MINIMAX SPARSE DETECTION BASED ON ONE-CLASS CLASSIFIERS

R.F.R. Suleiman, D. Mary and A. Ferrari

Laboratoire J-L. Lagrange, UMR7293,
Université de Nice Sophia Antipolis, CNRS, Observatoire de la Côte d'Azur,
Campus Valrose, 06108 Nice Cedex 02, FRANCE
Email: {raja.fazliza, david.mary, andre.ferrari}@unice.fr

ABSTRACT

We consider the problem of detecting a target signature which is known (up to an amplitude factor) to belong to a (possibly very) large library of signatures. Thus we know how each signature to be detected looks like, but we do not know which one is activated under $\mathcal{H}_1$. We propose a minimax approach for this problem aimed at maximizing the worst detection performance. Optimization issues and connections with One-Class classifiers are discussed and illustrated geometrically. Numerical results comparing the proposed approach to the classical sparse-coding dictionary learning technique K-SVD are provided on astrophysical hyperspectral data.

***Index Terms*—** Detection, minimax, sparsity, dictionary learning, SVM

1. PROBLEM STATEMENT AND PREVIOUS WORKS

We consider in this work the problem of detecting a target signature which is known (up to an amplitude factor) to belong to a (possibly very) large library of signatures. Thus we know how each signature to be detected looks like, but we do not know which one is activated under $\mathcal{H}_1$.

A typical illustration of this problem is the detection of known waveforms in data communications, in which case the man-made waveforms are designed using orthogonality principles to ease the discrimination between the active signals under $\mathcal{H}_1$ [1–4].

Other instances of this problem occur in Hyperspectral Imaging (HSI) for example, where the target signatures are usually imposed and cannot be optimized to improve detection. In some cases, the possible set of target spectra is naturally large (rare minerals detection, infected trees, mine detection, ...). In another set of cases, the possible target spectra are few but the corresponding data come with systematic perturbations that can be modeled, leading again to a large set of possible target spectra [5, 6].

The specific application considered in this article regards the detection of spectral lines of primordial galaxies in HSI data [7–11]. These sources have typical spectral shapes that can be simulated to generate a library of lines. In the data, each spectral line, appears, if present, shifted in wavelengths by an *a priori* unknown amount (which depends on the galaxies' distances), leading to a very large library of possible features under $\mathcal{H}_1$.

Detection methods can be divided in two main classes. The first class assumes that the target is known and leads to techniques relying on Matched Filter principles (those are called spectral matching (SM) methods in the HSI literature). When such knowledge is not available, the other class of approaches is aimed at discriminating the pixels that are different (in ways that are specific to each method) from the background. Those are called anomaly detection (AD) approaches [12].

When a reliable knowledge about the target features is available, SM techniques have larger detection power than AD techniques because SM indeed benefits from this knowledge. When this knowledge is encapsulated in a spectral library, and when this library is large, detection tests traditionally operate in subspaces of reduced dimensions which are obtained by various means (sample mean, SVD, or endmember selection techniques) [13, 14].

Dimension reduction has obvious interests in terms of complexity: the computational cost for testing the full library is, in some of the cases cited above, simply prohibitive. In addition, reducing the dimension may lead to globally improved detection power. However, and this observation is actually at the origin of the present work, this comes at a price: low-dimensionality tests usually present very low power for some targets – with obvious and potentially highly damageable consequences in health or mine detection surveys for instance.

In this framework, our contribution is the following. We propose a test that operates dimensionality reduction from a large spectral library and for which the target subspace is learned to maximize the worst detection case under $\mathcal{H}_1$. The test is thus optimized with respect to (w.r.t.) a minimax criterion, which consists in maximizing the worst probability of detection. This technique may be seen as a particular (i.e. robust, or minimax) case of subspace target Matched filters [14, 15] or Matched subspace detectors [2].

Minimaxity in detection has received a lot of attention in the litterature (*e.g.* chap. 8 and 9 from [16], [17]), but we are not aware of works addressing the specific problem introduced above. Note that we restrict the scope of this paper to techniques that exploit spectral information only. In the cases where the target spreads over several pixels, spatio-spectral techniques can further improve detection (see *e.g.* [18]).

The paper continues by defining in Sec.2 our model and the corresponding (constrained) Generalized Likelihood Ratio Test (GLRT). We also highlight in this Section the effects of dimension reduction on the detection power. Sec.3 formalizes and introduces the considered minimax detection criterion. The proposed optimization problem is then solved in the 1-dimensional case in Sec.4, where convexity issues are investigated. We also detail in this Section, the connections of the proposed approach with the well known Support Vector Machine (SVM). Sec.5 presents numerical results for astrophysics hyperspectral data.

