CONSTRAINED LIKELIHOOD RATIOS FOR DETECTING SPARSE SIGNALS IN HIGHLY NOISY 3D DATA

S. Paris, R.F.R. Suleiman, D. Mary, A. Ferrari

Laboratoire Lagrange, UMR7293,
Université de Nice Sophia-Antipolis, CNRS, Observatoire de la Côte d'Azur,
Campus Valrose, 06108 Nice - cedex 02, FRANCE
Email: {Silvia.Paris, Raja.Fazliza, David.Mary, Andre.Ferrari}@unice.fr

ABSTRACT

We propose a method aimed at detecting weak, sparse signals in highly noisy three-dimensional (3D) data. 3D data sets usually combine two spatial directions x and y (e.g. image or video frame dimensions) with an additional direction λ (e.g. temporal, spectral or energy dimension). Such data most often suffer from information leakage caused by the acquisition system's point spread functions, which may be different and variable in the three dimensions. The proposed test is based on dedicated 3D dictionaries, and exploits both the sparsity of the data along the λ direction and the information spread in the three dimensions. Numerical results are shown in the context of astrophysical hyperspectral data, for which the proposed 3D model substantially improves over 1D detection approaches.

***Index Terms*—** Detection, sparse, GLR, hyperspectral, dictionary learning.

1. INTRODUCTION AND PREVIOUS WORKS

An increasing number of modern applications exploits multidimensional data. For instance, hyperspectral data have been successfully used in different domains such as bio-medicine [1], astrophysics [2] and defense [3], thanks to their rich informative content (an entire spectral signature is available for each pixel of the image).

Multidimensional data sets often possess however intrinsic characteristics that lead to processing difficulties. First, the large data size: for example, hyperspectral images made up of hundreds spectral bands require increasing storage capacities and fast processing algorithms. Second, the poor signal-to-noise ratio (SNR): splitting the information over three dimensions and increasing the sampling resolution often leads to relatively stronger noise levels in the voxel elements. In addition, the features of interest may be intrinsically weak: examples are found in the detection of covert communications or in compressive sensing [4, 5]; another example is provided by the astrophysical data considered in Sec.4. In such cases, it may be extremely difficult to *detect* the presence of information, because it is both rare and weak. This is the situation considered in this paper.

Depending on the detection target, the information of interest may take the form of a natural or a manmade signal. Most often, signals can be compactly modeled as sparse using appropriate dictionaries of elementary features called atoms. In sparse synthesis models, the signal information is approximated as a linear combination of few atoms, with a vector of coefficients having few non-zero entries [6–8]. We will follow a sparse approach here.

Sparse estimation and sparse detection techniques share a common goal: finding low dimension subspaces in which the useful information is located. The literature in sparse detection is however far less abundant than that in sparse estimation. The next lines propose a survey of earlier works in the field.

Recently, the works [4, 5, 9–15] have proposed detection tests exploiting sparsity, mostly in the compressive sensing framework. In [9, 10], a best k-term approximation of the parameter vector is injected in the Likelihood Ratio. The same idea appears in [11], where a Maximum A Posteriori (MAP) estimate is used as a solution of the Basis Pursuit DeNoising (BPDN, or Lasso [16–18]). In [12], the sensing matrix is designed to correspond to the subspace of the sparse signals, which is assumed to be known. The works presented in [15] optimize Parseval frames for signal detection using the Generalized Likelihood Ratio (GLR) test [19]. In [14], sparse recovery is investigated for multiple testing. Following the seminal works [18], Fuchs describes in [13] the connection between the support detection performed by the BPDN and the GLR. In [4], the test called Higher Criticism (HC) is set in order to detect faint sparse vector means of multivariate normal vectors, and the *Max* test, based on a Bonferroni-type correction, is considered in [5]. Both tests are shown to be powerful in the case of very sparse and weak signals.

In [20–22], we introduced two detection tests based on MAP estimates, which were shown to be more powerful for sparse signals than classical methods, such as the unconstrained GLR test [19]. A connection was found with earlier works of Fan [23] and Basu [24], and with the *Max* test described in [5].

The sparse detection problem we consider here is that of detecting, in highly noisy 3D data (data-cubes), a few spatially localized and weak spectral features, whose energy furthermore spreads in the three dimensions, say x, y and λ. Here, *spatial* (resp. *spectral*) will refer to the x, y (resp. λ) dimensions, so that the proposed method can be generically applied regardless of the physical nature of x, y and λ. We will accordingly separate the 3D point spread function (PSF) into a spatial (SSF) and a wavelength spread function (WSF).

