SPARSITY-BASED COMPOSITE DETECTION TESTS. APPLICATION TO ASTROPHYSICAL HYPERSPECTRAL DATA.

Silvia Paris[1], *David Mary*[1], *André Ferrari*[1], *Sébastien Bourguignon*[2]

[1]UMR 6525 H. Fizeau, [2]UMR 6202 Cassiopée,
Université de Nice-Sophia Antipolis, CNRS, Observatoire de la Côte d'Azur, Campus Valrose, 06108 Nice - cedex 02, FRANCE

Email: [1]{Silvia.Paris, David.Mary, Andre.Ferrari}@unice.fr ; [2]Sebastien.Bourguignon@oca.eu

ABSTRACT

We propose an analysis of some connections existing between sparse estimation and detection tests. In addition to the Generalized Likelihood Ratio (GLR) and to the Bayes Factor, we consider two tests based on Maximum A Posteriori estimates of the sparse vector parameter. These detection tests are then set in order to take advantage of a redundant dictionary, and to account for instrument and noise characteristics specific to the MUSE integral field spectrograph, which will deliver astrophysical hyperspectral data. We use in this framework a specific representation dictionary, designed by finely discretising elementary spectral features (lines with various widths, steps, and continuum parameterisation). We show that the proposed detection strategy is efficient, and outperforms the GLR. Finally we present a possible improvement to this detection strategy, by exploiting spatial dependencies existing in the data cube.

1. INTRODUCTION

The literature in inverse problems has shown in the last decades a particular interest for various "sparsity aware" detection and estimation methods [8]. In an estimation framework, such methods are often based on thresholding functions, some of which can be interpreted in the Maximum A Posteriori (MAP) framework with appropriate priors [1]. Sparse estimation aims to identify spaces of reduced dimension where the information of interest is living, and is therefore related to the problem of detecting an active support. In a detection framework, specific test statistics have been introduced in order to focus on active components as well [5]. Despite the obvious similarity between the sparse estimation and detection approaches, analyses of their connections seem to be absent from the literature, except recently [6].

In this framework, a first objective of the present paper is to point out some connections existing between sparse estimation and detection tests. In particular, we show that the soft thresholding detection test statistics introduced by Fan in [5] correspond to the Posterior Density Ratio (PDR) test discussed by Basu in [2]. We also show that a particular case of the PDR and of another test introduced here (the Likelihood Ratio using a Maximum A Posteriori estimate, LRMAP) leads to consider the regularisation parameter involved in the Basis Pursuit Denoising (BPDN, [4]) problem as a per component detection threshold. Although such an interpretation has already been underlined in the literature (e.g. in [3] and [6]), we establish here its link with the PDR and LRMAP tests.

A second objective of this paper regards the application of "sparsity aware" detection tests to astrophysical hyperspectral data. The detection tests are set in order to take advantage of a redundant dictionary, and to account for instrument and noise characteristics specific to the MUSE instrument, which will deliver astrophysical hyperspectral data. We use in this framework a specific representation dictionary (originally used for spectral restoration in [3]), designed by finely discretising elementary spectral features to be detected in the data (lines with various widths, steps, and continuum parameterisation).

The paper continues with Sec. 2, where we introduce the considered detection tests and summarize some results from [9] which will be necessary for the sequel. The connection between Fan's soft-thresholding detection test and Basu's PDR is made in this section. A comparative analysis of these tests, along with a numerical illustration is proposed for a simple model. We turn to a more general model in Sec. 3, where a redundant dictionary is introduced in order to promote sparsity of the data in the dictionary column space. The Sec. 4 applies the proposed detection tests to hyperspectral astrophysical data. A dictionary specifically designed for the modeling of galactic spectra is exposed, and injected in the detection tests. Connection is made to the works [3] and [6] in Sec. 4.1. In a first step the spectra are considered independent, and spatial dependencies are exploited in a second step. Numerical results are exposed in Sec. 5, before concluding the paper.

2. DETECTION TESTS

We first consider the following hypothesis test:

$$\begin{cases} \mathscr{H}_0 : \underline{x} = \underline{w}, & \underline{w} \sim \mathscr{N}(\underline{0}, I) \\ \mathscr{H}_1 : \underline{x} = \underline{\theta} + \underline{w} \end{cases}, \qquad (1)$$

where $\underline{x}, \underline{\theta}$ and $\underline{w}$ are N-vectors and $\underline{\theta}$ has a few non-zero deterministic parameters. The detection tests we compare are the Generalized Likelihood Ratio (GLR), the Bayes Factor (BF), the Posterior Density Ratio (PDR, [2]) and a Likelihood Ratio (LR) using a MAP estimate. Denoting by $\pi(\underline{\theta})$ a prior probability distribution, the definitions of tests statistics are:

Generalized Likelihood Ratio: $$GLR(\underline{x}) = \frac{\max_{\underline{\theta} \neq \underline{0}} p(\underline{x} \mid \underline{\theta})}{p(\underline{x} \mid \underline{0})};$$

Bayes Factor: $$BF(\underline{x}) = \frac{\int_{\mathbb{R}^N} p(\underline{x} \mid \underline{\theta}) \pi(\underline{\theta}) d\underline{\theta}}{p(\underline{x} \mid \underline{0})};$$

Posterior Density Ratio [2]: $$PDR(\underline{x}) = \frac{\max_{\underline{\theta} \neq \underline{0}} p(\underline{\theta} \mid \underline{x})}{p(\underline{0} \mid \underline{x})};$$

LR using $\underline{\theta} | \mathscr{H}_1 = \hat{\underline{\theta}}_{MAP}$: $$LRMAP(\underline{x}) = \frac{p(\underline{x} \mid \hat{\underline{\theta}}_{MAP})}{p(\underline{x} \mid \underline{0})},$$

where $\hat{\underline{\theta}}_{MAP} = \arg\max_{\underline{\theta}} p(\underline{x} \mid \underline{\theta}) \pi(\underline{\theta})$.

