

Thèse

présentée pour l'obtention du grade de

Docteur
de l'Université de Technologie de
Compiègne

spécialité

Contrôle des Systèmes

par

Cédric RICHARD

**Une Méthodologie pour la Détection à
Structure Imposée**

Applications au plan temps-fréquence

Soutenue le 22 décembre 1998 devant le jury composé de :

MM. DENOEUX T.
DUBUISSON B.
DUVAUT P. (rapporteur)
FLANDRIN P. (rapporteur)
GAILLARD P. (président)
GONCALVES P.
LENGELLE R.
NIKIFOROV I.

A Hélène.

Remerciements

Ce travail a été réalisé au sein du Laboratoire de Modélisation et Sécurité des Systèmes de l'Université de Technologie de Troyes, où j'ai pu bénéficier d'un encadrement efficace et d'une ambiance chaleureuse. Que les membres de cette équipe reçoivent toute ma gratitude.

Je tiens à remercier en particulier M. Régis Lengellé, Professeur à l'UTT, d'avoir accepté de diriger cette thèse. Ses qualités scientifiques et humaines m'ont permises de la mener à bien, dans les meilleures conditions. Qu'il trouve ici l'expression de ma plus vive sympathie, et le souhait de conserver les liens que nous avons pu établir.

C'est avec chaleur que je remercie également M. Paul Gaillard, Professeur à l'UTT, pour m'avoir accueilli au sein de l'Université qu'il préside. Il m'a fait découvrir, avec M. Régis Lengellé, les fondements du traitement du signal et de la théorie de la décision à l'Université de Technologie de Compiègne. C'est donc pour moi un grand plaisir qu'il ait accepté de présider mon jury de thèse.

Je tiens à exprimer ma profonde reconnaissance à M. Patrick Flandrin, Directeur de Recherche à l'ENS Lyon, et M. Patrick Duvaut, Professeur à l'ENSEA, pour avoir accepté de rapporter le présent document. L'intérêt qu'ils ont porté à mon travail constituera longtemps pour moi une source de motivation, et les conseils et suggestions qu'ils ont pu me prodiguer autant de nouvelles perspectives de recherche. Je les en remercie vivement.

Je remercie également chaleureusement les autres personnalités scientifiques qui ont bien voulu accorder de l'intérêt à mes travaux en me faisant l'honneur de participer au jury, à savoir : M. Thierry Denoeux, M. Bernard Dubuisson, M. Paulo Gonçalves et M. Igor Nikiforov. En particulier, que M. Igor Nikiforov sache qu'il habité mes pensées tout au long de l'écriture du manuscrit. Son attachement pour les problèmes bien posés m'a encouragé à soigner la présentation de la démarche adoptée, ce qui donne la coloration de cette thèse.

Je ne peux pas non plus ignorer tous ceux qui ont su parfois me faire oublier le stress quotidien de ces trois dernières années, par le partage de moments de joie et de détente. J'adresse ainsi un grand merci à l'ensemble de mes collègues et amis.

Mes derniers remerciements iront enfin à ma compagne, Hélène, dont le soutien de chaque instant fut des plus réconfortants et la patiente sans limite. Sans elle, rien n'aurait sans doute été possible. C'est donc pour moi un grand plaisir de lui dédier l'ensemble de ce travail.

Table des matières

INTRODUCTION GÉNÉRALE	11
1. ÉLÉMENTS D'ANALYSE TEMPS-FRÉQUENCE	9
1.1 INTRODUCTION	15
1.2 LES REPRÉSENTATIONS LINÉAIRES	18
1.2.1 <i>La transformée de Fourier à court terme</i>	18
1.2.2 <i>La représentation de Gabor</i>	19
1.3 LES REPRÉSENTATIONS D'ÉNERGIE	19
1.3.1 <i>Motivations</i>	19
1.3.2 <i>Classe de Cohen</i>	21
1.3.3 <i>Exemples</i>	23
1.3.4 <i>Interférences</i>	24
1.3.5 <i>Lissage et distributions à interférences réduites</i>	25
1.3.6 <i>La méthode de réallocation</i>	26
1.4 CONCLUSION.....	28
2. IMPLÉMENTATION RÉCURSIVE DES REPRÉSENTATIONS TEMPS-FRÉQUENCE	26
2.1 PRINCIPES GÉNÉRAUX DE LA RÉCURSIVITÉ	32
2.1.1 <i>Récurtivité directe</i>	33
2.1.2 <i>Récurtivité indirecte</i>	34
2.2 LA TRANSFORMÉE DE FOURIER À COURT TERME.....	34
2.2.1 <i>Formulation récursive directe</i>	34
2.2.2 <i>Exemple de formulation récursive indirecte</i>	36
2.2.3 <i>Performances</i>	36
2.3 DISTRIBUTIONS DE LA CLASSE DE COHEN.....	38
2.3.1 <i>Formulation récursive directe</i>	38
2.3.2 <i>Distribution pseudo Wigner-Ville lissée récursive</i>	40
2.4 RÉCURSIVITÉ ET MÉTHODE DE RÉALLOCATION.....	43
2.4.1 <i>Spectrogramme réalloué</i>	44

2.4.2 Distributions pseudo Wigner-Ville lissées réallouées	46
2.5 CONCLUSION.....	50
2.6 ANNEXES DU CHAPITRE 2	51
2.6.1 Récursivité et représentations non réallouées.....	51
2.6.2 Récursivité et représentations réallouées.....	52
3. SYNTHÈSE DE DÉTECTEURS À STRUCTURE IMPOSÉE : MÉTHODOLOGIE	49
3.1 INTRODUCTION	55
3.2 DÉTECTION À STRUCTURE LIBRE	57
3.2.1 Hypothèses simples.....	57
3.2.2 Hypothèses composées	59
3.2.3 Exemples de problèmes de détection.....	61
3.3 DÉTECTION À STRUCTURE IMPOSÉE.....	63
3.3.1 Présentation du problème.....	63
3.3.2 Construction de la classe de détecteurs D	66
3.3.3 Choix et optimisation du critère $c(d, A_n)$	69
3.3.4 Optimisation d'un ensemble de critères de contraste	74
3.3.5 Expérimentations dans le plan temps-fréquence	78
3.3.6 Conclusion.....	82
4. DÉTECTEURS À STRUCTURE IMPOSÉE : OPTIMISATION DE LA COMPLEXITÉ	81
4.1 RECHERCHE D'UNE COMPLEXITÉ OPTIMALE	87
4.1.1 Convergence de la probabilité d'erreur empirique	87
4.1.2 Dimension de Vapnik-Chervonenkis	88
4.1.3 Le principe Structural Risk Minimization (SRM)	90
4.2 PRINCIPE SRM ET DÉTECTION TEMPS-FRÉQUENCE.....	91
4.2.1 Méthodes de sélection.....	91
4.2.2 Méthode de partitionnement.....	103
4.3 TRANSFORMÉE DE WIGNER-VILLE DISCRÈTE	107
4.3.1 Conventions et position du problème	108
4.3.2 Dimension des espaces Image de R_x et W_x	109
4.3.3 Exemples.....	114
4.4 CONCLUSION.....	117
5. DÉTECTION DU COMPLEXE K DANS L'EEG DE SOMMEIL	119
5.1 INTRODUCTION	119
5.2 PRÉSENTATION DU PROBLÈME.....	120

5.2.1 Macrostructure de l'électroencéphalogramme.....	120
5.2.2 Microstructure de l'électroencéphalogramme.....	122
5.3 SYNTHÈSE D'UN DÉTECTEUR DE COMPLEXES K.....	123
5.3.1 Description des données.....	123
5.3.2 Résultats et discussion.....	126
5.3.3 Conclusion.....	128
CONCLUSION GÉNÉRALE	131
BIBLIOGRAPHIE	135

Introduction générale

La théorie de la détection a pour objectif de conduire à une prise de décision optimale, choisie parmi un nombre fini d'alternatives possibles, étant donnée une réalisation d'un échantillon aléatoire. En particulier, dans le cadre de la détection d'un signal noyé dans un bruit, les décisions envisageables consistent à opter pour la validité de l'une des hypothèses suivantes : "l'observation $x(t)$ n'est constituée que de bruit" ou "le signal $s(t)$ est présent dans l'observation $x(t)$ ". Le critère fixant les objectifs de la structure de détection en terme d'optimalité peut alors consister, par exemple, à minimiser le coût moyen d'une décision ou encore à maximiser la probabilité de détection du signal $s(t)$ à probabilité de fausse alarme donnée. Pour ces deux critères en particulier, on montre que l'optimum est atteint en choisissant pour *statistique de détection* Λ le *rapport de vraisemblance*. La comparaison de celui-ci à un seuil λ_0 fournit alors la décision : le signal $s(t)$ est supposé présent dans l'observation $x(t)$ si Λ est supérieur à λ_0 , et absent dans le cas contraire. On parle alors de *détection à structure libre* puisqu'aucune contrainte ne pèse sur la structure de la statistique de détection, celle-ci reposant uniquement sur le choix d'une règle et la connaissance des lois de vraisemblance [Duv91].

Cependant, dans un grand nombre de situations, on ne dispose pas du confort absolu que constitue la connaissance des propriétés statistiques de l'échantillon pour chacune des hypothèses en compétition, ce qui rend inopérante l'approche précédente. Une démarche envisageable, néanmoins sous-optimale, consiste alors à définir préalablement la nature de la statistique de détection, puis à en optimiser les paramètres caractéristiques selon un critère donné. Pour cette raison, cette approche est qualifiée de *détection à structure imposée* [Duv91]. Dans ces circonstances, les problèmes rencontrés lors de l'élaboration d'une règle de détection sont généralement liés aux difficultés à répondre aux questions suivantes :

- Comment choisir la structure de la statistique de détection ?
- Quels sont les critères pertinents pour la détermination des paramètres caractéristiques de cette statistique ? Quelle procédure d'optimisation adopter ?

L'objectif de ce mémoire de thèse est d'apporter des éléments de réponse à chacune de ces interrogations lorsqu'on ne dispose, pour seule source d'information, que d'un ensemble de réalisations étiquetées des hypothèses en compétition. Cette hypothèse de travail a été retenue parce qu'elle est conforme aux conditions expérimentales que l'on a rencontrées dans le cadre de l'application ayant initialement motivée la présente étude : la détection du signal transitoire "complexe K" dans l'électroencéphalogramme de sommeil. Avant de poursuivre, il convient de présenter brièvement ce problème, qui nous a servi de référence pour isoler les diverses difficultés que l'on peut rencontrer avec des données réelles, afin de justifier certaines options adoptées au cours de nos travaux.

L'étude de l'électroencéphalogramme a débuté dans les années 30 [Loo35] avec la première classification en stades de l'activité cérébrale humaine durant le sommeil. Ce n'est qu'après une longue période de normalisation des analyses que les efforts se sont finalement concentrés sur l'analyse des phénomènes transitoires. Actuellement, l'un des principaux objectifs est la compréhension des mécanismes générateurs du sommeil et du rôle physiologique de l'activité électrique du cerveau. L'application qui nous concerne s'inscrit dans ce courant d'étude de la microstructure de l'électroencéphalogramme de sommeil puisqu'elle est relative à l'un de ses phénomènes transitoires : le complexe K. Cet événement nécessite le recours à des méthodes statistiques en raison de sa ressemblance marquée avec d'autres phénomènes non stationnaires observés en sommeil profond, parmi lesquels on recense les bouffées d'ondes delta. C'est en conséquence un problème de détection relativement difficile pour lequel aucune approche, jusqu'à maintenant, n'a pu être jugée satisfaisante en termes de performances. Toutefois, un certain nombre de travaux sur la non-stationnarité de l'électroencéphalogramme de sommeil ont permis de justifier l'intérêt des représentations temps-fréquence pour la caractérisation des phénomènes transitoires rencontrés [Cho87], [Sha93], [Bra96], [Cim97]. Aussi, on a jugé intéressant de recourir à cet espace de représentation pour l'élaboration d'un détecteur de complexes K, bien que le domaine d'application de la méthodologie proposée pour la synthèse de détecteurs à structure imposée ne se borne pas à celui-ci.

Ce mémoire est organisé ainsi :

- L'objectif du Chapitre 1 est de décrire, sans prétendre à l'exhaustivité, une catégorie particulière de représentations temps-fréquence : la classe de Cohen. En effet, devant le nombre de solutions envisageables, nous avons volontairement limité notre étude à ce vaste ensemble de représentations temps-fréquence non-paramétriques. Ainsi, après avoir exposé quelques propriétés de ces représentations, nous évoquons certaines difficultés

liées à leur mise en oeuvre. En particulier, nous posons le problème des termes interférentiels présents sur les représentations et inhérents à leur nature quadratique, puis nous exposons certaines solutions existantes telles que la méthode de réallocation.

- Le Chapitre 2 concerne les problèmes pratiques liés au calcul rapide des représentations, sujet relativement peu abordé si on considère que la mise au point d'algorithmes rapides a été élargi et continue d'élargir le champ des applications de l'analyse temps-fréquence. Dans le contexte des signaux de sommeil, le recours à de telles méthodes est impératif en raison du volume considérable de données à traiter.
- Dans le Chapitre 3, on présente brièvement la notion de détecteur à structure libre, en exposant plusieurs règles de décision, après avoir distingué le cas des hypothèses simples et composées. La suite du chapitre est alors uniquement consacrée à la définition d'une méthodologie pour l'élaboration d'un détecteur à structure imposée. Pour cela, on a notamment recours à des résultats de la théorie de l'apprentissage et de la reconnaissance des formes. L'efficacité de l'approche proposée est ensuite illustrée par la résolution de problèmes de détection dans le plan temps-fréquence.
- Afin de limiter les effets néfastes de la *malédiction de la dimensionnalité*, nous recherchons dans le Chapitre 4 une adéquation entre le nombre de paramètres caractéristique de la statistique de détection et le volume de données disponibles pour l'élaboration d'un test de détection. Pour cela, nous proposons diverses procédures d'optimisation dont on illustre l'efficacité par la résolution de problèmes de détection dans le plan temps-fréquence. Ces applications nécessitent en effet l'usage de telles techniques, les détecteurs obtenus à partir d'un ensemble d'apprentissage souffrant souvent d'un biais important en raison de la grande dimension de l'espace de représentation choisi. Finalement, on est amené, après certaines constatations expérimentales surprenantes, à se pencher sur la dimension effective de l'espace engendré par la distribution de Wigner-Ville discrète.
- Enfin, après une brève description de l'électroencéphalogramme de sommeil et du complexe K, on procède dans le dernier chapitre à la validation de la méthodologie proposée. L'accent est alors mis sur le fait que cette approche est immédiatement transposable à tout autre problème de détection de phénomènes transitoires.

Chapitre 1

Eléments d'analyse temps-fréquence

Les éléments de ce chapitre ont été principalement extraits de [Aug91], [Fla93] et [Coh95].

1.1 Introduction

Un signal peut être défini comme le support physique d'une information traduisant l'évolution d'un système. Les variations des grandeurs physiques caractérisant les états de ce système étant en général enregistrées séquentiellement, la *représentation temporelle* d'un signal constitue un espace de travail opportun. Toutefois, la *représentation fréquentielle* peut également offrir une alternative intéressante, en particulier pour l'analyse des systèmes linéaires invariants. En effet, on peut montrer que les fonctions exponentielles complexes e^{-pt} , $p \in \mathcal{C}$, sont fonctions propres de tout opérateur vérifiant ces propriétés. Cette représentation est obtenue au moyen de la transformée de Fourier qui, si elle existe, est définie ainsi pour les signaux continus :

$$X(\nu) = \int x(t) e^{-j2\pi\nu t} dt. \quad (1.1)$$

La représentation fréquentielle au sens de Fourier d'un signal repose sur la projection de celui-ci sur une base de fonctions mono-chromatiques à support infini. Si cette représentation fournit des informations sur les composantes fréquentielles du signal étudié, leur amplitude et leur phase, il en résulte cependant une perte d'information quant à leur localisation temporelle. Ceci peut s'avérer handicapant pour l'analyse des signaux non-stationnaires, dont les propriétés statistiques dépendent précisément du temps.

Signaux stationnaires et non-stationnaires

Un signal déterministe sera dit *stationnaire* s'il peut s'écrire sous la forme d'une somme discrète de fonctions sinusoïdales [Fla93] :

$$x(t) = \sum_{k \in \mathcal{N}} A_k \cos(2\pi\nu_k t + \phi_k) \quad \text{pour un signal réel ;} \quad (1.2)$$

$$x(t) = \sum_{k \in \mathcal{A}} A_k \exp j(2\pi\nu_k t + \phi_k) \quad \text{pour un signal complexe.} \quad (1.3)$$

Bien qu'il soit possible de définir des conditions plus fortes de stationnarité pour les signaux aléatoires [Duv91], on peut adopter en pratique la notion de stationnarité au sens large qui permet de traiter une large majorité des problèmes de façon satisfaisante. Ainsi, un signal sera dit *stationnaire au sens large*, ou encore *stationnaire jusqu'à l'ordre 2*, si son espérance mathématique est indépendante du temps et si sa fonction d'autocorrélation $\mathbf{E}\{x(t_1) x^*(t_2)\}$ dépend uniquement de $(t_2 - t_1)$. En conséquence, un signal sera dit *non-stationnaire* s'il ne vérifie pas l'une de ces propriétés.

Grandeurs locales

Bien que le spectre obtenu par la transformée de Fourier caractérise le comportement d'un signal sur toute sa durée, certaines grandeurs locales combinant des informations de nature temporelle et fréquentielle en ont été dérivées : la *fréquence instantanée* et le *retard de groupe*. Avant de présenter ces concepts, il convient de définir au préalable la notion de signal analytique.

A tout signal réel $x(t)$, on peut associer le signal complexe $z_x(t)$ suivant :

$$z_x(t) \equiv x(t) + j \cdot \mathcal{H}\{x(t)\}, \quad (1.4)$$

où $\mathcal{H}\{x\}$ désigne la transformée de Hilbert de x . Le signal z_x est appelé *signal analytique* associé à x . La définition énoncée ci-dessus admet une formulation plus explicite dans le domaine fréquentiel puisque :

$$Z_x(\nu) = 2 \cdot U(\nu) \cdot X(\nu), \quad (1.5)$$

où X et Z_x désignent respectivement les transformées de Fourier de x et z_x . La fonction U représente l'échelon unitaire d'Heaviside. Ainsi, le signal analytique z_x peut être obtenu par la transformée de Fourier inverse de X auquel on a préalablement annulé les composantes correspondant aux fréquences négatives. On notera que cette transformation n'affecte en aucun cas le contenu informationnel de x dans la mesure où la transformée de Fourier d'un signal réel vérifie la relation de symétrie hermitienne suivante :

$$X^*(\nu) = X(-\nu). \quad (1.6)$$

En considérant le signal analytique sous sa forme polaire :

$$z_x(t) = |z_x(t)| e^{j\varphi_x(t)}, \quad (1.7)$$

on définit la *fréquence instantanée* $\nu_x(t)$ selon :

$$\nu_x(t) \equiv \frac{1}{2\pi} \frac{d \arg z_x}{dt}(t) = \frac{1}{2\pi} \frac{d\varphi_x}{dt}(t). \quad (1.8)$$

Cette grandeur locale donne une information sur le contenu spectral instantané d'un signal.

De façon duale, on peut souhaiter décrire le comportement temporel local d'un signal en fonction de la fréquence. En considérant la transformée de Fourier du signal analytique sous sa forme polaire :

$$Z_x(\nu) = |Z_x(\nu)| e^{j\psi_x(\nu)}, \quad (1.9)$$

on peut obtenir une information sur l'instant d'arrivée d'une fréquence grâce au *retard de groupe* :

$$t_x(\nu) \equiv -\frac{1}{2\pi} \frac{d \arg Z_x(\nu)}{d\nu} = -\frac{1}{2\pi} \frac{d\psi_x(\nu)}{d\nu}. \quad (1.10)$$

Cependant, la fréquence instantanée et le retard de groupe ne conduisent à des représentations aisément interprétables que dans le cas des signaux mono-composantes pour lesquels, à un instant donné (resp. fréquence donnée), le signal n'existe qu'au voisinage d'une fréquence donnée (resp. instant donné). Dans le cas contraire, qui concerne la grande majorité des signaux non-stationnaires, ces fonctions mono-dimensionnelles locales s'avèrent généralement inadéquates, comme le montre la Figure 1.1. Il peut alors être intéressant de représenter les signaux conjointement dans les domaines temporel et fréquentiel afin de caractériser leur évolution spectrale au cours du temps. L'objectif du présent chapitre est de décrire, sans prétendre à l'exhaustivité, une catégorie particulière de représentations temps-fréquence : la classe de Cohen. En effet, devant le nombre de solutions envisageables, nous avons volontairement limité notre étude à ce vaste ensemble de représentations temps-fréquence non-paramétriques.

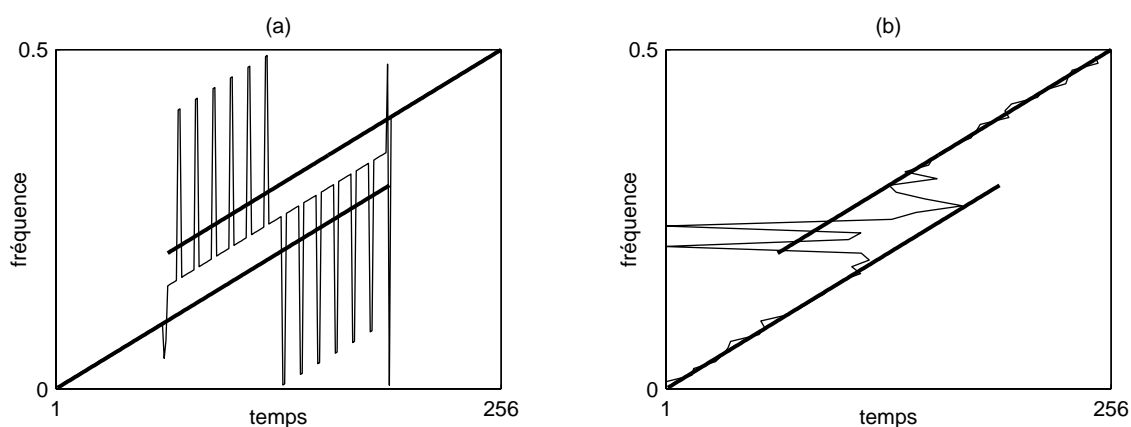


Figure 1.1 : Estimation de la *Fréquence instantanée* (a) et du *retard de groupe* (b) d'un signal constitué de deux composantes à modulation de fréquence linéaire, dont les lois figurent en gras sur les représentations.

Le Chapitre 1 est organisé ainsi. Dans un premier temps, nous introduisons la notion de représentation temps-fréquence par le biais de la décomposition d'un signal sur une famille de fonctions élémentaires bien localisées dans les domaines temporel et fréquentiel. Nous illustrons alors ce point de vue à l'aide de la transformée de Fourier à court terme et de la représentation de Gabor. Puis, nous sortons du cadre des représentations linéaires pour définir la classe de Cohen, qui réunit les distributions d'énergie covariantes par translations dans les domaines temporel et fréquentiel. Nous explorons alors les propriétés de ces représentations, et plus particulièrement celles du spectrogramme et de la distribution de Wigner-Ville. Enfin, nous abordons le problème des termes interférentiels présents sur les représentations temps-fréquence de la classe de Cohen et inhérents à leur nature quadratique.

1.2 Les représentations linéaires

1.2.1 La transformée de Fourier à court terme

a) *Définition.* Afin d'introduire une dépendance temporelle dans la transformée de Fourier, on remplace l'analyse spectrale globale par une succession d'analyses locales du signal observé au travers d'une fenêtre glissante $h(t)$. Cette représentation, usuellement appelée *transformée de Fourier à court terme*, est définie ainsi :

$$F_x(t, \nu; h) \equiv \int x(s) h^*(s-t) e^{-j2\pi\nu s} ds. \quad (1.11)$$

En général, la fenêtre $h(t)$ choisie est réelle, à support fini et centrée sur 0.

b) *Quelques propriétés.* Parmi les propriétés vérifiées par la transformée de Fourier à court terme, on peut mentionner que :

- La transformée de Fourier à court terme est covariante par translations en temps (à une modulation de phase près) et en fréquence :

$$y(t) = x(t-t_0) e^{-j2\pi\nu_0 t} \Rightarrow F_y(t, \nu; h) = F_x(t-t_0, \nu-\nu_0; h) e^{-j2\pi\nu_0 t}. \quad (1.12)$$

- Le signal $x(t)$ peut être reconstruit à l'aide d'une fenêtre de synthèse $g(t)$ différente de la fenêtre d'analyse $h(t)$ grâce à la relation :

$$x(t) = \iint F_x(s, \nu; h) g(t-s) e^{j2\pi\nu t} ds d\nu, \quad (1.13)$$

à condition que g et h vérifient :

$$\int g(t) h^*(t) dt = 1. \quad (1.14)$$

- La transformée de Fourier à court terme ne permet pas une localisation arbitrairement précise d'un événement dans les domaines temporel et fréquentiel. En effet, l'augmentation de la résolution temporelle (resp. fréquentielle) de cette représentation nécessite une fenêtre d'analyse mieux localisée en temps (en fréquence), ce qui a pour effet de dégrader sa résolution fréquentielle (resp. temporelle). Par conséquent, le choix de cette fenêtre repose sur la recherche d'un compromis entre l'hypothèse de quasi-stationnarité du signal sur l'intervalle d'analyse et les résolutions temporelle et fréquentielle de la représentation.

Pour un complément d'information sur la transformée de Fourier à court terme, le lecteur intéressé pourra consulter [Fla93] et [Coh95].

1.2.2 La représentation de Gabor

Dans le cas discret, la relation (1.13) de synthèse de la transformée de Fourier à court terme permet de définir la *représentation de Gabor* [Gab46] :

$$x(t) = \sum_n \sum_m F_x(nt_0, mv_0; h) g_{n,m}(t), \quad (1.15)$$

avec $g_{n,m}(t) = g(t - nt_0) e^{j2\pi mv_0 t}$, où $g(t)$ est généralement une fonction gaussienne.

Ainsi, les coefficients $F_x(nt_0, mv_0; h)$, appelés *coefficients de Gabor*, sont porteurs d'une information relative au contenu temps-fréquence du signal x au voisinage du point (nt_0, mv_0) . Les atomes $g_{n,m}(t)$, appelés *logons de Gabor*, sont associés dans le plan temps-fréquence à une maille rectangulaire d'aire unité centrée sur (nt_0, mv_0) .

Bien que séduisante, cette représentation linéaire est peu utilisée en pratique en raison des difficultés théoriques et pratiques qui accompagnent sa mise en oeuvre [Bas80], [Bal81].

1.3 Les représentations d'énergie

1.3.1 Motivations

Par rapport aux représentations linéaires qui offrent une décomposition des signaux analysés sur un ensemble de fonctions élémentaires, l'objectif des représentations quadratiques est de distribuer leur énergie conjointement dans les domaines temporel et fréquentiel. Cette perspective a été étudiée en premier lieu par Ville dans [Vil48], où celui-ci recherche une fonction $RTF_x(t, \nu)$, dépendant respectivement du temps et de la fréquence, et vérifiant les contraintes de distributions marginales définies ci-après. Notons que l'auteur semble s'être inspiré du formalisme de la mécanique quantique, sans pour autant y faire

référence et notamment établir l'équivalence entre le fruit de ses travaux et la fonction (position-impulsion) de Wigner [Wig32].

- Energie

L'énergie E_x du signal doit pouvoir être retrouvée par intégration de la représentation sur l'ensemble du plan temps-fréquence :

$$E_x = \iint RTF_x(t, \nu) dt d\nu. \quad (1.16)$$

- Marginales

La puissance instantanée et la densité spectrale d'énergie doivent correspondre aux distributions marginales temporelle et fréquentielle de la représentation :

$$\int RTF_x(t, \nu) d\nu = |x(t)|^2 \quad \text{et} \quad \int RTF_x(t, \nu) dt = |X(\nu)|^2. \quad (1.17)$$

Cependant, on peut imaginer d'autres contraintes en fonction d'un cahier des charges établi en termes d'interprétabilité de la représentation temps-fréquence, de cohérence avec les représentations temporelle et fréquentielle, ou encore de compatibilité avec les opérateurs classiques en traitement du signal tels que la modulation et la convolution [Aug91]. Voici une liste non-exhaustive des contraintes les plus couramment citées dans la littérature.

- Réalité

Dans un souci d'interprétation de la représentation obtenue, on peut exiger que :

$$RTF_x(t, \nu) \in \mathcal{R}, \quad \forall t, \quad \forall \nu. \quad (1.18)$$

De plus, on peut également souhaiter que la représentation soit positive, ce qui conférerait à RTF_x le statut de densité d'énergie.

- Covariance en translation

On peut imposer à RTF_x d'être covariante par translation dans le domaine temps-fréquence :

$$y(t) = x(t - t_0) e^{-j2\pi\nu_0 t} \Leftrightarrow RTF_y(t, \nu) = RTF_x(t - t_0, \nu - \nu_0). \quad (1.19)$$

- Conservation du support

La propriété de conservation du support temporel au sens large suppose que :

$$x(t) = 0, \quad |t| > T \Rightarrow RTF_y(t, \nu) = 0, \quad |t| > T, \quad \forall \nu. \quad (1.20)$$

Cette contrainte peut également être formulée en termes fréquentiels :

$$X(\nu) = 0, \quad |\nu| > B \Rightarrow RTF_y(t, \nu) = 0, \quad |\nu| > B, \quad \forall t. \quad (1.21)$$

- Inversibilité

L'inversibilité caractérise la possibilité de reconstruire x à partir de RTF_x . Cette propriété, qui garantit la conservation du contenu informationnel du signal, peut s'avérer fondamentale dans le contexte de la détection par représentation temps-fréquence.

- Unitarité

Il peut être souhaitable que le produit scalaire entre signaux soit conservé par changement d'espace de représentation :

$$\iint RTF_x(t, \nu) RTF_y^*(t, \nu) dt d\nu = \left| \int x(t) y^*(t) dt \right|^2. \quad (1.22)$$

Cette relation d'unitarité est appelée usuellement *formule de Moyal*.

Pour une énumération plus complète ou une présentation plus détaillée de ces propriétés, on se référera à [Fla93].

1.3.2 Classe de Cohen

a) *Définition*. Parmi les propriétés citées précédemment, le principe de covariance est fondamental. Il garantit en effet que le décalage temporel et la modulation d'un signal s'accompagnent de translations de sa représentation dans le plan temps-fréquence. L'ensemble des distributions quadratiques possédant cette propriété porte le nom de *Classe de Cohen*.

Les distributions de la classe de Cohen peuvent s'exprimer ainsi [Coh66], [Coh95] :

$$C_x(t, \nu; \phi) \equiv \iint \phi_{TR}(s-t, \tau) x\left(s + \frac{\tau}{2}\right) x^*\left(s - \frac{\tau}{2}\right) e^{-j2\pi\nu\tau} ds d\tau. \quad (1.23)$$

ϕ_{TR} désigne la *fonction de paramétrisation*, ou encore le *noyau* de la représentation, exprimée dans le domaine temps-retard (TR). On suppose que cette fonction est indépendante du signal de façon à garantir la nature quadratique de C_x .

Cette formulation n'est pas unique. En effet, on peut montrer que toute distribution temps-fréquence de la classe de Cohen peut également s'écrire comme la transformée de Fourier bidimensionnelle de la *fonction d'ambiguïté à bande étroite* A_x , pondérée par la fonction de paramétrisation exprimée dans le domaine Doppler-retard (DR) :

$$C_x(t, \nu; \phi) \equiv \iint \phi_{DR}(\xi, \tau) A_x(\xi, \tau) e^{-j2\pi(\xi t + \nu\tau)} d\xi d\tau. \quad (1.24)$$

On rappelle que la fonction d'ambiguïté A_x à bande étroite est définie par :

$$A_x(\xi, \tau) \equiv \int x\left(s + \frac{\tau}{2}\right) x^*\left(s - \frac{\tau}{2}\right) e^{j2\pi\xi s} ds, \quad (1.25)$$

De plus, ϕ_{DR} peut être évaluée par transformée de Fourier de ϕ_{TR} :

$$\phi_{DR}(\xi, \tau) = \int \phi_{TR}(t, \tau) e^{-j2\pi\xi t} dt.$$

En appliquant la transformée de Fourier à chaque terme de la définition (1.24), on obtient :

$$C_x(t, \nu; \phi) \equiv \phi_{TF}(t, \nu) * W_x(t, \nu) = \iint \phi_{TF}(s-t, \xi-\nu) W_x(s, \xi) ds d\xi, \quad (1.26)$$

où ϕ_{TF} est la fonction de paramétrisation exprimée dans le domaine temps-fréquence (TF). W_x , qui désigne la distribution de Wigner-Ville, sera définie dans la section 1.3.3. On notera que cette dernière définition permet d'interpréter chaque représentation de la classe de Cohen comme le résultat d'une opération de filtrage de la distribution de Wigner-Ville.

b) Propriétés. L'intérêt des différentes formulations présentées au paragraphe précédent réside dans le fait qu'elles permettent de traduire les propriétés des distributions de la classe de Cohen en conditions d'admissibilité de la fonction de paramétrisation. Ceci peut s'avérer utile pour l'étude d'une représentation particulière ou pour la caractérisation d'une classe de représentations devant répondre à un cahier des charges donné [Aug91]. A titre d'exemple, les contraintes sur le noyau associées aux propriétés énumérées dans la section 1.3.1 sont regroupées dans le Tableau 1.1. Pour la démonstration de chacune de ces assertions, on se reportera à [Fla93].

Propriétés de C_x	Contraintes sur ϕ_{DR}
Energie	$\phi_{DR}(0, 0) = 1$
Marginale temporelle	$\phi_{DR}(\xi, 0) = 1 \quad \forall \xi$
Marginale fréquentielle	$\phi_{DR}(0, \tau) = 1 \quad \forall \tau$
Réalité	$\phi_{DR}(\xi, \tau) = \phi_{DR}^*(-\xi, -\tau) \quad \forall \xi \quad \forall \tau$
Support temporel au sens large	$2 t > \tau \Rightarrow \int \phi_{DR}(\xi, \tau) e^{j2\pi\xi t} d\xi = 0$
Support fréquentiel au sens large	$2 \nu > \xi \Rightarrow \int \phi_{DR}(-\xi, \tau) e^{-j2\pi\nu\tau} d\tau = 0$
Inversibilité	$\phi_{DR}(\xi, \tau) \neq 0 \quad \forall \xi \quad \forall \tau$
Unitarité	$ \phi_{DR}(\xi, \tau) = 1 \quad \forall \xi \quad \forall \tau$

Tableau 1.1 : Propriétés de la représentation temps-fréquence et contraintes sur le noyau

1.3.3 Exemples

Outre les représentations dont les définitions ont précédé historiquement celle de la classe de Cohen et qui y ont été intégrées *a posteriori* (spectrogramme, distributions de Wigner-Ville, Page, Margenau-Hill, Born-Jordan, etc.), de nombreuses autres distributions ont été proposées depuis (distributions de Choi-Williams, Zhao-Atlas-Marks, etc.). Nous allons maintenant étudier en particulier le spectrogramme et la distribution de Wigner-Ville. Pour une définition des autres représentations citées et une analyse complète de leurs propriétés, on pourra se référer à [Fla93].

a) *Spectrogramme*. Le module au carré de la transformée de Fourier à court terme conduit à une représentation quadratique communément appelée spectrogramme. Cette représentation, qui distribue l'énergie du signal dans le plan temps-fréquence, est la plus ancienne qui ait été proposée, mais aussi l'une des plus utilisées. Elle est précisément définie ainsi :

$$SP_x(t, \nu; h) \equiv \left| \int x(\tau) h^*(\tau - t) e^{-j2\pi\nu\tau} d\tau \right|^2, \quad (1.27)$$

où h désigne une fenêtre d'analyse. Comme pour la transformée de Fourier à court terme, le choix de la fenêtre repose sur la recherche d'un compromis entre l'hypothèse de quasi-stationnarité du signal sur l'intervalle d'analyse et les résolutions temporelle et fréquentielle de la représentation.

Comme élément de la classe de Cohen, le spectrogramme est définie par une fonction de paramétrisation :

$$\phi_{DR}(\xi, \tau) = A_h^*(\xi, \tau), \text{ soit encore } \phi_{TF}(t, \nu) = W_h(t, \nu).$$

En conséquence, on peut montrer que les seules propriétés que possède cette représentation, en plus de la covariance par translations dans le plan temps-fréquence, sont : la conservation de l'énergie sous réserve que h soit d'énergie unité, la réalité et la positivité.

b) *Distribution de Wigner-Ville*. La distribution de Wigner-Ville peut être définie à partir de la représentation temporelle du signal :

$$W_x(t, \nu) \equiv \int x\left(t + \frac{\tau}{2}\right) x^*\left(t - \frac{\tau}{2}\right) e^{-j2\pi\nu\tau} d\tau, \quad (1.28)$$

ou, de manière duale, à partir de sa transformée de Fourier :

$$W_x(t, \nu) \equiv \int X\left(\nu - \frac{\xi}{2}\right) X^*\left(\nu + \frac{\xi}{2}\right) e^{-j2\pi t\xi} d\xi. \quad (1.29)$$

Cette distribution correspond à l'élément de la classe de Cohen pour lequel $\phi_{DR}(\xi, \tau) = 1$. En conséquence, elle vérifie les propriétés suivantes : conservation de l'énergie et des

supports au sens large, distributions marginales temporelle et fréquentielle, réalité, covariance en translation et unitarité. En revanche, elle n'est pas partout non-négative, ce qui limite l'analogie avec une distribution d'énergie.

1.3.4 Interférences

L'une des conséquences directes du caractère quadratique des distributions de la classe de Cohen est la présence de *composantes interférentielles* sur les représentations temps-fréquence, en des lieux où la puissance instantanée et la densité spectrale d'énergie du signal analysé peuvent être nulles. Ceci résulte du *principe de superposition quadratique*, qui implique que la représentation d'un signal constitué d'une somme de composantes élémentaires n'est pas la somme des représentations de chacune de ses composantes :

$$\text{si } x(t) = \sum_{n=1}^N x_n(t), \text{ alors } C_x(t, \nu; \phi) = \sum_{n=1}^N C_{x_n}(t, \nu; \phi) + 2 \operatorname{Re} \left\{ \sum_{n=1}^N \sum_{m=n+1}^N C_{x_n, x_m}(t, \nu; \phi) \right\}$$

$$\text{où } C_{x_n, x_m}(t, \nu; \phi) = \iiint e^{j2\pi\xi(s-t)} \phi_{DR}(\xi, \tau) x_n\left(s + \frac{\tau}{2}\right) x_m^*\left(s - \frac{\tau}{2}\right) e^{-j2\pi\nu\tau} d\xi ds d\tau.$$

L'usage des distributions de la classe de Cohen comme outil d'analyse nécessite de savoir distinguer les composantes interférentielles sur une représentation temps-fréquence. Leurs caractéristiques sont désormais bien connues et font l'objet d'une étude exhaustive dans [Aug91] et [Hla97]. En particulier, la géométrie des interférences de la représentation de Wigner-Ville peut se résumer ainsi, comme l'illustre la Figure 1.2 :

- Deux composantes d'un signal interagissent en créant des interférences localisées au voisinage du milieu géométrique du segment les joignant.
- Les interférences ont un caractère oscillant, de fréquence proportionnelle à la distance entre les composantes en interaction.
- L'axe d'oscillation des interférences est perpendiculaire à la droite joignant les composantes interagissantes.

Selon l'usage que l'on souhaite faire des représentations temps-fréquence, les interférences peuvent s'avérer utiles ou au contraire gênantes.

Faisant partie de la distribution, les termes interférentiels sont porteurs d'une information, en particulier de phase [Fla93], ce qui peut les rendre indispensables pour la détection ou la classification de signaux. De plus, leur présence contribue au respect de propriétés telles que l'unitarité.

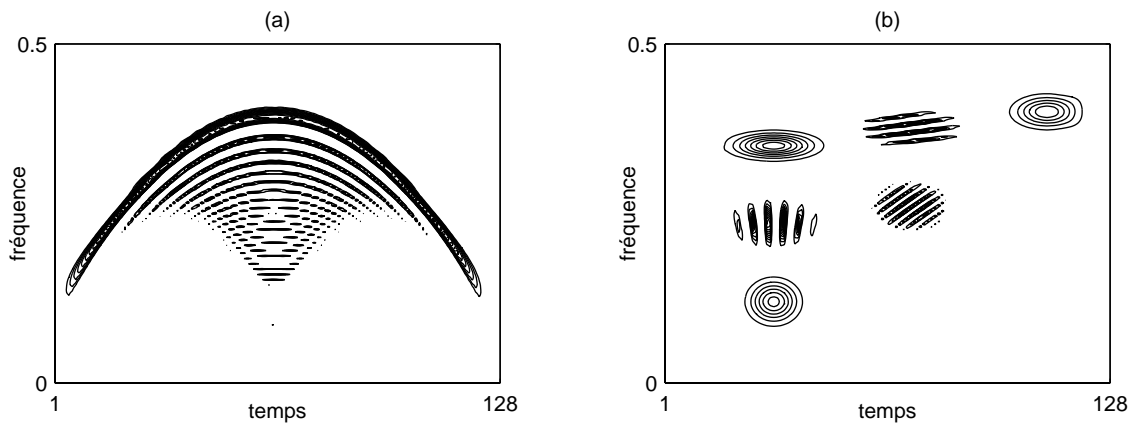


Figure 1.2 : Interférences de la représentation de Wigner-Ville. Exemple d'un signal à une composante à modulation de fréquence parabolique (a) et d'un signal à composantes gaussiennes multiples (b).

En revanche, leur importance en nombre et en amplitude fait que les interférences perturbent l'analyse des signaux à composantes multiples ou à modulation de fréquence non-linéaire. Pour cette raison, de nombreuses approches visant à les atténuer ont été proposées.

1.3.5 Lissage et distributions à interférences réduites

La lisibilité des représentations temps-fréquence étant un facteur essentiel pour l'analyse des signaux non-stationnaires, il peut être souhaitable de diminuer le nombre et l'amplitude des interférences. Une première étape consiste à ne représenter que des signaux analytiques afin d'éliminer les termes interférentiels provenant des interactions entre les composantes portées par les fréquences positives et celles portées par les fréquences négatives. On rappelle à cet effet qu'à tout signal réel, on peut associer un signal analytique en utilisant la transformation (1.4). Une seconde étape consiste alors à exploiter la structure oscillante des interférences, ce qui suggère d'introduire une opération de *lissage* dans le plan temps-fréquence. Appliqué à la représentation de Wigner-Ville, ce filtrage conduit à l'évaluation de l'expression suivante :

$$\phi_{TF}(t, \nu) * W_x(t, \nu),$$

où l'on reconnaît la définition générale (1.26) des distributions de la classe de Cohen. Pour le choix de ϕ_{TF} , différentes alternatives sont envisageables.

a) *Lissage fixe*. La solution la plus naturelle consiste à utiliser un filtre passe-bas bi-dimensionnel indépendant du signal analysé. Un certain nombre de représentations repose sur ce principe, parmi lesquelles on compte celle de Choi-Williams, Born-Jordan et Zhao-Atlas-Marks [Fla93]. On citera enfin l'exemple typique que constitue la distribution *pseudo Wigner-Ville lissée* :

$$PWL_x(t, \nu) \equiv \iint h(t - \tau) G(\nu - \xi) W_x(\tau, \xi) d\tau d\xi. \quad (1.30)$$

Dans ce cas, la séparabilité du filtre bi-dimensionnel utilisé permet de contrôler le lissage indépendamment en temps et en fréquence, et par conséquent la résolution de la représentation selon chacun de ces axes.

b) *Lissage adapté*. Une seconde solution consiste à paramétrer une fonction de lissage et à l'optimiser au sens d'un critère donné dépendant du signal à analyser. On peut alors distinguer deux approches :

- L'approche *globale*, qui consiste à utiliser la même fonction de lissage sur l'ensemble de la représentation [Bar91], [Bar93a], [Bar93b].
- L'approche *locale*, pour laquelle la fonction de lissage dépend du signal, mais également du temps et/ou de la fréquence [Gon98], [Jon90], [Jon95a], [Oeh97].

Il convient de noter que la méthode de Jones et Baraniuk [Jon95a], qui est certainement la procédure locale la plus performante de sa catégorie, est une extension de l'approche globale présentée dans [Bar93a]. Pour cela, les auteurs ont introduit une dépendance temporelle dans le processus d'optimisation originel de la fonction de lissage. Ils ont de plus proposé un *algorithme récursif* permettant de réduire significativement le temps de calcul de la représentation. Comme nous le verrons dans la section 2.4, nous avons également adopté une approche récursive pour l'implémentation rapide de la méthode de réallocation, dont le principe va être maintenant exposé.

1.3.6 La méthode de réallocation

Les distributions présentées précédemment constituent des solutions assez satisfaisantes pour l'amélioration de la lisibilité des représentations de la classe de Cohen. Cependant, quelle que soit l'approche retenue, l'opération de filtrage destinée à atténuer les composantes interférentielles se traduit en contrepartie par une diminution de la résolution dans le plan temps-fréquence. Certains travaux visant à limiter cette perte ont permis de dégager une méthodologie générale, la *méthode de réallocation*, dont nous exposons maintenant le principe.

a) *Réallocation du spectrogramme*. Avant d'être étendue à l'ensemble des représentations de la classe de Cohen [Aug95], la méthode de réallocation a été proposée dans le but d'améliorer les propriétés de localisation du spectrogramme [Kod78].

Le principe de la réallocation repose sur la formulation suivante du spectrogramme :

$$SP_x(t, \nu; h) \equiv \iint W_h(t-s, \nu-\xi) W_x(s, \xi) ds d\xi. \quad (1.31)$$

Soit C un point du plan temps-fréquence de coordonnées (t, ν) . Soit S le support de W_h centré en ce point. L'expression (1.31) montre que l'évaluation du spectrogramme en C consiste à affecter à ce point l'énergie moyenne (pondérée) du signal confinée dans S . Si ce lissage conduit à une atténuation des termes interférentiels, il entraîne en revanche un étalement de toutes les composantes du signal dans le plan temps-fréquence. Afin de pallier cet inconvénient, Kodera *et col.* ont suggéré de changer de point d'affectation en attribuant cette énergie moyenne au centre de gravité G des contributions énergétiques. Ce principe est analogue à celui couramment utilisé en mécanique qui consiste à affecter la masse totale d'un système à son centre de gravité et non à son centre géométrique. Dans le cadre de la méthode de réallocation, les coordonnées de G sont définies ainsi :

$$\begin{aligned} \hat{t}(t, \nu; x) &\equiv t - \frac{\iint s W_h(t-s, \nu-\xi) W_x(s, \xi) ds d\xi}{\iint W_h(t-s, \nu-\xi) W_x(s, \xi) ds d\xi}, \\ \hat{\nu}(t, \nu; x) &\equiv \nu - \frac{\iint \xi W_h(t-s, \nu-\xi) W_x(s, \xi) ds d\xi}{\iint W_h(t-s, \nu-\xi) W_x(s, \xi) ds d\xi}. \end{aligned} \quad (1.32)$$

Ce principe nous conduit ainsi à la définition du *spectrogramme réalloué*, dont la valeur en tout point est la somme des valeurs réaffectées à ce point :

$$SPR_x(t, \nu; h) \equiv \iint SP_x(t', \nu'; h) \delta(t - \hat{t}(t', \nu'; x)) \delta(\nu - \hat{\nu}(t', \nu'; x)) dt' d\nu'. \quad (1.33)$$

Les opérateurs de réallocation définis en (1.32) ne permettant pas une implémentation efficace du processus de réallocation, on préférera la formulation proposée par Auger et Flandrin [Aug95], qui nécessite seulement l'évaluation de deux transformées de Fourier à court terme supplémentaires :

$$\begin{aligned} \hat{t}(t, \nu; x) &= t - \mathbf{Re} \left\{ \frac{F_x(t, \nu; Th) F_x^*(t, \nu; h)}{|F_x(t, \nu; h)|^2} \right\}, \\ \hat{\nu}(t, \nu; x) &= \nu - \mathbf{Im} \left\{ \frac{F_x(t, \nu; Dh) F_x^*(t, \nu; h)}{|F_x(t, \nu; h)|^2} \right\}, \end{aligned}$$

où $Th(t) = t \cdot h(t)$ et $Dh(t) = dh(t)/dt$.

b) *Réallocation des distributions de la classe de Cohen.* Par analogie avec le spectrogramme et en partant de la définition (1.26), Auger et Flandrin ont généralisé le principe de la réallocation à l'ensemble des représentations temps-fréquence de la classe de Cohen [Aug95] :

$$CR_x(t, \nu; \phi) \equiv \iint C_x(t', \nu'; \phi) \delta(t - \hat{t}(t', \nu'; x)) \delta(\nu - \hat{\nu}(t', \nu'; x)) dt' d\nu'. \quad (1.34)$$

avec :

$$\hat{t}(t, \nu; x) = t - \frac{\int \int s \phi_{TF}(s, \xi) W_x(t-s, \nu-\xi) ds d\xi}{\int \int \phi_{TF}(s, \xi) W_x(t-s, \nu-\xi) ds d\xi},$$

$$\hat{\nu}(t, \nu; x) = \nu - \frac{\int \int \xi \phi_{TF}(s, \xi) W_x(t-s, \nu-\xi) ds d\xi}{\int \int \phi_{TF}(s, \xi) W_x(t-s, \nu-\xi) ds d\xi}.$$

De plus, ils ont également proposé une formulation des opérateurs de réallocation permettant une implémentation rapide de la procédure :

$$\hat{t}(t, \nu; x) = t - \frac{C_x(t, \nu; T\phi)}{C_x(t, \nu; \phi)}, \quad (1.35)$$

$$\hat{\nu}(t, \nu; x) = \nu + j \frac{C_x(t, \nu; D\phi)}{C_x(t, \nu; \phi)}, \quad (1.36)$$

avec $T\phi_{TR}(t, \tau) = t \cdot \phi_{TR}(t, \tau)$ et $D\phi_{TR}(t, \tau) = \partial \phi_{TR}(t, \tau) / \partial \tau$.

