observed data at wavelengths $\lambda_1 \dots \lambda_N$, $\mathbf{H}$ is the $N \times N$ matrix form of the LSF and $\boldsymbol{\epsilon}$ is a perturbation term accounting for noise and model errors.

The LSF describes how the "blurring" affects the spectrum, similarly to the point spread function for images. The nth column of $\mathbf{H}$ is the instrument response to a spectral line at wavelength λ_n, so spectrum $\boldsymbol{s}$ is "convolved":

$$\widetilde{s}_n = \sum\nolimits_m h_{n,m} s_m.$$

In the case of MUSE, spectral spreading varies with wavelength, so coefficients $h_{n,m}$ cannot be written in the usual form h_{m-n} and $\mathbf{H}$ is not a convolution matrix. Fig. 1(a) shows the LSF at both extremities of the spectrum, and a part of the corresponding matrix $\mathbf{H}$.

As far as noise is concerned, we suppose that data are collected at each spatio-spectral element of the cube following

$$d \sim \mathcal{P}(o + b) + r,$$

where $\mathcal{P}(\gamma)$ is the Poisson distribution with parameter γ, o and b are the emission levels of the object of interest and of the contaminating background, respectively, and r is the Gaussian instrumental (readout) noise with variance σ_r^2. Because b is highly energetic and long time exposures are considered, we assume that the light flux is high enough for the approximation $\mathcal{P}(o + b) \sim \mathcal{N}(o+b, o+b)$ to be valid, where $\mathcal{N}(m, \sigma^2)$ is the normal distribution with mean m and variance σ^2. Supposing correct background estimation $\widehat{b}$ (indeed, it can be performed from all 90 000 pixels since

[1] Note that MUSE is still under construction. Data used in this paper were simulated through the instrument numerical model described in [11].

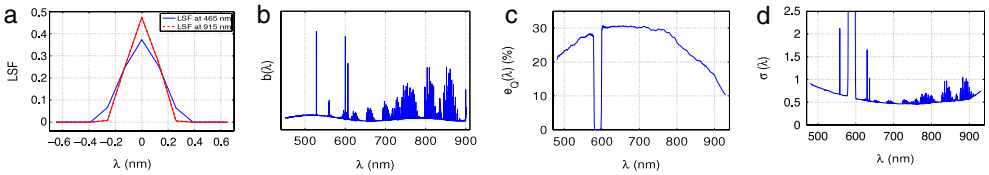


Fig. 1. MUSE observational specificities: (a) the expected LSF at both extremities of the spectrum; ((b), (c), (d)) the typical sky background emission b, the MUSE quantum efficiency e_Q and the typical noise level σ_ϵ respectively as functions of the wavelength. σ_ϵ corresponds to the light flux, in erg s^{-1} cm^{-2}.

it is constant within the field of view of the instrument), our available, background-subtracted, data y read

$$y = d - \widehat{b} \simeq o + \epsilon, \quad \text{where } \epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2),\ \sigma_\epsilon^2 = o + b + \sigma_r^2.$$

The background emission b is mainly caused by the atmospheric emission, which varies with the wavelength, as a typical model in Fig. 1(b) shows: one can see in particular bright emission lines at 558, 630 and 636 nm, corresponding to [OI] emission, and line packets beyond 700 nm, typical for OH emission. Moreover, the MUSE quantum efficiency e_Q (the ratio of emitted electrons to received photons) will vary spectrally (see Fig. 1(c)). A spectral band around $\lambda_{\text{Na}} = 589.2$ nm will also be reserved for laser guide star adaptive optics, such that no signal can be detected in a 20 nm bandwidth around λ_{Na}. All these effects contribute to a highly spectrally variable noise level, as Fig. 1(d) shows.

Note that the true noise variance is unknown, since it contains the unknown object contribution o. In practice, σ_ϵ^2 is estimated as $\widehat{\sigma}_\epsilon^2 = (\widehat{o} + \widehat{b} + \widehat{\sigma}_r^2)/\widehat{e}_Q^2$, where $\widehat{e}_Q$ and $\widehat{\sigma}_r^2$ are obtained from instrument calibration and $\widehat{o}$ is estimated from the data set itself: $\widehat{o} = y$. As regards model (1), we consider in the following that noise samples $\epsilon_1 \dots \epsilon_N$ are independent, following a centered Gaussian distribution with wavelength-dependent variance: $\epsilon_n \sim \mathcal{N}(0, \sigma_n^2)$, where variances σ_n^2 have been estimated as described above.

2.2. A sparsity-based prior model

Given the very low expected signal-to-noise ratio[2](from 15 dB for the few brightest objects to a negative SNR for most of the hundreds of fainter ones), exploiting MUSE spectra requires a restoration procedure. In this work, we consider a prior model on spectra based on sparsity constraints. Sparse approximation has been a very active field of research in the past fifteen years [12]. In this paper, a data set is said to be sparse in a given transform domain if, on applying such a transform to the data, only a few coefficients take significant values. Such an approach has been shown to efficiently perform noise reduction, because most of the signal energy is concentrated in these few, high-valued coefficients, that are less affected by noise than data points are in their original domain.