This work was supported by ANR project ANR-12-BS03-0003 Hypanema. R.F.R. Suleiman is funded by Majlis Amanah Rakyat (MARA).

2. SPARSE MODEL

The model discussed in the Introduction can in its simplest form be written as the following composite hypotheses :

$$\begin{cases} \mathcal{H}_0 & : \mathbf{x} = \mathbf{n}, \qquad\qquad \mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_N) \\ \mathcal{H}_1 & : \mathbf{x} = \mathbf{S}\boldsymbol{\alpha} + \mathbf{n}, \quad ||\boldsymbol{\alpha}||_0 = 1 \end{cases} \tag{1}$$

where $\mathbf{x}$ and $\mathbf{n} \in \mathbb{R}^N$. This model assumes no (or perfectly subtracted) background, and a covariance matrix that is known and equal under both hypotheses. In the case of a correlated noise model with known covariance matrix $\mathbf{R}$, i.e. $\mathcal{H}_1 : \mathbf{x} = \mathbf{S}\boldsymbol{\alpha} + \mathbf{n}$, $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$, considering the weighted data $\mathbf{R}^{-1/2}\mathbf{x}$ leads to a model of the form (1) [9].

We denote by $\mathbf{S} \in \mathbb{R}^{N \times L} = [\mathbf{s_1}, \mathbf{s_2}, \ldots, \mathbf{s_L}]$ the (known) set of the possible signal features, with L possibly much larger than N. $\mathbf{S}$ has normalized columns ($\|\mathbf{s}_i\|_2^2 = 1$, $\forall i = 1, \ldots, L$). Under $\mathcal{H}_1$, only one feature is activated, and this is expressed by imposing that the unknown vector $\boldsymbol{\alpha} \in \mathbb{R}^L$ has unit ℓ_0 pseudonorm.

The GLRT for model (1) with the constraint $||\boldsymbol{\alpha}||_0 = 1$ is:

$$\max_{\boldsymbol{\alpha}:||\boldsymbol{\alpha}||_0=1} \frac{P_r(\mathbf{x}|\mathbf{S}\boldsymbol{\alpha})}{P_r(\mathbf{x}|\mathbf{0})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma'. \tag{2}$$

This maximization should be performed over the index i, $i = 1, \ldots, L$, of the nonzero component, and over the corresponding value $\boldsymbol{\alpha}_i$. It is easy to see that maximizing on a given nonzero component $\boldsymbol{\alpha}_i$ yields $\boldsymbol{\alpha}_i^* = \mathbf{s}_i^T\mathbf{x}/\|\mathbf{s}_i\|_2^2 = \mathbf{s}_i^T\mathbf{x}$ which, injected into Likelihood Ratio (2), leads to

$$\text{Max Test} : T_{max}(\mathbf{x}, \mathbf{S}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma, \text{with } T_{max}(\mathbf{x}, \mathbf{S}) = \max_i |\mathbf{s}_i^T\mathbf{x}|. \tag{3}$$

In (3), superscript ^{T} denotes the transposition. This test is called Max Test in [3].

In the "one among many" detection problem set by (1), we expect that the probability of false alarm (P_{FA}) increases more than the probability of detection (P_{Det}) as the size L of the alternatives increases (assuming that the true alternative is represented in $\mathbf{S}$). Thus we expect that the power of the test degrades with increasing L. Assessing this problem analytically is difficult because it involves the distribution of the maximum of correlated variables. We illustrate this effect below numerically on a library of spectral lines (100 of which are displayed in Fig 3 (a)).

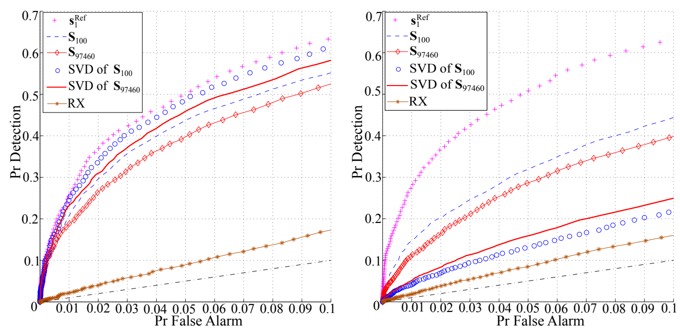
Starting from a library containing only the active spectrum under $\mathcal{H}_1$ (say, $\mathbf{s}_1$), we build, by adding other spectral lines to $\mathbf{S}$, two larger libraries of respectively 100 and 97460 lines ($\mathbf{S}_{100}$ and $\mathbf{S}_{97460}$), and we compare the ROC curves obtained by performing test (3) with $\mathbf{S} = \mathbf{s}_1, \mathbf{S}_{100}$ and $\mathbf{S}_{97460}$. We show the results of this experiment for two instances of atoms $\mathbf{s}_1$ for $P_{FA} \leq 0.1$. In the first case (Fig.1 (a)), the atoms added to $\mathbf{s}_1$ in $\mathbf{S}_{100}$ and $\mathbf{S}_{97460}$ have forms that are similar to $\mathbf{s}_1$ (i.e., the $\mathbf{s}_j$, $j = 2, \ldots, L$ are highly correlated to $\mathbf{s}_1$). In the second case (Fig.1 (b)), the considered atom $\mathbf{s}_1$ presents lower correlation with the other atoms of the library.