On peut montrer que le processus de réallocation préserve quelques propriétés intéressantes de la classe de Cohen. En effet, les représentations obtenues sont covariantes par translations dans le plan temps-fréquence et conservent l'énergie. De plus, elles permettent une localisation parfaite des modulations de fréquence linéaires et des impulsions. On notera que peu de représentations temps-fréquence vérifient cette dernière propriété. Au sein de la classe de Cohen en particulier, celle-ci est uniquement satisfaite par la représentation de Wigner-Ville [Aug95].

c) Exemples. Afin d'illustrer l'efficacité de la méthode de réallocation pour la représentation des signaux constitués de plusieurs composantes à bande étroite, un exemple est présenté en Figure 1.3. L'amélioration de la lisibilité apportée par cette méthode y est incontestable.

1.4 Conclusion

Au cours de ce chapitre, quelques rappels théoriques sur les représentations temps-fréquence ont été effectués. Ainsi, après avoir considéré certaines décompositions atomiques, qui sont des représentations linéaires du signal, nous avons étudié les représentations d'énergie de la classe de Cohen. Puis, nous avons posé le problème des interférences et enfin, certaines solutions comme la méthode de réallocation ont été évoquées. Cependant, les problèmes pratiques liés à l'évaluation de ces représentations, et en particulier au volume de calculs souvent prohibitif qu'induit l'analyse temps-fréquence d'un signal long, ont été laissés en suspens. Le prochain chapitre va donc être consacré à ces questions.

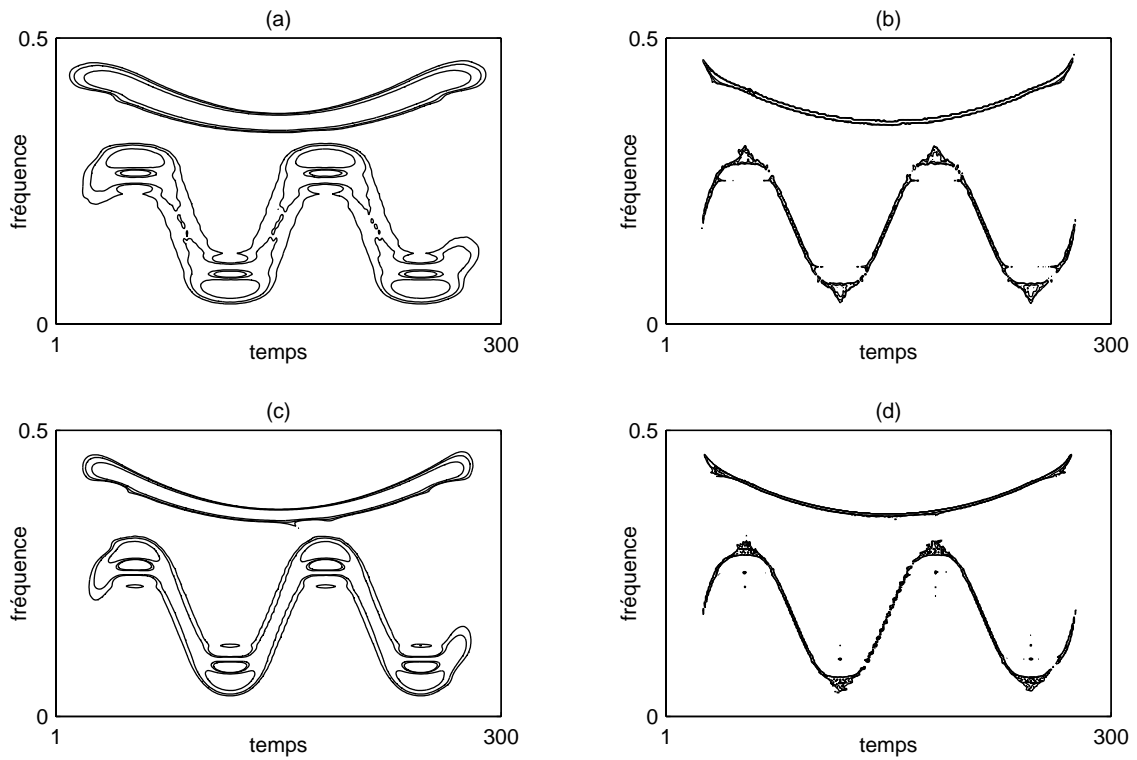


Figure 1.3 : Représentation temps-fréquence d'un signal constitué de deux composantes à bande étroite. (a) *spectrogramme*, (b) *spectrogramme réalloué*, (c) *distribution pseudo Wigner-Ville lissée*, (d) *distribution pseudo Wigner-Ville lissée réallouée*. Dans chacun de ces cas, les lignes de niveau représentent $1/5$, $1/10$ et $1/15$ de la valeur maximale de la représentation.

Chapitre 2

Implémentation récursive des représentations temps-fréquences

Le Chapitre 2 concerne les problèmes pratiques liés au calcul des représentations, sujet relativement peu abordé si on considère que la mise au point d'algorithmes économiques, rapides et élégants a élargi et continue d'élargir le champ des applications de l'analyse temps-fréquence. Dans ce domaine, on peut recenser deux approches.

La première vise à limiter la redondance des calculs. Pour cela, certains auteurs recherchent un ordonnancement optimal des opérations [Sun85], [Cun92], [Bar92]. D'autres en revanche exploitent les propriétés de symétrie de la fonction d'autocorrélation instantanée et de certaines fonctions de paramétrisation [Mar85], [Boa87], [Zie91]. Enfin, les distributions de la classe de Cohen peuvent faire l'objet d'une décomposition en une somme pondérée de spectrogrammes, seuls les termes les plus significatifs étant retenus [Cun94a], [Cun94b]. A l'exception de cette dernière approche, qui de plus n'aboutit qu'à une approximation de la représentation, les procédés cités conduisent à une réduction substantielle du temps de calcul. Pour cette raison, certains d'entre eux sont systématiquement associés aux algorithmes à stratégie récursive qui constituent la seconde classe de solutions que l'on a recensée.

Cette seconde approche consiste précisément à déterminer $RTF_x[n, \nu]$ à partir de $RTF_x[n-1, \nu]$, $RTF_x[\cdot, \nu]$ désignant une distribution temps-fréquence à temps discret. On conçoit aisément que toutes les représentations ne peuvent résulter d'un calcul de cette nature. Il apparaît ainsi dans [Hos80], [Ami87], [Rou93], [Tom96] et [Ric96b] que seules certaines fenêtres d'analyse autorisent l'évaluation récursive de la transformée de Fourier à court terme, sans qu'aucune condition nécessaire et suffisante d'admissibilité ne soit établie. Dans [Ami95], après avoir distingué les notions de récursivité directe et indirecte, qui feront l'objet d'un rappel au cours de ce chapitre, l'auteur définit une contrainte sur la fonction de paramétrisation afin que la distribution de la classe de Cohen associée satisfasse une relation récursive directe. Il expose également un exemple de formulation récursive indirecte, les

solutions de cette nature étant multiples [Ami96], [Ric97a]. De façon générale, tous ces travaux mettent l'accent sur la remarquable efficacité des algorithmes récursifs pour le calcul des représentations temps-fréquence. Ceci justifie l'intérêt dont ils font actuellement l'objet pour la mise en oeuvre de certains procédés complexes visant à limiter la présence des termes interférentiels dans les représentations [Jon95a], [Ric97a].

L'objectif de ce chapitre est de présenter de façon homogène et unifiée la propriété de récursivité de la transformée de Fourier à court terme F_x et de certains membres C_x de la classe de Cohen, en y incluant des contributions personnelles comme l'extension de cette propriété aux distributions modifiées par la méthode de réallocation. Ce chapitre est organisé ainsi. Après quelques rappels sur les notions de récursivité directe et indirecte [Uns83], nous établissons des conditions nécessaires et suffisantes d'admissibilité de la fenêtre d'analyse h et de la fonction de paramétrisation ϕ afin que les distributions $F_x[.;h)$ et $C_x[.;\phi)$ vérifient une relation récursive directe. Dans chacun de ces cas, nous proposons un exemple d'extension au cas récursif indirect et nous analysons l'efficacité des algorithmes correspondants. Enfin, nous proposons une stratégie récursive pour la mise en oeuvre du processus de réallocation de certaines représentations temps-fréquence.

2.1 Principes généraux de la récursivité

Considérons une séquence discrète $x[k]$, $k \in \mathcal{Z}$, observée au travers d'une fenêtre à court terme de largeur $2K+1$. A tout moment, la position de cette fenêtre est repérée par l'indice n de son échantillon central $x[n]$, comme le montre la Figure 2.1. Soit g une fonction dépendant des échantillons accessibles à un instant donné, que l'on note :

$$g[n] = \Gamma(x[n-K], \dots, x[n+K]), \quad (2.1)$$

où Γ est une fonction donnée.

Lorsqu'on déplace la fenêtre d'analyse de la position $n-1$ à la position n , on supprime $x[n-K-1]$ et on ajoute $x[n+K]$ à la liste des échantillons observables. Dans ce contexte, il est raisonnable d'envisager que, sous certaines conditions, $g[n]$ puisse être évaluée récursivement, c'est à dire à partir de $g[n-1]$:

$$g[n] = U(g[n-1], x[n-K-1], x[n+K]), \quad (2.2)$$

où U désigne une fonction de mise à jour à déterminer. Lorsque la fonction U ne dépend que des paramètres énumérés ci-dessus, la récursivité de g est qualifiée de *directe*. En revanche, si U ne dépend que de variables auxiliaires faisant l'objet d'une mise à jour récursive, la récursivité est dite *indirecte* [Uns83].

2.1.1 Récursivité directe

Nous allons maintenant énoncer une condition nécessaire et suffisante d'existence d'une relation de récursivité directe pour une fonction g donnée, en nous limitant au cas où la fonction de mise à jour U , présentée dans l'expression (2.2), dépend linéairement de $g[n-1]$.

Proposition 2.1. [Uns83]. Une fonction g de la variable n , dépendant des échantillons $x[n-K], \dots, x[n+K]$, vérifie la condition de récursivité directe :

$$g[n] = \alpha g[n-1] + \beta G(x[n+K]) + \gamma G(x[n-K-1]), \quad (2.3)$$

où α, β et γ sont des paramètres complexes et G une fonction arbitraire, si et seulement si :

$$g[n] = c \sum_{k=-K}^K \alpha^{-k} G(x[n+k]), \text{ avec } \beta = c \alpha^{-K} \text{ et } \gamma = -c \alpha^{K+1}. \quad (2.4)$$

Preuve. En partant de $g[n]$ et en appliquant $2K+1$ fois l'équation récursive (2.3), on obtient

$$g[n] = \alpha^{2K+1} g[n-(2K+1)] + \beta \sum_{i=0}^{2K} \alpha^i G(x[n+K-i]) + \gamma \sum_{i=0}^{2K} \alpha^i G(x[n-K-1-i]).$$

Après un changement de variables approprié, ce résultat peut être réécrit sous la forme suivante :

$$g[n] - \alpha^{2K+1} g[n-(2K+1)] = \beta \alpha^K \sum_{k=-K}^K \alpha^{-k} G(x[n+k]) + \gamma \alpha^K \sum_{k=-K}^K \alpha^{-k} G(x[n-(2K+1)+k]).$$

Puisque $g[n]$ et $g[n-(2K+1)]$ ne dépendent respectivement que des échantillons $x[n-K], \dots, x[n+K]$ et $x[n-3K-1], \dots, x[n-K-1]$, on en déduit que :

$$\begin{cases} g[n] = \beta \alpha^K \sum_{k=-K}^K \alpha^{-k} G(x[n+k]) \\ g[n-(2K+1)] = -\gamma \alpha^{-(K+1)} \sum_{k=-K}^K \alpha^{-k} G(x[n-(2K+1)+k]) \end{cases}$$

En comparant ces deux expressions, on trouve finalement que :

$$\beta \alpha^K = -\gamma \alpha^{-(K+1)} = c, \text{ où } c \text{ est une constante.} \quad \blacksquare$$

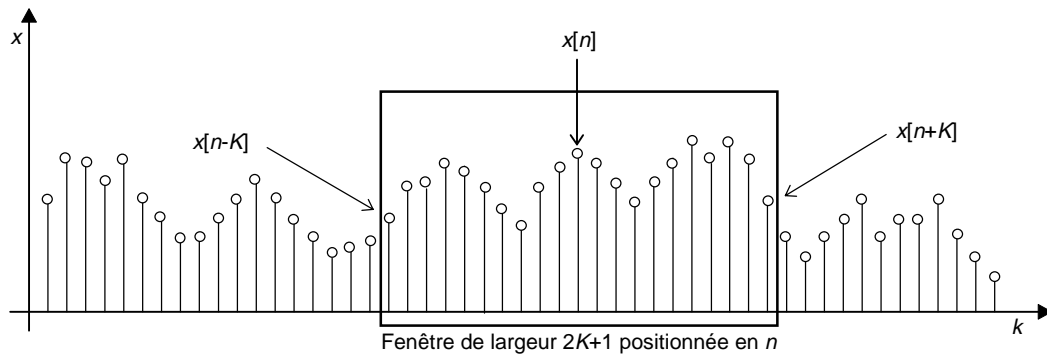


Figure 2.1 : Principe de l'analyse d'un signal au travers d'une fenêtre glissante de largeur $2K+1$.

Parmi les exemples les plus connus de récursivité directe, on peut citer celui du filtre à moyenne mobile :

$$m_x[n] \equiv \frac{1}{2K+1} \sum_{k=-K}^K x[n+k]. \quad (2.5)$$

En effet, m_x vérifie la condition (2.4) avec $G(x[n]) = x[n]$, $c = 1/(2K+1)$ et $\alpha = 1$, ce qui donne :

$$m_x[n] = m_x[n-1] + \frac{1}{2K+1} (x[n+K] - x[n-K-1]). \quad (2.6)$$

2.1.2 Récursivité indirecte

Il existe de nombreuses fonctions vérifiant la propriété de récursivité indirecte, parmi lesquelles figure la variance mobile :

$$\sigma_x^2[n] \equiv \frac{1}{2K+1} \sum_{k=-K}^K (x[n+k] - m_x[n])^2. \quad (2.7)$$

On peut en effet montrer que :

$$\sigma_x^2[n] = \frac{1}{2K+1} \sum_{k=-K}^K x^2[n+k] - m_{x^2}[n] = m_{x^2}[n] - m_x^2[n], \text{ où } m_{x^2} \text{ et } m_x \text{ vérifient la relation (2.6).}$$

Cependant, le cadre théorique de cette notion est moins bien défini que celui de la récursivité directe et il n'existe pas, à notre connaissance, de conditions nécessaires et suffisantes de récursivité indirecte. Le lecteur intéressé pourra néanmoins consulter [Uns83] pour de plus amples explications.

Dans la suite de ce chapitre, nous utiliserons la condition de récursivité directe (2.4) afin de présenter une formulation récursive de la transformée de Fourier à court terme et des distributions de la classe de Cohen. Puis, nous étendrons cette notion au processus de réallocation présenté dans la section 1.3.6.

2.2 La transformée de Fourier à court terme

2.2.1 Formulation récursive directe

En notant $x[nT_e]$, $n \in \mathcal{Z}$, un signal échantillonné à la cadence T_e que l'on prendra arbitrairement unitaire, on peut définir la transformée de Fourier à court terme d'un signal discret ainsi :

$$F_x[n, \nu; h] \equiv \sum_{k=-\infty}^{+\infty} x[n+k] h^*[k] e^{-j2\pi\nu k} = \sum_{k=1-N}^{N-1} x[n+k] h^*[k] e^{-j2\pi\nu k}, \quad (2.8)$$

où h désigne une fenêtre d'analyse glissante de largeur $2N-1$.

A chaque instant n , l'évaluation de F_x nécessite le calcul d'une transformée de Fourier, ce qui peut s'avérer coûteux en temps de calcul, en particulier lors de l'analyse de signaux longs. Ceci justifie notre intérêt pour une formulation récursive de l'expression (2.8), qui permettrait éventuellement de réduire le volume global des opérations à effectuer. En utilisant les notations et les résultats présentés dans la section 2.1, nous allons maintenant expliciter une condition nécessaire et suffisante d'admissibilité de la fenêtre d'analyse h , afin que F_x satisfasse la relation de récursivité directe (2.3). La proposition suivante complète les conditions suffisantes mentionnées dans [Hos80], [Ami87], [Rou93], [Tom96] et [Ric96b].

Proposition 2.2. La transformée de Fourier à court terme définie par l'expression (2.8) peut être formulée récursivement selon la relation (2.3) si et seulement si la fenêtre d'analyse h appartient à la classe générale $\mathcal{E}(F_x)$ suivante :

$$\mathcal{E}(F_x) = \{h : h^*[k] = c \rho^{-k}, k = 1 - N, \dots, N - 1, c \in \mathcal{C}, \rho \in \mathcal{C}\}, \quad (2.9)$$

Dans ce cas, la transformée de Fourier à court terme vérifie la relation :

$$F_x[n, \nu; h] = \rho e^{j2\pi\nu} F_x[n-1, \nu; h] + e^{-j2\pi\nu(N-1)} h^*[N-1] x[n+N-1] - \rho e^{j2\pi\nu N} h^*[1-N] x[n-N]. \quad (2.10)$$

Preuve. La comparaison des expressions (2.4) et (2.8) conduit à :

$$G(x[n+k]) = x[n+k] \text{ et } h^*[k] e^{-j2\pi\nu k} = c \alpha^{-k}, \text{ où } \alpha \text{ et } c \text{ sont des paramètres complexes.}$$

La fenêtre h et le paramètre c ne dépendant pas de ν , α est de la forme $e^{j2\pi\nu} \rho$, $\rho \in \mathcal{C}$. En conséquence, les fenêtres d'analyse autorisant une formulation récursive directe de la transformée de Fourier à court terme sont de la forme :

$$h^*[k] = c \rho^{-k}, k = 1 - N, \dots, N - 1, c \in \mathcal{C}, \rho \in \mathcal{C}. \quad (2.11)$$

L'identification des coefficients β et γ intervenant dans la relation (2.3) donne alors :

$$\alpha = \rho e^{j2\pi\nu}, \beta = h^*[N-1] e^{-j2\pi\nu(N-1)} \text{ et } \gamma = h^*[-N] e^{j2\pi\nu N} = \rho h^*[1-N] e^{j2\pi\nu N}. \quad \blacksquare$$

Remarque. Ce résultat nous conduit directement à une formulation récursive indirecte du spectrogramme puisque ce dernier est défini par le carré du module de F_x .

Afin d'illustrer la récursivité directe de la transformée de Fourier à court terme, supposons que h soit une fenêtre rectangulaire telle que $h[k] = 1$ si $k \in \{1 - N, \dots, N - 1\}$, et 0 sinon. Dans ce cas, on obtient :

$$F_x[n, \nu; h] = \left(F_x[n-1, \nu; h] + x[n+N-1] e^{-j2\pi\nu N} \right) e^{j2\pi\nu} - x[n-N] e^{j2\pi\nu N}, \text{ puisque } \rho = 1. \quad (2.12)$$

Ce cas particulier apparaît dans un certain nombre de travaux [Uns83], [Rou93], [Tom96].

2.2.2 Exemple de formulation récursive indirecte

L'objet de cette section est d'associer la récursivité de la transformée de Fourier à court terme à une classe de fenêtres d'analyse plus large que $\mathcal{E}(F_x)$. Pour cela, nous allons faire appel à la notion de récursivité indirecte en combinant plusieurs processus récursifs directs de type (2.10). Cette solution n'est cependant pas unique, des formulations plus spécifiques à certaines familles de fenêtres pouvant être établies [Tom96].

Ainsi, si la fenêtre d'analyse h appartient à la classe $\mathcal{E}_{\text{ext}}(F_x)$ définie ainsi :

$$\mathcal{E}_{\text{ext}}(F_x) = \left\{ h : h^*[k] = \sum_{p=1}^P c_p \rho_p^{-k}, k = 1-N, \dots, N-1, c_p \in \mathcal{C}, \rho_p \in \mathcal{C}, M \in \mathcal{N} \right\}, \quad (2.13)$$

la transformée de Fourier à court terme admet la formulation récursive indirecte suivante :

$$F_x[n, \nu; h] = \sum_{p=1}^P F_x[n, \nu; h_p], \text{ où } h_p^*[k] = c_p \rho_p^{-k}, \quad (2.14)$$

$F_x[n, \nu; h_p]$ étant évaluée récursivement grâce à la relation (2.10).

L'examen de $\mathcal{E}_{\text{ext}}(F_x)$ permet de constater que cette classe regroupe l'ensemble des fenêtres à court terme (périodisées) admettant une décomposition en série de Fourier finie. En particulier, les fenêtres de pondération semi-sinusoïdale ($P=2$), de Hamming et Hanning ($P=3$), et Blackman ($P=5$), très couramment utilisées en analyse spectrale, autorisent une formulation récursive indirecte de la transformée de Fourier à court terme [Ric97a].

2.2.3 Performances

Nous allons maintenant montrer que l'évaluation récursive de la transformée de Fourier à court terme nécessite souvent moins d'opérations que le calcul classique consistant à utiliser directement la définition (2.8). Pour cela, supposons par exemple que h soit une fenêtre rectangulaire de largeur $2N-1$, où N est une puissance de 2. Dans la suite de cette section, cette largeur sera systématiquement portée à $2N$ par ajout d'un échantillon nul, afin de permettre l'usage de la transformée de Fourier rapide lorsque cela est possible.

Calcul classique. L'évaluation de $F_x[n, k/2N; h]$, $k = 0, \dots, 2N-1$, résulte de la transformée de Fourier rapide d'un signal de $2N$ échantillons. Par conséquent, le calcul direct de la transformée de Fourier à court terme nécessite $2N \log_2(2N)$ additions complexes et $N \log_2 N$ multiplications complexes à chaque instant [Kun81]. Afin de pouvoir comparer ces résultats à ceux du calcul récursif, il est indispensable de convertir les opérations complexes en opérations réelles. Pour cela, on peut remarquer qu'une addition complexe (A_c) fait

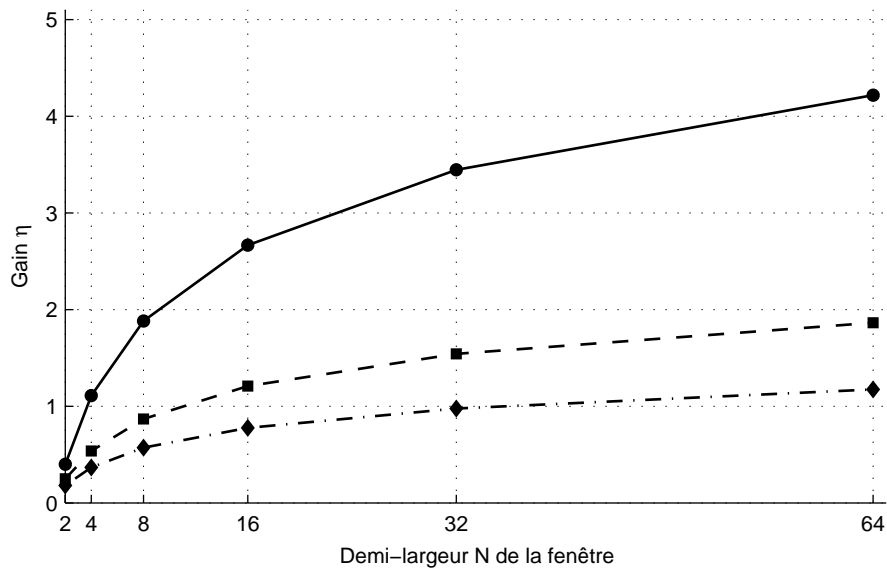


Figure 2.2 : Estimation du gain en temps de calcul lorsque l'approche récursive est préférée à la méthode classique pour l'évaluation de la *transformée de Fourier à court terme*. Différentes fenêtres d'analyse sont considérées : rectangulaire (●), semi-sinusoidale (■), Hamming et Hanning (◆).

intervenir deux additions réelles (A_R) et qu'une multiplication complexe (M_C) requiert quatre multiplications réelles (M_R) et deux additions réelles. Ainsi, le calcul direct de la transformée de Fourier à court terme nécessite, à chaque instant, le nombre d'opérations réelles suivant [Ric97a] :

$$A_R = 6N \log_2(N) + 4N \quad M_R = 4N \log_2(N).$$

Calcul récursif. Dans l'expression (2.12), l'évaluation du terme entre parenthèses nécessite $2N$ additions complexes. A ceci, il faut ajouter $2N$ multiplications et 1 addition complexes provenant respectivement du produit de ce terme par $\exp(jk\pi / N)$, $k = 0, \dots, 2N - 1$, et de la somme avec $x[n - N]$. Par conséquent, le volume global des opérations réelles mises en oeuvre par l'algorithme récursif est le suivant [Ric97a] :

$$A_R = 8N + 2 \quad M_R = 8N.$$

Ces résultats mettent en évidence la complexité linéaire de l'algorithme récursif, qui est par conséquent plus favorable que celle en $O(N \log_2 N)$ de l'approche classique. Cette conclusion reste valable lorsqu'on utilise d'autres fenêtres d'analyse, comme cela est indiqué en annexe de ce chapitre, page 51.

Afin de quantifier l'efficacité de l'approche récursive par rapport à la méthode classique, nous allons maintenant utiliser l'indicateur suivant, qui constitue une approximation du gain en temps de calcul [Kun81] :

$$\eta = \frac{OE_{classique}}{OE_{récursif}}, \quad (2.15)$$

où $OE_{classique}$ et $OE_{récursif}$ représentent le nombre d'opérations élémentaires requis par chacun des algorithmes. Nous rappelons qu'une *opération élémentaire* est classiquement définie comme une multiplication suivie d'une addition. Dans le cadre de cette étude, nous considérerons en première approximation que $OE \approx \max \{A_R ; M_R\}$ [Kun81]. Dans la Figure 2.2, l'indicateur (2.15) est représenté pour différentes fenêtres, en fonction de leur demi-largeur N . Ce résultat démontre clairement l'efficacité de l'approche récursive par rapport à la méthode classique, lorsque la largeur de la fenêtre excède une taille critique. De plus, on constate que le gain en temps de calcul dépend de N , mais également du nombre P de processus récursifs intervenant dans le calcul (2.14).

2.3 Distributions de la classe de Cohen

2.3.1 Formulation récursive directe

En temps discret, les distributions de la classe de Cohen peuvent s'exprimer ainsi :

$$C_x[n, \nu ; \phi] \equiv \sum_{l=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} \phi_{TR}[m, l] R_x[n+m, l] e^{-j4\pi\nu l}, \quad (2.16)$$

où $R_x[n+m, l] = x[n+m+l] x^*[n+m-l]$ désigne la fonction d'autocorrélation discrète du signal x , et ϕ_{TR} la fonction de paramétrisation de la représentation, exprimée dans le domaine temps-retard. On suppose que cette dernière est à support fini S_ϕ tel que :

$$S_\phi = \{(m, l) \in \mathcal{Z} \times \mathcal{Z} : |m| \leq M, |l| \leq L-1\}.$$

Comme pour la transformée de Fourier à court terme, nous allons maintenant expliciter une condition nécessaire et suffisante d'admissibilité de la fonction de paramétrisation ϕ_{TR} afin que C_x satisfasse une relation de récursivité directe.

Proposition 2.3. La distribution C_x définie par l'expression (2.16) vérifie la relation de récursivité directe suivante :

$$C_x[n, \nu ; \phi] = \alpha C_x[n-1, \nu ; \phi] + c \alpha^{-M} C_x[n+M, \nu ; \phi] - c \alpha^{M+1} C_x[n-M-1, \nu ; \phi], \quad (2.17)$$

si et seulement si la fonction de paramétrisation ϕ_{TR} , exprimée dans le plan temps-retard, appartient à la classe générale $\mathcal{E}(C_x)$ suivante :

$$\mathcal{E}(C_x) = \left\{ \phi_{TR} : \phi_{TR}[m, l] = c \sum_{k=-M}^M \alpha^{-k} \varphi[m-k, l], c \in \mathcal{C}, \alpha \in \mathcal{C}, |m| \leq M, |l| \leq L-1 \right\}, \quad (2.18)$$

où φ est une fonction de paramétrisation de support fini S_φ défini par :

$$S_\varphi = \{(m, l) \in \mathcal{Z} \times \mathcal{Z} : |m+l| \leq L-1\}.$$

Preuve. D'après la définition (2.16), lorsqu'on effectue une analyse temps-fréquence à court terme du signal x à l'instant n , on supprime $x[n-M-L]$ et on ajoute $x[n+M+L-1]$ à la liste des échantillons observables. Le processus récursif direct recherché est alors du type :

$$C_x[n, \nu; \phi] = U(C_x[n-1, \nu; \phi], X(n+M), X(n-M-1)), \quad (2.19)$$

où U désigne une fonction de mise à jour et $X(n+k)$ l'ensemble des échantillons de x intervenant dans la définition de la fonction d'autocorrélation R_x à l'instant $n+k$:

$$X(n+k) = \{x[n+k-L+1], \dots, x[n+k+L-1]\}.$$

Dans ce contexte, la relation de récursivité directe (2.3) se réécrit sous la forme :

$$C_x[n, \nu; \phi] = \alpha C_x[n-1, \nu; \phi] + \beta G(X(n+M)) + \gamma G(X(n-M-1)).$$

On notera que cette expression est une généralisation de la relation intervenant dans la Proposition 2.1 puisque $X(n)$ désigne désormais un ensemble d'échantillons du signal. La démarche conduisant à la formulation d'une condition nécessaire et suffisante reste cependant inchangée et on obtient :

$$C_x[n, \nu; \phi] = c \sum_{k=-M}^M \alpha^{-k} G(X(n+k)), \text{ avec } \beta = c \alpha^{-M} \text{ et } \gamma = -c \alpha^{M+1}. \quad (2.20)$$

Puisque C_x est une distribution de la classe de Cohen, la fonction G a la structure d'une représentation temps-fréquence bilinéaire. Ainsi, on pose :

$$G(X(n+k)) \equiv C_x[n+k, \nu; \phi] = \sum_{m=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} \phi_{TR}[m, l] x[n+k+m+L] x^*[n+k+m-L] e^{-j4\pi\nu l}. \quad (2.21)$$

Dans cette équation, ϕ_{TR} désigne une fonction de paramétrisation de support fini S_ϕ tel que :

$$S_\phi = \{(m, l) \in \mathcal{Z} \times \mathcal{Z} : |m+l| \leq L-1\},$$

car $X(n+k)$ ne désigne que les échantillons $x[n+k-L+1], \dots, x[n+k+L-1]$. Enfin, la comparaison de l'expression (2.20) et de la définition (2.16) nous conduit à :

$$\phi_{TR}[m, l] = c \sum_{k=-M}^M \alpha^{-k} \phi_{TR}[m-k, l]. \quad \blacksquare$$

Remarque. La Proposition 2.3 complète le résultat présenté dans [Ami95], le support de la fonction ϕ ne pouvant être quelconque.

D'après la relation (2.17), l'évaluation récursive de $C_x[n, \nu; \phi]$ nécessite le calcul de deux distributions de la classe de Cohen additionnelles, ce qui n'est *a priori* pas propice à une diminution du temps de calcul. Cependant, on note que le nombre d'opérations requis par la mise à jour de la distribution C_x est indépendant de la taille du support temporel $[-M, M]$ de sa fonction de paramétrisation ϕ . Ceci constitue certainement un avantage lorsque M est suffisamment grand. Dans la section suivante, nous montrons qu'il en est ainsi lorsque C_x désigne en particulier les distributions pseudo Wigner-Ville lissées [Ric96a], [Ric97a].

2.3.2 Distribution pseudo Wigner-Ville lissée récursive

a) *Réversivité directe.* D'après la définition (2.18) de la classe $\mathcal{E}(C_x)$, il est manifeste que certaines distributions pseudo Wigner-Ville lissées admettent une formulation récursive directe. En effet, si on suppose que φ est à variables séparables, il s'en suit que la fonction de paramétrisation ϕ est également à variables séparables. Jusqu'à la fin de ce chapitre, nous considérerons en particulier la classe de noyaux suivante :

$$\mathcal{E}(PWL_x) = \left\{ \phi : \phi_{TR}[m, l] = c \alpha^{-m} g[l], |m| \leq M, |l| \leq L-1 \right\}, \quad (2.22)$$

obtenues en posant $\varphi[m, l] = g[l] \delta[m]$, où g est une fonction quelconque définie sur l'intervalle $\{1-L, \dots, L-1\}$.

Il est alors intéressant de noter que les deux distributions additionnelles intervenant dans la relation (2.17) sont de type pseudo Wigner-Ville et que la relation récursive vérifiée par PWL_x est la suivante :

$$PWL_x[n, \nu; \phi] = \alpha PWL_x[n-1, \nu; \phi] + \sum_{l=-\infty}^{+\infty} \kappa[n, l] e^{-j4\pi\nu l}, \quad (2.23)$$

$$\text{avec } \kappa[n, l] = \phi_{TR}[M, l] R_x[n+M, l] - \alpha \phi_{TR}[-M, l] R_x[n-M-1, l]. \quad (2.24)$$

b) *Réversivité indirecte.* En adoptant la même démarche que pour la transformée de Fourier à court terme, la notion de réversivité peut être aisément étendue à une classe de fonctions de paramétrisation plus large que $\mathcal{E}(PWL_x)$. Ainsi, si le noyau d'une distribution appartient à :

$$\mathcal{E}_{\text{ext}}(PWL_x) = \left\{ \phi_{TR} : \phi_{TR}[m, l] = g[l] \sum_{p=1}^P c_p \alpha_p^{-m}, c_p \in \mathcal{C}, \alpha_p \in \mathcal{C}, |m| \leq M, |l| \leq L-1 \right\},$$

la représentation pseudo Wigner-Ville lissée associée admet une formulation récursive indirecte :

$$PWL_x[n, \nu; \phi] = \sum_{p=1}^P PWL_x[n, \nu; \phi_p], \text{ où } \phi_p[m, l] = c_p \alpha_p^{-m} g[l],$$

$PWL_x[n, \nu; \phi_p]$ étant évaluée récursivement grâce à la relation (2.23).

L'analyse de $\mathcal{E}_{\text{ext}}(PWL_x)$ permet de constater que cette classe regroupe l'ensemble des noyaux à variables séparables $g[l]h[m]$ tels que $h[m]$ (périodisée) admet une décomposition en série de Fourier finie. La fonction $g[l]$ quant à elle peut être quelconque.

c) *Performances.* D'après la relation (2.23), il apparaît que le nombre d'opérations requis pour l'évaluation récursive d'une distribution pseudo Wigner-Ville lissée est indépendant de la largeur du support temporel de sa fonction de paramétrisation. Nous allons montrer sur quelques exemples que cette propriété constitue l'atout majeur de l'approche récursive.

Les noyaux à variables séparables que nous allons utiliser vérifient tous la propriété de symétrie hermitienne suivante, par rapport à la seconde variable :

$$(g[l] h[m])^* = g[l] h[-m], \quad (2.25)$$

ce qui permet de reformuler la définition des distributions pseudo Wigner-Ville lissées ainsi :

$$PWL_x[n, \nu; gh] = 2 \operatorname{Re} \left\{ \Xi_x[n, \nu; gh] \right\} - \sum_{m=-M}^M g[0] h[m] R_x[n+m, 0], \quad (2.26)$$

$$\text{avec } \Xi_x[n, \nu; gh] = \sum_{l=0}^{L-1} \sum_{m=-M}^M g[l] h[m] R_x[n+m, l] e^{-j4\pi\nu l}. \quad (2.27)$$

On peut montrer que l'usage de ces expressions pour l'évaluation des distributions pseudo Wigner-Ville lissées conduit à une réduction du volume de calculs [Mar85], [Zie91]. Aussi, afin de tester les algorithmes classique et récursif dans les conditions les plus favorables, nous allons évaluer le nombre d'opérations requis sur la base des relations (2.26) et (2.27).

Considérons en particulier le cas où g est une fonction quelconque à valeurs complexes et h une fenêtre rectangulaire de largeur $2L-1$, où L est une puissance de 2. A chaque instant, les deux algorithmes débutent par l'évaluation de la fonctions d'autocorrélation instantanée ($M_c = L$). Puis, suivant l'approche adoptée, les calculs s'enchaînent ainsi :

Calcul classique. L'évaluation de l'expression (2.27) conduit à sommer l'ensemble des termes R_x ($A_c = 2ML$). Le résultat est alors fenêtré au moyen de la fonction à valeurs complexes g ($M_c = L$), puis une transformée de Fourier rapide lui est appliquée. Enfin, PWL_x est obtenu en utilisant l'expression (2.26) ($M_r = L; A_r = L$). En conclusion, le calcul classique de cette distribution pseudo Wigner-Ville lissée nécessite à chaque instant le nombre d'opérations réelles suivant [Ric97a] :

$$A_r = 4(M+1)L + 3L \log_2(L) \quad M_r = 7L + 2L \log_2(L).$$

Calcul récursif. Lorsque le noyau utilisé vérifie la propriété de symétrie hermitienne (2.25), l'approche faisant l'objet de cette section consiste à évaluer récursivement Ξ_x sur le modèle des relations (2.23) et (2.24), puis à utiliser la expression (2.27) afin d'obtenir le résultat final. L'évaluation de l'expression (2.24) nécessite L additions et $2L$ multiplications complexes. On applique alors une transformée de Fourier rapide, puis la partie réelle du résultat est ajoutée à la partie réelle de la valeur de Ξ_x obtenue à l'instant précédent ($A_r = L$). Enfin, comme pour l'approche classique, PWL_x résulte alors de l'application de la relation (2.26) ($M_r = L; A_r = L$). En conclusion, le calcul récursif de cette distribution pseudo Wigner-Ville lissée requiert à chaque instant le nombre d'opérations réelles suivant [Ric97a] :

$$A_r = 7L + 3L \log_2(L) \quad M_r = 7L + 2L \log_2(L).$$

Ces résultats mettent en évidence que le nombre d'opérations mises en oeuvre par l'approche récursive est indépendant de la taille du support temporel $[-M, M]$ de la fenêtre de lissage temporel h , ce que confirment également les résultats figurant en annexe de ce chapitre, page 51. Ceci constitue un atout indéniable par rapport à la méthode classique, comme nous le montre la Figure 2.3.(a). En revanche, on note sur la Figure 2.3.(b) que le gain en temps de calcul décroît lorsqu'on fait croître la largeur de la fenêtre de lissage fréquentiel g , pour une fenêtre de lissage temporel h donnée.

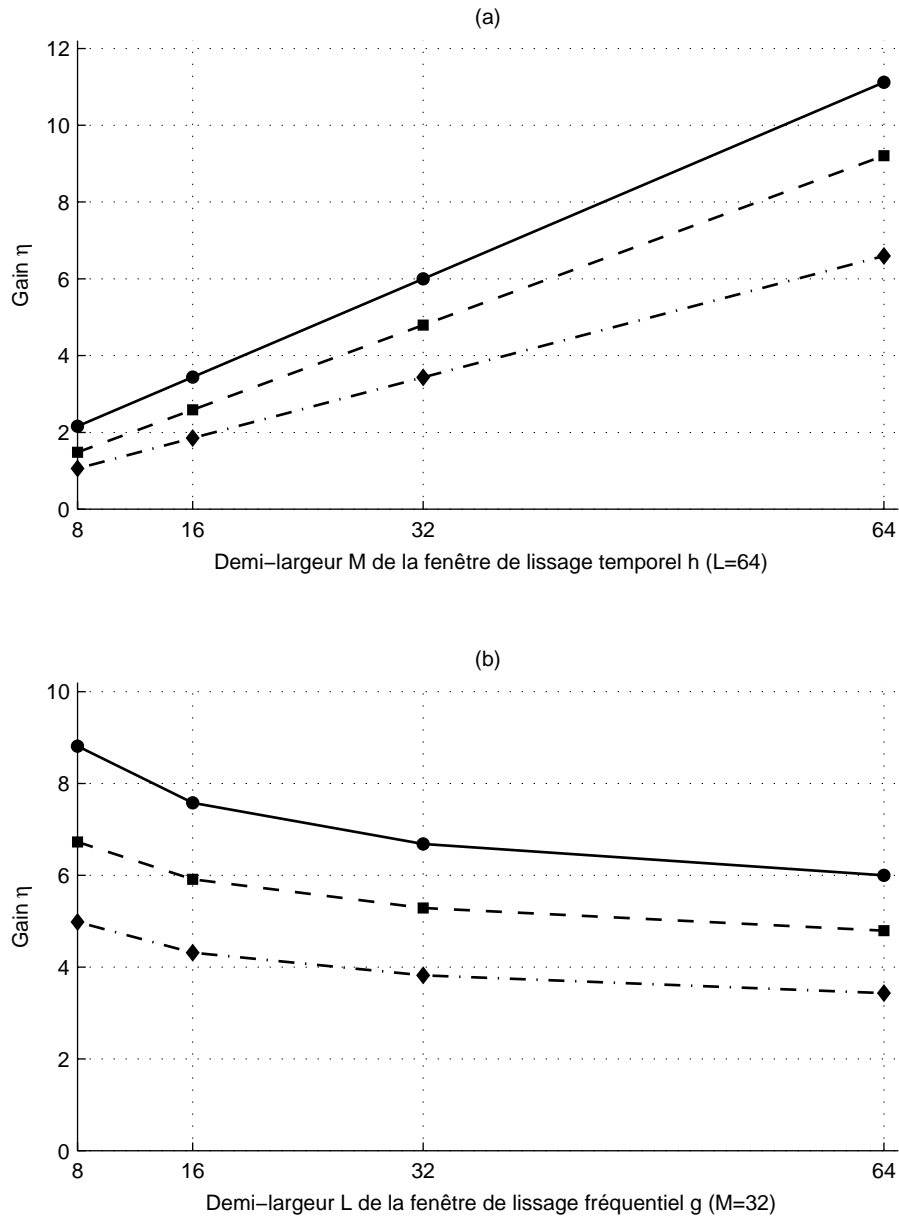


Figure 2.3 : Estimation du gain en temps de calcul lorsque l'approche récursive est préférée à la méthode classique pour l'évaluation de la *distribution pseudo Wigner-Ville lissée*. Différentes fenêtres de lissage temporel h sont considérées : rectangulaire (●), semi-sinusoidale (■), Hamming et Hanning (◆). (a) La demi-largeur L de la fenêtre de lissage fréquentiel g est fixée à 64. (b) La demi-largeur M de la fenêtre h est fixée à 32.

2.4 Récursivité et méthode de réallocation

L'amélioration de la lisibilité des représentations temps-fréquence constitue une étape cruciale pour l'analyse des signaux non-stationnaires. A cet effet, de nombreux éléments de solutions ont été proposés, parmi lesquels on compte la méthode de réallocation décrite dans la section 1.3.6. Ce procédé est resté peu utilisé jusqu'à ce qu'Auger et Flandrin démontrent

théoriquement son efficacité. Ces derniers ont alors présenté une nouvelle formulation des opérateurs de réallocation afin d'en faciliter l'implémentation [Aug95]. En nous appuyant sur ces résultats, nous proposons dans cette section une stratégie récursive visant à réduire le temps de calculs requis par la mise en oeuvre de la méthode. Ainsi, nous nous intéressons tout d'abord à l'introduction de la récursivité dans le processus de réallocation du spectrogramme, puis dans celui des distributions pseudo Wigner-Ville lissées. Dans ces deux cas, le gain en temps de calcul fait l'objet d'une estimation, que corroborent les résultats expérimentaux figurant dans [Ric97a].

2.4.1 Spectrogramme réalloué

a) *Récursivité et opérateurs de réallocation.* En temps discret, les opérateurs de réallocation associés au spectrogramme peuvent s'exprimer ainsi [Aug95] :

$$\hat{n}[n, \nu; x] \equiv n - \mathbf{Re} \left\{ \frac{F_x[n, \nu; Th] F_x^*[n, \nu; h]}{|F_x[n, \nu; h]|^2} \right\}, \quad (2.28)$$

$$\hat{\nu}[n, \nu; x] \equiv \nu - \mathbf{Im} \left\{ \frac{F_x[n, \nu; Dh] F_x^*[n, \nu; h]}{|F_x[n, \nu; h]|^2} \right\}, \quad (2.29)$$

où F_x désigne la transformée de Fourier à court terme définie par l'expression (2.8), $Dh[n]$ la dérivée de la fonction $h(t)$ à variable continue évaluée en n , et $Th[n] = n h[n]$.

Les fonctions \hat{n} et $\hat{\nu}$ de la variable discrète n ne vérifiant pas la condition de récursivité directe établie dans la Proposition 2.1, la recherche d'une stratégie récursive de nature indirecte s'impose. La structure particulière des expressions (2.28) et (2.29) nous amène ainsi à nous intéresser aux propriétés de récursivité des distributions $F_x[\cdot; h]$, $F_x[\cdot; Th]$ et $F_x[\cdot; Dh]$. Dans ces circonstances, en adoptant les mêmes notations que dans la section 2.2.1, on peut énoncer le résultat suivant [Ric96a], [Ric97a] :

Proposition 2.4. Soit h appartenant à la classe $\mathcal{E}(F_x)$ des fenêtres d'analyse autorisant une implémentation récursive directe de la transformée de Fourier à court terme $F_x[\cdot; h]$. Les relations suivantes établissent les récursivités directe et indirecte de $F_x[\cdot; Dh]$ et $F_x[\cdot; Th]$:

$$F_x[n, \nu; Dh] = \rho e^{j2\pi\nu} F_x[n-1, \nu; Dh] + e^{-j2\pi\nu(N-1)} Dh^*[N-1] x[n+N-1] - \rho e^{j2\pi\nu N} Dh^*[1-N] x[n-N], \quad (2.30)$$

$$F_x[n, \nu; Th] = \rho (F_x[n-1, \nu; Th] - F_x[n-1, \nu; h]) e^{j2\pi\nu} + (N-1) h^*[N-1] x[n+N-1] e^{-j2\pi\nu(N-1)} + \rho N h^*[1-N] x[n-N] e^{j2\pi\nu N}. \quad (2.31)$$

Preuve. La première relation résulte du fait que Dh appartient à $\mathcal{E}(F_x)$ puisque l'on a :

$Dh[k] = -\ln(\rho) h[k]$ où ρ , supposé non nul, est défini par l'expression (2.9).