Here we adopt a sparsity-based approach, where spectrum $\boldsymbol{s}$ is decomposed into the sum of a continuous spectrum $\boldsymbol{s}^c$ and a line spectrum $\boldsymbol{s}^\ell$, both of which are supposed to be sparse in an appropriate domain. Such a model was already proposed by, *e.g.*, Donoho and Huo [8], for the restoration of both spiky and continuous data. More precisely, we consider that:

- $\boldsymbol{s}^c$ can be approximated by a few sine waves, so the DCT of $\boldsymbol{s}^c$ is sparse. Let $\mathbf{W}_{\text{DCT}}$ represent the $N \times N$ DCT matrix and $\mathbf{W}^c = \mathbf{W}_{\text{DCT}}^T$ the inverse DCT matrix. We suppose that $\boldsymbol{s}^c = \mathbf{W}^c \boldsymbol{x}^c$, where only a few coefficients in $\boldsymbol{x}^c$ take significant values.
- $\boldsymbol{s}^\ell$ is composed of a few spectral lines, that is, $\boldsymbol{s}^\ell$ is sparse in the canonical basis. Note that we consider with such a model that lines are *unresolved* at the MUSE resolution, that is, their width does not exceed the spectral discretization step of MUSE, which is $\Delta\lambda \simeq 0.13$ nm. Hence, wider lines will be modeled by contiguous impulsions.

[2] We define the signal-to-noise ratio between noise-free data $\boldsymbol{x}$ and noisy data $\boldsymbol{y}$ as $\text{SNR}_{\text{dB}} = 10 \log_{10} \frac{\sum_n x_n^2}{\sum_n (x_n - y_n)^2}$.

In our experiments, the number of detected lines (of active coefficients in $\boldsymbol{s}^\ell$) varies between 0 and a few tens, depending on the SNR and the physical nature of the source. The number of DCT coefficients in $\boldsymbol{x}^c$ rarely exceeds 10. Compared to the size of the spectra (3000–4000 points), the solution that is searched for is extremely sparse. For homogeneous notation, we will use $\boldsymbol{s}^\ell = \boldsymbol{x}^\ell$ where $\boldsymbol{x}$ is the $2N$-point vector concatenating $\boldsymbol{x}^c$ and $\boldsymbol{x}^\ell$. Then we can write our prior model in matrix–vector form:

$$\boldsymbol{s} = \mathbf{W}\boldsymbol{x}, \quad \text{with } \mathbf{W} = [\mathbf{W}^c \quad \mathbf{I}_N], \text{ where } \boldsymbol{x} \text{ has a few non-zero values,} \tag{2}$$

and the observational model (1), parameterized by the unknown vector $\boldsymbol{x}$, reads

$$\boldsymbol{y} = \mathbf{H}\mathbf{W}\boldsymbol{x} + \boldsymbol{\epsilon}. \tag{3}$$

3. Estimation

Finding a solution $\widehat{\boldsymbol{s}}$ to problem (1) that satisfies the sparsity constraint (2) can be synthesized as the following problem:

$$\text{find sparse } \widehat{\boldsymbol{x}} \text{ that minimizes } d(\boldsymbol{y}, \mathbf{H}\mathbf{W}\boldsymbol{x}), \tag{4}$$

where $d(\boldsymbol{y}, \mathbf{H}\mathbf{W}\boldsymbol{x})$ is a measurement of the misfit between the data in $\boldsymbol{y}$ and their sparse approximation by $\mathbf{H}\mathbf{W}\boldsymbol{x}$.

3.1. ℓ^1-norm penalization: basis pursuit denoising

Many approaches have been proposed for addressing problem (4), which is not trivially solvable (see, *e.g.,* [13]). In this paper, we consider the widespread ℓ^1-norm penalization formulation

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} J(\boldsymbol{x}), \quad J(\boldsymbol{x}) = d(\boldsymbol{y}, \mathbf{H}\mathbf{W}\boldsymbol{x}) + R(\boldsymbol{x}), \text{ with } R(\boldsymbol{x}) = \alpha \sum\nolimits_m |x_m|.$$

This principle is often referred to as basis pursuit denoising (BPDN), following the work of Chen et al. [6]. With such a formulation, sparse estimation is defined as the optimization of a closed-form and convex functional, for which a strong theoretical background has been developed. In addition, many efficient optimization methods have been proposed in this context, that all converge to the global minimum of J.

Particularities affect our problem, however. According to the noise statistics introduced in Section 2.1, an appropriate misfit term is the Mahalanobis distance

$$d(\boldsymbol{y}, \mathbf{H}\mathbf{W}\boldsymbol{x}) = \|\boldsymbol{y} - \mathbf{H}\mathbf{W}\boldsymbol{x}\|^2_{\boldsymbol{\Sigma}} = (\boldsymbol{y} - \mathbf{H}\mathbf{W}\boldsymbol{x})\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \mathbf{H}\mathbf{W}\boldsymbol{x}),$$

where $\boldsymbol{\Sigma}$ is the noise covariance matrix – in our case, a diagonal matrix composed of variances $\sigma^2_{n,\, n=1,\ldots N}$. This corresponds to the opposite of the log-likelihood of the data under model (3) for the Gaussian $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ noise statistics.

Because the norm in the data misfit is weighted, $J(\boldsymbol{x})$ does not have the standard form of BPDN approaches, but can be written equivalently:

$$J(\boldsymbol{x}) = \left\| \boldsymbol{\Sigma}^{-1/2}\boldsymbol{y} - \boldsymbol{\Sigma}^{-1/2}\mathbf{H}\mathbf{W}\boldsymbol{x} \right\|^2 + R(\boldsymbol{x}),$$

so an equivalent model reads:

$$\boldsymbol{z} = \mathbf{D}\boldsymbol{x} + \boldsymbol{v}, \quad \text{with } \begin{cases} \text{“whitened” data } \boldsymbol{z} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{y}, \text{ i.e., } z_n = y_n/\sigma_n \\ \text{equivalent dictionary } \mathbf{D} = \boldsymbol{\Sigma}^{-1/2}\mathbf{H}\mathbf{W} \\ \boldsymbol{v} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\epsilon} \text{ is now } \mathcal{N}(0, \mathbf{I}_N). \end{cases} \tag{5}$$