In both cases (a) and (b), we see that the best detection performance is obtained as expected for the case $\mathbf{S} = \mathbf{s}_1$ (violet crosses), and that the detection power drops as the dimension of $\mathbf{S}$ increases ($\mathbf{S}_{100}$: blue dashes, $\mathbf{S}_{97460}$: red diamonds). This detection loss is much more important in case (b), where $\mathbf{s}_1$ is dissimilar to the other columns of $\mathbf{S}$.

Fig.1 also investigates the performance of a simple reduced dimension test of the form $|\mathbf{x}^T\mathbf{d}| \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma$ (see Sec.3), where $\mathbf{d}$ corresponds to the eigenvector associated to the largest singular value of $\mathbf{S}$, for $\mathbf{S} = \mathbf{S}_{100}$ and $\mathbf{S} = \mathbf{S}_{97460}$. In case (a), we see that this detection test of reduced dimension has larger power compared to the GLRT test (3) corresponding to model (1) with $\mathbf{S} = \mathbf{S}_{100}$ and $\mathbf{S} = \mathbf{S}_{97460}$ (compare the blue dashes to the blue circles, and the red diamonds to the red line). In case (b), the situation changes drastically, with a large power loss when the dimension is reduced.

Note finally that we have also considered the RX [19] test (which amounts here to $\|\mathbf{x}\|_2^2 \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma$, that is, an energy detector) because this test is a benchmark for AD methods [12, 13]. The loss in power of RX w.r.t. the other tests illustrates the penality incurred by not using *a priori* knowledge on $\mathcal{H}_1$.

This experiment shows that dimension reduction may result in a large power loss, or in other words, that some alternatives may become undetectable with low dimension tests. This poses the problem of designing a test of reduced dimension which is *robust* in the sense that large losses for such alternatives are minimized.



(a) (b)

Fig. 1. Compared ROC curves shown for $P_{FA} \leq 0.1$ for two instances of libraries $\mathbf{S} = \mathbf{S}_{100}$ and $\mathbf{S} = \mathbf{S}_{97460}$, in two different cases: (a) $\mathbf{s}_1$ under $\mathcal{H}_1$ is well correlated to the other atoms of $\mathbf{S}$ and (b): $\mathbf{s}_1$ is not well correlated to the other atoms.

3. MINIMAXITY AND DIMENSION REDUCTION

Instead of model (1), we consider now the following model:

$$\begin{cases} \mathcal{H}_0 & : \mathbf{x} = \mathbf{n}, \qquad\qquad \mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \\ \mathcal{H}_1 & : \mathbf{x} = \mathbf{D}\boldsymbol{\beta} + \mathbf{n}, \end{cases} \tag{4}$$

where $\mathbf{D} \in \mathbb{R}^{N \times r}$, $r < N$, and $\boldsymbol{\beta} \in \mathbb{R}^r$ is not sparse, as the dimensionality reduction is conducted by $\mathbf{D}$. The GLR test for (4) is :

$$T_{GLR} : \max_{\boldsymbol{\beta}} \frac{P_r(\mathbf{x}|\mathbf{D}\boldsymbol{\beta})}{P_r(\mathbf{x}|\mathbf{0})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \xi' \tag{5}$$

where under $\mathcal{H}_1$

$$P_r\left(\mathbf{x}\,|\mathbf{D}\boldsymbol{\beta}\right) = \frac{1}{(2\pi)^{N/2}} \exp\left(-\frac{1}{2}\|\mathbf{x} - \mathbf{D}\boldsymbol{\beta}\|_2^2\right). \tag{6}$$

