

Sur le Contrôle de la Complexité des Détecteurs opérant dans le Domaine Temps-Fréquence par le biais de la Fonction de Paramétrisation

Cédric RICHARD, Régis LENGELLE

Laboratoire de Modélisation et Sûreté des Systèmes, Université de Technologie de Troyes
12 rue Marie Curie, BP 2060, 10010 Troyes Cedex, France

cedric.richard@univ-troyes.fr, regis.lengelle@univ-troyes.fr

Résumé – Le problème considéré concerne la synthèse de détecteurs opérant dans le domaine temps-fréquence lorsque les seules informations disponibles a priori consistent en une base de réalisations étiquetées des hypothèses « bruit » et « signal + bruit ». En adoptant une argumentation propre à la théorie de l'apprentissage, on montre ainsi qu'un choix autre que la représentation de Wigner-Ville, au sein de la classe de Cohen, peut conduire à une amélioration significative des performances de la structure de décision, malgré la perte d'information statistique qui peut en résulter. Ce phénomène est alors illustré par un exemple.

Abstract – This paper deals with the design of time-frequency based detectors from labelled training data. Using elements of the Learning Theory, it is first shown that better performance can be obtained with a receiver operating on an other representation of Cohen's class than the Wigner-Ville one, even if statistical information is lost. Then a simple example is proposed to illustrate this phenomenon.

1. Introduction

En environnement non-stationnaire, la potentialité des méthodes d'analyse temps-fréquence de la classe de Cohen a été exploitée par de nombreux auteurs pour la résolution de problèmes de décision [1]-[3]. Un nombre important de ces approches nécessite cependant une connaissance parfaite des propriétés statistiques de l'échantillon pour chacune des hypothèses en compétition, à savoir les hypothèses « bruit » et « signal + bruit », que l'on note conventionnellement H_0 et H_1 . Cette connaissance a priori étant souvent inaccessible dans le cadre d'applications, Jones et Sayeed ont proposé dans un premier temps de maximiser un critère de contraste particulier, en l'occurrence celui de Fisher, sur un ensemble de réalisations des hypothèses H_0 et H_1 [4]. Puis, cette approche a été améliorée grâce à une procédure de recherche automatique du meilleur critère de contraste pour le problème rencontré, au sens de la probabilité d'erreur [5], [6]. Malheureusement, obtenir des performances satisfaisantes avec un détecteur déterminé à partir de signaux étiquetés nécessite de trouver une adéquation entre la complexité de celui-ci, formellement définie en théorie de l'apprentissage par la dimension de Vapnik-Chervonenkis, et le nombre de signaux dont on dispose pour l'apprentissage [7], [8].

L'objectif de cet article est de montrer que dans le cadre de la détection temps-fréquence pratiquée au sein de la classe de Cohen, la dimension de Vapnik-Chervonenkis des détecteurs peut être contrôlée par la fonction de paramétrisation de la représentation adoptée. Cet article est organisé ainsi. Dans un premier temps, nous décrivons brièvement la méthode évoquée précédemment pour la synthèse de détecteurs à partir d'un ensemble de signaux d'apprentissage. Puis, après avoir

présenté certains éléments de la théorie de l'apprentissage, nous exposons la raison pour laquelle un choix autre que la représentation de Wigner-Ville, au sein de la classe de Cohen, peut conduire à une amélioration des performances d'un détecteur opérant dans le plan temps-fréquence, malgré la perte d'information statistique qui peut en résulter. Enfin, nous illustrons ce phénomène par un exemple.