As both the Maximum Likelihood (ML) and the Minimum Mean Squared Error estimates of $\underline{\theta}$ are in general not sparse, we expect the GLR and the BF not to be optimal here. We shall nevertheless use these tests for the purpose of comparison. In contrast, and as now widely known, MAP estimates using a Laplacian prior (or more generally any probability density function which is strictly decreasing at 0) yields estimates $\widehat{\underline{\theta}}_{MAP}$ where 0 values are favored [1]. This makes such priors interesting for detection in the case of sparse parameters, and is indeed one motivation for considering the PDR and introducing the LRMAP as defined above.

The computation of each test statistic, considering a Laplacian prior $\pi(\underline{\theta}) = \prod_i \frac{1}{2\lambda_i} \exp^{-|\theta_i|/\lambda_i}$ for the BF, the PDR and the LRMAP,

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

yields, up to constants that can be included in the thresholds of the tests, to [9]:

- $T_{GLR}(\underline{x}) = \|\underline{x}\|^2$,
- $T_{BF}(\underline{x}) = \prod_{i=1}^{N} e^{\frac{x_i^2}{2}} \left\{ \left[2\cosh(\frac{x_i}{\lambda_i}) - \Phi\left(-x_i + \frac{1}{\lambda_i}\right) - \Phi\left(x_i + \frac{1}{\lambda_i}\right) \right] \right\}$,
- $T_{PDR}(\underline{x}) = \sum_{i \in \mathbf{NZ}} (|x_i| - \frac{1}{\lambda_i})^2$,
- $T_{LRMAP}(\underline{x}) = \sum_{i \in \mathbf{NZ}} (x_i^2 - \frac{1}{\lambda_i^2})$,

where $\Phi(\cdot)$ denotes the cumulative distribution function of a standard normal distribution, and $\mathbf{NZ}$ is the set of indices for which $\hat{\theta}_{MAP_i} \neq 0$. For model (1) with Laplacian prior, the MAP estimate is well known to correspond to a soft-thresholding of each x_i at thresholds $1/\lambda_i$ [8]. Consequently, using the indicator function $I(\cdot)$ the statistics of PDR and LRMAP further simplify to

$$T_{PDR}(\underline{x}) = \sum_{i=1}^{N} (|x_i| - \frac{1}{\lambda_i})^2 I(|x_i| > \frac{1}{\lambda_i}), \quad (2)$$

$$T_{LRMAP}(\underline{x}) = \sum_{i=1}^{N} (x_i^2 - \frac{1}{\lambda_i^2}) I(|x_i| > \frac{1}{\lambda_i}).$$

These expressions show that in order to set the same false alarm rate on each component, the $1/\lambda_i$ should be made equal to some value, say η.

Interestingly, the latter form of T_{PDR} corresponds to the soft thresholding statistics introduced in detection by Fan [5], with the objective of focusing the test on active components only, in order to mimic an Oracle to which the support of $\underline{\theta}$ is known. The PDR test introduced by Basu [2] consequently furnishes a precise framework to soft thresholding in detection.

Note finally that the results above hold not only for (1) but also for two slightly more general models, namely

$$\mathscr{H}_1 : \underline{z} = \underline{\alpha} + \underline{\varepsilon}, \quad (3)$$

where $\underline{\varepsilon} \sim \mathscr{N}(0, \Sigma)$, $\Sigma = \text{diag}\{\sigma_1^2, \ldots, \sigma_N^2\}$, and

$$\mathscr{H}_1 : \underline{z} = B\underline{\alpha} + \underline{w},$$

where $\underline{w} \sim \mathscr{N}(0, I)$ and B is an orthonormal dictionary (a wavelet basis for instance). Indeed, both models reduce to (1), the former by noting $\underline{x} = \Sigma^{-\frac{1}{2}} \underline{z}$, $\underline{\theta} = \Sigma^{-\frac{1}{2}} \underline{\alpha}$ and $\underline{w} = \Sigma^{-\frac{1}{2}} \underline{\varepsilon}$, and the latter by noting $\underline{x} = B^t \underline{z}$ and $\underline{\theta} = B^t \underline{\alpha}$. For (3), the hyperparameters λ_i should be taken as $\lambda_i = \sigma_i / \eta$ in order to have a uniform false alarm rate on all components (see Sec. 4.1 for a discussion on this point).