Maximizing (6) on $\boldsymbol{\beta}$ yields the optimal $\boldsymbol{\beta}^*$, which, by using $\Pi_{\mathbf{D}}$, the orthogonal projection of $\mathbf{x}$ on Im($\mathbf{D}$), reads:

$$\boldsymbol{\beta}^* = \mathbf{D}(\mathbf{D}^T\mathbf{D})^{-1}\,\mathbf{D}^T\mathbf{x} = \Pi_{\mathbf{D}}\mathbf{x}. \tag{7}$$

Substituting $\boldsymbol{\beta}^*$ in the GLR test (5) and taking the logarithm yields the test

$$T_{GLR}(\mathbf{x}, \mathbf{D}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \xi, \tag{8}$$

with $\xi = 2\ln \xi'$ and

$$T_{GLR}(\mathbf{x},\mathbf{D}) = \mathbf{x}^T\mathbf{D}(\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T\mathbf{x} = \mathbf{x}^T\,\Pi_\mathbf{D}\,\mathbf{x} = \|\Pi_\mathbf{D}\mathbf{x}\|_2^2\,. \tag{9}$$

For this test, we seek to optimize $\mathbf{D}$ in a minimax approach. We first investigate the behavior of P_{FA} and P_{Det} w.r.t. $\mathbf{D}$. Assume that under $\mathcal{H}_1$, the target signal is $\mathbf{S} = \mathbf{s}_\ell\boldsymbol{\alpha}_\ell$, then by (1) $\mathbf{x} \sim \mathcal{N}(\mathbf{s}_\ell\boldsymbol{\alpha}_\ell, \mathbf{I})$. Since the projection matrix $\Pi_\mathbf{D}$ is idempotent ($\Pi_\mathbf{D}\Pi_\mathbf{D} = \Pi_\mathbf{D}$), symmetric ($\Pi_\mathbf{D}^T = \Pi_\mathbf{D}$) and of rank r, the distributions of the test statistics $\mathbf{x}^T\,\Pi_\mathbf{D}\,\mathbf{x}$ in (8) - (9) under $\mathcal{H}_0$ and $\mathcal{H}_1$ are known (p.291 of [20]). These distributions can be written as Chi-squared distributions χ_r^2 with r degrees of freedom: $T_{GLR}(\mathbf{x}|\mathcal{H}_0) = \mathbf{n}^T\Pi_\mathbf{D}\mathbf{n} \sim \chi_r^2$ and $T_{GLR}(\mathbf{x}|\mathcal{H}_1) = (\mathbf{s}_\ell\boldsymbol{\alpha}_\ell+\mathbf{n})^T\Pi_\mathbf{D}(\mathbf{s}_\ell\boldsymbol{\alpha}_\ell+\mathbf{n}) \sim \chi^2_{r,\lambda^2}$, where $\lambda^2 = \|\Pi_\mathbf{D}\ \mathbf{s}_\ell\boldsymbol{\alpha}_\ell\|_2^2$ is the non centrality parameter (λ^2 obviously depends on the signal $\mathbf{s}_\ell$ under $\mathcal{H}_1$ and on $\mathbf{D}$). Thus if we denote by $\Phi_{\chi_r^2}$ and $\Phi_{\chi^2_{r,\lambda^2}}$ the corresponding cumulative distribution functions, and by Q the generalized Marcum Q-function [21],

$$\begin{aligned} P_{FA} &= P_r\ (T_{GLR} > \xi\ |\mathcal{H}_0) = 1 - \Phi_{\chi_r^2}(\xi),\\ P_{Det} &= P_r(T_{GLR} > \xi|\mathcal{H}_1) = 1 - \Phi_{\chi^2_{r,\lambda^2}}(\xi) = Q_{\frac{r}{2}}(\lambda, \sqrt{\xi}). \end{aligned} \tag{10}$$

The P_{FA} is independent of $\mathbf{D}$ ($\forall\mathbf{D}$: rank($\mathbf{D}$) = r). Since $Q_{\frac{r}{2}}(\lambda, \sqrt{\xi})$ is an increasing function of λ (Th.1 of [22]), maximizing P_{Det} is equivalent to maximizing λ. To address the problem posed at the end of Sec.2, we formulate our optimization criterion as determining the low dimension dictionary that maximizes the minimal P_{Det}, that is, that maximizes the $P_{Det}(\mathbf{s}_i,\mathbf{D})$ occurring for the worst case of $\{\mathbf{s}_i,\ i = 1,\ldots,L\}$:

$$\max_{\mathbf{D}:\|\mathbf{d}_i\|_2^2=1}\ \min_i P_{Det}(\mathbf{s}_i,\mathbf{D}) \Leftrightarrow \max_{\mathbf{D}:\|\mathbf{d}_i\|_2^2=1}\ \min_i \|\Pi_\mathbf{D}\ \mathbf{s}_i\|_2^2 \tag{11}$$

where $\mathbf{D} = [\mathbf{d}_1,\ldots,\mathbf{d}_r] \in \mathbb{R}^{N\times r}$ is the reduced dimension dictionary with r ($r < N < L$) unit-norm columns $\mathbf{d}_i$.