Quant à la deuxième relation, un calcul simple permet de la vérifier. ■

Afin de limiter le temps de calcul requis par l'évaluation de $F_x[\cdot; Dh]$, on préférera cependant à la relation (2.30) utiliser directement le fait que [Ric96a], [Ric97a] :

$$F_x[n, \nu; Dh] = -\ln(\rho) F_x[n, \nu; h], \text{ car } Dh[k] = -\ln(\rho) h[k]. \quad (2.32)$$

Enfin, en s'appuyant sur les résultats de la section 2.2.2, on notera que la stratégie récursive que nous proposons peut être étendue à l'ensemble des fenêtres de la classe $\mathcal{E}_{\text{ext}}(F_x)$ définie en section 2.2.2 [Ric96a], [Ric97a].

b) Algorithme et performances. L'algorithme présenté dans la Figure 2.4 adopte une stratégie récursive pour l'évaluation conjointe des spectrogrammes classique SP_x et réalloué SPR_x d'un signal x . On peut remarquer que le processus de réallocation n'y est activé que si $SP_x[n, \nu; h]$ est supérieur à un seuil ε donné [Aug95]. Ce critère garantit l'existence des expressions (2.28) et (2.29), et empêche la réallocation inutile des échantillons $SP_x[n, \nu; h]$ dont la valeur est proche de zéro.

Afin d'évaluer les performances de cet algorithme, une méthodologie similaire à celle employée dans la section 2.2.3 a été adoptée. Le nombre d'opérations requis par l'évaluation

PROGRAMME PRINCIPAL

Répéter pour chaque instant d'analyse n ,

Acquérir l'échantillon $x[n+N]$

Evaluer : $F_x[n, \nu; h]$ avec $F_x[n-1, \nu; h]$ cf. Eq. (2.10)

$F_x[n, \nu; Dh]$ avec $F_x[n, \nu; h]$ cf. Eq. (2.32)

$F_x[n, \nu; Th]$ avec $F_x[n-1, \nu; Th]$ et $F_x[n-1, \nu; h]$ cf. Eq. (2.31)

$SP_x[n, \nu; h] = F_x[n, \nu; h] \cdot F_x[n, \nu; h]^*$

Répéter pour chaque composante fréquentielle ν ,

Si $SP_x[n, \nu; h] > \varepsilon$ alors exécuter la FONCTION de REALLOCATION

Fin Répéter

Fin Répéter

FONCTION de REALLOCATION

Evaluer : $(\hat{n}, \hat{\nu})$ avec $F_x[n, \nu; h]$, $F_x[n, \nu; Th]$ et $F_x[n, \nu; Dh]$ cf. Eqs. (2.28), (2.29)

$SPR_x[\hat{n}, \hat{\nu}; h] = SPR_x[\hat{n}, \hat{\nu}; h] + SP_x[n, \nu; h]$

Figure 2.4 : Algorithme à stratégie récursive pour l'évaluation du *spectrogramme réalloué*.

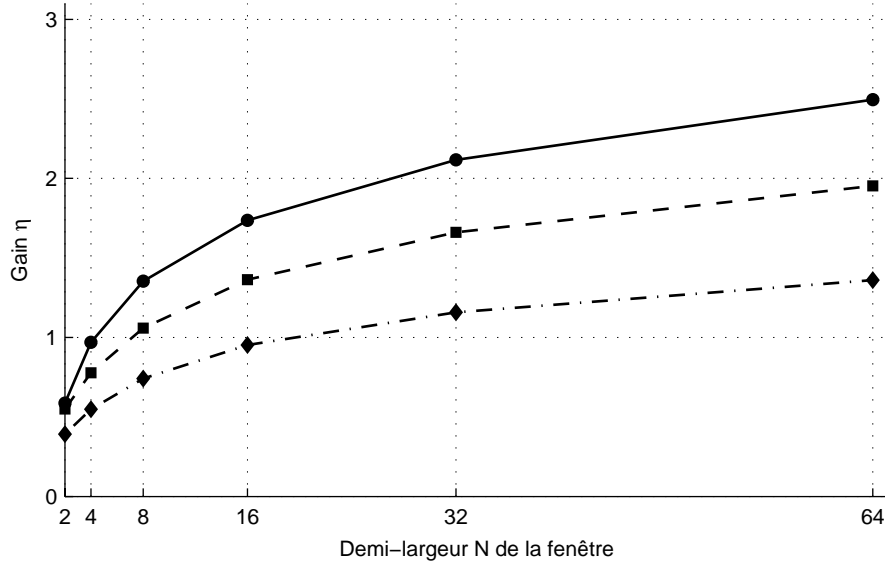


Figure 2.5 : Estimation du gain en temps de calcul lorsque l'approche récursive est préférée à la méthode classique pour l'évaluation du *spectrogramme réalloué*. Différentes fenêtres d'analyse sont considérées : rectangulaire (●), semi-sinusoidale (■), Hamming et Hanning (◆).

du spectrogramme réalloué lorsque la fenêtre d'analyse h est rectangulaire, semi-sinusoidale, de Hamming ou Hanning, est indiqué en annexe de ce chapitre, page 52. Le résultat présenté met en évidence la complexité linéaire de l'algorithme récursif, plus favorable que celle en $O(N \log_2 N)$ de l'approche classique consistant à appliquer directement les relations (2.8), (2.28) et (2.29). Dans la Figure 2.5, l'indicateur (2.15) du gain en temps de calcul est représenté pour différentes fenêtres d'analyse en fonction de leur demi-largeur N , lorsque tous les échantillons $SP_x[n, v; h]$ de la représentation font l'objet d'une réallocation. Ce résultat démontre l'efficacité de l'approche récursive lorsque la largeur de la fenêtre h excède une taille critique.

2.4.2 Distributions pseudo Wigner-Ville lissées réallouées

a) *Récursivité et opérateurs de réallocation*. Dans [Aug95], les auteurs formulent les opérateurs de réallocation associés aux distributions C_x de la classe de Cohen ainsi :

$$\hat{n}[n, v; x] = n - \frac{C_x[n, v; T\phi]}{C_x[n, v; \phi]}, \quad (2.33)$$

$$\hat{v}[n, v; x] = v + j \frac{C_x[n, v; D\phi]}{C_x[n, v; \phi]}, \quad (2.34)$$

où $T\phi_{TR}[m, l] = m \cdot \phi_{TR}[m, l]$ et $D\phi_{TR}[m, l]$ est la dérivée partielle $\partial \phi_{TR}(t, \tau) / \partial \tau$ évaluée en (m, l) .

En adoptant la même approche que pour le spectrogramme réalloué, on peut montrer que $C_x[\cdot; D\phi]$ et $C_x[\cdot; T\phi]$ peuvent être évaluées récursivement lorsque $C_x[\cdot; \phi]$ satisfait la relation de récursivité directe (2.17). Cependant, afin de limiter la complexité des différentes relations

récursives exposées, nous restreignons la présente étude aux cas des distributions pseudo Wigner-Ville lissées. Dans ces circonstances, en conservant les notations de la section 2.3, on peut énoncer le résultat suivant [Ric96a], [Ric97a] :

Proposition 2.5. Soit ϕ appartenant à la classe $\mathcal{E}(PWL_x)$ des fonctions de paramétrisation à variables séparables autorisant une implémentation récursive directe de la distribution pseudo Wigner-Ville lissée $PWL_x[\cdot; \phi]$. Les relations suivantes établissent respectivement les récursivités directe et indirecte de $PWL_x[\cdot; D\phi]$ et $PWL_x[\cdot; T\phi]$:

$$PWL_x[n, \nu; D\phi] = \alpha PWL_x[n-1, \nu; D\phi] + \sum_{l=1-L}^{L-1} \eta[n, l] e^{-j4\pi\nu l}, \quad (2.35)$$

avec $\eta[n, l] = D\phi_{TR}[M, l] R_x[n+M, l] - \alpha D\phi_{TR}[-M, l] R_x[n-M-1, l]$.

$$PWL_x[n, \nu; T\phi] = \alpha (PWL_x[n-1, \nu; T\phi] - PWL_x[n-1, \nu; \phi]) + \sum_{l=1-L}^{L-1} \psi[n, l] e^{-j4\pi\nu l}, \quad (2.36)$$

avec $\psi[n, l] = M \phi_{TR}[M, l] R_x[n+M, l] - \alpha (M+1) \phi_{TR}[-M, l] R_x[n-M-1, l]$.

Preuve. L'expression (2.35) résulte du fait que si ϕ appartient à $\mathcal{E}(PWL_x)$, alors $D\phi$ appartient également à $\mathcal{E}(PWL_x)$. En conséquence, la relation récursive directe (2.23) reste valide pour la distribution $PWL_x[\cdot; D\phi]$. Un calcul simple permet de vérifier la relation (2.36). ■

On peut noter que l'association de plusieurs processus récursifs de type (2.35) et (2.36) permet d'étendre la stratégie récursive qui vient d'être présentée à la classe des fonctions de paramétrisation $\mathcal{E}_{\text{ext}}(PWL_x)$. Dans ce cas, il est fait appel à la propriété de linéarité de la transformée de Fourier, comme cela est précisé dans la section 2.3.2.b [Ric96a], [Ric97a].

b) Performances. Les noyaux à variables séparables $g[l]h[m]$, notés gh , que nous allons utiliser pour démontrer l'intérêt de la stratégie récursive vérifient la propriété de symétrie hermitienne (2.25). Ceci a pour conséquence que [Ric97a] :

$$(Dgh[l, m])^* = -Dgh[l, -m] \text{ et } (Tgh[l, m])^* = Tgh[l, -m],$$

ce qui permet de reformuler les distributions pseudo Wigner-Ville lissées de noyaux Tgh et Dgh ainsi [Ric97a] :

$$PWL_x[n, \nu; Tgh] = 2 \mathbf{Re} \{ \Xi_x[n, \nu; Tgh] \} - \sum_{m=-M}^M Tgh[0, m] R_x[n+m, 0], \quad (2.37)$$

$$PWL_x[n, \nu; Dgh] = 2 \mathbf{Im} \{ \Xi_x[n, \nu; Dgh] \} - \sum_{m=-M}^M Dgh[0, m] R_x[n+m, 0], \quad (2.38)$$

où $\Xi_x[\cdot; \phi]$ est définie par :

$$\Xi_x[n, \nu; \phi] = \sum_{l=0}^{L-1} \sum_{m=-M}^M \phi[l, m] R_x[n+m, l] e^{-j4\pi\nu l} . \quad (2.39)$$

Tout comme $\Xi_x[\cdot; gh)$, les distributions $\Xi_x[\cdot; Dgh)$ et $\Xi_x[\cdot; Tgh)$ peuvent faire l'objet d'une évaluation récursive [Ric97a]. Les relations conservent alors l'allure générale des expressions (2.35) et (2.36).

Que la stratégie adoptée soit classique ou récursive, l'usage des relations (2.37), (2.38) et (2.39) permet de réduire le temps de calcul des distributions $PWL_x[\cdot; Tgh)$ et $PWL_x[\cdot; Dgh)$, et donc des opérateurs de réallocation. En conséquence, afin de comparer ces deux approches dans les conditions les plus favorables, les calculs de complexité algorithmique figurant en annexe de ce chapitre, page 52, ont été effectués sur la base de ces expressions. Ces résultats mettent en évidence que la complexité de l'algorithme récursif est indépendante de la taille du support temporel $[-M, M]$ de la fenêtre de lissage temporel h . Ceci constitue un atout indéniable pour l'approche proposée, comme nous le montre la Figure 2.7.(a). En revanche, on note sur la Figure 2.7.(b) que le gain en temps de calcul décroît lorsqu'on fait croître la largeur de la fenêtre de lissage fréquentiel g , pour une fenêtre de lissage temporel h donnée. Enfin, on peut montrer que le gain en temps de calcul est peu affecté par le nombre d'échantillons $PWL_x[n, \nu; gh)$ réalloués à chaque instant n , l'exécution de la fonction de réallocation précisée dans la Figure 2.6 nécessitant au maximum $3N$ additions et $2N$ multiplications réelles. Cette remarque est également valable pour le spectrogramme réalloué.

PROGRAMME PRINCIPAL

Répéter pour chaque instant d'analyse n ,

Acquérir l'échantillon $x[n+M+L-1]$

Evaluer : $PWL_x[n, \nu; \phi]$ avec $PWL_x[n-1, \nu; \phi]$ cf. Eq. (2.23)

$PWL_x[n, \nu; D\phi]$ avec $PWL_x[n, \nu; \phi]$ cf. Eq. (2.35)

$PWL_x[n, \nu; T\phi]$ avec $PWL_x[n-1, \nu; T\phi]$ et $PWL_x[n-1, \nu; \phi]$ cf. Eq. (2.36)

Répéter pour chaque composante fréquentielle ν ,

Si $PWL_x[n, \nu; \phi] > \varepsilon$, alors exécuter la FONCTION de REALLOCATION

Fin Répéter

Fin Répéter

FONCTION de REALLOCATION

Evaluer : $(\hat{n}, \hat{\nu})$ avec $PWL_x[n, \nu; \phi]$, $PWL_x[n, \nu; T\phi]$ et $PWL_x[n, \nu; D\phi]$

$PWLR_x[\hat{n}, \hat{\nu}; \phi] = PWLR_x[\hat{n}, \hat{\nu}; \phi] + PWL_x[n, \nu; \phi]$

Figure 2.6 : Algorithme à stratégie récursive pour l'évaluation de la distribution *pseudo Wigner-Ville lissée réallouée*.

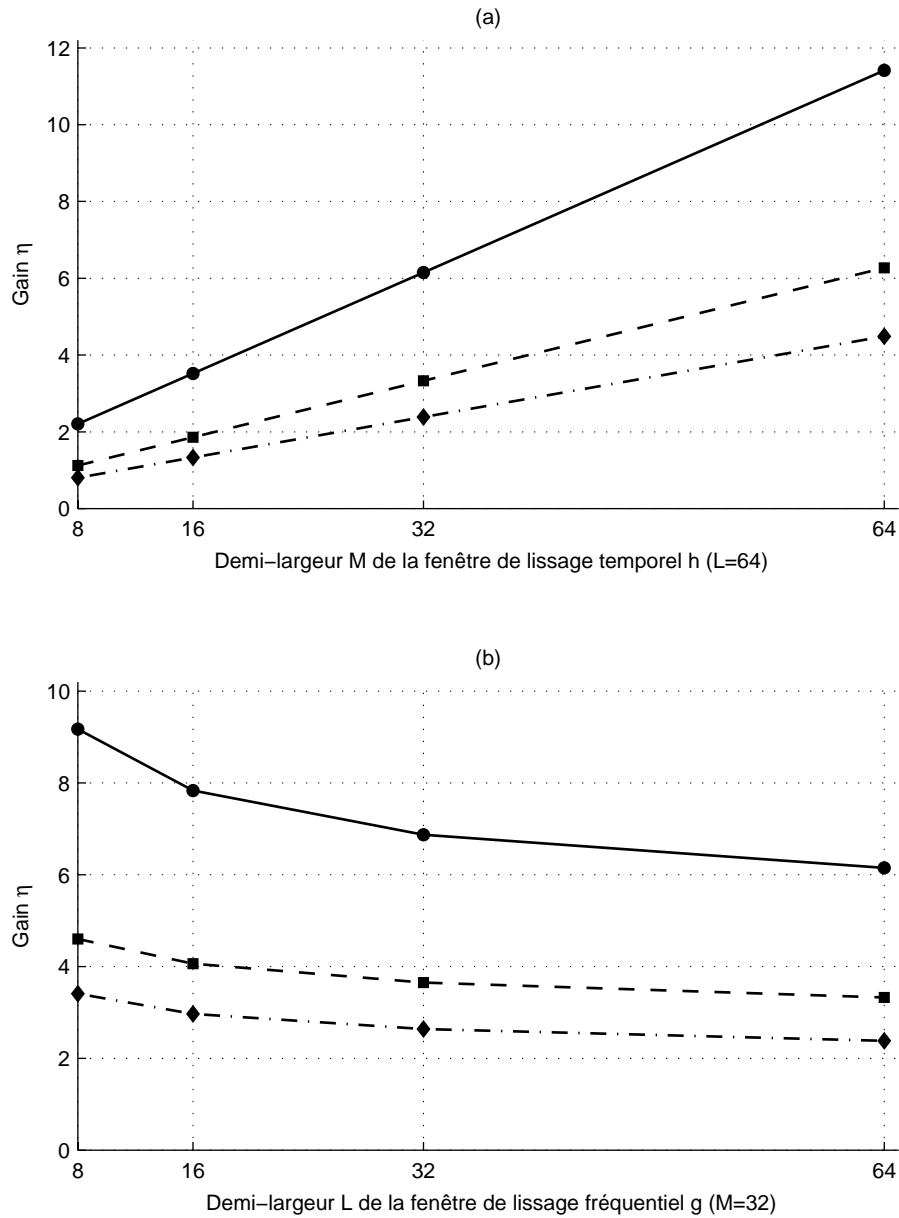


Figure 2.7 : Estimation du gain en temps de calcul lorsque l'approche récursive est préférée à la méthode classique pour l'évaluation de la *distribution pseudo Wigner-Ville lissée réallouée*. Différentes fenêtres de lissage temporel h sont considérées : rectangulaire (●), semi-sinusoidale (■), Hamming et Hanning (◆). (a) La demi-largeur L de la fenêtre de lissage fréquentiel g est fixée à 64. (b) La demi-largeur M de la fenêtre h est fixée à 32.

2.5 Conclusion

Après avoir effectué quelques rappels théoriques sur les représentations temps-fréquence au chapitre précédent, nous avons traité les problèmes pratiques liés à leur évaluation, en mettant plus particulièrement l'accent sur le potentiel des algorithmes récursifs. Ainsi, nous avons établi des conditions nécessaires et suffisantes d'admissibilité de la fenêtre d'analyse h et de la fonction de paramétrisation ϕ , afin que la transformée de Fourier à court terme et les distributions de la classe de Cohen puissent se prêter à ce type d'implémentation. Nous avons alors présenté une extension de cette propriété aux distributions modifiées par la méthode de réallocation. Enfin, l'efficacité de chacun des algorithmes récursifs proposés a été évaluée au moyen de calculs de complexité, qui ont permis de démontrer clairement l'intérêt que présente cette approche.

Fort des résultats de ces deux premiers chapitres, il est désormais possible de représenter les signaux non-stationnaires dans un espace mieux adapté : le plan temps-fréquence. Au cours des chapitres qui suivent, nous aurons fréquemment recours à celui-ci afin d'illustrer l'efficacité de la méthodologie proposée pour la synthèse de détecteurs à structure imposée. A cette occasion, nous insisterons sur les nouvelles perspectives qu'offre la classe de Cohen pour la résolution de ce type de problèmes, en nous appuyant sur des résultats de la théorie de l'apprentissage.

2.6 Annexes du Chapitre 2

2.6.1 Récursivité et représentations non réallouées

fenêtre h	additions réelles (A_R)		produits réels (M_R)	
	classique	récursif	classique	récursif
rectangulaire	$6N \log_2 N + 4N$	$8N + 2$	$4N \log_2 N$	$6N$
semi-sinusoidale	$6N \log_2 N + 8N$	$20N + 12$	$4N \log_2 N + 8N$	$16N + 16$
Hamming-Hanning	$6N \log_2 N + 8N$	$32N + 16$	$4N \log_2 N + 8N$	$24N + 20$

Tableau 1.2 : nombre d'opérations réelles mises en oeuvre, à chaque instant, pour le calcul de la transformée de Fourier à court terme. La fenêtre h d'analyse, rectangulaire, semi-sinusoidale, de Hamming ou Hanning, est supposée de largeur $2N$.

fenêtre h	additions réelles (A_R)		produits réels (M_R)	
	classique	récursif	classique	récursif
rectangulaire	$3L \log_2 L + 4(M+1)L$	$3L \log_2 L + 7L$	$2L \log_2 L + 7L$	$2L \log_2 L + 7L$
semi-sinusoidale	$3L \log_2 L + 4(2M+1)L$	$6L \log_2 L + 22L$	$2L \log_2 L + (8M+7)L$	$4L \log_2 L + 25L$
Hamming-Hanning	$3L \log_2 L + 4(2M+1)L$	$9L \log_2 L + 27L$	$2L \log_2 L + (8M+7)L$	$6L \log_2 L + 27L$

Tableau 1.3 : nombre d'opérations réelles mises en oeuvre, à chaque instant, pour le calcul de la distribution pseudo Wigner-Ville lissée. Les demi-largeurs des fenêtres h et g de lissage temporel et fréquentiel sont notées respectivement M et L . La fenêtre g est simplement supposée à valeurs complexes.

2.6.2 Récursivité et représentations réallouées

fenêtre h	additions réelles (A_R)		produits réels (M_R)	
	classique	récursif	classique	récursif
rectangulaire	$6N \log_2 N + 4N$	$16N + 2$	$4N \log_2 N + 6N$	$16N + 2$
semi-sinusoidale	$9N \log_2 N + 9N$	$32N + 16$	$6N \log_2 N + 16N$	$28N + 24$
Hamming-Hanning	$9N \log_2 N + 9N$	$46N + 20$	$6N \log_2 N + 16N$	$36N + 30$

Tableau 1.4 : nombre d'opérations réelles mises en oeuvre, à chaque instant, pour le calcul du spectrogramme réalloué. La fenêtre h d'analyse, rectangulaire, semi-sinusoidale, de Hamming ou Hanning, est supposée de largeur $2N$.

fenêtre h	additions réelles (A_R)		produits réels (M_R)	
	classique	récursif	classique	récursif
rectangulaire	$9L \log_2 L + (12M + 11)L$	$9L \log_2 L + 19L$	$6L \log_2 L + (8M + 15)L$	$6L \log_2 L + 19L$
semi-sinusoidale	$9L \log_2 L + (16M + 13)L$	$18L \log_2 L + 66L$	$6L \log_2 L + (16M + 19)L$	$12L \log_2 L + 70L$
Hamming-Hanning	$9L \log_2 L + (16M + 13)L$	$27L \log_2 L + 81L$	$6L \log_2 L + (16M + 19)L$	$18L \log_2 L + 82L$

Tableau 1.5 : nombre d'opérations réelles mises en oeuvre, à chaque instant, pour le calcul de la distribution pseudo Wigner-Ville lissée réallouée. Les demi-largeurs des fenêtres h et g de lissage temporel et fréquentiel sont notées respectivement M et L . La fenêtre g est simplement supposée à valeurs complexes.

Chapitre 3

Une méthodologie pour la synthèse de détecteurs à structure imposée

Illustrations dans le plan temps-fréquence

3.1 Introduction

Le problème de la détection d'un signal noyé dans un bruit additif est classiquement défini par le test statistique d'hypothèse suivant :

$$\begin{cases} H_0 : X = B \\ H_1 : X = S + B, \end{cases} \quad (3.1)$$

où X désigne l'observation, B le bruit additif et S le signal à détecter, S et B étant supposés indépendants. En pratique, le signal S peut être parfaitement connu, connu à l'exception d'un vecteur θ de paramètres éventuellement aléatoires, ou encore aléatoire. A ce problème, on associe généralement une partition $\{\mathcal{X}_0, \mathcal{X}_1\}$ de l'espace des observations \mathcal{X} , ainsi qu'une fonction booléenne $d(X)$, appelée *test de détection*, que l'on définit par :

$$d(X) = \begin{cases} 0 & \text{si } X \in \mathcal{X}_0 \text{ (} H_0 \text{ supposée vraie),} \\ 1 & \text{si } X \in \mathcal{X}_1 \text{ (} H_1 \text{ supposée vraie).} \end{cases}$$

En adoptant ce modèle, la résolution du problème (3.1) est alors équivalente à la recherche d'une partition de l'espace des observations, qui soit optimale au sens d'un critère défini à partir des objectifs à atteindre.

Lorsqu'on dispose des densités de probabilité conditionnelles de l'observation X , l'application d'un critère de décision, tel que celui de Bayes, conduit à une expression du test d'hypothèses (3.1) consistant à comparer une variable $\Lambda(X)$, appelée *statistique de détection*, à un seuil λ_0 . Le partitionnement de l'espace \mathcal{X} défini ci-dessus consiste alors à affecter les réalisations x aux régions \mathcal{X}_0 et \mathcal{X}_1 selon la règle suivante :

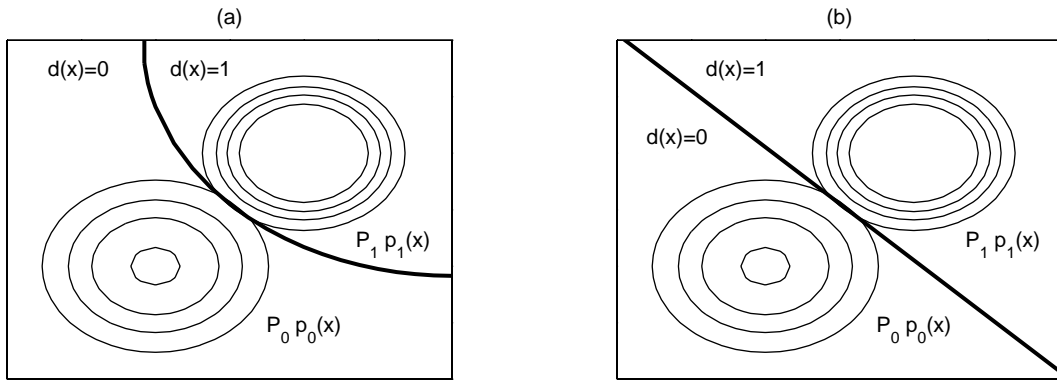


Figure 3.1 : Détection d'un signal aléatoire gaussien noyé dans un bruit blanc gaussien. Partition de l'espace des réalisations par (a) un *détecteur à structure libre* et (b) un *détecteur à structure linéaire imposée* maximisant le critère de Fisher. Sur ces représentations figurent les courbes de niveau des lois de vraisemblance conditionnelles, pondérées par les probabilités *a priori*.

$$\mathcal{X}_0 = \{x \in \mathcal{X} : \Lambda(x) < \lambda_0\},$$

$$\mathcal{X}_1 = \{x \in \mathcal{X} : \Lambda(x) > \lambda_0\}.$$

Dans ce cas, on parle de *détection à structure libre*, la fonction de décision recherchée ne faisant l'objet d'aucune contrainte d'ordre structurel, mais étant dictée par le choix d'un critère de décision et la connaissance des lois de probabilité conditionnelles.

En revanche, lorsque la connaissance *a priori* des hypothèses H_0 et H_1 se limite, par exemple, à des moments statistiques d'un certain ordre, les méthodes évoquées précédemment s'avèrent inopérantes. L'une des approches à laquelle on peut alors avoir recours consiste à définir préalablement la nature de la statistique de détection, linéaire ou quadratique le plus fréquemment, puis à en optimiser les paramètres caractéristiques selon un critère donné. Pour cette raison, cette approche est qualifiée de *détection à structure imposée*. À titre d'exemple, la Figure 3.1 illustre les notions de détection à structure libre et imposée en présentant les deux partitions $\{\mathcal{X}_0, \mathcal{X}_1\}$ obtenues pour la résolution d'un même problème. Dans ce chapitre, on se propose d'aborder successivement les questions propres au choix d'une statistique de détection, d'un critère lié de préférence aux performances susceptibles d'être atteinte par le détecteur, et d'une procédure dédiée à l'optimisation de celui-ci.

Le chapitre 3 est organisé ainsi. Dans un premier temps, on présente brièvement la notion de détecteur à structure libre, en exposant plusieurs règles de décision, après avoir distingué le cas des hypothèses simples et composées. La suite du chapitre est alors uniquement consacrée à la définition d'une méthodologie pour l'élaboration d'un détecteur à structure imposée. L'efficacité de celui-ci est ensuite illustrée par la résolution de problèmes de détection dans le plan temps-fréquence. Le recours à des résultats de la théorie de l'apprentissage et de la

reconnaissance des formes constitue, en plus de quelques contributions personnelles, l'une des originalités de ce travail.

3.2 Détection à structure libre

La détection à structure libre justifie son appellation par le fait qu'aucune contrainte ne pèse sur la structure de la fonction de décision recherchée. Celle-ci résulte uniquement du choix d'un critère de décision, qui est conditionné à la fois par les objectifs à atteindre et les informations dont on dispose. Cette liberté structurelle n'est en revanche possible que si on connaît les lois de probabilité du vecteur des observations X , conditionnellement aux hypothèses :

$p_0(x)$: densité de probabilité du vecteur aléatoire X conditionnellement à l'hypothèse H_0 ,

$p_1(x)$: densité de probabilité du vecteur aléatoire X conditionnellement à l'hypothèse H_1 .

Dans cette section, nous allons exposer les concepts généraux de la détection à structure libre que l'on a jugés importants pour la suite du document. La présentation est néanmoins loin d'être exhaustive mais la littérature sur le sujet est abondante [Tre68], [Poo94], [Don95].

3.2.1 Hypothèses simples

On suppose dans un premier temps que H_0 et H_1 sont des *hypothèses simples*, ce qui signifie qu'elles sont parfaitement précisées par une loi de vraisemblance conditionnelle unique. Les différents critères que l'on rencontre généralement pour l'élaboration d'un détecteur à structure libre sont alors :

Le critère de Bayes

Ce premier critère sanctionne les décisions erronées et récompense celles qui sont correctes. Ainsi, en associant le coût C_{ij} à la décision en faveur de l'hypothèse H_i lorsque H_j est effective, ceci nous amène à supposer que :

$$C_{00} < C_{01} \text{ et } C_{11} < C_{10}.$$

Le critère de Bayes consiste alors à minimiser le coût moyen d'une décision, qui est défini par :

$$C_{\text{moyen}} \equiv \sum_{i,j=0,1} C_{ij} \mathbf{P}\{d(X) = i; Y = j\},$$

où $Y = y$ est un booléen désignant l'hypothèse dont l'observation $X = x$ constitue une réalisation.

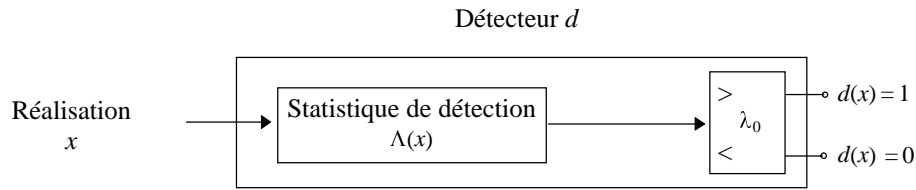


Figure 3.2 : Principe du détecteur à seuil.

On peut montrer que le test de détection reposant sur cette règle de décision équivaut à comparer le quotient $\Lambda_r(X)$ des lois de vraisemblance, appelé *rapport de vraisemblance*, à un seuil [Tre68], [Sch91], [Poo94], [Don95] :

$$\Lambda_r(X) \equiv \frac{p_1(X)}{p_0(X)} \begin{cases} > \\ < \end{cases} \begin{matrix} d(X)=1 \\ d(X)=0 \end{matrix} \frac{P_0 (C_{10} - C_{00})}{P_1 (C_{01} - C_{11})}, \quad (3.2)$$

où P_j est la probabilité *a priori* de l'hypothèse H_j .

Ainsi, le choix de la règle de Bayes induit la structure du test de détection, celle du détecteur à seuil présenté en Figure 3.2, à condition que les densités $p_0(x)$ et $p_1(x)$ soient connues de façon explicite, tout comme les coûts et les probabilités *a priori* P_0 et P_1 . Généralement, trois cas particuliers du critère de Bayes, correspondant chacun à un choix particulier de coûts, peuvent être rencontrés :

a) *Critère du maximum de probabilité a posteriori*. Il consiste à choisir les coûts de façon à ce qu'ils vérifient :

$$C_{10} - C_{00} = C_{01} - C_{11}.$$

Dans ce cas, l'hypothèse retenue maximise la probabilité *a posteriori*, la règle de décision définie par l'expression (3.2) se réécrivant ainsi :

$$P_1 p_1(X) \begin{cases} > \\ < \end{cases} \begin{matrix} d(X)=1 \\ d(X)=0 \end{matrix} P_0 p_0(X).$$

b) *Critère du minimum de probabilité d'erreur*. Ce critère suggère d'attribuer un coût nul à une décision correcte et un coût unitaire à une erreur. Le détecteur obtenu, que l'on note d_{Bayes} , minimise la probabilité d'erreur P_e :

$$d_{Bayes} = \arg \min_d P_e(d) \text{ avec } P_e(d) = \mathbf{P}\{d(X) \neq Y\}.$$

c) *Critère de type Minimax*. Ce critère se substitue à celui de Bayes lorsque les probabilités *a priori* des hypothèses sont inconnues. Elle implique la recherche de la probabilité P_0 la plus défavorable pour le détecteur de Bayes, c'est à dire celle qui maximise le coût moyen minimum.

Le critère de Neyman-Pearson

Lorsque les coûts et les probabilités *a priori* sont inconnus, ce qui ne permet alors pas d'appliquer les règles de décision décrites précédemment, on est amené à raisonner sur les probabilités conditionnelles. La règle de Neyman-Pearson consiste précisément à maximiser la probabilité de détection P_d en ayant fixé au préalable la probabilité de fausse alarme P_{fa} à un niveau donné, P_d et P_{fa} étant respectivement définis par :

$$P_d(d) = \mathbf{P}\{d(X) = 1 \mid Y = 1\},$$
$$P_{fa}(d) = \mathbf{P}\{d(X) = 1 \mid Y = 0\}.$$

On peut montrer que la résolution de ce problème d'optimisation avec contrainte d'inégalité conduit à la solution suivante [Sch91], [Poo94], [Don95] :

$$\Lambda_{rv}(X) \equiv \frac{p_1(X)}{p_0(X)} \begin{matrix} > \\ < \end{matrix} \begin{matrix} d(X)=1 \\ d(X)=0 \end{matrix} \lambda_0. \quad (3.3)$$

Dans ce cas, le seuil λ_0 est fixé par la probabilité P_{fa} , et peut être calculé grâce à l'expression :

$$P_{fa} = \int_{\lambda_0}^{+\infty} p_0(\Lambda(x)) d\Lambda,$$

où $p_0(\Lambda(x))$ désigne la loi de probabilité de $\Lambda(x)$, conditionnellement à l'hypothèse H_0 .

En conclusion, les stratégies bayésienne et de Neyman-Pearson coïncident : elles conduisent toutes deux à la comparaison du rapport de vraisemblance à un seuil λ_0 , seul ce dernier dépendant de la règle de décision choisie.

3.2.2 Hypothèses composées

Jusqu'à présent, nous avons supposé que les lois de vraisemblance conditionnelles étaient parfaitement connues. Il est cependant fréquent que cela ne soit pas le cas et que l'on soit tenu de prendre une décision sans connaître explicitement un vecteur de paramètres θ . Ceci nous amène à distinguer trois types d'hypothèses, dites *hypothèses composées* : θ aléatoire avec $p_i(\theta)$ connues, θ aléatoire avec $p_i(\theta)$ inconnues ou θ déterministe inconnu. Pour chacun de ces cas, des éléments de solutions vont être présentés.

Paramètres aléatoires, lois connues

La règle de Bayes peut être appliquée si les probabilités *a priori* P_i , les lois conditionnelles $p_i(x)$ et $p_i(\theta)$ ainsi que les coûts $C_{ij}(\theta)$ sont connus. Par probabilisation des hypothèses, on obtient alors :

$$\frac{\int [C_{01}(\theta) - C_{11}(\theta)] p_1(X | \theta) p_1(\theta) d\theta}{\int [C_{10}(\theta) - C_{00}(\theta)] p_0(X | \theta) p_0(\theta) d\theta} \underset{d(X)=0}{\overset{d(X)=1}{>}} \frac{P_0}{P_1}. \quad (3.4)$$

Lorsque les coûts C_{ij} sont indépendants du vecteur de paramètres θ , on reconnaît dans cette expression la structure du test de détection de Bayes présenté dans le cadre d'hypothèses simples.

Paramètres aléatoires, lois inconnues

Les lois conditionnelles $p_i(\theta)$ sont rarement connues et il est alors impossible d'appliquer le critère de Bayes. Du fait qu'il n'existe pas de solution analytique générale à ce problème, on détermine habituellement un type de loi susceptible d'être suivie par le vecteur aléatoire θ . On retient alors, pour paramètres caractéristiques de cette loi, ceux qui maximisent le coût moyen minimum de Bayes. Il s'agit par conséquent d'une approche de type *Minimax*.

Paramètres déterministes inconnus

Les paramètres des lois de vraisemblance étant supposés déterministes inconnus, il est alors impossible d'appliquer la règle de Bayes, le coût moyen ne pouvant être évalué. Ceci conduit généralement à adopter une démarche de type Neyman-Pearson, comme nous allons le voir.

a) *Hypothèse simple contre hypothèse composée.* On suppose que H_0 est une hypothèse simple et que H_1 est une hypothèse composée, $p_1(X|\theta)$ dépendant d'un vecteur θ de paramètres inconnus. Une démarche usuelle consiste à rechercher, s'il existe, le *test uniformément le plus puissant*. Celui-ci est optimal au sens de la règle de Neyman-Pearson, quel que soit θ : sa probabilité de détection est supérieure à celle de n'importe quel autre test, à probabilité de fausse alarme fixée et indépendamment du vecteur de paramètres inconnus. Lorsque ce test n'existe pas, on peut estimer θ par la méthode du maximum de vraisemblance, ce qui nous conduit à la structure de détection suivante :

$$\Lambda(X) = \frac{\max_{\theta} p_1(X | \theta)}{p_0(X)} \underset{d(X)=0}{\overset{d(X)=1}{>}} \lambda_0.$$

b) *Hypothèse composée contre hypothèse composée.* Les conditions d'existence d'un test uniformément le plus puissant sont extrêmement restrictives lorsque H_0 et H_1 sont des hypothèses composées [Sap90]. Pour cette raison, on a souvent recours à l'estimation des paramètres inconnus par la méthode du maximum de vraisemblance, sous chacune des hypothèses. Le test de détection équivaut alors à comparer le quotient $\Lambda_g(X)$, appelé *rapport de vraisemblance généralisé*, à un seuil :

$$\Lambda_g(X) \equiv \frac{\max_{\theta} p_1(X | \theta)}{\max_{\theta} p_0(X | \theta)} \underset{d(X)=0}{\overset{d(X)=1}{>}} \lambda_0. \quad (3.5)$$

Le seuil de détection λ_0 est généralement fixé par le choix de la probabilité de fausse alarme. Notons enfin que ce détecteur ne requiert aucune connaissance *a priori* sur le vecteur θ . Ceci justifie certainement l'intérêt qui lui est porté, bien qu'il ne soit pas nécessairement optimal.

L'objet de la prochaine section est d'illustrer les principes généraux qui viennent d'être exposés, par la résolution de deux problèmes classiques de détection. Dans le second exemple traité en particulier, on présente également une formulation temps-fréquence de la statistique de détection afin de montrer que ce point de vue n'est pas dénué d'intérêt. Pour ce faire, les signaux considérés sont supposés continus, cadre de travail partagé par l'abondante littérature traitant de la détection à structure libre dans le plan temps-fréquence [Fla88], [Lem95], [Say95], [Say96a]. On ne doit cependant pas ignorer que cette hypothèse nécessite le recours à des outils mathématiques spécifiques, tels que la notion de densité dans un espace de fonctions et autre théorème de Radon Nikodym. Le lecteur intéressé par ce sujet est invité à consulter la référence [Poo94].

3.2.3 Exemples de problèmes de détection

Le problème de détection considéré consiste à décider de la présence ou de l'absence d'un signal s noyé dans un bruit B , en optant pour l'une des hypothèses suivantes :

$$\begin{cases} H_0 : X(t) = B(t) \\ H_1 : X(t) = s(t) + B(t) \end{cases}, t \in T.$$

Le signal s à détecter durant l'intervalle d'observation T peut être supposé déterministe ou aléatoire, réel ou complexe.

Signal parfaitement connu

Le premier cas considéré suppose que le signal s est déterministe et parfaitement connu, et que le bruit additif B est blanc, gaussien et centré. Dans ces circonstances, on peut montrer que la statistique de détection optimale au sens du rapport de vraisemblance est définie par le produit scalaire de X par s [Poo94] :

$$\Lambda(X) = \int_0^T X(t) s^*(t) dt. \quad (3.6)$$

Historiquement, il a d'abord été établi que ce récepteur, appelé *filtre adapté*, constitue le système linéaire maximisant le rapport signal sur bruit [Nor63], [Lem95]. Ce n'est

qu'ultérieurement que son optimalité au sens du rapport de vraisemblance a été démontrée [Tre68]. Ceci constitue un cas intéressant où le choix d'une règle de décision, telle que celle de Bayes, conduit au même résultat que l'optimisation d'un détecteur à structure imposée par maximisation d'un critère de contraste, en l'occurrence le rapport signal sur bruit dans le cas présent.

Signal à paramètres inconnus

Dans la majorité des problèmes de détection rencontrés, la phase initiale du signal reçu reste cependant difficile à prédire. En conséquence, celle-ci est classiquement supposée uniformément distribuée sur l'intervalle $[0, 2\pi]$, [Tre68], [Lem95]. Conditionnellement à l'hypothèse H_1 , on note dans ce cas :

$$X(t) = s(t; \phi_0) + B(t), \text{ avec } s(t; \phi_0) = s(t) e^{j\phi_0}, \quad (3.7)$$

où ϕ_0 constitue l'unique paramètre inconnu du problème.

Moyennant certaines hypothèses sur les vitesses de modulation d'amplitude et de phase du signal s , on peut montrer que le développement du rapport de vraisemblance conduit à une statistique de détection équivalente à [Tre68], [Lem95] :

$$\Lambda(X) = \left| \int_0^T X(t) s^*(t) dt \right|^2, \quad (3.8)$$

qui correspond à la sortie d'un *filtre adapté avec détecteur d'enveloppe quadratique*. En appliquant la relation de Moyal (1.22), il est remarquable que cette statistique de détection peut être reformulée ainsi, dans le domaine temps-fréquence :

$$\Lambda(X) = \int \int_{t \in T} W_X(t, \nu) W_s^*(t, \nu) dt d\nu. \quad (3.9)$$

S'agissant d'un produit scalaire entre les représentations de Wigner-Ville de l'observation et du signal à détecter, la structure de détection (3.9) est qualifiée de *filtre adapté temps-fréquence*, par analogie avec le filtre adapté présenté précédemment.

Plus généralement, la correspondance dite de Weyl [Koz92], dont la relation de Moyal constitue un cas particulier, associe à toute statistique de détection quadratique donnée par :

$$\Lambda(X) = \int \int_t X(t) a(t, t') X^*(t') dt dt', \quad (3.10)$$

la formulation temps-fréquence équivalente suivante :

$$\Lambda(X) = \int \int_{t \nu} W_X(t, \nu) a_{TF}(t, \nu) dt d\nu, \quad (3.11)$$

$$\text{en définissant } a_{TF} \text{ ainsi : } a_{TF}(t, \nu) \equiv \int_{\tau} a\left(t + \frac{\tau}{2}, t - \frac{\tau}{2}\right) e^{-j2\pi\nu\tau} d\tau. \quad (3.12)$$

En offrant de la sorte, à toute statistique de détection quadratique, une formulation en termes de somme pondérée des composantes de la représentation de Wigner-Ville de l'observation, l'analyse temps-fréquence permet une interprétation aisée de la configuration des détecteurs concernés. Pour cette raison, la majorité des expérimentations présentées dans la suite de ce document, qui visent à illustrer l'efficacité de la méthodologie proposée pour la synthèse de détecteurs à structure imposée, repose sur la statistique de détection (3.11).

Mais là n'est pas le seul avantage de cette approche. Certains auteurs ont également montré que la formulation temps-fréquence (3.11) permet parfois de simplifier l'implémentation d'un détecteur quadratique, notamment lorsque les localisations temporelle et fréquentielle du signal à détecter sont inconnues [Lem95], [Say95], [Kro97].

3.3 Détection à structure imposée

Jusqu'à présent, les propriétés statistiques du signal et du bruit étaient parfaitement connues. La structure du détecteur ne nécessitait pas d'être préalablement fixée mais résultait de l'application d'un critère de décision et de la connaissance de $p_0(x)$ et $p_1(x)$. Cependant, dans un grand nombre de situations (données météorologiques, électroencéphalogrammes, etc.), on ne dispose pas du confort absolu que constitue la connaissance des lois de probabilité conditionnelles, ce qui rend impossible la détermination du rapport de vraisemblance. L'objet de cette section est de présenter une méthodologie pour l'élaboration d'un test de détection offrant certaines garanties en termes de probabilité d'erreur, lorsqu'on ne dispose que d'un ensemble de n réalisations étiquetées $(x^{(i)}, y^{(i)})$. Dans un tel contexte, la démarche naturellement adoptée consiste à rechercher dans une classe de détecteurs préalablement définie, une solution optimale au sens d'un critère ne dépendant que de ces données. Cette approche est qualifiée de détection à structure imposée.

3.3.1 Présentation du problème

Soit \mathcal{A}_n un ensemble de n échantillons $(x^{(i)}, y^{(i)})$. On suppose que ces données sont des réalisations de n couples aléatoires $(X^{(i)}, Y^{(i)})$ indépendants, identiquement distribués suivant la loi $\mathbf{P}\{(X; Y)\}$ inconnue. Soit \mathcal{D} une classe de tests de détection $d: \mathcal{R}^L \rightarrow \{0; 1\}$, où L est la dimension du vecteur aléatoire X .

Etant donné \mathcal{A}_n , la synthèse d'un détecteur optimal à structure imposée consiste à rechercher dans un ensemble \mathcal{D} donné, un détecteur minimisant la probabilité d'erreur conditionnelle suivante :

$$P_e(d_n) \equiv \mathbf{P}\{d_n(X, \mathcal{A}_n) \neq Y \mid \mathcal{A}_n\}. \quad (3.13)$$

Dans cette expression, $d_n(X, \mathcal{A}_n)$ désigne un détecteur de \mathcal{D} déterminé sur la base des n éléments (réalisations) de \mathcal{A}_n , au terme d'un processus que l'on nomme *apprentissage*. Le problème de la minimisation de $P_e(d_n)$ ne pouvant être résolu puisque la loi $\mathbf{P}\{(X; Y)\}$ est inconnue, on définit généralement un critère $c(d_n, \mathcal{A}_n)$ dont l'écriture ne dépend que des éléments de \mathcal{A}_n , et dont l'optimisation conduit à un test de détection offrant certaines garanties en termes de probabilité d'erreur. A titre d'exemple, la *probabilité d'erreur empirique* est l'un de ces critères qui possèdent les qualités requises :

$$P_{emp}(d_n, \mathcal{A}_n) \equiv \frac{1}{n} \sum_{i=1}^n \delta(d_n(x^{(i)}) - y^{(i)}). \quad (3.14)$$

A l'usage, l'optimisation de P_{emp} s'avère cependant problématique. Ceci justifie les nombreuses alternatives qui ont pu être proposées [Dud73], [Dub90], [Dev96], les plus significatives faisant l'objet de la section 3.3.3.