Coming back to the PDR and LRMAP for model (1), these tests are of the form $T(x) \gtrless_{H_0}^{H_1} \gamma$, with T_{PDR} and T_{LRMAP} positive as clearly visible from (2). For both tests, the maximal false alarm rate at a fixed η is obtained by setting $\gamma = 0$:

$$\begin{aligned} P_{FA_0} = Pr(T > 0 | \mathscr{H}_0) &= 1 - Pr(|x_i| < \eta \ \forall i | \mathscr{H}_0) \\ &= 1 - (2\Phi(\eta) - 1)^N. \end{aligned} \quad (4)$$

Denoting by $\mathbf{NZ}^*$ the set of indices for which $\theta_i \neq 0$, we obtain the detection probability at $\gamma = 0$ for both tests:

$$\begin{aligned} &P_{DET_0} = Pr(T > 0 | \mathscr{H}_1) \\ &= 1 - (2\Phi(\eta) - 1)^{[N - \text{Card}(\mathbf{NZ}^*)]} \prod_{i \in \mathbf{NZ}^*} [2 - \Phi(\eta - \theta_i) - \Phi(\eta + \theta_i)]. \end{aligned} \quad (5)$$

For $\gamma = 0$, the Receiver Operating Characteristics (ROC) of the PDR and LRMAP tests are thus precisely characterized by eq.(4) and (5).

Fig.1 compares the ROC curves of GLR, BF, PDR and LRMAP tests for a sparse vector with two non-zero components of amplitudes 3 and 3.79. The Neyman-Pearson test (for which the alternative $\underline{\theta}$ is known) is shown in cyan for reference. For the case $\eta = 2$ first, the ROC of the GLR, BF, LRMAP and PDR are respectively plotted in solid yellow, dashed dot black, solid blue and dotted blue. As expected, PDR and LRMAP outperform the BF and GLR. Both test statistics essentially focus on the largest components in $\underline{x}$, and therefore mimic the Oracle. Note that the PDR ROC curve is superior to that of the LRMAP in this case, except at the point given by the two equations (4) and (5). (This point is the top right here : $P_{FA_0} \approx P_{DET_0} \approx 1$ for $\eta = 2$). However, this is not systematic and it is parameter dependent. We will not pursue the comparison between PDR and LRMAP at $\gamma \neq 0$ for the reasons exposed at the end of this section.

For two other values of η (3.46: green and 3.8: red), we have plotted only the ROC of PDR and LRMAP (solid and dashed curves respectively). At high values of η, the ROC of both tests tend to be equivalent. They are strictly equivalent for $\gamma = 0$, which correspond to the magenta points. The locus of such points corresponds to the maximum probability of false alarm at a given η, and is written in closed form in eq. (4) and (5). In the general case, the PDR and

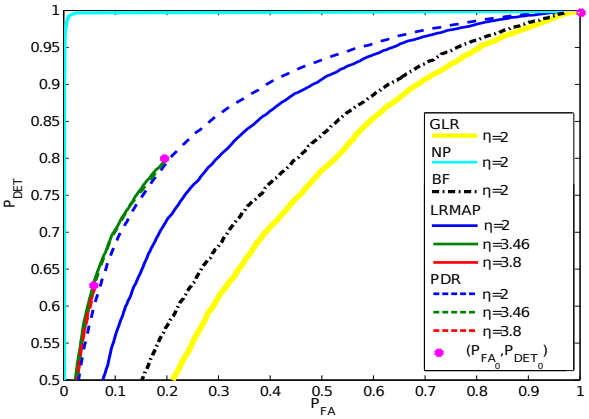


Figure 1: ROC of the NP, GLR, BF, LRMAP and PDR detectors. The NP, GLR and the BF tests are shown only for $\eta = 2$. At fixed η, the maximal extension in P_{FA} of the ROC curve of the PDR and LRMAP tests can be computed analytically through (4) and (5) (magenta points). These points correspond to set $\gamma = 0$, in which case the two tests become equivalent.

LRMAP tests depend on two parameters, η and γ. The strategy of setting $\gamma = 0$ allows to precisely control the P_{FA} and the P_{DET} of the test through the sole tuning of η via eq.(4) and (5).

We note however that numerical results (not displayed here) show that this strategy may not be optimal: PDR and LRMAP tests may achieve a higher P_{DET} for the same $P_{FA} = P_{FA_0}$ obtained for $\gamma = 0$, by using a different value of η and $\gamma \neq 0$. But since the distribution of T_{PDR} and T_{LRMAP} are not known for finite N [9], in practice the $P_{FA} = Pr(T(\eta) > \gamma | \mathscr{H}_0)$ of such tests cannot be computed analytically. An empirical evaluation is possible and requires to sample numerically the distribution of $T(\eta)$. To keep the tests simple, we will focus on the $\gamma = 0$ case, for which we show that satisfactory performances can be achieved in practice.

3. DETECTION TESTS BASED ON A REDUNDANT DICTIONARY

We now turn to a more general model where we aim to promote sparsity through a redundant dictionary. We test $\mathscr{H}_0 : \underline{x} = \underline{w}$ against $\mathscr{H}_1 : \underline{x} = D\underline{\theta} + \underline{w}$ where $D = [\underline{D}_1 \ldots \underline{D}_N]$ is a redundant normalized dictionary of size $(N \times L)$ with $L > N$ atoms, $\underline{\theta}$ is a column vector of L components, assumed to be sparse and $\underline{w} \sim \mathscr{N}(\underline{0}, I)$. We compare again the PDR and the LRMAP for a Laplacian prior with parameter $\lambda = \frac{1}{\eta}$, to the GLR.