The tests of [20–22] were applied to 3D (hyperspectral) data and relied on a dedicated one-dimensional (1D) redundant spectral dictionary. As an illustration of the proposed method, we show here that one can improve on the results of [20–22] by using spatio-spectral dictionaries. These dictionaries exploit both the sparsity of the data along the λ dimension and the information leakage caused by the 3D PSF over several voxels. We consider that the spectral pattern to be detected is not known, but that we dispose of a set of training signals that represents well the variety of features to be detected. The

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

The authors thank R.Bacon for providing MUSE data, and A. Schutz for help in displaying the 3D atoms.

spectral dictionary used in the proposed test is made of atoms that are learned using dictionary learning techniques [25,26], and further spread by the SSF and WSF, leading to 3D atoms. Note that we do not attempt to first deconvolve the data, and then detect. The reason is that in such a two-step approach, the dependence of deconvolution on a regularisation term would mess-up the data distribution. In addition, regularized deconvolution somewhat performs a detection, but without rigorously defining the realized detection test (nor controlling its size and power). In contrast, our approach is one-shot: the sparsity-based test essentially performs information re-concentration to improve detection (quite in the spirit of matched filtering principles), while keeping the statistical description of the data accurate.

To our knowledge, few related detection approaches exist. Among these, [2] proposes a detection/reconstruction method for low flux sources using a wavelet-based spatio-spectral dictionary, along with a multi-scale variance stabilization transform. In contrast to our approach, the dictionary of [2] is generic, and does not account for the instrumental spreading functions. We also note the use of (one-dimensional) specific spectral dictionaries in [27, 28]. The first paper considers the case of Poisson noise with a regularization term favouring group sparsity, and the second paper a dictionary of Gaussian lines in conjunction with a Bayesian version of the support-vector machine-learning technique [29].

In the following, we first introduce the considered detection approach for a 1D model in Sec.2. The 3D detection test is presented in Sec.3. Numerical results on astrophysical hyperspectral data are shown in Sec.4.

2. 1D SPECTRAL MODEL

We start by assuming a one-dimensional model where the columns of the data-cube along the λ dimension are considered independently:

$$\mathcal{H}_0 : \boldsymbol{s} = \boldsymbol{\epsilon}, \qquad \mathcal{H}_1 : \boldsymbol{s} = \boldsymbol{H}\boldsymbol{R}\boldsymbol{\alpha} + \boldsymbol{\epsilon}, \tag{1}$$

where $\boldsymbol{s} \in \mathbb{R}^{\Lambda}$ is a spectral column at spatial coordinates (x, y), $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ is the corresponding Gaussian noise vector, with known (or estimated) diagonal covariance matrix $\boldsymbol{\Sigma} = \mathrm{diag}[\sigma_1^2 \dots \sigma_\Lambda^2]$, $\boldsymbol{H}$ $(\Lambda \times \Lambda)$ denotes the composition matrix by the WSF (it is a convolution if the WSF is spectrally invariant), and $\boldsymbol{R} = [\boldsymbol{r}_1 \dots \boldsymbol{r}_L]$ is a $(\Lambda \times L)$ redundant dictionary with $L \geq \Lambda$. The question of the design of $\boldsymbol{R}$ will be addressed in Sec.3.2. $\boldsymbol{\alpha} \in \mathbb{R}^L$ represents a vector of unknown parameters. This model can be rewritten as

$$\mathcal{H}_1 : \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{s} = \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{H}\boldsymbol{R}\boldsymbol{\alpha} + \boldsymbol{w}, \tag{2}$$

where $\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{H}\boldsymbol{R} = \boldsymbol{D}_{\boldsymbol{\Sigma}\boldsymbol{H}}$ appears as an equivalent dictionary, and $\boldsymbol{w} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$. Noting $\boldsymbol{z} = \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{s}$, $\boldsymbol{D} = \boldsymbol{D}_{\boldsymbol{\Sigma}\boldsymbol{H}}\boldsymbol{N}_{\boldsymbol{D}_{\boldsymbol{\Sigma}\boldsymbol{H}}}^{-1}$ and $\boldsymbol{\theta} = \boldsymbol{N}_{\boldsymbol{D}_{\boldsymbol{\Sigma}\boldsymbol{H}}}\boldsymbol{\alpha}$, with $\boldsymbol{N}_{\boldsymbol{D}_{\boldsymbol{\Sigma}\boldsymbol{H}}}$ the diagonal matrix composed of the norms of the columns of $\boldsymbol{D}_{\boldsymbol{\Sigma}\boldsymbol{H}}$, we obtain

$$\mathcal{H}_1 \ : \ \boldsymbol{z} = \boldsymbol{D}\boldsymbol{\theta} + \boldsymbol{w}. \tag{3}$$

The approach above acts in the weighted (by $\boldsymbol{\Sigma}^{-\frac{1}{2}}$) data domain. This normalization simplifies the setting and interpretation of the detection tests. A test compares a statistic $T(\boldsymbol{z})$ to a threshold γ under the two hypotheses of the model ($T(\boldsymbol{z}) \gtrless_{H_0}^{H_1} \gamma$). The probability of false alarm is $\mathrm{P_{FA}} = Pr(T(\boldsymbol{z}) > \gamma \mid \mathcal{H}_0)$.

