SPARSITY-BASED DENOISING OF HYPERSPECTRAL ASTROPHYSICAL DATA WITH COLORED NOISE: APPLICATION TO THE MUSE INSTRUMENT

Sébastien Bourguignon, David Mary and Éric Slezak

University of Nice Sophia Antipolis, CNRS, Observatoire de la Côte d'Azur
B.P. 4229, 06304 Nice Cedex 4, France

ABSTRACT

This paper proposes a denoising method for hyperspectral astrophysical data, adapted to the specificities of the MUSE (Multi-Unit Spectroscopic Explorer) instrument, which will provide massive integral field spectroscopic observations of the far universe, characterized by very low signal-to-noise ratio and strongly non identically distributed noise.

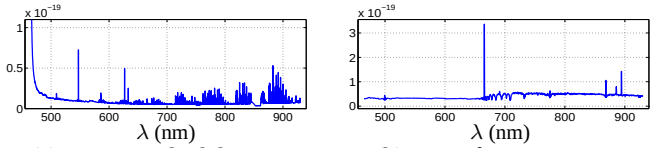
Data are considered as a collection of spectra. The proposed restoration procedure operates on each spectrum by minimizing a penalized data-fit criterion, which takes into account the noise spectral distribution, with additional constraints expressing prior sparsity information in a union of bases. Spectra are modeled as the sum of line and continuous spectra, which are supposed to be sparse in the canonical and the Discrete Cosine Transform bases, respectively. Dealing with colored noise requires specific methodological approaches regarding not only the estimator definition itself, but also hyperparameter tuning and optimization issues. These three points are successively investigated. Promising denoising results are obtained on realistic simulations of astrophysical observations.

***Index Terms*—** Denoising, overcomplete sparse representations, colored noise, astrophysical spectra.

1. INTRODUCTION

The MUSE instrument (Multi-Unit Spectroscopic Explorer) is an extremely powerful integral field spectrograph [1], which will be installed at ESO (European Southern Observatory) on the Very Large Telescope in Chile in 2012. It will provide hyperspectral data cubes with 300×300 spatial elements and up to 4000 spectral channels covering the essential part of the visible spectrum. In its wide field mode configuration, covering a field of view of 1arcmin^2, the expected performances of MUSE for a long time exposure (typically a few tens of hours) should allow the detection of much fainter galaxies than those observed by today's ground-based spectrographs.

A strong characteristic of MUSE data holds in a very low signal-to-noise ratio (SNR), where noise is *colored*, that is, the noise power distribution is highly variable with respect to wavelength. Figure 1(a) plots the expected noise standard deviation for such data as a function of wavelength. The powerful parasite emission of the atmosphere at specific wavelengths causes high spectral variations in the noise level, while the blue part of the spectrum is characterized by a very low SNR, due to lower instrumental efficiency at such wavelengths. Moreover, because of pre-processing steps such as previous subtraction of the background emission, the available spectra can even show negative values. Figures 1(b) and (c) show, respectively, the noise-free spectrum of a relatively bright light source and the corresponding simulated noise-contaminated spectrum as observed through the MUSE instrument. The noisy version shows a high number of spurious peaks caused by high noise variance at certain wavelengths, with amplitudes sometimes higher than true spectral lines. Hence, taking into account noise specificities is a very crucial point to achieve the ambitious scientific goals – in particular, low SNR source detections – of the MUSE instrument.



(a) Noise standard deviation (b) Noise-free spectrum $\boldsymbol{o}$

(c) Noisy spectrum $\boldsymbol{y}$ (d) Restored spectrum $\boldsymbol{u}$

Fig. 1. Noise level as a function of wavelength, and noise-free, noisy and denoised spectra of a galaxy. Intensities represent light fluxes, in $\text{erg.s}^{-1}\text{.cm}^{-2}\text{.pixel}^{-1}$.