Note that the columns of $\mathbf{D}$ do not all have the same norm, even if each column of $\mathbf{W}$ has unit norm. This is a crucial point that affects estimation quality: it was shown in [4] that using such a $\mathbf{D}$ leads to higher false alarm rates for coefficients u_m corresponding to atoms with *lower* norms. This problem can be solved by considering a regularization function of the form $R(\boldsymbol{x}) = \alpha \sum_m \|\boldsymbol{d}_m\| \, |x_m|$ where $\boldsymbol{d}_m$ is the mth column of $\mathbf{D}$ or, equivalently, by normalizing the dictionary and using the standard ℓ^1-norm

$$\widehat{\boldsymbol{u}} = \arg\min_{\boldsymbol{u}} J(\boldsymbol{u}) = \frac{1}{2}\left\| \boldsymbol{z} - \overline{\mathbf{D}}\boldsymbol{u} \right\|^2 + \alpha \|\boldsymbol{u}\|_1 \tag{6}$$

where each column of $\mathbf{D}$ has been normalized to form $\overline{\mathbf{D}}$: $\overline{\mathbf{D}} = \mathbf{D}\mathbf{N}_D{}^{-1}$, with $\mathbf{N}_D$ the diagonal matrix with elements $\|\boldsymbol{d}_m\|$, and $\boldsymbol{u} = \mathbf{N}_D\boldsymbol{x}$. Note that even if $\mathbf{W}$ is the concatenation of two orthonormal bases, the corresponding sub-dictionaries in $\overline{\mathbf{D}}$ are no longer orthogonal. Indeed, one has $\mathbf{D}^T\mathbf{D} = \mathbf{W}^T\mathbf{H}^T\boldsymbol{\Sigma}^{-1}\mathbf{H}\mathbf{W}$, where $\mathbf{H}^T\boldsymbol{\Sigma}^{-1}\mathbf{H} \not\propto \mathbf{I}$.

Optimization of (6) is performed with an iterative coordinate descent algorithm [14], which is a convergent and efficient strategy for tackling sparse optimization problems. The comparison of optimization methods is beyond the scope of this paper.

3.2. Hyperparameter selection

Estimate $\widehat{\boldsymbol{u}}$ in (6) crucially depends on the value of parameter α. Indeed, for $\alpha > \max_m |\overline{\boldsymbol{d}}_m^T\boldsymbol{z}|$, the solution is identically zero. The number of non-zero components in $\widehat{\boldsymbol{u}}$ generally increases as α decreases, so for too low α, $\widehat{\boldsymbol{u}}$ may not be sparse and may contain information due to noise. In this paper, we use the following characterization, where α can be viewed as a detection threshold. An equivalent formulation of (6) based on Karush–Kuhn–Tucker (KKT) optimality conditions for $\widehat{\boldsymbol{u}}$ reads [1]

$$\begin{cases} \text{for } m \text{ such that } \hat{u}_m \neq 0 : \overline{\boldsymbol{d}}_m^T(\boldsymbol{z} - \overline{\mathbf{D}}\widehat{\boldsymbol{u}}) = \alpha \operatorname{sign}(\hat{u}_m) \\ \text{for } m \text{ such that } \hat{u}_m = 0 : |\overline{\boldsymbol{d}}_m^T(\boldsymbol{z} - \overline{\mathbf{D}}\widehat{\boldsymbol{u}})| \leq \alpha. \end{cases}$$

Hence, α can be viewed as a threshold under which coefficients are set to zero. Suppose that the data contain just noise, that is, $H_0 : \boldsymbol{z} = \boldsymbol{v}$. If $\widehat{\boldsymbol{u}} = \mathbf{0}$; then $\boldsymbol{z} - \overline{\mathbf{D}}\widehat{\boldsymbol{u}} = \boldsymbol{v} \sim \mathcal{N}(0, \mathbf{I}_N)$ and $\overline{\boldsymbol{d}}_m^T(\boldsymbol{z} - \overline{\mathbf{D}}\widehat{\boldsymbol{u}})$ is also Gaussian with unit variance since $\|\overline{\boldsymbol{d}}_m\| = 1$. Therefore, one can choose typically $\alpha \simeq 3$ or 4, depending on the desired false alarm rate for each component $\widehat{u}_m$, which is written as $p_{\mathrm{FA}} = p(\widehat{u}_m \neq 0 | H_0) = 1 - \operatorname{erf}(\alpha/\sqrt{2})$, with erf the Gaussian error function.

3.3. Amplitude debiasing and final estimation

The use of ℓ^1-norm penalization in (6) is known to introduce bias on the values of amplitudes [10]. Hence, once the non-zero components in $\widehat{\boldsymbol{u}}$ have been selected, we perform a posterior least-squares amplitude re-estimation step. Let NZ index the non-zero component and $\overline{\mathbf{D}}_{\mathrm{NZ}}$ the matrix formed by the corresponding columns of $\overline{\mathbf{D}}$. Amplitudes are estimated by

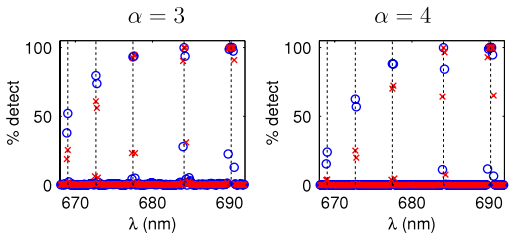
$$\widehat{\boldsymbol{u}}_{\mathrm{NZ}} = \left(\overline{\mathbf{D}}_{\mathrm{NZ}}^T\overline{\mathbf{D}}_{\mathrm{NZ}}\right)^{-1}\overline{\mathbf{D}}_{\mathrm{NZ}}^T\boldsymbol{z}$$

where the matrix inversion is properly defined if $\overline{\mathbf{D}}_{\mathrm{NZ}}$ is full rank (which is the case here if the number of non-zero values in $\widehat{\boldsymbol{x}}$ is lower than N).