A reference test for model (3) is the GLR test [19], whose test statistic arises from the maximum of the ratio of the likelihoods under the two model hypotheses. The test involves the Maximum Likelihood (ML) estimate of $\boldsymbol{\theta}$ under $\mathcal{H}_1$:

$$\mathrm{GLR} : \ \frac{\max_{\boldsymbol{\theta}} p(\boldsymbol{z} \mid \boldsymbol{\theta})}{p(\boldsymbol{z} \mid \boldsymbol{0})} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma. \tag{4}$$

The ML estimate of $\boldsymbol{\theta}$ is computed as

$$\hat{\boldsymbol{\theta}}_{ML} = \arg\max_{\boldsymbol{\theta}} p(\boldsymbol{z} \mid \boldsymbol{\theta}) = \arg\min_{\boldsymbol{\theta}} \frac{1}{2} \|\boldsymbol{z} - \boldsymbol{D}\boldsymbol{\theta}\|_2^2 . \tag{5}$$

Since $\boldsymbol{D}$ is $(\Lambda \times L)$ with $L > \Lambda$, $\hat{\boldsymbol{\theta}}_{ML}$ achieves $\boldsymbol{D}\hat{\boldsymbol{\theta}}_{ML} = \boldsymbol{z}$, and it is easy to see that the (unconstrained) GLR test (4) reduces to

$$\mathrm{GLR} : \ T_{\mathrm{GLR}}(\boldsymbol{z}) = \|\boldsymbol{z}\|_2^2 \ \underset{H_0}{\overset{H_1}{\gtrless}} \ \xi, \tag{6}$$

with $\xi = 2\ln\gamma$. The GLR test behaves as an energy detector, without taking any advantage from the use of the redundant dictionary.

An appropriate test for sparse data is the *Max* test [5], for which we propose a simple alternative formulation as a constrained GLR test. Let us consider the following model:

$$\mathcal{H}_1 : \boldsymbol{s} = \boldsymbol{H}\boldsymbol{R}\boldsymbol{\alpha} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}), \quad \|\boldsymbol{\alpha}\|_0 = 1, \tag{7}$$

which in the weighted domain leads to

$$\mathcal{H}_1 \ : \ \boldsymbol{z} = \boldsymbol{D}\boldsymbol{\theta} + \boldsymbol{w}, \quad \boldsymbol{w} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}), \quad \|\boldsymbol{\theta}\|_0 = 1. \tag{8}$$

In this new model, the signal component of the whitened data vector $\boldsymbol{z}$ is obtained as one of the $\boldsymbol{D}$ columns, selected and weighted by the sole non-zero entry of $\boldsymbol{\theta}$. We thus have

$$\mathcal{H}_1 \ : \ \boldsymbol{z} = \boldsymbol{d}_i\theta_i + \boldsymbol{w}, \tag{9}$$

with i and θ_i unknown.

According to model (9), we define the 1-sparse 1D GLR test:

$$\mathrm{GLR}_{1\mathrm{s}}^{(1\mathrm{D})} \ : \ \frac{\max_{j,\,\theta_j} p(\boldsymbol{z} \mid \boldsymbol{d}_j, \theta_j)}{p(\boldsymbol{z} \mid \boldsymbol{0})} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_{1\mathrm{s}}. \tag{10}$$

The computation of the $T_{\mathrm{GLR}_{1\mathrm{s}}}^{(1\mathrm{D})}$ test statistic gives

$$T_{\mathrm{GLR}_{1\mathrm{s}}}^{(1\mathrm{D})}(\boldsymbol{z}) = \min_{j,\,\theta_j} -\boldsymbol{z}^T\boldsymbol{d}_j\theta_j + \frac{1}{2}\theta_j^2\boldsymbol{d}_j^T\boldsymbol{d}_j. \tag{11}$$

For $\boldsymbol{d}_j$ fixed, the ML estimate of θ_j is $\hat{\theta}_{j_{ML}} = \boldsymbol{d}_j^T\boldsymbol{z}$. Replacing this expression in (11) we have

$$T_{\mathrm{GLR}_{1\mathrm{s}}}^{(1\mathrm{D})}(\boldsymbol{z}) \ = \ \min_j -\frac{1}{2}\left(\boldsymbol{d}_j^T\boldsymbol{z}\right)^2 = \max_j \frac{1}{2}\left\|\boldsymbol{d}_j^T\boldsymbol{z}\right\|_2^2. \tag{12}$$

This is equivalent to maximise $\left|\boldsymbol{d}_j^T\boldsymbol{z}\right|$ over j, which leads to

$$\mathrm{GLR}_{1\mathrm{s}}^{(1\mathrm{D})} \ : \ T_{\mathrm{GLR}_{1\mathrm{s}}}^{(1\mathrm{D})}(\boldsymbol{z}) = \max_j \left|\boldsymbol{d}_j^T\boldsymbol{z}\right| \ \underset{H_0}{\overset{H_1}{\gtrless}} \ \xi_{1\mathrm{s}}, \tag{13}$$

with $\xi_{1\mathrm{s}} = 2\ln\gamma_{1\mathrm{s}}$. This test is equivalent to the *Max* test [5]. The GLR (4) and the $\mathrm{GLR}_{1\mathrm{s}}^{(1\mathrm{D})}$ (13) tests for a 1D model are numerically compared in Sec.4, where it is shown that these methods are of insufficient power for the detection of very faint and highly sparse signals. Hence, next section focuses on a more accurate data model that takes into account the information smear caused not only by the WSF but also by the SSF.