In this paper, we propose a denoising scheme designed for such data, based on the hypothesis that data can be approximated by sparse representations in redundant transformation spaces [2, 3]. In Section 2, we define the denoised data as the solution of a quadratic data-fit criterion taking into account the noise distribution, and penalized by sparsity-inducing ℓ^1-norm constraints. Sparsity is expressed spectrally, where spectra are considered as the superposition of emission and absorption line spectra plus a continuum, which is supposed to have a sparse representation in the Discrete Cosine Transform (DCT) domain. Hyperparameter selection is addressed in Section 3 for colored noise, where such parameters are viewed as approximate detection thresholds, which can then be tuned automatically. An efficient algorithmic strategy based on Iterative- and Block-Coordinate Descent methods [4, 5] is proposed in Section 4. Finally, denoising results are presented in Section 5 on realistic deep-field simulated data provided by the MUSE consortium.

2. SPARSE DECOMPOSITIONS OF SPECTRA

Sparsity-based denoising relies on the hypothesis that most of the signal energy can be concentrated into a reduced number of coefficients with high values in some appropriate transform space,

This work was partially supported by ANR project 08-BLAN-0253-01 DAHLIA - Dedicated Algorithms for HyperspectraL Imaging in Astronomy.

whereas the noise distribution, expressed in such space, is concentrated around zero. Then, such high-valued coefficients are less affected by noise than signal samples are in their original space. Consequently, identifying such coefficients allows one to preserve most part of the useful information and to build a *denoised* version of the data [2]. In this paper, we consider the restoration of MUSE spectra at each spatial location by adding prior sparsity information. In a first approximation, the considered astrophysical spectra can be decomposed as a certain number of emission or absorption lines, superimposed on a continuum. Following previous works on sparse representations (*e.g.*, [3]), we consider that each spectrum has a sparse representation in the union of the canonical basis for spectral lines and of the DCT basis for the continuum.

2.1. Sparse estimation with ℓ^1-norm constraints

Let $\boldsymbol{y} = [y_1 \dots y_N]^T$ represent an observed spectrum discretized at wavelengths $\lambda_1, \dots, \lambda_N$. We want to reconstruct spectrum $\boldsymbol{o} = [o_1 \dots o_N]^T$ from:

$$\boldsymbol{y} = \boldsymbol{o} + \boldsymbol{\epsilon}, \tag{1}$$

where $\boldsymbol{\epsilon}$ collects noise samples that are supposed centered Gaussian with known covariance matrix $\boldsymbol{\Sigma} \triangleq \text{diag}(\sigma_1^2 \dots \sigma_N^2)$, where σ_n^2 is the noise variance at frequency λ_n shown in Figure 1. We model $\boldsymbol{o}$ as the sum of a *line* spectrum $\boldsymbol{o}^\ell$ and a *continuous* spectrum $\boldsymbol{o}^c$, where:

- $\boldsymbol{o}^\ell$ is sparse in the standard basis ($\boldsymbol{o}$ has only few emission and absorption lines),
- $\boldsymbol{o}^c$ has a sparse DCT representation ($\boldsymbol{x}^c = \text{DCT}(\boldsymbol{o}^c)$ has only few non-zero components).

Both $\boldsymbol{o}^\ell$ and $\boldsymbol{o}^c$ have to be estimated. Note that using such two bases with complementary properties is a key point for efficiently modelling the data [3]: a line spectrum is sparse in the canonical basis but not in the DCT space. Conversely, describing a continuous spectrum requires all coefficients of the canonical basis to be non-zero, whereas it can be approximated by few DCT coefficients. Hence, both bases are necessary to efficiently represent the considered astrophysical spectra with a small number of coefficients.

A well-known approach for enforcing sparsity properties consists in minimizing a data-fit criterion, penalized by ℓ_1-norm constraints [2, 3, 6]. We consider the quadratic data-fit:

$$\frac{1}{2}\|\boldsymbol{y} - \boldsymbol{o}^\ell - \boldsymbol{o}^c\|^2_{\boldsymbol{\Sigma}} = \frac{1}{2}(\boldsymbol{y} - \boldsymbol{o}^\ell - \boldsymbol{o}^c)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{o}^\ell - \boldsymbol{o}^c),$$

which corresponds to the neg-log-likelihood of data $\boldsymbol{y}$ given model (1) [7], and then statistically takes into account the specific noise structure shown in Figure 1. The sparsity constraint on the line spectrum can be written as the regularization term:

$$R^\ell(\boldsymbol{o}^\ell) = \alpha^\ell\|\boldsymbol{o}^\ell\|_1 = \alpha^\ell \sum_n |o_n^\ell|,\ \alpha^\ell > 0.$$