GLR: The ML estimate of $\underline{\theta}$ is

$$\underline{\hat{\theta}}_{ML} = \arg\max_{\underline{\theta}} p(\underline{x} \mid \underline{\theta}) = \arg\min_{\underline{\theta}} \frac{1}{2} \|\underline{x} - D\underline{\theta}\|^2 . \quad (6)$$

Since D is $(N \times L)$ with $L > N$, $\underline{\hat{\theta}}_{ML}$ achieves $D\underline{\hat{\theta}}_{ML} = \underline{x}$. Thus, it is easy to see that $T_{GLR}(\underline{x}) = \|\underline{x}\|^2$ as before.

LRMAP and PDR: Considering the model above, $\underline{\hat{\theta}}_{MAP}$ is the solution of

$$\underline{\hat{\theta}}_{MAP} = \arg\min_{\underline{\theta}} \frac{1}{2} \|\underline{x} - D\underline{\theta}\|^2 + \eta \|\underline{\theta}\|_1 , \quad (7)$$

which is also called BPDN [4]. In contrast to Sec.2, $\underline{\hat{\theta}}_{MAP}$ cannot be obtained by direct soft thresholding. If $\underline{\hat{\theta}}_{MAP}$ verifies (7), there exists $\underline{u}$ which verifies [7]

$$D^t (D\underline{\hat{\theta}}_{MAP} - \underline{x}) + \eta \underline{u} = 0, \quad (8)$$

where

$$\begin{cases} u_i = \text{sgn}(\hat{\theta}_{MAP_i}) & \text{if} \quad \hat{\theta}_{MAP_i} \neq 0 \\ u_i \leq 1 & \text{if} \quad \hat{\theta}_{MAP_i} = 0 \end{cases} . \quad (9)$$

As a consequence, for sufficiently large η, $\underline{\hat{\theta}}_{MAP}$ is identically zero. The first non-zero component of $\underline{\hat{\theta}}_{MAP}$ appears when η falls below $\max_i(|\underline{D_i}^t \underline{x}|)$: if $\eta > \max_i(|\underline{D_i}^t \underline{x}|)$, it is easy to see that taking $\underline{\hat{\theta}}_{MAP} = \underline{0}$ in (8) automatically yields a $\underline{u}$ which satisfies (9).

By multiplying (8) by $\underline{\hat{\theta}}^t_{MAP}$ and by noting $\underline{\hat{x}}_{MAP} = D\underline{\hat{\theta}}_{MAP}$, we obtain

$$\underline{\hat{x}}^t_{MAP}(\underline{\hat{x}}_{MAP} - \underline{x}) + \eta ||\underline{\hat{\theta}}_{MAP}||_1 = 0. \quad (10)$$

From the definition of the LRMAP we have

$$LRMAP(\underline{x}) = \frac{p(\underline{x} \mid \underline{\hat{\theta}}_{MAP})}{p(\underline{x} \mid \underline{0})} \Rightarrow T_{LRMAP} = \log(LRMAP) \quad (11)$$
$$= \underline{x}^t \underline{\hat{x}}_{MAP} - \frac{1}{2} \underline{\hat{x}}^t_{MAP} \underline{\hat{x}}_{MAP}.$$

By adding $\frac{1}{2}\underline{\hat{x}}^t_{MAP}\underline{\hat{x}}_{MAP}$ to both terms of (10) and comparing to T_{LRMAP} above, we obtain the test statistics for the LRMAP (and similarly for the PDR) :

$$T_{LRMAP} = \eta \left\|\underline{\hat{\theta}}_{MAP}\right\|_1 + \frac{1}{2} \underline{\hat{x}}^t_{MAP} \underline{\hat{x}}_{MAP}; \; T_{PDR} = \frac{1}{2} \underline{\hat{x}}^t_{MAP} \underline{\hat{x}}_{MAP}, \quad (12)$$

from which we see that both test statistics are strictly positive, or null if $\underline{\hat{\theta}}_{MAP} = \underline{0}$ (that is, if $\max_i(|\underline{D_i}^t \underline{x}|) \leq \eta$). For a fixed value of η, the maximal P_{FA} is thus given for both tests at $\gamma = 0$ by

$$P_{FA_0} = Pr(T > 0 \mid \mathscr{H}_0) = Pr(\max_i(|\underline{D_i}^t \underline{x}|) > \eta \mid \mathscr{H}_0). \quad (13)$$

Since D is redundant, the components of the L-vector $D^t\underline{x}$ are not independent, and finding P_{FA_0} analytically is thus a difficult problem (see [6] and references therein). We can nevertheless resort to Monte Carlo (MC) simulations to obtain an accurate correspondence between η and P_{FA_0} (as proposed in Sec. 5). Note finally that for $\gamma = 0$, it is not necessary to solve (7) to implement these detection tests, as they amount to compare $\max_i(|\underline{D_i}^t \underline{x}|)$ to η.

4. APPLICATION TO HYPERSPECTRAL DATA

We consider detection tests based on sparse estimation in the context of astrophysical spectra restoration for the forthcoming MUSE instrument. MUSE is an integral-field spectrograph, which will be installed at the Very Large Telescope (ESO, Chile) in 2012 and will deliver data cubes composed of 300×300 spectra sampled at ≈ 3400 wavelengths of the visible spectrum. One of the major challenges of MUSE concerns the detection and characterization of very distant galaxies. Such light sources are very faint, spatially localized within a few pixels and may show only a few spectrally salient features. Data will be acquired with very low signal-to-noise ratio. In particular, data will be strongly contaminated by the spectral signature of atmospheric molecules. In addition, data will be affected by a strong Poisson noise, which is indeed data dependent. Moreover, the instrumental detection efficiency is variable with wavelength. Consequently, the noise level is highly variable from one wavelength to another, and also from one spectrum to another. This has consequences on the setting of the detection tests, which will be made apparent in Sec. 4.1.