Estimates of the continuous and line spectra are finally obtained by reconstructing respectively $\widehat{\boldsymbol{s}}^c = \mathbf{W}^c\mathbf{N}_D^{c\,-1}\widehat{\boldsymbol{u}}^c$ and $\widehat{\boldsymbol{s}}^\ell = \mathbf{N}_D^{\ell\,-1}\widehat{\boldsymbol{u}}^\ell$, where $\widehat{\boldsymbol{u}}^c$ and $\mathbf{N}_D^c$ (resp. $\widehat{\boldsymbol{u}}^\ell$ and $\mathbf{N}_D^\ell$) are composed of the first N (resp. the last N) components of $\widehat{\boldsymbol{u}}$ and $\mathbf{N}_D$.

3.4. Line detection statistics with Monte Carlo simulations

We present MC simulations, showing the improvements yielded by the introduction of the LSF into the observational model. Indeed, the generic sparsity-based denoising method with models similar to those in [8] uses the model $\boldsymbol{y} = \boldsymbol{s} + \boldsymbol{\epsilon}$ instead of (1). We consider the spectrum for which experimental results will be presented in Section 5 (see the plots in the left panels of Fig. 6), and we focus on the spectral range [665, 695 nm], where the data show a series of spectral lines with different amplitudes, providing an ideal example for detection performance comparison. 1000 simulations of noisy data were generated with the corresponding spectrally variable noise variance. Then, estimation was performed by minimizing criterion (6), where $\overline{\mathbf{D}}$ is built with and without the LSF operator $\mathbf{H}$. Fig. 2 plots the average number of line detections in both cases, for $\alpha = 3$ and $\alpha = 4$. As α increases, the detection rates decrease. One can clearly see that the model with LSF yields better detection rates. The



# line	1	2	3	4	5
$\alpha = 3$, with LSF	41	72	99	100	100
$\alpha = 3$, without LSF	18	40	86	100	100
$\alpha = 4$, with LSF	10	34	86	100	100
$\alpha = 4$, without LSF	3	8	41	92	100

Fig. 2. Detection statistics of five spectral lines corresponding to models with LSF (blue circles) and without LSF (red crosses). Vertical dashed lines indicate the true lines. Right table: detection scores (in %) after merging adjacent detections for each line. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

wider line at 690 nm is modeled in both cases by several adjacent impulsions. Note that, depending on the noise simulation, lines may be estimated at slightly shifted wavelengths. This is visible in Fig. 2, where detection scores are spread on adjacent indexes (especially for the "cruder" model without LSF). Hence for a fair comparison, we include numerical values in Fig. 2, showing the detection scores obtained by grouping the detections obtained at adjacent wavelengths for each line. If the brightest line is always detected using both models, the detection rates are improved on the fainter lines with the convolution model.

4. Toward object detection: segmentation and model estimation

We consider in this section the use of the former decomposition for (i) object segmentation, and (ii) estimation of the corresponding spectrum and amplitude map.

4.1. Pixel aggregation

We consider a MUSE data cube as a collection of spectra, that have been estimated according to the scheme detailed in Section 3. We propose to group the connected pixels[3] that have *similar* spectra as belonging to the same object. A binary similarity is used: two spectra are declared similar if they share at least one detected spectral line at the same wavelength. More precisely, with the notation of Section 3.3, let $\widehat{\boldsymbol{u}} = [\widehat{\boldsymbol{u}}^c, \widehat{\boldsymbol{u}}^\ell]$ and $\widehat{\boldsymbol{v}} = [\widehat{\boldsymbol{v}}^c, \widehat{\boldsymbol{v}}^\ell]$ be the sparse approximations of two connected pixels. Pixels are grouped together if for some $n \in \{1 \dots N\}, \widehat{u}_n^\ell \neq 0$ and $\widehat{v}_n^\ell \neq 0$. This is a basic spectral distance, but it allows one to propagate the statistics on false detections that were obtained from the sparse decompositions of single spectra, as we show now.

Let p_{FA} be the false alarm probability at wavelength λ_n, that is, the probability that only noise generates a detected line in a spectrum at λ_n (from Section 3.2, one has $p_{\text{FA}} = 1 - \text{erf}(\alpha/\sqrt{2})$). Suppose that the spectra of two adjacent pixels are only due to noise. Then, the probability that both spectra contain a false detected line at λ_n is p_{FA}^2, and the probability that they do not share any detected line is $(1 - p_{\text{FA}}^2)^N$. Hence, the probability that $k = 2$ pixels are aggregated by error is $p_{\text{FA}}^{\text{aggreg.}}(2) = 1 - (1 - p_{\text{FA}}^2)^N \simeq N p_{\text{FA}}^2$ for small p_{FA}. As the number k of aggregated pixels increases, $p_{\text{FA}}^{\text{aggreg.}}(k)$ naturally decreases. Corresponding values of $p_{\text{FA}}^{\text{aggreg.}}(k)$ could be approximated, for a given p_{FA}, by using MC simulations. In practice, detected structures with less than three pixels are discarded as false alarms.

Pixel aggregation is performed in the following way. Let us assume that all spectra have been decomposed according to (6).

