

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR
ET DE LA RECHERCHE SCIENTIFIQUE

----- oOo -----

Ecole Nationale Polytechnique - ENP
Département de Génie Mécanique
Laboratoire de Génie Mécanique et Développement - LGMD



THÈSE DE DOCTORAT EN GENIE MECANIQUE

Mohammed BOUFENAR
Magister en Génie Mécanique

**APPROCHE COMPARATIVE DES TECHNIQUES DE DETECTION
ET D'ANALYSE EN PRESENCE DES DEFAUTS CONJUGUES
DANS LES MACHINES TOURNANTES**

Jury

Mr. Arezki SMAILI	Professeur, ENP, Alger	Président
Mr. Saïd RECHAK	Professeur, ENP, Alger	Directeur de thèse
Mr. Nouredine ZERHOUNI	Professeur, ENSMM, Besançon, Fr.	Examineur
Mr. Mohamed ouali SI CHAIB	Professeur, UMB, Boumerdes	Examineur
Mr. Abdelhamid MILOUDI	Professeur, USTHB, Alger	Examineur
Mr. Djamel BOUKHETALA	Professeur, ENP, Alger	Examineur

الهدف من هذه الأطروحة هو المساهمة في حل مشكلة استعادة الإشارات وتحديد مصادر الاهتزاز لتشخيص الآلات أو المنشآت الصناعية المعقدة تم إدخال المقاربة المبنية على تحسين تمثيل إشارات الاهتزاز من نوع النبض لحل مشكلة نقص دقة طرق وأساليب التحليل الزمن التردد ثم قمنا بإدماجها كمصدر للمعلومات القبلية لحل المسائل العكسية. بالفعل إن الدمج الناجح بين نموذج المعلومات القبلية للتمثيل الفراغي من جهة و نموذج النظرية التقديرية لبايز مكننا من تحقيق الهدف المرجو وهذا بفضل الطبيعة اللاخطية لعامل التخفيض لبايز الذي يستفيد من المنشأة الإحصائية الخاصة بالإشارة المبحوث عنها . عموما فإن النهج المبين يجمع بين خصائص الحساب السريع والأداء النظري الجيد و النجاعة الحسنة في التطبيق الافتراضي و الحقيقي

كلمات البحث: التردد الوقت، تبعثر، نهج بايز، المعلومات القبلية المبعثرة، المسائل العكسية

Abstract

This thesis is a contribution to solving the problem of signals estimating and noisy and blind vibration sources signal separation for machines or complex industrial facilities diagnosis. The approach based on sparsity improving of impulsive vibration signals was introduced to solve the problem of lack of accuracy and resolution in time-frequency methods. Then we interested in its integration as a source of prior knowledge in inverse problems. Indeed, the successful fusion between the prior model for the sparse representation on the one hand, and the theory of Bayesian estimation, enabled us to achieve the objective. This is accomplished due to the highly nonlinear nature of the Bayesian attenuation operator which takes advantage of the inherent statistical structure of the signal to be identified. Overall, the described approach combines the properties of fast calculation, good theoretical properties and good performance in simulations and in practice.

Keywords: time-frequency, sparsity, Bayesian approach, sparse priors, inverse problems

Résumé

Ce travail de thèse constitue une contribution à la résolution du problème d'estimation des signaux et de séparation aveugle des sources vibratoires bruitées en vue du diagnostic des machines ou d'installations industrielles complexes. L'approche basée sur l'amélioration de la parcimonie des signaux vibratoires de type impulsions a été introduite pour la résolution du problème de manque de précision et de résolution dans des méthodes temps-fréquence. Ensuite, nous nous sommes intéressés à son intégration comme source d'a priori dans des problèmes inverses. En effet, le mariage fructueux entre le modèle a priori pour la représentation parcimonieuse d'une part, et de la théorie de l'estimation Bayésienne, nous a permis d'atteindre l'objectif fixé. Ceci est réalisé grâce à la nature fortement non-linéaire de l'opérateur d'atténuation Bayésien qui tire profit de la structure statistique inhérente du signal à identifier. Dans l'ensemble, l'approche décrite combine les propriétés de calcul rapide, de bonnes propriétés théoriques et de bonnes performances dans les simulations et dans la pratique.

Mots-clés: temps-fréquence, parcimonie, approche Bayésienne, a priori parcimonieux, problèmes inverses

REMERCIEMENTS

Je tiens à exprimer ma profonde gratitude et toute la reconnaissance que je dois au Professeur Saïd RECHAK, directeur de cette thèse, pour ses nombreux encouragements, son encadrement dynamique et toute l'aide qu'il m'a apportée durant toute la durée de cette thèse. Je suis particulièrement redevable à lui pour sa confiance et son soutien et pour l'aide indispensable à l'avancement de mes travaux de recherches.

J'adresse mes plus sincères remerciements au Professeur Arezki SMAILI pour m'avoir fait l'honneur d'examiner mon travail et de présider le jury de ma soutenance.

Je remercie très chaleureusement le professeur Noureddine ZERHOUNI pour l'honneur qu'il m'a fait d'accepter d'être examinateur de cette thèse et pour la discussion fructueuse que nous avons eue. Par ses questions et remarques constructives, il m'a été d'une aide précieuse et m'a permis d'améliorer de manière significative certaines parties de mon manuscrit.

Mes remerciements et ma profonde gratitude s'adressent aussi aux professeurs Djamel BOUKHETALA, Abdelhamid MILOUDI et au professeur Mohamed ouali SI CHAIB pour l'honneur qu'ils m'ont fait d'accepter de faire partie de mon jury de soutenance. Je leur suis très reconnaissant du temps qu'ils ont consacré à l'évaluation de ce travail.

Je suis très reconnaissant envers mes amis et collègues : Rachid ZAMOUNE et Hamid BOUCHERIT pour l'immensité du travail qu'ils ont effectué pour mettre en marche notre banc d'essai malgré toutes les difficultés techniques que nous avons eues. Je voudrais également saluer Mr. Mohamed REZIG, Directeur de la Division DRN pour son soutien et son encouragement durant la période de cette thèse.

Enfin merci à tout mes amis et à tout ceux qui ont contribué de près ou de loin à l'aboutissement de ce modeste travail.

Je termine par un grand merci à toute ma famille pour le soutien qu'elle m'a apporté tout au long de la préparation de cette thèse.

SOMMAIRE

Chapitre I : Introduction générale et état de l'art

I.1	Introduction générale.	09
I.2	Objectifs du présent travail.	11
I.3	Etat de l'art.	12
I.4	Organisation de la thèse.	17

Chapitre II : Synthèse des méthodes temps-fréquence pour l'analyse des signaux de vibrations

II.1	Introduction	21
II.2	Origine des analyses temps-fréquence	21
II.3	Atome analysant temps-fréquence	22
II.4	Théorème d'incertitude d'Heisenberg.	22
II.5	La transformée de Fourier à fenêtre glissante (STFT)	23
II.5.1	Adaptativité de la représentation temps-fréquence de type STFT.	28
II.5.2	Critères d'optimalité	28
II.5.3	Optimisation globale.	30
II.5.4	Optimisation locale dans le plan temps-fréquence.	30
II.6	La Distribution de Wigner-Ville.	31
II.6.1	Introduction.	31
II.6.2	Définition.	31
II.6.3	Propriétés.	32
II.6.4	Mécanisme des interférences.	33
II.6.5	Lissage de la distribution de Wigner Ville	34
II.7	Conclusion.	36

Chapitre III : Décomposition en ondelettes et analyse multi-résolution (AMR)

III.1	Introduction	39
III.2	Ondelette analysante (Atome temps-fréquence).	40
III.3	Conditions d'admissibilité	41
III.4	La transformée d'ondelettes continue	42
III.5	Formule d'inversion.	44
III.6	Base d'ondelettes.	44
III.7	Quelques propriétés.	45
III.8	Transformée en Ondelettes Discrète.	45
III.9	L'analyse multi-résolution.	47
III.9.1	Approximation filtrée dans V_j	48
III.9.2	Espace de détails de $L^2(\mathbb{R})$	48

III.9.3	Base orthonormée de W_j	49
III.9.4	Voie de détails dans W_j	49
III.9.5	Théorie des filtres miroir en quadrature et algorithmes rapides de Mallat	50
III.9.6	Filtre miroir conjugué.	51
III.9.7	Filtre d'échelle.	51
III.9.8	Filtres miroir en quadrature.	52
III.9.9	Filtre de détail.	52
III.9.10	Algorithmes rapides de Mallat.	53
III.10	Conclusion	56

Chapitre IV : Parcimonie et représentation des signaux impulsionnels de vibrations

IV.1	Introduction	60
IV.2	Représentation parcimonieuse	60
IV.3	Bases et dictionnaires	61
	IV.3.1 Analyse et synthèse	62
	IV.3.2 Synthèse et reconstruction exacte	63
IV.4	Mesure de la parcimonie des coefficients d'analyse	63
IV.5	Algorithmes de décomposition	66
	IV.5.1 Poursuite adaptative (Matching Pursuit MP)	66
	IV.5.2 Orthonormal Matching Pursuit (OMP)	67
	IV.5.3 Basis Pursuit (BP)	68
IV.6	Parcimonie et entropies	68
	IV.6.1 Entropie de Rényi	68
	IV.6.2 Entropie de Shannon	69
IV.7	Formulations de la parcimonie dans un espace redondant adapté	69
IV.8	Choix de l'ondelette analysante adaptée au signal vibratoire impulsionnel	70
IV.9	Optimisation de l'ondelette de Morlet pour la détection des impulsions	71
IV.10	Application de la parcimonie dans les problèmes inverses	74
IV.11	Parcimonie et restauration des signaux (débruitage)	76
IV.12	Conclusion	78

Chapitre V : Approche Bayésienne pour le débruitage des signaux de vibration fortement bruités

V.1	Introduction	81
V.2	Estimation non-paramétrique (débruitage)	82
V.3	Approche classique sans <i>a priori</i>	83
	V.3.1 Choix du seuil	87
	V.3.2 Estimation du niveau de bruit	88
V.4	Formulation Bayésienne du problème d'identification	89
	V.4.1 Choix de la fonction de vraisemblance.	90

V.4.2	Choix de la distribution de probabilité <i>a priori</i> de parcimonie	91
V.5	Etude expérimentale	93
V.5.1	Analyse des signaux de synthèse	93
V.5.2	Analyse des signaux réels	95
V.6	Conclusion	99

Chapitre VI : Séparation aveugle des sources dans le cas des mélanges bruités

VI.1	Introduction	103
VI.2	Modélisation mathématique des mélanges.	104
VI.3	Modèle de mélanges linéaires instantanés	105
VI.3.1	L'indétermination de permutation	106
VI.3.2	L'indétermination du facteur d'échelle	107
VI.4	Principe de résolution de la SAS	107
VI.4.1	Méthode de traitement par bloc	107
VI.4.2	Méthodes adaptatives	108
VI.5	Analyse en Composantes Indépendantes (ACI)	109
VI.5.1	Mesures d'indépendance et algorithmes	111
VI.5.2	Prétraitements	113
VI.6	Critères d'indépendance utilisés en ACI	113
VI.6.1	Information mutuelle (IM)	114
VI.6.2	Maximum de vraisemblance	116
VI.6.3	Non-gaussianité et algorithme FastICA	117
VI.7	Critères de performances de la SAS.	120
VI.7.1	Erreur quadratique moyenne (EQM)	120
VI.7.2	Rapport signal sur résidus (SNR)	120
VI.8	Séparation des sources dans des mélanges bruités	121
VI.8.1	Approche par analyse en composante parcimonieuses SCA.	122
VI.8.2	Approche bayésienne de séparation aveugle des sources	123
VI.8.3	Principe d'estimation bayésienne de séparation de sources	124
VI.8.4	Formulation bayésienne simplifié d'un problème de séparation sources	126
VI.8.5	Approche basée sur le pré et post-traitement des coefficients.	128
VI.8.6	Nouvelle approche basé sur le débruitage adaptif des coefficients.	130
VI.8.7	Connexion avec l'analyse en composantes indépendantes.	130
VI.9	Etude expérimentale	132
VI.10	Conclusion	139
	Conclusion Générale	141
	Bibliographie	143

Liste des symboles et abréviations

$\psi(t)$	Fonction analysante ou Ondelette mère
$\psi_{a,b}(t)$	Atome temps-fréquence dilaté-translaté
τ_ψ	Position moyenne temporelle de la fonction analysante ψ
ξ_ψ	Position moyenne fréquentielle de la fonction analysante ψ
C_ψ	Constante d'admissibilité
$\chi_\psi(a, b)$	Transformée en ondelette continue
$\sigma_{\tau,\psi}$	Variances temporelle liée à la fonction analysante ψ
$\sigma_{\xi,\psi}$	Variances fréquentielle liée à la fonction analysante ψ
ϕ	Fonction d'échelle
ω_n	Fréquence naturelle non-amortie (rd/sec)
$h(t)$	Fenêtre temporelle (fonction analysante)
$A_x(b, \omega)$	Fonction d'ambiguïté
$\mathbb{E}\{A_x\}$	Fonction d'ambiguïté moyenne
$\rho_\gamma(\tau, \omega)$	Spectrogramme normalisé
ω_d	Fréquence naturelle amortie (rd/sec)
ζ	Facteur d'amortissement
$\delta(t)$	Impulsion instantanée de Dirac
β	Facteur de forme
f	Fréquence ($\omega = 2\pi f$) (Hz)
$H(\gamma)$	Entropie de Shannon
$R_\alpha(\gamma)$	Entropie de Rényi
ε	Bruit additif Gaussien
$y(t)$	Signal d'observation
$W(\cdot)$	Transformée en ondelettes
$d_{j,k}$	Coefficients de détail dans la décomposition en d'ondelettes
$S_{j,k}$	Coefficients d'approximation
α_i	Coefficients de décomposition dans le domaine temps-fréquence
ϕ_i	Atome temps-fréquence
\mathcal{D}	Dictionnaire d'atomes temps-fréquence
λ_j	Seuils de débruitage à l'échelle j
λ_{univ}	Seuil universel
$s(t)$	Vecteur des sources indépendantes
A	Matrice de mélange
$b(t)$	Vecteur des signaux bruits
$p(s)$	Densité de probabilité
$J(s)$	Négentropie
W	Matrice de séparation
w_j	Eléments de la Matrice de séparation
$L^2(\mathbb{R})$	Espace de Hilbert des fonctions de carré sommable

$\langle \cdot, \cdot \rangle_{L^2(\mathbb{R})}$	Produit scalaire sur $L^2(\mathbb{R})$ tel que pour toute fonction $(f, g) \in L^2(\mathbb{R})^2$, $\langle f, g \rangle_{L^2(\mathbb{R})} = \int_{\mathbb{R}} f(t)g(t)dt$
$\ \cdot\ _p$	Norme ℓ_p
δ_λ	Fonction de seuillage des coefficients d'ondelettes
$\tilde{\sigma}_\varepsilon$	Variance du bruit
$\mathcal{G}_{\gamma, \beta}(\theta)$	Distribution Gaussienne généralisée
$x(t)$	Vecteur des observations (Signal brut)
$I(y)$	Information mutuelle
KL_m	Divergence de Kullback-Leibler
$V(B)$	Fonction de vraisemblance
ACI	Analyse en composantes indépendantes
ACP	Analyse en composantes principale
AMR	Analyses multi-résolutions
CWT	Continuous Wavelet Transform
CI	Composants Indépendants
BPFO	Ball Passing Frequency Outer Race
BPFI	Ball passing Frequency Inner Race
FFT	Fast Fourier Transform
DLP	Diagonal Linear Projection (projection Diagonale linéaire)
DLS	En anglais Diagonal Linear Shrinker (contraction diagonale)
DWV	Distribution de Wigner-Ville
DWVL	Distribution de Wigner-Ville lissée
DWT	Transformée d'ondelette discrète
EQM	Erreur Quadratique Moyenne'
GGD	Generalized Gaussian Distribution
ICA	Independent Components Analysis
i.i.d	indépendants et identiquement distribués
MAP	maximum a posteriori
MAD	Valeur médiane absolue
MSE	Mean Squared Error
RSB	Rapport signal sur bruit
SAS	Séparation Aveugle des Sources
SCA	Sparse Components Analysis
SURE	Stein's Unbiased Risk Estimate
STFT	Short time Fourier transform
SNR	Rapport signal sur résidus
WVD	Wigner-Ville Distribution
TF	Temps-fréquence
MP	Matching Pursuit
BP	Basis Pursuit
FastICA	Fast Independent Component Analysis

Chapitre I

Introduction Générale & Etat de l'Art

I.1 Introduction Générale

Les travaux de recherche réalisés dans le cadre de cette thèse se trouvent à la croisée des chemins entre les mathématiques appliquées et le traitement numérique du signal. Ils s'articulent autour du triptyque : (i) Modélisation et estimation statistique ; (ii) Analyse harmonique computationnelle et représentations parcimonieuses ; (iii) Optimisation. Ces trois piliers constituent le socle théorique de l'essentiel des activités menées dans le cadre de ce travail afin de développer des approches originales capables de résoudre des problèmes inverses en restauration et en séparation de sources vibratoires. Les phénomènes physiques intervenant à l'intérieur d'un système mécanique complexe sont trop nombreux et complexes pour que tous leurs effets puissent être pris en compte sans hypothèses simplificatrices. Cette forte complexité se retrouve au niveau des signaux de vibration recueillis sur la machine. Ces derniers, très riches en information, ne sont que partiellement interprétables par les moyens traditionnels : analyse temporelle, analyse de Fourier, analyse cepstrale et analyse d'enveloppe. Ces méthodes, historiquement les plus anciennes, consistent en une analyse des caractéristiques temporelles et fréquentielles du signal enregistré. Cependant, bien que ces techniques puissent s'avérer très utiles dans des cas simples et permettre la formulation rapide d'un pré diagnostic, elles présentent un certain nombre d'inconvénients. Ces inconvénients conduisent souvent à des pertes de temps considérables ou à la formulation de diagnostics erronés. La localisation de l'origine des chocs et des phénomènes de modulation et, en particulier, des événements non stationnaires nécessite la mise en œuvre de techniques encore plus élaborées. Par ailleurs, les signaux de vibration transmis à partir des structures mécaniques vers les points de mesure se trouvent mélangés les uns aux autres et couplés avec les bruits de structure et d'environnement. Les différents composants contribuent à produire de l'énergie de vibration et il est difficile de surveiller un composant spécifique car les vibrations produites par les autres composants dissimulent la signature de ce dernier. Une manière de surmonter cet effet est de monter les capteurs de vibrations aussi près que possible des composants d'intérêt. Cependant, ce n'est souvent pas possible en raison des restrictions d'accès au composant concerné comme la complexité de la machine et la politique de garantie du fabricant. La seule alternative est de placer les capteurs sur la surface la plus interne possible de la structure. En conséquence, le signal recueilli par les capteurs n'est pas seulement due à l'élément spécifique ciblé. Les signaux recueillis par chaque capteur sont en effet la combinaison de l'énergie vibratoire produite par les différents composants à proximité en plus du bruit. La dissipation de l'énergie de vibration par voie de transmission complique encore davantage la situation. Les facteurs ci-dessus montrent la difficulté d'obtenir la signature réelle isolée de chaque composant. Face à cette lacune, l'apport des représentations temps-fréquence est a priori indéniable [1]. En effet, de telles représentations vont permettre d'accéder simultanément à trois types d'information significative :

- L'amplitude de la représentation fournit une information de puissance.
- Les fréquences qui apparaissent vont permettre d'identifier et d'étudier les différents éléments mécaniques en jeu par leurs modes et les contraintes auxquelles ils sont soumis.

- Les deux informations précédentes, et donc les différents événements que l'on peut distinguer dans la représentation, peuvent être localisés dans le temps, donnant ainsi une information sur le déroulement et l'enchaînement des divers processus.