4. RESOLUTION FOR A SINGLE ATOM

In the scope of this paper, we focus on the simplest case $r = 1$, for which $\mathbf{D}$ has only one atom: $\mathbf{D} = \mathbf{d}$. In this case, $\Pi_\mathbf{D} = \mathbf{d}\mathbf{d}^T$ and $\lambda^2(\mathbf{s}_i) = \boldsymbol{\alpha}_i^2(\mathbf{d}^T\mathbf{s}_i)^2$. Therefore, we seek the solution(s) of:

$$\max_{\mathbf{d}:\|\mathbf{d}\|_2=1}\ \min_i (\mathbf{d}^T\mathbf{s}_i)^2 \tag{12}$$

which is non convex because of the non convex constraint $\|\mathbf{d}\|_2 = 1$.

4.1. Solution of the Optimization Problem

First, note that problem (12) is invariant by rotation: if $\mathbf{d}^*$ is a solution, for any rotation matrix $\mathbf{Q}$, $\mathbf{Q}\mathbf{d}^*$ is also solution of (12) where all the $\mathbf{s}_i$ are replaced by $\mathbf{Q}\mathbf{s}_i$. We can verify this easily by replacing in (12) $\mathbf{d}$ and $\mathbf{s}_i$ by $\mathbf{Q}\mathbf{d}^*$ and $\mathbf{Q}\mathbf{s}_i$ respectively, where $\mathbf{Q}^T = \mathbf{Q}^{-1}$.

We will assume in the sequel that $\{\mathbf{s}_i,\ i = 1,2,\ldots,L\}$ are in the interior of a cone defined as an arbitrary rotation of the positive orthant $\mathbb{R}_+^N$. By the rotation argument above, we can indeed assume without loss of generality that all the $\mathbf{s}_i$ are in the interior of $\mathbb{R}_+^N$. Proposition 1 proves that (12) can be solved using a standard quadratic programming (QP) solver such as CVX [23].

Proposition 1. *The solutions of (12) are* $\{\mathbf{d}^*, -\mathbf{d}^*\}$*, where* $\mathbf{d}^* \in \mathbb{R}_+^N$ *is the solution of the QP:*

$$\begin{array}{ll} \textit{minimize} & -t \\ \textit{subject to} & t - \mathbf{d}^T\mathbf{s}_i \le 0,\ i = \{1,\ldots,L\} \\ & \|\mathbf{d}\|_2^2 \le 1 \end{array} \tag{13}$$

We only give the idea of the proof here for lack of space. The proposition relies on two Lemmas. The first Lemma shows that if $\mathbf{d}^*$ is solution of (12), $-\mathbf{d}^*$ is also solution, and $\mathbf{d}^*$ or $-\mathbf{d}^* \in \mathbb{R}_+^N$. To show that $\mathbf{d}^*$ is the solution of the QP (13), we consider now the problem (12) on $\mathbb{R}_+^N$ written in its epigraph form [24]. The second Lemma shows that the solution of this non convex problem can be obtained by the QP above.

4.2. Connection of Problem (13) with One-Class Classifiers

The proposed optimization problem can be represented geometrically on a unit-norm spherical boundary denoted by Σ_1, where all $\mathbf{s}_i \subset \Sigma_1$. This problem can be interpreted as finding the vector $\mathbf{d}$ on the unit-sphere that minimizes the largest angle θ_i or equivalently, finding the circle $\mathcal{C}$ of minimum radius R that contains all the points $\mathbf{s}_i$ (Fig.2). Furthermore, the proposed optimization criterion can also be solved through One-Class Classifiers of SVM type. SVM is an approach used to solve classification and regression problems, mainly for machine learning applications introduced by Vapnik *et al.* in 1992 [25]. Originally, it was a binary classifier, and later was developed to multi-class.

Schölkopf [26] proposed a one-class SVM (ν-SVC) that seeks an optimal hyperplane separating maximally the origin and the training samples (the class) in feature space. Suppose $\mathcal{P}_d$ is a plane perpendicular to $\mathbf{d}$ and ρ is the distance of $\mathcal{P}_d$ to the origin. We search the farthest plane which discards all $\mathbf{s}_i$ aside, as depicted in Fig.2. The function $\mathbf{y} := \mathbf{d}^T\mathbf{s}_i - \rho = 0$ denotes the decision function of the classification problem associated to $\mathcal{P}_d$.