Considered as a linear system, MUSE is characterized by its three-dimensional point spread function (PSF, the impulse response in both spatial and spectral domains), that can be separated into the spatial PSF, which typically covers 7×7 pixels, and the spectral line spread function (LSF), which spreads over 11 spectral elements. Note that MUSE is still under construction, so only simulated data are available until now. These data result from high-complexity astrophysical simulations, and account for noise and instrument characteristics.

Restoration of MUSE-like spectra was recently addressed in [3], where prior information was incorporated through sparsity constraints. A redundant dictionary R of elementary spectral features was built in accordance with astrophysical knowledge, so that the sparsely estimated non-zero components can be interpreted as physically meaningful features. More precisely, R concatenates three sub-dictionaries : $R = [R^\ell R^c R^b]$, each of which corresponds to a specific spectral component: a line spectrum, a continuous spectrum and a series of discontinuities. R^ℓ is a dictionary of discrete splines with several widths, which are centered along the reconstruction wavelength axis. Eleven width values were used, varying from 1 (delta functions) to 138 points. Delta functions and splines model respectively unresolved and resolved absorption or emission spectral lines. MUSE' spectral resolution equals 0.13 nm so that the maximal width equals $138 \times 0.13 \simeq 18$ nm. The continuous spectrum is composed of sine functions with low frequencies (reduced frequencies vary from $1/N$ to $8/N$, where N is the number of data) and 8 discretized phase shifts. It also includes the continuous component by means of a constant signal. Finally, dictionary R^b models a series of breaks in the spectrum, and is composed of step functions which are also centered along the wavelength axis. While the works [3] used this dictionary in the framework of spectral restoration, we use it for detection purpose here. We first consider MUSE spectra as spatially independent, and then suggest a strategy to exploit spatial dependencies.

4.1 Detection for independent spectra

Denoting by H the matrix form of the LSF, the observation model reads

$$\mathscr{H}_1 : \underline{y} = H\underline{s} + \underline{\varepsilon} = HR\underline{\alpha} + \underline{\varepsilon}, \quad (14)$$

where H is $(N \times N)$, R is $(N \times L)$, $\underline{\alpha}$ is a sparse L-vector, and $\underline{\varepsilon}$ is supposed Gaussian with known covariance matrix Σ. Working with weighted data $\Sigma^{-\frac{1}{2}}\underline{y}$, the model becomes

$$\mathscr{H}_1 : \Sigma^{-\frac{1}{2}}\underline{y} = \Sigma^{-\frac{1}{2}} HR\underline{\alpha} + \underline{w}, \quad (15)$$

where $\Sigma^{-\frac{1}{2}} HR = D_{\Sigma H}$ appears as an equivalent dictionary, and $\underline{w} \sim \mathscr{N}(\underline{0}, I)$. Noting finally $\underline{x} = \Sigma^{-\frac{1}{2}}\underline{y}$, $D = D_{\Sigma H} N^{-1}_{D_{\Sigma H}}$ and $\underline{\theta} = N_{D_{\Sigma H}}\underline{\alpha}$, with $N_{D_{\Sigma H}}$ the diagonal matrix composed of the norms of columns of $D_{\Sigma H}$, we obtain the same model as in Sec. 3, that is:

$$\mathscr{H}_1 \; : \; \underline{x} = D\underline{\theta} + \underline{w}. \quad (16)$$

Interestingly, it was shown in [3] that the normalization of the equivalent dictionary $D_{\Sigma H} = \Sigma^{-\frac{1}{2}} HR$ leads to uniform false alarm rate on

all coefficients of $\underline{\theta}$, which is controlled by the parameter η of the BPDN. A similar interpretation of η can be found in [6]. From the previous section, the PDR and LRMAP tests $T_{PDR/LRMAP}(\eta) \gtrless_{H_0}^{H_1} \gamma$, where $T_{PDR/LRMAP}$ are as in eq.(12), lead to this interpretation in the particular case where $\gamma = 0$.

Let us finally mention that for MUSE data, implementing the PDR and LRMAP tests as described above requires to compute and normalize one dictionary $D_{\Sigma H} = \Sigma^{-\frac{1}{2}} HR$ per spectrum, since, as discussed at the beginning of this Section, the noise power is strongly variable spatially.