Un modèle des signaux vibratoires est donné par l'équation (1.1). Ce modèle consiste en une somme de sinusoides modulées en fréquence et en amplitude :

$$y(t) = \sum_{k=1}^K A_k(t) \sin(\omega_k t + \varphi_k(t)) \quad (1.1)$$

Dans la représentation temps-fréquence idéale correspondante, chaque composante du modèle se traduit par une suite de maxima d'amplitude $(A_k(t))^2$ et de fréquence instantanée $\omega_k + \dot{\varphi}_k(t)$. Autrement dit, tous les paramètres du modèle, à savoir le nombre de composantes K et les fonctions modulantes $A_k(t)$ et $\varphi_k(t)$ sont directement accessibles à partir de la représentation temps-fréquence, qui est donc parfaitement adaptée à ce type d'analyse. De plus, les représentations temps-fréquence permettent de mettre en évidence des 'composantes complexes' de signaux, de façon très simple. En particulier, une variation locale de la fréquence, qui est une façon d'envisager la modulation de phase, se traduit simplement par une trajectoire dans le plan temps-fréquence. Bien sûr, l'utilisation des représentations temps-fréquence ne se limite pas au modèle précité. Des composantes supplémentaires, avec un support limité en temps, correspondant à une modification dynamique du nombre de composantes K , peuvent également être prises en compte. Globalement, ces représentations temps-fréquence sont considérées comme des outils puissants pour analyser les signaux non-stationnaires [2]. Dès lors, ce travail consiste, dans sa première partie, à étudier, comparer, modifier et adapter des techniques de représentation temps-fréquence qui peuvent être utilisées pour étudier le contenu fréquentiel d'un processus non stationnaire. Les méthodes courantes de diagnostic des machines basées sur la représentation du signal dans le domaine temps-fréquence sont présentées et les avantages et désavantages de chaque méthode sont notés. En premier, l'application de la transformée de Fourier à fenêtre glissante (STFT), la plus simple et la plus rapide des méthodes temps-fréquence est présentée. Les avantages et les inconvénients de cette méthode sont également expliqués. Aussi, la distribution de Wigner-Ville, est brièvement expliquées et leur avantages et désavantages examinés avec un programme en Matlab. Les signaux enregistrés durant les essais expérimentaux ont été utilisés pour vérifier la performance de chaque méthode. Ensuite, les plus récentes des méthodes pour une présentation temps-fréquence, soit les méthodes temps-échelle, sont introduites. L'une de ces méthodes, la transformée en ondelettes, donne la possibilité de compenser les faiblesses des méthodes temps-fréquence citées ci-dessus. La transformée en ondelettes possède la caractéristique particulière de parcimonie dans le plan temps-fréquence par laquelle elle est devenue un outil très efficace pour l'analyse des événements transitoires. La deuxième partie de cette thèse concerne les problèmes inverses. De manière très générale, le diagnostic et le contrôle à partir du signal peuvent être présenté comme des problèmes inverses, qui à l'observation d'un effet cherche sa cause. Il est bien connu que la résolution des problèmes

inverses est fondamentalement difficile étant donné le grand nombre d'inconnues qui les caractérisent et l'instabilité de leurs solutions. De nombreux efforts ont récemment été déployés dans ce domaine, avec des succès prometteurs. Nous distinguerons deux sous problèmes: l'identification aveugle et la séparation aveugle de sources (SAS). L'estimation aveugle consiste à estimer les paramètres qui décrivent un transfert au sein d'un milieu physique à partir de mesures externes seulement, c'est-à-dire sans connaître les excitations en entrée. En outre, lorsqu'elle est pratiquée en contexte industriel sur des systèmes complexes qui impliquent le fonctionnement de plusieurs organes et de sous-systèmes, l'analyse du signal vibratoire se heurte fréquemment à une difficulté pratique qui consiste à reconnaître au niveau des capteurs les contributions de plusieurs phénomènes physiques de natures différentes et que l'expérimentateur ne peut maîtriser de manière indépendante. En fonction du ou des phénomènes que l'expérimentateur cherche à isoler, il y a donc un intérêt certain à savoir décomposer les contributions de chacune des sources au niveau des mesures pour pouvoir ensuite les analyser individuellement. Un certain nombre de méthodes ont récemment été développées pour résoudre ce problème sans aucune connaissance a priori sur le milieu de propagation des signaux ni des sources qui les génèrent. Seule l'indépendance statistique des sources est supposée, ce qui permettrait de séparer les contributions de chacune des sources au niveau des signaux mesurés.

I.2 Objectifs du présent travail

Vu la difficulté de trouver les défauts caractérisés par des phénomènes transitoire ou non stationnaires dans la représentation temporelle ou fréquentielle, il est nécessaire d'aller vers la représentation mixte où la représentation temps-fréquence-énergie nous permettra :

- De détecter et de suivre l'évolution de défauts qui induisent une puissance vibratoire faible, qui cependant peut modifier considérablement la forme du signal. C'est notamment le cas des défauts qui induisent une modulation de l'amplitude ou de la fréquence de certaines composantes caractéristiques de la chaîne cinématique complexe.
- De surveiller les installations dont le processus de fonctionnement normal génère des chocs périodiques d'amplitudes élevées susceptibles de masquer l'apparition de défauts induisant des forces impulsionnelles.

Dans cette thèse nous prenons en charge les points suivants:

- a) Un problème que pose la représentation temps-fréquence est le manque de précision et de résolution dans des méthodes temps-fréquence et temps-échelle. Pour résoudre ce problème, nous développons une nouvelle approche basée sur la recherche de la meilleure base de décomposition permettant une représentation la plus parcimonieuse du signal vibratoire.
- b) Un autre problème de représentation temps-fréquence est le bruit qui nous empêche de voir le signal. Malgré le fait que la représentation temps-fréquence disperse le bruit sur le plan temps-fréquence donc garde le niveau du bruit bas, il reste que le niveau du bruit

complique l'identification des défauts. Pour résoudre ce problème, une approche Bayésienne est développée pour le débruitage (denoising) des signaux de vibrations.

- c) Enfin, le problème de la séparation aveugle des sources bruitées a été introduit et formulé afin de pouvoir séparer les effets des différentes sources excitatrices qui se combinent au niveau des mesures. La présence de bruit est néfaste pour l'estimation des sources. Une approche combinée basée sur l'amélioration de la parcimonie des signaux impulsionsnels et l'approche Bayésien est introduite.

I.3 Etat de l'art

La première méthode en temps-fréquence qui a été utilisée dès 1940 par Gabor pour la transmission de données était le spectrogramme ou la transformée de Fourier à fenêtre glissante (STFT). C'est une méthode simple et efficace qui s'utilise surtout pour analyser des signaux non stationnaires. L'inconvénient majeur de cette méthode est qu'elle dépend complètement de la longueur temporelle de la fenêtre. Une première idée, présentée par J. Ville (1948) pour surmonter le problème de résolution découle d'une méthode mathématique qui s'appelle (coherent states) en mécanique quantique. J. Ville a dérivé la distribution que Wigner avait élaborée en 1932 pour étudier la statistique de la mécanique quantique (quantum statistical mechanics). J. Ville, à la recherche du spectre instantané, voulait déployer l'énergie du signal dans le plan temps-fréquence et obtenir une densité d'énergie ayant au moins les propriétés suivantes:

- La somme des énergies pour toutes les fréquences à un temps particulier donne l'énergie instantanée.
- La somme des énergies pour tous les temps à une fréquence particulière donne le spectre instantané.

Ces deux propriétés s'appellent les conditions marginales en temps et en fréquence. On peut tout de suite constater que contrairement au spectrogramme, la transformée de Wigner-Ville donne la fréquence instantanée. Mais la distribution Wigner-Ville présente aussi des désavantages :

- a. Elle ne donne pas toujours une distribution positive.
- b. Pour un signal à composantes multiples, elle induit les termes rectangulaires (cross terms) qui sont des artefacts dont la localisation dans le plan temps-fréquence est à mi-chemin de celle des termes carrés (self terms) correspondants.

En 1989, Choi et Williams [3] ont présenté une nouvelle approche où ils se sont attaqués à l'inconvénient principal de la distribution de Wigner-Ville, soit les termes rectangulaires. Le noyau de leur distribution a une constante telle qu'en ajustant cette constante, on peut minimiser les termes rectangulaires. En augmentant la constante, on va vers la distribution de Wigner-Ville et en diminuant le paramètre, on élimine les termes rectangulaires mais en

même temps on perd de la résolution dans le temps et dans la fréquence. Dans le domaine du diagnostic de machine, Forrester [4] a utilisé la distribution de Wigner pour trouver des défauts dans le système de transmission d'un hélicoptère. Il a montré que la distribution Wigner-Ville peut révéler le défaut avec plus précision que les méthodes conventionnelles. Dans une série d'articles, Mcfaden et Wang [5, 6] ; ont expliqué comment ils ont appliqué les méthodes temps-fréquence au diagnostic des machines. Ils ont présenté la complexité de la représentation temps-fréquence produite par la méthode de Wigner-Ville et ont suggéré une fenêtre temporelle qui peut améliorer les résultats de Wigner-Ville. Rohrbaugh [7] a utilisé l'analyse temps-fréquence pour des machines marines. Il a comparé la STFT avec la représentation temps-fréquence de Cohen ayant un noyau cône (cone kernel). Rohrbaugh et Cohen [8] ont appliqué une nouvelle méthode de temps-fréquence qui est développée pour le diagnostic de pompes. Cette méthode est appelée (positive time-frequency distribution) et elle est plus efficace que la méthode de STFT pour le diagnostic de machines alternatives. Quelques applications des méthodes temps-fréquence à la surveillance de processus d'usinage comme le perçage et le broyage ont été présentées par Loughlin et al. [9]. Ils ont montré que les nouvelles méthodes peuvent fournir plus de détails, sur le signal que la méthode STFT. Atlas et al. [10] ont offert un résumé des applications de l'analyse temps-fréquence dans différents domaines de diagnostic des machines. Ils ont souligné l'importance d'applications des méthodes temps-fréquence dans le secteur industriel et la surveillance des machines. Ces techniques appartiennent à un ensemble plus général d'algorithmes qu'on retrouve aussi bien chez les mathématiciens que chez les spécialistes du traitement de signal. Dans cette optique, les transformées en ondelettes ont fait leur apparition et se sont largement imposées dans l'analyse et la résolution des problèmes liés aux sciences et aux techniques de l'ingénierie. Cet essor est dû principalement à deux propriétés spécifiques qui résultent des décompositions de signaux et de processus aléatoires par projection sur les bases d'ondelettes : la parcimonie de représentation et le blanchiment de données corrélées. La parcimonie d'une représentation de signal décrit la capacité de cette représentation à répartir l'énergie du signal sur un petit nombre de coefficients de grandes amplitudes. Les transformées en ondelettes permettent d'obtenir cette représentation parcimonieuse pour les signaux réguliers ou réguliers par morceaux. À l'opposé, la structure des composantes d'une certaine classe de perturbations reste sensiblement inchangée par cette même transformée. Cette capacité de compacter l'énergie de signaux est de nature à favoriser l'utilisation des bases d'ondelettes pour simplifier divers problèmes de traitement de signaux, d'analyse des données, compte tenu du fait que l'analyse de données non corrélées est souvent beaucoup plus simple que celle de données corrélées. L'utilisation des méthodes d'ondelettes pour le diagnostic des machines est très récente. L'identification des défauts dans les roulements à billes (Hongbin [11]) et dans les boîtes d'engrenage (Mcfadden et Wang [12]) sont des exemples d'application de cette méthode aux éléments essentiels des machines tournantes. Belaid et Miloudi ont exploité la redondance, caractéristique de la transformée en ondelettes continue, pour détecter des résonances dans le spectre d'un signal vibratoire. Cette approche leur a permis la localisation exacte de la résonance qui est un facteur important dans la procédure de diagnostic [13]. Dans une autre approche, Dalpiaz et Rivola ont utilisé la transformée en ondelettes pour la surveillance et le diagnostic dans les mécanismes contenant des cames [14].

Cependant, en traitement de signal vibratoire, les données observées sont très souvent composées de l'information utile (signal) corrompue par une perturbation généralement indépendante. La perturbation (bruit) peut être interne au système physique de traitement de l'information ou externe, due aux interactions entre le signal et les éléments de son milieu de propagation. L'estimation des signaux dans un milieu bruité est un cadre qui a été largement étudié dans les domaines de traitement du signal et de statistiques. Le problème a été donc abordé de différentes façons et plusieurs méthodes d'estimation ont été développées. Par exemple, on peut citer les travaux de Wiener [15], d'Ephraïm-Malah [16] ou Stein [17], pour n'en citer que quelques uns. Mais c'est dans les années 90, que les travaux de Donoho et son équipe ont révolutionné la théorie de l'estimation statistique, grâce notamment aux estimateurs de seuillage dans les bases d'ondelettes et paquets d'ondelettes [18-21]. Ces travaux exploitaient les progrès récents en théorie des ondelettes pour concevoir des opérateurs d'estimation non-linéaires simples et efficaces. Plus généralement, Donoho a montré le rôle important que joue la parcimonie dans l'efficacité des estimations dans le cas où la distribution du signal n'est pas connue et elle s'avère assez souvent difficile à estimer en pratique. Il en est de même concernant la nature de la perturbation, quand sa distribution n'est pas connue et est assez difficile à modéliser. Face à cette situation, Donoho et Johnstone ont développé des méthodes à coût minimum en termes de connaissances disponibles sur la nature des composantes de l'information reçue. On parle alors de décision et d'estimation (plus généralement de statistique) non-paramétriques au sens où l'on ne tient pas compte de connaissances *a priori* sur les distributions des signaux constituant l'observation. En effet, les performances obtenues pour l'estimation non-paramétrique avec ces méthodes sont comparables aux performances d'un oracle, lorsque ces fonctions de seuillage sont calibrées avec des seuils convenablement choisis [22]. Les techniques de régression non-paramétrique (ou débruitage) offrent un panel d'outils simples d'emploi permettant de récupérer une structure, sans imposer un modèle paramétrique au préalable. Cependant, toutes ces fonctions sont basées sur le seuillage, ce qui introduit un traitement singulier en ce sens où, de part et d'autre du seuil, les données sont traitées de manière différente. Notons, d'autant plus, que le principe de seuillage assume que la représentation du signal est parcimonieuse au sens strict. Si cela est souvent le cas pour des signaux réguliers, il n'en est pas de même pour les signaux naturels comme le signal vibratoire des machines complexes. Le fait que ces derniers signaux admettent une représentation en ondelettes parcimonieuse au sens large ne justifie pas l'adéquation du seuillage pour l'estimation de ces signaux. De ce fait, les approches bayésiennes ont également été envisagées dans la littérature pour l'estimation dans le domaine des ondelettes. Les approches bayésiennes s'avèrent plus efficaces que celles présentées dans le cadre classique (sans *a priori*) [23]. Dans le cadre d'une approche bayésienne, on suppose que la distribution des coefficients d'ondelettes est imposée comme *a priori*. Le choix d'une fonction de coût conditionne l'obtention d'une règle d'estimation bayésienne. Plusieurs modèles statistiques *a priori* adaptés à la modélisation statistique des coefficients d'ondelettes dans le cas bruité ont été envisagés. La combinaison de l'information contenue dans les données observées et celle contenue dans la loi *a priori* sur le signal à estimer permet d'obtenir la distribution conditionnelle *a posteriori*. Cette dernière permet d'obtenir l'estimateur bayésien et d'estimer la fonction inconnue.

Mallat a été le premier à proposer les distributions gaussiennes généralisées (GGD) [24]. Elles sont très communément adoptées comme *a priori* à des fins d'estimation et de compression dans le domaine des ondelettes [23]. Cependant, l'*a priori* GGD se révèle incapable de modéliser correctement les queues lourdes des distributions relevées, comportement typique des distributions des coefficients d'ondelettes. Plusieurs études ont ainsi mis en évidence que les propriétés statistiques, dans de telles représentations, sont loin d'avoir un comportement gaussien. En effet, choisir un espace muni d'une trame ou d'une base et projeter le signal dans un sous-espace de dimension inférieure dans lequel seules quelques composantes sont significatives sont loin d'avoir un comportement gaussien mais plutôt un comportement leptokurtique (un pic prononcé) avec des queues lourdes (queues relevées). C'est ce modèle d'*a priori* qui a été adopté dans le cadre de ce travail, pour modéliser les coefficients du signal vibratoire de type impulsif. Finalement, en poursuivant cette démarche, ce travail a la lourde ambition de fiabiliser le diagnostic et le suivi à travers l'intégration des problèmes inverses. L'intérêt est de séparer la contribution des différentes sources vibratoires généralement et directement liées à une défaillance plus ou moins importante d'un composant mécanique, à partir de plusieurs mesures réalisées par les accéléromètres. La séparation permettra non seulement de localiser les défauts sur les composants mais également de suivre l'évolution de l'endommagement de chacun d'entre eux. En effet à toutes les fois qu'il est nécessaire de surveiller n'importe quelle composante d'intérêt, les vibrations produites par les autres affectent le signal. Parmi les moyens pour contourner ce problème est de placer des capteurs aussi proches que possible des composantes données. Mais, certaines restrictions telles que la complexité, la politique de garantie du fabricant et l'inaccessibilité empêchent de tel emplacement. Du coup, dans la majorité des cas les capteurs sont placés sur la surface extérieure de la structure. Par conséquent les capteurs collectent non seulement des signaux de vibrations d'une composante spécifique mais plutôt, la combinaison de l'énergie vibratoire des différentes composantes, plus le bruit. L'approche adoptée dans le cadre de cette thèse pour résoudre ce problème, consiste à considérer ce cas comme un problème de séparation aveugle de sources et en tirer profit des méthodes statistiques et mathématiques développées à cet effet. L'objectif est de restituer les signaux sources émis en utilisant uniquement les signaux reçus. Les premiers travaux sur la séparation aveugle de sources ont été initiés par C. Jutten et J. Héroult en 1991, dans le cas d'un mélange instantané [25]. Depuis, ce thème a fait l'objet de nombreuses recherches dans divers domaines d'applications. Une multitude d'algorithmes a été proposée. Les algorithmes de séparation de sources peuvent être classés selon différents critères : mélange instantané ou mélange convolutif, linéaire ou non linéaire, batch ou adaptatif, statistiques d'ordre deux ou d'ordre supérieur. Pour se faire, les chercheurs dans ce domaine se basent essentiellement sur l'hypothèse de l'indépendance mutuelle des signaux sources, c'est à dire que, la densité de probabilité des signaux sources est égale au produit de ses densités de probabilités marginales. Dans ce cas le problème est connu aussi sous le nom de l'analyse en composantes indépendantes ACI (ou ICA, acronyme de l'anglais Independent Component Analysis). Dans [26], Pham et al. ont résolu le problème de séparation par une méthode fondée sur le principe du maximum de vraisemblance en considérant des observations non bruitées. De nombreux auteurs se sont tournés vers le diagnostic de machines tournantes.