2. Synthèse de la structure de détection

Si on écarte les solutions reposant sur l'extraction d'attributs des représentations (position d'un pic, encombrement temps-fréquence d'une composante, etc.), la majorité des travaux relatifs à la synthèse de détecteurs temps-fréquence à partir d'une base d'apprentissage concerne des structures de détection linéaires opérant sur la représentation de Wigner-Ville du signal observé [1]-[6]. Cette démarche est toujours motivée par la possibilité d'aboutir à un test optimal lorsque la solution du problème rencontré consiste en la comparaison d'une statistique de détection quadratique à un seuil. Pour cette même raison, nous restreignons le cadre de la présente étude à l'ensemble des règles d de détection définies ainsi :

$$d : \begin{cases} \text{si } \lambda(X) \equiv (a_{TF})^T W_X < \lambda_0, & \text{alors } H_0 \\ \text{sinon } H_1. \end{cases} \quad (1)$$

W_X désigne la représentation de Wigner-Ville de l'observation X écrite vectoriellement, et a_{TF} et λ_0 les paramètres du détecteur devant être déterminés à partir d'un ensemble de réalisations de H_0 et H_1 .

La synthèse du détecteur (1) consiste à déterminer le couple $(a_{TF}; \lambda_0)$ optimal au sens d'un critère à définir préalablement, étant donné un ensemble d'apprentissage. Ce problème étant NP difficile lorsque le critère retenu est la probabilité d'erreur p_e [8], on a généralement recours aux critères de contraste ne dépendant que des moments d'ordre 1 et 2 de la statistique de détection $\lambda(X)$ (critère de Fisher, rapport signal sur bruit, déflexion, etc.) parce qu'ils se prêtent à une optimisation plus aisée [5], [6]. En effet, quel que soit le critère de contraste c , on montre que la référence optimale a_{TF} est solution du système :

$$[\alpha \Sigma_0 + (1-\alpha) \Sigma_1] a_{TF} = M_1 - M_0. \quad (2)$$

M_i et Σ_i désignent respectivement les moments conditionnels d'ordre 1 et 2 de W_X , et α un paramètre réel dépendant uniquement de c . Subsiste peut-être le problème du choix d'un critère c particulier. La remarque suivante montre cependant qu'il n'en est rien. En optimisant les valeurs de α et λ_0 de sorte que le détecteur (1) minimise une estimation de la probabilité d'erreur \hat{p}_e , a_{TF} étant solution de l'équation (2), on obtient en effet la règle de détection optimale au sens du meilleur critère de contraste ne dépendant que des moments conditionnels d'ordre 1 et 2 de $\lambda(X)$. Pour cette raison, cette procédure est nommée *méthode du critère optimal*.

3. Optimisation de la complexité de la structure de détection

3.1 Principe

Comme l'illustre la Figure 1, les performances d'un détecteur établi à partir d'une base d'apprentissage sont conditionnées par l'adéquation entre la complexité de celui-ci, formellement définie en théorie de l'apprentissage par la dimension de Vapnik-Chervonenkis, et le nombre de signaux disponibles. En effet, les récepteurs dotés d'une complexité trop importante auront un faible pouvoir de généralisation. Dans le cas contraire, ces derniers seront incapables d'intégrer la totalité de l'information statistique présente dans l'ensemble d'apprentissage. Entre ces extrêmes, il existe une complexité optimale pour laquelle la probabilité d'erreur est minimale. Ce comportement de la probabilité d'erreur, que l'on qualifie souvent de *malédiction de la dimensionnalité*, a été formellement établi en théorie de l'apprentissage par Vapnik et Chervonenkis [7], [8]. Ils ont en effet montré qu'avec une probabilité égale à $(1-\varepsilon)$, l'inégalité suivante est satisfaite, quelle que soit la règle de détection d d'une classe \mathcal{D} donnée.

$$\hat{p}_e(d) - E(m, V_{\mathcal{D}}, \varepsilon) \leq p_e(d) \leq \hat{p}_e(d) + E(m, V_{\mathcal{D}}, \varepsilon) \quad (3)$$

avec :

$$E(m, V_{\mathcal{D}}, \varepsilon) = \sqrt{\frac{V_{\mathcal{D}}}{m} \left(1 + \ln \frac{2m}{V_{\mathcal{D}}} \right)} - \frac{1}{m} \ln \frac{\varepsilon}{4}.$$