- First, for each n, the image I_{λ_n} composed of all $\widehat{u}_n^\ell$ is formed: it is non-zero valued at pixels where a line has been detected at wavelength λ_n. Then, the connected components (we use an 8-connected neighborhood) corresponding to non-zero pixels in each I_{λ_n} are extracted. This provides a first

[3] Here, a pixel is defined as the spatial coordinates which index a spectrum in the cube.

segmentation map (that is, a list of pixel groups) with, however, many overlapping, redundant, structures, because pixels in such groups may share several lines.
- Thus, post-processing is performed, where two components are merged if they share the most part of their respective spatial extents. More precisely, for each component k, center coordinates (say, C_k) and the equivalent diameter (say, d_k) are computed. Then, components k and ℓ are merged if the distance between C_k and C_ℓ is less than $d_k + d_\ell$. Components are traveled in decreasing size order, and the procedure is repeated until no merging is performed.

4.2. Estimation of a separable model for an object

Suppose now that K pixels with spectra $\boldsymbol{y}_1 \dots \boldsymbol{y}_K$ have been identified as belonging to the same object. We assume that such data correspond to a *separable* model:

$$\boldsymbol{y}_k = a_k \mathbf{H}\boldsymbol{s} + \boldsymbol{\epsilon}_k, \quad \boldsymbol{\epsilon}_k \sim \mathcal{N}(0, \boldsymbol{\Sigma}_k), \text{ where } \boldsymbol{s} = \mathbf{W}\boldsymbol{x} \text{ with } \boldsymbol{x} \text{ sparse, } a_k \geq 0 \text{ and } \sum_{k=1}^{K} a_k = 1. \quad (7)$$

Indeed, for point sources, spatial convolution with the point spread function of the instrument spreads the spectrum on several pixels; hence such pixels have proportional spectra. For spatially extended galaxies, such a model assumes that there is no significant spatial variability in the source. This is a stronger assumption than in the first case, which appears to be roughly satisfied, especially for small, distant, objects. Note that, according to the noise description in Section 2.1, $\boldsymbol{\Sigma}_k$ may vary between the K pixels. Object angular diameters range from 0.5 arcsec, that is, approximately three pixels in each spatial dimension for point sources (where spreading is due to spatial convolution) to 4 arcsec, that is, twenty pixels.

We propose a two-step estimation procedure. First, $\boldsymbol{a}$ and $\boldsymbol{s}$ are jointly estimated by a least-squares method:

$$(\widehat{\boldsymbol{a}}_{\mathrm{LS}}, \widehat{\boldsymbol{s}}_{\mathrm{LS}}) = \arg\min_{\boldsymbol{a},\boldsymbol{s}} J_{\mathrm{LS}}(\boldsymbol{a}, \boldsymbol{s}) \quad \text{such that } a_k \geq 0 \text{ and } \sum\nolimits_k a_k = 1,$$
$$\text{where } J_{\mathrm{LS}}(\boldsymbol{a}, \boldsymbol{s}) = \sum\nolimits_k \|\boldsymbol{y}_k - a_k \boldsymbol{s}\|^2_{\boldsymbol{\Sigma}_k}, \quad (8)$$

where alternating minimizations of J_{LS} in $\boldsymbol{a}$ and $\boldsymbol{s}$ are performed until convergence is reached. Minimization in $\boldsymbol{s}$ is a simple least-squares problem that reads, after simple calculations,

$$\widehat{\boldsymbol{s}}_{\mathrm{LS}} = \boldsymbol{\Gamma} \sum\nolimits_k a_k \boldsymbol{\Sigma}_k^{-1} \boldsymbol{y}_k, \quad \text{with } \boldsymbol{\Gamma}^{-1} = \sum\nolimits_k a_k^2 \boldsymbol{\Sigma}_k^{-1}. \quad (9)$$

Minimization in $\boldsymbol{a}$ under the constraints $a_k \geq 0$ and $\sum_k a_k = 1$ is a convex linearly constrained quadratic program, that can be solved by well-known fast algorithms (for example Matlab's `lsqlin` function). Note that J_{LS} is not convex in $(\boldsymbol{a}, \boldsymbol{s})$, so optimization may converge to a local minimum in $(\boldsymbol{a}, \boldsymbol{s})$. However, in our simulations, we noticed that this problem is quite well conditioned in terms of amplitudes: for example, a simple average along the spectral axis already gives fairly acceptable estimates of amplitudes, that can be used as a good starting point.

In a second step, a sparsity-based restoration procedure similar to that described in Section 3 is applied to $\widehat{\boldsymbol{s}}_{\mathrm{LS}}$. Let us first remark that, on combining (7) and (9), $\widehat{\boldsymbol{s}}_{\mathrm{LS}}$ reads

$$\widehat{\boldsymbol{s}}_{\mathrm{LS}} = \mathbf{H}\mathbf{W}\boldsymbol{x} + \boldsymbol{\xi}, \quad \text{with noise } \boldsymbol{\xi} = \boldsymbol{\Gamma} \sum\nolimits_k a_k \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\epsilon}_k.$$

Assuming independence between noise samples $\boldsymbol{\epsilon}_k$ and $\boldsymbol{\epsilon}_\ell$ for $k \neq \ell$, one can show that the covariance matrix of the noise term $\boldsymbol{\xi}$ is $\boldsymbol{\Gamma}$; hence the correct data misfit to consider reads $\|\widehat{\boldsymbol{s}}_{\mathrm{LS}} - \mathbf{H}\mathbf{W}\boldsymbol{x}\|^2_{\boldsymbol{\Gamma}}$. Let $\mathbf{D}_{\boldsymbol{\Gamma}} = \boldsymbol{\Gamma}^{-1/2}\mathbf{H}\mathbf{W}$ and $\overline{\mathbf{D}}_{\boldsymbol{\Gamma}}$ be obtained by normalizing the columns of $\mathbf{D}_{\boldsymbol{\Gamma}}$ to unity. Like in the procedure detailed in Section 3.1, the proper way to perform the sparse decomposition of $\widehat{\boldsymbol{s}}_{\mathrm{LS}}$ consists in solving

$$\widehat{\boldsymbol{u}} = \min_{\boldsymbol{u}} \frac{1}{2} \left\| \boldsymbol{\Gamma}^{-1/2} \widehat{\boldsymbol{s}}_{\mathrm{LS}} - \overline{\mathbf{D}}_{\boldsymbol{\Gamma}} \boldsymbol{u} \right\|^2 + \beta \sum\nolimits_m |u_m|, \quad (10)$$

where β has the same interpretation as α in (6). Like in the procedure of Section 3.3, one can then proceed to amplitude debiasing and further restoration of continuous and line spectra.