Soit d_{Bayes} le détecteur de Bayes obtenu par minimisation de la probabilité d'erreur lorsque les lois de probabilité sous les hypothèses H_0 et H_1 sont parfaitement connues. Soit $d_{n,opt}(X, \mathcal{A}_n)$ le détecteur optimal au sens d'un critère $c(d_n, \mathcal{A}_n)$ donné. Comme l'illustre la Figure 3.3, l'*erreur de modélisation* définie par l'expression (3.15) est représentative de l'efficacité de $d_{n,opt}(X, \mathcal{A}_n)$.

$$E_{mod}(d_{n,opt}) \equiv P_e(d_{n,opt}) - P_e(d_{Bayes}), \quad (3.15)$$

que l'on peut réécrire ainsi :

$$E_{mod}(d_{n,opt}) = \underbrace{\left[P_e(d_{n,opt}) - \inf_{d \in \mathcal{D}} P_e(d) \right]}_{E_{estim}} + \underbrace{\left[\inf_{d \in \mathcal{D}} P_e(d) - P_e(d_{Bayes}) \right]}_{E_{approx}}. \quad (3.16)$$

La reformulation de la définition de E_{mod} met en évidence l'influence de deux sources d'erreur de natures différentes : l'*erreur d'estimation* et l'*erreur d'approximation*. La première, notée E_{estim} , reflète conjointement la pertinence du critère $c(d_n, \mathcal{A}_n)$ choisi et l'efficacité du processus d'apprentissage qui lui est associé, étant donnés un ensemble de données \mathcal{A}_n et une famille de tests de détection \mathcal{D} . La seconde, notée E_{approx} , ne dépend que du choix initial de \mathcal{D} .

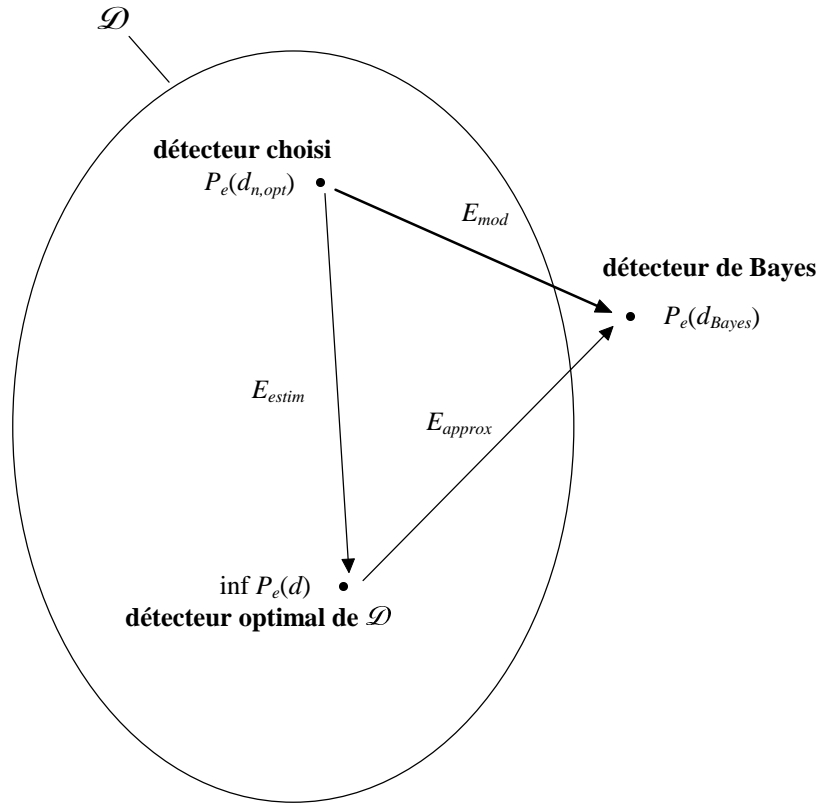


Figure 3.3 : Recherche d'un détecteur $d_{n,opt}$ dans une classe \mathcal{D} donnée par optimisation d'un critère dont l'écriture ne dépend que des n échantillons d'un ensemble d'apprentissage.

La minimisation de l'erreur de modélisation repose sur la recherche d'un compromis entre ces deux termes antagonistes : l'augmentation du nombre de tests de \mathcal{D} conduit à un accroissement de E_{estim} tandis que E_{approx} décroît, et inversement. Pour se convaincre de l'existence de ce compromis, qui a été formellement étudié par Vapnik et Chervonenkis dans [Vap71] et qui sera abordé dans le Chapitre 4, supposons que \mathcal{D} regroupe l'ensemble des fonctions mesurables $d: \mathcal{R}^L \rightarrow \{0;1\}$ et que $c(d, \mathcal{A}_n)$ désigne la probabilité d'erreur empirique. Dans ces circonstances, et quelle que soit la loi $\mathbf{P}\{X; Y\}$, on peut trouver un test de détection dont la probabilité d'erreur empirique est nulle, et dont le résultat est quelconque pour toute réalisation $X=x$ n'appartenant pas à \mathcal{A}_n :

$$d_n(X, \mathcal{A}_n) = \begin{cases} y^{(i)} & \text{si } X = x^{(i)}, i = 1, \dots, n \\ 0 \text{ ou } 1 & \text{de façon aléatoire, sinon.} \end{cases}$$

Ce détecteur conduit ainsi à l'annulation de l'erreur d'approximation au détriment de l'erreur d'estimation. On notera que le cas dual consistant à maximiser E_{approx} et à annuler E_{estim} peut être observé en restreignant \mathcal{D} à un test de détection unique.

La suite de ce mémoire est consacrée aux problèmes qui se posent lors de la conception d'un détecteur à structure imposée, et que cette introduction a permis d'identifier :

1. Le choix de la classe \mathcal{D} autorisant une erreur d'approximation nulle, sans pour autant être dans le cas trivial présenté précédemment, est exposé en section 3.3.2.
2. La recherche de critères $c(d, \mathcal{A}_n)$ pertinents et d'algorithmes d'apprentissage pour la minimisation de l'erreur d'estimation fait l'objet de la section 3.3.3.
3. La recherche automatique d'un compromis entre l'erreur d'approximation et l'erreur d'estimation est présentée au Chapitre 4.

Le recours à des résultats de la théorie de l'apprentissage et de la reconnaissance des formes afin d'apporter des éléments de réponses à ces questions constitue, en plus de quelques contributions personnelles, l'une des originalités de ce travail.

3.3.2 Construction de la classe de détecteurs \mathcal{D}

On distingue au moins deux types d'approches pour la résolution d'un problème de détection. La première concerne la détection dite à structure libre, pour laquelle l'expression du test résulte de l'application d'un critère de détection et de la connaissance des lois de probabilité de l'observation, conditionnellement aux hypothèses. La deuxième, appelée détection à structure imposée, suppose de définir préalablement une classe \mathcal{D} de tests de détection, pour ne retenir que celui qui est optimal au sens d'un critère donné. L'objet de cette section est de guider le lecteur dans son choix de \mathcal{D} .

a) *Structures linéaires.* Le test de détection le plus simple consiste à supposer que la statistique $\Lambda(X)$ est une fonction linéaire des composantes de l'observation :

$$\Lambda(X) = \sum_{i=1}^L a_i X[i] - \lambda_0 = a^T X - \lambda_0 \underset{d(X)=0}{\overset{d(X)=1}{>}} 0. \quad (3.17)$$

Ce type de détecteur présente l'avantage d'être directement lié à la notion de filtrage, ce qui en facilite l'interprétation. Cependant, il n'offre des solutions optimales que pour une classe réduite de problèmes. En particulier, lorsque H_0 et H_1 suivent des lois gaussiennes, il est impératif que celles-ci aient les mêmes propriétés statistiques d'ordre deux [Don95]. Si cette hypothèse est raisonnable dans le cadre de certaines applications, la détection dans certains systèmes de communication par exemple, celle-ci reste inappropriée dans la majorité des cas.

Afin de répondre de façon satisfaisante à un plus grand nombre de problèmes, on peut envisager d'enrichir la classe \mathcal{D} en adjoignant au test de détection défini par l'expression

(3.17) l'information relative aux produits croisés des composantes $X[i]$ du vecteur X . On obtient dans ce cas une famille de détecteurs reposant sur une statistique de détection quadratique :

$$\Lambda(X) = \sum_{i=1}^L a_i X[i] + \sum_{i=1}^L \sum_{j=1}^L b_{ij} X[i] X^*[j] - \lambda_0 = a^T X + X^T b X^* - \lambda_0 \begin{matrix} \geq \\ < \end{matrix} \begin{matrix} d(X)=1 \\ d(X)=0 \end{matrix} 0. \quad (3.18)$$

Ces tests peuvent conduire en particulier à une situation d'optimalité au sens du critère de Bayes lorsque les hypothèses H_0 et H_1 suivent des lois gaussiennes de paramètres distincts [Don95], en supposant bien entendu que l'on puisse accéder à a et b .

Plus généralement, on peut démontrer que, moyennant certaines hypothèses sur $p_0(x)$ et $p_1(x)$, la statistique du détecteur de Bayes admet des décompositions de la même nature que celles qui viennent d'être évoquées.

Proposition 3.1. [Dev96]. Soit p la loi de vraisemblance de la variable aléatoire X . Soit $\{z^{(i)}\}_i$ un ensemble orthonormé complet de fonctions bornées appartenant à $\mathcal{L}_2(\lambda)$, où λ désigne la mesure de Lebesgue sur \mathcal{R}^L . Si p appartient à $\mathcal{L}_2(\lambda)$, alors les densités conditionnelles appartiennent également à $\mathcal{L}_2(\lambda)$, et une forme équivalente $\Lambda_{Bayes}(X)$ de la statistique de détection du détecteur de Bayes admet le développement suivant :

$$\Lambda_{Bayes}(X) \equiv P_1 p_1(X) - (1 - P_1) p_0(X) = \sum_{i=1}^{+\infty} a_i z^{(i)}(X), \quad (3.19)$$

où les coefficients a_i sont donnés par :

$$a_i = \int z^{(i)}(x) \Lambda_{Bayes}(x) dx.$$

La convergence du développement (3.19) est à considérer en moyenne quadratique :

$$\lim_{K \rightarrow +\infty} \int \left(\Lambda_{Bayes}(x) - \sum_{i=1}^K a_i z^{(i)}(x) \right)^2 dx = 0.$$

Remarque. Parmi les ensembles $\{z^{(i)}\}_i$ possédant les propriétés énoncées ci-dessus, on peut recenser par exemple la base standard des fonctions trigonométriques, les polynômes de Legendre ou encore les bases de Laguerre et de Haar [Dev96].

Soit \mathcal{D} la classe des *détecteurs linéaires généralisés* définie par :

$$d(x) = \begin{cases} 0 & \text{si } \sum_{i=1}^K a_i z^{(i)}(x) < 0 \\ 1 & \text{sinon,} \end{cases} \quad (3.20)$$

où $z^{(i)}(x)$ est une fonction quelconque de la variable x .

La Proposition 3.1 établit donc la convergence vers 0 de l'erreur d'approximation définie par l'expression (3.16), lorsque K tend vers l'infini et que $\{z^{(i)}\}_{1 \leq i \leq K}$ est un ensemble orthonormé complet de fonctions bornées de $\mathcal{L}_2(\lambda)$. Bien que séduisant, ce résultat nécessite cependant une certaine connaissance *a priori* de la loi de vraisemblance de l'observation X , qui est inconnue par hypothèse. Ceci justifie le rejet provisoire de la classe des détecteurs linéaires généralisés, les hypothèses requises par la Proposition 3.1 n'étant pas satisfaisante. Le résultat présenté dans le prochain paragraphe est beaucoup plus intéressant puisqu'il concerne la convergence de l'erreur de modélisation grâce à un choix approprié de l'ensemble $\{z^{(i)}\}_{1 \leq i \leq K}$, sans qu'aucune hypothèse ne soit faite sur $p_0(x)$ et $p_1(x)$.

b) *Notions de consistance* [Dev96]. Il est généralement impossible d'atteindre les performances du détecteur de Bayes avec un détecteur à structure imposée déterminé à partir d'un ensemble d'apprentissage de dimension finie n . Cependant, on peut espérer qu'il existe dans la classe \mathcal{D} , une suite $\{d_{n,opt}(X, \mathcal{A}_n)\}_{n>0}$ de détecteurs obtenus par optimisation d'un critère donné sur \mathcal{A}_n , telle que la probabilité d'erreur

$$P_e(d_{n,opt}) \equiv \mathbf{P}\{d_{n,opt}(X, \mathcal{A}_n) \neq Y \mid \mathcal{A}_n\}$$

puisse être rendue arbitrairement proche de $P_e(d_{Bayes})$ en faisant tendre n vers l'infini. La notion de *consistance* repose sur cette idée.

Définition 3.2. Une suite $\{d_{n,opt}(X, \mathcal{A}_n)\}_{n>0}$ de détecteurs déterminés par optimisation d'un critère donné sur \mathcal{A}_n est dite consistante pour une loi $\mathbf{P}\{X; Y\}$ si :

$$\lim_{n \rightarrow +\infty} \mathbf{E}\{P_e(d_{n,opt})\} \equiv \lim_{n \rightarrow +\infty} \mathbf{P}\{d_{n,opt}(X, \mathcal{A}_n) \neq Y\} = P_e(d_{Bayes}),$$

et fortement consistante si :

$$\lim_{n \rightarrow +\infty} P_e(d_{n,opt}) = P_e(d_{Bayes}) \text{ avec une probabilité égale à 1.}$$

La propriété de consistance garantit que la probabilité d'erreur d'un détecteur, obtenu par optimisation d'un critère sur une base d'apprentissage donnée de taille n , converge vers $P_e(d_{Bayes})$ lorsque n tend vers l'infini. La suite $\{d_{n,opt}(X, \mathcal{A}_n)\}_{n>0}$ est dite fortement convergente si cette propriété est vraie quelle que soit la réalisation de \mathcal{A}_n . On peut enfin distinguer le cas où la propriété de consistance n'est vérifiée que pour une loi de probabilité $\mathbf{P}\{X; Y\}$ donnée, du cas où elle reste vraie indépendamment de celle-ci. La définition suivante prend en compte le caractère parfois universel de la consistance.

Définition 3.3. La suite $\{d_{n,opt}(X, \mathcal{A}_n)\}_{n>0}$ est dite universellement (fortement) consistante si elle est (fortement) consistante pour toute loi $\mathbf{P}\{X; Y\}$.

Fort de ces définitions issues de la théorie de l'apprentissage, on peut désormais exposer le résultat principal de cette section. Celui-ci émane d'une proposition plus générale, énoncée et démontrée dans [Dev96], que l'on a adaptée au problème qui motive cette section :

Proposition 3.4. (démonstration dans [Dev96]). La suite $\{d_{n,opt}(X, \mathcal{A}_n)\}_n$ des détecteurs linéaires généralisés résultant de la minimisation de l'erreur empirique sur \mathcal{A}_n est universellement fortement consistante si :

$K \rightarrow +\infty$ et $K/n \rightarrow 0$ lorsque $n \rightarrow +\infty$,
et si les fonctions $z^{(1)}, z^{(2)}, \dots, z^{(L)}, z^{(L+1)}, \dots, z^{(K)}$, désignent (en particulier) les polynômes $X[1], X[2], \dots, X[L], X[1] X[2], \dots$. On rappelle que $X[j]$ correspond à la $j^{\text{ème}}$ composante du vecteur aléatoire X .

Cette proposition donne ainsi une légitimité à une classe particulière des détecteurs linéaires généralisés, à condition que la taille de la base d'apprentissage croisse plus vite que le degré des polynômes. En conséquence, dans la suite de ce document et sauf indication contraire, les tests de détection considérés seront de la forme :

$$d(Z) = \begin{cases} 0 & \text{si } \Lambda(Z) \equiv a^T Z - \lambda_0 < 0 \\ 1 & \text{sinon,} \end{cases} \quad (3.21)$$

$Z[1], \dots, Z[L], Z[L+1], \dots$ désignant les polynômes $X[1], \dots, X[L], X[1] X[2], \dots$, et a un vecteur à composantes réelles.

En conclusion, en plus des garanties de performances que représente la consistance universellement forte de cette classe, la détermination d'un détecteur est facilitée par le fait qu'elle s'apparente à la recherche d'un discriminant linéaire dans l'espace des transformées Z . Ce thème, qui fait l'objet d'une abondante littérature en reconnaissance des formes, est abordé dans la prochaine section.

3.3.3 Choix et optimisation du critère $c(d, \mathcal{A}_n)$

a) Présentation du problème. L'étude menée précédemment a permis de caractériser une classe \mathcal{D} de détecteurs autorisant une erreur d'approximation nulle, en l'occurrence celle des structures linéaires généralisées définies par la Proposition 3.4. Le problème se posant désormais est celui de la recherche d'un critère dont l'optimum correspond au minimum de l'erreur d'estimation. On rappelle que E_{estim} est définie ainsi :

$$E_{estim} \equiv P_e(d_{n,opt}) - \inf_{d \in \mathcal{D}} P_e(d), \quad (3.22)$$

avec $d_{n,opt} \equiv \arg \min_{d \in \mathcal{D}} c(d, \mathcal{A}_n)$, ou $\arg \max_{d \in \mathcal{D}} c(d, \mathcal{A}_n)$, selon la nature de c . (3.23)

Il est démontré dans [Dev96] que la probabilité d'erreur empirique P_{emp} correspond au critère c recherché. Ceci signifie que le test de \mathcal{D} le plus performant que l'on peut obtenir, à partir de la seule connaissance d'un ensemble d'apprentissage \mathcal{A}_n , est celui qui minimise la probabilité d'erreur empirique sur cet ensemble. Si ce résultat est intuitivement satisfaisant, on peut néanmoins regretter que pour ce critère, il est impossible de résoudre l'équation (3.23) de façon exacte. En effet, la recherche d'une dichotomie optimale de n points étiquetés de \mathcal{R}^L est un problème démontré NP difficile [Joh78], qui nécessite le recours à des méthodes de résolution approchées. Dans ces circonstances, une démarche naturelle pour la détermination d'un optimum local pourrait consister à utiliser une méthode d'optimisation classique. Cependant, P_{emp} est une fonction constante par morceaux en raison de la nature échantillonnée de \mathcal{A}_n , ce qui rend l'opération de minimisation particulièrement délicate. Afin de pallier cette difficulté, de nombreux auteurs ont proposé de modifier quelque peu ce critère, au moyen d'un lissage par exemple, afin de limiter le nombre de minima locaux [Was72], [Dot75], [Fri76], [Wid88], [Sk179]. Dans [Dev96], il est toutefois fait état du caractère parfois discutable des solutions obtenues.

b) *Critères de contraste.* Une autre approche consiste à adopter un critère, dit *de contraste*, qui quantifie le caractère discriminant d'une statistique de détection $\Lambda(Z)$ pour un problème donné, à la manière d'une mesure de distance entre les lois de probabilité $p_0(\Lambda(z))$ et $p_1(\Lambda(z))$. Ce principe est illustré par la Figure 3.4. Certains de ces critères tels que les divergences de Jeffreys et de Kullback-Leibler, l'affinité de Bhattacharyya ou encore la distance de Kolmogorov [Dev96], nécessitent la connaissance de $p_0(Z)$ et $p_1(Z)$ [Dev96]. En contrepartie,

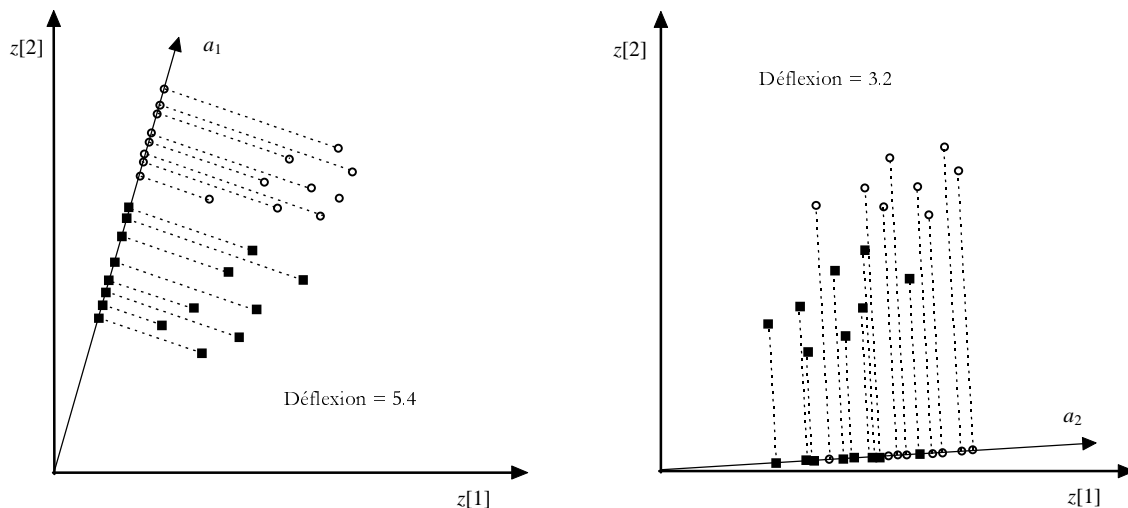


Figure 3.4 : Projection des éléments étiquetés {■ ; ○} d'une base d'apprentissage sur des droites de vecteurs directeurs notés a_1 et a_2 . L'évaluation du critère de contraste que constitue la déflexion laisse apparaître que la statistique linéaire associée à a_1 est plus discriminante pour le problème posé que celle associée à a_2 .

ils fournissent des informations de qualité sur les performances que l'on est en droit d'attendre d'un détecteur, étant donné qu'ils sont étroitement liés à l'erreur de Bayes. D'autres critères se contentent en revanche d'une quantité d'information plus modeste, qui se limite généralement aux moments conditionnels d'ordre un et deux de la statistique de détection $\Lambda(\mathbf{Z})$. Dans ce cas, des arguments simples basés sur des considérations géométriques suffisent bien souvent à justifier le bien-fondé de leur définition, que renforcent des propriétés théoriques qui seront évoquées dans la suite de cette section. Ces critères, qui nous intéressent plus particulièrement en raison du peu d'information *a priori* qu'ils requièrent, font l'objet d'une abondante littérature. Pour cette raison, on ne citera dans ce travail que les plus connus, laissant le soin au lecteur de consulter par exemple [Dud73], [Gar80], [Dub90] et [Dev96].

Soient η_i et σ_i^2 l'espérance mathématique et la variance conditionnelles de la statistique $\Lambda(\mathbf{Z})$. Le *rapport signal sur bruit*, la *déflexion* et le *critère de Fisher*, dont les appellations varient selon les ouvrages, sont respectivement définis par :

$$c_{RSB}(\eta_0, \eta_1, \sigma_0^2, \sigma_1^2) = \frac{(\eta_1 - \eta_0)^2}{\sigma_0^2}, \quad (3.24)$$

$$c_{Deflex}(\eta_0, \eta_1, \sigma_0^2, \sigma_1^2) = 2 \frac{(\eta_1 - \eta_0)^2}{\sigma_0^2 + \sigma_1^2}, \quad (3.25)$$

$$c_{Fisher}(\eta_0, \eta_1, \sigma_0^2, \sigma_1^2) = \frac{(\eta_1 - \eta_0)^2}{P_0 \sigma_0^2 + (1 - P_0) \sigma_1^2}. \quad (3.26)$$

L'optimisation de ces critères peut être interprétée comme la maximisation de la dispersion interhypothèse Δ_{inter} effectuée conjointement avec la minimisation de la dispersion intrahypothèse Δ_{intra} , en supposant que l'on adopte les définitions suivantes :

$$\Delta_{inter} = (\eta_1 - \eta_0)^2,$$

$$\Delta_{intra} = p \sigma_0^2 + (1 - p) \sigma_1^2,$$

où $\{p; 1-p\}$ désigne une pondération donnée.

L'interprétation de c_{RSB} , c_{Deflex} et c_{Fisher} s'avère par conséquent satisfaisante en termes de mesure de contraste entre les hypothèses H_0 et H_1 , lorsque $p_0(\Lambda(\mathbf{Z}))$ et $p_1(\Lambda(\mathbf{Z}))$ sont unimodales comme sur la Figure 3.5. Dans le cas contraire, les résultats obtenus peuvent prêter à discussion. On conseille dans ce cas d'augmenter le degré de la statistique de détection polynomiale définie par l'équation (3.21).

Le résultat théorique suivant contribue à justifier l'intérêt que présentent de tels critères de contraste ne dépendant que des moments d'ordre 1 et 2 de la statistique de détection $\Lambda(\mathbf{Z})$, lorsqu'on ne dispose pas de la connaissance de $p_0(\mathbf{Z})$ et $p_1(\mathbf{Z})$.

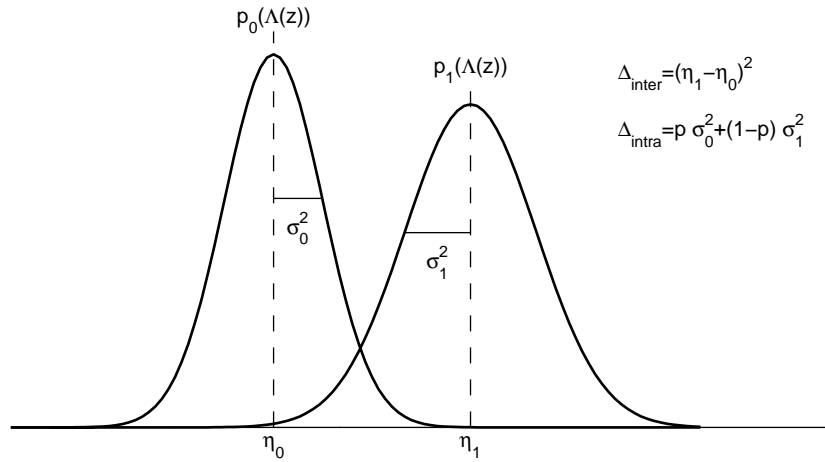


Figure 3.5 : Définition des dispersions intrahypothèse et interhypothèse. Principe d'un critère de contraste ne dépendant que des moments d'ordre 1 et 2 de la statistique de détection.

Proposition 3.5. [Fuk90]. Le rapport de vraisemblance maximise tout critère de contraste c ne dépendant que des moments d'ordre 1 et 2 de la statistique de détection.

Preuve. Soit d un test de détection défini par :

$$d(Z) = \begin{cases} 0 & \text{si } \Lambda(Z) < 0 \\ 1 & \text{sinon.} \end{cases} \quad (3.27)$$

On suppose que la statistique de détection $\Lambda(Z)$ est optimale au sens d'un critère $c(\eta_0, \eta_1, s_0^2, s_1^2)$ quelconque, tel que :

$$\eta_i = E\{\Lambda(Z) | H_i\} = \int \Lambda(z) p_i(z) dz,$$

$$s_i^2 = E\{\Lambda^2(Z) | H_i\} = \int \Lambda^2(z) p_i(z) dz.$$

Pour des raisons de commodité d'écriture, le moment s_i^2 a été préféré à la variance σ_i^2 . Le résultat final n'en est pas pour autant affecté, c demeurant une fonction de η_i et σ_i^2 , à l'image des différents critères évoqués précédemment. En effet, on rappelle que :

$$s_i^2 = \eta_i^2 + \sigma_i^2.$$

Le calcul des variations de c , notées δc , en fonction de celles de $\Lambda(Z)$, notées $\delta\Lambda(Z)$, conduit à :

$$\delta c = \frac{\partial c}{\partial s_0^2} \delta s_0^2 + \frac{\partial c}{\partial s_1^2} \delta s_1^2 + \frac{\partial c}{\partial \eta_0} \delta \eta_0 + \frac{\partial c}{\partial \eta_1} \delta \eta_1, \quad (3.28)$$

avec :

$$\delta \eta_i = \int \delta \Lambda(z) p_i(z) dz, \quad (3.29)$$

$$\delta s_i^2 = \int 2 \Lambda(z) \delta \Lambda(z) p_i(z) dz. \quad (3.30)$$

En utilisant les résultats (3.29) et (3.30), l'expression de δc peut être réécrite ainsi :

$$\delta c = \int \left[2 \left\{ \frac{\partial c}{\partial s_0^2} p_0(z) + \frac{\partial c}{\partial s_1^2} p_1(z) \right\} \Lambda(z) + \left\{ \frac{\partial c}{\partial \eta_0} p_0(z) + \frac{\partial c}{\partial \eta_1} p_1(z) \right\} \right] \delta \Lambda(z) dz. \quad (3.31)$$

L'optimalité de $\Lambda(Z)$ au sens de c implique que δc est nul, indépendamment de $\delta \Lambda(Z)$. Par conséquent, le terme $[\cdot]$ est égal à 0, ce qui permet d'établir l'expression suivante après quelques calculs :

$$\Lambda(Z) = \beta_0 \frac{p_0(Z)}{\alpha p_0(Z) + (1-\alpha) p_1(Z)} + \beta_1 \frac{p_1(Z)}{\alpha p_0(Z) + (1-\alpha) p_1(Z)}, \quad (3.32)$$

où :

$$\alpha = \frac{\partial c / \partial s_0^2}{\partial c / \partial s_0^2 + \partial c / \partial s_1^2} = \frac{\partial c / \partial \sigma_0^2}{\partial c / \partial \sigma_0^2 + \partial c / \partial \sigma_1^2}, \quad (3.33)$$

$$\beta_0 = -\alpha \frac{\partial c / \partial \eta_0}{2 \partial c / \partial s_0^2} = -\alpha \frac{\partial c / \partial \eta_0}{2 \partial c / \partial \sigma_0^2}, \quad (3.34)$$

$$\beta_1 = (\alpha - 1) \frac{\partial c / \partial \eta_1}{2 \partial c / \partial s_1^2} = (\alpha - 1) \frac{\partial c / \partial \eta_1}{2 \partial c / \partial \sigma_1^2}. \quad (3.35)$$

La combinaison des résultats (3.32) et (3.27) conduit finalement à la règle de décision suivante :

$$\frac{p_1(Z)}{p_0(Z)} \underset{d(X)=0}{\overset{d(X)=1}{>}} \lambda(\alpha, \beta_0, \beta_1), \quad (3.36)$$

qui consiste à comparer le rapport de vraisemblance à un seuil $\lambda(\alpha, \beta_0, \beta_1)$, celui-ci dépendant du critère de contraste c choisi. On notera que ce résultat ne révèle rien sur le moyen de déterminer la statistique $\Lambda(Z)$, celle-ci étant supposée connue au début de la démonstration. ■

La Proposition 3.5 justifie donc pleinement l'usage des critères de contraste ne dépendant que des moments d'ordre 1 et 2 de la statistique $\Lambda(Z)$, lorsqu'une pénurie relative d'information sur les hypothèses H_0 et H_1 nécessite d'imposer la structure du détecteur. Cependant, les détecteurs linéaires optimaux au sens de certains de ces critères peuvent parfois conduire à des résultats contestables, comme le montre la Figure 3.6. La méthode de recherche du critère optimal étudiée dans la section suivante vise précisément à combler cette lacune.

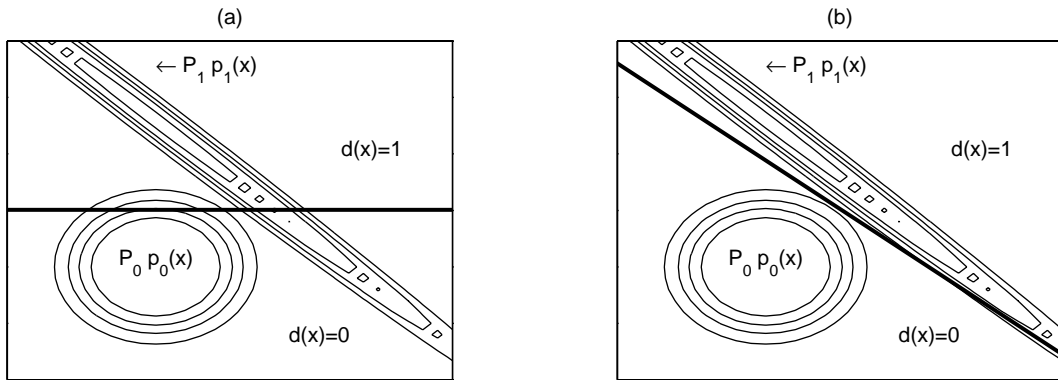


Figure 3.6 : Détection d'un signal aléatoire gaussien, de probabilité *a priori* P_0 égale à 0.7, noyé dans un bruit blanc gaussien. Partition de l'espace des réalisations par un détecteur à structure imposée maximisant (a) le rapport signal sur bruit et (b) la déflection. Sur ces représentations figurent les courbes de niveau des lois de vraisemblance conditionnelles, pondérées par les probabilités *a priori*.

3.3.4 Optimisation d'un ensemble de critères de contraste

L'étude qui vient d'être menée a permis de démontrer théoriquement l'intérêt que présente la classe \mathcal{S} des critères de séparabilité ne dépendant que des moments d'ordre 1 et 2 de la statistique de détection, pour la synthèse de détecteurs linéaires généralisés. Le choix d'un critère particulier n'en demeure pas moins un problème crucial rarement abordé dans la littérature, les solutions arbitrairement retenues étant souvent le rapport signal sur bruit ou la déflection. La méthodologie présentée dans cette section répond à cette attente en permettant de déterminer le meilleur critère de \mathcal{S} , c'est à dire celui pour lequel le détecteur à une probabilité d'erreur minimale. Cette approche, relativement méconnue, a été initialement proposée pour la recherche de discriminants linéaires en reconnaissance des formes [Fuk90]. Ce n'est que très récemment qu'elle a été adaptée au problème de la synthèse de détecteurs à structure imposée [Ric97b], [Ric97c], et qu'elle a connu de nouveaux développements. Ces derniers feront l'objet du Chapitre 4 [Ric98a], [Ric98b], [Ric99].

a) *Recherche du critère optimal.* Soit $\Lambda(Z)$ une statistique linéaire généralisée définie par :

$$\Lambda(Z) = a^T Z - v,$$

où Z est un vecteur dépendant de l'observation X , dont les éléments sont à valeur dans \mathcal{R} . Le problème de la synthèse d'un détecteur à structure imposée, tel qu'il a été défini à la section précédente, consiste à rechercher a et v optimaux au sens d'un critère de contraste c ne dépendant que des moments conditionnels suivants :

$$\begin{aligned} \eta_i &= \mathbf{E}\{\Lambda(Z) | H_i\} = \mathbf{E}\{a^T Z - v | H_i\} = a^T m_i - v, \\ \sigma_i^2 &= \mathbf{Var}\{\Lambda(Z) | H_i\} = a^T \mathbf{E}\{(Z - m_i)(Z - m_i)^T | H_i\} a = a^T \Sigma_i a, \end{aligned}$$

avec $m_i = \mathbf{E}\{Z | H_i\}$, et $\Sigma_i = \mathbf{E}\{(Z - m_i)(Z - m_i)^T | H_i\}$.

Dans ces circonstances, les dérivées de c par rapport à a et ν doivent être nulles, c'est à dire :

$$\begin{cases} \frac{\partial c}{\partial a} = \frac{\partial c}{\partial \sigma_1^2} \frac{\partial \sigma_1^2}{\partial a} + \frac{\partial c}{\partial \sigma_2^2} \frac{\partial \sigma_2^2}{\partial a} + \frac{\partial c}{\partial \eta_1} \frac{\partial \eta_1}{\partial a} + \frac{\partial c}{\partial \eta_2} \frac{\partial \eta_2}{\partial a} = 0 \\ \frac{\partial c}{\partial \nu} = \frac{\partial c}{\partial \sigma_1^2} \frac{\partial \sigma_1^2}{\partial \nu} + \frac{\partial c}{\partial \sigma_2^2} \frac{\partial \sigma_2^2}{\partial \nu} + \frac{\partial c}{\partial \eta_1} \frac{\partial \eta_1}{\partial \nu} + \frac{\partial c}{\partial \eta_2} \frac{\partial \eta_2}{\partial \nu} = 0, \end{cases} \quad (3.37)$$

où les dérivées partielles de η_i et σ_i^2 sont données par :

$$\frac{\partial \sigma_i^2}{\partial a} = 2 \Sigma_i a \quad \frac{\partial \eta_i}{\partial a} = m_i \quad \frac{\partial \sigma_i^2}{\partial \nu} = 0 \quad \frac{\partial \eta_i}{\partial \nu} = -1. \quad (3.38)$$

Ces résultats permettent de réécrire le système (3.37) ainsi :

$$\begin{cases} 2 \left[\frac{\partial c}{\partial \sigma_0^2} \Sigma_0 + \frac{\partial c}{\partial \sigma_1^2} \Sigma_1 \right] a = - \left[\frac{\partial c}{\partial \eta_0} M_0 + \frac{\partial c}{\partial \eta_1} M_1 \right] \\ \frac{\partial c}{\partial \eta_0} + \frac{\partial c}{\partial \eta_1} = 0. \end{cases} \quad (3.39)$$

En introduisant la deuxième équation du système (3.39) dans la première, et en notant que a peut être défini à un coefficient multiplicatif près, on aboutit finalement au système linéaire suivant :

$$[\alpha \Sigma_0 + (1 - \alpha) \Sigma_1] a = (m_1 - m_0), \quad (3.40)$$

avec :

$$\alpha = \frac{\partial c / \partial \sigma_0^2}{\partial c / \partial \sigma_0^2 + \partial c / \partial \sigma_1^2}. \quad (3.41)$$

Ce résultat est particulièrement intéressant puisqu'il établit que le choix du critère de contraste n'intervient dans l'expression de a que par l'intermédiaire d'un paramètre α , qu'il est aisé d'optimiser. A titre de remarque, on notera que ce dernier est compris entre 0 et 1 si on limite l'étude aux critères de contraste dont les sens de variation par rapport à σ_0^2 et σ_1^2 sont identiques ($\partial c / \partial \sigma_0^2$ et $\partial c / \partial \sigma_1^2$ de même signe), propriété par ailleurs légitimement souhaitable. Quant au terme constant ν , on peut montrer qu'il dépend explicitement de c par le biais de la deuxième équation du système (3.39). Ceci ne constitue cependant pas un obstacle pour la procédure d'optimisation à laquelle il vient d'être fait allusion, l'influence de ν étant annulée par le choix du seuil de détection. Etant donné α , ce dernier est en effet déterminé de sorte que la probabilité d'erreur $P_e(d_\alpha)$ du détecteur d_α suivant soit minimale :

$$d_\alpha(Z) = \begin{cases} 0 & \text{si } a_\alpha^T Z - \lambda_{0,\alpha} < 0 \\ 1 & \text{sinon,} \end{cases}$$

où a_α est solution du système (3.40) et $\lambda_{0,\alpha}$ le seuil à déterminer.

Dans le cadre de cette méthode, dite *du critère optimal*, la synthèse d'un détecteur repose ainsi sur l'optimisation conjointe du paramètre α et du seuil $\lambda_{0,\alpha}$, afin de minimiser la probabilité d'erreur $P_{e,\alpha}$. En conclusion, cette méthodologie conduit au détecteur linéaire généralisé optimal au sens du meilleur critère de contraste de \mathcal{S} , sans qu'il soit nécessaire d'exhiber celui-ci. L'algorithme correspondant est présenté en Figure 3.7.

En pratique, la probabilité d'erreur P_e ne peut être évaluée. On doit alors disposer d'une estimation, que peut fournir par exemple l'erreur empirique P_{emp} sur l'ensemble d'apprentissage \mathcal{A}_n . D'autres procédures sont cependant préférables, parmi lesquelles on recense les méthodes de bootstrap ou encore de validation croisée [Fuk90], [Dev96]. Dans ces circonstances, et conformément aux notations qui ont été utilisées jusqu'à présent, on pose alors :

$$d_{n,opt} = \arg \min_{0 \leq \alpha \leq 1} \hat{P}_e(d_\alpha, \mathcal{A}_n), \text{ où } \hat{P}_e \text{ désigne une estimation de la probabilité d'erreur } P_e.$$

b) *Quelques caractéristiques.* Le détecteur obtenu par la méthode du critère optimal est au moins aussi performant que ceux résultant de la maximisation du rapport signal sur bruit, de la

Déclaration des variables

\mathcal{A}_n base d'apprentissage de n éléments
 \hat{m}_i estimation de l'espérance de Z sous H_i
 $\hat{\Sigma}_i$ estimation de la matrice de variance covariance de Z sous H_i
 \hat{P}_e estimation de la probabilité d'erreur
 α variable comprise entre 0 et 1
 $\Delta\alpha$ incrément associé à la variable α

Procédure itérative de recherche de α

Initialisation
 Estimer \hat{m}_i et $\hat{\Sigma}_i$ à partir des réalisations de \mathcal{A}_n
 Initialisation de α à 0

Répéter tant que $\alpha \leq 1$,
 Résoudre $(\alpha \hat{\Sigma}_0 + (1-\alpha) \hat{\Sigma}_1) a_\alpha = (\hat{m}_1 - \hat{m}_0)$
 Déterminer le seuil $\lambda_{0,\alpha}$ pour lequel \hat{P}_e passe par un minimum noté $\hat{P}_{e,\alpha}$
 Mise à jour de α : $\alpha \leftarrow \alpha + \Delta\alpha$

Fin Répéter

Sélectionner $(a_\alpha; \lambda_{0,\alpha})$ minimisant $\hat{P}_{e,\alpha}$

Figure 3.7 : Algorithme pour la *détermination d'un détecteur linéaire généralisé* à partir d'une base d'apprentissage.

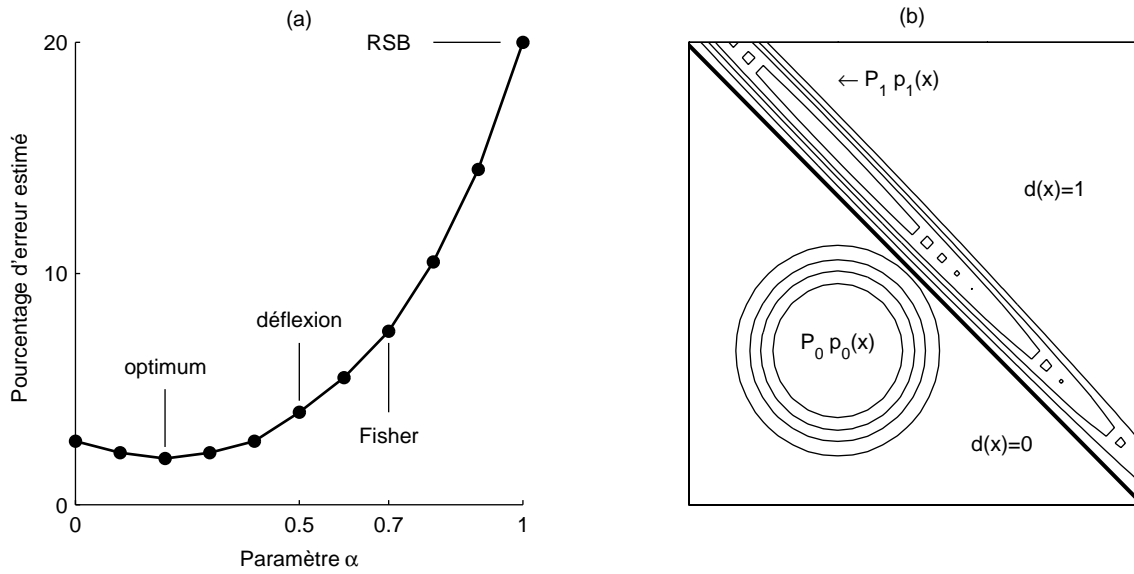


Figure 3.8 : Détection d'un signal aléatoire gaussien, de probabilité *a priori* P_0 égale à 0.7, noyé dans un bruit blanc gaussien. (a) La méthode du critère optimal, qui consiste à rechercher le paramètre α minimisant la probabilité d'erreur, conduit à un détecteur plus performant que ceux résultant de la maximisation des critères classiques. (b) La partition de l'espace des réalisations réalisée par le détecteur optimal s'avère plus satisfaisante que celles présentées en Figure 3.6.

déflexion ou encore du critère de Fisher. Ces propos sont illustrés par la Figure 3.8. En effet, c_{RSB} , $c_{Déflex}$ et c_{Fisher} étant des éléments de \mathcal{O} , il leur correspond à chacun une valeur particulière du paramètre α . Ainsi, en appliquant la relation (3.41) à la définition (3.24) du rapport signal sur bruit par exemple, on retrouve le résultat bien connu suivant :

$$a_{RSB} = \Sigma_0^{-1} (M_1 - M_0) \text{ car } \alpha_{RSB} = 1, \text{ en supposant } \Sigma_0 \text{ inversible.}$$

De façon analogue, on peut établir respectivement pour la déflexion et le critère de Fisher :

$$a_{Déflex} = 2 (\Sigma_0 + \Sigma_1)^{-1} (M_1 - M_0), \text{ car } \alpha_{Déflex} = 0.5.$$

$$a_{Fisher} = (P_0 \Sigma_0 + P_1 \Sigma_1)^{-1} (M_1 - M_0), \text{ puisque } \alpha_{Fisher} = P_0.$$

Ceci démontre donc clairement que les détecteurs obtenus par maximisation du rapport signal sur bruit, de la déflexion ou encore du critère de Fisher présentent *a priori* moins de garanties en termes de performances que celui résultant de la méthode du critère optimal.

L'objet de la prochaine section est d'illustrer la méthode qui vient d'être présentée en résolvant plusieurs problèmes de détection à structure imposée dans le plan temps-fréquence. Ce dernier constitue en effet un espace de représentation privilégié, qui nous permettra de porter un jugement qualitatif sur la configuration des structures de détection obtenues.

3.3.5 Expérimentations dans le plan temps-fréquence

La synthèse de détecteurs par maximisation d'un critère de contraste sur un ensemble d'apprentissage est une approche relativement récente dans le domaine de l'analyse temps-fréquence [Jon95b]. 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 rencontrés concerne des détecteurs linéaires opérant sur la représentation de Wigner-Ville du signal observé [Jon95b], [Say96b], [Sam96], [Ric97b], [Ric97c], [Ric98d]. En général, cette démarche est motivée par le désir de préserver toute possibilité d'aboutir à un test optimal, lorsque la solution du problème rencontré fait appel à une statistique de détection quadratique [Pic88]. Pour cette raison, on va s'intéresser dans un premier temps à l'ensemble des règles de décision définies par l'expression (3.42), afin d'illustrer la méthode du critère optimal.

$$d(X) = \begin{cases} 0 & \text{si } (a_{TF})^T W_X - \lambda_0 < 0 \\ 1 & \text{sinon,} \end{cases} \quad (3.42)$$

où a_{TF} et λ_0 doivent être déterminés à partir d'un ensemble de réalisations des hypothèses en compétition. W_X désigne la représentation de Wigner-Ville de l'observation X , écrite sous forme vectorielle.

a) *Première expérimentation.* Le premier problème considéré concerne la détection d'un signal aléatoire S noyé dans un bruit blanc, gaussien, centré et identiquement distribué sous les hypothèses H_0 et H_1 :

$$\begin{cases} H_0 : X = B \\ H_1 : X = S + B, \end{cases} \quad (3.43)$$

où $S(k) = k \exp(-0.45k) \sin(0.5k\pi + \Theta)$, $k \in \{1, \dots, 16\}$, Θ 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 -3 dB. Dans ces circonstances, il est établi en Section 3.2.3 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. On rappelle que cette structure est appelée *filtre adapté temps-fréquence*.

Afin de caractériser l'efficacité de la méthode du critère optimal pour ce type de scénario, l'apprentissage du détecteur (3.42) a été effectué sur un ensemble de 20000 réalisations de chacune des hypothèses H_0 et H_1 , préalablement transformées en signaux analytiques. Il apparaît sur la Figure 3.9 que la référence temps-fréquence \hat{a}_{TF} obtenue est semblable à la représentation de Wigner-Ville du signal S , ce qui est conforme au résultat théorique rappelé

ci-dessus. La Figure 3.10 présente quant à elle les courbes COR de cette structure de détection et du filtre adapté temps-fréquence, ce dernier fournissant une borne supérieure en termes de performances atteignables. La concordance de ces caractéristiques confirme le jugement qualitatif que l'on a porté sur la référence \hat{a}_{TF} , et illustre l'efficacité de la méthode du critère optimal pour traiter le problème de détection considéré.