The sparsity constraint on the continuous spectrum can be written as the regularization term:

$$R^c(\boldsymbol{x}^c) = \alpha^c\|\boldsymbol{x}^c\|_1,\ \alpha^c > 0$$

with $\boldsymbol{x}^c = \text{DCT}(\boldsymbol{o}^c)$. We note $\mathbf{W}^c$ the orthonormal inverse DCT matrix, so that $\boldsymbol{o}^c = \mathbf{W}^c\boldsymbol{x}^c$. Let $\mathbf{W} = [\mathbf{I}_N\ \mathbf{W}^c]$ be the $N \times 2N$ matrix formed by the concatenation of Identity matrix $\mathbf{I}_N$ and $\mathbf{W}^c$, and $\boldsymbol{x} = [\boldsymbol{x}^{\ell^T}\ \boldsymbol{x}^{c^T}]^T$ with $\boldsymbol{x}^\ell = \boldsymbol{o}^\ell$. Our sparsity-based spectral denoising method lies in the following optimization problem:

$$\begin{aligned} \widehat{\boldsymbol{x}} &= [\widehat{\boldsymbol{x}}^{\ell^T}\ \widehat{\boldsymbol{x}}^{c^T}]^T = \arg\min_{\boldsymbol{x}^\ell, \boldsymbol{x}^c} J(\boldsymbol{x}^\ell, \boldsymbol{x}^c), \text{ with} \\ J(\boldsymbol{x}^\ell, \boldsymbol{x}^c) &= \frac{1}{2}\|\boldsymbol{y} - \mathbf{W}\boldsymbol{x}\|^2_{\boldsymbol{\Sigma}} + R^\ell(\boldsymbol{x}^\ell) + R^c(\boldsymbol{x}^c). \end{aligned} \tag{2}$$

For adequately chosen values of α^ℓ and α^c, the minimization of J yields sparse estimates [3, 6]. Note that J can be written as a more classical sparse representation problem, under the form:

$$J(\boldsymbol{x}^\ell, \boldsymbol{x}^c) = \frac{1}{2}\|\boldsymbol{z} - \mathbf{M}\boldsymbol{x}\|^2_{\mathbf{I}_N} + R^\ell(\boldsymbol{x}^\ell) + R^c(\boldsymbol{x}^c), \tag{3}$$

with $\boldsymbol{\Sigma}^{-1/2} = \text{diag}(\sigma_1^{-1} \dots \sigma_N^{-1})$, $\boldsymbol{z} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{y}$ and $\mathbf{M} = \boldsymbol{\Sigma}^{-1/2}\mathbf{W}$. In the vocabulary of sparse representations [3, 6], the minimization of (3) aims at approximating weighted data $\boldsymbol{z}$ by the shortest linear combination of *atoms* taken from *dictionary* $\mathbf{M} = [\boldsymbol{\Sigma}^{-1/2}\ \mathbf{M}^c]$ with $\mathbf{M}^c = \boldsymbol{\Sigma}^{-1/2}\mathbf{W}^c$. In our case, however, the equivalent dictionary $\mathbf{M}$ is not composed of orthogonal blocks since $\mathbf{M}^c$ is not orthogonal, nor are its columns normalized to 1. Consequently, important differences hold between the problem considered here and classical sparse representations in unions of orthogonal bases, as for example in [3, 4, 6] that consider $\boldsymbol{\Sigma} = \mathbf{I}_N$. In particular, consequences on hyperparameter tuning and optimization will be discussed in Sections 3 and 4, respectively.