3. 3D SPATIO-SPECTRAL MODEL

3.1. Model

The spread of information in the three dimensions can be represented by the composition of the signal feature with both the WSF and the SSF. For the simplicity of exposition we consider that the SSF is spectrally invariant, but this is not necessary. Model (1) accounts only for the WSF, through the use of $\boldsymbol{H}$: the spectral content of each 1D atom $\mathbf{r}_i$, $i = 1, \dots, L$ of the $\boldsymbol{R}$ dictionary is spread by $\boldsymbol{H}$ in the λ direction. However, at every single wavelength λ, the SSF also dilutes spatially the information over a number of $(N_x \times N_y)$ adjacent

columns. Thus, to each spectral column $\boldsymbol{s}(x,y)$ with $x=1,\ldots,X$ and $y=1,\ldots,Y$, is associated a data sub-cube of dimension $(N_x \times N_y \times \Lambda)$ where the energy is smeared because of the SSF.

We now explicit this model in vector form. All the $(N_x \times N_y)$ Λ-dimensional spectra in each such data sub-cube are stacked one on top of the other to form a $N_x N_y \Lambda$ column vector $\boldsymbol{s}_V$. This leads to the following expression:

$$\mathcal{H}_1 : \boldsymbol{s}_V(x,y) = \boldsymbol{F}(x,y)\boldsymbol{H}\boldsymbol{R}\boldsymbol{\alpha} + \boldsymbol{\epsilon}_V, \tag{14}$$

where $\boldsymbol{s}_V$ and $\boldsymbol{\epsilon}_V \in \mathbb{R}^{N_x N_y \Lambda}$, $\boldsymbol{\epsilon}_V \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_V(x,y))$ is a Gaussian noise vector characterized by a $(N_x N_y \Lambda \times N_x N_y \Lambda)$ diagonal covariance matrix $\boldsymbol{\Sigma}_V$, in which the diagonal is the vectorized form of the associated sub-cube containing the noise variances of each voxel. $\boldsymbol{H}$, $\boldsymbol{R}$, and $\boldsymbol{\alpha}$ are the same as in model (7) and $\boldsymbol{F}(x,y)$ represents the $(N_x N_y \Lambda \times \Lambda)$ matrix composition by the SSF. If we denote by

$$\boldsymbol{f} = \begin{bmatrix} f_{11} & \cdots & f_{1N_y} \\ f_{21} & \cdots & f_{2N_y} \\ \vdots & & \vdots \\ f_{N_x 1} & \cdots & f_{N_x N_y} \end{bmatrix} \tag{15}$$

the array of the SSF coefficients at position (x,y) (that we first assume for simplicity invariant in the λ dimension), and by

$$\boldsymbol{\phi} = \text{vec}\boldsymbol{f} = [f_{11} \ldots f_{N_x 1}, \ \ldots \ , f_{1N_y} \ldots f_{N_x N_y}]^T \tag{16}$$

its vector form, then $\boldsymbol{F} = [f_{11}\boldsymbol{I}_\Lambda \ \ldots \ f_{N_x N_y}\boldsymbol{I}_\Lambda]^T = \boldsymbol{\phi} \otimes \boldsymbol{I}_\Lambda$, where $\otimes$ denotes the Kronecker product and $\boldsymbol{I}_\Lambda$ the Λ-dimensional Identity matrix. Note that when the SSF is not spectrally invariant, $\boldsymbol{F}$ should be replaced by $\boldsymbol{F} = [\text{diag}(\boldsymbol{f}_{11}), \ldots, \text{diag}(\boldsymbol{f}_{N_x N_y})]^T$, where $\boldsymbol{f}_{ij} = [f_{ij}(\lambda_1), \ldots, f_{ij}(\lambda_\Lambda)]$ with $i = 1,\ldots,N_x$, $j = 1,\ldots,N_y$.

According to (14), in the weighted data domain we obtain for each spectral column $\boldsymbol{s}(x,y)$ the following model for the associated vectorized data-cube $\boldsymbol{s}_V$:

$$\mathcal{H}_1 : \boldsymbol{\Sigma}_V^{-\frac{1}{2}}\boldsymbol{s}_V = \boldsymbol{\Sigma}_V^{-\frac{1}{2}}\boldsymbol{F}\boldsymbol{H}\boldsymbol{R}\boldsymbol{\alpha} + \boldsymbol{w}_V, \quad \|\boldsymbol{\alpha}\|_0 = 1, \tag{17}$$

