

Parcimonie, apprentissage et classification pour une approche minimax en détection

Raja Fazliza RAJA SULEIMAN, David MARY, André 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
{raja.fazliza, david.mary, andre.ferrari}@unice.fr

Résumé – Cet article propose, étudie et compare plusieurs tests de détection minimax. Ces tests allient une approche de type GLR (*Generalized Likelihood Ratio*) à une réduction de dimension basée sur des modèles parcimonieux. L’analyse géométrique et statistique des performances de ces tests nous conduit à proposer des méthodes d’apprentissage de dictionnaires minimax. Ces méthodes sont comparées par simulations numériques à d’autres méthodes minimax, dont la partie classification est issue de la littérature.

Abstract – This article proposes, studies and compares several minimax detection tests. These tests rely on a GLR (*Generalized Likelihood Ratio*) approach allied to sparse models. The geometrical and statistical analysis of the performances of the tests leads us to propose minimax dictionary learning algorithms. These algorithms are compared to other minimax methods whose classification part is taken from the literature.

1 Introduction

Nous considérons dans cet article le problème de détection suivant. Sous \mathcal{H}_1 est activé un signal $\mathbf{s}_i \in \mathbb{R}^{N \times 1}$ appartenant à un dictionnaire $\mathbf{S} = [\mathbf{s}_1, \dots, \mathbf{s}_N]$ de *grande dimension* ($\mathbf{S} \in \mathbb{R}^{N \times L}$, avec $L \gg N$). L’amplitude et l’indice i du signal activé sous \mathcal{H}_1 sont inconnus. Ce type de problème se produit par exemple en imagerie hyperspectrale. Dans de nombreux cas, l’ensemble des signatures spectrales possibles est grand en raison de la diversité naturelle des éléments à détecter (minéraux, cibles militaires, etc...), et/ou parce que les observations font intervenir des perturbations systématiques, connues et multiples [1, 2]. Dans ce cadre, les signatures cibles ne peuvent évidemment pas être optimisées pour la détection, et des tests impliquant la totalité des alternatives peuvent présenter un coût de calcul prohibitif.

Une démarche standard pour la détection est d’opérer à une réduction de dimension, en focalisant le test sur un ou des sous-espaces des signaux cibles. Ce genre d’approches est efficace en moyenne mais, comme nous l’illustrerons, il tend à produire des tests dont la puissance s’effondre pour certaines alternatives par rapport au cas où seule l’alternative activée sous \mathcal{H}_1 est testée – ce qui peut être fort dommageable dans certaines applications (médicales ou militaires par exemple). Nous nous intéressons dans cette étude à des tests d’hypothèses de dimension réduite visant à minimiser la plus grande perte de puissance sous \mathcal{H}_1 , c’est-à-dire à des tests de type *minimax*. Le problème peut aussi être formulé comme un *maximin*, où on cherche à maximiser la plus faible puissance sous \mathcal{H}_1 .

La réduction de dimension envisagée ici repose sur l’ex-

ploitation de modèles parcimonieux. Une première approche dans ce cadre, développée dans [3], fait intervenir l’optimisation d’un dictionnaire de dimension réduite $r \ll N$. Dans le cas particulier où $r = 1$, nous avons montré que la solution du problème minimax peut être obtenue en résolvant un problème d’optimisation convexe. Comme nous allons l’illustrer, une telle réduction de dimension peut cependant s’avérer excessive, le cas $r > 1$ pouvant permettre d’obtenir une puissance minimax (et moyenne) supérieure à celle du cas $r = 1$.

Le cas $r > 1$ mène à plusieurs difficultés que nous soulignons dans la Sec. 2. Celles-ci sont liées en partie au problème d’optimisation qui n’est pas convexe. Nous proposons et étudions en Sec. 3 une variante du test considéré dans [3], qui est équivalente à ce dernier pour $r = 1$. Pour le cas $r > 1$, nous proposons des méthodes d’apprentissage adaptées (Sec. 4) dont nous montrons en Sec. 5 l’aptitude à augmenter la puissance minimax (et moyenne) par rapport au cas $r = 1$. Ces méthodes sont comparées par simulations numériques à d’autres méthodes minimax dont la partie classification est issue de la littérature (moVMF [8] et Spherical K-Means [7]).