2.2. Amplitude bias correction and spectrum restoration

The ℓ^1-norm penalization efficiently induces sparsity by locating few non-zero components in $\widehat{\boldsymbol{x}}$ [3, 6]. However, it systematically introduces bias in amplitudes [6], and posterior amplitude re-estimation is necessary to obtain satisfactory estimates for line and continuous spectra. This can be done simply by least-squares. Let NZ index the non-zero components in $\widehat{\boldsymbol{x}}$, let $\mathbf{W}_{\text{NZ}}$ be the matrix formed by columns of $\mathbf{W}$ with indexes NZ and let $\boldsymbol{u}_{\text{NZ}}$ collect the non-zero amplitudes to be retrieved. Minimization of $\|\boldsymbol{y} - \mathbf{W}_{\text{NZ}}\boldsymbol{u}_{\text{NZ}}\|^2_{\boldsymbol{\Sigma}}$ yields:

$$\widehat{\boldsymbol{u}}_{\text{NZ}} = \left(\mathbf{W}_{\text{NZ}}^T\boldsymbol{\Sigma}^{-1}\mathbf{W}_{\text{NZ}}\right)^{-1}\mathbf{W}_{\text{NZ}}^T\boldsymbol{\Sigma}^{-1}\ \boldsymbol{y}.$$

Note that Card NZ $\ll N$ if $\widehat{\boldsymbol{x}}$ is sparse, hence in that case the latter equation is well-defined. Let $\widehat{\boldsymbol{u}} = [\widehat{\boldsymbol{u}}^{\ell^T}\ \widehat{\boldsymbol{u}}^{c^T}]^T$ denote the $2N \times 1$ vector obtained by filling $\boldsymbol{u}_{\text{NZ}}$ with zeros outside NZ. Then, line and continuous estimated spectra are obtained, respectively, by $\widehat{\boldsymbol{o}}^\ell = \widehat{\boldsymbol{u}}^\ell$ and $\widehat{\boldsymbol{o}}^c = \mathbf{W}^c\widehat{\boldsymbol{u}}^c$, and the restored spectrum is $\widehat{\boldsymbol{o}} = \widehat{\boldsymbol{o}}^\ell + \widehat{\boldsymbol{o}}^c$.

3. HYPERPARAMETER TUNING FOR COLORED NOISE

Obviously, estimate $\widehat{\boldsymbol{x}}$ defined by (2) highly depends on the values of parameters α^ℓ and α^c, that control the trade-offs between data fidelity and sparsity constraints. It can be shown that for $\alpha^\ell, \alpha^c > \max|\mathbf{W}^T\boldsymbol{y}|$, $\widehat{\boldsymbol{x}}$ is identically zero. Conversely, for too small α^ℓ and α^c, the solution may not be sparse at all. We base the selection of α^ℓ and α^c on the following result, that extends a result in [8] to the case of colored noise and several hyperparameters:

Property 1 *$\widehat{\boldsymbol{x}} = [\widehat{\boldsymbol{x}}^{\ell^T}, \widehat{\boldsymbol{x}}^{c^T}]^T$ minimizes (2) if and only if:*

i) *for n such that $\hat{x}_n^\ell = 0$:* $|e_n| < \alpha^\ell$
ii) *for n such that $\hat{x}_n^\ell \neq 0$:* $e_n = \alpha^\ell \text{sign}(x_n^\ell)$
iii) *for n such that $\hat{x}_n^c = 0$:* $|\boldsymbol{w}_n^{c^T}\boldsymbol{e}| < \alpha^c$
iv) *for n such that $\hat{x}_n^c \neq 0$:* $\boldsymbol{w}_n^{c^T}\boldsymbol{e} = \alpha^c \text{sign}(x_n^c)$,

where $\boldsymbol{e} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \mathbf{W}\widehat{\boldsymbol{x}})$ is the $N \times 1$ vector containing the estimation residuals, weighted by the noise variances σ_n^2.

The proof follows the same scheme as the proof in [8], although we consider here a more general problem than the one in [8].