4.2 Improved detection exploiting spatial dependencies

As seen in the introduction of this Section, two spatially contiguous spectra are likely to share some spectral information, either because the galactic source is resolved (i.e. spread over several pixels), or because of the spatial PSF. After performing the detection tests described in eq.(12) and (13) for a target $P_{FA_0}(\eta)$ on the MUSE data cube, we are left with a set $\mathbf{\Gamma}$ of spectra flagged as "detected", and a complementary set $\bar{\mathbf{\Gamma}}$ of spectra in which no feature was found at the significance level set by P_{FA_0}. We propose the following strategy to improve the detection on some spectra of $\bar{\mathbf{\Gamma}}$: we use the spectral feature(s) estimated in neighbor spectra, in which significant features have been detected, to set a "second round" detection test. To be more specific, let us denote by $\underline{x}_b \in \mathbf{\Gamma}$ and $\underline{x}_f \in \bar{\mathbf{\Gamma}}$ two neighbor weighted spectra, the former having bright (detected) spectral features, and the latter no detected spectral feature. In a first (crude) approximation, our model will assume that the two noiseless spectra are the same. A meaningful estimate $\underline{\hat{x}}_f$ of $\underline{x}_f$ can be computed as $\underline{\hat{x}}_f = D\underline{\hat{\theta}}(\underline{x}_b) = \underline{\hat{x}}_b$, where $\underline{\hat{\theta}}(\underline{x}_b)$ is a sparse parameter vector estimate based on the bright neighbor spectrum. $\underline{\hat{\theta}}(\underline{x}_b)$ can for instance be obtained as the solution of eq.(7) for $\underline{x} = \underline{x}_b$ (efficient strategies are discussed in [3]). A fast alternative is to use a greedy algorithm such as the Matching Pursuit (MP) [8] for instance. Since in the first approximation considered here we assume that the neighbor spectra are the same, we have for the faint spectrum, the following simple model under $\mathscr{H}_1$:

$$\underline{x}_f = \underline{\hat{x}}_f + \underline{w}, \qquad \text{with} \qquad \underline{\hat{x}}_f = D\underline{\hat{\theta}}(\underline{x}_b). \tag{17}$$

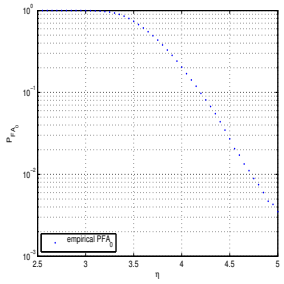
This model accounts for spatial dependency.
Injecting the estimate $\underline{\hat{x}}_f$ in the Likelihood Ratio, a matched filter-like second round detection test for $\underline{x}_f$ is therefore

$$\frac{p(\underline{x}_f|\underline{\hat{x}}_f)}{p(\underline{x}_f|\underline{0})} \gtrless_{H_0}^{H_1} \gamma' \qquad \Leftrightarrow \qquad \underline{x}_f^t \underline{\hat{x}}_f - \frac{1}{2}\underline{\hat{x}}_f^t \underline{\hat{x}}_f \gtrless_{H_0}^{H_1} \gamma, \tag{18}$$

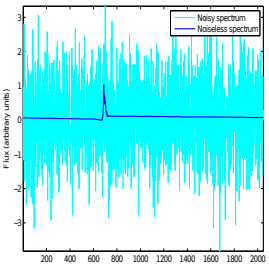
which is obtained similarly as in eq.(12) and where $\gamma = \log \gamma'$. The corresponding $P_{FA}(\gamma)$ is $1 - \Phi(\frac{\gamma + \frac{1}{2}\underline{\hat{x}}_f^t \underline{\hat{x}}_f}{||\underline{\hat{x}}_f||})$.

5. NUMERICAL RESULTS

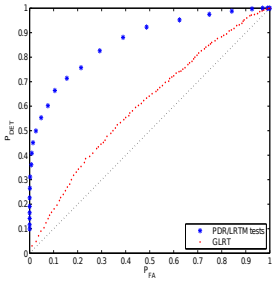
For the purpose of MC simulation we will show results for spectra shorter than those of MUSE ($N = 2048$ instead of ≈ 3400). To quantify how the noise $\underline{\varepsilon}$ affects a particular convolved spectrum $\underline{s}_H = H\underline{s}$ in (14), we define the Signal-to-Noise Ratio as $SNR(\underline{s}_H) = 10\log_{10} \frac{||\underline{s}_H||^2}{\mathrm{Tr}\{\Sigma\}}$. To fix the ideas, a SNR of 0dB corresponds to a noise that has the same power as the convolved spectrum, and a SNR of -20dB corresponds to a noise power hundred times greater than the power of the convolved spectrum.
We first compare the detection performance of GLR and PDR/LRMAP on one of the spectra simulated by the MUSE consortium. The model is (14), and in this simulation $\Sigma = I$ (spectrally variable noise will be considered in the next experiments). The first step is to compute via MC simulation the correspondence $P_{FA_0} \Leftrightarrow \eta$ (cf eq.(13)). This is shown in Fig 2(a). For instance, $P_{FA} = 0.01 \Leftrightarrow \eta = 4.72$. The considered spectrum is shown



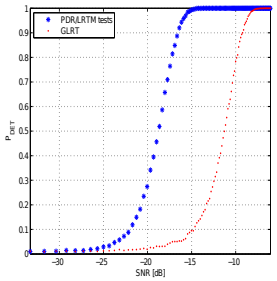
(a) Empirical $P_{FA_0}(\eta)$.



(b) $H\underline{s}$ (blue) and corresponding data (cyan) at $SNR = -19.3dB$.



(c) ROC of PDR/LRMAP vs GLR at fixed $SNR = -19.3dB$.



(d) P_{DET} vs SNR at fixed $P_{FA} = 0.01$.