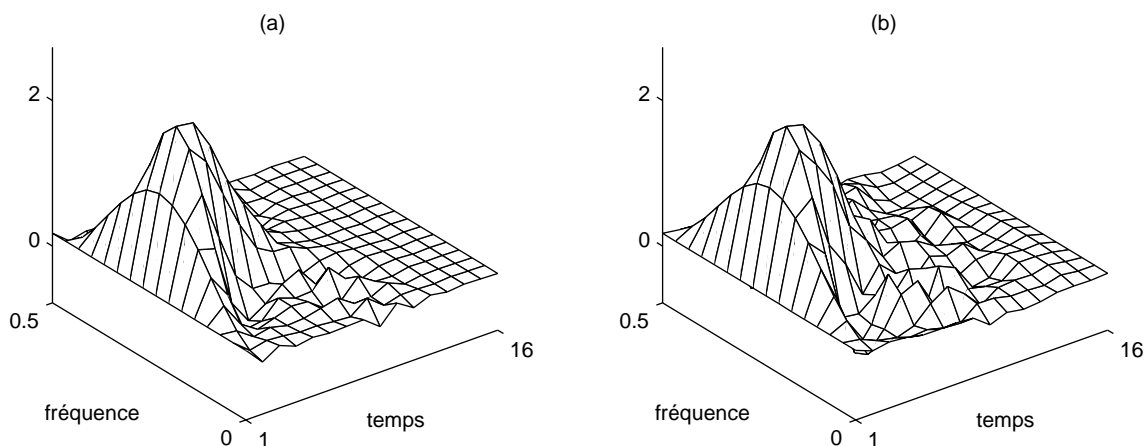


Figure 3.9 : Détection d'un signal à phase initiale aléatoire noyé dans un bruit blanc, gaussien et centré. (a) Représentation de Wigner-Ville du signal à détecter. (b) Configuration du détecteur linéaire opérant dans le plan temps-fréquence. Ce dernier a été obtenu par la méthode du critère optimal à partir de 20000 réalisations de chacune des hypothèses H_0 et H_1 .

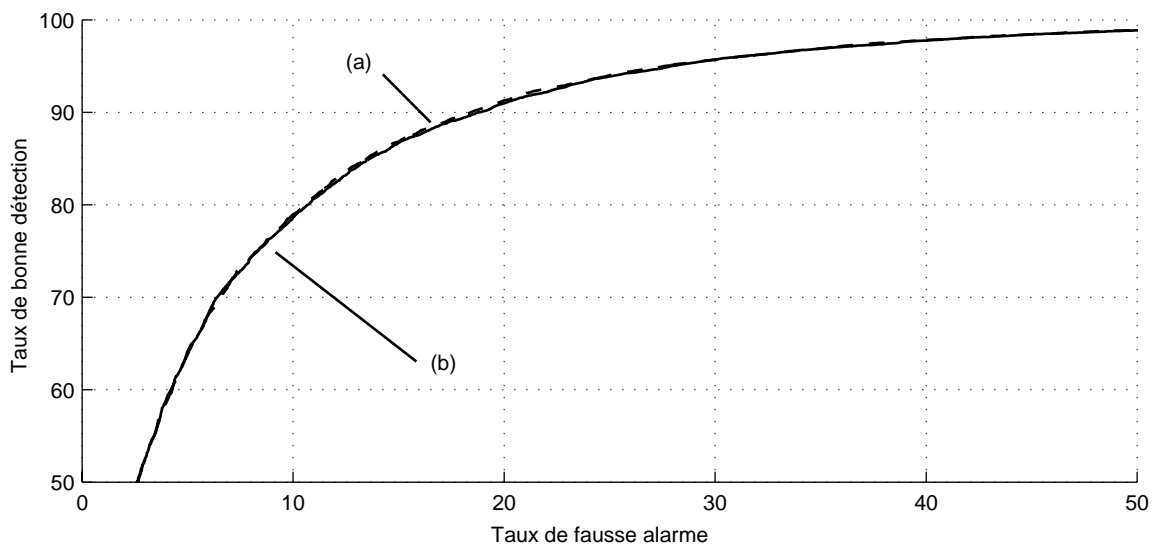


Figure 3.10 : Détection d'un signal à phase initiale aléatoire noyé dans un bruit blanc gaussien. (a) Courbe COR du filtre adapté opérant dans le plan temps-fréquence. (b) Courbe COR du détecteur linéaire opérant dans le plan temps-fréquence, obtenu par la méthode du critère optimal. Un ensemble de 4000 réalisations de chacune des hypothèses H_0 et H_1 , indépendantes des données d'apprentissage, a été utilisé pour estimer les performances de cette structure.

b) *Deuxième expérimentation.* Le second problème étudié est celui de la détection d'un signal déterministe s noyé dans un bruit blanc, gaussien et centré, dont la variance dépend de l'hypothèse H_0 ou H_1 considérée :

$$\begin{cases} H_0 : X = \gamma \cdot B \\ H_1 : X = s + B, \end{cases} \quad (3.44)$$

où γ est un réel donné et $s(k) = k \exp(-0.45k) \sin(0.5k\pi)$, $k \in \{1, \dots, 16\}$. Comme pour l'expérimentation précédente, la variance de B est fixée de sorte que le rapport signal sur bruit soit égal à -3 dB. L'écriture du rapport de vraisemblance pour le problème (3.44) conduit à l'expression suivante de la statistique de détection [Poo94] :

$$\Lambda(X) = \frac{1-\gamma^2}{2\gamma^2} X^T X + s^T X, \quad (3.45)$$

qui doit être comparée à un seuil λ_0 pour la prise de décision. Afin de faciliter l'analyse ultérieure de la statistique obtenue par la méthode du critère optimal, on choisit d'ores et déjà d'adopter une formulation temps-fréquence pour le terme quadratique de l'expression (3.45). En adoptant une écriture vectorielle pour la représentation de Wigner-Ville d'un signal X composé de L échantillons, on peut montrer que :

$$\frac{1}{L} \sum_{k=1}^L W_X[k] = \|X\|^2, \quad (3.46)$$

ce qui conduit à :

$$\Lambda(X) = \frac{1-\gamma^2}{2L\gamma^2} \sum_{k=1}^L W_X[k] + \sum_{k=1}^L s[k] X[k]. \quad (3.47)$$

La statistique de détection définie par l'expression (3.45) peut ainsi être réécrite sous la forme $\Lambda(X) = (a_{TF})^T W_X + (a_T)^T X$. Dans ce cas, les éléments de la composante temps-fréquence a_{TF} de $\Lambda(X)$ sont identiquement égaux à $(1-\gamma^2)/2L\gamma^2$, la composante temporelle a_T correspondant quant à elle au signal s à détecter.

La méthode du critère optimal a été mise en oeuvre pour l'optimisation du test de détection suivant :

$$d(X) = \begin{cases} 0 & \text{si } (\hat{a}_{TF})^T W_X + (\hat{a}_T)^T X - \lambda_0 < 0 \\ 1 & \text{sinon,} \end{cases} \quad (3.48)$$

à partir d'un ensemble d'apprentissage de 20000 réalisations de chacune des hypothèses du problème (3.44), γ étant fixé à 4. Les références temporelle \hat{a}_T et temps-fréquence \hat{a}_{TF} résultant de la procédure d'apprentissage sont proposées en Figure 3.11. On note que celles-ci présentent une bonne adéquation avec a_T et a_{TF} . La comparaison des courbes COR

représentées dans la Figure 3.12 confirme cette analyse et illustre une fois de plus l'efficacité de la méthode du critère optimal.

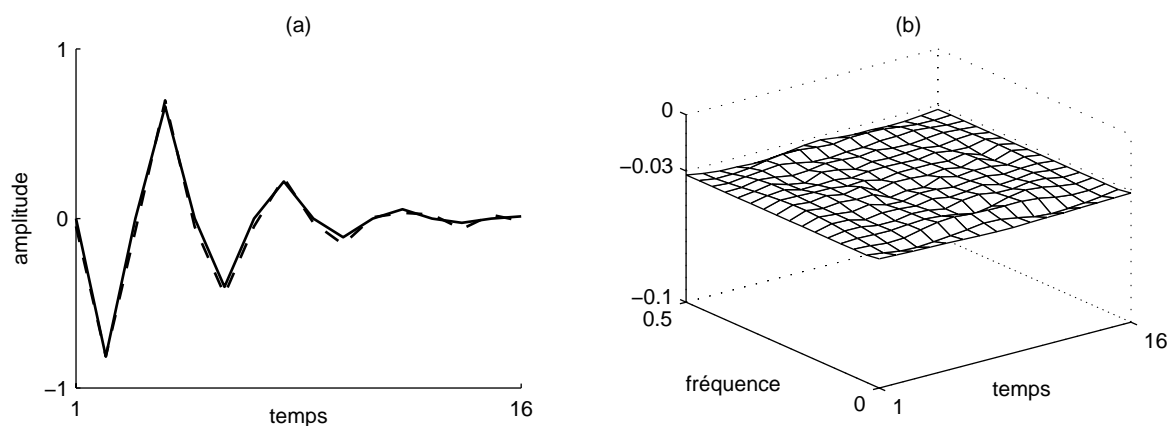


Figure 3.11 : Détection d'un signal déterministe noyé dans un bruit blanc, gaussien et centré, dont la variance dépend de l'hypothèse H_0 ou H_1 considérée. (a) Comparaison du signal à détecter (trait continu) et de la composante temporelle a_T du détecteur obtenu par la méthode du critère optimal (trait discontinu). (b) Configuration de la composante temps-fréquence a_{TF} de ce même détecteur. Ces résultats ont été obtenus à partir d'un ensemble de 20000 réalisations de chacune des hypothèses H_0 et H_1 .

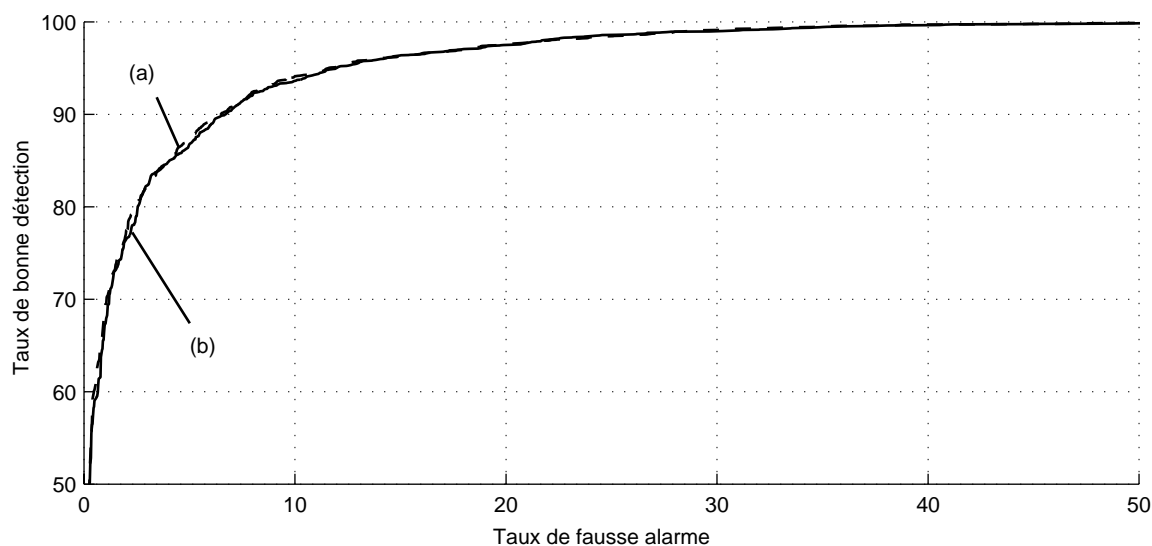


Figure 3.12 : Détection d'un signal déterministe noyé dans un bruit blanc, gaussien et centré, dont la variance dépend de l'hypothèse H_0 ou H_1 considérée. (a) Courbe COR du détecteur optimal au sens du rapport de vraisemblance. (b) Courbe COR du détecteur obtenu par la méthode du critère optimal. Un ensemble de 4000 réalisations de chacune des hypothèses H_0 et H_1 , indépendantes des données d'apprentissage, a été utilisé pour estimer les performances de cette structure.

3.3.6 Conclusion

Pour la résolution d'un problème de détection, deux approches au moins peuvent être recensées. La première concerne la détection à structure libre, pour laquelle l'expression du test de détection résulte de l'application d'un critère de décision et de la connaissance des lois de probabilité de l'observation. La deuxième, appelée détection à structure imposée, nécessite de définir préalablement une classe \mathcal{D} de tests de détection, avant de rechercher celui qui est optimal au sens d'un critère donné.

Le Chapitre 3 a été principalement consacré à la définition d'une méthodologie pour l'élaboration de détecteurs à structure imposée. En particulier, on a montré que deux arguments au moins plaident pour un choix en faveur de la classe \mathcal{D} des détecteurs polynomiaux définis par la Proposition 3.4. Le premier est fondé sur les garanties de performance que représente la consistance universellement forte de cette classe. Le second, plus pratique, repose sur la constatation que la recherche d'un détecteur au sein de cette classe est équivalente à l'optimisation d'un discriminant linéaire, thème faisant l'objet d'une abondante littérature en reconnaissance des formes. Pourtant, après avoir démontré théoriquement l'intérêt que présente, pour la synthèse d'un détecteur linéaire généralisé, la classe \mathcal{S} des critères de séparabilité ne dépendant que des moments d'ordre un et deux de la statistique de détection, certaines constatations expérimentales ont montré que le choix de l'un d'eux en particulier demeure crucial. La méthode d'apprentissage présentée dans le cadre de ce chapitre a permis de répondre à cette attente, en conduisant à la détermination du meilleur critère de \mathcal{S} pour le problème traité, c'est à dire celui pour lequel le détecteur obtenu a une probabilité d'erreur minimale. Les résultats expérimentaux proposés confirment par ailleurs les qualités de la méthodologie, lorsque la taille de l'ensemble d'apprentissage est importante par rapport au degré de la statistique de détection. L'objet du prochain chapitre est de proposer une méthodologie permettant d'ajuster automatiquement ce dernier paramètre en fonction du problème posé, afin d'obtenir les meilleures performances possibles.

Chapitre 4

Optimisation de la complexité des détecteurs à structure imposée

Applications au plan temps-fréquence

La Proposition 3.4 du chapitre précédent établit la consistance universellement forte de la suite $\{d_{n,opt}(X, \mathcal{A}_n)\}_{n>0}$ des détecteurs linéaires généralisés $d(X)$ définis par

$$d(X) = \begin{cases} 0 & \text{si } \sum_{i=1}^{K_n} a_i z^{(i)}(X) < 0 \\ 1 & \text{sinon,} \end{cases} \quad (4.1)$$

lorsqu'ils sont optimaux au sens de la probabilité d'erreur empirique sur \mathcal{A}_n , $\{z^{(i)}(X)\}$ désignant une classe particulière de polynômes. Ce résultat n'est cependant valide que si la croissance de K_n est maîtrisée par rapport à celle du nombre de données d'apprentissage n :

$$\lim_{n \rightarrow +\infty} K_n/n = 0.$$

Hormis cette première contrainte portant sur K_n , la question du choix de ce paramètre n'a pas encore été abordée dans le présent document, lorsque n est fixé. L'analogie de ce problème d'apprentissage avec celui de la modélisation laisse cependant présager que la sélection de l'ordre du modèle (4.1) repose sur la recherche d'un compromis. En effet, on constate expérimentalement que l'obtention d'une solution performante en termes de probabilité d'erreur nécessite l'adaptation de l'ordre K_n , que l'on nommera *complexité*, au nombre de données de l'ensemble d'apprentissage \mathcal{A}_n . Ainsi, comme l'indique la Figure 4.1, les récepteurs dotés d'un nombre trop important de degrés de liberté ont généralement un faible pouvoir de généralisation. Dans le cas contraire, il leur est souvent impossible d'intégrer la totalité de l'information statistique présente dans \mathcal{A}_n . Entre ces extrêmes, il existe une complexité optimale $K_{n,opt}$, dépendant en particulier de n , pour laquelle la probabilité d'erreur est minimale. L'objectif de ce chapitre est de fournir des outils au lecteur afin qu'il soit en mesure d'approcher cette situation d'optimalité.

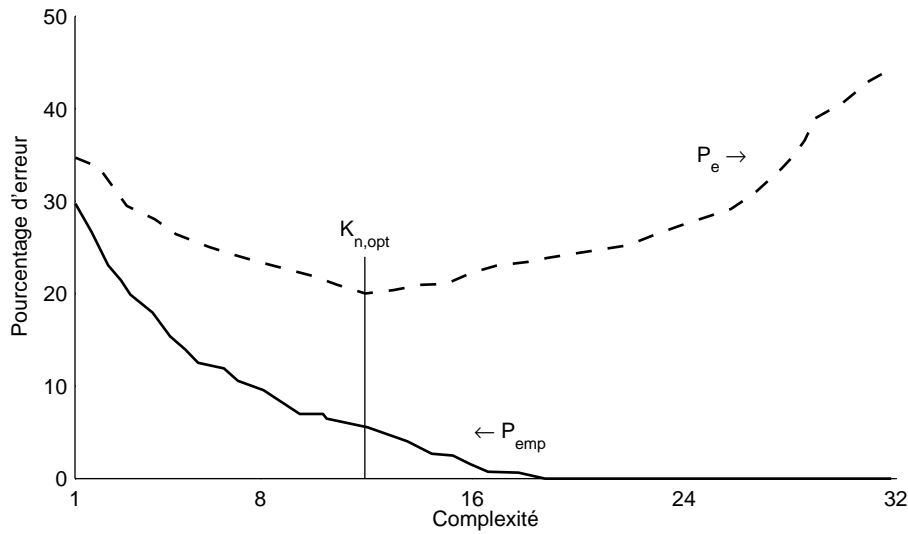


Figure 4.1 : Evolution de P_{emp} , et d'une estimation de P_e obtenue par une procédure de type leave-one-out, en fonction de la complexité d'un détecteur linéaire. Dans le cas présent, cette complexité correspond au nombre de termes de la décomposition de Karhunen-Loève des signaux EEG qui ont été utilisés pour effectuer l'apprentissage d'un détecteur linéaire.

Le Chapitre 4 est organisé ainsi. Tout d'abord, on s'interroge sur les relations qui lient la probabilité d'erreur empirique P_{emp} et la probabilité d'erreur P_e , et qui conditionnent les performances que peut atteindre un détecteur établi à partir de \mathcal{A}_n . On est ainsi amené à rechercher une condition suffisante de convergence uniforme en probabilité de P_{emp} vers P_e lorsque n tend vers l'infini. Une contribution essentielle à la résolution de ce problème, que l'on doit à Vapnik et Chervonenkis [Vap71], permet de déduire un principe très connu en théorie de l'apprentissage, dit *Structural Risk Minimization Principle*, qui peut guider efficacement l'expérimentateur dans sa quête de la complexité optimale $K_{n,opt}$. Après une brève présentation de ce concept, celui-ci est mis en œuvre par le biais de procédures d'optimisation diverses et originales [Ric98a], [Ric98b], [Ric99], que l'on illustre à chaque fois par la résolution de problèmes de détection dans le plan temps-fréquence. Ces applications nécessitent en effet l'usage de telles techniques, les détecteurs obtenus à partir d'un ensemble d'apprentissage souffrant souvent d'un biais important en raison de la grande dimension de l'espace des réalisations \mathcal{X} . Finalement, on est amené, après certaines constatations expérimentales surprenantes, à se pencher sur la dimension effective de l'espace engendré par la distribution de Wigner-Ville discrète [Ric98c].

Hormis les éléments de la théorie de l'apprentissage faisant l'objet de la section 4.1, que l'on doit à Vapnik et Chervonenkis et qui sont repris dans de nombreux ouvrages [Vap82], [Dev96], [Gas97], les résultats présentés dans ce chapitre constituent en grande majorité des contributions originales.

4.1 Recherche d'une complexité optimale

4.1.1 Convergence de la probabilité d'erreur empirique

Le détecteur $d_{n,opt}$ de la classe \mathcal{D} minimisant la probabilité d'erreur empirique sur l'ensemble d'apprentissage \mathcal{A}_n offre de meilleures garanties de performances que toute solution satisfaisant un autre critère de sélection, comme cela a été évoqué à la section 3.3.3. Dans l'expérimentation illustrée par la Figure 4.2, il est cependant manifeste que la minimisation de P_{emp} n'a de sens que lorsque l'ensemble d'apprentissage atteint une taille critique, pour une complexité donnée du détecteur. Il convient donc de s'assurer maintenant de la convergence de P_{emp} vers P_e , quelle que soit la loi $\mathbf{P}\{(X; Y)\}$. Si celle-ci est établie, la vitesse à laquelle elle s'effectue livrera certainement des informations précieuses sur la démarche à suivre pour déterminer $K_{n,opt}$.

Plusieurs types de convergence peuvent être souhaités, parmi lesquels on recense la convergence faible en probabilité :

$$\lim_{n \rightarrow +\infty} \mathbf{P}\{|P_e(d_{n,opt}) - P_{emp}(d_{n,opt}, \mathcal{A}_n)| \geq \varepsilon\} = 0. \quad (4.2)$$

Celle-ci signifie que la probabilité d'avoir un écart important entre P_{emp} et P_e décroît lorsque le nombre de données d'apprentissage augmente. Des contraintes beaucoup plus fortes peuvent être cependant envisagées. On peut ainsi souhaiter que l'écart entre P_{emp} et P_e puisse être rendu arbitrairement faible, simultanément pour tous les tests de la classe \mathcal{D} . La convergence uniforme en probabilité traduit cette propriété :

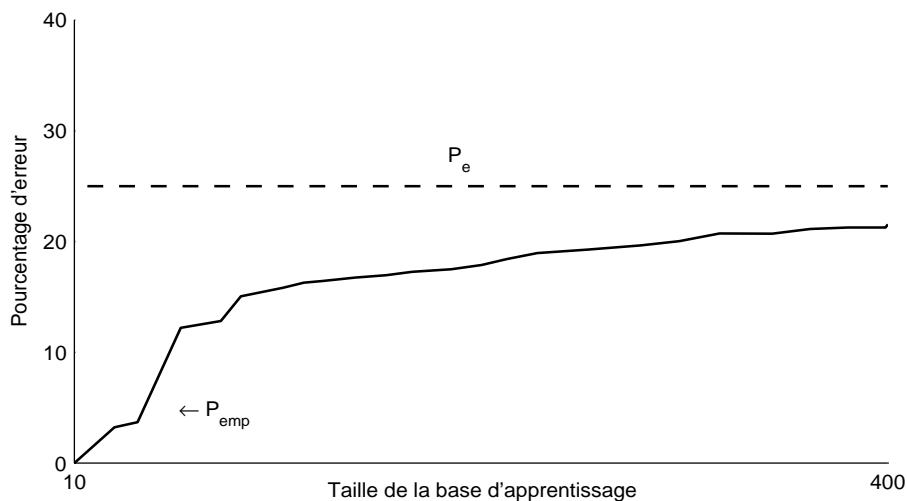


Figure 4.2 : Evolution de P_{emp} en fonction de la taille n de l'ensemble d'apprentissage, pour un détecteur linéaire de complexité fixée. La probabilité P_e est estimée au moyen d'une procédure de leave-one-out. Comme pour la Figure 4.1, les signaux utilisés sont de nature EEG.

$$\lim_{n \rightarrow +\infty} \mathbf{P} \left\{ \max_{d \in \mathcal{D}} |P_e(d) - P_{emp}(d, \mathcal{A}_n)| \geq \varepsilon \right\} = 0. \quad (4.3)$$

Une condition suffisante, qui est longtemps restée en vigueur, pour que la relation (4.3) soit satisfaite est que \mathcal{D} soit de cardinal fini [Dev96]. Il s'agit néanmoins d'une exigence très restrictive que ne vérifie pas, par exemple, la famille des structures de détection linéaires généralisées. Il a fallu attendre les travaux de Vapnik et Chervonenkis pour que ce résultat soit étendu à certaines classes de cardinalité infinie [Vap71]. Ces auteurs ont en effet démontré une inégalité fondamentale pour la théorie de l'apprentissage, dont la forme actuelle est la suivante [Gas97] :

$$\mathbf{P} \left\{ \max_{d \in \mathcal{D}} |P_e(d) - P_{emp}(d, \mathcal{A}_n)| \geq \varepsilon \right\} \leq 4 \Pi_{\mathcal{D}}(2n) e^{-n\varepsilon^2}. \quad (4.4)$$

$\Pi_{\mathcal{D}}$, qui désigne la *fonction de croissance* de la classe \mathcal{D} , est définie ainsi :

$$\Pi_{\mathcal{D}}(n) \leq \left(\frac{en}{V_{\mathcal{D}}} \right)^{V_{\mathcal{D}}} \text{ lorsque } n \geq V_{\mathcal{D}}, \quad (4.5)$$

où $V_{\mathcal{D}}$ est un paramètre à valeur entière, dit *dimension de Vapnik-Chervonenkis* ou encore *VC-dimension* de \mathcal{D} , caractérisant la *capacité d'apprentissage* des détecteurs de cette classe. L'inégalité (4.4) établit par conséquent la convergence uniforme de la probabilité d'erreur empirique lorsque la VC-dimension de \mathcal{D} est finie. Inversement, on peut montrer que cette condition est nécessaire : il existe en effet certaines lois $\mathbf{P}\{(X; Y)\}$ pour lesquelles la propriété (4.3) n'est pas satisfaite lorsque $V_{\mathcal{D}}$ est infinie [Ehr88]. L'objet de la prochaine section est de donner une définition de la dimension de Vapnik-Chervonenkis d'une classe donnée de tests de détection, et de montrer en particulier qu'elle est finie lorsqu'il s'agit des détecteurs linéaires généralisés.

4.1.2 Dimension de Vapnik-Chervonenkis

Dans le contexte de la détection à structure imposée, la dimension de Vapnik-Chervonenkis caractérise la capacité d'apprentissage des tests de détection d'une classe \mathcal{D} donnée. Comme l'indique la définition suivante, celle-ci représente en effet le nombre maximum de données d'apprentissage que les détecteurs de \mathcal{D} peuvent apprendre sans erreur, quel que soit leur étiquetage.

Définition 4.1. La dimension de Vapnik-Chervonenkis d'une classe \mathcal{D} donnée, notée $V_{\mathcal{D}}$, est définie par le plus grand nombre d'éléments de l'espace des réalisations \mathcal{X} dont les détecteurs de \mathcal{D} peuvent réaliser toutes les dichotomies.

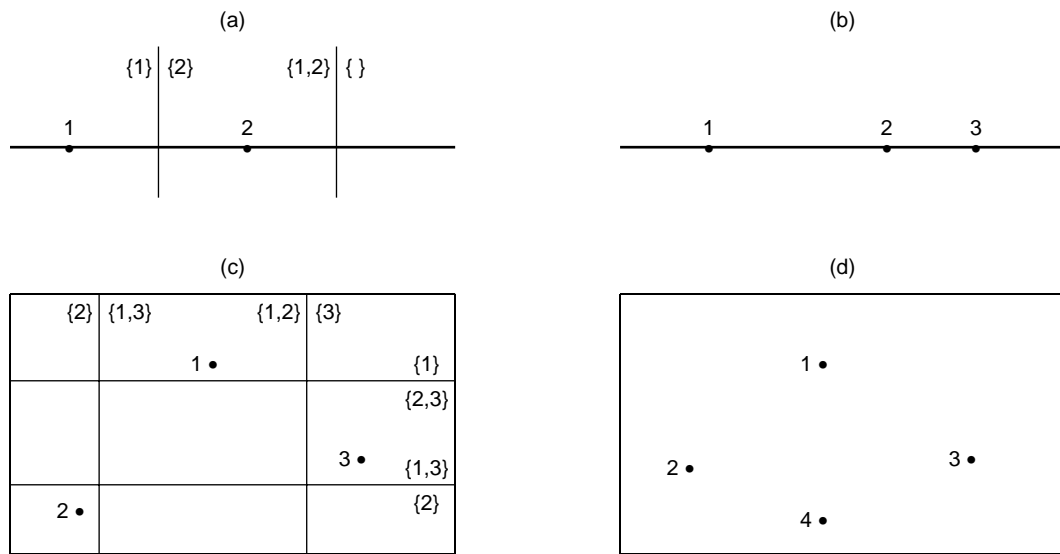


Figure 4.3 : Détermination de la dimension de Vapnik-Chervonenkis de la classe \mathcal{D} des discriminants linéaires de \mathcal{R}^L . Il est manifeste sur (a) que les discriminants de \mathcal{R} peuvent réaliser toutes les dichotomies de 2 points. Sur l'exemple (b) où figurent 3 points, il est en revanche impossible d'obtenir les regroupements suivants : $\{1,3\}$ et $\{2\}$. Dans \mathcal{R} , la VC-dimension de \mathcal{D} est donc égale à 2. En considérant les figures (c) et (d), on établit de même que la VC-dimension de la classe des discriminants linéaires de \mathcal{R}^2 est égale à 3. Sur l'exemple (d), il est en effet impossible d'obtenir les regroupements $\{1,4\}$ et $\{2,3\}$.

La Figure 4.3 permet d'établir que la VC-dimension de la classe des détecteurs linéaires de \mathcal{R} (resp. \mathcal{R}^2) est égale à 2 (resp. 3). Ce résultat peut être aisément étendu à \mathcal{R}^L , pour lequel on montre que : $V_{\mathcal{D}} = L + 1$. D'après [Gas97], l'importance d'un tel paramètre pour l'apprentissage a été initialement pressentie par Cover dans le cadre des discriminants linéaires [Cov65], bien avant d'être formalisé par Vapnik et Chervonenkis. Celui-ci a en effet observé qu'un discriminant linéaire doté d'un nombre de degrés de liberté du même ordre que la taille n de l'ensemble d'apprentissage présente généralement de piètres performances, celui-ci ne reflétant pas nécessairement la structure des données. Par le biais de l'inégalité (4.4), la dimension de Vapnik-Chervonenkis est ainsi un paramètre fondamental caractérisant l'adéquation entre la complexité d'un détecteur et la taille de l'ensemble d'apprentissage. Son évaluation reste néanmoins délicate lorsqu'on sort du cadre des détecteurs linéaires généralisés, et l'on doit alors avoir recours à des méthodes d'approximation [Bot97].

La section suivante fournit un principe basé sur l'inégalité (4.4) pour le choix de la complexité d'un détecteur, à condition de disposer d'un procédé pour contrôler celle-ci.

4.1.3 Le principe *Structural Risk Minimization* (SRM)

L'inégalité de Vapnik-Chervonenkis fournit un intervalle de confiance liant la probabilité d'erreur P_e et la probabilité d'erreur empirique P_{emp} pour tout détecteur d'une classe \mathcal{D} donnée. D'après l'expression (4.4), la relation suivante est en effet satisfaite avec une probabilité $1 - \varepsilon$:

$$P_e(d) - P_{emp}(d, \mathcal{A}_n) \leq E(n, V_{\mathcal{D}}, \varepsilon), \quad \forall d \in \mathcal{D}, \quad (4.6)$$

$$\text{avec } E(n, V_{\mathcal{D}}, \varepsilon) \equiv \sqrt{\frac{V_{\mathcal{D}}}{n} \left(1 + \ln \left(\frac{2n}{V_{\mathcal{D}}} \right) \right)} - \frac{1}{n} \ln \left(\frac{\varepsilon}{4} \right).$$

Ce résultat indique que la minimisation de l'erreur empirique s'avère satisfaisante lorsque le rapport $n/V_{\mathcal{D}}$ est significatif, E étant alors voisin de zéro. Dans le cas contraire, il convient de trouver un compromis entre les termes antagonistes P_{emp} et E . L'augmentation de la capacité d'apprentissage d'une classe de détecteurs entraîne en effet la décroissance de P_{emp} et la croissance de E . Inversement, un appauvrissement de \mathcal{D} permet de diminuer la largeur de l'intervalle de confiance E aux dépens de la probabilité d'erreur empirique P_{emp} .

Afin de trouver un compromis satisfaisant, le principe SRM préconise la construction, au sein de la classe \mathcal{D} , d'une séquence de sous-ensembles imbriqués $\mathcal{D}^{(r)}$ de VC-dimension strictement croissante [Dev96] :

$$\mathcal{D}^{(1)} \subset \mathcal{D}^{(2)} \subset \dots \subset \mathcal{D}^{(r)} \subset \dots \subset \mathcal{D}, \text{ avec } V_{\mathcal{D}}^{(1)} < V_{\mathcal{D}}^{(2)} < \dots < V_{\mathcal{D}}^{(r)} < \dots < V_{\mathcal{D}}. \quad (4.7)$$

La phase d'apprentissage est alors réalisée en deux étapes :

1. Recherche du détecteur minimisant la probabilité d'erreur empirique, dans chaque sous-ensemble $\mathcal{D}^{(r)}$:

$$d_{n,opt}^{(r)} = \arg \min_{d \in \mathcal{D}^{(r)}} P_{emp}(d, \mathcal{A}_n). \quad (4.8)$$

2. Sélection du détecteur présentant l'erreur garantie E_{gar} la plus favorable :

$$d_{n,opt} = \arg \min_{r \geq 1} E_{gar}(d_{n,opt}^{(r)}, \mathcal{A}_n, V_{\mathcal{D}}^{(r)}, \varepsilon), \quad (4.9)$$

avec : $E_{gar}(d, \mathcal{A}_n, V_{\mathcal{D}}, \varepsilon) = P_{emp}(d, \mathcal{A}_n) + E(n, V_{\mathcal{D}}, \varepsilon)$, définie d'après l'expression (4.6).

Bien que séduisant, ce principe pose néanmoins quelques problèmes pratiques lors de sa mise en oeuvre. En particulier, celui-ci nécessite la connaissance de la VC-dimension de \mathcal{D} alors que ce paramètre ne peut être évalué analytiquement que dans les cas les plus simples. Les détecteurs linéaires généralisés faisant parti de cette exception, cette première limitation s'avère donc sans conséquence dans le cadre de ce travail. En revanche, la borne supérieure définie par l'erreur garantie est généralement surestimée et le calcul d'un intervalle de

confiance plus étroite pose de réelles difficultés. Afin de pallier ceci, on préconise généralement l'usage de données indépendantes de \mathcal{A}_n afin de sélectionner $d_{n,opt}$ sur la base d'une estimation de la probabilité d'erreur, plutôt que d'adopter la relation (4.9). On parle alors de principe SRM pris au sens large.

4.2 Principe SRM et détection temps-fréquence

L'application du principe SRM suppose que l'on soit en mesure de construire une séquence (4.7) pour la classe de détecteurs \mathcal{D} choisie, thème que nous n'avons pas encore abordé. Afin de combler cette lacune, nous présentons maintenant trois stratégies permettant de faire varier la VC-dimension des détecteurs linéaires généralisés, définis par :

$$d(Z) = \begin{cases} 0 & \text{si } \Lambda(Z) \equiv a^T Z - \lambda_0 < 0 \\ 1 & \text{sinon,} \end{cases}$$

où a et Z sont des vecteurs de \mathcal{R}^L . Conformément au résultat énoncé à la section 4.1.2, on notera que dans le cas présent : $V_\varphi = L + 1$.

Les deux premières méthodes présentées, qui s'inspirent de procédés de simplification des réseaux de neurones artificiels, consistent en une sélection des composantes $a[i]$ influençant significativement le processus d'apprentissage. Leur efficacité est illustrée par la synthèse d'un détecteur linéaire opérant dans le plan temps-fréquence, bien que leur champ d'application ne se limite pas à cette application. En revanche, la dernière méthode est uniquement dédiée au contrôle de la complexité des détecteurs temps-fréquence, celle-ci reposant sur un moyennage local des représentations. Ces trois méthodes originales font l'objet des références [Ric98a], [Ric98b] et [Ric99].

Avertissement. Dans le cadre de ce document, la méthode utilisée pour l'élaboration de toute structure de détection, à partir d'un ensemble d'apprentissage, est celle du critère optimal. On rappelle qu'elle repose conjointement sur la résolution du système linéaire (3.40) et l'optimisation du paramètre α qui y intervient. Par soucis de clarté, on supposera tout au long de la Section 4.2 que ce dernier est fixé à une valeur donnée. En pratique, les procédures qui vont être présentées doivent cependant être intégrées au processus d'apprentissage du détecteur, selon le schéma de principe indiqué en Figure 4.4.

4.2.1 Méthodes de sélection

Une première alternative permettant de fixer la VC-dimension des détecteurs linéaires généralisés à une valeur r donnée est d'annuler $L - r$ composantes du vecteur a_α , solution du

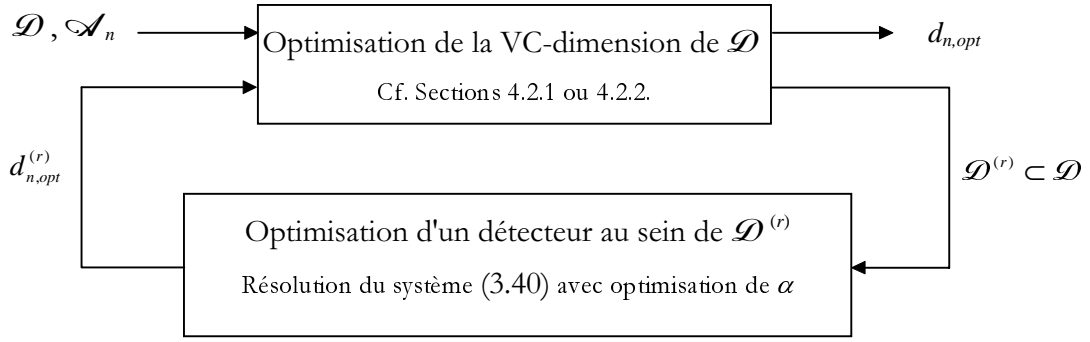


Figure 4.4. Schéma de principe du contrôle de la complexité d'un détecteur durant la phase d'apprentissage.

système (3.40). Si on exclut l'idée naïve d'éliminer les plus petites en valeur absolue, deux principes au moins peuvent être envisagés.

Le premier consiste à annuler les composantes $a_\alpha[i]$ entraînant les plus faibles variations de l'erreur quadratique EQ_α suivante, définie à partir de l'équation d'apprentissage (3.40) :

$$EQ_\alpha(a) \equiv \left\| (\alpha \Sigma_0 + (1 - \alpha) \Sigma_1) a - (m_1 - m_0) \right\|^2, \quad (4.10)$$

avec $m_i \equiv \mathbf{E}\{Z | H_i\}$, et $\Sigma_i \equiv \mathbf{E}\{(Z - m_i)(Z - m_i)^T | H_i\}$, étant donné α .

Cette méthode, dite *variationnelle*, s'inspire d'une technique d'élagage des réseaux de neurones, appelée *Optimal brain damage* [Her94].

Le second repose sur la minimisation de l'erreur quadratique EQ_α , sous la contrainte $\|a\| \leq c$, $c \in \mathcal{R}^+$. S'il n'y a pas d'annulation de composantes de la solution à proprement parler, certaines d'entre elles sont cependant condamnées de tendre vers zéro, suivant le choix de la constante c . Le principe de cette *méthode de pénalisation* s'inspire également d'un procédé issu du domaine des réseaux de neurones artificiels, appelé *Weight decay* [Vap82].

Méthode variationnelle [Ric98b], [Ric99]

a) *Principe*. Le calcul des variations de EQ_α , notées δEQ_α , en fonction de celles des composantes $a[i]$ du vecteur a , notées $\delta a[i]$, conduit à :

$$\begin{aligned} \delta EQ_\alpha(a) = & \sum_i \frac{\partial EQ_\alpha(a)}{\partial a[i]} \delta a[i] + \frac{1}{2} \sum_i \frac{\partial^2 EQ_\alpha(a)}{\partial a[i]^2} \delta a[i]^2 \\ & + \frac{1}{2} \sum_{i \neq j} \frac{\partial^2 EQ_\alpha(a)}{\partial a[i] \partial a[j]} \delta a[i] \delta a[j] + \mathcal{O}(\|a\|^2). \end{aligned} \quad (4.11)$$

Le premier terme du membre de droite de cette équation est nul lorsque $a = a_\alpha$, où a_α est solution du système (3.40), étant donné α . Le troisième est également égal à zéro lorsqu'on se place dans une base de vecteurs propres normalisés $\{\Phi_\alpha^{(i)}\}_{1 \leq i \leq L}$ de $\alpha \Sigma_0 + (1 - \alpha) \Sigma_1$. On

rappelle que cette matrice est diagonalisable puisqu'elle est symétrique. Dans cette base, l'erreur quadratique associée au vecteur de projection a_α est définie par :

$$EQ_\alpha(\tilde{a}_\alpha) = \sum_i \left(\mu_\alpha^{(i)} \tilde{a}_\alpha[i] - \tilde{m}_\alpha[i] \right)^2, \quad (4.12)$$

$$\text{où } \tilde{a}_\alpha = P_\alpha^T a_\alpha \text{ et } \tilde{m}_\alpha = P_\alpha^T (m_1 - m_0), \quad (4.13)$$

la $i^{\text{ème}}$ colonne de la matrice de passage P_α étant le vecteur propre normalisé $\Phi_\alpha^{(i)}$ associé à la $i^{\text{ème}}$ valeur propre $\mu_\alpha^{(i)}$. En combinant les expressions (4.11) et (4.12), on obtient finalement :

$$\delta EQ_\alpha(\tilde{a}_\alpha) = \sum_i \left(\mu_\alpha^{(i)} \delta \tilde{a}_\alpha[i] \right)^2. \quad (4.14)$$

Par conséquent, l'accroissement $\delta EQ_\alpha^{(i)}$ correspondant à l'annulation de la $i^{\text{ème}}$ composante de \tilde{a}_α , obtenu pour $\delta \tilde{a}_\alpha[i] = \tilde{a}_\alpha[i]$, est égal à :

$$\delta EQ_\alpha^{(i)}(\tilde{a}_\alpha) = \left(\mu_\alpha^{(i)} \tilde{a}_\alpha[i] \right)^2.$$

En constatant que $\mu_\alpha^{(i)} \tilde{a}_\alpha[i] = \tilde{m}_\alpha[i]$ lorsque $\mu_\alpha^{(i)} \neq 0$, on a finalement :

$$\delta EQ_\alpha^{(i)}(\tilde{a}_\alpha) = \tilde{m}_\alpha[i]^2, \text{ si } \mu_\alpha^{(i)} \neq 0. \quad (4.15)$$

Les composantes $\tilde{a}_\alpha[j]$ associées aux valeurs propres nulles peuvent quant à elles être systématiquement écartées de la sélection, l'accroissement $\delta EQ_\alpha^{(j)}$ associée étant nul.

En conclusion, la méthode variationnelle permet de fixer la VC-dimension des détecteurs linéaires généralisés à une valeur r donnée, en annulant les $L-r$ composantes du vecteur $\tilde{a}_\alpha \equiv P_\alpha^T a_\alpha$ les moins significatives au sens de l'erreur quadratique EQ_α .

b) Propriété fondamentale. On se propose de montrer maintenant que l'accroissement $\delta EQ_\alpha^{(i)}$ correspondant à l'annulation de la $i^{\text{ème}}$ composante de \tilde{a}_α est indépendant du paramètre α , à condition que Σ_0 et Σ_1 vérifient certaines propriétés peu restrictives. Ce résultat repose sur le théorème préliminaire suivant :

Proposition 4.1. [Gol93]. Soient A et B deux matrices symétriques, à valeurs dans \mathcal{R} et de dimension $(n \times n)$. Soit $C(\mu)$ la matrice définie ainsi :

$$C(\mu) = \mu A + (1 - \mu) B, \quad \mu \in \mathcal{R}.$$

S'il existe une valeur de μ appartenant à l'intervalle $[0, 1]$ telle que $C(\mu)$ est une matrice semi-définie positive et que $\ker(C(\mu)) = \ker(A) \cap \ker(B)$, alors il existe une matrice non-singulière P telle que $P^T A P$ et $P^T B P$ sont diagonales.

Preuve. Soit μ appartenant à l'intervalle $[0, 1]$ telle que $C(\mu)$ est une matrice semi-définie positive et que $\ker(C(\mu)) = \ker(A) \cap \ker(B)$. Soit $Q_1^T C(\mu) Q_1$ la décomposition de Schur de $C(\mu)$ définie ainsi :

$$Q_1^T C(\mu) Q_1 = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}, \text{ avec } D = \text{diag}(d_1, \dots, d_k), d_i > 0.$$

En posant $X_1 = Q_1 \text{diag}(D^{-1/2}, I_{n-k})$, $B_1 = X_1^T B X_1$ et $C_1 = X_1^T C(\mu) X_1$, on obtient alors :

$$C_1 = \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix} = \mu A_1 + (1-\mu) B_1.$$

Puisque $\ker(C(\mu)) = \ker(A) \cap \ker(B)$, les matrices A_1 et B_1 ont la structure suivante :

$$A_1 = \begin{pmatrix} A_{11} & 0 \\ 0 & 0 \end{pmatrix}, B_1 = \begin{pmatrix} B_{11} & 0 \\ 0 & 0 \end{pmatrix} \text{ où } A_{11} \text{ et } B_{11} \text{ sont des matrices de dimension } (n-k \times n-k).$$

Si on suppose que μ est différent de 0 et que $Z^T B_{11} Z = \text{diag}(b_1, \dots, b_k)$ est la décomposition de Schur de B_{11} , en posant $X = X_1 \text{diag}(Z, I_{n-k})$ on obtient finalement que :

$$X^T B X = \text{diag}(b_1, \dots, b_k, 0, \dots, 0) \equiv D_B,$$

$$X^T A X = \frac{1}{\mu} X^T (C(\mu) - (1-\mu) B) X = \frac{1}{\mu} \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix} - \frac{1-\mu}{\mu} D_B \equiv D_A.$$

En revanche, si μ est égal à 0, on considère alors la décomposition de Schur $Z^T A_{11} Z = \text{diag}(a_1, \dots, a_k)$ de la matrice A_{11} et on procède de la même façon que précédemment. ■

De cette proposition découle directement le résultat original suivant :

Proposition 4.2. Soient Σ_0 et Σ_1 les matrices de variance covariance du vecteur aléatoire Z , conditionnellement aux hypothèses H_0 et H_1 . Soit $C(\mu)$ la matrice définie ainsi :

$$C(\mu) = \mu \Sigma_0 + (1-\mu) \Sigma_1, \mu \in \mathcal{R}.$$

S'il existe un réel μ de l'intervalle $[0, 1]$ pour lequel $\ker(C(\mu)) = \ker(\Sigma_0) \cap \ker(\Sigma_1)$, l'accroissement $\delta EQ_\alpha^{(i)}$ de l'erreur quadratique EQ_α correspondant à l'annulation de la $i^{\text{ème}}$ composante de \tilde{a}_α est indépendant du paramètre α .

Preuve. D'après la proposition 4.1, il existe une matrice de passage P telle que $P^T \Sigma_0 P$ et $P^T \Sigma_1 P$ soient toutes deux diagonales. Celle-ci étant indépendante de α , l'accroissement $\delta EQ^{(i)}$ correspondant à l'annulation de la $i^{\text{ème}}$ composante de $\tilde{a}_\alpha = P^T a_\alpha$ est également indépendant de α . D'après la relation (4.15), $\delta EQ_\alpha^{(i)}$ est en effet défini ainsi :

$$\delta EQ^{(i)}(\tilde{a}_\alpha) = \tilde{m}[i]^2, \text{ si } \mu_\alpha^{(i)} \neq 0, \text{ avec } \tilde{m} = P^T (m_1 - m_0). \quad \blacksquare$$

On peut noter que la condition $\ker(C(\mu)) = \ker(\Sigma_0) \cap \ker(\Sigma_1)$ est peu restrictive en pratique, l'usage d'un détecteur à structure imposée se justifiant surtout lorsque les sous-espaces $\ker(\Sigma_0)$ et $\ker(\Sigma_1)$ sont égaux. Dans le cas contraire, on préférera en effet adopter

une approche plus classique de type filtrage adapté à un sous-espace [Sch91], en l'occurrence ici à $\ker(\Sigma_0) - [\ker(\Sigma_0) \cap \ker(\Sigma_1)]$ ou $\ker(\Sigma_1) - [\ker(\Sigma_0) \cap \ker(\Sigma_1)]$.