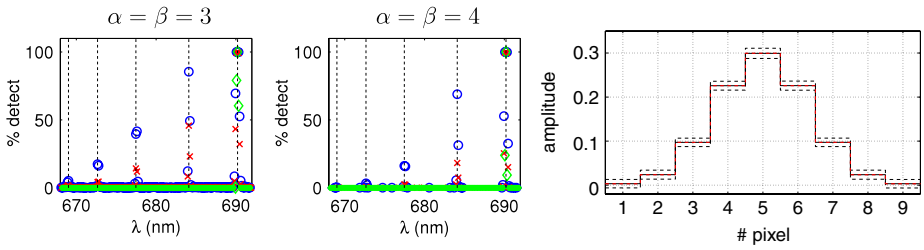


Fig. 3. Left and center: detection statistics for five spectral lines for a simulated separable model. Detection rates obtained with model (7) (blue circles), by considering only the brightest pixel (red crosses) and by averaging the spectra (green diamonds). Right: true and estimated amplitude maps (dotted red and solid black lines, respectively). Dotted black lines represent ±5 standard error levels. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.3. Line detection and amplitude statistics with Monte Carlo simulations

MC simulations are performed, where a separable model was built according to (7), with the spectrum already used in Section 3.4, $K = 9$ pixels and the Gaussian amplitude map shown in Fig. 3 right. 1000 simulations of noisy data were generated with the corresponding spectrally variable noise variance. For each simulation, estimation is performed according to (8) and (10). Fig. 3 shows the detection rates achieved, compared to the results obtained by performing single decompositions on the brightest spectrum and on the spectrum averaged over the K pixels (in the latter two cases, we set α in (6) equal to β in (10) for the joint model). Best results are achieved for the joint model, and almost no detection is achieved from the averaged spectrum, because averaging gives too much importance to the noisiest pixels (those with smallest amplitudes). Note that worse detection rates are achieved compared to the single-spectrum case in Section 3.4. This is an expected result, since the noise level affecting each pixel in these simulations is the same as in Section 3.4, whereas the signal is spread on the K pixels (recall that the sum of the amplitudes equals 1).

The amplitude estimation is shown in Fig. 3 right, and is very accurate, as expected: estimating amplitudes in model (8) is very well conditioned, since KN data points are used to estimate K values. Indeed, the variance of the *unconstrained* amplitude estimation corresponding to (8) reads, with the notation of Section 4.2, $\mathrm{Var}\,\widehat{a}_k = \left\| \boldsymbol{\Sigma}_k^{-1/2} \mathbf{H} \widehat{\mathbf{s}}_{\mathrm{LS}} \right\|^{-2}$, where the latter equation can be derived from usual least-squares estimation theory.[4] In our example, this yields $\forall k$, $\mathrm{std}(\widehat{a}_k) \simeq 2 \times 10^{-3}$.

5. Experimental results with MUSE-like simulations

The methods detailed in Sections 3 and 4 were applied to a MUSE-like deep field simulated data cube. Such data were generated by the MUSE consortium and rely on high-complexity astrophysical simulations. In summary, an astronomical scene is simulated as a distribution of galaxies with different computed spatial and spectral profiles, where the number density of objects with respect to the redshift is chosen in accordance with the Hubble Space Telescope's Ultra Deep Field counts. Then, data enter another simulation code corresponding to the MUSE instrument numerical model [11]. Generating such data cubes is an extremely burdensome operation, and only one simulated deep field scene has been available up to now. This data cube has 300×300 pixels and $N = 3463$ wavelengths, and for each element of the cube the noise variance is estimated as explained in Section 2.1.

The sparse decomposition procedure of Section 3 was applied to the 90 000 spectra. Parameter α in criterion (6) was set deliberately to a low value, in order to favor the detection of small lines for the pixel aggregation procedure, at the cost of false detections. We set $\alpha = 3$, for which $p_{\mathrm{FA}} \simeq 2.7 \times 10^{-3}$; then on average one has $Np_{\mathrm{FA}} \simeq 9$ false line detections for each spectrum. Then, the pixel aggregation

[4] The variance of the *constrained* problem should be slightly less than such a value.

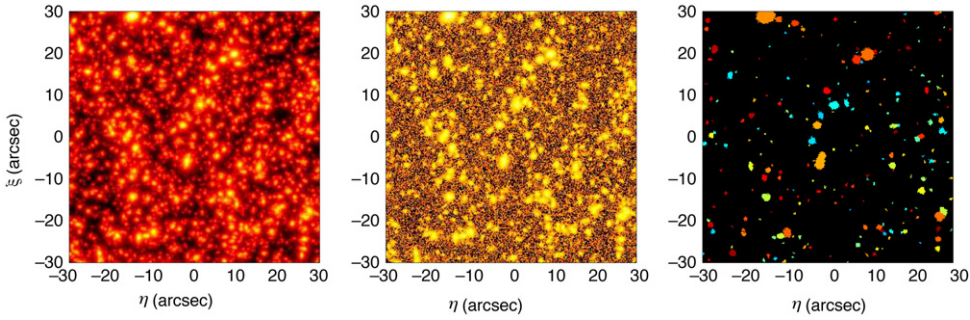


Fig. 4. Left and center: images obtained by averaging the noise-free and noisy data cubes, respectively, along the spectral axis. Right: detected objects after sparse estimation and pixel aggregation. The color mapping of objects is arbitrary.