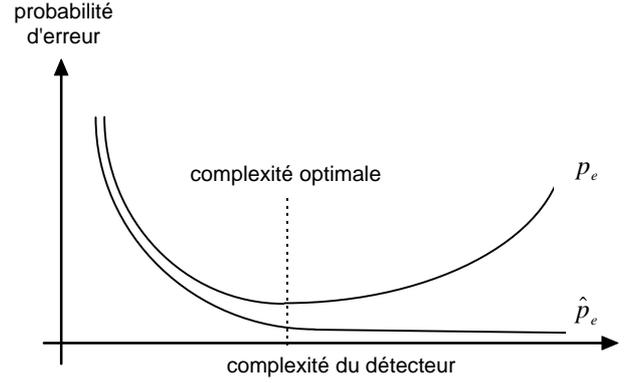


FIG 1 : Performance d'un détecteur obtenu à partir d'un ensemble d'apprentissage de taille finie. p_e désigne la probabilité d'erreur de celui-ci, et \hat{p}_e sa valeur estimée à partir des signaux d'apprentissage.

Dans cette expression, $p_e(d)$ désigne la probabilité d'erreur de d , et $\hat{p}_e(d)$ sa valeur estimée à partir des signaux de l'ensemble d'apprentissage, supposés au nombre de m . Le paramètre $V_{\mathcal{D}}$, dit *dimension de Vapnik-Chervonenkis* de \mathcal{D} , représente quant à lui l'aptitude des détecteurs de cette classe à résoudre des problèmes plus ou moins complexes. Le lecteur intéressé en trouvera une définition formelle dans [8].

Si l'estimation de $V_{\mathcal{D}}$ est en général relativement difficile, on retiendra cependant que $V_{\mathcal{D}} = n + 1$ lorsque \mathcal{D} désigne la classe des détecteurs linéaires opérant dans \mathcal{R}^n [8]. Il en résulte que le mode opératoire de certaines procédures visant à améliorer les performances de telles structures de détection consiste en une projection des observations X dans un espace adéquat de dimension $n' < n$, l'objectif étant alors d'établir un meilleur compromis entre $V_{\mathcal{D}}$ et m [5].

Transposé au cadre de la détection linéaire dans le domaine temps-fréquence, il en émane que des résultats équivalents peuvent être attendus en élaborant le détecteur sur une représentation C_X^o non inversible et convenablement choisie, dont la fonction de paramétrisation est désignée par ϕ . Le contrôle de la dimension de l'espace des observations, initialement engendré par W_X , est dans ce cas assuré par l'annulation partielle de ϕ dans le domaine des ambiguïtés. Ainsi, bien que d'autres solutions soient envisageables pour la synthèse de la fonction de paramétrisation ϕ , ce qui constitue une perspective intéressante de ce travail, ce principe simple peut aisément être mis en oeuvre en procédant de la sorte :

1. Choix d'un type de fonction de paramétrisation ϕ , les considérations prises en compte pouvant alors être diverses (amélioration de la lisibilité, possibilité d'une implémentation récursive, etc.)
2. Optimisation du support de ϕ afin de minimiser une estimation de la probabilité d'erreur.

Dans la section qui suit, l'efficacité potentielle de cette approche est vérifiée par la résolution d'un problème de détection particulier.

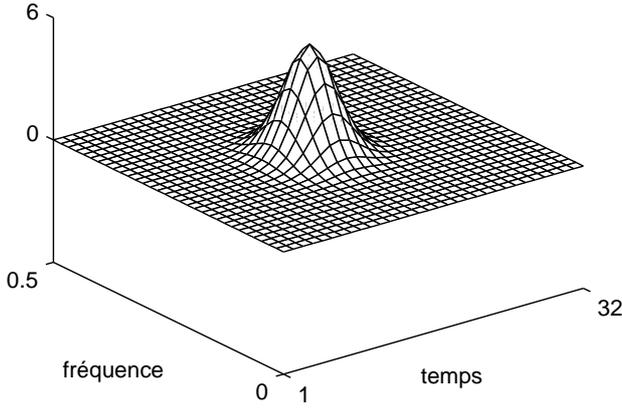


FIG 2 : Représentation de Wigner-Ville de l'événement transitoire S à détecter.