Cheng et al. [27] ont utilisé l'ACI pour évaluer la contribution des sources vibratoires d'une machine tournante. Sur une plateforme où deux moteurs fonctionnent simultanément, El Rhabi et al: [28], appliquent la SAS aux signaux vibratoires pour diagnostiquer l'état de fonctionnement de chaque moteur lors qu'ils fonctionnent simultanément. Afin d'extraire la contribution des vibrations générées par un passage des billes sur un défaut localisé sur la bague interne d'un roulement, Randall et al [29] et Bouguerriou et al [30] appliquent la SAS pour traiter le problème. Depuis ces travaux pionniers en séparation de sources, ce domaine de recherche a pris une importance grandissante au point où un workshop annuel lui est dédié (ICA Workshop). Néanmoins, ces méthodes de séparation de sources supposent implicitement que le modèle d'observation (mélange instantané/convolutif ou linéaire/non-linéaire) est exact et que les données ne sont pas bruitées. En pratique, lors de l'application, ces méthodes se heurtent à des difficultés car la réalité est souvent plus complexe. Il est donc impératif de prendre en compte le bruit, ce qui entraîne un problème d'estimation de la matrice de séparation et, même si l'estimation est exacte, la restauration des sources est bruitée. Une première étude sur l'influence du débruitage par ondelettes préalablement à la séparation a été menée par Paraschiv-Ionescu et al. [31] qui ont proposé le schéma de débruitage pré-séparation. Ainsi, pour obtenir de bonnes performances dans des environnements fortement bruités, les observations sont débruitées en appliquant un seuillage doux avant d'appliquer l'algorithme de séparation de sources. Une amélioration de la méthode existante peut être apportée en ajoutant après la séparation un débruitage post-séparation. Cette amélioration permet de compléter le débruitage pré-séparation dans le cas où on aurait fait le choix de ne pas trop débruiter les observations afin de ne pas perdre trop de signal ce qui serait néfaste pour l'estimation de la matrice de séparation. En effet, un débruitage pré-séparation trop énergique risquerait de modifier la structure du mélange et de nuire à la qualité de l'estimation de la matrice de séparation. En fin, pour conclure, nos contributions dans le cadre de ce travail s'articulent sur l'achèvement de la démarche suivante: Pour résoudre le problème le manque de précision et de résolution dans des méthodes temps-fréquence, nous développons une nouvelle approche basée sur la recherche de la meilleure base de décomposition permettant une représentation parcimonieuse du signal vibratoire impulsionnel. Particulièrement, nous nous intéresserons à son intégration dans des problèmes inverses. A partir de ces problèmes inverses, on verra comment une approche par l'utilisation des représentations redondantes, alliée au principe de parcimonie par son utilisation comme source d'*a priori*, permet d'attaquer ces problèmes en vue de leur inversion. Dans un second lieu, nous exposerons le formalisme général du problème de débruitage dans le domaine des transformées en ondelettes en présence d'un bruit additif gaussien. Ainsi, deux types d'approches seront introduites : l'approche classique de type estimation non-paramétrique (seuillage de coefficients) et l'approche bayésienne basé sur des modèles statistiques adaptés à la modélisation des coefficients. Nous traitons le choix du seuil pour la séparation entre les coefficients qui seront attribués au bruit et ceux qui seront attribués au signal. Ensuite, dans le contexte Bayésien et afin de reconstruire le signal non bruitée, il est nécessaire d'imposer des contraintes sur la solution reconstruite, ce qui consiste à choisir un modèle a priori leptokurtique. Pour conclure cet état de l'art, nous avons proposé un bilan comparatif entre ces deux types d'approches tout en soulignant leurs limitations.

En dernier, afin de pouvoir séparer les effets des différentes sources excitatrices qui se combinent au niveau des mesures, le problème de la séparation aveugle des sources bruitées (SAS) a été abordé comme extension naturelle du cadre du débruitage traité dans le chapitre précédent. Plusieurs ingrédients seront empruntés aux chapitres précédents notamment ceux traitant de la modélisation et la parcimonie des sources. L'approche adoptée dans le cadre de ce travail repose sur un traitement post opératoire qui combine l'utilisation de la parcimonie comme source d'*a priori* pour les signaux impulsionnels et l'approche Bayésien. Cette méthode présente l'avantage de pouvoir adapter le débruitage post-séparation, opérés individuellement sur chaque signal, à la nature des sources recherchées et tire profit du fait qu'on possède plus d'information sur les sources recherchées que sur la nature des mélanges observés.

I.4 Organisation de la thèse

Comme on vient de voir, cet état de l'art comporte une brève revue historique suivie par une bibliographie des méthodes temps-fréquence et temps-échelle pour la surveillance des machines. L'objectif du travail et l'organisation de cette thèse sont également définis dans ces pages. Dans le **chapitre 2** les avantages et les inconvénients des méthodes traditionnelles de diagnostic des machines sont soulignés et la nécessité d'appliquer des méthodes temps-fréquence est discutée par quelques exemples. L'application de la méthode de la transformée de Fourier à fenêtre glissante et de la distribution de Wigner-Ville au diagnostic de machines sont également présentées dans ce chapitre, il s'agit de rappeler l'intérêt de telles transformations en insistant sur les problèmes pratiques posés par leurs mises en œuvre et leurs interprétations. On discute particulièrement le choix des fonctions de lissage à utiliser pour obtenir des performances supérieures de la distribution de Wigner-Ville comparée à la transformée de Fourier à court terme. **Le chapitre 3** présente l'application de la transformée en ondelettes dans le domaine du diagnostic des machines. Les avantages et désavantages de cette méthode sont montrés par quelques exemples. De plus les fondements mathématiques de l'analyse multi-résolution et les avantages d'une telle décomposition sont détaillés. Nous verrons que l'analyse multi-résolution permet de mettre en relief les échelles caractéristiques d'un signal en le décomposant dans la base des fonctions d'ondelettes. **Le quatrième chapitre** de ce manuscrit constitue une introduction à la parcimonie. Il permettra de mieux appréhender ce domaine et fournira ainsi les outils nécessaires au développement des algorithmes proposés dans ce manuscrit. Particulièrement, nous nous intéresserons à son intégration dans des problèmes inverses. A partir de ces problèmes inverses, on verra comment une approche par l'utilisation des représentations redondantes, alliée au principe de parcimonie par son utilisation comme source d'*a priori*, permet d'attaquer ces problèmes en vue de leur inversion. Dans **le cinquième chapitre** de cette thèse, nous exposerons le formalisme général du problème de débruitage dans le domaine des transformées en ondelettes en présence d'un bruit additif gaussien. Ainsi, deux types d'approches seront introduites : l'approche classique de type estimation non-paramétrique (seuillage de coefficients), où nous traitons le choix du seuil pour la séparation entre les coefficients qui seront attribués au bruit et ceux qui seront attribués au signal et l'approche bayésienne basé

sur des modèles statistiques adaptés à la modélisation des coefficients. L'inférence Bayésienne consiste à estimer le signal inaccessible et non observé x à partir de sa version dégradée observée y en exploitant la loi *a priori* sur x et la loi conditionnelle de y sachant x déduite généralement de l'équation de dégradation. Ceci est accompli en minimisant un risque moyen au sens de la loi conditionnelle *a posteriori* de x/y . Typiquement, le risque quadratique correspond à l'estimateur du maximum a posteriori (MAP) et nécessite de résoudre un problème d'optimisation. De façon plus générale, bon nombre de problèmes en traitement de signal se ramènent à la formulation d'une énergie que l'on doit minimiser, cette énergie possédant un terme d'attache aux données et un ou plusieurs termes de régularisation. Cependant, dans le contexte Bayésien, il est nécessaire d'imposer des contraintes sur la solution reconstruite, ce qui consiste à choisir un modèle *a priori* adaptée. L'une des hypothèses les plus abondamment adoptées est celle où le signal d'intérêt est supposé *a priori* Gaussien, et où le bruit est additif blanc Gaussien indépendant du signal utile. Ce modèle est toutefois très simplificateur des situations réelles et ne reflète ni la complexité des propriétés statistiques des signaux naturelles, ni celle des dispositifs d'acquisition rencontrés en pratique. Il apparaît donc crucial d'établir à la fois un modèle approprié du signal de dégradation (du bruit), mais aussi de concevoir un *a priori* de type leptokurtique le plus subtile pour capturer des propriétés statistiques complexes du signal que l'on veut récupérer. Pour conclure cet état de l'art, nous avons proposé un bilan comparatif entre ces deux types d'approches tout en soulignant leurs limitations. Dans **le sixième chapitre**, le problème de la séparation aveugle des sources bruitées (SAS) a été abordé comme extension naturelle du cadre du débruitage traité dans le chapitre précédent. Plusieurs ingrédients seront empruntés aux chapitres précédents notamment ceux traitant de la modélisation et la parcimonie des sources. L'approche est basée sur l'analyse en composantes indépendantes (ACI). Cependant, avant d'entamer cette approche, il nous a semblé utile de passer en revue au préalable l'état de l'art du problème de la SAS qui ne peut être exhaustif. Nous rappelons le problème de la SAS et les éléments qui s'y rattachent, la modélisation du mélange, l'objectif et les difficultés de la résolution de ce problème. Ensuite une approche post opératoire qui combine l'utilisation de la parcimonie comme source d'*a priori* pour les signaux impulsionnels et l'approche Bayésien est introduite. Cette méthode présente l'avantage de pouvoir adapter le débruitage post-séparation, opérés individuellement sur chaque signal, à la séparation des sources recherchées et tire profit du fait qu'on possède plus d'information sur les sources recherchées que sur la nature des mélanges observés.

Chapitre II

Synthèse des méthodes temps-fréquence pour l'analyse des signaux de vibrations

Chapitre 2

Synthèse des méthodes temps-fréquence pour l'analyse des signaux de vibrations

Résumé

Dans ce chapitre, la théorie des méthodes temps-fréquence est brièvement expliquée et des exemples sur des signaux simulés et des signaux réels à partir de banc d'essais des paliers à roulement défectueux sont analysés. L'objectif est d'étudier, comparer et modifier les techniques de représentation temps-fréquence à savoir: la transformée de Fourier à fenêtre glissante (STFT) et la distribution de Wigner-Ville (WVD). Les avantages et les inconvénients de ces méthodes sont également expliqués et des solutions pour améliorer leurs performances sont exposées.

Sommaire

II.1	Introduction	21
II.2	Origine des analyses temps-fréquence	21
II.3	Atome analysant temps-fréquence	22
II.4	Théorème d'incertitude d'Heisenberg.	22
II.5	La transformée de Fourier à fenêtre glissante (STFT)	23
	II.5.1 Adaptativité de la représentation temps-fréquence de type STFT.	28
	II.5.2 Critères d'optimalité.	28
	II.5.3 Optimisation globale.	30
	II.5.4 Optimisation locale dans le plan temps-fréquence.	30
II.6	La Distribution de Wigner-Ville.	31
	II.6.1 Introduction.	31
	II.6.2 Définition.	31
	II.6.3 Propriétés.	32
	II.6.4 Mécanisme des interférences.	33
	II.6.5 Lissage de la distribution de Wigner Ville	34
II.7	Conclusion.	36

II.1 Introduction

La manifestation initiale d'un défaut dans une machine peut produire des phénomènes transitoires dans le signal de vibration. Le passage d'une bille sur un défaut localisé sur l'une des pistes d'un roulement, le contact à l'engrènement d'une dent endommagée avec d'autres dents dans un réducteur sont des exemples de problèmes industriels bien connus générant des événements transitoires. Les méthodes d'analyse des vibrations dans le domaine fréquentiel, telles que le spectre de puissance, atténuent par moyennage, les événements transitoires et de ce fait ils n'apparaissent pas clairement dans les lignes spectrales. Les méthodes d'analyse dans le domaine temporel sont également utilisées pour analyser des signaux transitoires, mais l'information concernant la fréquence des différents composants de la machine est perdue. Ainsi, il apparaît, par conséquent, nécessaire d'utiliser de nouvelles techniques qui combinerait l'information fréquentiel et les variations d'amplitude dans le temps. Dans ce chapitre, la théorie des méthodes temps-fréquence est brièvement expliquée et des exemples sur des signaux simulés et des signaux réels à partir de banc d'essais de paliers à roulement défectueux sont analysés. L'objectif de ce travail, tel qu'il avait été défini au départ, consiste à évaluer l'apport des représentations temps-fréquence (TF) à l'analyse des signaux vibratoires et par là même, à la compréhension des phénomènes physiques sous-jacents conduisant ainsi à la sélection et l'interprétation physique des composantes significatives.

II.2 Origine des analyses temps-fréquence

L'analyse de Fourier demeure encore aujourd'hui un outil incontournable dans le domaine du traitement du signal. Plus encore, elle fournit un formalisme mathématique d'excellence pour l'analyse des systèmes différentiels linéaires. Ceci s'explique par le fait que les motifs analysants de Fourier, que sont les fonctions exponentielles complexes, forment naturellement une base de vecteurs propres orthogonaux pour toute application différentielle linéaire à coefficients constants. De façon intrinsèque, cette analyse est tournée vers l'étude des comportements stationnaires d'un signal. Par conséquent, un phénomène local intervenant dans un signal sera décomposé en base de Fourier comme la superposition d'une infinité de phénomènes stationnaires; l'événement local étant inextricablement intégré dans les déphasages des fonctions sinusoïdales. Dès lors il est devenu nécessaire de construire des outils mathématiques plus sophistiqués permettant une analyse pertinente des phénomènes transitoires des systèmes non linéaires. En 1948, Wigner et Ville ouvrent la voie en proposant la distribution de Wigner-Ville. Cette transformation intégrale associe à un signal unidimensionnel un champ bidimensionnel de valeurs scalaires défini dans le plan temps-fréquence. Les performances de cet outil se concentrent sur la localisation de fréquences pures pour lesquelles la transformation de Wigner-Ville réalise une collocation. Les analyses en ondelettes définies comme la transformation intégrale d'un signal par un ensemble de fonctions analysantes corrélées, appelées atomes temps-fréquence, fournissent à la fois une solution opérationnelle au problème de la localisation d'événements transitoires mais également à celui de la représentation efficace de signaux.

II.3 Atome analysant temps-fréquence

Afin de se rendre compte du comportement fréquentiel instantané des signaux étudiés, une fonction analysante $\psi \in L^2(\mathbb{R})$ doit posséder des propriétés de localisation fréquentielle et temporelle adaptées à l'information qu'elle doit mesurer par le biais d'un produit scalaire. Les quantités qui permettent d'apprécier les dispositions d'une telle fonction ψ quant à l'analyse d'un signal dans le plan temps-fréquences sont ses positions moyennes τ_ψ et ξ_ψ dans le domaine temporel, respectivement dans le domaine fréquentiel, données par les relations :

$$\tau_\psi = \frac{1}{\|\psi\|^2} \int_{\mathbb{R}} t |\psi(t)|^2 dt \quad (2.1)$$

$$\xi_\psi = \frac{1}{2\pi\|\psi\|^2} \int_{\mathbb{R}} \xi |\hat{\psi}(\xi)|^2 d\xi \quad (2.2)$$

Ainsi que ces variances temporelles $\sigma_{\tau,\psi}$ et fréquentielle $\sigma_{\xi,\psi}$ données par les relations :

$$\sigma_{\tau,\psi}^2 = \frac{1}{\|\psi\|^2} \int_{\mathbb{R}} (t - \tau_\psi)^2 |\psi(t)|^2 dt \quad (2.3)$$

$$\sigma_{\xi,\psi}^2 = \frac{1}{2\pi\|\psi\|^2} \int_{\mathbb{R}} (\xi - \xi_\psi)^2 |\hat{\psi}(\xi)|^2 d\xi \quad (2.4)$$

Ces dernières caractéristiques géométriques sont contraintes par une loi naturelle, mise en évidence par W. Heisenberg et traduite en des termes mathématiques dans le théorème qui suit. De termes, Gabor introduit la terminologie de logon ou atome temps-fréquence [32].

II.4 Théorème d'incertitude d'Heisenberg

Le principe d'incertitude permet de lier la durée d'un signal à sa largeur de bande. Il tire son nom de la mécanique quantique, où il exprime l'impossibilité de connaître avec précision la position et le moment cinétique d'une particule en mouvement. Son introduction en théorie du signal, due à Gabor, traduit l'impossibilité d'avoir un signal dont les supports temporels et fréquentiels sont arbitrairement petits. Le principe d'incertitude fournit alors une borne inférieure sur l'encombrement TF d'un signal.

Théorème 2.1 : *Soient une fonction $\psi \in L^2(\mathbb{R}, \mathbb{C})$ de variance temporelle $\sigma_{\tau,\psi}$ et de variance fréquentielle $\sigma_{\xi,\psi}$. Alors le produit des variances admet un minorant strictement positif :*

$$\sigma_{\tau,\psi} \sigma_{\xi,\psi} \geq 1/4 \quad (2.5)$$

L'égalité est obtenue dans le cas où la fonction ψ est un chirp de Gabor donnée par l'équation (2.6).

Un chirp est un signal concentré sur une fréquence instantanée qui varie dans le temps) :

$$\psi(t) = A(t)e^{-j\varphi(t)} \rightarrow \omega(t) = \varphi'(t) \quad (2.6)$$

D'après ce théorème, aucune fonction analysante ne peut fournir une collocation parfaite à la fois en temps et en fréquence.

II.4.1 Théorème de Paley-Wiener

Théorème 2.2 : *Soit une fonction $\psi \in L^2(\mathbb{R})$ non nulle et à support compact alors sa transformée de Fourier $\hat{\psi}$ ne peut pas s'annuler sur tout un intervalle ; de même si $\hat{\psi}$ est non nulle et à support compact alors ψ ne peut pas s'annuler sur tout un intervalle.*

Cet argument sur la compacité des supports temporel et fréquentiel des fonctions de $L^2(\mathbb{R})$ indique que des compromis sont impératifs et que l'évaluation d'une véritable fréquence instantanée est impossible. En raison de ces deux théorèmes, la localisation dans le plan temps-fréquence d'une fonction analysante est alors évaluée en écart-type en définissant dans le plan temps-fréquence un domaine, appelé 'Boîtes de Heisenberg' décrit par la définition suivante.

II.4.2 Boîte de Heisenberg

Soit une fonction $\psi \in L^2(\mathbb{R})$, alors on appelle boîte de Heisenberg le rectangle du plan temps-fréquence centré sur les positions moyennes temporelle τ_ψ et fréquentielle ξ_ψ dont l'aire est définie par les variances temporelle $\sigma_{\tau,\psi}$ et fréquentielle $\sigma_{\xi,\psi}$.

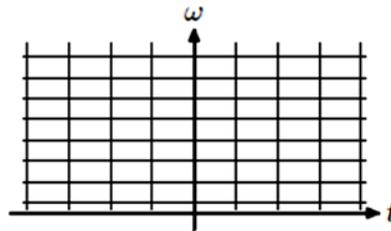


Fig. 2.1 Pavages du plan temps-fréquence (boîtes de Heisenberg)

L'inconvénient d'une telle représentation réside dans la largeur constante de la boîte : toutes les fréquences sont analysées avec la même résolution. L'inégalité de Heisenberg-Gabor peut donc être vue comme une limitation mathématique qui interdit une localisation arbitrairement petite dans les domaines temporel et fréquentiel.

II.5 La transformée de Fourier à fenêtre glissante (STFT)

La transformation de Fourier occupe une place privilégiée dans la théorie et le traitement des signaux. Néanmoins, la nature même de cette transformation ne peut donner d'un signal temporel qu'une information fréquentielle globale dans laquelle toute notion de chronologie a disparu : son champ d'application naturel est l'analyse des signaux stationnaires.

$$X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t} dt \quad (2.7)$$

Ainsi, dès lors que l'on considère des signaux modulés ou des processus non stationnaires, une analyse spectrale classique fournit une information moyennée sur la durée d'observation et donc peu caractéristique de la structure instantanée du signal étudié. L'exemple représenté sur la figure (2.2) correspondant à un palier avec un défaut de roulement, ne peut être diagnostiqué avec certitude car la représentation fréquentielle fournit des pics situés aux fréquences caractéristiques du défaut, dont les amplitudes du bruit dominant l'effet des impulsions périodiques.