where the dependence on (x,y) of $\boldsymbol{\Sigma}_V$ and $\boldsymbol{F}$ is implicit. Similarly to Sec.2, we refer to $\boldsymbol{\Sigma}_V^{-\frac{1}{2}}\boldsymbol{F}\boldsymbol{H}\boldsymbol{R} = \boldsymbol{D}_{\boldsymbol{\Sigma}\boldsymbol{F}\boldsymbol{H}}$ as an equivalent dictionary and $\boldsymbol{w}_V \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{I}_{N_x N_y \Lambda}\right)$. Noting $\boldsymbol{z}_V = \boldsymbol{\Sigma}_V^{-\frac{1}{2}}\boldsymbol{s}_V$, $\boldsymbol{D}_V = \boldsymbol{D}_{\boldsymbol{\Sigma}\boldsymbol{F}\boldsymbol{H}}\boldsymbol{N}^{-1}_{\boldsymbol{D}_{\boldsymbol{\Sigma}\boldsymbol{F}\boldsymbol{H}}}$ and $\boldsymbol{\theta} = \boldsymbol{N}_{\boldsymbol{D}_{\boldsymbol{\Sigma}\boldsymbol{F}\boldsymbol{H}}}\boldsymbol{\alpha}$, with $\boldsymbol{N}_{\boldsymbol{D}_{\boldsymbol{\Sigma}\boldsymbol{F}\boldsymbol{H}}}$ the diagonal matrix composed of the norms of the columns of $\boldsymbol{D}_{\boldsymbol{\Sigma}\boldsymbol{F}\boldsymbol{H}}$, (17) leads to the model

$$\mathcal{H}_1 \ : \ \boldsymbol{z}_V = \boldsymbol{D}_V\boldsymbol{\theta} + \boldsymbol{w}_V, \quad \|\boldsymbol{\theta}\|_0 = 1, \tag{18}$$

which corresponds to the expression of model (8), with the crucial difference that this time both the WSF and SSF are taken into account in $\boldsymbol{D}_V$.

The constrained GLR test for (18) gives

$$\text{GLR}_{1\text{s}}^{(3\text{D})} \ : \ T_{\text{GLR}_{1\text{s}}}^{(3\text{D})}(\boldsymbol{z}_V) = \max_j \left|\boldsymbol{d}_{V\,j}{}^T\boldsymbol{z}_V\right| \underset{H_0}{\overset{H_1}{\gtrless}} \xi_{1\text{s}}, \tag{19}$$

where $\xi_{1\text{s}} = 2\ln\gamma_{1\text{s}}$. For (13) and (19), the threshold $\xi_{1\text{s}}$ controlling the P_{FA} is computed numerically (see [20], Sec.5).

3.2. Design strategies for $\boldsymbol{R}$

There are several ways to choose $\boldsymbol{R}$. Of course, this dictionary can be taken as a generic one (for instance wavelets). However, if some atoms of such dictionary are known not to well represent the data to be detected, they should be removed, as they will only increase the P_{FA}. In addition, when a representative data set of the signals under $\mathcal{H}_1$ is available, it may be more efficient to design a specific dictionary. In such cases, $\boldsymbol{R}$ is often simply taken as the mean or the first singular vector obtained by a SVD of the data set [30], or it can be optimized using now classical sparse dictionary learning techniques, for instance K-SVD [25]. Another set of approaches called minimax seeks a dictionary of reduced dimension, which maximizes the worst probability of detection under $\mathcal{H}_1$ [26, 31].

Note an important issue in the case where a spectral library is available. Assume this (possibly very large) library $\boldsymbol{\mathcal{L}}$ is so accurate that the signal under $\mathcal{H}_1$ can be considered to belong to $\boldsymbol{\mathcal{L}}$. In this case, it may be not only computationally prohibitive (see Sec.4), but also suboptimal (w.r.t. detection power) to implement a test of the form (19) with $\boldsymbol{R} = \boldsymbol{\mathcal{L}}$ (*i.e* the concatenation of all possible alternatives). This is because the P_{FA} may increase wildly with the number of alternatives [26]. Hence, dictionaries with reduced dimensions such as those cited above are very useful.

As a final remark, we comment on the one-sparse hypothesis made in (7) and (17). If the signals under $\mathcal{H}_1$ are sparsely decomposed in $\boldsymbol{R}$ using a few atoms with one main salient feature, then mainly this feature will be detectable in highly noisy data. On the other hand, if $\boldsymbol{R}$ is sufficiently well designed/large and thus contains (up to an amplitude factor) an atom equal to the signal under $\mathcal{H}_1$, we can also consider that the signals are one-sparse in $\boldsymbol{R}$.

Next section compares the 1D (13) and 3D (19) one-sparse GLR tests for the K-SVD [25] and the minimax approaches of [26].

4. ILLUSTRATION ON HYPERSPECTRAL DATA

The proposed detection strategies are illustrated in the context of the astrophysical hyperspectral data of the ESO's Multi Unit Spectroscopic Explorer (MUSE) instrument [8]. MUSE will deliver datacubes composed of $X \times Y = 300 \times 300$ spectra, sampled at approximately $\Lambda = 3600$ wavelengths of the visible spectrum.

One of the major challenges of this instrument concerns the detection of very distant galaxies. Those galaxies present spectra composed by a sole, faint and narrow line, known as Lyman-alpha line (Ly-α). Fig.1 reports examples of Ly-α spectra (blue, magenta and green) and the corresponding data (grey) in the MUSE data-cube for one of them (see also Fig.4). The highly noisy character of these data is striking: the lines are totally buried in noise (the SNR of the spectra is: $\text{SNR} = 10\log_{10}\frac{\|\boldsymbol{F}\boldsymbol{H}\boldsymbol{R}\boldsymbol{\alpha}\|_2^2}{\text{Tr}\{\boldsymbol{\Sigma}\}} \approx -30\text{dB}$).