Properties 1 i) and ii) stipulate that $\hat{x}_n^\ell$ is zero if and only if $|e_n| < \alpha^\ell$: α^ℓ can be viewed as a threshold on $|e_n|$ under which no line is detected at wavelength λ_n. Suppose that estimate $\widehat{\boldsymbol{x}}$ is such that $\mathbf{W}\widehat{\boldsymbol{x}}$ efficiently approximates data $\boldsymbol{o}$. Ideally, one is left with $\boldsymbol{y} - \mathbf{W}\widehat{\boldsymbol{x}} = \boldsymbol{\epsilon}$, and $e_n = \sigma_n^{-2}\epsilon_n$ is Gaussian with zero mean and variance $\sigma_e^2(n) = \sigma_n^{-2}$. To perform a detection test with a given false alarm rate, one would then choose a threshold on $|e_n|$ at, say, $k\sigma_e(n)$, that yields a false alarm rate of $\tau_{\text{FA}} = 1 - \text{erf}(k/\sqrt{2})$, where erf is the Gaussian error function. For example, $k = 3$ yields a false alarm rate of approximately 0.27%. *That is, α^ℓ should depend on σ_n, with relation $\alpha_n^\ell = k/\sigma_n$.* In other words, penalization term $\alpha^\ell \|\boldsymbol{x}^\ell\|_1$ should be replaced by:

$$R^\ell(\boldsymbol{x}^\ell) = \sum_n \alpha_n^\ell |x_n^\ell|, \text{ with } \alpha_n^\ell = k/\sigma_n.$$

Note, however, that estimation residual $|\boldsymbol{y} - \mathbf{W}\widehat{\boldsymbol{x}}|$ may take higher values than $|\boldsymbol{\epsilon}|$, in particular because of amplitude estimation bias in $\widehat{\boldsymbol{x}}$ (see § 2.2). Then, choosing $\alpha_n^\ell = k/\sigma_n$ may yield a false detection rate slightly higher than τ_{FA}.

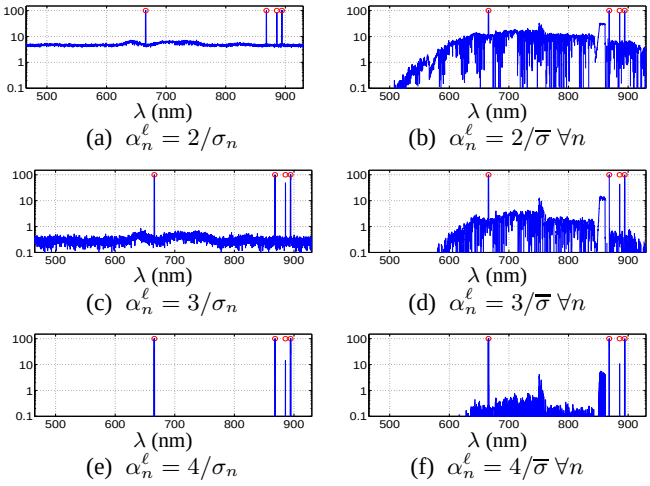
Similarly, it comes from Properties 1 iii) and iv) that α^c is a threshold on $|\mathbf{W}^{c\,T}\boldsymbol{e}|$, under which DCT coefficients are set to zero. Variances associated to coefficients $\{\boldsymbol{w}_n^{c\,T}\boldsymbol{e}\}_n$ are $\gamma_n^{-2} = \boldsymbol{w}_n^{c\,T}\boldsymbol{\Sigma}^{-1}\boldsymbol{w}_n^c$, which depend on n. Hence, a proper way to penalize DCT coefficients consists in replacing penalization term $\alpha^c\|\boldsymbol{x}^c\|_1$ by:

$$R^c(\boldsymbol{x}^c) = \sum_n \alpha_n^c |x_n^c|, \text{ with } \alpha_n^c = k/\gamma_n,$$
$$\text{where } \gamma_n = (\boldsymbol{w}_n^{c\,T}\boldsymbol{\Sigma}^{-1}\boldsymbol{w}_n^c)^{-1/2}.$$

From the Bayesian point of view [7], such tuning corresponds to setting Laplace prior distributions on each x_n with different scale parameters, which may appear unusual. In classical regularization problems with Laplace priors, however, the scale factor is often selected with regard to the noise level (*e.g.*, [2]). Hence, for colored noise, adapting hyperparameters to each noise level is a quite natural way to proceed. The importance of correctly tuning parameter α^ℓ is illustrated in Figure 2. Denoising is applied on 5000 noisy realizations of a synthetic spectrum with a continuum and the four main spectral lines of the spectrum in Figure 1(b). The detection rate (that is, the average number of non-zero occurrences in $\widehat{\boldsymbol{x}}^\ell$) is given at each wavelength for different tunings of α^ℓ. In this example, $\alpha_n^c = 4/\gamma_n$. As desirable, an almost uniform false detection rate is achieved by choosing $\alpha_n^\ell = k/\sigma_n$ (left column), with more false detections (and more true detections) for lower k. On the contrary, choosing α^ℓ uniformly yields worse detection statistics (right column): note in particular more false detections around 860 nm, in the spectral band where the noise variance is lower (see Figure 1(a)).