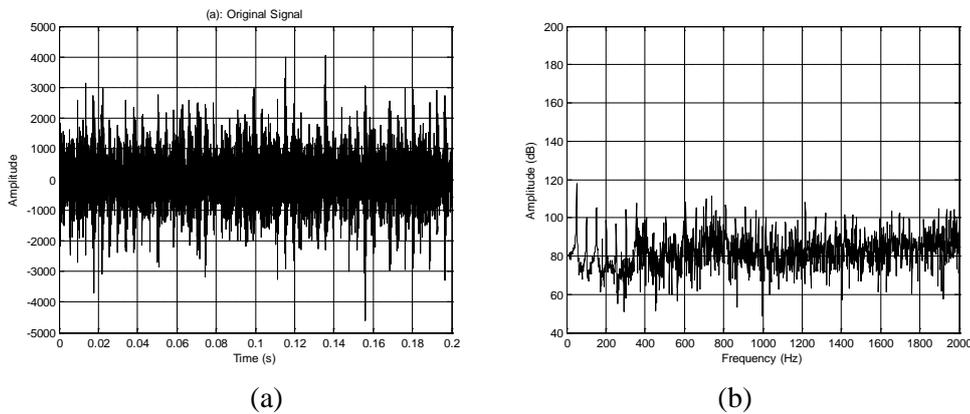


Fig. 2.2 (a) Signal temporel d'un palier avec défaut de roulement, (b) spectre de puissance associé

Une solution partielle à ce problème, la plus intuitive, consiste à associer à un signal non stationnaire une suite de transformées de Fourier à court terme en essayant d'adapter les fenêtres d'observation successives aux variations de structure du signal de telle sorte que les hypothèses de stationnarité, limitatives de ce genre de méthode, soient localement satisfaites. Pour permettre l'accès et la localisation des composants du signal, la transformée de Fourier est modifiée de sorte que pour chaque fraction temporelle, une transformée de Fourier (*FFT*) est appliquée. Le signal est découpé au moyen d'une fenêtre $h(t - b)$ où l'indice b représente le positionnement temporel de cette fenêtre et donc le positionnement du spectre correspondant. Cette nouvelle transformée appelée transformée de Fourier à fenêtre glissante (*STFT*: Short Time Fourier Transform) utilise une fonction fenêtre pour localiser la sinusoïde complexe. Elle est définie comme suit :

$$F(b, \omega) = \int_{-\infty}^{\infty} x(t)h(t - b)e^{-j\omega t} dt \quad (2.8)$$

L'expression $h(t - b)$ est la fonction fenêtre qui confine la sinusoïde complexe $e^{-j\omega t}$. On peut utiliser plusieurs formes de fonctions pour définir la fonction fenêtre, à savoir ; les fenêtres de : Hamming, Hanning, Cosinus, Kaiser, Gausse, ... La série de spectres ainsi constitués représente une forme de transformée temps-fréquence du signal appelé Spectrogramme. Ce traitement fait l'hypothèse de stationnarité durant la durée de la fenêtre h

quelque soit la partie du signal considéré. La longueur de la fenêtre est cependant choisie pour respecter cette hypothèse. Ce choix influence directement les propriétés de résolution de la décomposition ; plus la fenêtre h est petite, plus la résolution temporelle est meilleure mais plus la résolution fréquentielle est mauvaise. Si une haute résolution fréquentielle est nécessaire alors une longue fenêtre temporelle h sera utilisée et il sera difficile de respecter les hypothèses de stationnarité. La forme, la longueur de cette fenêtre ainsi que le pas d'incrément sont des paramètres fixés avant l'analyse. Il est présumé une bonne connaissance *a priori* du signal à analyser. Dans le cas où il s'agit de la fonction fenêtre de Gausse, la *STFT* est appelée particulièrement la transformée de Gabor et la fonction analysante est appelée Gaborette, elle est définie par l'expression suivante :

$$h(t) = \frac{1}{\sqrt{\sigma}\pi^{1/4}} \exp\left(-\frac{t^2}{2\sigma^2}\right) \quad (2.9)$$

Où ; σ est un paramètre fixe qui indique la déviation standard de la fenêtre de Gausse sur l'axe de temps. Le produit de la fenêtre et de la sinusoïde complexe est appelé atome de Fourier fenêtré ou généralement atome temps-fréquence. Il est défini comme suit :

$$h_{f,b}(t) = h(t - b)e^{-j\omega t} \quad (2.10)$$

La *STFT* est donc, le produit de convolution du signal $x(t)$ et du conjugué de l'atome ; nous obtenons ainsi :

$$F(b, f) = \int_{-\infty}^{\infty} x(t)h_{f,b}^*(t)dt \quad (2.11)$$

La transformée de Fourier à fenêtre glissante (*STFT*) est la plus rapide et la méthode la plus simple en termes de mise en œuvre et de la représentation visuelle des résultats. Cependant, la limitation de la taille de la bande passante imposée par la fonction de la fenêtre signifie que si nous voulons obtenir une meilleure résolution dans le temps nous devons sacrifier la résolution en fréquence et vice versa. Le principe d'incertitude de Heisenberg-Gabor traduit l'impossibilité d'avoir un signal dont les supports temporels et fréquentiels sont arbitrairement petits. Il fournit alors une borne inférieure sur l'encombrement temps-fréquence d'un signal. L'inégalité de Heisenberg-Gabor peut donc être vue comme une limitation mathématique qui interdit une localisation arbitrairement petite dans le domaine temporel et fréquentiel. Les figures (2.4, 2.6 et 2.9) illustrent l'influence de la taille de la fenêtre sur la résolution fréquentielle et temporelle du spectrogramme associé.

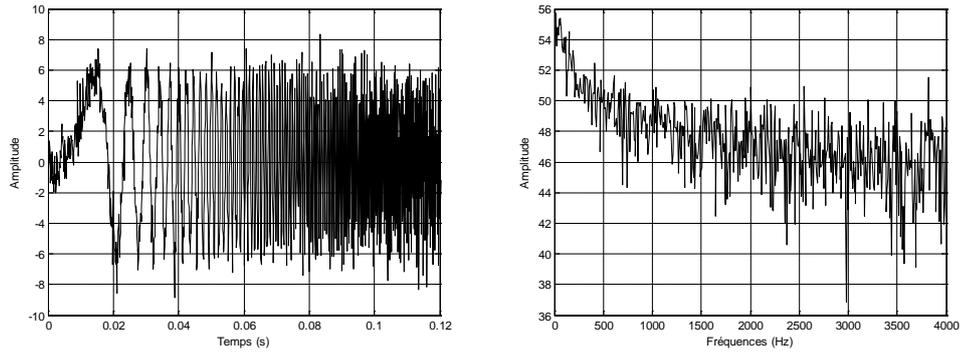


Fig. 2.3 Signal chirp bruité et son spectre associé

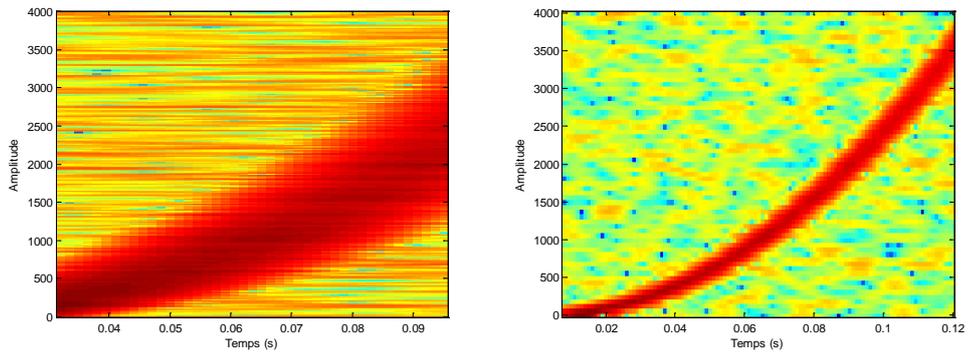


Fig. 2.4 Spectrogramme du Signal chirp bruité avec deux largeurs différentes de la fenêtre temporelle

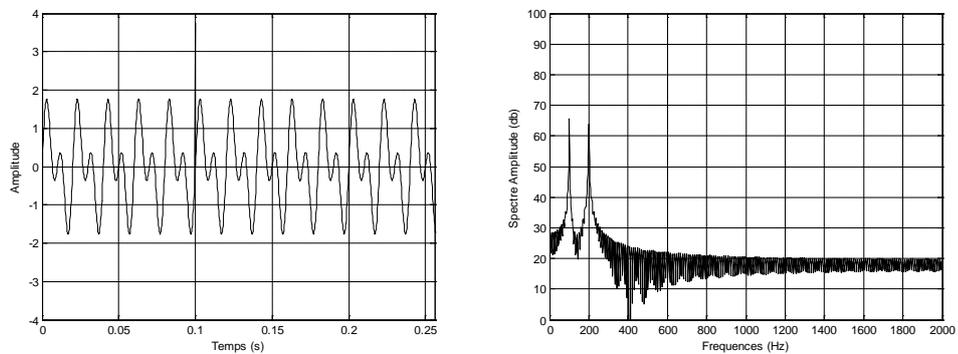


Fig. 2.5 Signal simulé constitué de la somme de deux sinusoides et d'une impulsion

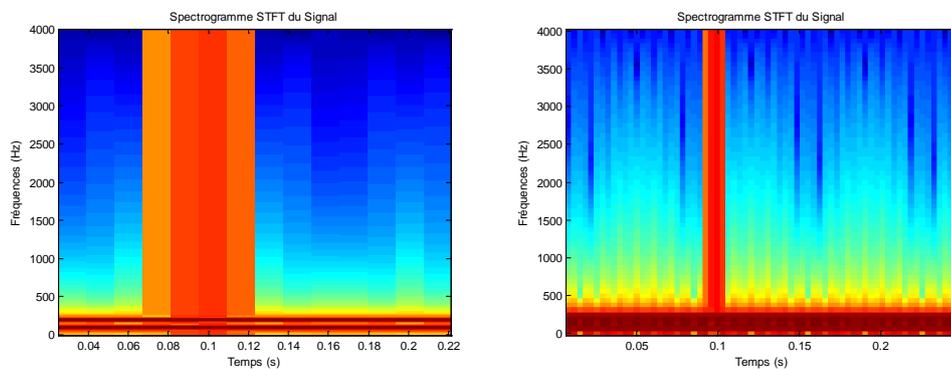


Fig. 2.6 Spectrogramme du signal simulé avec deux différentes largeurs de la fenêtre temporelle

Un autre exemple concernant un train d'impulsion artificiellement corrompu par un bruit gaussien (figure 2.8) est utilisé pour tester l'efficacité du spectrogramme. Les impulsions générées par des roulements à billes endommagés présentent souvent la forme représentée sur la figure (2.7).

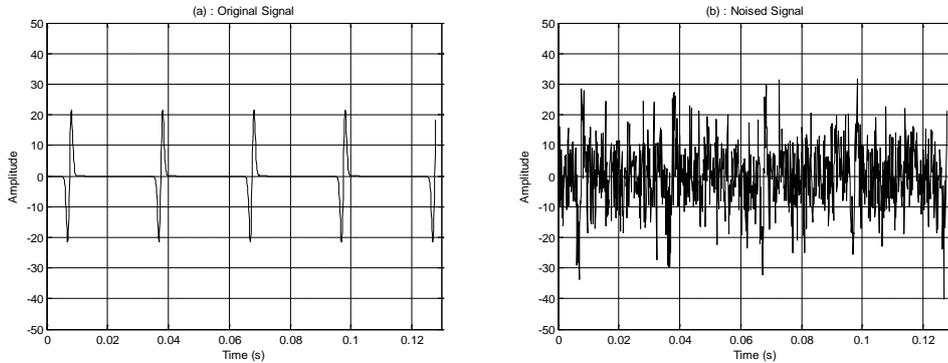


Fig. 2.7 Impulsions simulées **Fig. 2.8** Impulsions simulées avec un bruit lourd

Les figures (2.9) et (2.10) montrent, respectivement, la STFT avec différentes largeurs de la fenêtre $h(t)$, du signal composé d'un train d'impulsions et le même signal corrompu artificiellement par un bruit gaussien. Les deux spectrogrammes ont été calculés en utilisant la fenêtre rectangulaire. Les spectrogrammes donnent une assez bonne description du comportement du signal dans les deux cas. Cependant, une grande différence dans la localisation temps-fréquence de la représentation STFT du signal est évidente si la largeur de la fenêtre change. La limitation de la bande passante imposée par la fonction de fenêtre signifie que si nous voulons obtenir une meilleure résolution dans le temps, nous devons sacrifier la résolution en fréquence et vice versa.

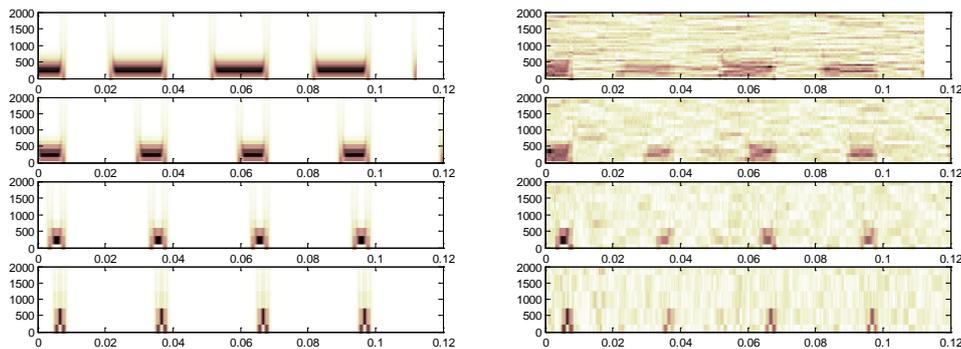


Fig. 2.9 Spectrogramme des impulsions simulées **Fig. 2.10** Spectrogramme des impulsions bruitées

Toutefois, dans le cas d'un signal à composants multiples (figure 2.2a), en raison de la mauvaise résolution temps-fréquence, les composants du signal recueilli à partir du banc d'essais sont mal représentés sur le spectrogramme (figure 2.11). Donner une bonne représentation de tous les composants du signal nécessite une bonne résolution en temps et en fréquence, ceci ne peut être réalisé en utilisant le spectrogramme.

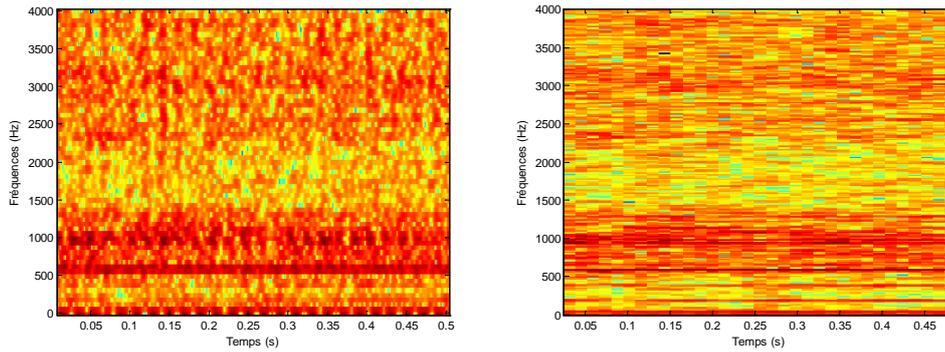


Fig. 2.11 Spectrogrammes du signal d’acquisition du palier avec défaut de roulement calculés avec deux différentes largeurs de la fenêtre temporelle $h(t)$

Une autre limitation est liée à l’utilisation d’une fenêtre unique. Elle peut s’appréhender de manière intuitive en considérant un signal constitué de la somme de plusieurs objets TF occupant des surfaces TF différentes. Les objets définissant une surface approximativement identique à celle de la fenêtre d’analyse seront bien localisés. Si, en revanche, tel n’est pas le cas, les objets TF seront déformés par la fenêtre d’analyse. Un résultat optimal nécessite alors plusieurs fenêtres de dimensions adaptées aux composantes du signal qui sont inconnues a priori.

II.5.1 Adaptativité de la représentation temps-fréquence de type STFT

Il est démontré d’après les exemples illustrés précédemment que les représentations temps-fréquence ne permettent de donner qu’une image relativement ‘subjective’ d’un signal, dans la mesure où celle-ci dépend de choix préalables (type de fenêtre). Se pose donc le problème de la détermination de la ‘meilleure’ représentation pour un signal donné ou une classe donnée de signaux, si tant est que cette notion ait un sens. Il ne s’agit cependant que de cas très particuliers, difficilement généralisables à des situations plus complexes. Nous nous intéressons dans cette section à une approche adaptative, en cherchant donc à adapter à un signal donné une fenêtre appropriée, en fonction d’un critère à optimiser. Plus précisément, nous considérons deux approches pour la sélection de la fenêtre optimale dans le cas d’une représentation de Fourier à court terme : l’une exploitant des critères d’adéquation de la fenêtre au signal étudié, l’autre des critères de parcimonie de la représentation dans l’esprit des algorithmes de (*Basis Pursuit*) [33]. Dans un cadre comme dans l’autre, le domaine d’applicabilité de telles approches s’avère confiné à des classes de signaux simples qui adaptent la représentation à des composantes isolées du signal analysé.

II.5.2 Critères d’optimalité

L’adaptation de la représentation de Fourier à court terme est effectuée en sélectionnant parmi une famille de fenêtres données celle qui optimise un critère donné d’optimalité. On a souvent recours aux deux familles suivantes de critères, qui exploitent deux idées différentes : l’adéquation (‘matching’) de la fenêtre au signal analysé, ou la parcimonie (‘sparsity’) de la représentation obtenue.

a. Critères d'adéquation

La fenêtre h est sélectionnée de manière à ce que sa longueur "étalement" dans le plan temps-fréquence soit le plus proche possible de celui du signal. Comme l'a montré W. Kozek [34], cette démarche aboutit à une relation liant la fonction d'ambiguïté de la fenêtre à celle du signal analysé. Dans les formulations les plus simples, la fenêtre optimale est solution d'un problème d'optimisation du type:

$$h = \arg \max_{\gamma} \left(K * |A_x|^2, |A_{\gamma}|^2 \right) \quad (2.12)$$

où A_x est la fonction d'ambiguïté du signal x et A_{γ} celle de la fenêtre γ .

$$A_x(b, \omega) = \int_{-\infty}^{+\infty} x(t) \bar{x}(t - \tau) e^{-j\theta t} dt \quad (2.13)$$

La fonction d'ambiguïté peut s'interpréter comme une corrélation TF puisqu'elle mesure la ressemblance du signal avec sa version décalée en temps (τ) et en fréquence (θ). On appelle donc la variable τ , le décalage temporel et la variable θ , le décalage fréquentiel. K est un noyau de lissage bidimensionnel. Cette optimisation est alors effectuée numériquement. L'extension au cas des signaux x aléatoires s'effectue en substituant à A_x la 'fonction d'ambiguïté moyenne' $\mathbb{E}\{A_x\}$.

b. Critères de parcimonie

Ce critère concerne la recherche d'une fenêtre qui rend la représentation la plus parcimonieuse, c'est à dire telle que l'énergie du signal soit concentrée sur un petit nombre de coefficients significatifs. Cette démarche est proche de celle utilisée dans des algorithmes de choix de meilleure base [33]. Les mêmes types de critères peuvent être utilisés, par exemple des critères d'entropie. Pour cela, on définit un spectrogramme normalisé ρ_{γ} du signal x par

$$\rho_{\gamma}(\tau, \omega) = \frac{|F_{\gamma}(\tau, \omega)|^2}{2\pi \|x\|^2 \|\gamma\|^2} \quad (2.14)$$

avec F_{γ} la transformée de Fourier à court terme du signal x calculée avec la fenêtre γ . ρ_{γ} est alors une densité de probabilités. On introduit l'entropie de Shannon par

$$H(\gamma) = - \iint \rho_{\gamma}(\tau, \omega) \log(\rho_{\gamma}(\tau, \omega)) d\tau d\omega \quad (2.15)$$

et l'entropie de Rényi par

$$R_{\alpha}(\gamma) = \frac{1}{1 - \alpha} \log \left(\iint \rho_{\gamma}(\tau, \omega)^{\alpha} d\tau d\omega \right), \quad \alpha \in (0, 1) \quad (2.16)$$

La recherche de la fenêtre optimale est effectuée numériquement par l'optimisation paramétrée d'une des deux expressions suivantes : $h = \arg \min H(\gamma)$ ou $h = \arg \min R_\alpha(\gamma)$.

II.5.3 Optimisation globale

Comme on a mentionné précédemment le domaine d'applicabilité de telle approche s'avère confiné à des classes de signaux simples ; et il s'avère que cette approche est rapidement mise en difficulté lorsque la composition temps-fréquence du signal se complexifie. La taille optimale de fenêtre dépend amplement du critère utilisé et des valeurs du paramètre α dans le cas de l'entropie de Rényi ; dès que le signal contient des composantes ayant des caractéristiques temps-fréquence différentes. La fenêtre 'optimale' résulte donc, d'un compromis entre les différentes 'composantes' présentes dans le signal, et des valeurs du paramètre α , (les faibles valeurs de α ont tendance à privilégier la localisation temporelle, alors que les grandes valeurs privilégient la localisation fréquentielle). Ceci limite donc l'intérêt d'une optimisation globale de la fenêtre pour ces signaux.