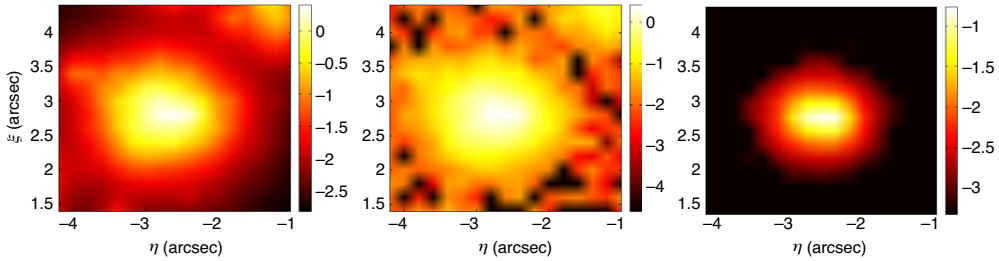


Fig. 5. Amplitude estimation for the detected object at ($\eta = -2.5$ arcsec, $\xi = 3$ arcsec) in Fig. 4. Left and center: spectral averaging of noise-free and noisy data, respectively. Right: the estimated amplitude map. Amplitudes are on a log scale. Note that the map in the right panel is, by definition, normalized to unit sum, whereas data in the left and center panels are not, so the amplitudes cannot be directly compared (see the text).

process of Section 4.1 was performed. One has then $p_{\mathrm{FA}} \simeq 2.7 \times 10^{-3}$ and $p_{\mathrm{FA}}^{\mathrm{aggreg.}} \simeq 2.5 \times 10^{-2}$. Fig. 4 shows the map of detected objects, jointly with the image obtained by averaging the noise-free and noisy data cubes along the spectral axis. For noisy data, spectral averaging is performed by adequately weighting the data: for spectrum $\boldsymbol{y} = [y_1, \ldots, y_N]^T$, where y_n is affected by Gaussian noise with variance σ_n^2, the average is computed using

$$\bar{y} = \left(\sum_{n=1}^{N} \frac{1}{\sigma_n^2} \right)^{-1} \sum_{n=1}^{N} \frac{y_n}{\sigma_n^2}. \tag{11}$$

Eq. (11) statistically corresponds to the maximum-likelihood estimate of a *mean value* μ under the model $y_n = \mu + \epsilon_n$, where $\epsilon_n \sim \mathcal{N}(0, \sigma_n^2)$. This is obviously not the case here, since the spectra are not constant. It was observed, however, that such weighting significantly improves the resulting image.

Let us remark that the pixel aggregation was based only on the detection of common spectral lines in coefficients $\widehat{\boldsymbol{u}}^{\ell}$, so the (numerous) objects with a strong continuous spectrum and without lines in the data cube cannot be detected by such a scheme. Conversely, it is adapted to the detection of fainter objects, whose spectrum is mainly composed of one (or a few) line(s), and which do not appear significantly in the average image—look, for example, at the very top right corner of the scene in Fig. 4.

Once such object segmentation was performed, the estimation of separable models described in Section 4.2 was applied, with $\beta = 4$ in Eq. (10). Results are shown for a detected object with center coordinates around ($\eta = -2.5$ arcsec, $\xi = 3$ arcsec) in Fig. 4. $K = 76$ pixels are associated with this object by the aggregation method. Fig. 5, right, shows the estimated amplitudes associated with each pixel of the object, whose shape corresponds quite well to the amplitudes in the noise-free image obtained by averaging all spectral bands (left panel). Averaging the noisy data along the spectral

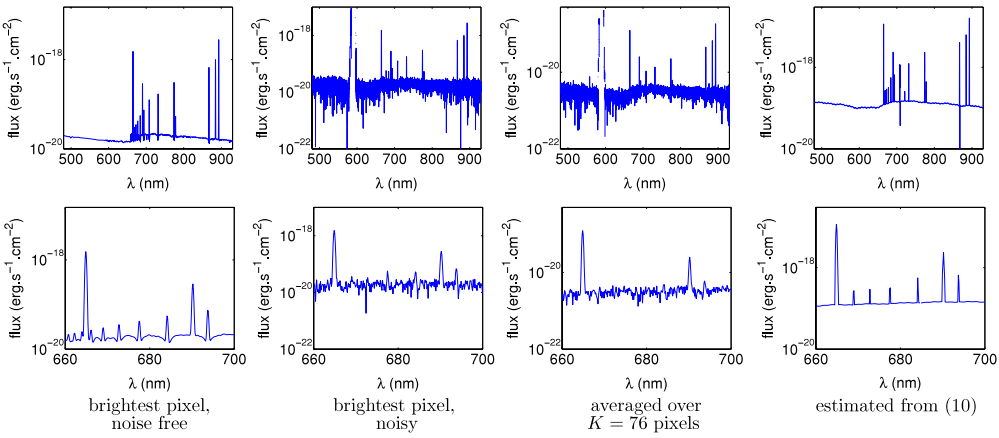


Fig. 6. Different spectra associated with the object defined by the 76 flagged pixels in Fig. 5. Top: whole spectral range. Bottom: zoom on a series of emission lines between 660 and 700 nm.