2 Approche minimax et parcimonie

Le problème de détection présenté en introduction mène au modèle d’hypothèses composites suivant :

$$\begin{cases} \mathcal{H}_0 & : \mathbf{x} = \mathbf{n}, & \mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_N) \\ \mathcal{H}_1 & : \mathbf{x} = \mathbf{S}\boldsymbol{\alpha} + \mathbf{n}, & \|\boldsymbol{\alpha}\|_0 = 1 \end{cases} \quad (M_1)$$

où \mathbf{x} et $\mathbf{n} \in \mathbb{R}^{N \times 1}$. Sans perte de généralité, les colonnes de \mathbf{S} sont supposées normalisées ($\|\mathbf{s}_i\|_2 = 1, \forall i = 1, \dots, L$). La contrainte sur la pseudonorme $\ell_0 = 1$ exprime le fait que sous \mathcal{H}_1 , seulement un signal \mathbf{s}_i d’amplitude α_i est activé. Une ap-

*Ce travail est financé par le projet Hypanema : ANR-12-BS03-0003.

proche standard en détection est de suivre le GLR, qui pour (M_1) donne le test : $\max_{i=1,\dots,L} |\mathbf{s}_i^\top \mathbf{x}| \geq_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma$, où γ est un seuil et l'exposant \top dénote la transposition [3]. Les résultats de [3] montrent que ce modèle peut présenter deux inconvénients quand $L \gg N$. Le premier est que même si le signal activé sous \mathcal{H}_1 fait bien partie de la série de filtres adaptés réalisés par le test, le grand nombre d'alternatives testées s'accompagne d'un taux élevé de fausses alarmes, particulièrement pour les alternatives qui sont les plus faiblement corrélées aux autres. Le deuxième est que ce test peut présenter un coût de calcul trop important pour être implémenté en pratique. Ces observations nous ont conduit au modèle de dimension réduite suivant :

$$\begin{cases} \mathcal{H}_0 & : \mathbf{x} = \mathbf{n}, & \mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \\ \mathcal{H}_1 & : \mathbf{x} = \tilde{\mathbf{D}}\boldsymbol{\beta} + \mathbf{n} \end{cases} \quad (M_2)$$

avec $\tilde{\mathbf{D}} \in \mathbb{R}^{N \times r}$ un dictionnaire optimisé de dimension réduite $r \ll N$ et $\boldsymbol{\beta}$ non parcimonieux. Maximiser la probabilité de détection minimale est alors équivalent à rechercher un dictionnaire $\tilde{\mathbf{D}}^* = [\tilde{\mathbf{d}}_1^*, \dots, \tilde{\mathbf{d}}_r^*]$ défini par :

$$\tilde{\mathbf{D}}^* = \arg \max_{\tilde{\mathbf{D}}: \|\tilde{\mathbf{d}}_i\|_2=1} \min_i \text{P}_{\text{Det}}(\mathbf{s}_i, \tilde{\mathbf{D}}) = \arg \max_{\tilde{\mathbf{D}}: \|\tilde{\mathbf{d}}_i\|_2=1} \min_i \|\Pi_{\tilde{\mathbf{D}}} \mathbf{s}_i\|_2^2 \quad (1)$$

où $\Pi_{\tilde{\mathbf{D}}}$ est la matrice de projection orthogonale sur $\tilde{\mathbf{D}}$. Dans le cas où $r = 1$ ($\tilde{\mathbf{D}} = \mathbf{d}$), notre problème minimax (formulé comme un maximin ici) devient

$$\mathbf{d}^* = \arg \max_{\mathbf{d}: \|\mathbf{d}\|_2=1} \min_i |\mathbf{d}^\top \mathbf{s}_i|. \quad (2)$$

Dans tous les cas $r > 0$, la contrainte entraîne un problème non convexe. Pour $r = 1$ cependant, on a montré que sous certaines conditions peu contraignantes en pratique la solution \mathbf{d}^* peut être obtenue en résolvant un problème de programmation quadratique via un algorithme QP standard (cf. [3] et CVX [4]).