4. OPTIMIZATION ALGORITHM

We propose an optimization strategy derived from the Block Coordinate Relaxation method (BCR) in [4]. The original BCR was proposed for minimizing criteria such as (3) with ℓ^1-norm constraints in union of orthogonal bases. It iteratively performs blockwise optimization steps with respect to each set of orthogonal components, where each step has an analytical explicit solution. In our case, as discussed at the end of § 2.1, orthogonality in the DCT domain is lost. Hence, we propose an algorithm that alternates BCR steps on $\boldsymbol{x}^\ell$ and Iterative Coordinate Descent (ICD) steps, that is, scalar optimizations, on each x_n^c. ICD was shown to be computationally efficient to retrieve sparse solutions [9], although other algorithmic



(a) $\alpha_n^\ell = 2/\sigma_n$ (b) $\alpha_n^\ell = 2/\overline{\sigma}\ \forall n$
(c) $\alpha_n^\ell = 3/\sigma_n$ (d) $\alpha_n^\ell = 3/\overline{\sigma}\ \forall n$
(e) $\alpha_n^\ell = 4/\sigma_n$ (f) $\alpha_n^\ell = 4/\overline{\sigma}\ \forall n$

Fig. 2. Detection rates (%) for different tunings of α^ℓ. Parameter $\overline{\sigma}$ is chosen as $\overline{\sigma}^{-2} = \frac{1}{N}\sum_n \sigma_n^{-2}$. Circles locate the true lines.

options are possible among the many algorithms that have been recently proposed for the minimization of such functionals as (2).

One can show that minimization of J in $\boldsymbol{x}^\ell$ yields N independent scalar problems with soft-thresholding solutions:

$$\tilde{\boldsymbol{x}}^\ell = \arg\min_{\boldsymbol{x}^\ell} J(\boldsymbol{x}^\ell, \boldsymbol{x}^c) \Leftrightarrow \forall n, x_n^\ell = \sigma_n^2 \phi^{\text{st}}_{\alpha_n^\ell}(r_n^c/\sigma_n^2), \quad (4)$$

where r_n^c is th n^{th} component of $\boldsymbol{r}^c = \boldsymbol{y} - \mathbf{W}^c\boldsymbol{x}^c$ and ϕ_τ^{st} is the soft-thresholding function:

$$\begin{cases} \phi_\tau^{\text{st}}(u) &= 0 \text{ if } |u| \leq \tau \\ \phi_\tau^{\text{st}}(u) &= u - \text{sign}(u)\tau \text{ if } |u| > \tau. \end{cases}$$

On the contrary, optimization with respect to $\boldsymbol{x}^c$ cannot be performed jointly: let $\tilde{\boldsymbol{x}}^c = \arg\min_{\boldsymbol{x}^c} J(\boldsymbol{x}^\ell, \boldsymbol{x}^c)$, then:

$$\tilde{\boldsymbol{x}}^c = \arg\min_{\boldsymbol{x}^c} \frac{1}{2}\|\boldsymbol{r}^\ell - \mathbf{M}^c\boldsymbol{x}^c\|^2_{\mathbf{I}_N} + \sum_n \alpha_n^c |x_n^c|, \quad (5)$$

with $\boldsymbol{r}^\ell = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{y} - \boldsymbol{x}^\ell)$ and $\mathbf{M}^c = \boldsymbol{\Sigma}^{-1/2}\mathbf{W}^c$. Matrix $\mathbf{M}^c$ is not orthogonal, so (5) does not have an analytic solution. However, scalar optimizations in x_n^c still do. Let $\widetilde{x}_n^c = \arg\min_{x_n^c} J(\boldsymbol{x}^\ell, \boldsymbol{x}^c)$. One can show that:

$$\widetilde{x}_n^c = \frac{1}{\boldsymbol{m}_n^T\boldsymbol{m}_n}\phi^{\text{st}}_{\alpha_n^c}\big(\boldsymbol{m}_n^T(\boldsymbol{r}^\ell - \sum_{k\neq n} x_k^c\boldsymbol{m}_k)\big). \quad (6)$$

