Détection de sources astrophysiques dans les données du spectrographe intégral de champ MUSE

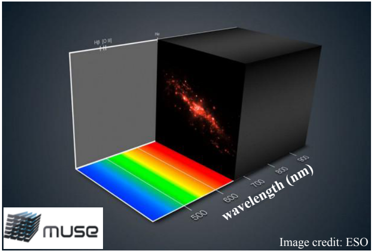
David Mary, André Ferrari and Raja Fazliza R.S.

Université de Nice Sophia Antipolis, France

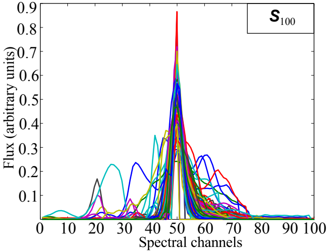
Porquerolles, 15 mai 2014

Hyperspectral Imaging with MUSE: detection of galaxies and emission lines

The spectrograph MUSE delivers data cube of 300×300 *pixels at* 3600 *wavelength channels. Lyman-*α *emitters are very distant and faint; contain essentially one emission line, whose profile varies with the considered object.*



MUSE (Multi Unit Spectroscopic Explorer) data cube



100 possible alternatives under $\mathcal{H}_1$

Detect one target signature $\boldsymbol{s}_i \in \mathbb{R}^{N\times 1}$ which belongs to a large known library $\boldsymbol{S} = [\boldsymbol{s}_1, \ldots, \boldsymbol{s}_L]$ ($\boldsymbol{S} \in \mathbb{R}^{N\times L}$, with $L \gg N$).

The amplitude and the alternative $\boldsymbol{s}_\ell$ activated under $\mathcal{H}_1$ are unknown.

Strategy for dimension reduction: Average vs Minimax performance

Standard approach when having a very large library is to test in subspaces of reduced dimensions (*e.g.*: sample mean, SVD, K-SVD, etc).

+ less computationally complex with similar average performances.

— might induce drop of performance for some targets.

Proposed approach

Learn a reduced size library with the objective of minimizing the maximum power loss.

Exact model and associated GLRT

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

$\boldsymbol{n} \in \mathbb{R}^{N\times 1}$: gaussian noise (covariance matrix is known and equal under $\mathcal{H}_0$ and $\mathcal{H}_1$),
$\boldsymbol{S} \in \mathbb{R}^{N\times L} = [\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_L]$ with $||\boldsymbol{s}_i||_2^2 = 1\ \forall i$, $L \gg N$, and $\boldsymbol{\alpha}$: unknown vector.

GLR test using $\boldsymbol{S}$ under constraint $||\boldsymbol{\alpha}||_0 = 1$:

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

Corresponding GLRT: *Max* test over *S*

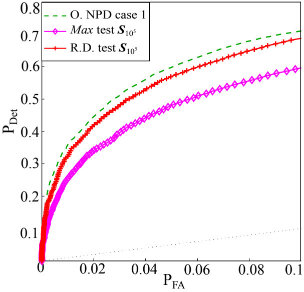
$$\Rightarrow T_{\text{Max}} : \max_{i=1,\ldots,L} |\boldsymbol{s}_i^\top \boldsymbol{x}| \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma\,.$$

Issues when testing with a reduced size library

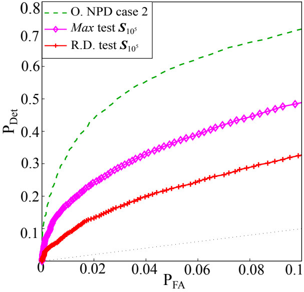
Assuming in all cases that $\boldsymbol{s}_\ell$ is the active alternative under $\mathcal{H}_1$:

- Oracle Neyman-Pearson detector (O. NPD): $|\boldsymbol{x}^\top \boldsymbol{s}_\ell| \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma$.
- A reduced one-dimension GLR test: $|\boldsymbol{x}^\top \boldsymbol{v}| \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma$, (*e.g.* $\boldsymbol{v}$: SVD of $\boldsymbol{S}$).