II.5.4 Optimisation locale dans le plan temps-fréquence

Lorsque l'optimisation globale n'a pas un sens bien défini, l'adaptation peut être rendue locale dans le plan temps-fréquence, limitant ainsi l'influence du critère utilisé. Après identification (et sélection) d'une région d'intérêt du plan temps-fréquence, l'optimisation de fenêtre est effectuée en se limitant aux composantes correspondantes du signal. Plus précisément, étant donnée une fenêtre d'analyse initiale h , un signal correspondant $x_{h,\Omega}$, est reconstruit par

$$x_{h,\Omega}(t) = \frac{1}{2\pi\|h\|^2} \int_{\Omega} F_h(b, \omega) h(t-b) e^{j\omega t} db d\omega \quad (2.17)$$

et la recherche de fenêtre optimale est effectuée comme précédemment, mais à partir du signal $x_{h,\Omega}$, cette fois. De nouveau, la fenêtre optimale pour la région est celle qui optimise le critère choisi, et fournit une représentation temps-fréquence pour le signal (complet) x , qui a été optimisée pour une composante donnée de x . On affine ainsi la représentation non seulement de la composante sélectionnée, mais aussi de toutes les autres composantes présentant des caractéristiques similaires. Dans toutes les situations où sont présentes des composantes 'séparées', des calculs similaires à ceux présentés dans la section précédente peuvent être effectués. Essentiellement, dès qu'une composante peut être isolée, les différents critères fourniront des résultats concordants pour la fenêtre optimale. L'approche locale présentée ici constitue une bonne alternative aux approches globales. Elle permet notamment d'envisager, via des synthèses partielles de composantes que l'adaptation aura permis de définir précisément, des représentations adaptatives faisant intervenir plusieurs fenêtres simultanément. En cela, elle offre une alternative simple et intéressante pour l'amélioration de la résolution dans le plan temps-fréquence.

II.6 La Distribution de Wigner-Ville

II.6.1 Introduction

Les représentations temps-fréquence permettant de caractériser un signal simultanément dans les domaines temporels et fréquentiels, jouent un rôle de plus en plus important pour l'étude des signaux non stationnaires. De telles représentations se heurtent malheureusement à une nécessité de compromis insurmontable entre les résolutions temporelle et fréquentielle. Dans cet esprit, un regain particulier d'intérêt se manifeste pour une transformation proposée en mécanique quantique par E. P. Wigner. Cette transformation permet de définir ce que nous conviendrons d'appeler la distribution de Wigner-Ville (*DWV*) en référence et hommage à J. Ville qui, le premier, introduisit cette même notion en théorie du signal. En effet, elle apparaît comme une représentation temps-fréquence de base possédant des propriétés mathématiques intéressantes. Elle fournit une décomposition temps fréquence sans aucune restriction sur les résolutions temporelles et fréquentielles. Elle apparaît tout à fait adaptée à l'analyse des signaux non stationnaires puisque qu'elle ne nécessite pas d'hypothèses sur le signal lui-même. Un grand nombre de travaux ont permis d'apprécier son comportement théorique vis-à-vis de certaines classes de signaux, son potentiel d'applications et ses particularités par rapport à d'autres représentations de ce type. De plus, son application à des signaux physiques et son interprétation dans le cas des problèmes pratiques ont fait l'objet d'études récentes. La *DWV* transforme un signal monodimensionnel en une représentation bidimensionnelle. La *DWV* est maintenant bien connue mais, quoique ses avantages aient été souvent remarqués, certains des problèmes d'interprétation qu'elle soulève ont semblé freiner son utilisation effective comme alternative avantageuse aux analyses spectrales à court-terme classique. La bilinéarité de cette transformée a des conséquences désastreuses qui se manifestent par l'apparition d'interférences et d'énergies négatives dans la distribution temps-fréquence de l'énergie du signal. Ces phénomènes sont remarquables lorsqu'il y a présence de deux événements distincts en temps et en fréquence. Ce phénomène est totalement irréaliste d'un point de vue physique et laisse entrevoir des difficultés d'interprétation de la *DWV*. C'est pourquoi, afin d'extraire des résultats fiables, il est nécessaire d'éviter ces énergies négatives par un post traitement qui affectera les propriétés de résolution de la transformée. On se propose ici d'en présenter les principes de base ainsi que les modifications nécessaires conduisant à la possibilité d'une telle utilisation. On discute particulièrement le choix des fonctions de lissage à utiliser pour obtenir des performances supérieures à celles des méthodes classiques à base de transformée de Fourier à court terme. L'ensemble de cette étude s'appuie sur des résultats pour la plupart déjà connus, le souci premier étant de les rassembler d'une manière homogène au service d'une présentation générale et d'un mode d'emploi de la méthode.

II.6.2 Définition

Dans ce qui suit nous rappelons la définition de la *DWV* continue et les problèmes liés à sa discrétisation. La *DWV* continue associée à un signal temporel complexe d'énergie finie $x(t)$

une fonction réelle $W_x(t, f)$ du temps t et de la fréquence f définie par l'une des deux définitions équivalentes suivantes :

$$W_x(t, f) = \frac{1}{2\pi} \int_{-\infty}^{\infty} x\left(t + \frac{\tau}{2}\right) x^*\left(t - \frac{\tau}{2}\right) e^{-i2\pi f\tau} d\tau \quad (2.18)$$

$$W_x(t, f) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X\left(f + \frac{\xi}{2}\right) X^*\left(f - \frac{\xi}{2}\right) e^{i2\pi\xi t} d\xi \quad (2.19)$$

Où $X(f)$ représente la transformée de Fourier de $x(t)$. Cette distribution est bilinéaire, car le signal est multiplié par lui-même. $W_x(t, f)$, appelée aussi, Distribution de Wigner-Ville (*DWV*) est une représentation énergétique du signal. Il nous paraît important de mettre en parallèle les définitions de la *DWV* à partir du signal temporel $x(t)$ et à partir de son spectre $X(f)$. En effet, leur similitude fait pressentir les rôles du temps t et de la fréquence f dans cette représentation. Ceci est mis en évidence par l'examen des propriétés de la *DWV* dont nous rappelons ici brièvement quelques-unes de ces propriétés.

II.6.3 Propriétés

Les propriétés importantes pour l'analyse des signaux sont essentiellement la résolution TF et l'interprétabilité. C'est donc sur celles-ci que nous focaliserons notre attention. Les propriétés supplémentaires ne sont pas présentées ici. La résolution TF de la distribution de Wigner-Ville est excellente. La distribution de Wigner-Ville préserve l'encombrement d'une gaussienne défini par le principe d'incertitude. Dans ce sens, elle possède donc une résolution TF maximale [35]. Néanmoins, On ne peut pas garantir la positivité de la distribution de Wigner-Ville en toute généralité. Cette non-propriété engendre des problèmes d'interprétation. Ainsi une valeur négative de la distribution de Wigner-Ville invalide l'interprétation physique directe d'énergie du signal en ce point. D'autre part, les résultats fournis par la distribution de Wigner-Ville sont très satisfaisants dès lors qu'on considère un signal mono-composant de géométrie TF simple. En revanche, dès qu'il s'agit de signaux multi-composants ou de géométrie TF plus compliquée, la distribution de Wigner-Ville fait apparaître des termes d'interférence. Ces termes sont localisés dans le plan TF à des endroits où les signaux qui les engendrent n'ont pas d'énergie. De ce fait, ils compliquent notablement l'analyse jusqu'à la rendre pratiquement impossible. Le mécanisme des interférences sera présenté plus en détail au paragraphe II.6.4.

- **Conservation d'énergie :** La *DWV* est une fonction réelle répartissant l'énergie d'un signal dans le plan temps-fréquence. En intégrant la *DWV* de $x(t)$ sur l'ensemble du plan temps-fréquence on obtient l'énergie de $x(t)$:

$$E_x = \iint_{-\infty}^{\infty} W_x(t, f) dt df = \int_{-\infty}^{\infty} |x(t)|^2 dt \quad (2.20)$$

Ce qui explique que l'on tende à interpréter la DWV comme la densité d'énergie du signal dans le plan temps-fréquence.

- **Propriétés Marginales:** Ses distributions marginales s'identifient à la puissance instantanée et à la densité spectrale de puissance du signal :

$$\int_{-\infty}^{\infty} W_x(t, f) df = |x(t)|^2 \quad (2.21)$$

$$\int_{-\infty}^{\infty} W_x(t, f) dt = |X(f)|^2 \quad (2.22)$$

Toutefois, il est démontré que toutes les propriétés de la DWV continue n'étaient pas préservées par discrétisation du fait de problèmes de recouvrements. Ceci a conduit ainsi les auteurs à proposer de nouvelles définitions permettant de conserver les propriétés de la DWV continue. La prise en charge de ces remarques a permis d'introduire une version de la DWV à la fois discrète en temps et en fréquence, définie par l'une des deux définitions équivalentes :

$$W(n, m) = \frac{1}{2N} \sum_{k=0}^{N-1} z_k z_{n-k}^* e^{-i\pi m/N} (2k - n) \quad (2.23)$$

$$W(n, m) = \frac{1}{2N} \sum_{l=0}^{N-1} Z_l Z_{m-l}^* e^{i\pi m/N} (2l - m) \quad (2.24)$$

Où N est le nombre d'échantillons, z_0, z_1, \dots, z_{N-1} (resp. Z_0, Z_1, \dots, Z_{N-1}) du signal (resp. le spectre) discret et où les indices sont considérés modulo N . Considérant cette définition, la plupart des propriétés de la DWV continue sont préservées.

II.6.4 Mécanisme des interférences

Malgré les propriétés souhaitables de la distribution de Wigner-Ville, il s'avère quelle n'est pas en mesure de produire une représentation satisfaisante de signaux multi-composants en raison de la présence de termes croisés. Le mécanisme des interférences est en fait inhérent à la nature bilinéaire de la distribution de Wigner-Ville. La bilinéarité de la DWV crée dans le plan temps-fréquence des termes d'interaction sans signification physique immédiate selon le principe de superposition quadratique qui est défini comme suit:

$$W_{x+y}(t, f) = W_x(t, f) + W_y(t, f) + 2\Re_e\{W_{x,y}(t, f)\} \quad (2.25)$$

$$W_{x,y}(t, f) = \int_{-\infty}^{\infty} x\left(t + \frac{\tau}{2}\right) y^*\left(t - \frac{\tau}{2}\right) e^{-i2\pi f\tau} d\tau \quad (2.26)$$

Les deux premiers sont les distributions de Wigner-Ville des signaux $x(t)$ et $y(t)$ que l'on désigne par auto-termes et le troisième, qui provient de l'interaction entre les deux

composantes $x(t)$ et $y(t)$, constitue le terme d'interférence. Ce principe peut être généralisé à N composants. Les termes d'interférence compliquent la lecture de la représentation de la DWV . Les interférences présentes dans la DWV rendent difficile l'interprétation du signal. Cependant, si on se fixe comme objectif la construction d'une distribution TF à interférences réduites, il convient dans un premier temps de comprendre le mécanisme de création des interférences, pour être à même, dans un second temps, de choisir le noyau de manière adéquate par un lissage approprié. Une étude attentive de ces limitations indique en fait quelles modifications apporter à la DWV pour rendre son utilisation plus souple et ses résultats plus facilement interprétables. Les figures ci-dessous montrent l'effet de 'termes croisés' dans la DWV . La figure (2.12) montre la DWV du signal composé d'un train d'impulsions et la figure (2.13) le même signal corrompu artificiellement par un bruit gaussien. D'autres impulsions ont été ajoutées au signal dans les deux cas. Bien qu'il y ait une bonne représentation du comportement des deux signaux. Le seul avantage évident que la DWV a sur la STFT est sa meilleure résolution temps-fréquence. La concentration de l'énergie à la fréquence instantanée et d'autres propriétés souhaitables théoriquement fait de la DWV également un outil préféré à la STFT dans de nombreuses applications de traitement du signal.

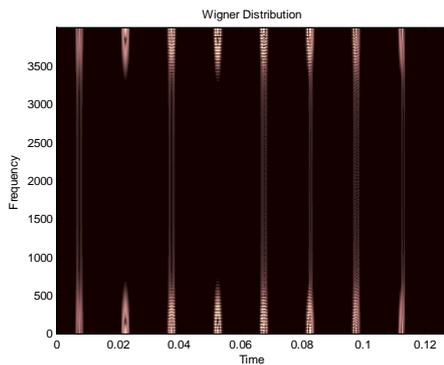


Fig. 2.12 WVD des impulsions simulées

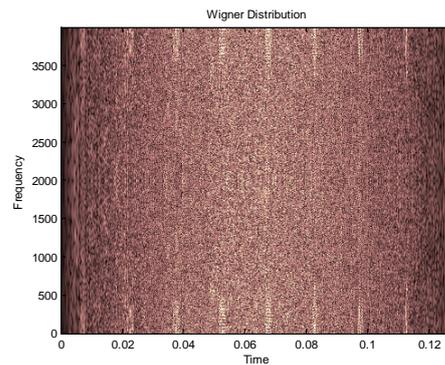


Fig. 2.13 WVD des impulsions bruitées

II.6.5 Lissage de la distribution de Wigner Ville

L'approche la plus cohérente pour la réduction des termes d'interférences consiste à utiliser les connaissances disponibles sur leur mécanisme de création pour développer une stratégie appropriée. Compte tenu de l'aspect oscillant des termes d'interférences, une piste peut être exploitée en lissant la distribution de Wigner-Ville par un filtre passe-bas bidimensionnel. Un exemple typique est donné par la distribution de Choi-Williams [3]. Néanmoins, le point de vue adopté dans cet article, pour la réduction des interférences, consiste à définir un lissage fixe dont l'action est identique en tout point TF en ne privilégiant aucune orientation donnée dans le plan TF. Une autre alternative consiste à paramétrer une fonction de lissage (ou un noyau), et à déterminer pour chaque point TF, la valeur optimale selon un critère donné, du paramètre libre. Dans ce sens, on peut parler d'une adaptation locale du lissage.

$$C_x(t, f, F) = \iint_{-\infty}^{\infty} W_x(\hat{t}, \hat{f}) F(t - \hat{t}, f - \hat{f}) d\hat{t} d\hat{f} \quad (2.27)$$

où F est une fonction de lissage normalisée avec $\iint_{-\infty}^{\infty} F(t, f) dt df = 1$. Selon les propriétés bidimensionnelles de cette fonction, la distribution de Wigner-Ville lissée (*DWVL*) diminue les niveaux d'énergies négatives et fournit suivant les cas une distribution d'énergie complètement positive. Cette condition est souhaitée quel que soit le signal. D'un autre côté, le lissage doit préserver le caractère local du signal. Des compromis sont alors à trouver. Le noyau gaussien suivant est souvent utilisé comme fonction de lissage de la *DWVL*:

$$F(t, f) = \exp -(\gamma t)^2 \exp -(2\pi\delta f)^2 \quad \text{avec} \quad \gamma\delta = 1 \quad (2.28)$$

Là encore, les propriétés de telles représentations ont été abondamment discutées pour traduire sur la fonction F les contraintes imposées par l'utilisateur ; nous n'y reviendrons pas ici. Le point important est qu'on peut montrer que la représentation ainsi formulée définit la classe la plus générale des représentations énergétiques bilinéaires admissibles, sous la seule contrainte d'invariance vis-à-vis des translations dans le plan temps-fréquence. En revanche, ceci permet de retrouver la plupart des représentations proposées antérieurement comme cas particuliers. Citons le cas des spectrogrammes, pour lesquels la fonction de lissage de la *DWV* s'identifie par la fenêtre d'observation temporelle :

$$F_s(t, f) = W_h(t, f) \quad (2.29)$$

$$C_x(t, f, F_s) = \left| \int_{-\infty}^{\infty} x(u) h^*(t - u) e^{-i\omega u} du \right|^2 \quad (2.30)$$

Le choix d'une pondération détermine donc crucialement la méthode d'analyse retenue comme approximation lissée de la *DWV* initiale. De plus, les propriétés de la représentation résultante sont en un sens résumées dans la pondération choisie, ce qui, en retour, guide dans le choix de cette dernière. En tout état de cause, quelque soit l'approche retenue, il convient de souligner que le lissage imposé pour la réduction des interférences se traduit en contrepartie par une perte de résolution TF. On sait bien que les méthodes du type spectrogrammes sont fondamentalement limitées par la nécessité d'un compromis entre les résolutions temporelle et fréquentielle: améliorer la résolution fréquentielle ne peut se faire qu'au prix d'un élargissement de la durée d'observation, i.e. d'une diminution de la résolution temporelle dans une situation non stationnaire. Ceci provient directement du fait que la fonction de lissage associée à la *DWV* (fenêtre d'observation), possède une répartition temporelle et fréquentielle ne pouvant être contrôlée séparément. On montre en effet que, pour tout signal h d'énergie unité sur toute la durée T , l'inégalité d'Heisenberg-Gabor se traduit sur la *DWV* par la relation :

$$\iint_{-\infty}^{\infty} \left[\frac{(t - t_0)^2}{T^2} + T^2(f - f_0)^2 \right] W_h(t, f) dt df \geq \frac{1}{2\pi} \quad (2.31)$$

Où t_0 et f_0 sont le temps et la fréquence moyennes du signal h . Les résolutions temporelles et fréquentielle étant ainsi liées.

II.7 Conclusion

Dans ce chapitre, les idées principales de l'analyse temps-fréquence (TF) sont présentées, et l'intérêt d'une telle approche est justifié sur la base du modèle des signaux vibratoires utilisés. En comparant les résultats obtenus à partir de l'analyse des différents signaux de vibration, nous pouvons conclure que: L'analyse temps-fréquence présente des avantages certains comparée à l'analyse des vibrations séparément en fonction du temps ou dans le domaine fréquentiel. Ces avantages font de cette méthode un outil intéressant dans la surveillance des machines. Par ailleurs, la STFT peut donner une représentation satisfaisante d'un signal dans le plan temps-fréquence à condition qu'une longueur appropriée de la fenêtre temporelle soit choisie pour découper le signal. Le problème d'amélioration de la résolution indispensable à la lisibilité de la représentation STTF y est aussi étudié. Pour pallier les inconvénients de manque de précision, il est proposé l'introduction un noyau (atome) optimal. A notre sens, ceci est actuellement ce qui fournit le meilleur compromis entre la qualité des résultats et le coût de calcul. Cette approche a cependant des limitations qui sont également mises en évidence. D'autre part, les résultats expérimentaux obtenus de l'analyse des signaux synthétiques bruités, montrent que le seul avantage évident de la *DWV* comparée à la *STFT* est sa meilleure résolution temps-fréquence. La concentration de l'énergie à la fréquence instantanée et d'autres propriétés théoriques souhaitables font de la *DWV* également préférable à la *STFT*. L'ensemble des propriétés de la *DWV* confère à la transformation de Wigner-Ville un rôle central comme outil d'analyse dans le plan temps-fréquence. A l'instar de toute méthode, son utilisation nécessite cependant des précautions d'emploi et d'interprétation. On a essayé ici de cerner l'essentiel de celles-ci, justifiant à la fois les limitations propres des analyses à court terme classiques et l'introduction pour la *DWV* d'une approximation satisfaisante et souple d'emploi: la *DWVL*. Celle-ci, outre le fait qu'elle exploite au mieux les deux degrés de liberté du plan temps-fréquence, permet une justification simple de ses bonnes propriétés de résolution, guidant ainsi dans le choix des paramètres d'analyse. D'une manière un peu schématique, on peut considérer que la transformation de Wigner-Ville permet de dépasser les limitations propres aux traitements linéaires de situations non stationnaires, en forçant en quelque sorte une 'stationnarisation' locale par une opération non linéaire. Le prix à payer pour cette dernière (valeurs négatives de la représentation et termes d'interférence) peut alors être minimisé par un choix adéquat des lissages appropriés.

Chapitre III

Décomposition en ondelettes et analyse multi-résolution (AMR)

Chapitre 3

Décomposition en ondelettes et analyse multi-résolution (AMR)

Résumé

La transformée en ondelettes permet de réaliser une analyse des structures locales d'un signal avec un zoom qui dépend de l'échelle considérée. Les fondements mathématiques de cette étude sont développés dans ce chapitre, dans lequel sont détaillés le principe de base de l'analyse et les avantages d'une telle décomposition. Ce chapitre, essentiellement bibliographique, introduit successivement les éléments fondamentaux des analyses en ondelettes et des analyses multi-résolutions (AMR). Une attention particulière est portée sur les fonctions analysantes et aborde successivement les analyses en ondelettes continues puis les analyses en ondelettes discrètes. Nous verrons que l'analyse multi-résolution permet de mettre en relief les échelles caractéristiques d'un système en le décomposant dans la base des fonctions d'ondelettes. La transformée en ondelettes s'est avérée particulièrement utile pour l'analyse des signaux qui sont considérés apériodique, bruités, intermittents, transitoires, et autres. Sa capacité d'examiner le signal simultanément dans le plan temps-fréquence est bien meilleure de celle des autres méthodes traditionnelles telle que la transformée de Fourier à court terme (STFT).