La Fig.1 (b) présente une interprétation géométrique du cas $r = 1$, intéressante pour la suite. Les alternatives étant normées elles appartiennent à la sphère unité (Σ_1). Le problème est alors de trouver le plus petit cercle \mathcal{C} entourant toutes les alternatives sur la sphère: l'atome minimax \mathbf{d}^* est dans la direction du centre de ce cercle. Le problème est aussi équivalent à maximiser la marge (plan \mathcal{P}_d) qui repousse tous les \mathbf{s}_i de l'autre côté, ou à trouver la sphère Σ de plus petit rayon qui englobe tous les \mathbf{s}_i [3]. Les alternatives "marginales", i.e. pour lesquelles la puissance détection est minimale, sont sur le cercle \mathcal{C} .

Pour des données réelles comme des spectres en longueur d'onde, où $N \gg 3$, de telles alternatives peuvent être nombreuses, alors que dans le cas $N = 3$ la probabilité d'en avoir plus de deux est faible. Ceci est illustré sur la Fig.1 : (a) représente un ensemble de 100 alternatives possibles (\mathbf{S}_{100}) sous \mathcal{H}_1 . Ces alternatives sont issues de raies d'émission de spectres astrophysiques. Le panneau (c) compare \mathbf{d}^* à la solution obtenue par l'algorithme K-SVD pour $r = 1$ [6]. Clairement, \mathbf{d}^* reflète mieux les comportements marginaux (raies larges et/ou "multimodales") que K-SVD, qui modélise bien le comportement moyen. Le panneau (d) montre les alternatives sur les bords de \mathcal{C} dans ce cas. Il y en 5 ici, et elles présentent claire-

ment des profils différents des autres. Ceci suggère qu'en augmentant le nombre r de colonnes du dictionnaire optimisé afin de mieux représenter les alternatives marginales, il peut être possible d'augmenter la probabilité de détection minimale. Un autre intérêt à augmenter r est qu'un meilleur échantillonnage de la distribution des alternatives devrait s'accompagner d'une augmentation de la performance moyenne.

Ces observations vont en faveur d'une contrainte moins forte sur la réduction de dimension et ont motivé la présente étude. Cependant, le test (1) issu de (M_2) considéré dans [3] s'avère présenter des difficultés majeures dans ce cadre : premièrement, l'optimisation de (1) ne semble pas se prêter pas à une reformulation dans un cadre d'optimisation convexe ; deuxièmement, l'application d'algorithmes d'apprentissage, de la littérature ou *ad hoc* pour approcher une solution de (1) dans le cas $r > 1$ ont mené à des résultats peu concluants. Ceci nous a conduit à nous tourner vers une variante de (M_2) , plus proche du modèle exact (M_1) et donc plus facilement interprétable.

3 Modèle et test proposés

Considérons à présent le modèle :

$$\begin{cases} \mathcal{H}_0 & : \mathbf{x} = \mathbf{n}, & \mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \\ \mathcal{H}_1 & : \mathbf{x} = \mathbf{D}\boldsymbol{\beta} + \mathbf{n}, & \|\boldsymbol{\beta}\|_0 = 1 \end{cases} \quad (M_3)$$

avec $\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_r]$ normalisé, $\mathbf{D} \in \mathbb{R}^{N \times r}$, $r \ll N$ et $\boldsymbol{\beta}$ inconnu et 1-parcimonieux. Le GLR pour (M_3) mène à :

$$\max_{j=1,\dots,r} |\mathbf{d}_j^\top \mathbf{x}| \geq_{\mathcal{H}_0}^{\mathcal{H}_1} \xi \Leftrightarrow \max_{j=1,\dots,r} (\mathbf{d}_j^\top \mathbf{x})^2 \geq_{\mathcal{H}_0}^{\mathcal{H}_1} \xi^2. \quad (3)$$