Hence, we replace minimization (5) by N iterative minimizations in x_n^c. Finally, optimization works by performing alternately one step (4) on $\boldsymbol{x}^\ell$ and N steps (6) on each x_n^c. Algorithmic efficiency is improved by mostly (say 90% of iterations) updating only the non-zero components of the current iterate. Since a sparse solution is searched, the algorithm is initialized by the zero vector. Convergence proofs of such a strategy towards the minimum of J can be found in [5]. Property 1 gives an explicit characterization of any minimizer of $\widehat{\boldsymbol{x}}$, which we use as a convergence test. In practice, convergence is usually reached after less than 10 outer iterations.

Criterion J is a convex functional in $\boldsymbol{x}$, although not strictly. Hence, the question of unicity of the minimizer $\widehat{\boldsymbol{x}}$ is not trivial. Usual criteria such as *mutual incoherence*, positive *Exact Recovery Coefficient* or *Restricted Isometry Property*) are not directly transposable

to our setting with weighted data-fit criterion and non-uniform hyperparameters. One can also show that the *Unique Representation Property* (that is, any N columns of the dictionary are linearly independent), which also characterizes unicity, is not satisfied here. The study of the unicity of $\widehat{\boldsymbol{x}}$ and of the equivalence between ℓ^1 and ℓ^0 problems in such a framework is out of the scope of this paper, which is limited to practical considerations: the minimizers of J form a convex set [6], possibly not reduced to one element, but all minimizers are equally acceptable solutions and our optimization strategy is ensured to converge towards one of them.

5. APPLICATION TO SIMULATED DATA

Denoising is applied on each spectrum of a $51 \times 51 \times 3578$ datacube simulated by the MUSE consortium, containing 25 astrophysical objects. Figure 3 shows both noise-free and noisy images at wavelength $\lambda = 600$ nm, jointly with the restored image, and their corresponding SNR. We define the SNR between noise-free data $O(\boldsymbol{r}, \lambda)$ and noisy data $Y(\boldsymbol{r}, \lambda)$, indexed by pixels $\boldsymbol{r}$ and wavelengths λ, as:

$$\mathrm{SNR}_{\mathrm{dB}} = 10 \log_{10} \frac{\sum_{\boldsymbol{r},\lambda} O(\boldsymbol{r},\lambda)^2}{\sum_{\boldsymbol{r},\lambda} \left(O(\boldsymbol{r},\lambda) - Y(\boldsymbol{r},\lambda)\right)^2}.$$

Although no spatial correlation has been accounted for, the corresponding image is improved, showing that all spectra have been satisfactorily denoised. Note, however, a strong artifact in the bottom left of the restored image caused by the false detection of a spectral line. The denoised spectrum obtained from the noisy spectrum in Figure 1(b) is shown in Figure 1(d), with $\alpha_n^\ell = 4/\sigma_n$ and $\alpha_n^c = 4/\gamma_n$ (see Section 3). The four main spectral lines are retrieved and the continuous part of the spectrum is globally restored with, in this example, 31 non-zero DCT coefficients. The denoised spectrum has 20.5 dB SNR, to be compared to 2.8 dB for the noisy spectrum. Table 1 gives the SNR of noisy and restored data cubes, for several values of parameters α^ℓ and α^c, showing approximately 30 dB improvement for well-chosen hyperparameters. Setting $\alpha_n^c = 4/\gamma_n$ yields the best results, while SNR results are less sensitive to α_n^ℓ above $4/\sigma_n$: as α_n^ℓ increases, non-detections of true lines are more frequent, but do not affect so much the SNR, since the error on the (few) lines in the spectrum is averaged on the whole wavelength axis.