3.2 Expérimentation

Le problème de détection considéré concerne la détection d'un signal aléatoire S noyé dans un bruit blanc B , gaussien, centré et identiquement distribué sous H_0 et H_1 :

$$\begin{cases} H_0 : X = B \\ H_1 : X = S + B, \end{cases} \quad (4)$$

où $S(k) = \exp(-(k-16)^2/18) \cos(k\pi/2 + \Theta_0)$, $1 \leq k \leq 32$, Θ_0 étant une variable aléatoire uniformément distribuée sur l'intervalle $[0, 2\pi]$. La variance de B est fixée de sorte que le rapport signal sur bruit soit égal à -6 dB. Dans ces circonstances, il est établi que le test de détection optimal est obtenu pour $a_{TF} = W_S$, où W_S désigne la représentation de Wigner-Ville du signal à détecter [1]-[4], qui est par ailleurs présentée en Figure 2. On rappelle que ce détecteur est appelé *filtre adapté temps-fréquence*.

Afin d'illustrer les effets néfastes d'un ensemble d'apprentissage de faible cardinalité sur les performances

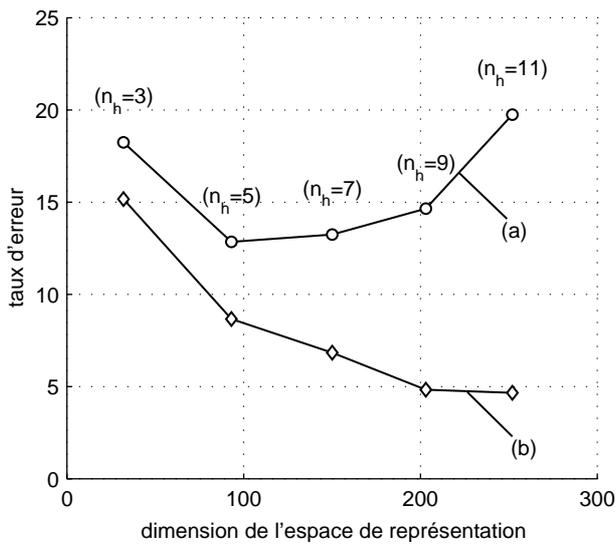


FIG 4 : Evolution du taux d'erreur du détecteur en fonction de la largeur n_h de la fenêtre d'analyse h .

- (a) estimation du taux d'erreur sur $\mathcal{E}_{\text{test}}$
- (b) estimation du taux d'erreur sur \mathcal{E}_{app}

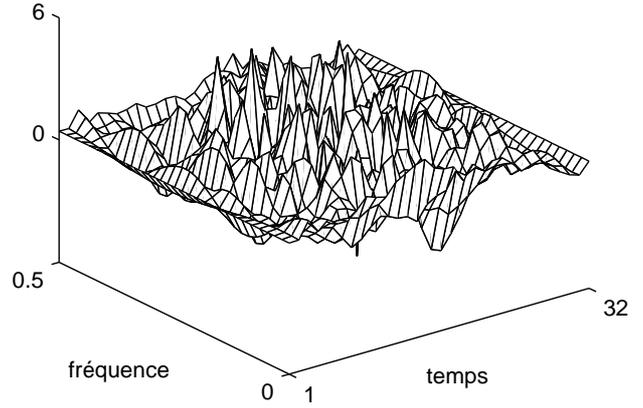


FIG 3 : Configuration \hat{a}_{TF} du détecteur opérant sur la représentation de Wigner-Ville de l'observation X .

d'une structure de détection, le détecteur (1) a été optimisé grâce à la méthode du critère optimal, sur la base de 300 réalisations de chacune des hypothèses H_0 et H_1 du problème (4). Il apparaît sur la Figure 3 que la référence \hat{a}_{TF} obtenue ne reflète pas le contenu temps-fréquence du signal S , ce qui s'avère être une manifestation de la malédiction de la dimensionnalité, comme cela est montré dans [5].