Il ne paraît pas souhaitable d'imposer l'orthogonalité des colonnes de \mathbf{D} afin de laisser suffisamment de liberté dans le processus d'adaptation. Par conséquent, les variables $c_j = \mathbf{d}_j^\top \mathbf{x}$, $j = 1, \dots, r$ sont corrélées. Sous \mathcal{H}_0 , $c_j = \mathbf{d}_j^\top \mathbf{n}$ et $\mathbb{E}c_j c_k = \mathbf{d}_j^\top \mathbf{d}_k$. Sous \mathcal{H}_1 , on suppose dans l'analyse qui suit que les alternatives ont même rapport signal à bruit donc même amplitude, et on posera sans perte de généralité $\alpha = 1$. Dans ce cas $\mathbf{x} = \mathbf{s}_i + \mathbf{n}$, et $\text{P}_{\text{Det}}(\mathbf{D}, \mathbf{s}_i) = \mathbb{P}(\max_{j=1,\dots,r} c_j^2 > \xi^2 | \mathbf{s}_i)$ avec $c_j \sim \mathcal{N}(\mathbf{d}_j^\top \mathbf{s}_i, 1)$

et $\mathbb{E}c_j c_k = \mathbf{d}_j^\top \mathbf{s}_i \mathbf{s}_i^\top \mathbf{d}_k + \mathbf{d}_j^\top \mathbf{d}_k$. Malheureusement, le maximum de variables aléatoires corrélées n'admet pas de caractérisation statistique facilement exploitable. Ceci implique que la probabilité de fausse alarme $\text{P}_{\text{FA}} = \mathbb{P}(\max_{j=1,\dots,r} c_j^2 > \xi^2 | \mathcal{H}_0)$

doit être calculée par Monte Carlo, et ceci ôte surtout tout espoir d'élaborer une stratégie analytique pour maximiser sur \mathbf{D} $\min_{i=1,\dots,L} \text{P}_{\text{Det}}(\mathbf{D}, \mathbf{s}_i)$. On peut néanmoins faire les observations suivantes. Sous \mathcal{H}_1 , $c_j^2 | \mathbf{s}_i \sim \chi_1^2((\mathbf{d}_j^\top \mathbf{s}_i)^2)$, et par conséquent :

$$\mathbb{P}(c_j^2 > \xi^2 | \mathbf{s}_i) = 1 - \Phi_{\chi_{1, (\mathbf{d}_j^\top \mathbf{s}_i)^2}} = Q_{\frac{1}{2}}(|\mathbf{d}_j^\top \mathbf{s}_i|, \xi),$$

où $\Phi_{\chi_{1, (\mathbf{d}_j^\top \mathbf{s}_i)^2}}$ est la fonction de répartition d'un χ^2 décentré à un degré de liberté, et $Q_{\frac{1}{2}}(\mathbf{d}_j^\top \mathbf{s}_i, \xi)$ la fonction Marcum Q généralisée. Cette fonction est croissante en son premier argument, ici $|\mathbf{d}_j^\top \mathbf{s}_i|$ [9]. Ceci établi, l'approche simplifiée suivie ci-dessous considère que le comportement de $\max_{j=1,\dots,r} |c_j^2|$ est

essentiellement dominé par celui de la variable c_j^2 ayant la plus forte moyenne. Cette approximation signifie, en nommant $\mathbf{d}_{j^*} = \arg \max_{\mathbf{d}_1, \dots, \mathbf{d}_r} |\mathbf{d}_j^\top \mathbf{s}_i|$, que $P_{\text{Det}}(\mathbf{s}_i) = \mathbb{P}(\max_{j=1, \dots, r} c_j^2 > \xi^2 | \mathbf{s}_i)$ peut être approchée par $\mathbb{P}((c_{j^*})^2 > \xi^2 | \mathbf{s}_i)$, qui est d'autant plus élevée que l'est le paramètre de non-centralité $(\mathbf{d}_{j^*}^\top \mathbf{s}_i)^2$. Dans le cadre de cette approximation, l'alternative \mathbf{s}_{i^*} la plus "mal représentée" par le dictionnaire :