Another approach that describes our optimization criterion is the Support Vector Data Description (SVDD). This method finds the minimum volume of a closed boundary sphere Σ, of center $\boldsymbol{a}$ and radius R [27]. R is minimized such that Σ contains all points $\mathbf{s}_i$ similarly to the smallest circle approach described above (see Fig.2). If the $\mathbf{s}_i$ are normalized $\|\mathbf{s}_i\|_2^2 = 1$, SVDD is equivalent to ν-SVC.

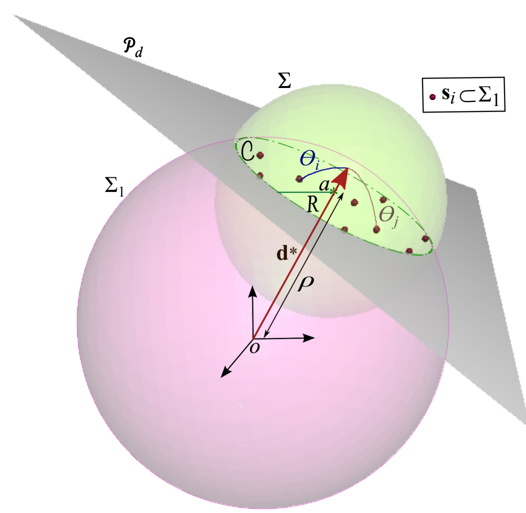


Fig. 2. Geometrical view of One-Class Classifiers for the considered minimax detection problem (12)-(13), with Σ_1 denotes a unit-sphere: the problem is equivalent to minimizing the largest angle θ_i between $\mathbf{d}$ and $\mathbf{s}_i$, to finding the circle $\mathcal{C}$ of minimum radius R that contains all the points $\mathbf{s}_i$, to maximizing the distance ρ of $\mathcal{P}_d$ to the origin, or to minimizing the radius R of the sphere Σ that contains all the data points $\mathbf{s}_i$ (Σ admits $\mathcal{C}$ as a great circle).

5. APPLICATION TO ASTROPHYSICAL SPECTRA

We present here numerical results obtained for the detection of spectral lines in HSI data that will be acquired by the integral field spectrograph named MUSE (Multi Unit Spectroscopic Explorer). This instrument will deliver data cubes made of 300×300 pixels at 3600 wavelengths. Highly realistic data were simulated and provided by the MUSE consortium.

An important challenge in such data is to detect faint line spectra emitted by distant galaxies (Lyman Alpha emitters). Such sources are expected to be relatively rare (at most one line in each spectrum), and have very specific signatures that can be simulated using astrophysical models (hence a large library), leading to the "one among many" data model (1). In our case, astrophysical simulations lead to a spectral library $\mathbf{S} \in \mathbb{R}^{3600\times 9746}$. Because the lines can essentially be centered at *any* of the 3600 wavelength channels, the effective size of $\mathbf{S}$ is $3600 \times (9746 \times 3600) = 1.2 \times 10^{11}$, and this should be tested over each of the 90000 spectra of the data cube. Using the full library leads to a test of prohibitive complexity. The minimax approach is interesting here to detect better the most marginal (potentially most interesting) spectral forms.

For the purpose of computing ROC curves over the whole library, we make two simplifications. First, we consider atoms of smaller dimension ($N = 100$, see Fig.3 (a)). Then, we train the considered test dictionaries for the set of the 9746 atoms, centered at wavelength channel 50 (Fig.3 (b)), and we do not consider possible translations under $\mathcal{H}_1$. These simplifications have no impact on the presented results but allow an exhaustive statistical analysis. Full size detection tests are described in [18].

In Fig.4 and Table 1, we compare the AUC obtained for several alternatives (100 of them are shown in Fig.4, while the Table addresses the 9746 alternatives) and for 5 dictionaries. The first dictionary is $\mathbf{D} = \mathbf{s}_i$, where $\mathbf{s}_i$ is the atom under $\mathcal{H}_1$. This is indeed the reference. The second dictionary is $\mathbf{S} = \mathbf{S}_{9746}$, that is, the full library. The third one is the 1-dimension dictionary obtained by the proposed minimax approach ($\mathbf{D} = \mathbf{d}^*$, Sec.4). The last two employ the K-SVD algorithm [28], a dictionary learning technique that has been proven useful for sparse signal modeling. We consider here the K-SVD dictionaries for $r = 1$ (leading to $\mathbf{D} = \mathbf{d}_1^{\text{(k)svd}}$, in which case this is equivalent to the SVD approach discussed in Sec.2), and $r = 7$ ($\mathbf{D} = \mathbf{D}_7^{\text{ksvd}}$). Fig.3 (b) shows the $\mathbf{d}_1^{\text{(k)svd}}$ and $\mathbf{d}^*$.