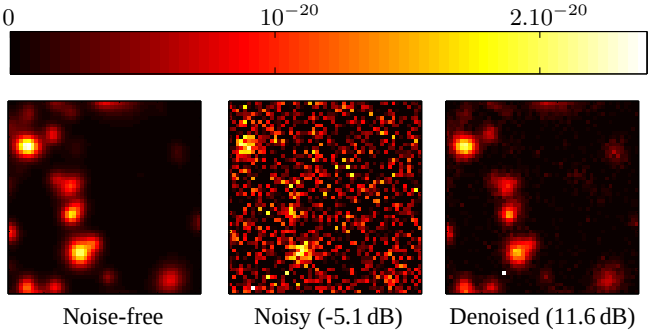


Fig. 3. Noise-free, noisy and denoised images at $\lambda = 600$ nm. The three images use the same color map.

6. CONCLUSION

We have proposed a denoising method for astrophysical hyperspectral data, based on a sparse representation of spectra in the union of two bases, the canonical one and the DCT basis. The strong variability of the noise level with respect to wavelength was taken into account by weighting a data-fit measure by corresponding variances. Noise variability was also shown to affect hyperparameter tuning, for which specific selection rules were derived for automatic tuning. An efficient and convergent optimization strategy was used in this context, which alternates block-wise and coefficient-wise optimization steps. Results on MUSE-like simulated data showed very promising results in terms of data denoising and line detection.

Direct extensions of this work regard the enrichment of the proposed model with other transforms than DCT that are adapted to astrophysical spectral shapes, in order to detect specific sources in still lower SNR. MUSE deep-field images are mostly composed of galaxies which may exhibit structured spatial features, so relevant information in images should also concentrate in a few coefficients in appropriate transformation spaces. Hence, further work also regards the coupling of spectral and spatial constraints in a whole data cube.

	$\alpha^c = 3/\gamma_n$	$\alpha^c = 4/\gamma_n$	$\alpha^c = 5/\gamma_n$
$\alpha_n^\ell = 3/\sigma_n$	4.2	4.8	4.8
$\alpha_n^\ell = 4/\sigma_n$	9.8	13.1	12.7
$\alpha_n^\ell = 5/\sigma_n$	10	13.5	13.1
$\alpha_n^\ell = 6/\sigma_n$	9.9	13.3	12.6

Table 1. $\mathrm{SNR}_{\mathrm{dB}}$ after denoising, for several values of hyperparameters. The SNR on the noisy data cube is -16.9 dB.

7. REFERENCES

[1] R. Bacon and *et al.*, "Probing unexplored territories with MUSE: a second generation instrument for the VLT," in *Proceedings of the SPIE*, July 2006, vol. 6269 of *Ground-based and Airborne Instrumentation for Astronomy*, p. 62690J.

[2] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[3] D.L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inform. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov. 2001.

[4] S. Sardy, A. G. Bruce, and P. Tseng, "Block coordinate relaxation methods for nonparametric wavelet denoising," *J. Comput. Graph. Statist.*, vol. 9, pp. 361–379, 2000.

[5] P. Tseng, "Convergence of a block coordinate descent method for nondifferentiable minimization," *J. Optim. Theory Appl.*, vol. 109, pp. 475–494, 2001.

[6] J.-J. Fuchs, "On sparse representations in arbitrary redundant bases," *IEEE Trans. Inform. Theory*, vol. 50, no. 6, pp. 1341–1344, June 2004.

[7] J. Idier, Ed., *Bayesian Approach to Inverse Problems*, ISTE Ltd and John Wiley & Sons Inc, Apr. 2008.

[8] S. Alliney and S.A. Ruzinsky, "An algorithm for the minimization of mixed ℓ^1 and ℓ^2 norms with application to bayesian estimation," *IEEE Trans. Signal Process.*, vol. 42, no. 3, pp. 618–627, Mar. 1994.

[9] S. Bourguignon, H. Carfantan, and J. Idier, "A sparsity-based method for the estimation of spectral lines from irregularly sampled data," *IEEE J. Selected Topics Sig. Proc.*, vol. 1, no. 4, pp. 575–585, Dec. 2007.