$$\mathbf{s}_{i^*} = \arg \min_{\mathbf{s}_1, \dots, \mathbf{s}_L} \max_{j=1, \dots, r} |\mathbf{d}_j^\top \mathbf{s}_i|,$$

est aussi celle créant la probabilité de détection minimale. La stratégie des algorithmes d'apprentissage minimax ci-dessous sera donc de construire des dictionnaires $\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_r]$ tels que la corrélation minimale :

$$\rho^{(r)} = \min_{i=1, \dots, L} \max_{j=1, \dots, r} |\mathbf{d}_j^\top \mathbf{s}_i|, \quad (4)$$

soit rendue aussi grande que possible à r fixé, et tels que $\rho^{(r)}$ soit une fonction aussi rapidement croissante que possible de r .

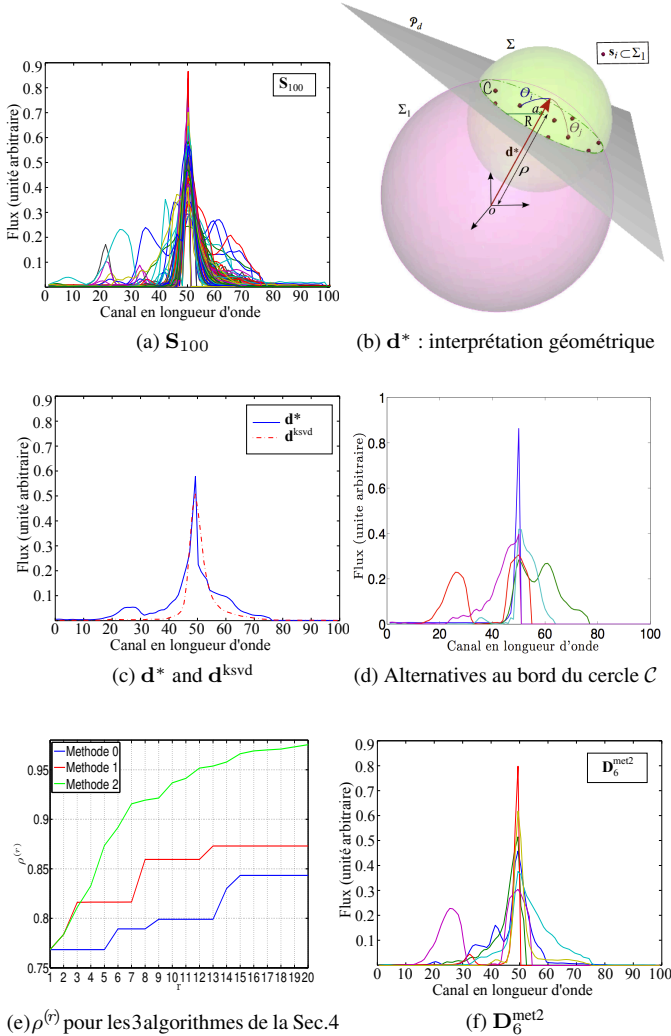


FIGURE 1 – Illustrations du problème minimax considéré.

4 Apprentissage minimax

Par manque de place on ne fournira pas ici le pseudo-code des méthodes présentées, mais celles-ci sont détaillées ci-après.