Sommaire

III.1	Introduction	39
III.2	Ondelette analysante (Atome temps-fréquence).	40
III.3	Conditions d'admissibilité	41
III.4	La transformée d'ondelettes continue	42
III.5	Formule d'inversion.	44
III.6	Base d'ondelettes.	44
III.7	Quelques propriétés.	45
III.8	Transformée en Ondelettes Discrète.	45
III.9	L'analyse multi-résolution.	47
III.10	Conclusion	56

III.1 Introduction

Les premiers travaux concernant l'analyse en ondelettes sont accordés à Morlet qui en 1975, dans ses recherches, tente d'apporter des réponses à la localisation de couches pétrolifères. Il se trouve confronté à la difficulté d'identifier la composition fréquentielle instantanée d'un signal. L'outil phare de l'époque se trouve être l'analyse de Fourier particulièrement bien adaptée à l'analyse des phénomènes stationnaires mais qui n'apporte pas de réponse satisfaisante quant à l'étude d'évènements transitoires comme les réflexions d'une onde de choc sur une frontière entre deux couches de sol. Dans ses recherches sur l'identification de phénomènes transitoires, Morlet envisage dans un premier temps de travailler sur le signal lui-même en considérant des fenêtres de plus en plus restreintes pour réaliser une étude de plus en plus locale mais sans succès notable. Dans un second temps, il s'inspire des travaux de Gabor proposant une analyse de Fourier à fenêtre glissante dans laquelle les limitations temporelles sont appliquées au signal analysant. Quoique pertinente à l'époque, l'analyse de Gabor souffre cependant d'une imprécision de localisation en temps relative aux événements de fréquences élevées ainsi que de l'absence d'une propriété d'involution (analyse/synthèse) classique aux analyses de Fourier. Morlet entreprend alors de développer des motifs analysants résolument tournés vers l'analyse et la synthèse d'évènements transitoires. En basant sa stratégie de recherche sur la construction de motifs analysants caractérisées par un nombre d'oscillations donné dans une fenêtre temporelle, en effectuant des dilatations et des translations du support d'une unique fonction de référence. Ainsi, Morlet, aidé de Grossmann, donne naissance au concept d'ondelettes. Les analyses effectuées avec de tels motifs lui fournissent à partir d'un signal unidimensionnel un champ de valeurs bidimensionnel à partir duquel il peut reconstruire le signal via une intégrale double. Cette stratégie de construction de fonctions analysantes permet de décrire continûment le plan temps-fréquence et d'envisager la complétude d'une analyse de $L^2(\mathbb{R})$. Une analyse en ondelettes peut être perçue comme un microscope mathématique. La puissance de résolution de ce microscope dépend alors d'une unique fonction à partir de laquelle vont être construits les atomes temps-fréquence fonctionnant comme les messagers de l'information transitoire siégeant dans un signal. Au delà d'un formalisme de transformation intégrale, les véritables contributions apportées par la prolifique littérature ondelette sont les définitions de ces fonctions alors adaptées à des problèmes précis. Leurs facultés à ne pas percevoir les mouvements constants leur ont données cette dénomination d'ondelette. Le coût et la précision des calculs mobilisés limitent néanmoins les performances de l'outil. En 1985 Jean Morlet, motivé par la détection de singularité dans des signaux sismiques, et aidé par un mathématicien, Alex Grossman formalisent le concept mathématique d'ondelettes et donnent un nouvel élan aux analyses en ondelettes en définissant le cadre des analyses multirésolutions des ondelettes orthogonales. La formalisation de l'analyse a été faite ensuite par quelques mathématiciens jusqu'aux travaux en commun de Stéphane Mallat (motivé par des considérations pratiques en imagerie) et Yves Meyer qui ont généralisé l'analyse en ondelettes à l'analyse multirésolution basée sur une hiérarchie d'espaces d'approximation imbriqués, ce qui a conduit à des algorithmes rapides réalisant la transformée discrète en ondelettes. Puis Ingrid Daubechies a généralisé l'analyse multirésolution en proposant l'obtention de bancs de filtres de tailles finies

conduisant à des ondelettes à support compact performantes numériquement. La découverte de ces fonctions non analytiques, construites par des algorithmes récursifs rapides, marque l'avènement des analyses multirésolutions dyadiques engendrées par les ondelettes à support compact. Les quelques personnalités, J. Morlet, A. Grossmann, Y. Meyer, S. Mallat, I. Daubechies et A. Cohen constituent le tronc fondamental de la 'Saga Ondelettes' à partir duquel la généalogie des auteurs ayant contribué à cet édifice se ramifie à la mesure de l'intégration de l'outil ondelette dans les domaines de la science de l'ingénieur. Quelques détails concernant l'histoire de cette épopée mathématique des années 1980-1990 se trouvent dans l'ouvrage de vulgarisation en Ref. [36]. Nous présentons dans ce chapitre les analyses multirésolutions de type ondelettes. Cette présentation ne se veut pas exhaustive. Le lecteur peut se référer à la Ref. [37] pour un *tour* plus complet du sujet. Quelques exemples d'applications sont passés en revue. Les connexions, essentiels dans notre étude, sont introduites à cette occasion et leurs expressions sont données pour aborder dans les chapitres qui suivent, l'utilisation des ondelettes à des fins d'estimation des signaux vibratoires bruités.

III.2 Ondelette analysante (Atome temps-fréquence)

Soit une fonction $\psi \in L^2(\mathbb{R})$ de positions moyennes (τ_ψ, ξ_ψ) et de variances $(\sigma_{\tau,\psi}, \sigma_{\xi,\psi})$ et fréquentielle finies, soient $(a, b) \in \mathbb{R}^+ \times \mathbb{R}$ on appelle atome temps-fréquence le motif dilaté-translaté $\psi_{a,b}$ tel que :

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \quad (3.1)$$

avec $\frac{1}{\sqrt{a}}$ est un facteur de normalisation qui assure que $\psi_{a,b}$ a une norme ℓ_2 égale à 1. Alors la boîte de Heisenberg du motif dilaté translaté $\psi_{a,b}$ définit un rectangle du plan temps-fréquence centré sur le point $(a\tau_\psi + b, f/a)$ de résolution temporelle $a\sigma_{\tau,\psi}$ et de résolution fréquentielle $\sigma_{\xi,\psi}/a$. La figure (3.1b) représente respectivement les boîtes de Heisenberg d'un atome temps-fréquence basse fréquence et d'un atome temps-fréquence haute fréquence.

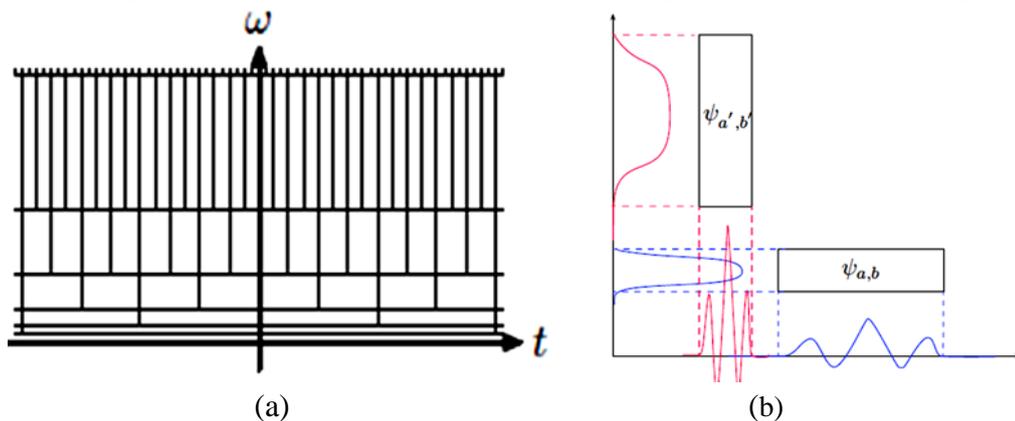


Fig. 3.1 (a) Pavages du plan temps-fréquence de la représentation en ondelettes (b) Boîtes de Heisenberg d'un atome temps-fréquence $\psi_{a,b}$ basse fréquence et d'un atome temps-fréquence $\psi_{a',b'}$ haute fréquence

Le paramètre b est un paramètre de position de l'ondelette le paramètre a est un paramètre d'échelle qui est inversement proportionnel à la fréquence.

- Si $a > 1$ (grand) l'ondelette sera plus étalée et correspondra à une fréquence plus faible, la résolution devient alors bonne en fréquence et mauvaise en temps.
- Quand $a < 1$ (petit), l'ondelette sera plus contractée et correspondra aux fréquences plus élevées que celle de l'ondelette mère (figure 3.1b).

La transformation en ondelettes s'adapte aux fréquences étudiées : elle offre, à la fois, une bonne résolution fréquentielle en basses fréquences et une bonne résolution temporelle en hautes fréquences.

Remarque 3.1: On peut comparer les différents effets de la transformée de Fourier à fenêtre glissante et de la transformée en ondelettes sur les boîtes d'Heisenberg. La transformée de Fourier à fenêtre glissante choisit le centre de la boîte, mais par contre, la forme et la surface sont imposées : Pour la transformée en ondelettes, on choisit la localisation temporelle et la forme de la boîte tandis que la surface de la boîte et la fréquence moyenne sont imposées, la surface de la boîte est de toute façon minorée par le principe d'incertitude d'Heisenberg ; par contre la fréquence moyenne est liée à l'ondelette mère choisie.

III.3 Conditions d'admissibilité

Une ondelette est une fonction $\psi(t)$ qui a la forme d'une petite onde localisée en temps et doit satisfaire certains critères mathématiques. Il existe un grand nombre d'ondelettes qui peuvent être utilisées dans l'analyse des signaux. La meilleure pour une application donnée dépend en même temps de la nature du signal et de l'information à extraire du signal. Les critères mathématiques que doit satisfaire une ondelette sont les suivants :

1. Une ondelette doit avoir une énergie finie :

$$E = \int_{-\infty}^{+\infty} |\psi(t)|^2 dt < \infty \quad (3.2)$$

2. La condition d'admissibilité suivante doit être satisfaite :

$$C_\psi = \int_0^\infty \frac{|\hat{\psi}(\xi)|^2}{|\xi|} d\xi < \infty \quad (3.3)$$

Avec : $\hat{\psi}(\xi)$ est la transformée de Fourier de $\psi(t)$, i.e : $\hat{\psi}(\xi) = \int_{-\infty}^{+\infty} \psi(t)e^{-i\omega t} dt$. Ce qui implique qu'on doit avoir : $\hat{\psi}(\xi = 0) = 0$, autrement dit ; l'ondelette $\psi(t)$ doit avoir une moyenne nulle. La constante C_ψ est appelée la constante d'admissibilité.

$$\hat{\psi}(0) = 0 \Leftrightarrow \int_{\mathbb{R}} \psi(t) dt = 0 \quad (3.4)$$

3. Un critère additionnel pour les ondelettes complexes est que sa transformée de Fourier doit être réelle et s'annule pour les fréquences négatives.

Les ondelettes qui satisfassent les conditions d'admissibilité sont en fait, des filtres passe-bande. Ceci signifie, en termes simples, qu'elles ne laissent passer à travers que les composants d'une bande fine de fréquences.

Régularité : Une ondelette ψ est dite de classe C^m , $m \in \mathbb{N}$ si elle satisfait les propriétés suivantes :

- a. ψ ainsi que toutes ses dérivées, jusqu'à l'ordre m , sont à décroissance rapide à l'infini.
- b. $\int_{-\infty}^{+\infty} t^k \psi(t) dt = 0$ pour $0 \leq k \leq m$

Ces conditions expriment respectivement la régularité, la localisation et le caractère oscillant de l'ondelette. Aussi impose-t-on souvent des contraintes supplémentaires de régularité, de décroissance rapide ou de compacité suivant le besoin [38].

III.4 Transformée en ondelettes continue

De façon analogue à l'analyse de Fourier, les analyses en ondelettes réalisent la transformation intégrale d'un signal donné par projection sur un ensemble (continu ou dénombrable) de fonctions. Dans le cas des analyses en ondelettes, ces dernières fonctions sont toutes déduites par dilatation et translation d'une unique fonction de référence localisée en temps et en fréquence. De telles analyses fournissent une caractérisation d'un signal donné dans le plan temps-fréquence. Soit une ondelette ψ , on définit la transformée en ondelettes continue d'un signal par l'application linéaire χ_ψ donnée par la formule suivante :

$$\chi_\psi(a, b) = \frac{1}{\sqrt{C_\psi}} \int_{-\infty}^{\infty} x(t) \psi_{a,b}^*(t) dt \quad (3.5)$$

Cette transformation intégrale peut s'interpréter comme le produit scalaire d'un signal $x(t)$ de avec l'ensemble des atomes temps-fréquence $\psi_{a,b}$ obtenus par dilatation et translation d'une ondelette admissible ψ pour des coefficients de résolution a et de translation b . Cette transformée est dite continue car c'est une fonction continue de paramètres de dilatation et de translation et on verra que ces paramètres peuvent être discrétisés ; on obtient alors la transformée en ondelettes discrète. La transformée en ondelettes continue renseigne sur le comportement du signal autour du point b dans un voisinage (intervalle) de taille qui est de l'ordre de a . Toutefois, le mécanisme d'analyse de l'information transitoire contenue dans le signal $x(t)$ par les atomes temps-fréquence s'explique en remarquant que l'ondelette ψ est un

filtre passe-bande et que la voie de coefficients ondelettes $\chi_{\psi}(a, \cdot)$ s'exprime comme une convolution du signal $x(t)$ et du motif ψ dilaté à la résolution a . Le coefficient $\chi_{\psi}(a, b)$ est alors appelé coefficient ondelettes ou coefficient de détails de résolution a et de translation b . L'astérisque (*) indique que le conjugué complexe de la fonction d'ondelette qui est utilisé. Les paramètres a et b permettent la dilatation (échelle) et la translation (retard) sur l'axe des temps de l'ondelette. La transformée d'ondelette est appelée communément 'microscope mathématique' qui peut 'zoomer' sur les composantes du signal par le biais du paramètre de l'échelle a et le déplacement le long de l'axe du temps par le biais du paramètre b .

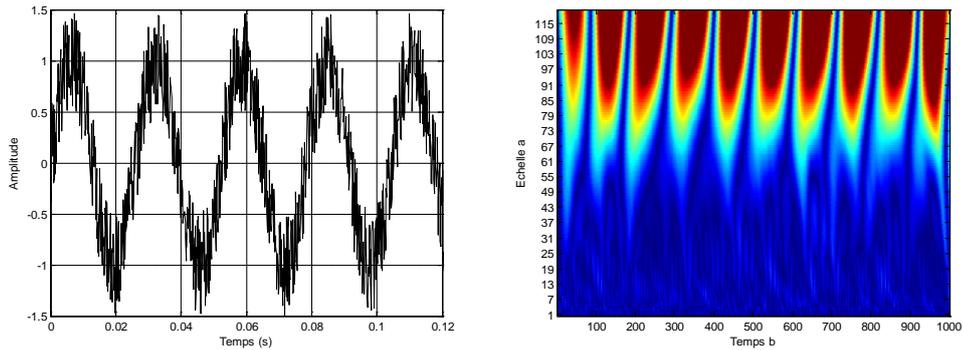


Fig. 3.2 Sinusoïde bruitée et Scalogramme associé

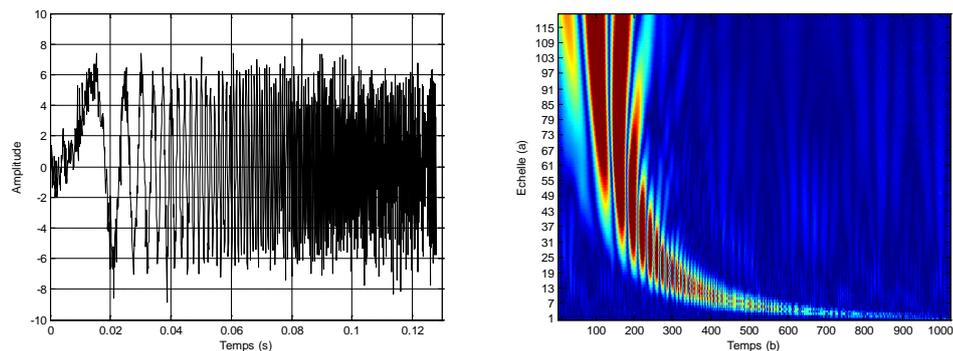


Fig. 3.3 Signal chirp bruité et Scalogramme associé

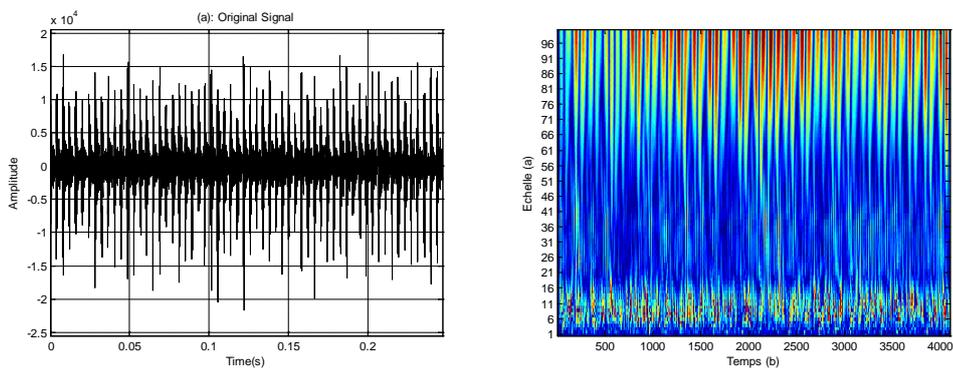


Fig. 3.4 Signal temporel d'un palier avec un défaut de roulement et Scalogramme associé

III.5 Formule d'inversion

Soit ψ une ondelette qui en plus des propriétés oscillante et de normalisation, vérifie la condition d'admissibilité, ce qui signifie que la transformée en ondelettes possède la propriété de conservation de l'énergie, c'est à dire qu'il n'y a pas de perte d'information entre la fonction $x(t)$ et sa transformée en ondelettes. Cette propriété est très importante car elle garantit que la transformée en ondelettes possède des propriétés de stabilité. En partant d'un signal qui a une caractéristique physique donnée (énergie finie), on obtient une fonction qui a les mêmes propriétés. La formule de synthèse (de reconstruction) est définie par l'expression suivante :

$$x(t) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_0^{\infty} \chi_\psi(a, b) \psi_{a,b}(t) \frac{dadb}{a^2} \quad (3.6)$$

Cela signifie que le signal d'origine peut être reconstruit à partir de la transformée d'ondelette par intégration sur l'ensemble des échelles et des localisations, a et b . Si on limite l'intégration seulement sur une bande spécifique d'échelles a , alors on opère ainsi un filtrage du signal original. Dans la pratique, il est parfois intéressant de faire une décomposition du signal par une ondelette analysante g et de reconstruire le même signal à partir d'une autre ondelette h différente de g , dans ce cas nous avons :

La décomposition d'un signal $x(t)$ avec l'ondelette g :

$$\chi_\psi(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) g^* \left(\frac{t-b}{a} \right) dt \quad (3.7)$$

et la synthèse avec une autre ondelette de reconstruction h :

$$x(t) = \frac{1}{C_{g,h}} \int_{-\infty}^{\infty} \int_0^{\infty} \chi_g(a, b) h \left(\frac{t-b}{a} \right) \frac{dadb}{a^2} \quad (3.8)$$

La condition d'admissibilité dans ce cas sur les fonctions g et h ($g, h \in L^2(\mathbb{R})$) est donnée par:

$$C_{g,h} = \int_0^{\infty} \frac{\hat{g}(\xi) \hat{h}(\xi)}{|\xi|} d\xi < +\infty \quad (3.9)$$

III.6 Base d'ondelettes

La transformée en ondelettes continue d'un signal est une représentation de celui-ci sur la famille non dénombrable $(\psi_{a,b})_{a>0, b \in \mathbb{R}}$, alors la famille $(\psi_{a,b})_{a>0, b \in \mathbb{R}}$ est une base continue de $L^2(\mathbb{R})$. Cette représentation est continue et infiniment redondante [39]. La redondance d'information est un désavantage en matière de compression de données, elle est en revanche très appréciée en analyse des vibrations. Une telle propriété est fondamentale pour l'analyse des données en particulier des vibrations où la redondance d'information est souhaitable,

comme nous le verrons plus tard, en plus une telle base conduit en théorie des ondelettes à l'introduction de bases d'ondelettes orthogonales non redondante.