The possible shapes of the signals under $\mathcal{H}_1$ can, in this case, be accurately simulated using astrophysical models: we dispose of a library of $\mathcal{L} \approx 10^4$ line profiles (100 of which are shown on Fig.2(a)), which extend on about $N_\lambda = 100$ spectral channels. The location of the lines is however arbitrary and unknown, because it depends on the distance of the source (Doppler effect). Using the full library in the tests (13) and (19) entails performing $\mathcal{L} \times \Lambda \times X \times Y \times N_\lambda \approx 3 \times 10^{14}$ multiplications and additions. If we assume $3 \cdot 10^9$ operations/sec, this represents 6 months of computation time for the 3D test. Our approach to design $\boldsymbol{R}$ is thus as follows. We optimize three dictionaries (SVD, K-SVD with 7 atoms and minimax [26])

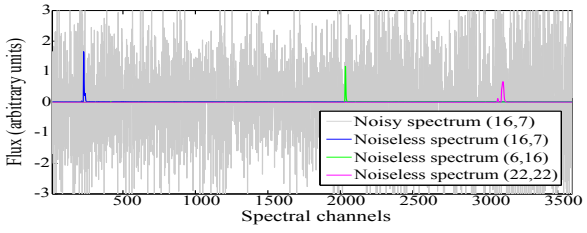
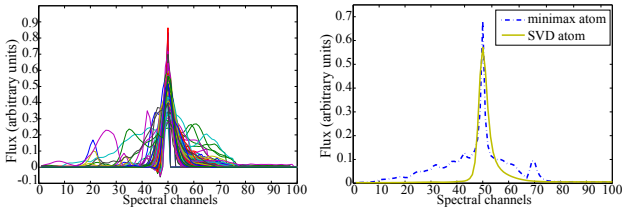


Fig. 1. Typical spectra under $\mathcal{H}_1$ and corresponding data for $(16,7)$.

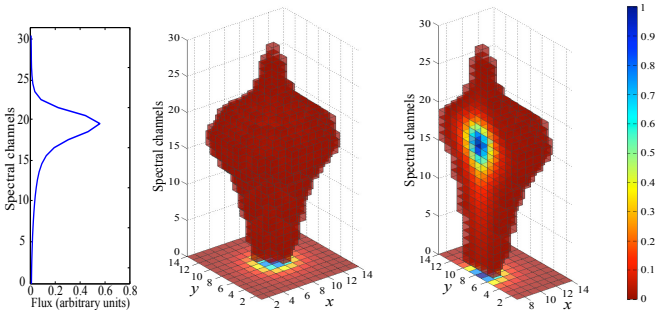
of reduced dimensions (resp. $\Lambda \times 1$, $\Lambda \times 7$ and $\Lambda \times 1$) on the whole set of lines centered at a fixed wavelength. The resulting SVD and minimax atoms are shown in Fig.2(b). Then, we generate translation invariant versions of these dictionaries by shifting the corresponding atoms at all possible wavelengths, which yields $\boldsymbol{R}_{\text{SVD}}$, $\boldsymbol{R}_{\text{KSVD}_7}$ and $\boldsymbol{R}_{\text{minimax}}$.

MUSE is characterized by a 3D PSF that can be separated into two distinct functions: a WSF, which spreads over 7 spectral elements, and a SSF, which covers (13×13) pixels in the spatial domain. Both the WSF and the SSF of the instrument slightly vary with wavelength. In this work we consider a constant approximation, obtained by computing the mean over all wavelengths. The combination of the three dictionaries above with these spread functions are used in (7)-(13) and (17)-(19). An example of atom obtained in the 3D case is shown in Fig.3.



(a) 100 spectral lines in $\boldsymbol{\mathcal{L}}$ (b) SVD and minimax atoms

Fig. 2. Examples of spectral lines and trained atoms.



(a) (b)

Fig. 3. Example of 3D atom: (a) SVD atom (b) Left: corresponding 3D atom after composition by the SSF and the WSF. Right: a cut of the 3D atom.

We compare the detection tests on a small MUSE data-cube[1] of size $X \times Y = 26\times28$ with $\Lambda = 3600$. Fig.4(a) illustrates the *noiseless* reference scene (arbitrary units), which contains three spatially localized Ly-α sources, highlighted by white circles. These sources appear spread because of the SSF. The spectra associated to the blue, magenta and green-crossed pixels are those shown in Fig.1.

We have run the detection tests at a same $\text{P}_{\text{FA}} = 10^{-3}$ for each spectral pixel. Due to slight variations of the whitening matrix $\boldsymbol{\Sigma}^{-\frac{1}{2}}$ from one spectrum to another, and to possible estimation errors of the background and of the covariance matrices, we refer to this P_{FA} as an indicative mean value for all the spectra of the data-cube. As visible in Fig.4 this estimated P_{FA} is fairly accurate.