- **Méthode 0** : L'approche la plus simple dans l'optique présentée ci-dessus est de construire \mathbf{D} en concaténant \mathbf{d}^* (cf. (2)) et les $r - 1$ atomes les moins corrélés à \mathbf{d}^* .
- **Méthode 1** : Un raffinement à la méthode 0 est d'identifier progressivement les alternatives les plus mal représentées par le dictionnaire en cours, puis de classifier les données en r classes, chacune étant représentée par la solution minimax $r = 1$ associée :
 1. **Aggrégation des alternatives**. Nous commençons par calculer sur l'ensemble des L spectres l'atome minimax \mathbf{d}^* , et initialisons $\tilde{\mathbf{D}}_1^{\text{Met1}} = \mathbf{d}^*$. Le spectre \mathbf{s}_i^* de \mathbf{S} le moins corrélé à $\tilde{\mathbf{D}}_1^{\text{Met1}}$ est identifié et ajouté au dictionnaire : $\tilde{\mathbf{D}}_2^{\text{Met1}} = [\mathbf{d}^* \ \mathbf{s}_i^*]$. Le spectre \mathbf{s}_i^* de \mathbf{S} le moins corrélé à $\tilde{\mathbf{D}}_2^{\text{Met1}}$ est alors identifié et ajouté au dictionnaire. Le processus est répété jusqu'à obtenir un dictionnaire $\tilde{\mathbf{D}}_r^{\text{Met1}}$ de r atomes.
 2. **Classification et minimax par classe**. Un spectre \mathbf{s}_i est attribué à la classe de l'atome $\tilde{\mathbf{d}}_j^{\text{Met1}}$ de $\tilde{\mathbf{D}}_r^{\text{Met1}}$ si $|\mathbf{s}_i^\top \tilde{\mathbf{d}}_j^{\text{Met1}}| > |\mathbf{s}_i^\top \tilde{\mathbf{d}}_k^{\text{Met1}}|, \forall k \neq j$. Afin de mieux représenter chacune des r classes C_i ainsi obtenues, on génère l'atome minimax de chaque classe en calculant $\mathbf{d}^*(C_i) = \arg \max_{\mathbf{d}: \|\mathbf{d}\|_2=1} \min_{\mathbf{s}_i \in C_i} |\mathbf{d}^\top \mathbf{s}_i|$. Le dictionnaire résultant est $\mathbf{D}_r^{\text{Met1}} = [\mathbf{d}^*(C_1), \dots, \mathbf{d}^*(C_r)]$.
- **Méthode 2** : Il s'agit d'un raffinement à la méthode 1. Disposant de $\tilde{\mathbf{D}}_j^{\text{Met2}}$, initialisé à \mathbf{d}^* également, cet algorithme procède, avant l'identification de l'alternative \mathbf{s}_i^* la plus mal représentée par $\tilde{\mathbf{D}}_j^{\text{Met2}}$ (cf 1. ci-dessus), à une étape de mise à jour de ce dictionnaire (classification en j classes et minimax par classe) comme décrit en 2. ci-dessus.

5 Résultats numériques

Les approches d'apprentissage proposées cherchent à améliorer la puissance minimax en augmentant le nombre de colonnes, de telle sorte à ce que la plus faible corrélation maximale des alternatives avec le dictionnaire (i.e., $\rho^{(r)}$ en (4)) augmente avec r . Cette évolution est représentée pour les trois méthodes considérées en Fig.1 (e). Les fonctions $\rho^{(r)}$ obtenues sont croissantes, ce qui montre que les méthodes proposées remplissent bien toutes l'objectif visé. La capacité à représenter rapidement (suivant r) et efficacement les alternatives les plus marginales est aussi assez logiquement fonction du degré d'élaboration de la méthode d'apprentissage considérée.

La validation réelle de l'approche concerne la puissance minimax elle-même. Celle-ci est évaluée à travers les AUC (*Area Under the Curve*) des courbes CORE : les AUC intègrent la puissance à toutes les P_{FA} pour les 100 signatures spectrales activées sous \mathcal{H}_1 de la Fig.1(a). Ces AUC sont présentées en Fig. 2 et dans la Table pour 8 dictionnaires, dont 6 sont optimisés. Pour chaque dictionnaire \mathbf{D} , le test effectué est (3), où ξ est réglé pour couvrir des P_{FA} de 0 à 1. Le premier dictionnaire considéré est $\mathbf{S} = \mathbf{s}_i$, où \mathbf{s}_i est l'atome activé sous \mathcal{H}_1 :