En conclusion, les différents sous-espaces sur lesquels le vecteur a_α est projeté, de sorte que la complexité du détecteur linéaire généralisé associé puisse être contrôlée par la méthode variationnelle, sont indépendants du paramètre α . L'association de cette approche à la méthode du critère optimal constitue donc un gage de robustesse pour la solution obtenue.

Méthode de pénalisation [Ric98b], [Ric99]

La méthode de pénalisation repose sur la construction, au sein de la classe \mathcal{D} des détecteurs linéaires généralisés, d'une séquence de sous-ensembles imbriqués $\mathcal{D}^{(r)}$:

$$\mathcal{D}^{(1)} \subset \mathcal{D}^{(2)} \subset \dots \subset \mathcal{D}^{(r)} \subset \dots \subset \mathcal{D}, \quad (4.16)$$

$$\text{où } \mathcal{D}^{(r)} = \left\{ d : a^T Z - \lambda_0 \begin{array}{l} \underset{d(Z)=0}{>} \\ \underset{d(Z)=1}{<} \end{array} 0, \|a\|^2 \leq c^{(r)}, c^{(r)} \in \mathcal{R} \right\}, \quad (4.17)$$

avec $0 < c^{(1)} < c^{(2)} < \dots < c^{(r)} < \dots$.

Bien que la procédure d'optimisation en deux étapes requise par cette approche soit analogue à celle du principe SRM, la définition de $\mathcal{D}^{(r)}$ est toutefois quelque peu différente : elle repose désormais sur un classement des détecteurs selon le module de leur vecteur de projection a , et non selon la VC-dimension de la classe qui les regroupe. Pourtant, comme cela va être montré, la finalité visant à faire varier la capacité d'apprentissage des structures de détection, afin de déterminer celle qui semble la mieux adaptée à une base d'apprentissage donnée, demeure inchangée.

Afin de simplifier l'exposé qui suit, on se place à présent dans une base de vecteurs propres normalisés $\{\Phi_\alpha^{(i)}\}_{1 \leq i \leq L}$ de $\alpha \Sigma_0 + (1 - \alpha) \Sigma_1$. Les notations adoptées sont identiques à celles employées pour décrire la méthode variationnelle. Soit $\tilde{a}_\alpha^{(r)}$ le vecteur de projection minimisant l'erreur quadratique EQ_α définie par l'équation (4.10), sous la contrainte $\|\tilde{a}_\alpha^{(r)}\| \leq \sqrt{c^{(r)}}$ de sorte que le détecteur linéaire généralisé associé appartienne à $\mathcal{D}^{(r)}$. D'après le théorème de Kuhn-Tucker, il existe un multiplicateur de Lagrange $\rho^{(r)}$, associé au paramètre $c^{(r)}$, tel que :

$$\tilde{a}_\alpha^{(r)} \equiv \arg \min_{\tilde{a} \in \mathcal{R}^L} \left\{ EQ_\alpha(\tilde{a}) + \rho^{(r)} \|\tilde{a}\|^2 \right\}.$$

En annulant les dérivées partielles de la fonction $EQ_\alpha(\tilde{a}) + \rho^{(r)} \|\tilde{a}\|^2$ par rapport à $\tilde{a}[i]$, on peut établir la relation (4.18).

$$\tilde{a}_\alpha^{(r)}[i] \equiv \frac{\mu_\alpha^{(i)}}{(\mu_\alpha^{(i)})^2 + \rho^{(r)}} \tilde{m}_\alpha[i] = \frac{(\mu_\alpha^{(i)})^2}{(\mu_\alpha^{(i)})^2 + \rho^{(r)}} \tilde{a}_\alpha[i], \quad (4.18)$$

où $\tilde{a}_\alpha = P_\alpha^T a_\alpha$, a_α étant solution du système linéaire (3.40).

Ce résultat montre ainsi que les composantes $\tilde{a}_\alpha[i]$ ne sont pas affectées par la méthode de pénalisation lorsque $\mu_\alpha^{(i)} \gg \sqrt{\rho^{(r)}}$, mais tendent en revanche vers 0 lorsque $\mu_\alpha^{(i)} \ll \sqrt{\rho^{(r)}}$. Par conséquent, l'introduction d'une contrainte dans le processus d'apprentissage entraîne une modification de la capacité des détecteurs, que l'on peut alors estimer par l'expression (4.19) en adaptant à notre problème le résultat donné dans [Moo92] :

$$V_\rho^{(r)} = \mathbf{Int} \left\{ \sum_i \frac{(\mu_\alpha^{(i)})^2}{(\mu_\alpha^{(i)})^2 + \rho^{(r)}} \right\}, \text{ où } \mathbf{Int}\{\} \text{ désigne l'opérateur } \textit{partie entière}. \quad (4.19)$$

Ce résultat n'est cependant valide que si la matrice $\alpha \Sigma_0 + (1-\alpha) \Sigma_1$ admet un spectre de valeurs propres étroit, conformément à [Moo92].

En agissant sur la dynamique de certaines composantes du vecteur de projection a durant l'apprentissage, les méthodes de sélection qui viennent d'être présentées procèdent de façon analogue pour contrôler la capacité d'apprentissage des détecteurs linéaires généralisés. Bien que très efficace, comme le montre l'exemple traité dans la section suivante, la méthode de pénalisation est néanmoins plus délicate à mettre en oeuvre que l'approche variationnelle, puisqu'elle nécessite l'optimisation conjointe des paramètres α et $\rho^{(r)}$.

Expérimentations dans le plan temps-fréquence

L'objet de cette section est d'illustrer l'efficacité des méthodes variationnelle et de pénalisation, en optimisant le détecteur (4.20) à partir d'un ensemble d'apprentissage de taille modeste par rapport au nombre de paramètres à déterminer.

$$d(X) = \begin{cases} 0 & \text{si } (a_{TF})^T W_X - \lambda_0 < 0 \\ 1 & \text{sinon,} \end{cases} \quad (4.20)$$

où W_X désigne la représentation de Wigner-Ville du signal X , écrite sous forme vectorielle. Le problème considéré dans le cadre de cette expérimentation est identique à celui figurant en Section 3.3.5.a. Il concerne la détection d'un signal aléatoire S noyé dans un bruit blanc, gaussien, centré et identiquement distribué sous les hypothèses H_0 et H_1 :

$$\begin{cases} H_0 : X = B \\ H_1 : X = S + B, \end{cases} \quad (4.21)$$

où $S(k) = k \exp(-0.45k) \sin(0.5k\pi + \Theta)$, $k \in \{1, \dots, 16\}$, Θ 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 -3 dB. Dans ces circonstances, on rappelle que le test de détection optimal est obtenue pour $a_{TF} = W_S$, où W_S désigne la représentation de Wigner-Ville du signal à détecter.

Afin d'illustrer les effets néfastes d'un ensemble d'apprentissage de faible cardinalité sur les performances d'une structure de détection, le détecteur (4.20) 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 , préalablement transformées en signaux analytiques. Pour cela, chacun des moments statistiques conditionnels $M_{i=0,1}$ et $\Sigma_{i=0,1}$ du vecteur aléatoire W_X a été estimé au moyen de 200 échantillons. Les autres données ont quant à elles été destinées à l'évaluation des performances du détecteur au cours du processus d'apprentissage, afin de guider le choix du paramètre α . Il apparaît sur la Figure 4.5.(a) que la composante \hat{a}_{TF} obtenue ne reflète pas le contenu temps-fréquence du signal S à détecter. Cette constatation est par ailleurs confirmée par les piètres performances que présente le détecteur, comme l'indique la Figure 4.5.(b). A titre de comparaison, on pourra consulter la Section 3.3.5.a, où le cas d'un ensemble d'apprentissage de 40000 données est utilisé pour résoudre le problème de détection (4.21).

De façon à maîtriser l'influence de la taille de l'ensemble d'apprentissage sur les performances des détecteurs, les procédures variationnelle et de pénalisation ont été successivement associées à la méthode du critère optimal. Dans le cadre de la méthode variationnelle, l'optimisation de la probabilité d'erreur, estimée comme dans le cas précédent à partir de 100 réalisations de chacune des hypothèses, conduit à un minimum pour $\alpha = 0.4$ et une VC-dimension réduite à 5. Le résultat présenté en Figure 4.6 laisse apparaître que la référence temps-fréquence \hat{a}_{TF} obtenue est similaire à la représentation de Wigner-Ville du signal à détecter. En ce qui concerne la méthode de pénalisation, la probabilité d'erreur atteint un minimum pour $\alpha = 0.32$ et une VC-dimension approximativement égale à 8, si on adopte l'estimateur (4.19). Le résultat présenté en Figure 4.7 montre également que dans ce cas, la référence \hat{a}_{TF} obtenue reflète le contenu temps-fréquence du signal S .

Finalement, les performances de ces détecteurs ont été estimées sur un ensemble de 4000 réalisations de chacune des hypothèses. Les résultats sont indiqués en Figure 4.8. Pour une probabilité de fausse alarme fixée à 0.1, les méthodes variationnelle et de pénalisation conduisent respectivement à des probabilités de bonne détection égales à 0.75 et 0.77. A titre de comparaison, ces résultats sont voisins de la probabilité de bonne détection du filtre adapté opérant dans le plan temps-fréquence qui est égale à 0.78, et qui constitue la performance de référence. Ces performances sont en revanche très supérieures à la probabilité de 0.65 obtenue lorsque la complexité de la structure de détection ne fait l'objet d'aucun contrôle. Ceci illustre l'efficacité des procédures proposées.

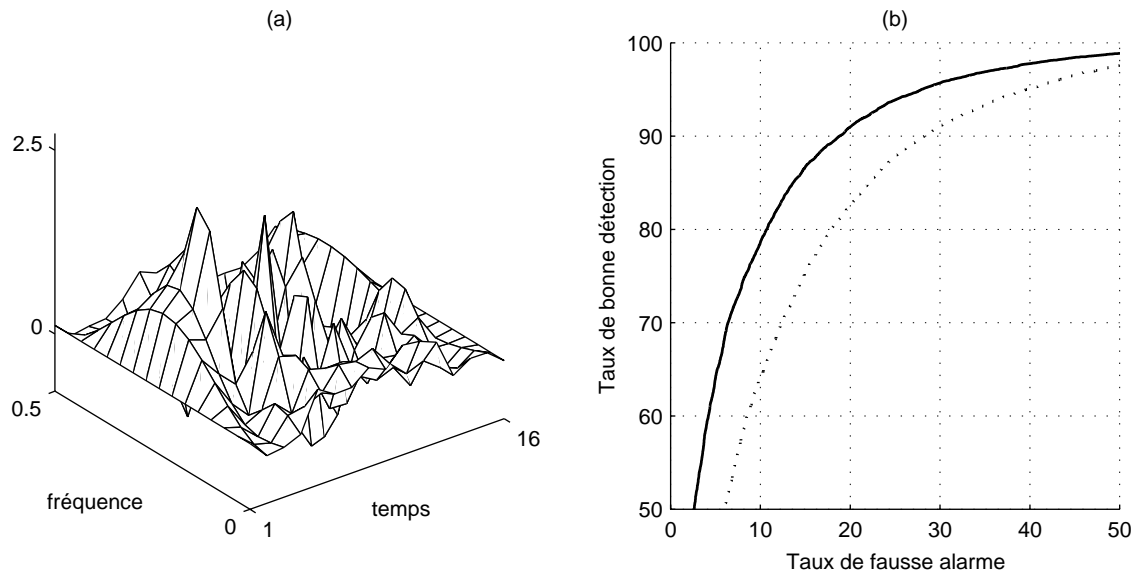


Figure 4.5 : Détection d'un signal à phase initiale aléatoire noyé dans un bruit blanc, gaussien et centré. (a) Configuration du détecteur linéaire opérant dans le plan temps-fréquence obtenu par la méthode du critère optimal. L'ensemble d'apprentissage est composé de 300 réalisations des hypothèses H_0 et H_1 . (b) Courbes COR du filtre adapté opérant dans le plan temps-fréquence (trait plein) et du détecteur obtenu par la méthode du critère optimal (trait discontinu). Un ensemble de 4000 données a été utilisé pour estimer ces performances.

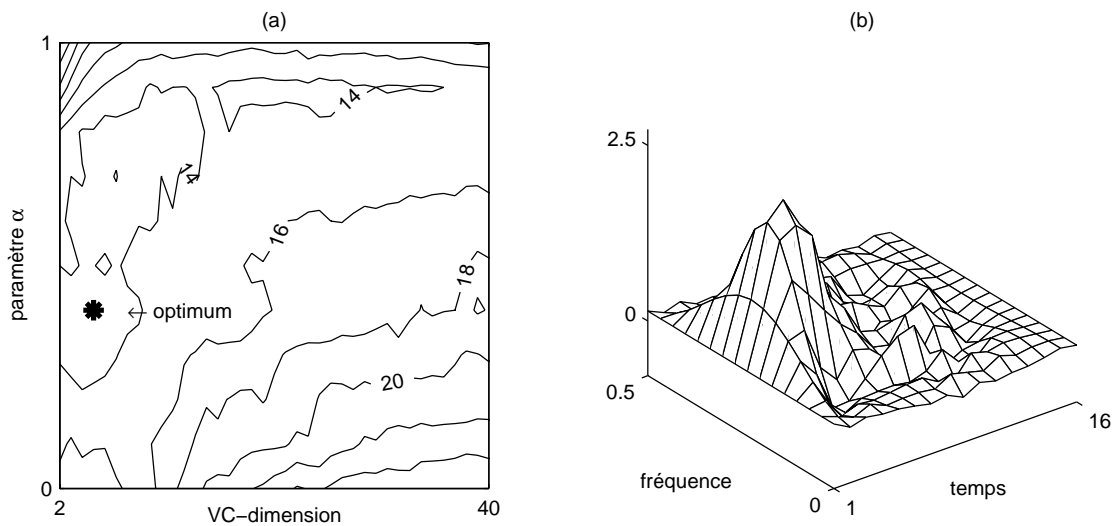


Figure 4.6 : Détection d'un signal à phase initiale aléatoire noyé dans un bruit blanc, gaussien et centré. (a) Optimisation de la structure de détection par association des méthodes du critère optimal et variationnelle. Les courbes de niveaux sont celles d'une estimation de la probabilité d'erreur. La probabilité d'erreur minimale est obtenue pour une VC-dimension égale 5. (b) Configuration du détecteur linéaire opérant dans le plan temps-fréquence. L'ensemble d'apprentissage est composé de 300 réalisations de chacune des hypothèses H_0 et H_1 .

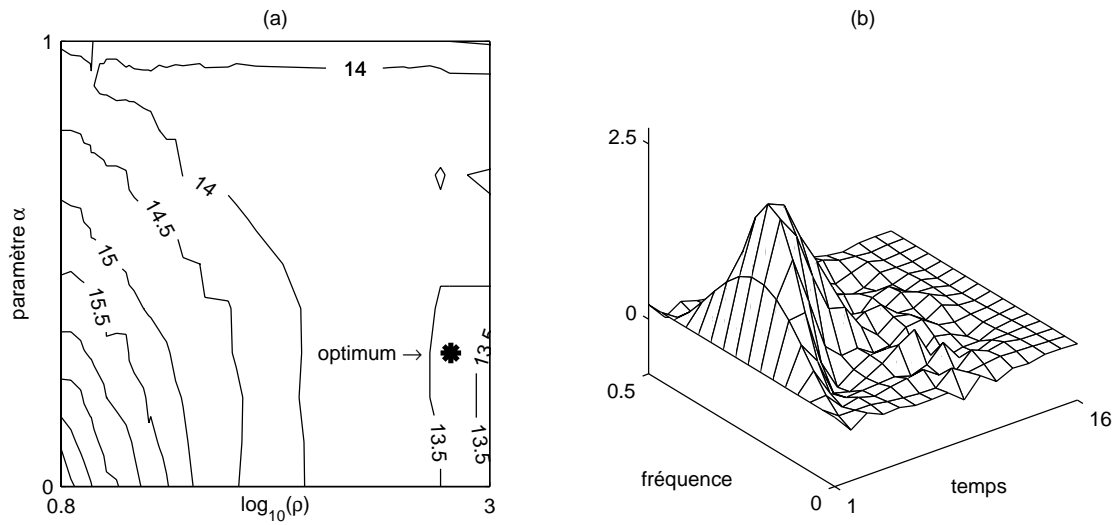


Figure 4.7 : Détection d'un signal à phase initiale aléatoire noyé dans un bruit blanc, gaussien et centré. (a) Optimisation de la structure de détection par association des méthodes du critère optimal et de pénalisation. Les courbes de niveaux sont celles d'une estimation de la probabilité d'erreur. (b) Configuration du détecteur linéaire temps-fréquence. L'ensemble d'apprentissage est composé de 300 réalisations de chacune des hypothèses H_0 et H_1 .

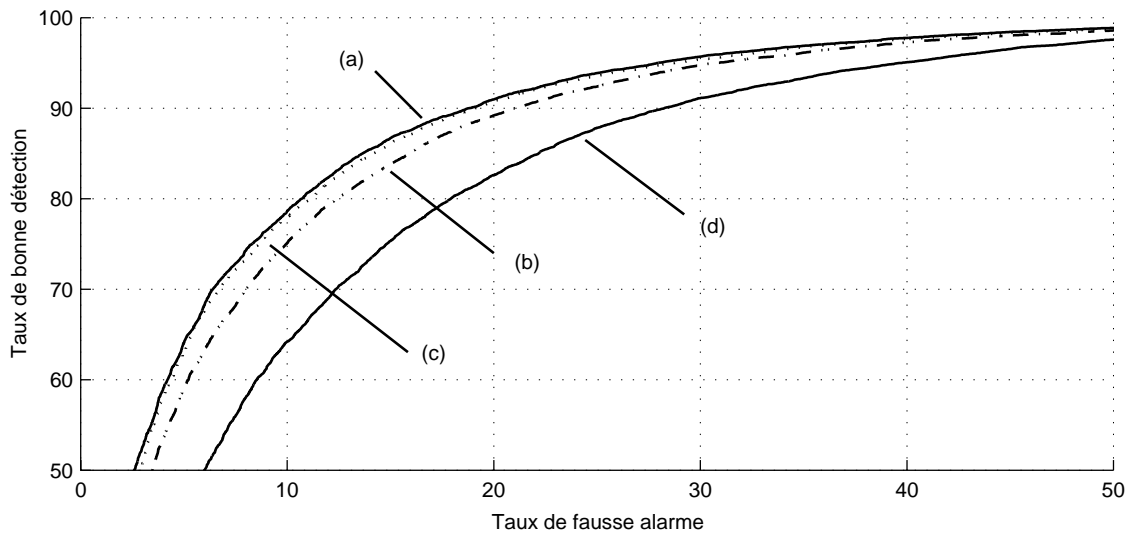


Figure 4.8 : Détection d'un signal à phase initiale aléatoire noyé dans un bruit blanc, gaussien et centré. (a) Courbe COR du filtre adapté opérant dans le plan temps-fréquence. Courbes COR des détecteurs linéaires opérant dans le plan temps-fréquence, obtenus par association de la méthode du critère optimal et des méthodes (b) variationnelle ou (c) de pénalisation, lorsqu'on dispose d'un ensemble d'apprentissage constitué de 300 réalisations de chacune des hypothèses H_0 et H_1 . (d) Courbe COR du détecteur obtenu dans les mêmes conditions que (b) et (c), lorsqu'il n'est fait appel à aucune procédure de contrôle de la complexité.

Classe de Cohen et détection à structure imposée

Au cours d'un certain nombre d'expérimentations que l'on a pu effectuer, on a été amené à constater qu'un filtrage préalable de la représentation de Wigner-Ville des signaux étudiés peut conduire à une amélioration substantielle des performances d'un détecteur linéaire opérant dans le plan temps-fréquence. Dans le cadre de cette section, on s'attache à expliquer ce phénomène en considérant dans un premier temps, les détecteurs linéaires obtenus par résolution du système (3.40). Une extrapolation à un cadre plus général est brièvement envisagée dans un second temps.

Conformément aux éléments présentés en Section 1.3.2, on rappelle que le filtrage de la représentation de Wigner-Ville W_X se traduit dans le domaine Doppler-retard par une simple pondération de la fonction d'ambiguïté à bande étroite A_X . C'est précisément pour bénéficier de cette propriété que le raisonnement qui suit suppose la mise en oeuvre des méthodes du critère optimal, variationnelle et de pénalisation dans cet espace de représentation. La nature du problème de détection demeure cependant inchangée en raison de la bijectivité de l'application linéaire liant W_X et A_X , dont on rappelle qu'il s'agit de la transformée de Fourier bidimensionnelle. En particulier, si on note a_{DR} la transformée de Fourier bidimensionnelle de la référence temps-fréquence a_{TF} , le test de détection (4.20) se reformule ainsi, dans le domaine Doppler-retard :

$$d(X) = \begin{cases} 0 & \text{si } \Lambda(X) = (a_{TF})^T W_X - \lambda_0 = (a_{DR})^T A_X - \lambda_0 < 0 \\ 1 & \text{sinon,} \end{cases} \quad (4.22)$$

en adoptant une notation vectorielle pour W_X , A_X , a_{TF} et a_{DR} . Dans un premier temps, on suppose que la référence a_{DR} résulte de l'équation (3.40), de sorte que les enseignements apportés par l'étude des méthodes variationnelle et de pénalisation puissent être exploités dans la suite de cette section.

Les travaux de Vapnik et Chervonenkis sur la théorie de l'apprentissage ont mis en évidence qu'un contrôle adapté de la complexité d'une structure de décision peut conduire à une amélioration de ses performances. En particulier, les méthodes variationnelle et de pénalisation présentées dans ce document remplissent parfaitement ce rôle, comme le confirment les expérimentations effectuées dans le plan temps-fréquence. Au prix d'un changement de base défini par une matrice de passage orthogonale notée P , on a ainsi montré que leur mode d'action consiste en une pondération $\phi_{DR}[i] \tilde{a}_{DR}[i]$ des composantes de la référence $\tilde{a}_{DR} = P^T a_{DR}$. Plus précisément, la référence Doppler-retard \tilde{a}_{DR}^ϕ résultant de chacune de ces deux procédures s'écrit :

$$\tilde{a}_{DR}^\phi = D_\phi \tilde{a}_{DR}, \text{ avec } D_\phi = \text{diag}(\phi_{DR}[1], \dots, \phi_{DR}[L^2]), \quad (4.23)$$

où L désigne le nombre d'échantillons de l'observation X . Dans la définition de la matrice diagonale D_ϕ , les composantes $\phi_{DR}[i]$ sont soit booléennes, soit données par l'expression (4.18), selon que l'on fasse usage de la méthode variationnelle ou de pénalisation. La statistique de détection linéaire $\Lambda^\phi(X)$ mettant en oeuvre la référence Doppler-retard \tilde{a}_{DR}^ϕ est alors donnée par l'expression suivante :

$$\Lambda^\phi(X) \equiv (a_{DR}^\phi)^T A_X - \lambda_0 = (P D_\phi P^T a_{DR}^\phi)^T A_X - \lambda_0 = a_{DR}^{T} (\Phi_{DR} A_X) - \lambda_0, \quad (4.24)$$

avec $\Phi_{DR} = P D_\phi P^T$.

Ainsi, $\Lambda^\phi(X)$ consiste, à une constante additive près, en un produit scalaire entre la référence a_{DR} et l'image de A_X par l'application linéaire associée à la matrice Φ_{DR} . Cette dernière n'étant pas diagonale dans le cas général, il est malheureusement impossible de qualifier l'opération $\Phi_{DR} A_X$ de pondération de la fonction d'ambiguïté, et par conséquent de l'associer à une représentation temps-fréquence de la classe de Cohen. L'expression (4.24) suggère cependant qu'une pondération appropriée de A_X , avant l'exécution de la procédure d'apprentissage, offre certainement une alternative intéressante aux méthodes de sélection, pour le contrôle de la capacité du détecteur. Pour cela, il est suffisant que la matrice diagonale Φ_{DR} conduise à l'annulation de certaines composantes de A_X , ou tout moins à leur atténuation, à l'image des modes opératoires respectifs des méthodes variationnelle et de pénalisation. En conséquence, ces premiers éléments laissent prévoir l'existence de détecteurs opérant sur des représentations de la classe de Cohen, telles que le spectrogramme par exemple, qui s'avèrent plus performants que celui agissant sur la représentation de Wigner-Ville.

Afin de vérifier cette conjecture, le cadre expérimental proposé pour illustrer l'efficacité des méthodes dites de sélection a été conservé. Le détecteur suivant, opérant sur un spectrogramme, a alors été optimisé sur la base de 300 réalisations de chacune des hypothèses du problème (4.21), grâce à la méthode du critère optimal :

$$d(X) = \begin{cases} 0 & \text{si } (a_{TF})^T S_X^h - \lambda_0 < 0 \\ 1 & \text{sinon,} \end{cases} \quad (4.25)$$

où S_X^h désigne le spectrogramme de l'observation X , pondéré par une fenêtre de Hanning de largeur $L/4$, notée h . On rappelle que L représente le nombre d'échantillons de X .

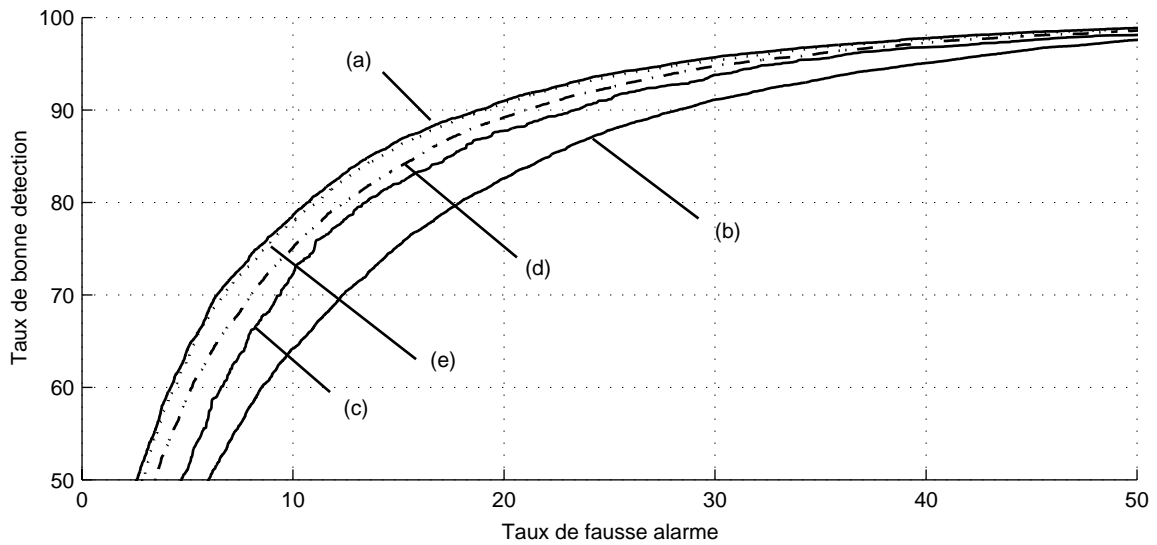


Figure 4.9 : Détection d'un signal à phase initiale aléatoire noyé dans un bruit blanc, gaussien et centré. (a) Courbe COR du filtre adapté opérant dans le plan temps-fréquence. Courbes COR des détecteurs linéaires opérant sur (b) la représentation de Wigner-Ville et (c) le spectrogramme de l'observation X , sans contrôle de la complexité lors de la mise en oeuvre de la méthode du critère optimal. Courbes COR des détecteurs linéaires opérant sur la représentation de Wigner-Ville, obtenus par association de la méthode du critère optimal et des méthodes (d) variationnelle ou (e) de pénalisation. L'ensemble d'apprentissage est constitué de 300 réalisations de chacune des hypothèses H_0 et H_1 . Les performances sont estimées sur un ensemble comprenant 8000 données.

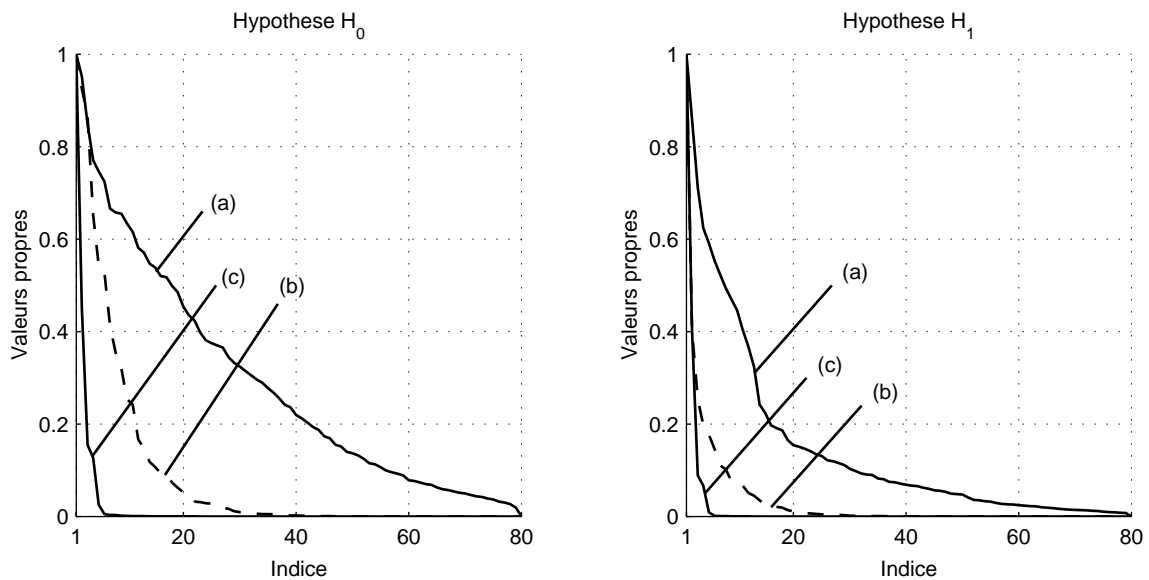


Figure 4.10. Valeurs propres normalisées des matrices de variance covariance conditionnelles des composantes de (a) la représentation de Wigner-Ville W_X , (b) du spectrogramme S_X^h et (c) de $\Phi_{TF} W_X$. Dans ce dernier cas, Φ_{TF} provient de l'application de la méthode de pénalisation, comme l'indique l'expression (4.23). Afin de ne pas nuire à la clarté de la figure, les valeurs propres obtenues après la mise en oeuvre de la procédure variationnelle n'ont pas été indiquées. Les résultats sont cependant semblables aux cas (b) et (c).

Les performances de ce détecteur, indiquées en Figure 4.9, ont été estimées sur un ensemble comprenant 8000 données. Pour une probabilité de fausse alarme fixée à 0.1, il apparaît ainsi que les structures de détection opérant sur la représentation de Wigner-Ville et le spectrogramme S_X^h ont respectivement une probabilité de bonne détection de 0.65 et 0.72. Cette issue, à l'avantage du spectrogramme, va à l'encontre de nombreuses idées reçues en matière de détection dans le plan temps-fréquence, l'opération $S_X^h = W_X * W_h$ pouvant s'accompagner d'une perte d'information. Le résultat présenté est toutefois justifié par la théorie de Vapnik et Chervonenkis, le filtrage de W_X à l'aide W_h se traduisant par une réduction de la dimension de l'espace des observations, accompagnée d'une atténuation de certaines composantes, comme l'indique la Figure 4.10.

En conclusion, si ces travaux sont encore à un stade préliminaire, ils laissent cependant entrevoir la possibilité d'appliquer le principe SRM sans avoir recours à des procédures externes telles que les méthodes variationnelle, de pénalisation ou encore de partitionnement, qui fait l'objet de la prochaine section. C'est là une illustration supplémentaire et originale des potentialités de la classe de Cohen pour résoudre certains problèmes de décision.

4.2.2 Méthode de partitionnement

Contrairement aux procédures variationnelle et de pénalisation, la méthode de partitionnement est uniquement dédiée au contrôle de la VC-dimension des détecteurs linéaires opérant dans le plan temps-fréquence :

$$d(X) = \begin{cases} 0 & \text{si } \Lambda(X) = \sum_{n,m} a_{TF}[n,m] W_X[n,m] - \lambda_0 < 0 \\ 1 & \text{sinon,} \end{cases} \quad (4.26)$$

où la matrice W_X désigne la représentation de Wigner-Ville discrète du signal X . En effet, cette approche consiste à contraindre la référence a_{TF} d'être constante sur chacune des r régions A_i d'une partition du plan temps-fréquence, où elle prend une valeur notée π_i [Ric98a]. Ce principe est illustré par la Figure 4.11. En vertu du fait que la statistique de détection $\Lambda(X)$ s'exprime alors sous la forme suivante

$$\Lambda(X) = \sum_{i=1}^r \sum_{[n_i, m_i] \in A_i} a_{TF}[n_i, m_i] W_X[n_i, m_i] = \sum_{i=1}^r \sum_{[n_i, m_i] \in A_i} \pi_i W_X[n_i, m_i] = \sum_{i=1}^r \pi_i \sum_{[n_i, m_i] \in A_i} W_X[n_i, m_i], \quad (4.27)$$

on impose de ce fait au détecteur d'exploiter, à un coefficient multiplicatif près, la valeur moyenne de la représentation de Wigner-Ville sur chacune des r régions A_i . Par cette heuristique, on espère ainsi contrôler la complexité de la structure de détection en limitant la redondance informationnelle due, entre autre, à la corrélation d'éléments $W_X[n, m]$ adjacents

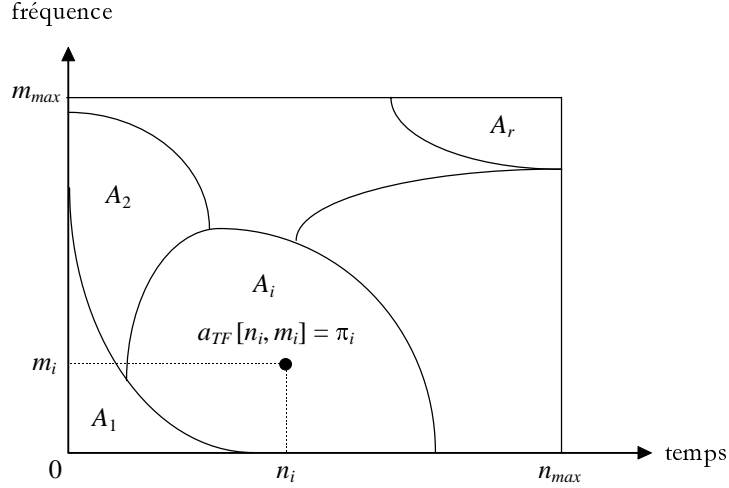


Figure 4.11 : Principe du partitionnement du plan temps-fréquence pour la détection.

[Fla93], tout en préservant en partie de l'information statistique nécessaire à la résolution du problème.

Le concept de *partitionnement du plan temps-fréquence* a parfois été exploité à d'autres fins. Dans [Cha97] par exemple, les auteurs recherchent une partition permettant de décomposer le signal analysé en composantes élémentaires, le débruitage et la classification étant des applications potentielles. Dans le cadre de la détection linéaire dans le plan temps-fréquence, la mise en oeuvre de ce concept, telle qu'elle a été proposée, permet de définir les classes de détecteurs $\mathcal{D}^{(r)}$ suivantes [Ric98a] :

$$\mathcal{D}^{(r)} = \left\{ d : \sum_{n,m} a_{TF}[n,m] W_x[n,m] - \lambda_0 \begin{array}{l} \underset{d(X)=0}{>} \\ \underset{d(X)=1}{<} \end{array} 0, a_{TF}[n,m] = \pi_i \text{ sur } A_i, i = 1, \dots, r \right\}, r \geq 1, \quad (4.28)$$

qui satisfont la relation $\mathcal{D}^{(1)} \subset \mathcal{D}^{(2)} \subset \dots \subset \mathcal{D}^{(r)} \subset \dots$ préconisée pour l'application du principe SRM. Conformément à ce dernier, la recherche de la structure $d_{n,opt}^{(r)}$ au sein de $\mathcal{D}^{(r)}$ se traduit alors par l'optimisation conjointe de la partition $\{A_1, \dots, A_r\}$, des paramètres $\{\pi_1, \dots, \pi_r\}$ et du seuil λ_0 , au sens de la probabilité d'erreur empirique. Selon que l'on adopte le principe SRM au sens strict ou large, le détecteur $d_{n,opt}$ finalement retenu est celui minimisant l'erreur garantie E_{gar} définie par l'expression (4.9), ou toute estimation de la probabilité d'erreur sur un ensemble de données indépendantes de \mathcal{A}_n .

Malheureusement, la détermination de $d_{n,opt}^{(r)}$ dans un tel contexte est un problème NP difficile, consistant à rechercher une partition du plan qui soit optimale au sens d'un critère donné. Toutefois, afin de trouver une solution approchée, on propose de coupler la méthode du critère optimal avec un algorithme classique de partitionnement, comme l'indique la Figure 4.12. Celui qui a été mis en oeuvre dans le cadre de l'expérimentation qui va suivre, repose sur la modélisation de la partition $\{A_1, \dots, A_r\}$ par un diagramme de Voronoï [Dev96],

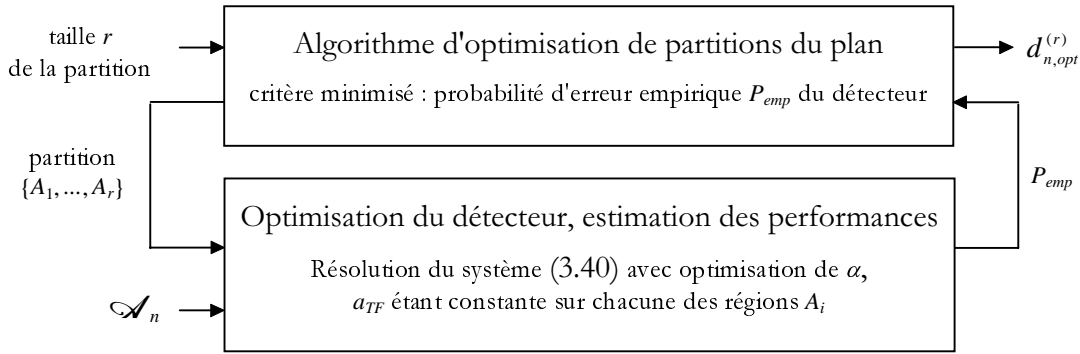


Figure 4.12. Contrôle de la capacité d'apprentissage d'un détecteur linéaire opérant dans le plan temps-fréquence, par la méthode de partitionnement.

que l'on optimise grâce à un algorithme de type génétique [Bac97]. L'étude détaillée de cette procédure de partitionnement, ainsi que des alternatives existantes, sort cependant du cadre de ce mémoire. En conséquence, le lecteur intéressé est invité à consulter [Sch96] par exemple, s'il souhaite obtenir d'avantage de précisions sur la modélisation et l'optimisation des partitions.

Afin de caractériser l'efficacité de la méthode de partitionnement pour le contrôle de la complexité des détecteurs opérant dans le plan temps-fréquence, le cadre d'expérimentation déjà commun aux méthodes variationnelle et de pénalisation a été conservé. On rappelle qu'il concerne la détection d'un signal aléatoire S noyé dans un bruit blanc, gaussien, centré et identiquement distribué sous les hypothèses H_0 et H_1 :

$$\begin{cases} H_0 : X = B \\ H_1 : X = S + B, \end{cases} \quad (4.29)$$

où $S(k) = k \exp(-0.45k) \sin(0.5k\pi + \Theta)$, $k \in \{1, \dots, 16\}$, Θ 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 -3 dB.

Le détecteur (4.26) a été optimisé sur la base de 300 réalisations de chacune des hypothèses, préalablement transformées en signaux analytiques, selon le principe décrit par la Figure 4.12. Etant donné une partition $\{A_1, \dots, A_r\}$, la valeur moyenne locale des représentations W_x de l'ensemble d'apprentissage a tout d'abord été évaluée sur chacune des r régions A_i , conformément à l'expression (4.27). Avant d'appliquer la méthode du critère optimal, chacun des moments statistiques conditionnels $M_{i=0,1}$ et $\Sigma_{i=0,1}$ des vecteurs de \mathcal{R}^r ainsi obtenus a alors été estimé au moyen de 200 données. Les autres réalisations ont quant à elles été réservées à l'évaluation des performances du détecteur au cours du processus

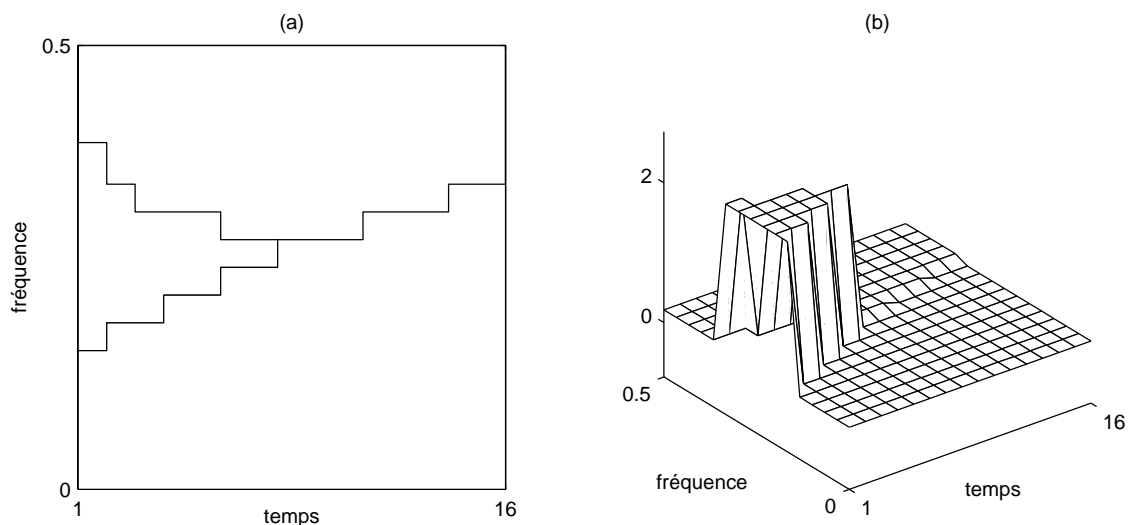


Figure 4.13 : Détection d'un signal à phase initiale aléatoire noyé dans un bruit blanc, gaussien et centré. La structure de détection est optimisée en associant les méthodes du critère optimal et de partitionnement. (a) Configuration de la partition obtenue. (b) Configuration du détecteur linéaire temps-fréquence. L'ensemble d'apprentissage est composé de 300 réalisations de chacune des hypothèses H_0 et H_1 .

d'apprentissage. Celles-ci ont ainsi permis d'ajuster d'une part le paramètre α intervenant dans l'équation (3.40), d'autre part de guider l'algorithme de partitionnement dans sa quête d'une meilleure solution $\{A_1, \dots, A_r\}$ pour l'itération suivante.

Le meilleur détecteur au sens de la probabilité d'erreur a été obtenu pour une partition composée de 3 régions, dont la configuration est donnée par la Figure 4.13.(a). Il apparaît que l'une des zones correspond approximativement au support temps-fréquence du signal S à détecter, ce qui est intuitivement satisfaisant. De plus, le résultat présenté par la Figure 4.13.(b) montre que la référence \hat{a}_{TF} associée à cette partition possède certains points communs avec W_s , à l'image d'une représentation temps-fréquence simplifiée. Afin de juger de son efficacité, les performances de ce détecteur ont été estimées sur un ensemble de 4000 réalisations de chacune des hypothèses. Les courbes COR présentées par la Figure 4.14 montre que les structures de détection obtenues grâce aux méthodes variationnelle, de pénalisation et de partitionnement sont satisfaisantes, au regard des performances du filtre adapté temps-fréquence. L'avantage se porte toutefois en faveur des méthodes dites de sélection, à la fois pour les résultats auxquels elles mènent et les propriétés théoriques qu'elles vérifient. Du point de vue de leur mise en oeuvre, les différentes approches présentées possèdent cependant des spécificités qu'il convient de prendre en compte au moment du choix. Ainsi, si les matrices de variance covariance Σ_i manipulées par la procédure de partitionnement sont de dimension $r \times r$, où r est la taille de la partition considérée, elles sont de dimension $L^2 \times L^2$ lorsqu'on adopte l'une des méthodes de sélection, où L^2 est le nombre

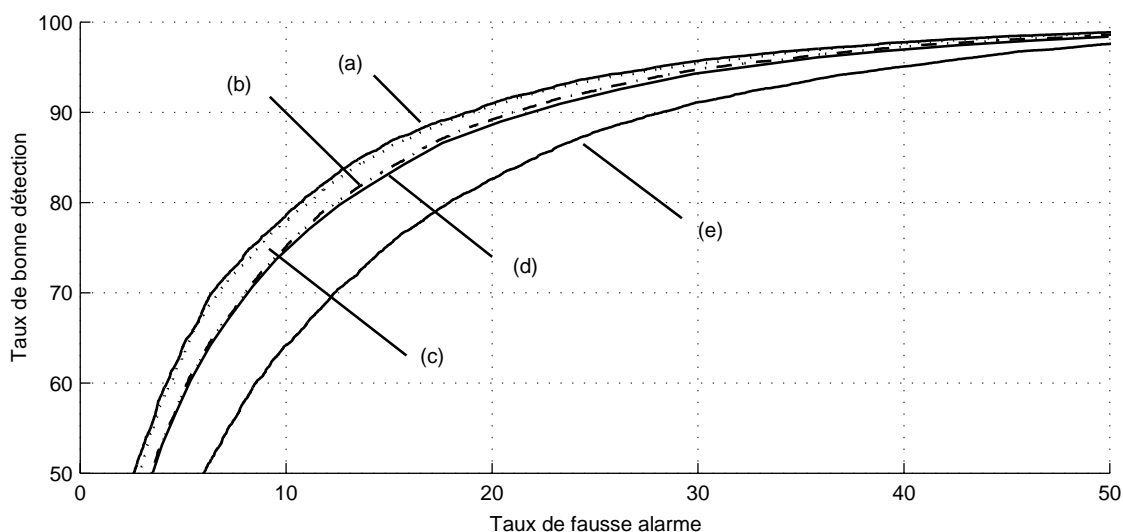


Figure 4.14 : Détection d'un signal à phase initiale aléatoire noyé dans un bruit blanc, gaussien et centré. (a) Courbe COR du filtre adapté opérant dans le plan temps-fréquence. Courbes COR des détecteurs linéaires opérant dans le plan temps-fréquence, obtenus par association de la méthode du critère optimal et des méthodes (b) variationnelle, (c) de pénalisation ou (d) de partitionnement, lorsqu'on dispose d'un ensemble d'apprentissage constitué de 300 réalisations de chacune des hypothèses H_0 et H_1 . (e) Courbe COR du détecteur obtenu dans les mêmes conditions que (b), (c) et (d), lorsqu'il n'est fait appel à aucune procédure de contrôle de la complexité.

d'éléments de W_x . Dans le cadre de la synthèse de détecteurs opérant dans le plan temps-fréquence, ce dernier point limite par conséquent l'usage des procédures variationnelle et de pénalisation aux signaux raisonnablement courts. L'objet de la prochaine section est de montrer qu'il est néanmoins envisageable de repousser quelque peu cette limite, en raison de la redondance informationnelle présente dans la représentation de Wigner-Ville.