Fig.4(b) and Fig.4(c) report the 1D model-based tests' outcomes, respectively for the unconstrained GLR test (6) and the $\text{GLR}_{1\text{s}}^{(1\text{D})}$ test (13), using $\boldsymbol{R}_{\text{KSVD}_7}$. In these figures, the values of the test statistics above $\xi(\text{P}_{\text{FA}})$ are shown in color: in the black pixels,

[1]Note that, as MUSE is still under construction, we have used a data set specifically simulated and provided by the MUSE consortium (the first real hyperspectral data-cubes are expected in the course of the year 2013).

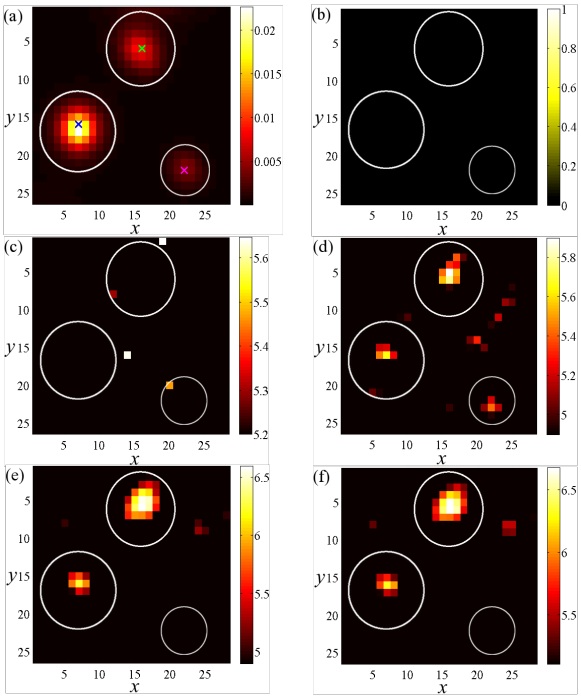


Fig. 4. Compared detection performances of the unconstrained GLR (6) and the $\text{GLR}_{1\text{s}}$ ((13) and (19)) on a simulated MUSE data sub-cube. (a): Reference map (mean over all wavelengths). (b,c): GLR (6) and GLR_{1s} (13) using $\boldsymbol{R}_{\text{KSVD}_7}$. (d-f): 3D GLR_{1s} (19) using $\boldsymbol{R}_{\text{minimax}}$, $\boldsymbol{R}_{\text{SVD}}$ and $\boldsymbol{R}_{\text{KSVD}_7}$.

no detection is found at $\text{P}_{\text{FA}} = 10^{-3}$. As clearly visible, at this P_{FA}, none of the three objects in the scene is detected (similar null results are obtained with 1D tests using $\boldsymbol{R}_{\text{SVD}}$ and $\boldsymbol{R}_{\text{minimax}}$).

The detection results obtained with the 3D model-based $\text{GLR}_{1\text{s}}^{(3\text{D})}$ test are shown in Fig.4(d) using $\boldsymbol{R}_{\text{minimax}}$, in Fig.4(e) using $\boldsymbol{R}_{\text{SVD}}$, and in Fig.4(f) using $\boldsymbol{R}_{\text{KSVD}_7}$, at $\text{P}_{\text{FA}} = 10^{-3}$. This time, two among the three Ly-α *objects* in the scene are clearly detected. We note that some pixels (in the top right corner of the scene) lead to a slight increase in false alarm, probably because of estimation errors of the background. The minimax and (K)-SVD approaches yield essentially comparable results (with slightly worse performances for the minimax approach on these three sources). The third source (bottom right corner of the scene) remains undetected: the features detected around pixel $(x, y) = (22, 22)$ do actually not correspond to real spectral features (the two other do). We believe that this kind of signal illustrates a detection limit in this application.

5. CONCLUSION

In the context of highly noisy 3D data, the proposed method allies dictionary learning techniques, sparse models and spread functions to concentrate information prior to detection. The method is currently generalized to the case where sources are spatially structured (instead of punctual).

In the considered application, this 3D method clearly improves on a 1D approach. We believe there is not much room for amelioration, because of the signals' unknown location in λ and (x, y). To decrease the P_{FA}, while keeping the power essentially constant, possible optimizations are to put cosmological constraints on the translations of the sources in λ, and to shift the learned atoms at multiples of the sampling channels (instead of all channels). In addition, all tests could use the true (instead of approximated) SSF and WSF, to the detriment of a considerable increase of the computational complexity.

6. REFERENCES

[1] T. Vo-Dinh et al., "A hyperspectral imaging system for in vivo optical diagnostics," in *Engineering in Medicine and Biology Magazine, IEEE*, 2004, vol. 23, pp. 40–49.

[2] J.L. Starck, J.M. Fadili, S. Digel, B. Zhang, and J.Chiang, "Source detection using a 3D sparse representation: application to the fermi gamma-ray space telescope," *Astronomy & Astrophysics*, vol. 504, no. 2, pp. 641–652, 2009.

[3] D. Manolakis, D. Marden, and G. A. Shaw, "Hyperspectral image processing for automatic target detection applications," *Lincoln Lab. J.*, vol. 14, no. 1, pp. 79–116, 2003.