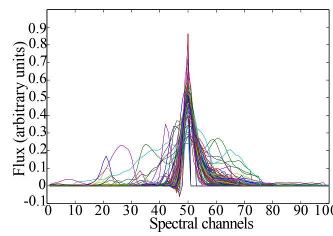
In Fig.4, we can see that $\mathbf{d}_1^{\text{(k)svd}}$'s (black crosses) performances are nearly as good as the reference (green dots) for most atoms. However, for some alternatives (for instance $i = 18$ and $i = 91$) the performances of $\mathbf{d}_1^{\text{(k)svd}}$ and $\mathbf{D}_7^{\text{ksvd}}$ (cyan line) drop drastically. This is due to the high model error of some $\mathbf{s}_i$ which are not well correlated with the atoms of the K-SVD dictionaries. The corresponding spectral lines present a large probability to remain undetected. In contrast, the proposed minimax approach $\mathbf{d}^*$ (blue dash-dots) maintains a more stable detection power in such cases.

The full library $\mathbf{S}_{9746}$ (red stars) represents inferior overall performances to the K-SVD dictionaries even though it contains the active alternative. The reason is that that the atoms of the K-SVD dictionaries represent well most alternatives (*cf* situation depicted in Fig.1(a)), but not all of them (as $i = 18$ and $i = 91$, situation Fig.1(b)).

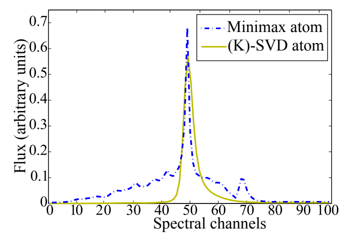
Turning now to the results over the whole set of 9746 alternatives, Table 1 shows that the highest minimum AUC is obtained by the full library $\mathbf{S}_{9746}$ (but in practice, this test is not implementable on full size data), followed by the proposed dictionary $\mathbf{d}^*$. The $\mathbf{d}_1^{\text{(k)svd}}$ dictionary has the lowest minimum AUC because some atoms are discarded by this test. In fact we computed that there are no alternative for which the loss in AUC w.r.t. the reference (row 1 in Table 1) is superior to 14% for the minimax approach, while this happens for 15 alternatives for K-SVD1 and the loss goes up to 26% in some cases. Of course, the minimax approach has a lower average detection power than K-SVD because a trade-off has to be made between average versus minimax performance.

Dictionary used in (3)	Ranking Criterion	
	Min AUC (minimax)	Average AUC
Atom under $\mathcal{H}_1$	Ref : 0.84	Ref : 0.85
$\mathbf{S}_{9746}$	1^{st} : 0.73	3^{rd} : 0.80
$\mathbf{d}^*$	2^{nd} : 0.71	4^{th} : 0.78
$\mathbf{d}_1^{\text{(k)svd}}$	4^{th} : 0.59	1^{st} : 0.84
$\mathbf{D}_7^{\text{ksvd}}$	3^{rd} : 0.66	2^{nd} : 0.83

Table 1. Results over the whole set of alternatives



(a) 100 spectral lines



(b) Atoms: $\mathbf{d}^*$ and $\mathbf{d}_1^{\text{(k)svd}}$

Fig. 3. The spectral lines and the trained dictionaries.

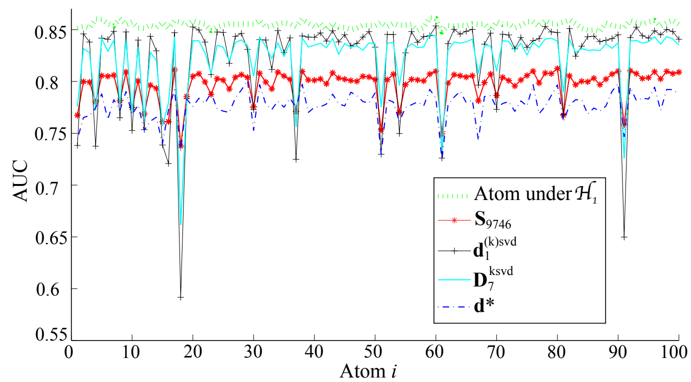


Fig. 4. AUC for the first 100 atoms.