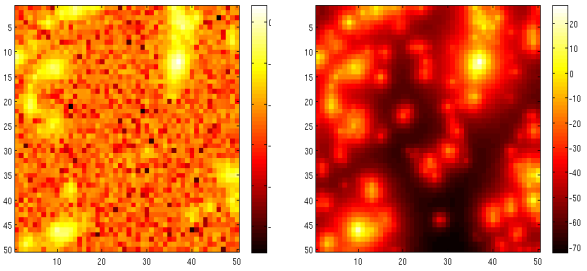
Figure 2: Compared detection performances of PDR/LRMAP and GLR on a MUSE spectrum.

in Fig 2(b) in blue, and the corresponding noisy data are shown in cyan. The SNR equals -19.3dB here, which is representative of MUSE spectra (compare with Fig. 3(b)). Fig 2(c) compares the ROC curves of PDR/LRMAP (blue) vs GLR (red) at fixed $SNR = -19.3$dB. For the GLR, the threshold γ corresponding to a desired P_{FA} is easily obtained by

$$P_{FA}(\gamma) = Pr(||\underline{x}||^2 > \gamma) = 1 - \Phi_{\chi^2_N}(\gamma) \Rightarrow \gamma = \Phi^{-1}_{\chi^2_N}(1 - P_{FA}). \tag{19}$$

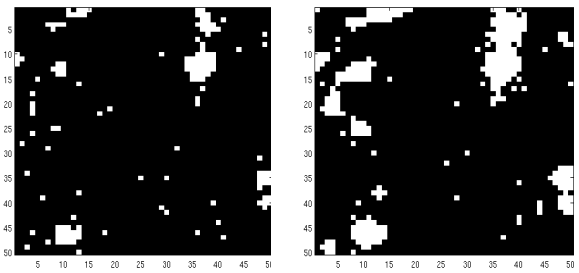
Clearly the PDR/LRMAP tests are superior to the GLR. While the latter test is "blind" to the dictionary model and acts as an energy detector, the former tests allow to detect more efficiently particular spectral features (the emission line visible in 2(b) is detected by a wide spline in the dictionary here). Turning now to Fig. 2(d), the detection performances of PDR/LRMAP are compared at fixed $P_{FA} = 0.01$ versus SNR. The tests allow better detection w.r.t. GLR for a rather wide range of SNR (-10 to -20dB for this spectrum). The tests are compared on a reduced MUSE data cube of size 50×50 in Fig 3, which is illustrated in Fig 3(a).The SNR of each spectrum is shown in Fig 3(b). Most spectra are buried in noise and have SNR below -20dB. The PDR/LRMAP and the GLR were run on each noisy spectrum, using one dictionary $D_{\Sigma H} = \Sigma^{-\frac{1}{2}} HR$ per spectrum (there is one noise matrix Σ per spectrum; an example of the noise variances involved in the diagonal of Σ is shown if Fig. 4, top). The FA rate is 0.01 for both tests. Fig 3(c) shows in white the location of the spectra which have been detected using GLR, and Fig 3(d) shows the equivalent for the PDR/LRMAP. At the same FA, the GLR has rejected $\mathscr{H}_0$ in 146 cases, while the PDR/LRMAP has detected at least one significant spectral feature in 304 spectra.

We finally turn to detection using spatial dependencies. The relevance of the approach is illustrated on a synthetic bright (weighted) spectrum $\underline{x}_b = D\underline{\theta}_b + \underline{\varepsilon}$, where $\underline{\varepsilon} \sim \mathscr{N}(\underline{0}, I)$ and $\underline{\theta}_b$ is a sparse parameter vector of only 4 non-zero components (corresponding to one step, one smooth global oscillation, plus two lines). The neighbor spectrum is taken with parameters $\underline{\theta}_f = 0.5\underline{\theta}_b$. The factor 0.5 is introduced in order to make model (17) only approximately accurate and this experiment slightly more realistic, as in practice neighbor spectra will not have the same parameters. The MUSE noise model displayed in Fig.4 top was used to simulate noisy data. The



(a) Noisy MUSE subcube : Absolute value of the mean over wavelengths (logscale).

(b) SNR of each spectrum.



(c) Detection using the GLR

(d) Detection using the PDR/LRMAP

Figure 3: Compared detection performances of PDR/LRMAP and GLR on a simulated MUSE datacube.

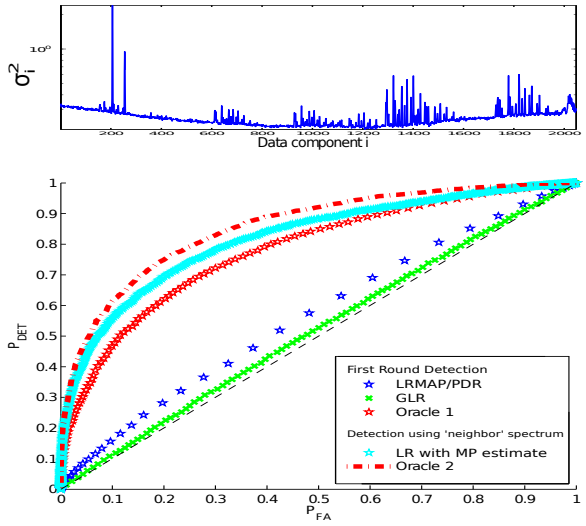


Figure 4: Top : Example of noise variances for a MUSE spectrum. Bottom : compared ROC by considering contiguous spectra as independent ("First Round Detection") and by exploiting spatial dependencies ("Using neighbor spectrum").