1) $\boldsymbol{s}_\ell$ similar shape to $\boldsymbol{v}$ ($\boldsymbol{s}_\ell^\top \boldsymbol{v} = 0.93$)



2) $\boldsymbol{s}_\ell$ dissimilar shape to $\boldsymbol{v}$ ($\boldsymbol{s}_\ell^\top \boldsymbol{v} = 0.50$)



R.D. test: advantageous w.r.t. *Max* test iff $\boldsymbol{s}_\ell$ is well correlated to the learned subspace.
$\Rightarrow$ possibility to devise robust (minimax) R.D. tests.

Model of reduced dimension and corresponding GLRT

Proposed R.D. model

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

$\boldsymbol{D} \in \mathbb{R}^{N \times K}$ with $||\boldsymbol{d}_j||_2^2 = 1$ is a low dimension ($K \ll L$) dictionary to be optimized.

Corresponding GLRT: *Max* test over *D*

$$T_{\boldsymbol{D}}(\boldsymbol{x}) = \max_{j=1,\dots,K} |\boldsymbol{d}_j^\top \boldsymbol{x}| \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \xi.$$

Aim: optimize ***D*** to maximize the worst-case detection performance.

Minimax optimization criterion

Optimal minimax dictionary: exact criterion

$$\boldsymbol{D}^* = \arg\max_{\boldsymbol{D}} \min_{i=1,\dots,L} \mathbb{P}\left(\max_{j=1,\dots,K} |\boldsymbol{d}_j^\top \boldsymbol{x}| > \xi | \mathcal{H}_1, \boldsymbol{s}_i\right)$$

$$\text{subject to} \begin{cases} \mathbb{P}\left(\max_{j=1,\dots,K} |\boldsymbol{d}_j^\top \boldsymbol{x}| > \xi | \mathcal{H}_0\right) \leq \mathrm{P}_{\mathrm{FA}_0}, \\ \|\boldsymbol{d}_j\|_2 = 1,\ j = \{1,\dots,K\}. \end{cases}$$

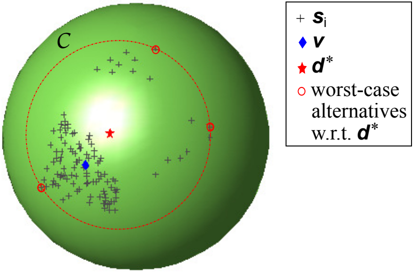
Solving $\boldsymbol{D}^* (K > 1)$ involves distributions of the max. of correlated variables $\boldsymbol{d}_j^\top \boldsymbol{x}$, $j = \{1,\dots,K\}$, which is an extremely intricate task.

Exact criterion for $K = 1$

$$\boldsymbol{d}^* = \arg\max_{\boldsymbol{d}:\|\boldsymbol{d}\|_2=1} \min_{i=1,\dots,L} |\boldsymbol{d}^\top \boldsymbol{s}_i|$$

Illustration of the minimax atom $\boldsymbol{d}^*$ $(K = 1)$

Proposition:
Assuming that $\boldsymbol{S} \in \mathbb{R}_+^N$, then $\boldsymbol{d}^* \in \mathbb{R}_+^N$ is the solution of a QP problem.



- On this example, $\boldsymbol{d}^*$ is held by three marginal alternatives (at the border of the smallest enclosed circle $\mathcal{C}$). Those "isolated" $\boldsymbol{s}_i$ induce the worst P_{Det}.
- $\boldsymbol{v}$ (SVD) represents well the most populated area of the alternatives.
- Representing $\boldsymbol{S}$ by a single atom may be insufficient w.r.t. the intrinsic diversity of $\boldsymbol{S}$.