6. CONCLUSION

We have shown that the proposed minimax approach fulfills the desired objective of maximizing the worst detection performance. In the 1-dimensional case, the corresponding criterion is non convex, but we showed that optimal atom $\mathbf{d}^*$ is obtained as the solution of the QP problem (13), where $\mathbf{d}^*$ belongs to the positive orthant. Moreover, we found connections between this approach and the SVM One-Class Classifiers and we provided the corresponding geometrical illustration. Further investigations are carried out to optimize the minimax approach over r-dimensional dictionaries with $r > 1$.

7. REFERENCES

[1] B. Sklar, *Digital Communications: Fundamentals & Applications*, Prentice Hall, 2011.

[2] L. L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Trans. Signal Process.*, vol. 42, no. 8, pp. 2146–2457, 1994.

[3] 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(5), pp. 2533–2556, 2010.

[4] D. Donoho and J. Jin, "Higher criticism for detecting sparse heterogeneous mixtures," *The Annals of Statistics*, vol. 32(3), pp. 962–994, 2004.

[5] A. Berk *et al*, "MODTRAN5: a reformulated atmospheric band model with auxiliary species and practical multiple scattering options: update," in *Proc. SPIE 5806, Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XI*. SPIE, 2005, pp. 662–667.

[6] G. Healey and D. Slater, "Models and methods for automated material identification in hyperspectral imagery acquired under unknown illumination and atmospheric conditions," *IEEE Trans. On Geoscience and Remote Sensing.*, vol. 37, no. 6, pp. 2706–2717, November 1999.

[7] S. Bourguignon, D. Mary, and E. Slezak, "Processing MUSE hyper spectral data: Denoising deconvolution and detection of astrophysical sources," *Statistical Methodology*, April 2011.

[8] 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.

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

[10] S. Paris, D. Mary, and A. Ferrari, "PDR and LRMAP detection tests applied to massive hyper spectral data," in *Computation Advances in Multi-Sensor Adaptive Processing (CAMSAP)*. IEEE, 2011, pp. 93–96.

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

[12] S. Matteoli, F. Carnesecchi, M. Diani, G. Corsini, and L. Chiarantini, "Comparative analysis of hyperspectral anomaly detection strategies on a new high spatial and spectral resolution data set," in *Proc. SPIE 6748, Image and Signal Processing for Remote Sensing XIII*, 2007, pp. 67480E.1–67480E.11.

[13] 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.

[14] 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.

[15] S. Kraut, L. L. Scharf, and L.T. McWhorter, "Adaptive subspace detectors," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 1–16, January 2001.

[16] E.L. Lehmann and J.P. Romano, *Testing statistical Hypotheses*, Springer, 2005.

[17] S.A. Kassam and H.V. Poor, "Robust techniques for signal processing: A survey," in *Proc. of IEEE*. IEEE, 1985, vol. 73, pp. 433–481.

[18] S. Paris, R.F.R. Suleiman, D. Mary, and A. Ferrari, "Constrained likelihood ratios for detecting sparse signals in highly noisy 3D data," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2013, Accepted.

[19] I.S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. on Acoustics, Speech, and Signal Process.*, vol. 38, pp. 1760–1770, October 1990.

[20] A. Sen and M. Srivastava, *Regression Analysis: Theory, Methods and Applications*, Springer-Verlag New York Inc., 1990.

[21] A.H. Nutall, "Some integrals involving the (Q sub M) - function," Tech. Rep. AD-779846, Naval Underwater Systems Center New London, Connecticut, May 1974.

[22] Y. Sun, A. Barics, and S. Zhou, "On the mono city, log-concavity and tight bounds of the generalized Marcum and Nutall Q-functions," *IEEE Trans. on Information Theory*, vol. 56, no. 3, pp. 1166–1186, March 2010.

[23] CVX Research Inc., "CVX: Matlab software for disciplined convex programming, version 2.0 beta," `http://cvxr.com/cvx`, September 2012.

[24] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[25] B.E. Boser, I.M. Guyon, and V.N. Vapnik, "A training algorithm for optimal margin classifiers," in *COLT '92*. ACM, 1992, pp. 144–152.

[26] B. Schölkopf, *Support Vector Learning*, Ph.D. thesis, Technishen Universitat Berlin, 1997.

[27] D.M.J. Tax and R.P.W. Duin, "Support vector data description," *Machine Learning*, vol. 54, pp. 45–66, 2004.

[28] M. Aharon, M. Elad, and A.Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54(11), pp. 4311–4322, November 2006.