En conséquence, afin d'établir un meilleur compromis entre le cardinal de l'ensemble d'apprentissage et la complexité du détecteur temps-fréquence, celui-ci a été synthétisé de sorte à opérer sur le spectrogramme de l'observation X , qui constitue un exemple de représentation C_X^ϕ non inversible. Dans ces circonstances, la dimension de l'espace de représentation, estimée en évaluant le rang de Σ_0 ou Σ_1 , a alors pu être contrôlée en recherchant la largeur n_h optimale de la fenêtre d'analyse h . Pour ce faire et étant donnée h , un ensemble \mathcal{E}_{app} de 200 réalisations de chacune des hypothèses H_0 et H_1 a été réservé à l'optimisation du

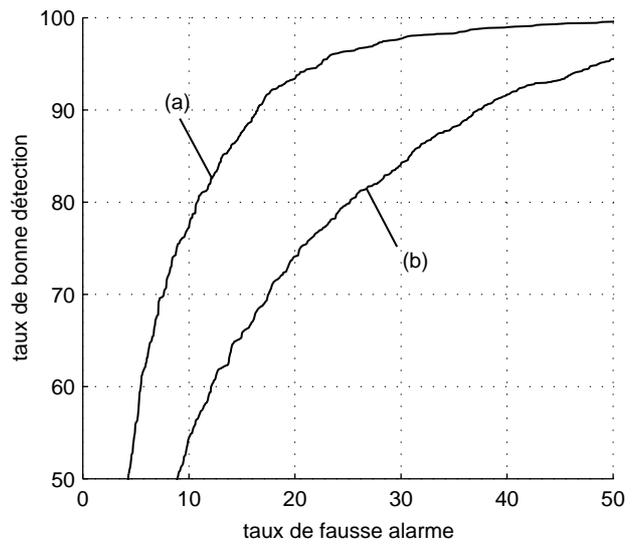


FIG 5 : Courbes COR des détecteurs linéaires opérant sur :

- (a) le spectrogramme à fenêtre de Hanning de largeur 5,
- (b) la représentation de Wigner-Ville.

Pour ces estimations, 2000 données ont été utilisées.