c'est la référence. Le deuxième est la bibliothèque de spectres, \mathbf{S}_{100} . (La dimension réelle de notre bibliothèque de spectres est beaucoup plus grande [3], et interdit l'usage du GLR sur (M_1). Nous n'en présentons qu'une partie à des fins illustratives ici, et pour faire des simulations exhaustives par la suite.) Le troisième dictionnaire, \mathbf{d}^* , correspond à l'approche minimax (2) pour $\mathbf{S} = \mathbf{S}_{100}$. Les deux dictionnaires suivants sont appris par l'algorithme K-SVD [6] pour $r = 1$ et $r = 6$ atomes (\mathbf{d}^{ksvd} et $\mathbf{D}_6^{\text{ksvd}}$ respectivement). Les dictionnaires correspondants aux méthodes présentées en Sec.4 sont $\mathbf{D}_6^{\text{met0}}$, $\mathbf{D}_6^{\text{met1}}$ et $\mathbf{D}_6^{\text{met2}}$. A titre d'illustration, la Fig.1 montre les atomes de $\mathbf{D}_6^{\text{met2}}$. Ces signaux couvrent bien la diversité des alternatives du panneau (a) de la Fig.1 ; pour $\mathbf{D}_6^{\text{met2}}$, $\rho^{(6)} = 0, 9$, cf (e).

On donne aussi les résultats obtenus pour deux dictionnaires formés par la combinaison de méthodes minimax et classification sur la sphère (les atomes du dictionnaire étant normés) de la littérature sur 5 et 6 classes (moVMF [8] : $\mathbf{D}_5^{\text{vmf}}$ et Spherical K-Means [7] : $\mathbf{D}_6^{\text{skm}}$), où on a appliqué l'approche minimax sur chaque classe obtenue par ces algorithmes.

La Fig. 2 et la Table montrent que l'approche K-SVD à 1 atome (ligne grise) donne la meilleure performance moyenne mais que dans quelques cas (s_4, s_{10} et s_{90} par exemple) la puissance de ce test s'effondre. En revanche, les approches minimax \mathbf{d}^* et $\mathbf{D}_6^{\text{met2}}$ maintiennent des performances plus stables dans de tels cas, ce qui est bien sûr leur objectif.

Dictionnaire	Classement	
	AUC min	AUC moyenne
Alternative sous \mathcal{H}_1	Réf : 0.886	Réf : 0.887
\mathbf{S}_{100}	1 ^{er} : 0.813	3 ^{ème} : 0.847
\mathbf{d}^*	7 ^{ème} : 0.768	6 ^{ème} : 0.836
\mathbf{d}^{ksvd}	8 ^{ème} : 0.700	1 ^{er} : 0.863
$\mathbf{D}_6^{\text{ksvd}}$	6 ^{ème} : 0.753	4 ^{ème} : 0.843
$\mathbf{D}_6^{\text{Met0}}$	3 ^{ème} : 0.788	5 ^{ème} : 0.823
$\mathbf{D}_6^{\text{Met1}}$	4 ^{ème} : 0.776	3 ^{ème} : 0.848
$\mathbf{D}_6^{\text{Met2}}$	2 ^{ème} : 0.794	3 ^{ème} : 0.846
$\mathbf{D}_6^{\text{skm}}$	3 ^{ème} : 0.788	3 ^{ème} : 0.848
$\mathbf{D}_5^{\text{vmf}}$	5 ^{ème} : 0.772	2 ^{ème} : 0.860

TABLE 1 – Résultats sur 100 alternatives (incertitude: ± 0.001).