Approximation criterion to optimize $\boldsymbol{D}$ for $K > 1$: deriving bounds

$$\mathrm{P}_{\mathrm{Det}}(\boldsymbol{s}_\ell, \boldsymbol{D}) = \mathbb{P}\left(\max_{j=1,\dots,K}(\boldsymbol{d}_j^\top \boldsymbol{x})^2 > \xi^2 | \mathcal{H}_1, \boldsymbol{s}_\ell, \boldsymbol{D}\right).$$

CDF of the max of continuous random variables $(X_1, \dots, X_N)$:
$F_{\max(X_1,\dots,X_N)}(t) \leq \min\left(F_1(t), \dots, F_N(t)\right).$

$$\mathrm{P}_{\mathrm{Det}}(\boldsymbol{s}_\ell, \boldsymbol{D}) \geq \max_{j=1,\dots,K} Q_{\frac{1}{2}}(|\boldsymbol{d}_j^\top \boldsymbol{s}_\ell|, \xi)$$

Approximation of the optimal dictionary

$$\boldsymbol{D}^* \approx \arg \max_{\boldsymbol{D}:\|\boldsymbol{d}_j\|_2=1} \rho^{(K)}(\boldsymbol{D})$$

where $\rho^{(K)}(\boldsymbol{D}) = \min_{i=1,\dots,L} \max_{j=1,\dots,K} |\boldsymbol{d}_j{}^\top \boldsymbol{s}_i|$ is the minimax correlation function.

Approximation criterion to optimize $\boldsymbol{D}$ for $K > 1$

$$\mathrm{P}_{\mathrm{FA}}(\boldsymbol{D}) = \mathbb{P}\left(\max_{j=1,\dots,K}(\boldsymbol{d}_j^\top \boldsymbol{x})^2 > \xi^2 | \mathcal{H}_0, \boldsymbol{D}\right).$$

CDF of multivariate normals (C. Khatri, 1968):

$$\mathbb{P}(|u_i| \leq c_i, i = 1, \dots, m) \geq \prod_{i=1}^{m} \mathbb{P}(|u_i| \leq c_i)$$

when $\boldsymbol{u} = (u_1, \dots, u_m)^\top$ is distributed as multivariate normal $\boldsymbol{u} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$.

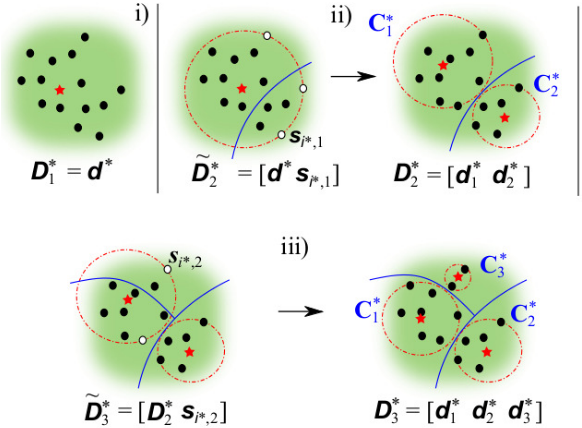
$$\mathrm{P}_{\mathrm{FA}}(\boldsymbol{D}) \leq 1 - \prod_{j=1}^{K} \mathbb{P}(|u_j| \leq \xi)$$

$$\mathrm{P}_{\mathrm{FA}}(\boldsymbol{D}) \leq 1 - \Phi_{\chi_1^2}^{K}(\xi^2).$$

If $\boldsymbol{D}$ is orthogonal, the upper bound is the exact $\mathrm{P}_{\mathrm{FA}}(\boldsymbol{D})$.

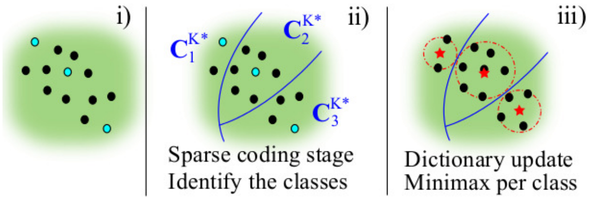
Ilustrations of the proposed learning methods to optimize *D*

- **Greedy minimax**: ($K = 3$)
  Heuristic optimization based on the approximation criterion. Dictionary update stage by 1-dimensional minimax optimization. Samples the distribution in a greedy manner to open new classes.