4.3 Transformée de Wigner-Ville discrète : Dimension de l'espace Image

La distribution de Wigner-Ville joue un rôle majeur dans le domaine de l'analyse temps-fréquence, du fait des nombreuses propriétés qu'elle vérifie. Cependant, en raison du volume important de données qui la caractérise, son usage demeure délicat dans certaines applications telles que la détection et la classification de signaux non-stationnaires. L'objet de cette section est de montrer que l'information véhiculée par cette représentation est toutefois très redondante du fait de l'existence de relations linéaires entre les composantes $W_x[n,m]$. Plus précisément, on démontre que la totalité de la représentation de Wigner-Ville d'un signal réel, complexe ou analytique peut être linéairement reconstituée à partir d'un nombre limité d'éléments $W_x[n,m]$ particuliers. La considération d'une telle propriété peut évidemment

mener à de nouveaux algorithmes dans le domaine de l'analyse temps-fréquence, plus rapides et moins coûteux en mémoire, cette redondance informationnelle étant non négligeable. A titre d'exemple, il apparaît en effet qu'environ un quart seulement des composantes $W_x[n,m]$ de la représentation de Wigner-Ville d'un signal analytique sont linéairement indépendantes.

Cette section est organisée ainsi. Avant de poser le problème, les conventions adoptées sont préalablement précisées. Puis, le calcul du nombre maximum d'éléments $W_x[n,m]$ linéairement indépendants est effectué, en distinguant le cas des signaux réels, complexes et analytiques. En guise d'illustration, un détecteur linéaire opérant uniquement sur des composantes linéairement indépendantes et aptes à engendrer la totalité de la représentation de Wigner-Ville est enfin synthétisé. Celui-ci présente évidemment les mêmes performances que la structure de détection opérant sur l'ensemble des éléments $W_x[n,m]$.

Les résultats présentés dans cette section sont, à notre connaissance, tous originaux. Ils font l'objet de la référence [Ric98c].

4.3.1 Conventions et position du problème

Soit $x = (x[0], \dots, x[L-1])^T$ un signal constitué de L échantillons, appartenant à un espace linéaire donné que l'on note \mathbf{E} . Le paramètre L est supposé pair, sans que cela nuise au caractère général du raisonnement qui suit. Dans ces circonstances, on rappelle que la transformée de Wigner-Ville discrète du signal x est alors donnée par :

$$W_x[n,m] = 2 \sum_{k=1-L/2}^{L/2-1} R_x[n,k] \exp\left(-\frac{4j\pi mk}{L}\right), \quad (n,m) \in [0, \dots, L-1]^2, \quad (4.30)$$

$$\text{où } R_x[n,k] = x[n+k] x^*[n-k]. \quad (4.31)$$

En s'appuyant sur la définition (4.30), on définit les applications f et g suivantes, telles que $g \circ f$ fasse correspondre à tout élément x de l'espace linéaire \mathbf{E} , l'élément $W_x = (\dots, W_x[n,m], \dots)^T$, $(n,m) \in [0, \dots, L-1]^2$:

$$x \xrightarrow{f} R_x = (\dots, R_x[n,k], \dots)^T, \quad (n+k, n-k) \in [0, L-1]^2, \quad (4.32)$$

$$R_x \xrightarrow{g} W_x = (\dots, W_x[n,m], \dots)^T, \quad (n,m) \in [0, L-1]^2. \quad (4.33)$$

Soient \mathbf{F} et \mathbf{G} les espaces Image respectifs de f et $g \circ f$, c'est à dire :

$$\mathbf{F} = \{R_x, x \in \mathbf{E}\}, \quad (4.34)$$

$$\mathbf{G} = \{W_x, x \in \mathbf{E}\}. \quad (4.35)$$

\mathbf{F} et \mathbf{G} sont respectivement nommés espaces Image de la fonction d'autocorrélation instantanée et de la transformée de Wigner-Ville discrètes.

La somme de deux transformées de Wigner-Ville discrètes n'étant pas une transformée de Wigner-Ville valide, \mathbf{G} n'est pas un espace linéaire. En conséquence, on est amené à reconsidérer la notion de dimension pour cet espace. Dans le cadre de cet exposé, on définit la dimension de \mathbf{G} , que l'on note $\dim(\mathbf{G})$, par le nombre minimum de vecteurs linéairement indépendants requis pour engendrer tous les éléments de cet espace, au moyen de combinaisons linéaires. L'addition n'étant pas non plus une loi de composition interne pour \mathbf{F} , on adoptera en conséquence la même définition pour $\dim(\mathbf{F})$ que pour $\dim(\mathbf{G})$.

L'équation (4.30) prenant la forme d'une transformée de Fourier discrète, étant donné n , les espaces \mathbf{F} et \mathbf{G} sont isomorphes et par conséquent de même dimension. Puisque $\dim(\mathbf{G})$ est en particulier égal au nombre recherché de composantes $W_x[n, m]$ linéairement indépendantes dans une représentation de Wigner-Ville discrète, on s'attache à déterminer $\dim(\mathbf{F})$ dans les paragraphes qui suivent. Pour cela, on est amené à distinguer le cas des signaux réels, complexes et analytiques.

4.3.2 Dimension des espaces Image de \mathbf{R}_x et \mathbf{W}_x

Les trois propositions qui suivent établissent la dimension de l'espace Image de la fonction d'autocorrélation instantanée discrète, et par conséquent celle de la transformée de Wigner-Ville discrète.

Proposition 4.3. Soit $x = (x[0], x[2], \dots, x[L-1])^T$ un signal réel constitué de L échantillons, L étant supposé pair. La dimension de l'espace Image \mathbf{F} de la fonction d'autocorrélation instantanée discrète $R_x = (\dots, x[n+k] x^*[n-k], \dots)^T$, $(n+k, n-k) \in [0, L-1]^2$, est donnée par :

$$\dim(\mathbf{F}) = L^2 / 4 + L / 2 . \quad (4.36)$$

Preuve. En s'appuyant sur la définition des composantes de R_x , on construit une base de vecteurs $e^{(i,j)}$ de \mathbf{F} , comme l'indique la Figure 4.15. Leur dénombrement, auquel on va procéder, fournit alors la dimension $\dim(\mathbf{F})$ recherchée, qui satisfait :

$$\dim(\mathbf{F}) \equiv \text{card} \{e^{(i,j)}\} = \text{card} \{(n+k, n-k) \in [0, \dots, L-1]^2, k \geq 0\} . \quad (4.37)$$

La combinaison des conditions $0 \leq n+k \leq L-1$ et $0 \leq n-k \leq L-1$, que définit l'expression (4.37), conduit à :

$$\max \{-k ; k\} \leq n \leq \min \{L-k-1 ; L+k-1\}, \text{ étant donné } k \geq 0,$$

que l'on peut encore exprimer ainsi :

$$0 \leq n \leq L-2k-1, \text{ à condition que } 0 \leq L-2k-1, \text{ c'est à dire que } 0 \leq k \leq \text{Int}\{(L-1)/2\}.$$

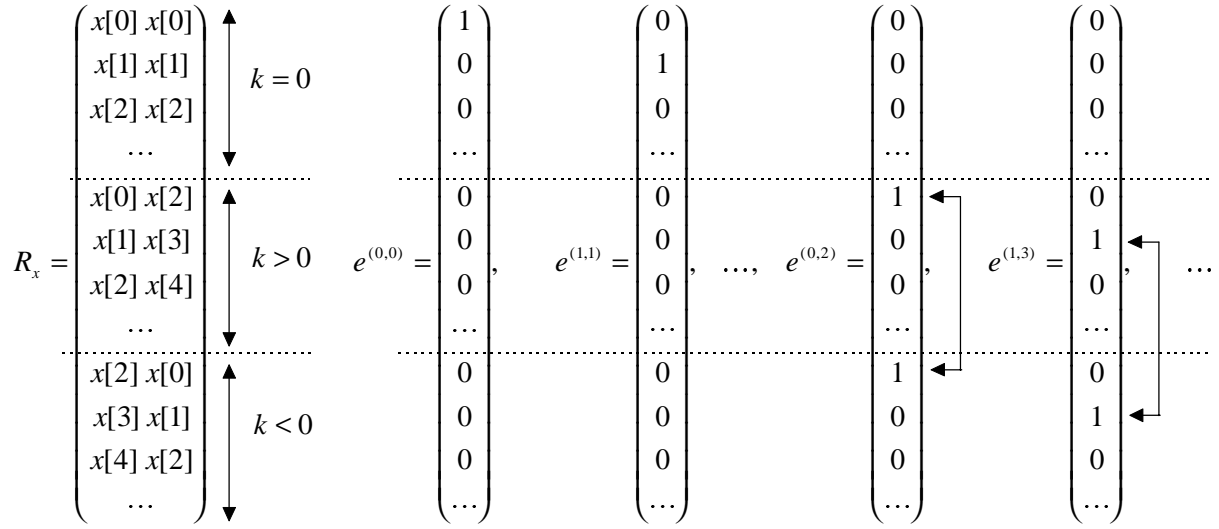


Figure 4.15. Détermination d'une base de l'espace \mathbf{F} lorsque \mathbf{E} désigne l'espace des signaux réels.

Sachant que $\text{Int}\{(L-1)/2\} = (L-1)/2$ puisque le paramètre L est pair, les couples (n, k) valides vérifient les deux conditions suivantes : $0 \leq n \leq L-2k-1$ avec $0 \leq k \leq (L-1)/2$.

Le dénombrement des solutions envisageables conduit alors au résultat suivant :

$$\dim(\mathbf{F}) = \dim(\mathbf{G}) = \sum_{k=0}^{(L-1)/2} (L-2k) = (L^2/4) + (L/2).$$

En conclusion, la représentation de Wigner-Ville d'un signal réel de L échantillons peut être générée à partir de $L^2/4 + L/2$ composantes $W_x[n, m]$ linéairement indépendantes. ■

Proposition 5.4. Soit $x = (x[0], x[2], \dots, x[L-1])^T$ un signal complexe non analytique de L échantillons, L étant supposé pair. La dimension de l'espace Image \mathbf{F} de la fonction d'autocorrélation instantanée $R_x = (\dots, x[n+k]x^*[n-k], \dots)^T$, $(n+k, n-k) \in [0, L-1]^2$, est donnée par :

$$\dim(\mathbf{F}) = L^2/2. \quad (4.38)$$

Preuve. La procédure adoptée pour déterminer la dimension de \mathbf{F} , lorsque x est complexe mais non analytique, consiste également à dénombrer les vecteurs d'une base de cet espace. En particulier, la base présentée en Figure 4.16 rend compte, d'une part que les doubles produits $x[k]x^*[k]$ appartiennent à \mathcal{R} , d'autre part que les termes complexes $x[n+k]x^*[n-k]$ et $x[n-k]x^*[n+k]$, $k \neq 0$, sont conjugués.

En adoptant les notations de la Figure 4.16, on a ainsi :

$$\begin{aligned} \dim(\mathbf{F}) &\equiv \text{card}\{e^{(i,j)}\} + \text{card}\{e_{\text{réel}}^{(i,j)}\} + \text{card}\{e_{\text{imag}}^{(i,j)}\} \\ &= \text{card}\{(n, n) \in [0, \dots, L-1]^2\} + 2 \text{card}\{(n+k, n-k) \in [0, \dots, L-1]^2, k > 0\}. \end{aligned} \quad (4.39)$$

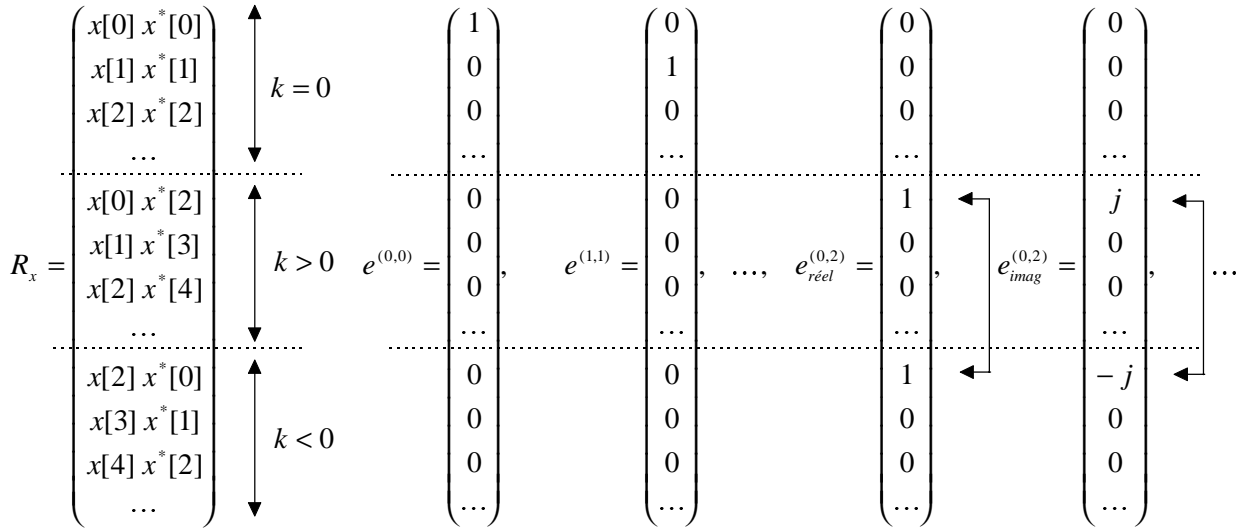


Figure 4.16. Détermination d'une base de l'espace \mathbf{F} lorsque \mathbf{E} désigne l'espace linéaire des signaux complexes, non analytiques.

Or, d'après la Proposition 4.3, on sait que :

$$\text{card}\{(n+k, n-k) \in [0, \dots, L-1]^2, k > 0\} = \text{card}\{(n+k, n-k) \in [0, \dots, L-1]^2, k \geq 0\} - L \quad (4.40)$$

$$= (L^2 / 4) - (L / 2).$$

En combinant les expressions (4.39) et (4.40), on en déduit donc que :

$$\dim(\mathbf{F}) = L^2 / 2$$

En conclusion, la représentation de Wigner-Ville d'un signal complexe mais non analytique, de L échantillons, peut être générée à partir de $L^2 / 2$ éléments $W_x[n, m]$ linéairement indépendants. ■

Le cas des signaux analytique est certainement le plus important puisqu'en conduisant à l'annulation des composantes portées par les fréquences négatives, l'usage de la transformée de Hilbert permet une amélioration de la lisibilité des représentations de la classe de Cohen.

Proposition 4.5. Soit $x = (x[0], x[2], \dots, x[L-1])^T$ un signal analytique constitué de L échantillons, L étant supposé pair. La dimension de l'espace Image \mathbf{F} de la fonction d'autocorrélation instantanée $R_x = (\dots, x[n+k] x^*[n-k], \dots)^T, (n+k, n-k) \in [0, L-1]^2$, est donnée par :

$$\dim(\mathbf{F}) = L^2 / 4 + L - 1. \quad (4.41)$$

Preuve. Soient $y = (y[0], \dots, y[L/2], 0, \dots, 0)^T$ la transformée de Fourier du signal analytique x , et \mathbf{H} l'espace engendré par les doubles produits des composantes de y :

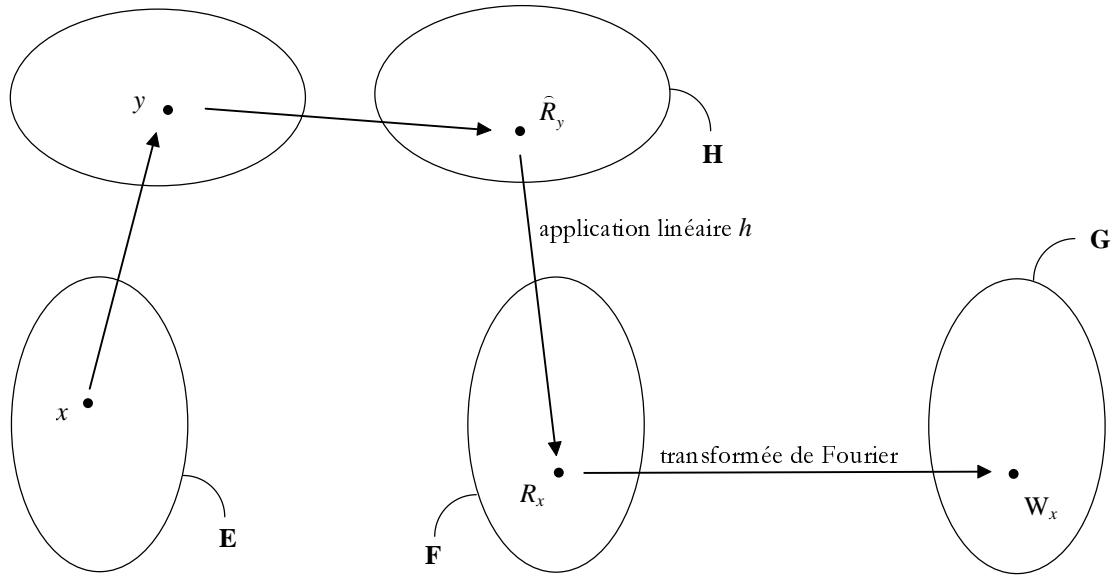


Figure 4.17 : Principe de l'évaluation de la dimension de l'espace **G** lorsque **E** désigne l'espace des signaux analytiques.

$$\mathbf{H} = \left\{ \hat{R}_y = [\dots, y[p] y^*[q], (p, q) \in [0, \dots, L/2]^2] \right\}. \quad (4.42)$$

Soit h l'application linéaire qui, à tout élément \hat{R}_y de **H**, fait correspondre l'élément R_x de **F** défini ainsi :

$$x[n+k] x^*[n-k] = \frac{1}{L^2} \sum_{p=0}^{L/2} \sum_{q=0}^{L/2} y[p] y^*[q] \exp\left(2j\pi \frac{(n+k)p - (n-k)q}{L}\right). \quad (4.43)$$

La stratégie adoptée pour évaluer la dimension de **F** consiste, comme l'indique la Figure 4.17, à introduire entre **E** et **F** un espace intermédiaire **H**, dont la définition est donnée par l'expression (4.42). En désignant par $\{e^{(m,n)}\}$ une base de celui-ci, le nombre de vecteurs $h(e^{(m,n)})$ linéairement indépendants fournit alors les dimensions $\dim(\mathbf{F})$ et $\dim(\mathbf{G})$ recherchées.

La démarche suivie pour déterminer une base de l'espace **H** est identique à celle qui a été adoptée dans le cas des signaux réels et complexes. Le principe de construction donné par la Figure 4.18 prend ainsi en compte les trois propriétés suivantes :

1. Les doubles produits $y[p] y^*[q]$, $p = q$, sont réels ;
2. Les doubles produits $y[0] y^*[L/2]$ et $y[L/2] y^*[0]$ sont réels, $y[0]$ et $y[L/2]$ étant réels puisque x est un signal analytique ;

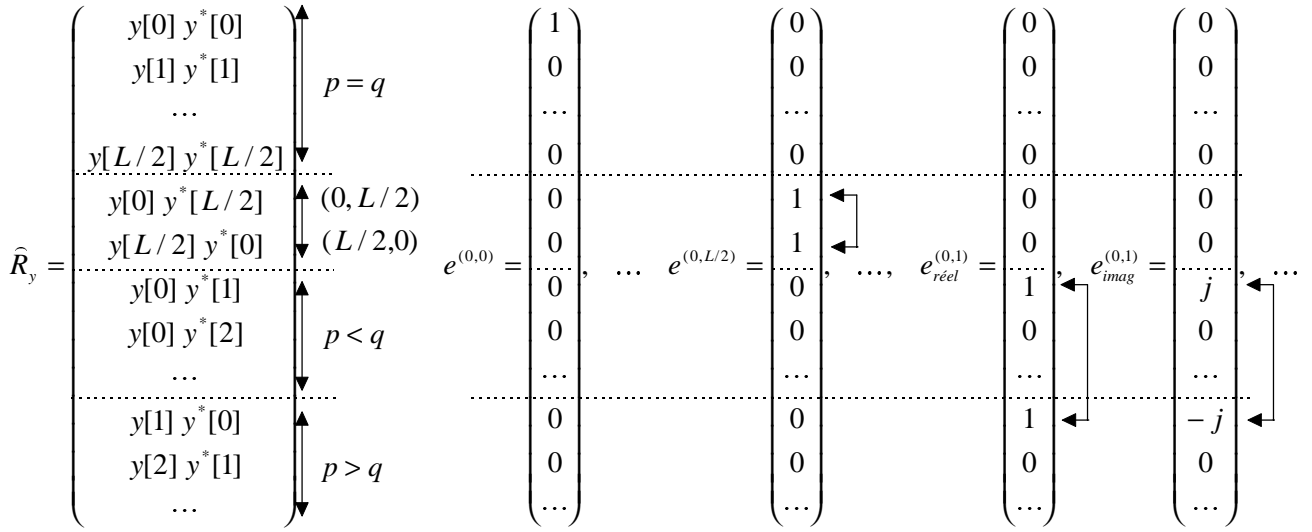


Figure 4.18. Détermination d'une base de l'espace \mathbf{F} lorsque \mathbf{E} désigne l'espace linéaire des signaux analytiques.

3. Les doubles produits n'entrant pas dans l'une des deux catégories précédentes sont complexes.

Dans ces circonstances, la dimension de l'espace \mathbf{H} est donnée par l'expression suivante :

$$\begin{aligned}
 \dim(\mathbf{H}) &\equiv \text{card} \left\{ e_{\text{réel}}^{(p,q)}, (p,q) \in [0, \dots, L/2]^2 - (0, L/2), p < q \right\} \\
 &\quad + \text{card} \left\{ e_{\text{imag}}^{(p,q)}, (p,q) \in [0, \dots, L/2]^2 - (0, L/2), p < q \right\} \\
 &\quad + \text{card} \left\{ e^{(p,p)}, p \in [0, \dots, L/2] \right\} + \text{card} \left\{ e^{(0, L/2)} \right\} \\
 &= 2 \text{card} \left\{ (p,q), (p,q) \in [0, \dots, L/2]^2 - (0, L/2), p < q \right\} + (L/2 + 1) + 1.
 \end{aligned} \tag{4.44}$$

Or, l'évaluation du premier terme de la somme indiquée ci-dessus donne :

$$\text{card} \left\{ (p,q), (p,q) \in [0, \dots, L/2]^2 - (0, L/2), p > q \right\} = \sum_{i=1}^{L/2+1} i - (L/2 + 1) - 1 = L^2 / 8 + L / 4 - 1,$$

ce qui implique que la dimension de l'espace \mathbf{H} est égale à :

$$\dim(\mathbf{H}) = L^2 / 4 + L. \tag{4.45}$$

En conséquence, l'espace \mathbf{F} engendré par l'ensemble des vecteurs $\{h(e^{(m,n)})\}$ est au plus de dimension $L^2 / 4 + L$. En fait, on peut montrer qu'il existe une unique relation linéaire entre ces vecteurs, qui est : $h(e^{(0,0)}) = h(e^{(L/2, L/2)})$. Ce résultat conduit à :

$$\dim(\mathbf{F}) = L^2 / 4 + L - 1.$$

En conclusion, la représentation de Wigner-Ville d'un signal analytique de L échantillons peut être générée à partir de $L^2 / 4 + L - 1$ composantes $W_x[n, m]$ linéairement indépendantes. ■

4.3.3 Exemples

Afin d'illustrer les propriétés qui ont été énoncées, on présente tout d'abord un exemple de reconstruction de la représentation de Wigner-Ville d'un signal analytique de L échantillons, à partir de $L^2/4 + L - 1$ composantes $W_x[n, m]$ linéairement indépendantes. Puis, un détecteur linéaire nécessitant uniquement ce type d'information est généré. On montre alors qu'il présente les mêmes performances que la structure de détection linéaire opérant sur la totalité de la représentation.

a) *Reconstruction d'une représentation.* Les composantes temps-fréquence de couleur claire, sur la Figure 4.19.(a), sont linéairement indépendantes et véhiculent l'information nécessaire à la reconstruction de la représentation de Wigner-Ville de tout signal analytique de 16 échantillons. Celles-ci ont été exhibées en générant préalablement une matrice, notée M , dont chacune des 2000 lignes correspond à la représentation temps-fréquence d'un bruit de L échantillons, initialement blanc avant d'être transformé en un signal analytique. Les composantes recherchées, au nombre de $L^2/4 + L - 1$ dans ce cas, ont alors été isolées en étudiant la décroissance du rang de la matrice M , consécutive à l'annulation successive de chacune de ses colonnes. Les éléments $W_x[n, m]$ ainsi identifiés dépendent évidemment de l'ordre dans lequel s'effectue cette analyse de l'interdépendance linéaire des colonnes de M .

Au prix d'éventuelles permutations, on suppose dorénavant que les $L^2/4 + L - 1$ premières colonnes de M sont linéairement indépendantes, tout comme les $L^2/4 + L - 1$ premières composantes des représentations W_x , pour lesquelles on adopte une écriture vectorielle. Afin

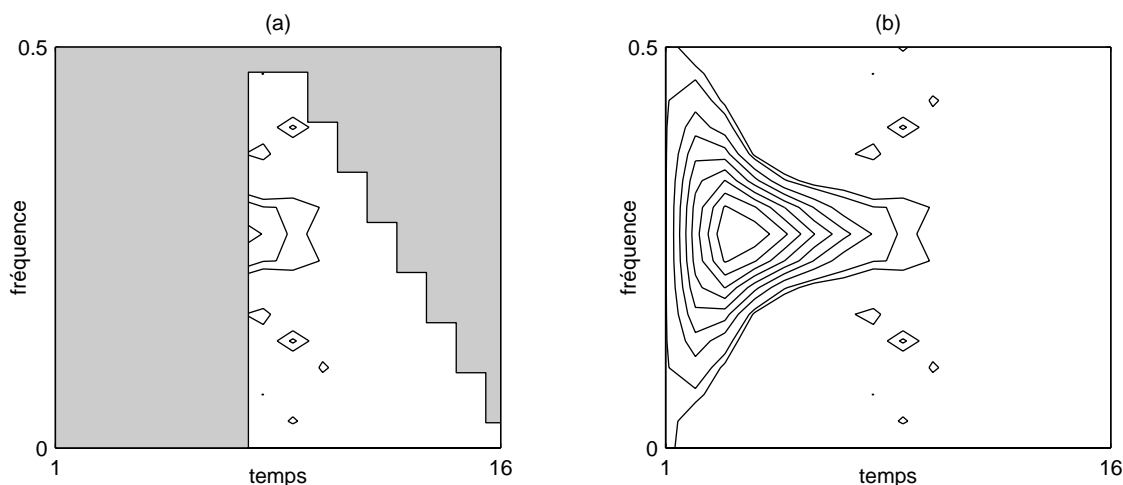


Figure 4.19 : L'information véhiculée par les composantes temps-fréquence de couleur claire sur (a) permet de reconstruire une représentation de Wigner-Ville dans sa totalité, comme l'illustre (b). Le signal étudié dans cet exemple est défini ainsi : $s(k) = k \exp(-0.45k) \sin(0.5k\pi)$, $k \in \{1, \dots, 16\}$.

d'estimer la matrice Δ qui permet, à partir des éléments identifiés précédemment, de reconstituer la totalité d'une représentation de Wigner-Ville ainsi :

$$W_x[i] = \sum_{j=1}^{L^2/4+L-1} \Delta[i, j] W_x[j], \quad i \in \{L^2/4+L, \dots, L^2\}, \quad (4.46)$$

les $L^2 - (L^2/4 + L - 1)$ problèmes aux moindres carrés suivants ont été résolus :

$$\hat{\Delta}[i, \cdot] = \arg \min_{\delta \in \mathcal{R}^{L^2/4+L-1}} \left\| M[\cdot, i] - \sum_{j=1}^{L^2/4+L-1} \delta[j] M[\cdot, j] \right\|^2, \quad i \in \{L^2/4+L, \dots, L^2\}. \quad (4.47)$$

L'erreur quadratique associée à la résolution de chacun d'eux est théoriquement nulle, ce que l'on a pu vérifier expérimentalement.

b) *Synthèse d'un détecteur.* Le second exemple proposé concerne la synthèse d'un détecteur linéaire opérant uniquement sur les $L^2/4 + L - 1$ premières composantes de W_x , que l'on suppose linéairement indépendantes :

$$d(X) = \begin{cases} 0 & \text{si } \Lambda(X) = \sum_{i=1}^{L^2/4+L-1} \hat{a}_{TF}[i] W_x[i] - \lambda_0 < 0 \\ 1 & \text{sinon.} \end{cases} \quad (4.48)$$

Avant de procéder à cette expérimentation, il convient d'établir les relations liant \hat{a}_{TF} et la référence a_{TF} opérant sur la totalité de la représentation de Wigner-Ville de l'observation X .

La statistique de détection $\Lambda(X)$, mettant en oeuvre la référence temps-fréquence a_{TF} , peut être reformulée ainsi :

$$\Lambda(X) = \sum_{i=1}^{L^2} a_{TF}[i] W_x[i] - \lambda_0 = \sum_{i=1}^{L^2/4+L-1} \left\{ a_{TF}[i] + \sum_{j=L^2/4+L}^{L^2} \Delta[i, j] a_{TF}[j] \right\} W_x[i] - \lambda_0, \quad (4.49)$$

L'identification terme à terme des statistiques définies par les expressions (4.48) et (4.49) conduit au résultat suivant :

$$\hat{a}_{TF}[i] = a_{TF}[i] + \sum_{j=L^2/4+L}^{L^2} \Delta[i, j] a_{TF}[j], \quad i \in \{1, \dots, L^2/4+L-1\}. \quad (4.50)$$

Réciproquement, les $L^2/4 + L - 1$ premières composantes de a_{TF} peuvent être évaluées à partir de \hat{a}_{TF} en résolvant le système linéaire suivant, obtenu en combinant les équations (4.46) et (4.50) :

$$\hat{a}_{TF}[i] = a_{TF}[i] + \sum_{j=L^2/4+L}^{L^2} \Delta[i, j] a_{TF}[j], \quad i \in \{1, \dots, L^2/4+L-1\}, \quad (4.51)$$

$$\text{où } a_{TF}[j] = \sum_{k=1}^{L^2/4+L-1} \Delta[k, j] a_{TF}[k], \quad j \in \{L^2/4+L, \dots, L^2\}. \quad (4.52)$$

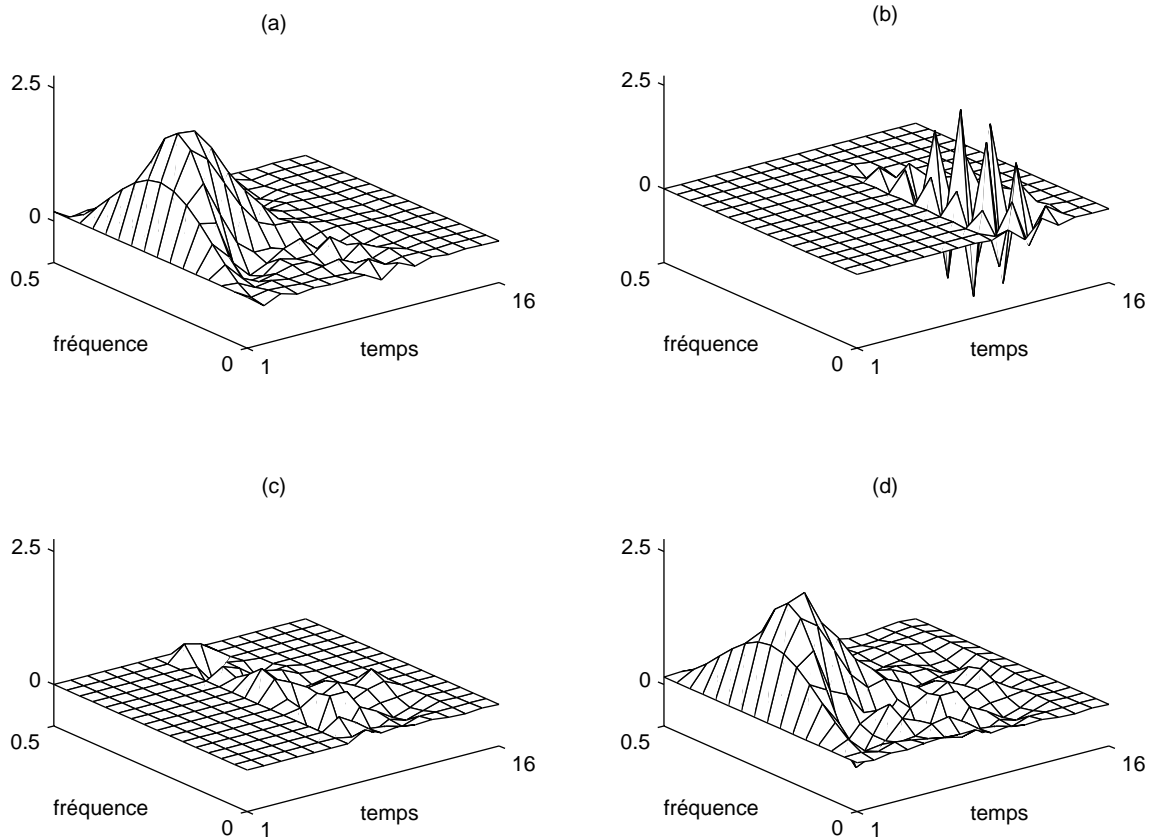


Figure 4.20 : Synthèse d'un détecteur linéaire opérant uniquement sur une région particulière du plan temps-fréquence. (b) La référence obtenue dans ces circonstances, grâce à la méthode du critère optimal, ne présente *a priori* aucun point commun avec (a) la représentation de Wigner-Ville du signal à détecter. Et pourtant, le développement de cette référence sur la totalité du plan temps-fréquence, obtenu en appliquant successivement les relations (4.50) et (4.46), permet de retrouver en deux étapes (c) et (d), une configuration satisfaisante.

Afin de retrouver $a_{TF}[i]$, $i \in \{L^2/4 + L, \dots, L^2\}$, il ne reste alors plus qu'à appliquer la relation (4.52).

Pour illustrer ces résultats, la référence temps-fréquence \hat{a}_{TF} , opérant sur les composantes déterminées lors de la section précédente, a été optimisée pour la résolution du problème de détection (4.21). Pour cela, la méthode du critère optimal a été mise en oeuvre sur la base de 10000 réalisations de chacune des hypothèses H_0 et H_1 , préalablement transformées en signaux analytiques. On note que la référence \hat{a}_{TF} obtenue, présentée en Figure 4.20.(b), ne semble avoir aucun point commun avec la représentation de Wigner-Ville du signal à détecter, qui est donnée en Figure 4.20.(a). Le système linéaire (4.51)-(4.52) a alors été résolu afin de déterminer $a_{TF}[k]$, $k \in \{1, \dots, L^2/4 + L - 1\}$, qui est représenté en Figure 4.20.(c). Ce résultat a enfin été développé sur l'ensemble du plan temps-fréquence en appliquant la relation (4.52). La référence a_{TF} résultant de ces opérations, qui est donnée par la

Figure 4.20.(d), s'avère convaincante puisqu'elle est caractéristique de la configuration du filtre adapté temps-fréquence, ici optimal. Avec des probabilités de bonne détection égales à 0.78 pour une probabilité de fausse alarme de 0.1, les références a_{TF} et \hat{a}_{TF} présentent évidemment les mêmes performances.

4.4 Conclusion

Dans un premier temps, ce chapitre a été consacré à la question du contrôle de la complexité des détecteurs linéaires généralisés en vue d'une optimisation de leur performances, lorsqu'ils sont établis à partir de données d'apprentissage. Afin d'apporter des éléments de solution à ce problème, on a ainsi fait appel au principe SRM qui constitue, par le biais de l'inégalité de Vapnik-Chervonenkis sur laquelle il repose, l'un des fondements de la théorie de l'apprentissage. La mise en oeuvre de celui-ci suppose cependant que l'on ait la maîtrise de la VC-dimension de la classe de détecteurs \mathcal{D} considérée. En conséquence, on a été amené à présenter trois méthodes remplissant cette fonction, deux d'entre elles s'inspirant de techniques d'élagage des réseaux de neurones artificiels. A cette occasion, on a montré que le mode opératoire de ces dernières repose sur une pondération appropriée des composantes de l'observation. Appliquée au cadre de la détection linéaire dans le plan temps-fréquence, cette propriété a permis de mettre en évidence que le contrôle de la capacité d'apprentissage des détecteurs est également envisageable par l'intermédiaire de la fonction de paramétrisation ϕ de la représentation $C_x[n, m; \phi]$ sur laquelle ils opèrent. C'est là une illustration originale des potentialités de la classe de Cohen pour la résolution de certains problèmes de détection à structure imposée, qu'il conviendra d'approfondir ultérieurement.

Enfin, ce chapitre a permis de mettre en évidence que la dimension de l'espace engendré par la représentation de Wigner-Ville des signaux réels, complexes ou analytiques, est inférieure au nombre de composantes $W_x[n, m]$ qui la constitue. A titre d'illustration, cette propriété a été mise en oeuvre pour la reconstruction complète de la représentation de Wigner-Ville, à partir d'une connaissance partielle, mais non quelconque, de celle-ci. Hormis les perspectives offertes en termes d'algorithmes plus rapides et moins coûteux en mémoire, il semble que cette propriété puisse être utilisée à des fins de débruitage. En effet, on peut envisager de sélectionner une région du plan temps-fréquence où le rapport signal sur bruit paraît favorable, puis de la filtrer avant de procéder à la reconstruction de la représentation.

Chapitre 5

Détection des complexes K dans l'électroencéphalogramme de sommeil

5.1 Introduction

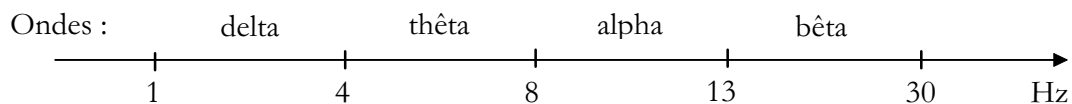
L'étude des phénomènes transitoires de l'électroencéphalogramme a débuté dans les années 30 [Loo35] avec la première description des fuseaux de sommeil. L'objectif de ces travaux était alors d'établir une première classification, en stades, de l'activité cérébrale humaine durant le sommeil. A l'heure actuelle, les efforts se concentrent d'avantage sur l'analyse des phénomènes transitoires, l'un des principaux objectifs étant la compréhension des mécanismes générateurs du sommeil et du rôle physiologique de l'activité électrique du cerveau. Dans ce domaine de recherche, l'utilisation de moyens automatiques de traitement de l'information demeure un facteur de progrès important en raison du volume considérable de données à analyser [Gai73], [Ray86], [Rey87], [Sha93], [Cim97]. En effet, l'expertise des signaux de sommeil, extrêmement coûteuse et astreignante lorsqu'elle doit être pratiquée visuellement, ne peut être raisonnablement envisagée à grande échelle, à moins d'être automatisée.

L'objet de ce chapitre est de proposer un outil destiné à la détection d'un signal transitoire particulier de l'électroencéphalogramme de sommeil : le *complexe K*. Le choix de ce problème se justifie par une assez grande difficulté à le résoudre en raison d'une ressemblance marquée de cet événement avec d'autres phénomènes observés en sommeil profond : les *bouffées d'ondes delta*. Ce chapitre, qui fait l'objet des références [Ric98d] et [Ric98e], est organisé ainsi. Dans un premier temps, on procède à une description succincte de l'électroencéphalogramme de sommeil, de ses phénomènes transitoires et plus particulièrement du complexe K. Puis, une présentation de la méthodologie adoptée pour synthétiser un détecteur de complexes K est effectuée. Enfin, les performances de la structure de détection obtenue sont analysées, puis comparées à d'autres solutions existantes.

5.2 Présentation du problème

5.2.1 Macrostructure de l'électroencéphalogramme

L'exploration physiologique du sommeil implique l'étude de plusieurs signaux, tels que l'électroencéphalogramme, l'électrooculogramme ou encore l'électromyogramme. Le résultat de cette analyse prend alors la forme d'une classification, en cinq stades, de séquences successives de 30 secondes de sommeil. Ces stades sont essentiellement définis par la nature des signaux alors rencontrés dans l'électroencéphalogramme, parmi lesquels on recense les bouffées d'ondes delta, thêta, alpha ou bêta, que l'on identifie par leur support fréquentiel [Cim97] :



En s'appuyant sur cette nomenclature, les cinq stades de sommeil venant d'être évoqués, et illustrés par la Figure 5.1, sont définis ainsi :

Stade 1 : fragmentation des ondes alpha et survenue des ondes thêta.

Stade 2 : apparition des fuseaux de sommeil et des complexes K.

Stade 3 : présence, sur 20 à 50 % du tracé, d'ondes lentes de grande amplitude, dites delta.

Stade 4 : présence de bouffées d'ondes delta sur plus de 50% du tracé.

Sommeil paradoxal : atonie musculaire et mouvements oculaires rapides.

Conventionnellement, le *sommeil léger* regroupe les stades 1 et 2, tandis que le *sommeil profond* est composé des stades 3 et 4.

Si cette classification fournit des informations essentielles pour l'identification de certaines anomalies du sommeil, elle conduit néanmoins à un diagnostic incomplet. En effet, elle ne tient pas explicitement compte, par exemple, de la fréquence d'apparition des phénomènes transitoires, pourtant révélatrice de quelques pathologies. L'objet de la prochaine section est précisément de décrire certains de ces événements, dit phasiques, parmi lesquels on recense le complexe K.

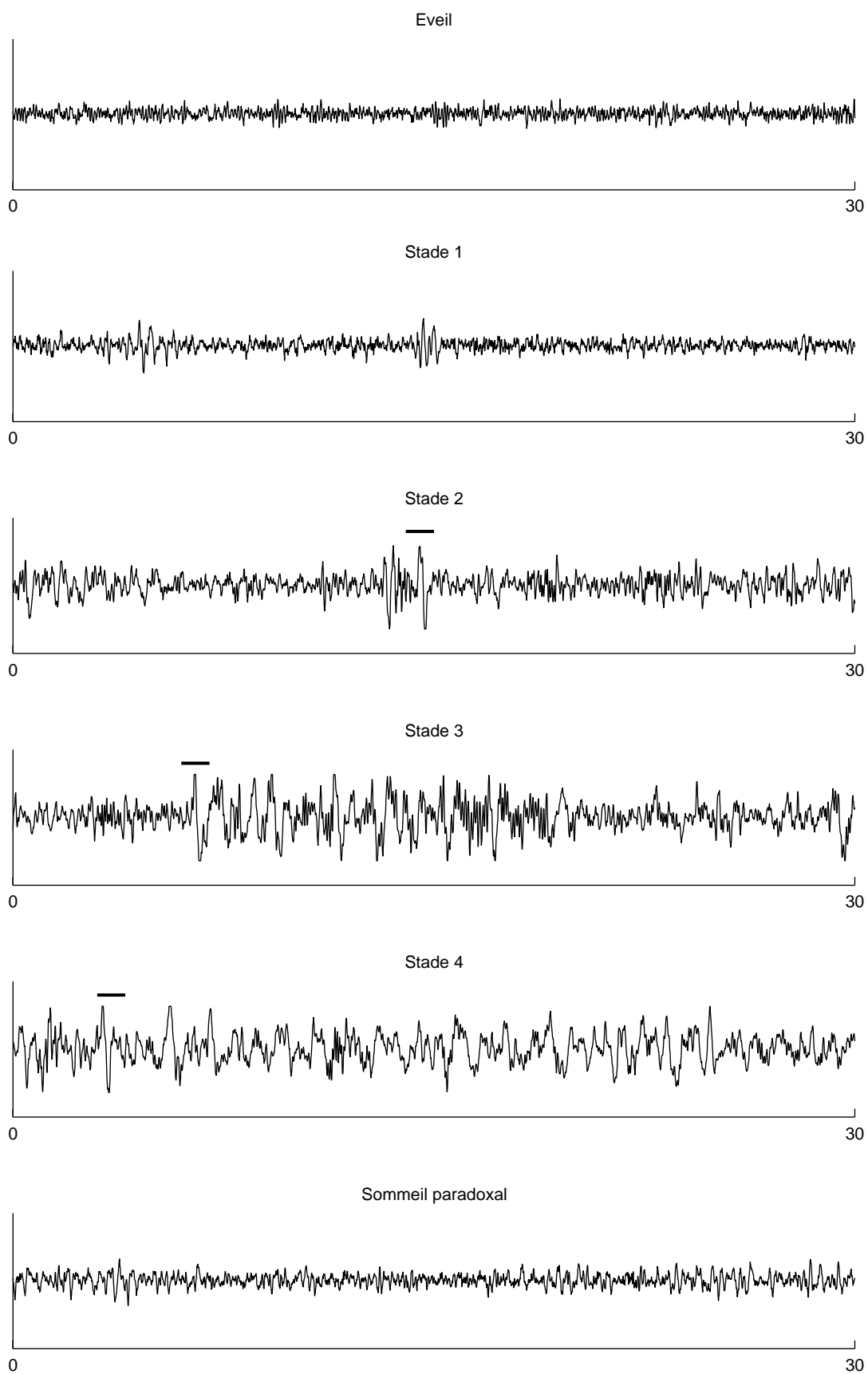


Figure 5.1. Illustration des 5 stades de sommeil, au moyen de séquences de 30 secondes d'électroencéphalogramme. Les segments horizontaux figurant sur les tracés des stades 2, 3 et 4 indiquent la présence de complexes K.

5.2.2 Microstructure de l'électroencéphalogramme

Dans cette section, on décrit succinctement les trois événements phasiques les plus connus de l'électroencéphalogramme de sommeil, sans s'attarder sur leur rôle physiologique qui est parfois encore débattu.

a) *Les pointes Vertex* [Cim97]. Comme l'indique la Figure 5.2, ce sont des impulsions négatives qui surviennent en stade 1, durant la phase d'endormissement, et dont l'amplitude augmente avec l'approfondissement du sommeil. Ces phénomènes phasiques peuvent apparaître en réponse à des stimuli externes, où se manifester spontanément.

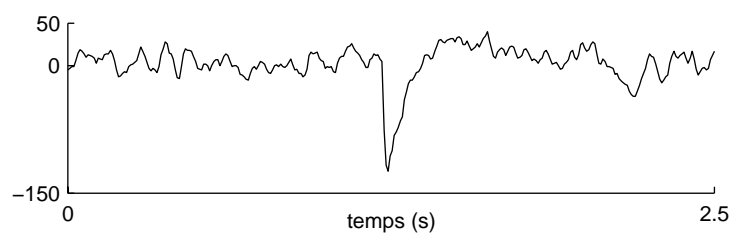


Figure 5.2. Exemple de pointe Vertex

b) *Les fuseaux de sommeil* [Cim97]. Ces événements constituent l'un des critères de définition du stade 2. Ils apparaissent sur l'électroencéphalogramme sous la forme de signaux transitoires quasi-sinusoïdaux, de fréquence comprise entre 12 et 14 Hz, et de durée variant entre 0.5 et 1 seconde chez l'adulte. Un exemple est présenté en Figure 5.3. On les associe volontiers à un mécanisme neurologique protégeant l'organisme des perturbations externes du sommeil, bien que cette fonction ne fasse pas toujours l'unanimité.

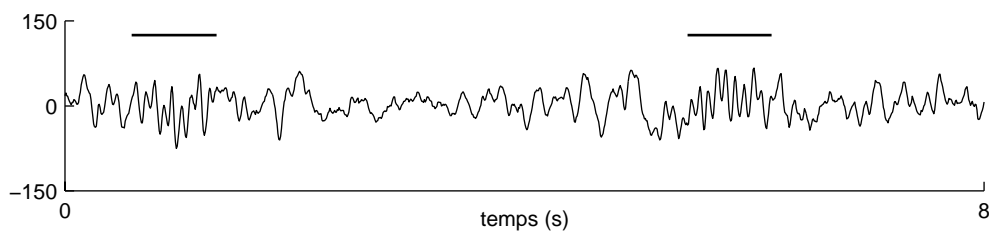


Figure 5.3. Exemples de fuseaux de sommeil.

c) *Le complexe K* [Cim97]. Le complexe K constitue, avec les fuseaux de sommeil, l'un des principaux marqueurs du début de sommeil puisqu'il apparaît dès le stade 2. Il est généralement d'aspect biphasique, comprenant une onde positive rapide suivie d'une onde négative de grande amplitude, comme on peut le constater sur la Figure 5.4. Il dure environ 1

seconde, présente un support fréquentiel compris entre 1 et 4 Hz, et se distingue par son amplitude de l'électroencéphalogramme de fond du stade 2, comme l'indique la Figure 5.1. Cependant, en raison d'une similitude marquée avec d'autres phénomènes non stationnaires observés en sommeil profond, tels que les bouffées d'ondes delta présentées en Figures 5.1 et 5.5, il reste très difficile à isoler en stades 3 et 4. Ceci justifie l'intérêt de méthodes statistiques pour tenter d'apporter une réponse satisfaisante à ce problème de détection.