Concernant les méthodes proposées, on voit que la méthode 2 se compare avantageusement aux méthodes 0 et 1 en terme de puissance minimax. La méthode 2 permet aussi ici d'obtenir la meilleure performance minimax toutes méthodes confondues (l'utilisation du dictionnaire total étant en pratique exclue). Il faut aussi souligner que la méthode 2 proposée apporte dans le cas $r = 6$ une amélioration de la performance minimax (et moyenne) par rapport au cas $r = 1$ (\mathbf{d}^*) : ce résultat était l'objectif principal de la présente étude. Il confirme en particulier la pertinence de l'analyse au premier ordre basée sur la corrélation (4) (cf Fig.1 (e)). Des tests non reportés ici montrent qu'on peut obtenir un accroissement plus élevé encore de la puissance minimax à plus grand r (AUC min de 0,808 pour $r = 11$

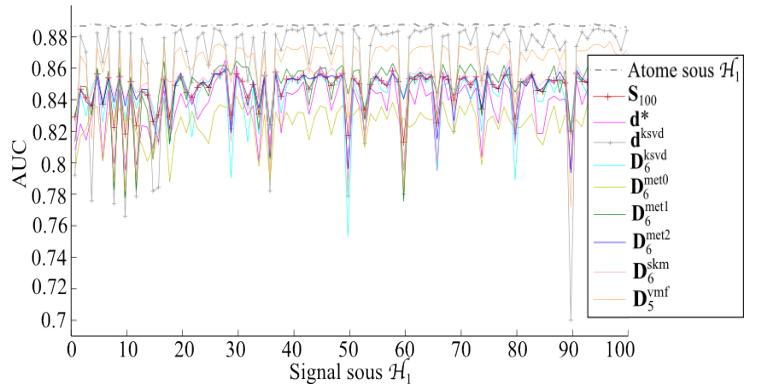


FIGURE 2 – AUC pour 100 alternatives.

avec la méthode 2 par exemple). Enfin, on notera que les méthodes combinant classification sur la sphère et approche minimax présentent également de bonnes performances moyenne et minimax, notamment Spherical K-Means.

En conclusion, ce travail a proposé, étudié et comparé un ensemble de tests et d'approches d'apprentissage minimax en détection. Bien que le cas d'une réduction de dimension extrême ($r = 1$) par rapport au modèle exact permette de résoudre exactement le problème minimax dans le cadre du GLR, cette étude montre que des dictionnaires appris à l'aide de méthodes sous-optimales (i.e. qui ne résolvent pas le critère mais cherchent à en approcher la solution) peuvent néanmoins améliorer la puissance minimax (et moyenne) de ce type de tests.

Références

- [1] A. Berk *et al* *MODTRAN5 : A reformulated atmospheric band model with auxiliary species and practical multiple scattering options : Update*. Proc. SPIE 5806, Alg. and Tech. for Multisp., Hypersp., and Ultrasp. Imagery XI, pp. 662-667, 2005.
- [2] 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 Geosc. and Rem. Sens., vol. 37(6), pp. 2706-2717, Nov.1999.
- [3] R.F.R. Suleiman, D. Mary et A. Ferrari. *Minimax sparse detection based on one-class classifiers*. ICASSP, May 2013.
- [4] CVX Research Inc. *CVX : Matlab software for disciplined convex programming, version 2.0 beta*. <http://cvxr.com/cvx>.
- [5] K. Engan *et al*. *Frame based signal compression using method of optimal directions (MOD)*. ISCAS, vol.4, pp. 1-4, Jul. 1999
- [6] M. Aharon, M. Elad, et A. Bruckstein *K-SVD : An algorithm for designing overcomplete dictionaries for sparse representation*. IEEE Trans. Sig. Proc., vol. 54(11), pp. 4311-4322, Nov. 2006.
- [7] I.S. Dhillon et D.S. Modha. *Concept decompositions for large sparse text data using clustering*. Journal of Machine Learning, vol. 42(1), pp. 143-175, Jan. 2001.
- [8] A. Banerjee, I.S. Dhillon, J. Ghosh et S. Sra. *Clustering on the unit hypersphere using Von Mises-Fisher distribution*. Journal of Machine Learning Research, 2005.
- [9] A.H. Nutall. *Some Integrals Involving the (Q sub M) - Function*, Naval Underwater System Center New London, Connecticut, Tech. Report AD-779846, 1974.