Ilustrations of the proposed learning methods to optimize *D*

- **K-minimax**: ($K = 3$)
  The dictionary update stage of the K-SVD algorithm [Aharon *et al.*, 2006] is replaced by the exact solution of the 1-dimensional minimax problem.



$\hookleftarrow$ ii)

Subspace learning of faces

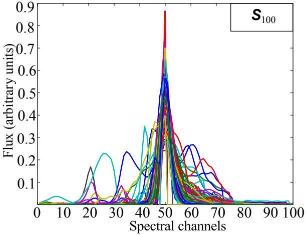
Library ***S***: 40 subjects in front position (possible alternatives under $\mathcal{H}_1$), taken from the ORL Database of Faces by AT&T Laboratories Cambridge.

K-SVD
($K = 3$)

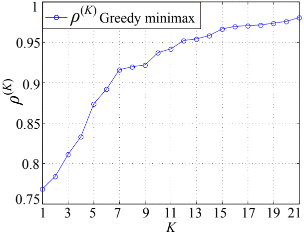
Greedy minimax
($K = 3$)

Minimax approach captures marginal features, (K)-SVD tends to represent “average face(s)”.

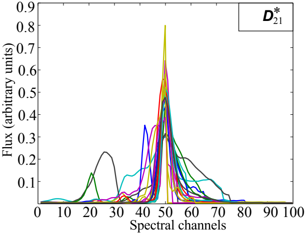
Subspace learning of 100 spectral profiles



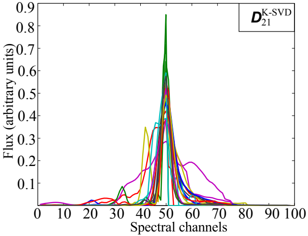
i. Spectral profiles, $\boldsymbol{S} \in \mathbb{R}^{100 \times 100}$



ii. Minimax correlation function, $\rho^{(K)}$

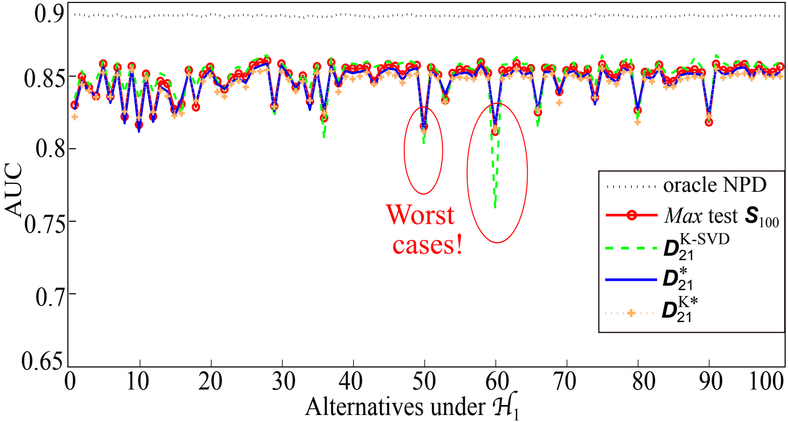


iv. $\boldsymbol{D}^*_{21}$ (minimax)