axis[5] also provides a satisfactory image, because such a spectrum has a sufficiently high continuum, which brings in energy at all wavelengths—this will not be the case for more distant, and therefore interesting, objects. Note that the average image in Fig. 5 (left) is used here as ground truth; however, since it is obtained from the data cube, it suffers from the contamination of other objects. In contrast, in the estimated amplitude map, pixels are zero outside the selected area. In other words, Fig. 5 (right) is a *segmented* image, whereas Fig. 5 (left) and (center) are raw images extracted from the cube. Results on estimated spectra are shown in Fig. 6, and reveal the interest of taking into account the joint observation model (7). Estimation results have to be compared with the noise-free (but convolved) spectrum at the center of the object, plotted in the left panels. In particular, the zooms in the bottom panels show that six faint emission lines are retrieved—and even deconvolved. In contrast, only the two or three brightest lines may be detected from the noisy spectrum at the brightest pixel, plotted in parts (b). Note also that averaging the noisy spectra of all 76 flagged pixels yields still poorer results. The continuous part of the spectrum is also quite well estimated and denoised.

6. Conclusion

Methodological developments were presented for processing MUSE-like hyperspectral data. A first part concerned the sparse representation of spectra in a union of two structurally different bases, which can be viewed as a decomposition of spectra into a continuous component and a line spectrum. Sparsity was addressed in the classical ℓ^1-penalization framework. Observational specificities were taken into account, which also required modifications in the penalization term, and a false alarm probability was associated with the value of the corresponding ℓ^1-norm weight. A second part of our work considered object detection and characterization. The former sparse decompositions were exploited to aggregate connected pixels for which spectra are similar, where a binary similarity was defined, that propagates the formerly obtained false alarm characterizations. Last, a method was proposed for the joint estimation of the amplitudes and the spectrum of a separable model in spatially segmented sub-cubes. Results for a realistic MUSE-like simulated data cube were presented with both object segmentation and further joint estimation, giving satisfactory results. In particular, joint estimation in segmented regions was shown to significantly improve the quality of the spectral estimation.

Possibilities for work on improving such results are numerous. First, the proposed decomposition of spectra in the union of the DCT basis and the canonical basis is probably far from optimal for

[5] Averaging is also performed by weighting all data by their appropriate variances; see Eq. (11).

the spectra of galaxies, and should be enriched with more refined dictionaries. Then, the approach for pixel grouping is mainly heuristic, and is based on a binary measurement of similarity between spectral lines. More sophisticated segmentation methods should be investigated, based *e.g.* on spectral similarity measurements used in hyperspectral imaging (spectral angle, spectral correlation [3,5]), which should also allow one to detect objects on the basis of their continuous spectra. This is a hard task, however, because of the high dimensionality of the problem considered (in particular, the high number of objects that are searched for).

From a practical point of view, assessing the performance of the method in terms of object detection (completeness, spatial and spectral characteristics, biases) is a priority for further work. Automatically associating detected objects with ground truth is nothing but a trivial operation, however, because the object density in the simulation is high, generating cross-talks between close objects. Last, the noise variances used in these simulations are estimated variances in an ideal setting, in particular where optimal background subtraction was supposed. Analyzing the impact of the variance estimation error on all the processing steps developed is also of major importance.

References

[1] S. Alliney, S. Ruzinsky, An algorithm for the minimization of mixed ℓ^1 and ℓ^2 norms with application to Bayesian estimation, IEEE Trans. Signal Process. 42 (1994) 618–627.
[2] R. Bacon, et al. Probing unexplored territories with MUSE: a second generation instrument for the VLT, in: Proceedings of the SPIE, 2006, p. 62690J.
[3] S. Bourguignon, Spectral similarity measurements, Technical Report, Observatoire de la Côte d'Azur, Available on demand, 2009.
[4] S. Bourguignon, D. Mary, E. Slezak, Sparsity-based denoising of hyperspectral astrophysical data with colored noise: application to the MUSE instrument, in: Proc. IEEE Whispers, 2010, pp. 1–4. doi:10.1109/WHISPERS.2010.55949021-4.
[5] C.I. Chang, Hyperspectral Imaging: Techniques for Spectral Detection and Classification, Kluwer Academic, Plenum Publishers, New York, NY, 2003.
[6] S.S. Chen, D.L. Donoho, M.A. Saunders, Atomic decomposition by basis pursuit, SIAM J. Sci. Comput. 20 (1998) 33–61.
[7] P. Ciuciu, J. Idier, J.F. Giovannelli, Regularized estimation of mixed spectra using a circular Gibbs–Markov model, IEEE Trans. Signal Process. 49 (2001) 2201–2213.
[8] D. Donoho, X. Huo, Uncertainty principles and ideal atomic decomposition, IEEE Trans. Inform. Theory 47 (2001) 2845–2862.
[9] D.L. Donoho, I.M. Johnstone, Ideal spatial adaptation by wavelet shrinkage, Biometrika 81 (1994) 425–455.
[10] J.J. Fuchs, On sparse representations in arbitrary redundant bases, IEEE Trans. Inform. Theory 50 (2004) 1341–1344.
[11] A. Jarno, R. Bacon, P. Ferruit, A. Pécontal-Rousset, Numerical simulation of the VLT/MUSE instrument, in: G.Z. Angeli, M.J. Cullum (Eds.), Proceedings of the SPIE, 2008, p. 701710.
[12] S. Mallat, A Wavelet Tour of Signal Processing: The Sparse Way, Academic Press Inc., 2008.
[13] J.L. Starck, F. Murtagh, J. Fadili, Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity, Cambridge University Press, 2010.
[14] P. Tseng, Convergence of a block coordinate descent method for nondifferentiable minimization, J. Optim. Theory Appl. 109 (2001) 475–494.