5.3 Synthèse d'un détecteur de complexes K

5.3.1 Description des données

L'acquisition des électroencéphalogrammes utilisés dans le cadre de cette étude a été effectuée sur la dérivation dite Cz, à une fréquence d'échantillonnage de 128 Hz. Un ensemble de complexes K a d'abord été constitué par un groupe de 5 experts. Ces électrophysiologistes confirmés, qui ont effectué visuellement la cotation des signaux, ont travaillé individuellement sur trois enregistrements de huit heures environ. Du fait d'une certaine variabilité des résultats obtenus par ces cotateurs, on n'a retenu pour cette étude que les événements sélectionnés par la majorité d'entre eux, afin de constituer un ensemble de référence pour le complexe K. Puis, une base de phénomènes présentant des similitudes marquées avec ce dernier (bouffées d'ondes delta, etc.) a été constituée. Tous ces événements transitoires ont alors été synchronisés avant d'être segmentés sur des durées de 2 secondes. Finalement, les données ont été décimées d'un facteur 4, une fréquence d'échantillonnage de 32 Hz étant compatible avec les supports fréquentiels du complexe K, des ondes rapides qui peuvent se superposer à lui, ainsi que des bouffées d'ondes delta.

De cette façon, une base de données composée de 609 complexes K et de 1196 événements de nature différente, que l'on désignera désormais de façon impropre par ondes delta, a été constituée pour la présente étude. Quelques exemples issus de cet ensemble d'apprentissage sont représentés en Figures 5.4 et 5.5.

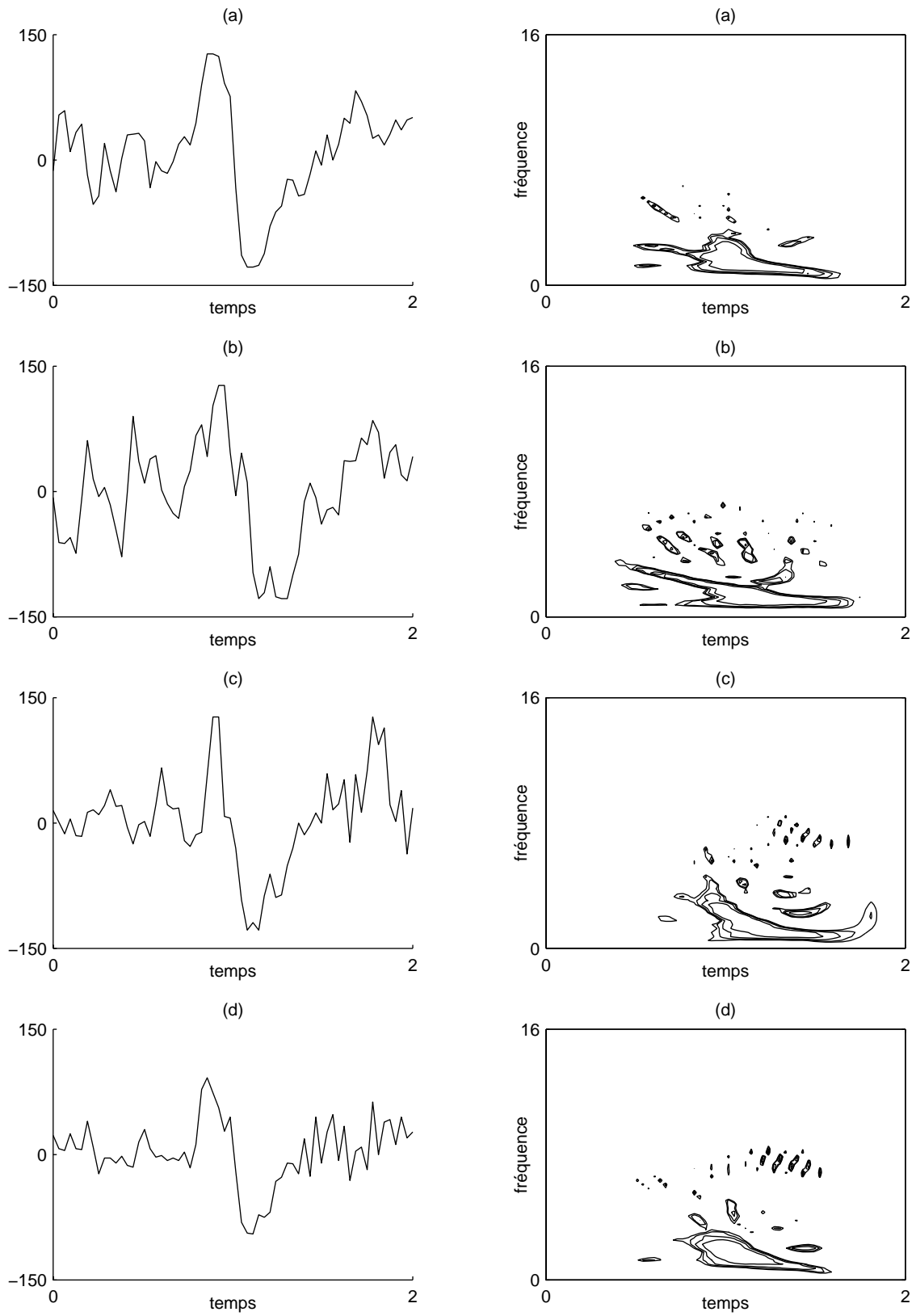


Figure 5.4. Représentations de complexes K dans les domaines temporel et temps-fréquence (Wigner-Ville).

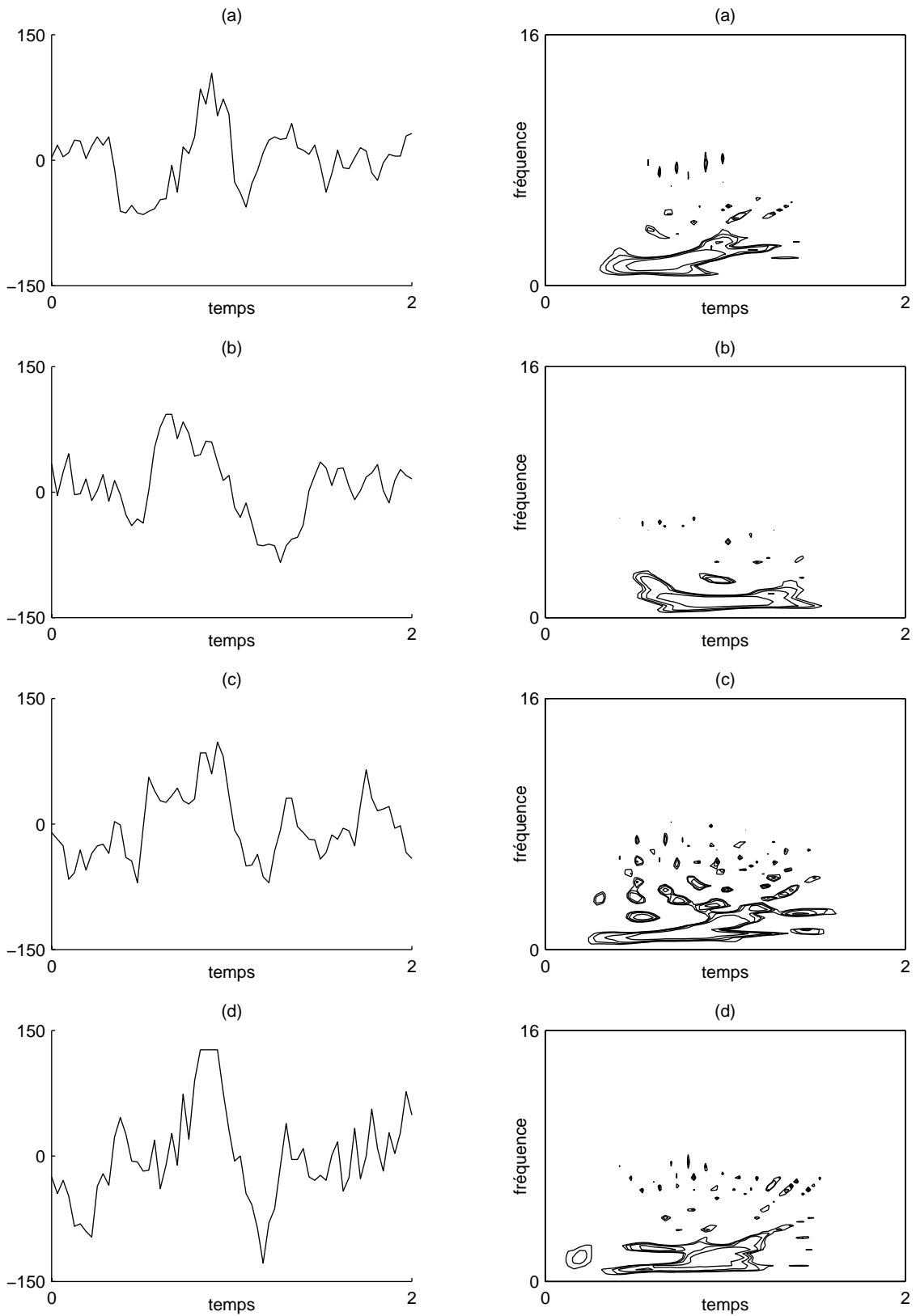


Figure 5.5. Représentations d'ondes delta dans les domaines temporel et temps-fréquence (Wigner-Ville).

5.3.2 Résultats et discussion

Un certain nombre de travaux sur la non-stationnarité de l'électroencéphalogramme de sommeil justifie l'intérêt des représentations temps-fréquence de la classe de Cohen, pour la caractérisation des phénomènes transitoires rencontrés [Cho87], [Sha93], [Bra96], [Cim97]. Aussi, on a jugé intéressant de recourir aux composantes $W_x[n, m]$ de la représentation de Wigner-Ville de l'observation X , pour l'élaboration d'un détecteur linéaire généralisé de complexe K . A ces données de nature quadratique, on a couplé les échantillons $X[n]$ de sorte que toute possibilité de converger vers une solution de type filtre adapté soit préservée. Ainsi, la classe de détecteurs qui a été considérée s'écrit :

$$\mathcal{D} = \left\{ d : \Lambda(X) = (a_{TF})^T W_X + (a_T)^T X - \lambda_0 \begin{array}{c} d(X)=1 \\ > \\ < \\ d(X)=0 \end{array} 0 \right\}.$$

La VC-dimension de \mathcal{D} étant déjà conséquente par rapport au nombre de données disponibles, on n'a pas jugé opportun pour le moment de faire appel à des statistiques de détection polynomiales de degré plus élevé.

L'optimisation du test de détection a été effectuée grâce à la méthode du critère optimal, présentée en Section 3.3.4, associée à la procédure de partitionnement faisant l'objet de la Section 4.2.2. Cependant, les approches basées sur les principes variationnel, de pénalisation ou consistant à lisser les représentations temps-fréquence, ayant donné de bons résultats en simulation, on envisage de les tester ultérieurement sur ce problème. La partition $\{A_1, \dots, A_r\}$ du plan temps-fréquence a été optimisée sur la base de 300 complexes K et de 600 ondes delta, en suivant le mode opératoire décrit par la Figure 4.12. Le meilleur détecteur, au sens d'une estimation de la probabilité d'erreur, a été obtenu pour une partition composée 30 régions, comme le montre la Figure 5.6. Sur la base de cette partition, on a alors utilisé les 309 complexes K et les 596 ondes delta restants afin d'enchaîner 250 phases d'apprentissage, grâce à la méthode du critère optimal, et de test. Ainsi, 220 complexes K et 500 ondes delta ont été tirés aléatoirement pour la synthèse du détecteur, à chaque itération, les autres échantillons étant destinés à l'estimation de la courbe COR. La Figure 5.7 présente les contributions relatives des composantes temporelle $(a_T)^T X$ et temps-fréquence $(a_{TF})^T W_X$ dans la statistique $\Lambda(X)$, par le biais de la projection des données sur le plan défini par les vecteurs a_T et a_{TF} . Cette représentation illustre explicitement l'intérêt de combiner les informations de nature linéaire et quadratique pour la résolution de ce problème.

Dans un premier temps, l'efficacité de l'approche présentée a été évaluée en comparant les résultats obtenus à ceux d'un détecteur opérant dans le domaine temporel, puis à ceux d'un détecteur opérant dans le plan temps-fréquence, dont la complexité avait été préalablement

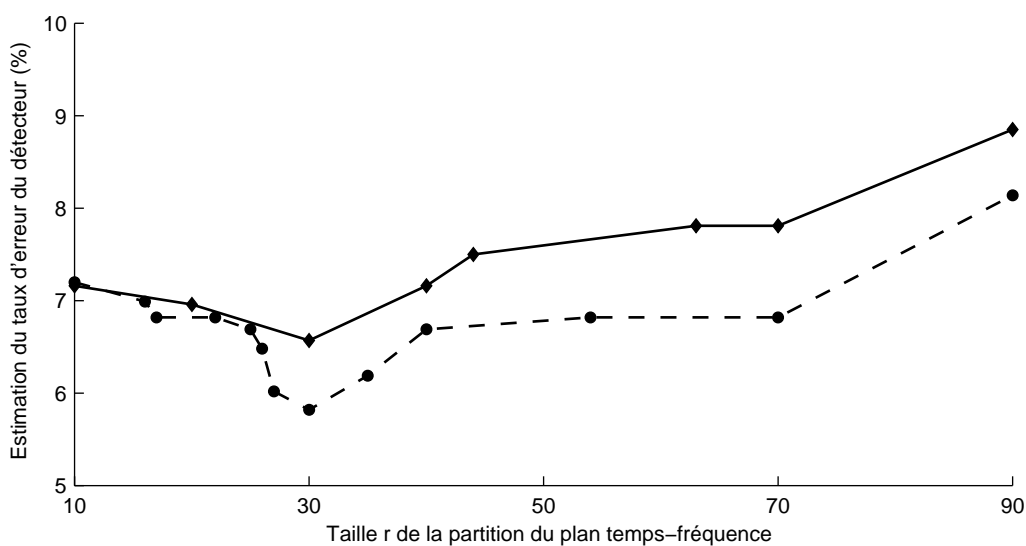


Figure 5.6. Estimation de l'erreur de généralisation en fonction de la taille r de la partition du plan temps-fréquence, pour les détecteurs opérant dans le domaine temps-fréquence (trait continu) et conjointement dans les domaines temporel et temps-fréquence (trait discontinu).

optimisée. Pour ce dernier, la taille de la partition a ainsi été fixée à 30, conformément aux résultats présentés en Figure 5.6. Les courbes COR moyennes résultant des 250 apprentissages décrits précédemment, représentées en Figure 5.8, indiquent les pertes de performances dues à l'usage de statistiques de détection purement linéaires ou quadratiques pour résoudre ce problème. Dans un second temps, les performances du détecteur ont été comparées à celles de quelques solutions déjà existantes. Comme cela est montré dans le Tableau 5.1, les meilleurs résultats ont été rapportés dans [Ban92] : un réseau de neurones a permis d'obtenir 90% de bonnes détections pour 8.1% de fausses alarmes, à comparer aux 4% de fausses alarmes que l'on a obtenu pour le même niveau de sensibilité. Tout en demeurant prudent pour la comparaison de ces résultats, qui dépendent fortement de la constitution de la base de données, on peut affirmer que la solution à laquelle on a abouti compte parmi les plus performantes.

	Approche	Bonnes détections (%)	Fausses alarmes (%)
Da Rosa et col. [Dar91]	Modélisation	89	49
Destiné et col. [Des94]	Réseaux de neurones	64 et 76	pour 4 et 12, resp.
Bankman et col. [Ban92]	Réseaux de neurones	90 et 95	pour 8.1 et 14.1, resp.
La présente étude [Ric98d]	Structure imposée	90 et 95	pour 4 et 7.5, resp.

Tableau 5.1. Performances de quelques détecteurs de complexes K existants.

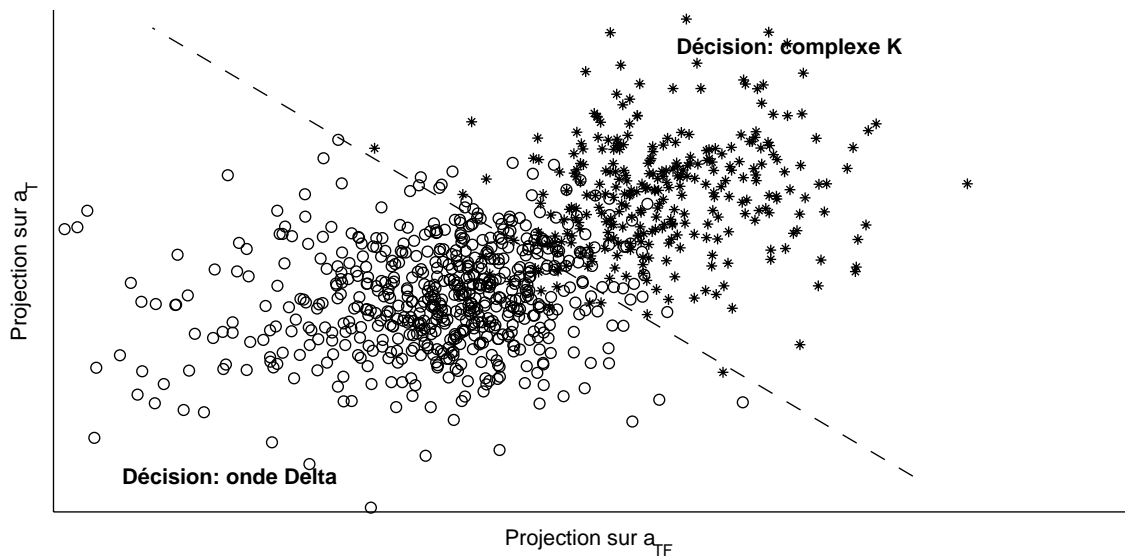


Figure 5.7. Projection des données sur le plan défini par les vecteurs a_T et a_{TF} . Les complexes K et les ondes delta sont respectivement matérialisés par * et O.

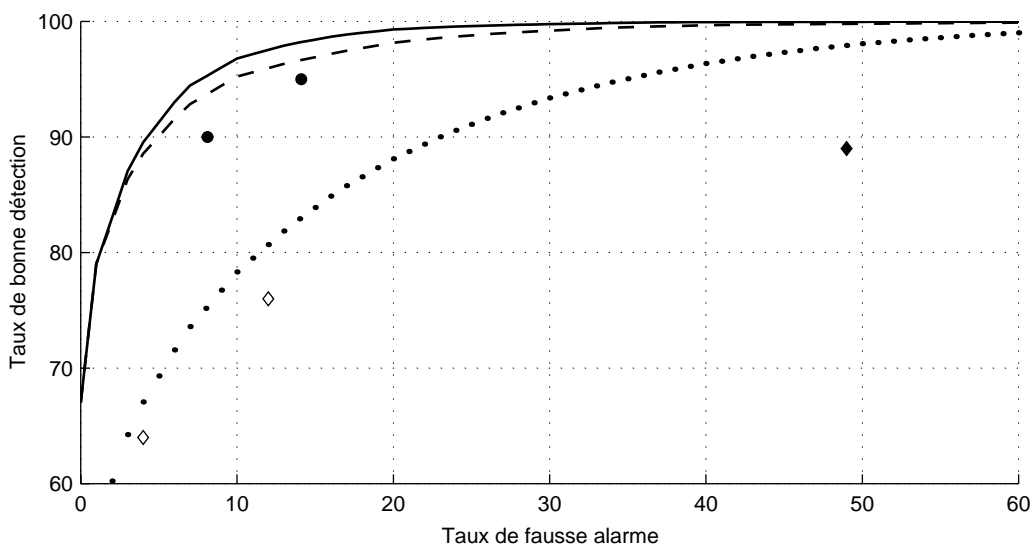


Figure 5.8. Détecteurs de complexes K. Courbes COR obtenues respectivement pour les détecteurs opérant dans les domaines temporel (trait pointillé), temps-fréquence (trait discontinu), et conjointement dans les deux (trait continu). Performances des détecteurs de Da Rosa (◆), Destiné (◇) et Bankman (●).

5.3.3 Conclusion

L'objet de ce chapitre était de valider, sur un problème réel réputé difficile, certains des principes présentés dans ce document. Il résulte ainsi de la présente étude sur la détection automatique du complexe K dans l'électroencéphalogramme de sommeil que la solution proposée compte parmi les plus performantes actuellement rapportées dans la littérature. De

plus, il semble que ces résultats puissent encore faire l'objet d'améliorations significatives en combinant les informations provenant de plusieurs dérivations, démarche généralement adoptée par les électrophysiologistes. Enfin, l'un des avantages de la méthodologie présentée réside dans le fait qu'elle est immédiatement transposable à tout autre problème de détection de phénomènes transitoires.

Conclusion générale

La définition d'une méthodologie pour la synthèse de détecteurs à structure imposée, avec application à la détection par représentation temps-fréquence, a motivé l'écriture de ce mémoire. Aussi, avant de clore celui-ci, nous allons en rappeler les principaux résultats en nous attachant à dégager quelques perspectives de recherche.

Dans un premier temps, nous avons présenté la notion de représentation temps-fréquence en nous intéressant principalement aux distributions d'énergie de la classe de Cohen. Nous avons alors évoqué certains problèmes pratiques liés à leur usage, la présence de termes interférentiels nuisant à l'interprétabilité des représentations en étant un exemple. Parmi les autres difficultés qui peuvent se présenter, nous nous sommes intéressés en particulier au volume de calculs prohibitif qu'induit l'analyse de signaux longs. Ainsi, après avoir mis l'accent sur le potentiel des algorithmes récursifs pour le calcul rapide des représentations temps-fréquence, nous avons présenté de façon homogène et unifiée les propriétés de récursivité directe et indirecte de la transformée de Fourier à court terme et de certains éléments de la classe de Cohen. Enfin, nous les avons étendues aux distributions modifiées par la méthode de réallocation.

Puis, nous avons procédé aux premiers choix stratégiques concernant la définition d'une méthodologie pour l'élaboration de détecteurs à structure imposée. On rappelle que cette approche consiste à sélectionner le meilleur test de détection d'un ensemble \mathcal{D} donné, au sens d'un critère c à définir. Pour \mathcal{D} , notre choix s'est porté sur la classe des détecteurs polynomiaux définis par la Proposition 3.4, principalement pour sa propriété de consistance universellement forte, qui constitue une garantie de performance. En ce qui concerne c , nous avons tout d'abord montré l'intérêt théorique que présente la classe \mathcal{O} des critères de contraste ne dépendant que des moments d'ordre un et deux de la statistique de détection. Nous avons cependant été amené à constater que la qualité de chacun de ces critères dépend du problème traité. La méthode présentée dans le cadre de ce chapitre a permis d'apporter une réponse à ce problème, en conduisant à la détermination analytique du meilleur critère de \mathcal{O} pour le problème traité, c'est à dire celui pour lequel le détecteur obtenu a une probabilité

d'erreur minimale. Finalement, le plan temps-fréquence, qui constitue un espace privilégié pour l'analyse de la configuration de certains détecteurs, a permis d'illustrer la validité des options adoptées.

Afin de limiter les effets néfastes de la *malédiction de la dimensionnalité*, nous avons recherché une adéquation entre la richesse de la classe \mathcal{D} , formellement définie par la notion de VC-dimension empruntée à la théorie de l'apprentissage, et le nombre de données disponibles pour l'élaboration d'un test de détection. Nous avons ainsi proposé trois méthodes, dont le mode opératoire consiste en un contrôle de la dimension de l'espace \mathcal{X} des observations. Transposé au cadre de la détection linéaire dans le plan temps-fréquence, ces procédures ont permis de montrer qu'un choix approprié de la représentation $C_x[n, m; \phi]$ conditionne significativement les performances susceptibles d'être atteintes par un détecteur. C'est là une illustration originale des potentialités de la classe de Cohen pour la résolution de certains problèmes de détection à structure imposée, qui confère à la représentation un rôle décisif dans le processus décisionnel. Ce concept étant novateur, nous pensons qu'il peut être intéressant de poursuivre les investigations dans cette voie, en caractérisant notamment les effets de la fonction de paramétrisation ϕ d'une représentation donnée sur la dimension de l'espace \mathcal{X} des réalisations. On pourra alors envisager une formulation temps-fréquence du principe SRM, ne nécessitant plus le recours à des procédures externes telles que celles qui ont été évoquées ci-dessus.

En toute logique, nous avons débuté cette étude par la distribution de Wigner-Ville sans pouvoir, faute de temps, l'étendre à d'autres distributions. Néanmoins, nous avons montré, contre toute attente, que la dimension de l'espace engendré par la représentation de Wigner-Ville est inférieure au nombre de composantes $W_x[n, m]$ qui la constituent, que le signal analysé soit réel, complexe ou analytique. A titre d'illustration, cette propriété a été mise en oeuvre pour la reconstruction complète de la représentation de Wigner-Ville, à partir d'une connaissance partielle, mais non quelconque, de celle-ci. Mises à part les perspectives offertes en termes d'algorithmes plus rapides et moins coûteux en mémoire, il semble que cette propriété puisse également être utilisée à des fins de débruitage.

Finalement, la méthodologie présentée dans ce mémoire a été validée sur le problème réputé difficile de la détection du complexe K dans l'électroencéphalogramme du sommeil. Il résulte de cette étude que la solution proposée compte parmi les plus performantes décrites dans la littérature. De plus, ces résultats sont susceptibles de pouvoir être améliorés en combinant les informations de plusieurs dérivations de l'électroencéphalogramme.

Bibliographie

- [Ami87] M. AMIN. A New Approach to Recursive Fourier Transform. *Proceedings of the IEEE*, 1987, vol. 75, n° 11, p. 1537-1538.
- [Ami95] M. G. AMIN. The Running Time-Frequency Distributions. *Circuits, Systems and Signal Processing*, 1995, vol. 14, n° 3, p. 401-414.
- [Ami96] M. G. AMIN. Recursive Kernels for Time-Frequency Signal Representations. *IEEE Signal Processing Letters*, 1996, vol. 3, n° 1, p. 16-18.
- [Aug91] F. AUGER. *Représentation Temps-Fréquence des Signaux Non-Stationnaires : Synthèse et Contribution*. - 221 p. Th. : ECN, Nantes, 1991.
- [Aug95] F. AUGER et P. FLANDRIN. Improving the Readability of Time-Frequency and Time-Scale Representations by the Reassignment Method. *IEEE Transactions on Signal Processing*, 1995, vol. 43, n° 5, p. 1068-1089.
- [Bac97] T. BACK, U. HAMMEL et H. P. SCHWEFEL. Evolutionary Computation : Comments on the History and Current State. *IEEE Transactions on Evolutionary Computation*, 1997, vol. 1, n° 1, p. 1-14.
- [Bal81] R. BALIAN. Un Principe d'Incertitude Fort en Théorie du Signal ou en Mécanique Quantique. *C. R. Acad. Sc. Paris II*, 1981, vol. 292, n° 20, p. 1357-1362.
- [Ban92] I. N. BANKMAN, V. G. SIGILLITO, R. A. WISE et P. L. SMITH. Feature-Based Detection of K-Complex Wave in the Human Electroencephalogram using Neural Networks. *IEEE Transactions on Biomedical Engineering*, 1992, vol. 39, n° 12, p. 1605-1610.
- [Bar91] R. G. BARANIUK et D. L. JONES. A Radially-Gaussian, Signal-Dependent Time-Frequency Representation. *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 1991, p. 3181-3184.

- [Bar93a] R. G. BARANIUK et D. L. JONES. Signal-Dependent Time-Frequency Analysis using a Radially Gaussian Kernel. *Signal Processing*, 1993, vol. 32, p. 263-284.
- [Bar93b] R. G. BARANIUK et D. L. JONES. A Signal-Dependent Time-Frequency Representation : Optimal Kernel Design. *IEEE Transactions on Signal Processing*, 1993, vol. 41, p. 1589-1602.
- [Bar92] D. T. BARRY. Fast Calculation of the Choi-Williams Time-Frequency Distribution. *IEEE Transactions on Signal Processing*, 1992, vol. 40, n° 2, p. 450-455.
- [Bas80] M. J. BASTIAANS. Gabor's Expansion of a Signal into Gaussian Elementary Signals. *Proceedings of the IEEE*, 1980, vol. 68, n° 4, p. 538-539.
- [Boa87] B. BOASHASH et P. J. BLACK. An Efficient Real-time Implementation of the Wigner-Ville Distribution. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1987, vol. 35, n° 11, p. 1611-1619.
- [Bot97] L. BOTTOU. La Mise en Oeuvre des Idées de Vladimir N. Vapnik. In S. CANU, O. GASCUEL, Y. LECHEVALLIER et S. THIRIA (éd.). *Statistique et Méthodes Neuronales*. Paris : Dunod, 1997, p. 262-274.
- [Bra96] J. C. BRAUN, D. F. HANLEY et N. V. THAKOR. Detection of Neurological Injury Using Time-Frequency Analysis of the Somatosensory Evoked Potentials. *Electroencephalography and Clinical Neurophysiology*, 1996, vol. 100, p. 310-318.
- [Cha97] E. CHASSANDE-MOTTIN, F. AUGER, I. DAUBECHIES et P. FLANDRIN. Partition du Plan Temps-Fréquence et Réallocation. *Proc. Seizième Colloque GRETSI*, 1997, p. 1447-1450.
- [Cho87] H. I. CHOI, W. J. WILLIAMS et H. ZAVERY. Analysis of Event Related Potentials: Time-Frequency Energy Distributions. *Biomedical Sciences Instrumentation*, 1987, vol. 23, p. 251-258.
- [Cim97] A. CIMETIERE. *Méthodes Temporelles et Temps-Fréquence pour la Reconnaissance Automatique des Complexes K de l'EEG de Sommeil*. - 152 p. Th. : UTC, Compiègne, 1997.
- [Coh66] L. COHEN. Generalized Phase Space Distribution Functions. *Journal of Mathematical Physics*, 1966, vol. 7, p. 781-786.

- [Coh95] L. COHEN. *Time-Frequency Analysis*. Englewood Cliffs (NJ, USA) : Prentice Hall, 1995.
- [Cov65] T. M. COVER. Geometrical and Statistical Properties of Systems of Linear Inequalities with Applications in Pattern Recognition. *IEEE Trans. Electron. Comp.*, vol. 10, 1965, p. 326-334.
- [Cun92] G. S. CUNNINGHAM et W. J. WILLIAMS. Fast Computation of the Wigner Distribution for Finite Length Signal. *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 1992, p. 177-180.
- [Cun94a] G. S. CUNNINGHAM et W. J. WILLIAMS. Fast Implementations of Generalized Discrete Time-Frequency Distributions. *IEEE Transactions on Signal Processing*, 1994, vol. 42, n° 6, p. 1496-1508.
- [Cun94b] G. S. CUNNINGHAM et W. J. WILLIAMS. Kernel Decomposition of Time-Frequency Distributions. *IEEE Transactions on Signal Processing*, 1994, vol. 42, n° 6, p. 1425-1442.
- [Dar91] A. C. DA ROSA, B. KEMP et T. PAIVA. A Model-based Detector of Vertex Sharp Waves and K-Complexes in Sleep Electroencephalogram. *Electroencephalography and Clinical Neurophysiology*, 1991, vol. 78, p. 71-79.
- [Des94] J. DESTINE, B. BECKERS, M. FOMBELLIDA, R. POIRRIER, D. DIVE et G. FRANCK. Utilisation des Réseaux de Neurones Artificiels pour la Reconnaissance de Grapho-Elements Phasiques dans le Cadre de l'Analyse du Sommeil. *Proc. Symposium International 25 ans d'Analyse Automatique du Sommeil (Genève)*, 1994.
- [Dev96] L. DEVROYE, L. GYORFI et G. LUGOSI. *A Probabilistic Theory of Pattern Recognition*. New York : Springer Verlag, 1996.
- [Don95] R. N. MC DONOUGH et A. D. WHALEN. *Detection of Signals in Noise*. London : Academic Press, 1995.
- [Dot75] H. DO-TU et M. INSTALLE. On Adaptive Solution of Piecewise linear Approximation Problem. *Proc. Milwaukee Symposium on Automatic Computation and Control*, 1975, p. 1-6.
- [Dub90] B. DUBUISSON. *Diagnostic et Reconnaissance des Formes*. Paris : Hermès, 1990.

- [Dud73] R. O. DUDA et P. E. HART. *Pattern Classification and Scene Analysis*. A Wiley-Interscience Publication, 1973.
- [Duv91] P. DUVAUT. *Traitement du Signal - Concepts et Applications*. Paris : Hermès, 1991.
- [Ehr88] A. EHRENFUCHT, D. HAUSSLER, M. KEARNS et G. VALIANT. A General Lower Bound on the Number of Examples needed for Learning. *Proc. Workshop on Computational Learning Theory*, 1988, p. 42-55.
- [Fla88] P. FLANDRIN. A Time-Frequency Formulation of Optimum Detection. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1988, vol. 36, n° 9, p. 1377-1384.
- [Fla93] P. FLANDRIN. *Temps-Fréquence*. Paris : Hermès, 1993.
- [Fri76] J. FRITZ et L. GYORFI. On the Minimization of Classification Error Probability in Statistical Pattern Recognition. *Problems of Control and Information Theory*, 1976, vol. 5, p. 371-382.
- [Fuk90] K. FUKUNAGA. *Introduction to Statistical Pattern Recognition*. London : Academic Press, 1990.
- [Gab46] D. GABOR. Theory of Communication. *Journal of the IEE*, 1946, vol. 93, p. 429-457.
- [Gai73] J. M. GAILLARD et R. TISSOT. Principles of Automatic Analysis of Sleep Records with Hybrid System. *Computers and Biomedical Research*, 1973, vol. 6, p. 1-13.
- [Gar80] W. A. GARDNER. A Unifying View of Second Order Measures of Quality for Signal Classification. *IEEE Transactions on Communication*, 1980, vol. 28, p. 807-816.
- [Gas97] O. GASCUEL. La Dimension de Vapnik-Chervonenkis - Application au Réseaux de Neurones. In S. CANU, O. GASCUEL, Y. LECHEVALLIER et S. THIRIA (éd.). *Statistique et Méthodes Neuronales*. Paris : Dunod, 1997, p. 244-261.
- [Gol93] G. H. GOLUB et C. F. VAN LOAN. *Matrix Computations*. London : The Johns Hopkins University Press, 1993.
- [Gon98] P. GONCALVEZ et E. PAYOT. Adaptive Diffusion Equation for Time-Frequency Representations. *Proc. IEEE Digital Signal Processing Workshop*, Bryce Canyon, Utah, USA, 1998.

- [Her94] J. HERAULT et C. JUTTEN. *Réseaux Neuronaux et Traitement du Signal*. Paris : Hermès, 1994.
- [Hla97] F. HLAWATSCH et P. FLANDRIN. The Interference Structure of the Wigner Distribution and Related Time-Frequency Signal Representations. In W. Mecklenbrauker et F. Hlawatsch (éd.). *The Wigner Distribution - Theory and Applications in Signal Processing*. Amsterdam : Elsevier, 1997.
- [Hos80] G. H. HOSTETTER. Recursive Discrete Fourier Transformation. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1980, vol. 28, n° 2, p. 184-190.
- [Joh78] D. JOHNSON et F. PREPARATA. The Densest Hemisphere Problem. *Theoretical Computer Science*, 1978, vol. 6, p. 93-107.
- [Jon90] D. L. JONES et T. W. PARKS. A High Resolution Data Adaptive Time-Frequency Representation : Optimal Kernel Design. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1990, vol. 38, p. 2127-2135.
- [Jon95a] D. L. JONES et R. G. BARANIUK. An Adaptive Optimal Kernel Time-Frequency Representation. *IEEE Transactions on Signal Processing*, 1995, vol. 43, n° 10, p. 2361-2371.
- [Jon95b] D. L. JONES et A. M. SAYEED. Blind Quadratic and Time-Frequency Based Detectors from Training Data. *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 1995, p. 1033-1036.
- [Kod78] K. KODERA, R. GENDRIN et C. De VILLEDARY. Analysis of Time-Varying Signals with Small BT Values. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1978, vol. 34, n° 1, p. 64-76.
- [Koz92] W. KOZEK. Time-Frequency Signal Processing based on the Wigner-Weyl Framework. *Signal Processing*, 1992, vol. 29, n° 1, p. 77-92.
- [Kro97] B. S. KRONGOLD, A. M. SAYEED, M. MOEHRING, J. A. RITCEY, M. SPENCER et D. L. JONES. Quadratic Time-Scale Detection of Microemboli in Flowing Blood with Doppler Ultrasound. *Proc. SPIE's 42nd Meeting*, 1997.
- [Kun81] M. KUNT. *Traitement Numérique des Signaux*. Paris : Bordas, 1981.

- [Lem95] O. LEMOINE. *Détection de Signaux Non Stationnaires par Représentation Temps-Fréquence*. - 194 p. Th. : Université de Nice-Sophia Antipolis, 1995.
- [Loo35] A. L. LOOMIS, E. N. HARVEY et G. A. HOBART. Potentials Rhythms of the Cerebral Cortex during Sleep. *Science*, 1935, vol. 81, p. 597-598.
- [Mar85] W. MARTIN et P. FLANDRIN. Wigner-Ville Spectral Analysis of Nonstationary Processes. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1985, vol. 33, n° 6, p. 1461-1469.
- [Moo92] J. E. MOODY. Generalization, Weight Decay and Architecture Selection for Non-Linear Learning Systems. In J. E. MOODY, S. J. HANSON et R. P. LIPPMANN (éd.). *Advances in Neural Information Processing 4*. San Mateo : Morgan Kaufmann, 1992.
- [Nor63] D. NORTH. An Analysis of the Factors which Determine Signal/Noise Discrimination in Pulsed Carrier Systems. *Proceedings of the IEEE*, 1963.
- [Oeh97] H. OEHLMANN et D. BRIE. Distribution de Wigner-Ville Locale pour la Réduction des Interférences. *Proc. Seizième Colloque GRETSI*, 1997, p. 667-670.
- [Pic88] B. PICINBONO et P. DUVAUT. Optimal Linear-Quadratic Systems for Detection and Estimation. *IEEE Transactions on Information Theory*, 1988, vol. 43, p. 304-311.
- [Poo94] H. V. POOR. *An Introduction to Signal Detection in Noise*. New-York : Springer-Verlag, 1994.
- [Ray86] S. RAY, W. D. LEE, C. D. MORGAN et W. AIRTH-KINDREE. Computer Sleep Stage Scoring. An expert System Approach. *International Journal of Bio-Medical Computing*, 1986, vol. 19.
- [Rey87] C. REYNOLDS, C. F. KUPFER, D. J. HOUCK, C. C. STACK, S. R. BERMAN et B. ZIMER. Sleep Research in Affective Illness: State of the Art. *Sleep*, 1987, vol. 10.
- [Ric96a] C. RICHARD et R. LENGELLE. Fast Implementation of Time-Frequency Representations Modified by the Reassignment Method. *Proc. International Conference on Signal Processing (Beijing)*, 1996, p. 343-346.

- [Ric96b] C. RICHARD et R. LENGELLE. Recursive Implementation of some Time-Frequency Representations. *Proc. IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, 1996, p. 313-316.
- [Ric97a] C. RICHARD et R. LENGELLE. Joint Recursive Implementation of Time-Frequency Representations and their Modified Version by the Reassignment Method. *Signal Processing*, 1997, vol. 60, n° 2, p. 163-179.
- [Ric97b] C. RICHARD et R. LENGELLE. Une Nouvelle Approche pour la Détection Linéaire Optimale dans le Plan Temps-Fréquence. *Proc. Seizième Colloque GRETSI*, 1997, p. 659-662.
- [Ric97c] C. RICHARD et R. LENGELLE. Joint Time and Time-Frequency Optimal Detection. *Proc. IEEE UK Symposium on Applications of Time-Frequency and Time-Scale Methods*, 1997, p. 29-32.
- [Ric98a] C. RICHARD et R. LENGELLE. Structural Risk Minimization for Reduced-Bias Time-Frequency-Based Detectors Design. *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1998, p. 2397-2400.
- [Ric98b] C. RICHARD et R. LENGELLE. Two Algorithms for Designing Optimal Reduced-bias Data-Driven Time-Frequency Detectors. *Proc. IEEE International Symposium on Time-Frequency and Time-Scale Analysis*, 1998, p. 601-604.
- [Ric98c] C. RICHARD et R. LENGELLE. On the Dimension of the Discrete Wigner-Ville Transform Range Space. Application to Time-Frequency Detectors Design. *Proc. IEEE International Symposium on Time-Frequency and Time-Scale Analysis*, 1998, p. 5-8.
- [Ric98d] C. RICHARD et R. LENGELLE. Joint Time and Time-Frequency Optimal Detection of K-complexes in Sleep EEG. *Computers and Biomedical Research*, 1998, vol. 31, n° 3, p. 209-229.
- [Ric98e] C. RICHARD et R. LENGELLE. Détection Automatique des Phénomènes Transitoires de l'EEG par Analyse Temps-Fréquence. *Innov. Techn. Biol. Med.*, 1998, vol. 19, n° 3, p. 167-177.
- [Ric99] C. RICHARD et R. LENGELLE. Data-driven design and complexity control of time-frequency detectors. A paraître dans *Signal Processing*, 1999.

- [Rou93] J. F. ROUX. Discrete Recursive Fourier Transform with a Non-Symmetrical Exponential Window. *Traitement du Signal*, 1993, vol. 10, n° 2, p. 185-192.
- [Sam96] B. SAMIMY, G. RIZZONI, A. M. SAYEED et D. L. JONES. Design of Training Data-Based Quadratic Detectors with Application to Mechanical Systems. *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 1996, p. 1767-1770.
- [Sap90] G. SAPORTA. *Probabilités, Analyse des Données et Statistique*. Paris : Technip, 1990.
- [Say95] A. M. SAYEED et D. L. JONES. Optimal Detection using Bilinear Time-Frequency and Time-Scale Representations. *IEEE Transactions on Signal Processing*, 1995, vol. 43, n° 12, p. 2872-2883.
- [Say96a] A. M. SAYEED et D. L. JONES. Time-Frequency Detectors, *Proc. CISS'96*, 1996.
- [Say96b] A. M. SAYEED et D. L. JONES. Optimal Reduced-Rank Time-Frequency / Time-Scale Quadratic Detectors. *Proc. IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, 1996, p. 209-212.
- [Sha93] N. SCHALTENBRAND, R. LENGELLE et J.-P. MACHER. Neural Network Model: Application to Automatic Analysis of Human Sleep. *Computers and Biomedical Research*, 1993, vol. 26, p. 157-171.
- [Sch91] L. L. SCHARF. *Statistical Signal Processing*. Reading (MA, USA) : Addison Wesley, 1991.
- [Sch96] M. SCHOENAUER, L. KALLEL et F. JOUVE. Mechanical Inclusions Identification by Evolutionary Computation. *Revue Européenne des Eléments Finis*, 1996, vol. 5, p. 619-648.
- [Skl79] J. SKLANSKY et G. WASSEL. *Pattern Classifiers and Trainable Machines*. New York : Springer-Verlag, 1979.
- [Sun85] M. SUN, C. C. LI, L. N. SEKHAR et R. J. SCLABASSI. Efficient Computation of the Discrete Pseudo-Wigner Distribution. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1985, vol. 37, n° 11, p. 1735-1742.
- [Tom96] S. TOMAZIC. On Short-Time Fourier Transform with Single-Sided Exponential Window. *Signal Processing*, 1996, vol. 55, p. 141-148.

- [Uns83] M. UNSER. Recursion in Short-Time Signal Analysis. *Signal Processing*, 1983, vol. 5, p. 229-240.
- [Tre68] H. L. VAN TREES. *Detection, Estimation, and Modulation - Part I*. New York : Wiley, 1968.
- [Vap71] V. N. VAPNIK et A. Y. CHERVONENKIS. On the Uniform Convergence of Relative Frequencies of Events to their Probabilities. *Theory of Probabilities and its Applications*, 1971, vol. 16, p. 264-280.
- [Vap82] V. N. VAPNIK. *Estimation of Dependences Based on Empirical Data*. New-York : Springer-Verlag, 1982.
- [Vil48] J. VILLE. Théorie et Application de la Notion de Signal Analytique. *Câbles et Transmissions*, 1948, vol. 2A, p. 61-74.
- [Was72] G. WASSEL et J. SKLANSKY. Training a one-dimensional Classifier to Minimize the Probability of Error. *IEEE Transactions on Systems, Man and Cybernetics*, 1972, vol. 2, p. 533-541.
- [Wid88] B. WIDROW et M. HOFF. Adaptive Switching Circuits. In J. A. ANDERSON et E. ROSENFELD (éd.). *Neurocomputing: Foundations of Research*. Cambridge (MA, USA) : MIT Press, 1988.
- [Wig32] E. P. WIGNER. On the Quantum Correction for Thermodynamic Equilibrium. *Physical Review*, 1932, vol. 40, p. 749-759.
- [Zie91] T. P. ZIELINSKI. Note on Computation of some Bilinear Time-Frequency Signal Representations. *Proc. International Conference on Digital Signal Processing*, 1991, p. 40-44.

RESUME :

Ce mémoire traite de la détection des signaux non-stationnaires lorsque les seules informations disponibles pour l'élaboration du détecteur consistent en une base de signaux étiquetés, auquel cas la structure de celui-ci doit être choisie a priori. La potentialité des approches temps-fréquence (TF) pour la résolution de tels problèmes est en particulier considérée. Les Chapitres 1 et 2 justifient l'intérêt de l'analyse TF pour l'étude des signaux non-stationnaires. Certains problèmes pratiques liés à l'usage des représentations de la classe de Cohen sont présentés et leur résolution évoquée. En particulier, afin de limiter le volume de calcul prohibitif qu'induit l'analyse des signaux longs, une stratégie récursive est proposée. Le Chapitre 3 est consacré à la détection à structure imposée des signaux non-stationnaires. On justifie théoriquement l'intérêt des détecteurs linéaires généralisés puis on présente un algorithme permettant d'en optimiser les paramètres. Celui-ci repose sur la recherche du meilleur critère de contraste pour le problème traité, au sens de la probabilité d'erreur. Afin d'en optimiser les performances, on vise dans le Chapitre 4 à adapter la complexité des détecteurs linéaires généralisés, notion définie en théorie de l'apprentissage par la dimension de Vapnik-Chervonenkis VC, à la taille de la base de signaux utilisée. En s'appuyant sur le principe de "Structural Risk Minimization" de Vapnik, on propose ainsi diverses méthodes d'optimisation de VC reposant sur un contrôle approprié de la dimension de l'espace de représentation des signaux. En particulier, dans le cadre de la détection par représentation TF, on démontre que l'opération de lissage peut remplir cette fonction puis que la dimension de l'espace engendré par la distribution de Wigner-Ville est plus faible qu'attendue. Enfin, le Chapitre 5 est consacré à l'application des méthodes proposées au problème de la détection du complexe K dans l'EEG de sommeil.

MOTS-CLES :

théorie de la détection
analyse temps-fréquence
distribution de Wigner-Ville
algorithme récursif

critère de contraste
théorie de l'apprentissage
dimension de Vapnik-Chervonenkis
complexe K