v. $\boldsymbol{D}^{\text{K-SVD}}_{21}$

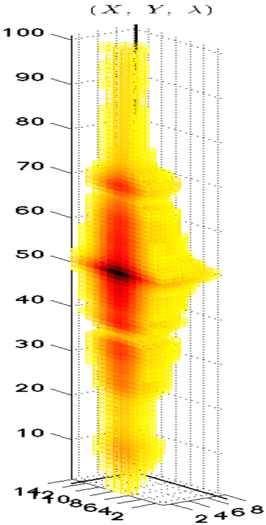
AUC (Area under curve) of the ROCs

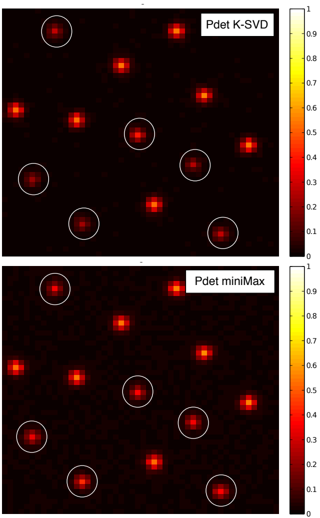


Performances ranking: Minimax vs Average

Dictionary	Ranking Criterion	
	min AUC	Average AUC
O. NPD	Ref : 0.886	Ref : 0.887
$\boldsymbol{S}_{100}$	1st : 0.813	3rd : 0.847
$\boldsymbol{d}^*$	2nd : 0.794	6th : 0.836
$\boldsymbol{v}$ (SVD)	4th : 0.699	1st : 0.863
$\boldsymbol{D}_{21}^{\text{K-SVD}}$	3rd : 0.764	2nd : 0.849
$\boldsymbol{D}_{21}^*$	1st : 0.812	4th : 0.845
$\boldsymbol{D}_{21}^{\text{K}*}$	1st : 0.813	5th : 0.843

MUSE simulation. 3D atoms, SNR = -11 dB, $P_{FA} = 0.01$





Conclusions

- Propose a detection test of reduced dimension where dictionary $\boldsymbol{D}$ is learned with the objective of maximizing the worst $\mathrm{P_{Det}}$.
- Case $K = 1$, exact solution to the minimax problem, but might be insufficient to represent well all the intrinsic diversity of $\boldsymbol{S}$.
- Case $K > 1$:
  - We derived proxies that are used to elaborate two algorithms for minimax detection.
  - better sampling of $\boldsymbol{S}$, thus improves the minimax performance over the case $K = 1$.

Perspectives

- Minimax modification of learning algorithm involving sparsity can be easily achieved by the use of use of $\boldsymbol{d}^*$ in the dictionary update stage.
- Find the optimal value of K, which depends on the intrinsic diversity of $\boldsymbol{S}$.

Thank you

Loss of performances of *Max* test as L increases

Assuming that $\boldsymbol{S} \in \mathbb{R}^{N\times L}$ is orthonormal ($\boldsymbol{s}_i^\top \boldsymbol{s}_j = \delta_{i,j}$) and that the active alternative under $\mathcal{H}_1$ is $\boldsymbol{s}_\ell$ (we assume that $\alpha = 1$), one can prove that:

$$\mathrm{P_{Det}}(\boldsymbol{s}_\ell) = 1 - \Phi_{\chi^2_{1,1}}\left(\Phi^{-1}_{\chi^2_1}\left((1-\mathrm{P_{FA}})^{\frac{1}{L}}\right)\right)\,(1-\mathrm{P_{FA}})^{\frac{L-1}{L}}.$$

(Φ_ν is the cumulative distribution function of ν).

- As L grows, the *Max* test correlates increasingly many alternatives to the data, with only marginal improvement of $\mathrm{P_{Det}}$ but substantially increased of $\mathrm{P_{FA}}$.
- Asymptotically, the GLR has no power in this setting.

$$\lim_{L\to+\infty} \Phi_{\chi^2_{1,1}}\left(\Phi^{-1}_{\chi^2_1}\left((1-\mathrm{P_{FA}})^{\frac{1}{L}}\right)\right) = 1,$$

$$\lim_{L\to+\infty} \mathrm{P_{Det}}(\boldsymbol{s}_\ell) = \mathrm{P_{FA}},\ \forall \ell : 1,\dots,L.$$