SNR corresponding to the two convolved spectra are -20.9dB and -26.9dB respectively. The strictly sparse parameter vectors considered here allow us to compare the performance of the proposed detection tests with those of Oracles, which we define as follows. The Oracle considering only the faint spectrum (Oracle 1 for short) has knowledge of $\mathbf{NZ}^*$, the support of $\underline{\theta}_f$ (which is the same as that of $\underline{\theta}_b$), but not of the amplitudes, which are estimated by least squares. Denoting by $D_{\mathbf{NZ}^*}$ the restriction of D to the column corresponding to $\mathbf{NZ}^*$, this leads to $\hat{\underline{\theta}}_{Or,f} = (D^t_{\mathbf{NZ}^*} D_{\mathbf{NZ}^*})^{-1} D^t_{\mathbf{NZ}^*} \underline{x}_f$, and to a spectrum estimate $\hat{\underline{x}}_{Or,f} = D_{\mathbf{NZ}^*} \hat{\underline{\theta}}_{Or,f}$. This estimate can be used in a matched filter-like test where the model is: $\underline{x} = \hat{\underline{x}}_{Or,f} + \underline{\varepsilon}$ under $\mathscr{H}_1$ and $\underline{x} = \underline{\varepsilon}$ under $\mathscr{H}_0$. This yields $\underline{x}^t_f \hat{\underline{x}}_{Or,f} - \frac{1}{2} \hat{\underline{x}}^t_{Or,f} \hat{\underline{x}}_{Or,f} \gtrless^{H_1}_{H_0} \gamma$. The ROC of the Oracle considering only the faint spectrum is indeed an upper bound to any realizable detection test which considers this data spectrum independently from others. Consider now the case of an Oracle which estimates the parameter amplitudes using $\underline{x}_b$ (Oracle 2). Since both parameters $\underline{\theta}_f$ and $\underline{\theta}_b$ are similar, the resulting estimate will still be meaningful, up to a scaling, for the faint spectrum. This leads to an estimate $\hat{\underline{\theta}}'_{Or,f} = (D^t_{\mathbf{NZ}^*} D_{\mathbf{NZ}^*})^{-1} D^t_{\mathbf{NZ}^*} \underline{x}_b$, and to a spectrum estimate $\hat{\underline{x}}'_{Or,f} = D_{\mathbf{NZ}^*} \hat{\underline{\theta}}'_{Or,f}$. Since the bright spectrum is relatively less noisy than the faint one, we expect the Oracle 2 to perform better than Oracle 1.

Turning now to the numerical results in Fig.4, the ROC obtained when considering contiguous spectra as independent ("First Round Detection") are compared for the GLR (green crosses), PDR/LRMAP (blue stars), and Oracle 1 (red stars). The PDR/LRMAP and GLR were set to a false alarm of 0.01 ($\eta = 4.72$). As expected, the LRMAP is superior to the GLR, and Oracle 1 is superior to both tests. As a second round detection, we implemented the detection test as in eq.(18), where $\hat{\underline{x}}_f$ was estimated by a MP estimate of $\underline{x}_b$ with threshold $\eta = 4.72$. The resulting ROC is plotted in cyan stars. The improvement w.r.t. the PDR/LRMAP tests is significant and interestingly, the corresponding ROC is even better than the ROC of Oracle 1.

6. CONCLUSIONS

We pointed out some connections existing between sparse estimation and detection tests. In particular, we showed the connection between soft thresholding detection test statistics and PDR. Another test, a LR using the MAP estimate was proposed. These detection tests were then set in order to take advantage of a redundant dictionary. They were applied to astrophysical hyperspectral data, using a dictionary specifically designed for such spectra. We showed that the proposed tests outperform the GLR. Finally, an improvement of the proposed detection method, which takes advantages from spatial dependencies, was discussed and illustrated in a toy case. In practice indeed, hyperspectral neighbor spectra have more complex relationship than assumed here, and the total P_{FA} when using two successive tests must be worked out. These points will be investigated in further works.

7. ACKNOWLEDGEMENTS

The authors acknowledge the work of B.Trémoulhéac on the dictionary R and thank I. Smith for mentionning the existence of reference [2].

REFERENCES

[1] A. Antoniadis and J. Fan. Regularization of wavelet approximations. *Journal of the American Statistical Association*, 96(455):939–955, 2001.

[2] S. Basu. Bayesian hypotheses testing using posterior density function. *Statistics and Probability Letters*, (30), 1996.

[3] S. Bourguignon, D. Mary, and E. Slezak. Restoration of astrophysical spectra with sparsity constraints: models and algorithms. *Accepted for publication to IEEE J. Sel. Topics in Signal Processing*.

[4] S.S. Chen, D.L. Donoho, and M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM Review*, 43(1):129–159, 2001.

[5] J. Fan. Test of significance based on wavelet thresholding and Neyman's truncation. 1994.

[6] J. J. Fuchs. The generalized likelihood ratio test and the sparse representation approach. *ICISP 2010, LNCS 6134*, 2010.

[7] J.J. Fuchs. On sparse representations in arbitrary redundant bases. *IEEE Transactions Inf. Theory*, 50(6):1341–1344, 2004.

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

[9] S. Paris, D. Mary, and A. Ferrari. Composite hypothesis tests for sparse parameters. *IEEE SSP 2011*.