[4] D. L. Donoho and J. Jin, "Higher criticism for detecting sparse heterogeneous mixtures," *Annals of Statistics*, vol. 32, no. 3, pp. 962–994, 2004.

[5] E. Arias-Castro, E. J. Candès, and Y. Plan, "Global testing under sparse alternatives: ANOVA, multiple comparisons and the higher criticism," *Annals of Statistics*, vol. 39, no. 5, pp. 2533–2556, 2010.

[6] S. Bourguignon, D. Mary, and E. Slezak, "Restoration of astrophysical spectra with sparsity constraints: models and algorithms," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, pp. 1002 – 1013, September 2011.

[7] S. Mallat, *A wavelet tour of signal processing : the sparse way (3rd Edition)*, Academic Press, 3rd edition, 2008.

[8] S. Bourguignon, D. Mary, and E. Slezak, "Processing MUSE hyperspectral data: Denoising, deconvolution and detection of astrophysical sources," *Statistical Methodology*, vol. 9, pp. 32–43, 2012.

[9] J. Theiler and B. Wohlberg, "Detection of spectrally sparse anomalies in hyperspectral imagery," in *Proc. IEEE Southwest Symposium on Image Analysis and Interpretation*, 2012, pp. 117–120.

[10] J. Haupt and R. Nowak, "Compressive sampling for signal detection," in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, 2007, vol. 3, pp. 1509–1512.

[11] Z. Wang, G. R. Arce, and B. M. Sadler, "Subspace compressive detection for sparse signals," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, 2008, pp. 3873–3876.

[12] J. Paredes, Z. Wang, G. R. Arce, and B. M. Sadler, "Compressive matched subspace detection," in *17th European Signal Processing Conference (EUSIPCO 2009)*, 2009, pp. 120–124.

[13] J. J. Fuchs, "The generalized likelihood ratio test and the sparse representation approach," in *ICISP'10 Proceedings of the 4th International Conference on Image and Signal Processing*, 2010, pp. 245–253.

[14] M. Malloy and R. Nowak, "Sequential analysis in high dimensional multiple testing and sparse recovery," in *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, 2011, pp. 2661–2665.

[15] W. U. Bajwa and A. Pezeshki, "Finite frames for sparse signal processing," in *Finite Frames: Theory and Applications*. Springer, to appear.

[16] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[17] S.S. Chen, D.L. Donoho, and M.A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Review*, vol. 43, no. 1, pp. 129–159, 2001.

[18] J.J. Fuchs, "On sparse representations in arbitrary redundant bases," *IEEE Transactions on Information Theory*, vol. 50, no. 6, pp. 1341 – 1344, June 2004.

[19] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*, vol. 2, Prentice-Hall PTR, 1st edition, 1998.

[20] S. Paris, D. Mary, and A. Ferrari, "Sparsity-based composite detection tests. Application to astrophysical hyperspectral data," in *19th European Signal Processing Conference (EUSIPCO 2011)*, 2011, pp. 1909–1913.

[21] S. Paris, D. Mary, and A. Ferrari, "PDR and LRMAP detection tests applied to massive hyperspectral data," in *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2011 4th IEEE International Workshop on*, 2011, pp. 93–96.

[22] S. Paris, D. Mary, and A. Ferrari, "Detection tests using sparse models, with application to hyperspectral data," *IEEE Trans. Signal Processing*, vol. 61, no. 6, pp. 1481–1494, 2013.

[23] J. Fan, "Test of significance based on wavelet thresholding and Neyman's truncation," *Journal of the American Statistical Association*, vol. 91, no. 434, pp. 674–688, June 1996.

[24] S. Basu, "Bayesian hypotheses testing using posterior density function," *Statistics and Probability Letters*, vol. 30, no. 1, pp. 79–86, 1996.

[25] M. Aharon, M. Elad, and A. M. Bruckstein, "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Processing*, vol. 54, no. 11, pp. 4311–4322, November 2006.

[26] R.F.R. Suleiman, D. Mary, and A. Ferrari, "Minimax sparse detection based on one-class classifiers," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013, Accepted.

[27] K. Chan, J. Li, W. Eichinger, and E. Bai, "A new physics-based method for detecting weak nuclear signals via spectral decomposition," *Nuclear Instruments and Methods in Physics Research A*, vol. 667, pp. 16–25, 2012.

[28] A. Asensio Ramos and R. Manso Sainz, "Signal detection for spectroscopy and polarimetry," *Astronomy & Astrophysics*, vol. 547, 2012.

[29] M.E. Tipping, "The relevance vector machine," in *Advances in Neural Information Processing Systems 12*, S. A. Solla, T. K. Leen, and K.-R. Müller (Eds.), Eds. 2012, pp. 652–658, MIT Press.

[30] D. Manolakis, R. Lockwood, T. Cooley, and J. Jacobson, "Is there a best hyperspectral detection algorithm?," in *Proc. SPIE 7334, Algorithms and Technologies for Multispectral, Hyperspectral and Ultraspectral Imagery XV*, 2009, pp. 733402.1–733402.16.

[31] E.L. Lehmann ad J.P. Romano, *Testing statistical hypotheses*, Springer, 2005.