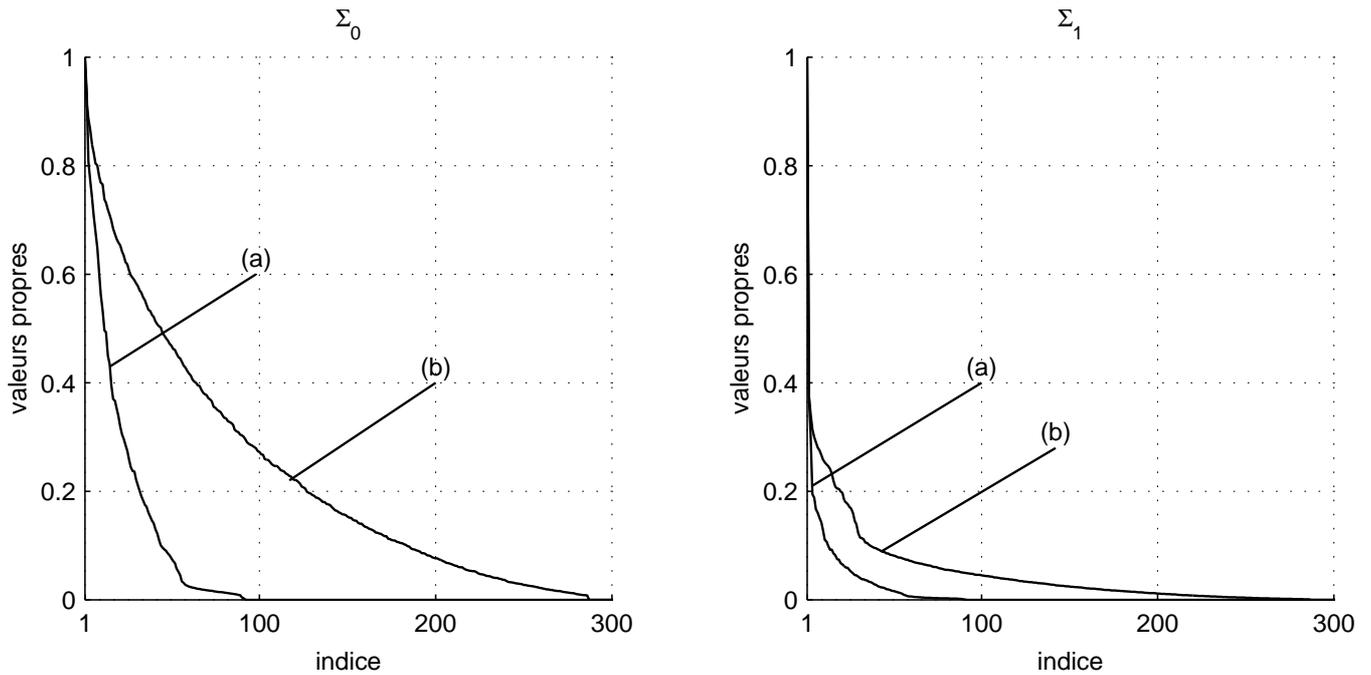


FIG 6 : Valeurs propres normalisées des matrices de variance-covariance Σ_0 et Σ_1 des composantes (a) du spectrogramme et (b) de la représentation de Wigner-Ville des observations. Ce résultat illustre le fait que la dimension de l'espace engendré est plus faible dans le cas (a) que dans le cas (b).

détecteur selon la méthode du critère optimal, la base $\mathcal{E}_{\text{test}}$ constituée des 100 échantillons restants étant consacrée à l'estimation des performances de celui-ci. La Figure 4 indique que les meilleurs résultats ont été obtenus pour une fenêtre de Hanning de largeur égale à 5.

Finalement, les performances des deux détecteurs synthésés dans le cadre de cette étude ont été estimées grâce à un ensemble de 2000 réalisations de chacune des hypothèses H_0 et H_1 en compétition. Les résultats indiqués en Figure 5 montrent que la structure de détection opérant sur le spectrogramme est plus performante que celle opérant sur la représentation de Wigner-Ville. Comme le montre la Figure 6, ce phénomène est justifié par une dimension de l'espace de représentation plus en adéquation avec le faible cardinal de l'ensemble d'apprentissage.

4. Conclusion

En adoptant une argumentation propre à la théorie de l'apprentissage, on a montré qu'un choix approprié de la représentation C_x^o conditionne significativement les performances susceptibles d'être atteintes par une structure de détection opérant dans le plan temps-fréquence. C'est là une illustration supplémentaire des potentialités de la classe de Cohen pour la résolution de certains problèmes de décision, qui confère à la représentation un rôle actif et décisif dans le processus décisionnel.

Références

- [1] P. FLANDRIN. *A time-frequency formulation of optimum detection*. IEEE Trans. on ASSP, 1988, vol. 36, n°9, p. 1377-84.
- [2] A. SAYEED et D. JONES. *Optimal detection using bilinear time-frequency an time-scale representations*. IEEE Trans. on SP, 1995, vol. 43, n°12, p. 2872-83.
- [3] G. MATZ et F. HLAWATSCH, *Time-frequency formulation and design of optimal detectors*. Proc. IEEE TFTS, 1996, p. 213-6.
- [4] D. JONES et A. SAYEED. *Blind quadratic and time-frequency based detectors from training data*. Proc. IEEE ICASSP, 1995, p. 1033-6.
- [5] C. RICHARD. *Une méthodologie pour la détection à structure imposée. Applications au plan temps-fréquence*. Th. : UTC, Compiègne, 1998.
<http://www-isis.enst.fr/Kiosque/theses/>
- [6] C. RICHARD et R. LENGELLE. *Une nouvelle approche pour la détection linéaire optimale dans le plan temps-fréquence*. Proc. GRETSI, 1997, p. 659-62.
- [7] V. VAPNIK, A. CHERVONENKIS. *On the uniform convergence of relative frequencies events to their probabilities*. Theory of probabilities and its applications, 1971, vol. 16, p.264-80.
- [8] L. DEVROYE, L. GYORFI et G. LUGOSI. *A probabilistic theory of pattern recognition*. New York : Springer Verlag, 1996.