III.7 Quelques propriétés

Linéarité : La transformée en ondelettes est linéaire, c'est à dire :

$$\chi_{\alpha x_1 + \beta x_2}(a, b) = \chi_{\alpha x_1}(a, b) + \chi_{\beta x_2}(a, b) \quad \forall (\alpha, \beta) \in \mathbb{C}^2, \quad \forall (x_1, x_2) \in \mathbb{R}^2 \quad (3.10)$$

Translation : Nous avons la relation de translation suivante :

$$\chi_{\tau_{b'}x}(a, b) = \chi_x(a, b - b') \quad (3.11)$$

Où $\tau_{b'}x$ désigne la translation de x par le vecteur b' : $\tau_{b'}x(t) = x(t - b')$.

Cette propriété d'invariance par translation de la transformée en ondelettes continue est très importante dans l'analyse de certains signaux tels que les signaux de vibrations et les signaux sismiques où il est intéressant en pratique de faire une comparaison des transformées en ondelettes du signal et le même signal décalé dans le temps.

Dilatation : Pour une normalisation ℓ_1 , de la famille d'ondelettes, nous avons la relation d'invariance par dilatation :

$$\chi_{\delta_k x}(a, b) = \chi_x(ka, kb) \quad (3.12)$$

Avec $\delta_k x(t) = x(kt)$

III.8 Transformée en Ondelettes Discrète

Dans les calculs pratiques, les variables a et b de la transformée en ondelettes continue ne varient pas continument dans \mathbb{R} , mais sont discrétisées, ainsi la transformée en ondelettes discrète consiste à trouver une sous famille dénombrable $(\psi_{a_j, b_k})_{a_j, b_k \in \mathbb{R}}$ de la famille $(\psi_{a, b})_{a, b \in \mathbb{R}}$ et formant une base de $L^2(\mathbb{R})$. Deux algorithmes (rapides) sont envisageables :

- On prend $a_j = 2^j, b_k = k2^j, j, k \in \mathbb{Z}$.

C'est l'algorithme de S. Mallat sur lequel nous reviendrons plus en détail dans le cadre de l'analyse multirésolution.

- On prend $a_j = 2^j, b_k = k, j, k \in \mathbb{R}$.

C'est l'algorithme à trous, en raison de non orthogonalité des familles d'ondelettes, dans ce cas, la transformée en ondelettes sera redondante et également invariante en translation. Un tel algorithme est particulièrement adapté aux applications où la redondance est souhaitable.

Morlet a proposé de construire des bases ou des frames de fonctions construites, en utilisant une discrétisation logarithmique des coefficients a et b , selon le modèle suivant :

$$\psi_{j,k}(t) = \frac{1}{\sqrt{a_0^j}} \psi\left(\frac{t - kb_0 a_0^j}{a_0^j}\right) \quad (3.13)$$

Où les valeurs entières j et k contrôlent la dilatation et la translation de l'ondelette. La transformée d'ondelette d'un signal continu $x(t)$, en utilisant les ondelettes discrètes, est:

$$d_{j,k} = \int_{-\infty}^{\infty} x(t) \psi_{j,k}(t) dt = \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{a_0^j}} \psi(a_0^{-j} t - kb_0) dt \quad (3.14)$$

Dans le cas de la transformée d'ondelette discrète, les valeurs de $d_{j,k}$ sont appelés les coefficients d'ondelettes ou coefficients du détail. Pour déterminer la qualité de la décomposition du signal dans l'espace des ondelettes, nous introduisons alors, la théorie des frames qui fournit un cadre structurel pour l'étude des ondelettes discrets. Les frames d'ondelettes sont construits par discrétisation des paramètres d'échelle et de translation comme il a été introduit ci-dessus. La famille des fonctions d'ondelettes qui constitue une frame est telle que l'énergie résultante des coefficients d'ondelettes auxquelles sont liés est limitée dans une gamme de l'énergie du signal original, i.e.

$$AE \leq \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} |d_{j,k}|^2 \leq BE \quad (3.15)$$

Où $d_{j,k}$ sont les coefficients d'ondelettes discrets, A et B sont les limites de la structure (frame), et E est l'énergie du signal original $x(t)$. Les valeurs de A et B dépendent des paramètres a_0 et b_0 choisis précédemment pour l'analyse. Si $A = B$ la formule de reconstruction du signal est la suivante :

$$x(t) = \frac{1}{A} \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} d_{j,k} \psi_{j,k}(t) \quad (3.16)$$

Cependant, si $A = B = 1$ la famille d'ondelette définie ainsi constitue une base orthogonale. Un choix commun de $a_0 = 2$ et $b_0 = 1$ est largement utilisé en littérature. Le choix de l'échelle logarithmique (puissance de 2) pour la dilatation et la translation définit la transformée dyadique. Cette discrétisation permet la construction d'une base d'ondelette ortho-normale. En substituant $a_0 = 2$ et $b_0 = 1$ dans l'équation d'ondelette, nous obtenons l'expression suivante :

$$\psi_{j,k}(t) = \frac{1}{\sqrt{2^j}} \psi\left(\frac{t - k2^j}{2^j}\right) = 2^{-j/2} \psi(2^{-j} t - k) \quad (3.17)$$

III.9 L'analyse multirésolution (AMR)

Cette représentation débute par l'introduction des bases orthonormées d'ondelettes qui sont associées ou non à la notion d'analyse multirésolution. Cette notion implique l'imbrication d'espaces d'approximations, et elle permet de définir des algorithmes performants de transformée en ondelettes rapides. Sans ces performances numériques, les analyses multirésolutions seraient beaucoup moins prisées. Les analyses multirésolutions de $L^2(\mathbb{R})$ ont été définies par Mallat dans [40] et Meyer dans [41]. Elles correspondent à une suite de sous-espaces emboîtés $\{V_j\}$ de $L^2(\mathbb{R})$, dits espaces d'approximation, vérifiant les propriétés suivantes :

$$\text{Espaces emboîtés:} \quad \dots \subset V_{j-1} \subset V_j \subset \dots \subset L^2(\mathbb{R}) \quad (3.18)$$

$$\text{Fermeture:} \quad \bigcap_{j \in \mathbb{Z}} V_j = \{0\}, \quad \overline{\bigcup_{j \in \mathbb{Z}} V_j} \equiv L^2(\mathbb{R}) \quad (3.19)$$

$$\text{Dilatation:} \quad \forall x \in V_j, \quad x(2t) \in V_{j+1} \quad (3.20)$$

$$\text{Translation:} \quad \forall x \in V_j, \quad \forall k \in \mathbb{Z}, \quad x(t - k2^{-j}) \in V_j \quad (3.21)$$

Base orthonormée:

$$\forall \varphi \in V_0, \quad \{\varphi(t - k)\}_{k \in \mathbb{Z}} \quad \text{forment une base orthonormale de } L^2(\mathbb{R}) \quad (3.22)$$

Les espaces V_j sont appelés espace d'approximation. L'approximation multirésolution de x à la résolution 2^{-j} est définie comme sa projection orthogonale sur l'espace V_j . La propriété (3.21) signifie que V_j est invariant par toute translation proportionnelle à l'échelle 2^{-j} ; la propriété (3.20) garantit que la dilatation par 2 d'une fonction x de V_j (donc le grossissement par deux des détails) définit une approximation de cette fonction à l'échelle plus grossière 2^{-j+1} .

La théorie présentée par Meyer [41] et Mallat [40] permet de construire des fonctions d'échelle ϕ dont les motifs translatés $\{x(t - k)\}_{k \in \mathbb{Z}}$ sont orthogonaux deux à deux. De telles fonctions d'échelle, dites orthogonales, permettent alors la construction de bases orthonormées pour les espaces d'approximation d'une analyse multirésolutions de $L^2(\mathbb{R})$.

Théorème 3.1: Soit $\{V_j\}_{j \in \mathbb{Z}}$ une analyse multirésolutions de $L^2(\mathbb{R})$, alors il existe une unique fonction ϕ (fonction d'échelle) dont les motifs $\{x(t - k)\}_{k \in \mathbb{Z}}$ forment une famille orthonormale de $L^2(\mathbb{R})$; alors pour toute échelle j entière les motifs dilatés-translatés $\{\phi_{j,k}\}_{k \in \mathbb{Z}}$ forment une base orthonormée d'un espace d'approximation V_j tels que :

$$\forall t \in \mathbb{R}, \quad \phi_{j,k}(t) = 2^{-j/2} \phi(2^{-j}t - k) \quad k \in \mathbb{Z} \quad (3.23)$$

Ces motifs ont la propriété suivante :

$$\int_{-\infty}^{\infty} \phi_{0,0}(t) dt = 1 \quad (3.24)$$

La fonction d'échelle peut être convoluée avec le signal $x(t)$ pour produire les coefficients d'approximation, comme suit :

$$S_{j,k} = \int_{-\infty}^{\infty} x(t)\phi_{j,k}(t)dt \quad (3.25)$$

Ces coefficients d'approximation à une échelle j sont appelés l'approximation discret du signal à cet échelle.

III.9.1 Approximation filtrée dans V_j

Une approximation continue du signal à l'échelle j peut être générée par sommation d'une séquence des fonctions d'échelle factorisée par les coefficients d'approximation. Soient ϕ une fonction d'échelle orthogonale admissible et $x(t)$ une fonction de $L^2(\mathbb{R})$; alors on appelle voie d'approximation à l'échelle $j \in \mathbb{Z}$, notée $x_j(t)$, la projection de la fonction $x(t)$ sur l'espace d'approximation V_j admettant la décomposition en série suivante :

$$x_j(t) = \sum_{k=-\infty}^{\infty} S_{j,k} \phi_{j,k}(t) \quad \text{tel que} \quad \lim_{j \rightarrow +\infty} x_j(t) = x(t) \quad (3.26)$$

Où : $S_{j,k} = \langle x, \phi_{j,k} \rangle_{L^2(\mathbb{R})}$ sont les coefficients de la transformée en échelle de la fonction x calculés avec la fonction père ϕ associés à la résolution 2^{-j} et à la translation $k2^{-j}$. Pour une fonction d'échelle ϕ (filtre passe-bas) de variance fréquentielle σ , les éléments d'un espace V_j relatifs à une échelle $j \in \mathbb{Z}$, sont à spectre à bande limitée de largeur $2^j\sigma$ en écart-type. Les projections successives d'un signal $x(t)$ sur les espaces V_j , pour des échelles j croissantes, permettent de compléter la voie d'approximation $x_{j_0 < j}$ avec des détails fréquentiels. Ces derniers signaux vivant entre deux approximations successives x_j et x_{j+1} sont appelés (voies de détails). Leur existence est associée à celle d'espaces de détails. On définit alors, l'espace W_j complémentaire de V_j dans V_{j+1} tel que $V_j \oplus W_j$.

III.9.2 Espace de détails de $L^2(\mathbb{R})$

On appelle espaces de détails d'une analyse multirésolutions $\{V_j\}_{j \in \mathbb{Z}}$ de $L^2(\mathbb{R})$ la suite $\{W_j\}_{j \in \mathbb{Z}}$ de sous-espaces de $L^2(\mathbb{R})$ vérifiant les propriétés suivantes :

Orthogonalité: $\forall j \in \mathbb{Z}, V_{j+1} = V_j \oplus W_j \quad (3.27)$

Superposition de détails : $\forall v \in \mathbb{Z}, \bigoplus_{j \in \mathbb{Z}} W_j = V_v \oplus \bigoplus_{j \geq v} W_j \equiv L^2(\mathbb{R}) \quad (3.28)$

Dilatation: $\forall x \in W_j, \varphi(2t) \in W_{j+1}$ (3.29)

Translation: $\forall x \in W_j, \forall k \in \mathbb{Z}, \varphi(t - k2^{-j}) \in W_j$ (3.30)

Base orthonormée:

$$\forall \varphi \in W_0, \{\varphi(t - k)\}_{k \in \mathbb{Z}} \text{ forment une base orthonormale de } L^2(\mathbb{R}) \quad (3.31)$$

III.9.3 Base orthonormée de W_j

Soit ϕ une fonction d'échelle admissible engendrant une analyse multirésolutions $\{V_j\}_{j \in \mathbb{Z}}$ de $L^2(\mathbb{R})$, soit ψ une ondelette admissible dont les motifs $\{x(t - k)\}_{k \in \mathbb{Z}}$ forment une famille orthonormale de $L^2(\mathbb{R})$ alors, pour toute échelle $j \in \mathbb{Z}$, les motifs dilatés-translatés $\{\psi_{j,k}\}_{k \in \mathbb{Z}}$ forment une base orthonormée d'un espace d'approximation W_j tels que :

$$\forall t \in \mathbb{R}, \psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j}t - k) \quad (3.32)$$

De plus, l'ensemble $\{\psi_{j,k}\}_{k \in \mathbb{Z}}$ forme un frame orthonormal de $L^2(\mathbb{R})$

III.9.4 Voie de détails dans W_j

Soient ψ une ondelette orthogonale admissible et x une fonction de $L^2(\mathbb{R})$; alors on appelle voie de détails à l'échelle $j \in \mathbb{Z}$, notée d_j , la projection de la fonction x sur l'espace de détails W_j admettant la décomposition en série suivante :

$$d_j(t) = \sum_{k=-\infty}^{\infty} d_{j,k} \psi_{j,k}(t) \quad (3.33)$$

Où : $d_{j,k} = \langle x, \psi_{j,k} \rangle_{L^2(\mathbb{R})}$ sont les coefficients de la transformée en ondelettes de la fonction x , calculés avec l'ondelette ψ , relatifs à la résolution 2^j et à la translation $k/2^j$. La formule de synthèse mixte d'une fonction de $L^2(\mathbb{R})$, dans le cadre des analyses en ondelettes continues trouve naturellement un équivalent discret dans le formalisme des analyses multirésolutions. Les sommes continues de l'équation (3.8) cèdent alors leur place à des sommes discrètes définissant la superposition d'une voie d'approximation à une résolution donnée et des voies de détails de résolutions supérieures. Soit (ϕ, ψ) un couple de fonctions d'échelle et de détails admissibles, engendrant une analyse multirésolutions de $L^2(\mathbb{R})$ déterminée par la suite d'espaces d'approximation $\{V_j\}_{j \in \mathbb{Z}}$ et la suite d'espaces de détails $\{W_j\}_{j \in \mathbb{Z}}$; alors toute fonction x de $L^2(\mathbb{R})$ peut être décomposée en la somme d'une voie d'approximation $x_\nu(t)$ à une échelle $\nu \in \mathbb{Z}$ et des voies de détails $d_j(t)$ aux échelles $j \geq \nu$ telle que:

$$x(t) = x_\nu(t) + \sum_{j=\nu}^{+\infty} d_j(t) \quad (3.34)$$

Nous pouvons conclure alors, à partir de cette dernière équation, que le signal original continu est exprimé comme la combinaison de son approximation à une échelle donnée v et des détails du signal à cette échelle. Nous pouvons alors, représenter le signal $x(t)$ en utilisant des séries d'expansion combinées de coefficients d'approximation et de coefficients d'ondelette appelés coefficients de détail comme suit :

$$x(t) = \sum_{n=-\infty}^v S_{v,n} \phi_{v,n}(t) + \sum_{j=v}^{+\infty} \sum_{k=-\infty}^{+\infty} d_{j,k} \psi_{j,k}(t) \quad (3.35)$$

La discrétisation de la transformée en ondelettes continue, dans le cadre des analyses multirésolutions, fournit un outil puissant permettant une analyse hiérarchisée des fonctions de $L^2(\mathbb{R})$. De manière très synthétique ces dernières peuvent être décomposées en une voie d'approximation agrégeant toutes les informations fréquentielles jusqu'à une résolution donnée et la somme de voies de détails vivant aux résolutions supérieures. Cependant c'est en mettant à jour des similitudes avec des algorithmes de traitement d'images et ceux de la synthèse de la parole [42] que Mallat [40] donne aux analyses multirésolutions leurs lettres de noblesse. En s'inspirant de la théorie des filtres miroir en quadrature [43] ainsi que des algorithmes de Burt et Adelson [44], il munit les analyses multirésolutions d'algorithmes rapides pour le calcul des coefficients d'échelles et de détails ; les inscrivant alors définitivement dans un contexte numérique. Le paragraphe qui suit en présente les notions élémentaires.

III.9.5 Théorie des filtres miroir en quadrature et algorithmes rapides de Mallat

Dans le cadre des analyses multirésolutions la paternité de la construction des espaces d'approximation et de détails est accordée à la fonction d'échelle. Par ailleurs, la propriété d'inclusion vérifiée par les espaces d'approximation d'une analyse multirésolutions, donnée en équation (3.18), implique que les éléments de V_0 sont également des éléments de V_1 et que par conséquent la fonction d'échelle ϕ admet une décomposition unique sur la base orthonormée de V_1 formée par les atomes dilatés-translatés de la fonction ϕ telle que :

$$\phi(t) = \sqrt{2} \sum_{k \in \mathbb{Z}} h_n \phi(2t - n) \quad \forall t \in \mathbb{R} \quad (3.36)$$

Cette relation établit l'existence d'un filtre discret h_n intrinsèque lié à la définition de la fonction d'échelle ϕ . Les h_n sont les coordonnées de $\phi(t)$ dans la base $\{\sqrt{2}\phi(2t - k), k \in \mathbb{Z}\}$. Pour une base orthonormée les coordonnées vérifient :

$$h_n = \sqrt{2} \int_{-\infty}^{+\infty} \phi(t) \phi(2t - n) dt \quad \text{et} \quad \sum_{n=-\infty}^{+\infty} h_n^2 = 1 \quad (3.37)$$

De façon plus générale, il est étudié dans [45] des conditions nécessaires et suffisantes pour qu'un filtre discret de $L^2(\mathbb{R})$ engendre une fonction d'échelle admissible; une telle condition

permettant alors la construction d'algorithmes rapides pour le calcul des transformées en échelle et en ondelettes discrétisées. Le théorème suivant, dont une démonstration a été proposée par Mallat [46] ainsi que par Cohen [45], fournit les conditions nécessaires et suffisantes d'existence de tels filtres alors appelés 'filtres miroir conjugués'.

III.9.6 Filtre miroir conjugué

Théorème 3.2 : Soit ϕ une fonction de $L^2(\mathbb{R})$; alors elle engendre une analyse multirésolutions de $L^2(\mathbb{R})$ si et seulement si le filtre discret $\{h_n\}_{n \in \mathbb{Z}}$ voit sa transformée de Fourier \hat{h} , périodique, continûment différentiable dans un voisinage de $\xi = 0$ et telle que :

$$\begin{cases} \forall \xi \in \mathbb{R}, & |\hat{h}(\xi)|^2 + \left| \hat{h}\left(\xi + \frac{1}{2}\right) \right|^2 = 2 \\ & \hat{h}(0) = 2 \\ & \inf_{\xi \in [-1/2, 1/2]} |\hat{h}(\xi)| > 0 \end{cases} \quad (3.38)$$

de plus ;

$$\forall \xi \in \mathbb{R}, \quad \hat{\phi}(\xi) = \prod_{p=1}^{+\infty} \hat{h}(\xi/2^p)/\sqrt{2} \quad (3.39)$$

La portée de ce théorème est telle que les fonctions d'échelle disparaissent au profit des seuls filtres qui les engendrent dans un contexte numérique. Ces derniers sont alors appelés filtres d'échelle et sont renormalisés par le corollaire suivant afin d'en simplifier l'utilisation formelle.

III.9.7 Filtre d'échelle

Corollaire 3.1: Soit ϕ une fonction d'échelle et $\{h_n\}_{n \in \mathbb{Z}}$ son filtre miroir conjugué vérifiant les hypothèses du théorème précédent ; alors la fonction ϕ admet la décomposition suivante :

$$\phi(t) = \sum_{n \in \mathbb{Z}} h_n \phi(2t - n) \quad (3.40)$$

soit de manière équivalente en introduisant le filtre d'échelle périodique m_h tel que :

$$m_h(f) = \frac{1}{2} \hat{h}(f) = \frac{1}{2} \sum_{n \in \mathbb{Z}} h_n \exp(-2\pi i n f) \quad (3.41)$$

$$\forall f \in \mathbb{R}, \quad \hat{\phi}(f) = m_h(f/2) \hat{\phi}(f/2) = \prod_{j=1}^{+\infty} m_h(f/2^j) \quad (3.42)$$

Le filtre d'échelle m_h déterminant complètement la fonction ϕ selon la propriété donnée en équation (3.42) est un filtre passe-bas. Il intervient dans les produits de convolution permettant de calculer les voies d'approximation x_j d'une fonction x de $L^2(\mathbb{R})$ aux

échelles $j \in \mathbb{Z}$. Ceci explique la coupure fréquentielle réalisée par les espaces d'approximation d'une analyse multirésolutions. On peut de la même façon, que la fonction d'échelle, caractériser l'ondelette $\psi(t)$ qui est $W_0 \subset W_1$ en introduisant le filtre discret g_n . Soit ψ une ondelette associée à une fonction d'échelle ϕ engendrant une analyse multirésolutions $\{V_j\}_{j \in \mathbb{Z}}$ de $L^2(\mathbb{R})$. La fonction ψ se décompose alors de manière unique sur l'espace d'approximation V_1 selon la relation :

$$\psi(t) = \sqrt{2} \sum_{n \in \mathbb{Z}} g_n \phi(2t - n) \quad \forall t \in \mathbb{R} \quad (3.43)$$

On montre que la fonction ψ est nécessairement déterminée à partir de ϕ par un filtre discret g . Les paragraphes qui suivent présentent les conditions nécessaires et suffisantes que doit vérifier un filtre pour engendrer une ondelette associée à une fonction d'échelle donnée. Cette démarche calquée sur celle des filtres d'échelle plonge les analyses multirésolutions dans la théorie des filtres miroir en quadrature initialement développée par Esteban et Galand dans [47] et appliquée aux cas des analyses en ondelettes discrètes par Mallat et Feauveau dans [48].

III.9.8 Filtres miroir en quadrature

Théorème 3.3: *Soit ϕ une fonction d'échelle déterminée par son filtre miroir conjugué $\{h_n\}_{n \in \mathbb{Z}}$ et engendrant une analyse multirésolutions $\{V_j\}_{j \in \mathbb{Z}}$ de $L^2(\mathbb{R})$; alors le filtre $\{g_n\}_{n \in \mathbb{Z}}$ détermine une ondelette ψ engendrant les espaces de détails $\{W_j\}_{j \in \mathbb{Z}}$ si et seulement si :*

$$\begin{cases} \hat{h}(0) = 0 \quad \text{et} \quad \hat{h}(1/2) = \sqrt{2} \\ |\hat{g}(\xi)|^2 + \left| \hat{g}\left(\xi + \frac{1}{2}\right) \right|^2 = 2 \\ \hat{h}(\xi) \overline{\hat{g}(\xi)} + \hat{h}\left(\xi + \frac{1}{2}\right) \overline{\hat{g}\left(\xi + \frac{1}{2}\right)} = 0 \end{cases} \quad (3.44)$$

Les filtres $\{h_n\}_{n \in \mathbb{Z}}$ et $\{g_n\}_{n \in \mathbb{Z}}$ sont alors appelés filtres miroir en quadrature. Comme pour les fonctions d'échelle, un filtre de détails peut être mis en évidence qui détermine alors complètement une ondelette admissible donnée.

III.9.9 Filtre de détail

Corollaire 3.2: *Soit ψ une ondelette donnée par son filtre $\{g_n\}_{n \in \mathbb{Z}}$ miroir en quadrature avec le filtre $\{h_n\}_{n \in \mathbb{Z}}$ d'une fonction d'échelle ϕ ; alors la fonction ψ admet la décomposition suivante :*

$$\forall t \in \mathbb{R}, \quad \psi(t) = \sqrt{2} \sum_{n \in \mathbb{Z}} g_n \phi(2t - n) \quad (3.45)$$

soit de manière équivalente en introduisant le filtre de détail périodique m_g tel que :

$$m_g(f) = \frac{1}{2} \hat{g}(f) = \frac{1}{2} \sum_{n \in \mathbb{Z}} g_n \exp(-2\pi i n f) \quad (3.46)$$

$$\forall f \in \mathbb{R}, \quad \hat{\phi}(f) = m_g(f/2) \hat{\phi}(f/2) = m_g(f/2) \prod_{j=2}^{+\infty} m_h(f/2^j) \quad (3.47)$$

Les réponses impulsionnelles des filtres miroir en quadrature h et g vérifient la relation :

$$\forall n \in \mathbb{Z}, \quad g(n) = (-1)^{1-n} \overline{h_{1-n}} \quad (3.48)$$

L'extension de la théorie des filtres miroirs en quadrature, issue des travaux de Esteban et Galand pour la synthèse de la parole [47], au cadre des analyses en ondelettes discrétisées sur une grille dyadique a permis d'inscrire définitivement les analyses en ondelettes dans un contexte numérique. Désormais une analyse multirésolutions peut être entièrement déterminée par la donnée du filtre miroir d'une fonction d'échelle. Ce formalisme a ouvert la voie à d'innombrables travaux portant sur la construction de bases d'ondelettes alors désignées par Meyer comme de (purs objets numériques) et servant des objectifs spécifiques. A titre d'exemple, on peut citer les ondelettes à support compact développées par I. Daubechies dans [49], devenues populaires parce qu'elles sont déterminées à partir de filtres à réponses impulsionnelles finies.

III.9.10 Algorithmes rapides de Mallat

L'imbrication des espaces V_j et l'utilisation de convoluées discrètes par les filtres h et g ont permis la construction d'algorithmes rapides en cascade, capables de réaliser la projection orthogonale d'un signal sur un espace d'ondelettes, c'est-à-dire de faire correspondre à n valeurs échantillonnées du signal des coefficients d'échelle et d'ondelette, et l'opération inverse : la reconstruction d'un signal à partir de la donnée de coefficients d'échelle et d'ondelette par l'utilisation de filtres duaux \bar{h} et \bar{g} . La taille du vecteur des coefficients d'échelle et d'ondelettes varie en fonction de la taille du filtre h , car l'application du filtre consiste en une convolution et la taille des filtres varient en fonction des AMR. Les convolutions par les filtres sont suivies ou précédées respectivement de décimation-sous-échantillonnage (décomposition) ou d'insertion de zéros (reconstruction).

a. Algorithme d'analyse

En travaillant sur la vision artificielle, P. Burt et E. Adelson [44] ont développé des algorithmes pyramidaux empiriques. Mallat leur donne une justification mathématique dans le cadre de la théorie des filtres miroir en quadrature et les relie à la théorie des analyses multirésolutions [37]. Les algorithmes mis en évidence permettent de réaliser l'analyse (décomposition) et la synthèse (recomposition), sans perte, d'un signal x_j d'un espace d'approximation V_j dont les coefficients d'approximation sont supposés connus. L'algorithme d'analyse permet un calcul rapide des coefficients d'échelle $S_{j,k}$ et de détails d_j du signal x

pour des échelles $j < J$. La méthode d'implémentation des coefficients réalise un calcul pyramidal. Alors via un calcul de convolutions discrètes suivies de décimations, la voie de coefficients x_j est décomposée en une voie de coefficients d'approximation x_{j-1} et une voie de coefficients de détails d_{j-1} . La représentation pyramidale provient du fait que chacune des voies x_{j-1} et d_{j-1} possède deux fois moins de coefficients que la voie x_j .

Théorème 3.4: Soient (ϕ, ψ) un couple de fonctions d'échelle et de détail admissibles engendrant une analyse multirésolutions. Soient h et g leurs filtres miroir en quadrature, soient x une fonction de $L^2(\mathbb{R})$ et x_j sa voie de coefficients d'approximation à une échelle j ; alors les voies de coefficients d'approximation x_{j-1} et de détails d_{j-1} vérifient les relations :

$$\forall n \in \mathbb{Z}, \begin{cases} x_{j-1,n} = \sum_{p \in \mathbb{Z}} h_{p-2n} x_{j,p} = [x_j * h^*](2n) \\ d_{j-1,n} = \sum_{p \in \mathbb{Z}} g_{p-2n} x_{j,p} = [x_j * g^*](2n) \end{cases} \quad (3.49)$$

Les relations de convolution, données en équation (3.49), mettent en évidence une sélection passe-bas réalisée par le filtre h ainsi qu'une troncature passe-haut réalisée par le filtre g . Le calcul final nécessite cependant une décimation des voies $[x_j * h^*]$ et $[x_j * g^*]$ qui ne retient que les coefficients de positions paires. Cet algorithme d'analyse multirésolutions rapide est représenté dans la figure (3.5) en ce qui concerne la décomposition d'un signal x d'une échelle fine J à une échelle grossière J_0 . Cette procédure, appliquée de manière itérative aux voies de coefficients d'approximation du signal x , permet de proche en proche de décomposer la voie de coefficients d'approximation x_j en la pyramide de coefficients $\{x_{J_0}, d_{J_0}, d_{J_0+1}, \dots, d_{J-1}\}$ tel qu'il est indiqué dans la figure (3.5).

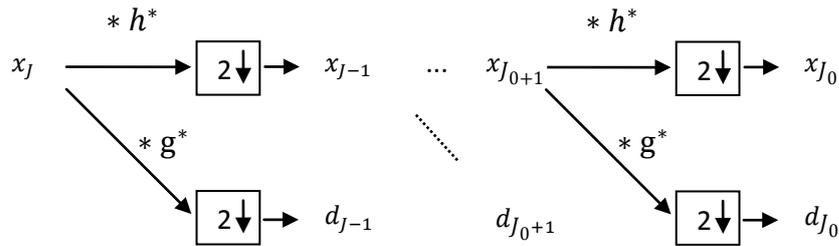


Fig. 3.5 Algorithme d'analyse pyramidale d'une voie de coefficients d'approximation x_j en ses voies de coefficients d'approximation x_{J_0} et de coefficients de détails $\{d_j\}_{J_0 \leq j \leq J-1}$, le symbole * indique un schéma de convolution discrète

Réciproquement, un algorithme rapide permet la synthèse d'une voie de coefficients d'approximation à une échelle donnée $j \in \mathbb{Z}$ à partir des voies de coefficients d'approximation et de détails d'échelles inférieures. Cet algorithme rapide est donné dans le théorème suivant.

b. Algorithme de synthèse

Théorème 3.5: Soient (ϕ, ψ) un couple de fonction d'échelle et de détail admissibles engendrant une analyse multirésolutions de $L^2(\mathbb{R})$, soient h et g leurs filtres miroir en quadrature, soient x une fonction de $L^2(\mathbb{R})$ et (x_{j-1}, d_{j-1}) ses voies de coefficients d'approximation et de détails implémentées à une échelle $j - 1$; alors la voie de coefficients d'approximation x_j vérifie la relation :

$$x_{j,n} = \sum_{p \in \mathbb{Z}} h_{n-2p} x_{j,p} + \sum_{p \in \mathbb{Z}} g_{n-2p} d_{j,p} = [x_j^* * h^*](n) + [d_j^* * g^*](n) \quad (3.50)$$

En généralisant le principe et en supposant que la pyramide des coefficients du signal x ait été implémentée d'une échelle fine $J - 1$ jusqu'à une échelle grossière J_0 , une voie de coefficients d'approximation x_j peut être reconstituée à partir des voies de coefficients d'approximation x_{j_0} et de détails d_j relatives aux échelles $J_0 \leq j \leq J - 1$, tel qu'il est représenté dans la figure suivante.

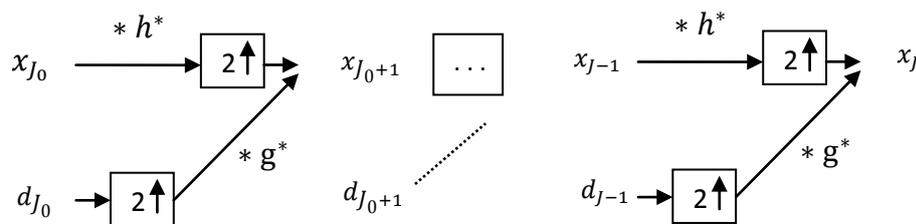


Fig. 3.6 Algorithme de synthèse pyramidale de la voie de coefficients d'approximation x_j à partir de la voie de coefficients de détails $\{d_j\}_{J_0 \leq j \leq J-1}$ et de la voie de coefficients d'approximation x_{j_0}

Les figures (3.7 - 3.9) montrent l'analyse multirésolutions de signaux réalisés des échelles 1 à 3, avec un couple de fonctions d'échelle et de détails de Daubechies d'ordre 4. Ces figures présentent le résultat de la décomposition de x en voies de détails $\{d_1, d_2, d_3\}$ et en voie d'approximation x_4 . Cet exemple met en relief l'existence d'oscillations sur les voies de détails.

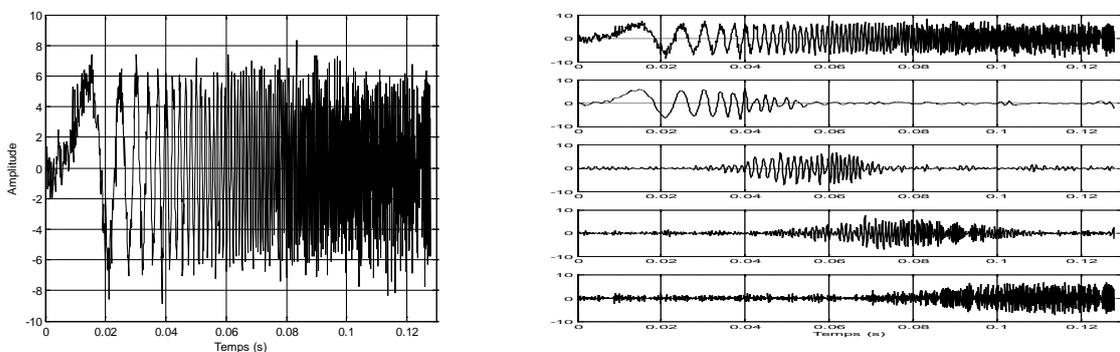


Fig. 3.7 Analyse multi-résolution du signal chirp

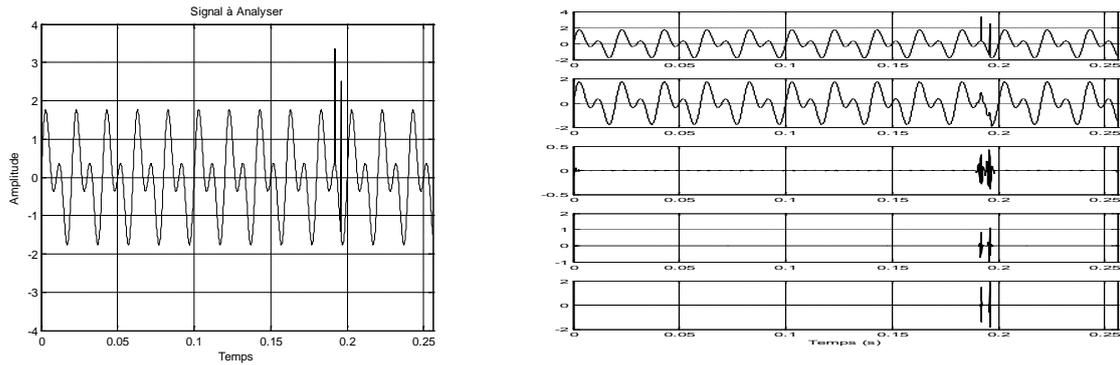


Fig. 3.8 Analyse multi-résolution du signal simulé avec présence de deux impulsions

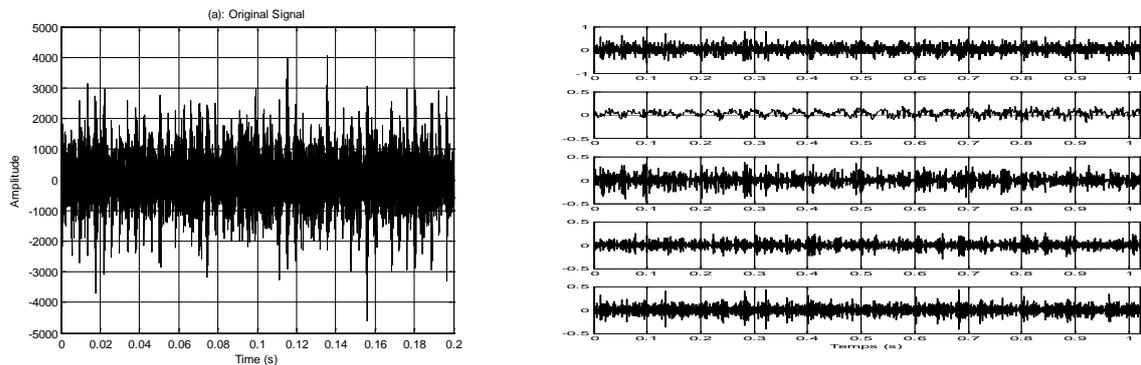


Fig. 3.9 Analyse multi-résolution du signal d'acquisition avec à un défaut de roulement

III.10 Conclusion

Les analyses en ondelettes continues sont définies comme la projection d'un signal donné sur un ensemble de fonctions, bien localisées en temps et en fréquence, permettant de décrire continûment le plan temps-fréquence via des dilatations et des translations. Ces analyses fournissent par conséquent une information qui se trouve être localement redondante pour la description sans perte de phénomènes fréquentiels transitoires siégeant dans un signal. L'analyse par ondelettes fait dans ce sens mieux que l'analyse de Fourier à fenêtre glissante ; cette dernière a l'inconvénient de proposer une fenêtre d'observation de taille fixe, localisant donc mal les discontinuités à des échelles plus petites que l'échelle d'observation, et étant limité dans les basses fréquences par cette même taille de fenêtre. La transformée en ondelettes propose une fenêtre d'observation qui s'adapte aux échelles des fluctuations présentes. Dés lors, nous avons présenté dans ce chapitre quelques exemples classiques d'ondelettes, chacune d'elles possédant des avantages et des inconvénients. Dans la pratique, le choix d'une telle ou telle ondelette n'est pas facile et dépend beaucoup de l'objectif à atteindre et de la situation donnée. En revanche, les analyses multirésolutions, qui en découlent, offrent une représentation très condensée d'un signal qui s'avère pertinente pour pallier au problème de redondance de l'information. L'orthogonalité des bases considérées permet une discrétisation du plan temps-fréquence selon une grille dyadique minimale.

Dès lors un signal est analysé en voies d'approximation et de détails d'échelles successives. L'extension de la théorie des filtres-miroir en quadrature à la théorie des analyses en ondelettes proposée par Mallat, a permis une réduction considérable des coûts de calculs. Cependant, dans des situations où il est souhaitable d'avoir plus d'information, par exemple en analyse des signaux naturels tel que les vibrations la redondance est alors appréciée et l'utilisation de la transformée en ondelettes continue (ou de frame) est plus adaptée. Dans notre cas, il est plus intéressant de construire des ondelettes adaptées qui peuvent donner des résultats numériques plus satisfaisants que les ondelettes classiques.

Chapitre IV

Parcimonie et représentation des signaux impulsionnels de vibrations

Chapitre 4

Parcimonie et représentation des signaux impulsionnels

Résumé

En mathématique, le concept de parcimonie revient à considérer qu'un signal peut être exprimé à l'aide d'un faible nombre d'éléments. L'usage de ce concept a permis de grandes avancées dans la modélisation et la manipulation de la plupart des signaux naturels. Dans ce chapitre, nous établissons les méthodes de représentation et de quantification permettant d'identifier un signal comme parcimonieux, c'est-à-dire que l'on peut décomposer de manière parcimonieuse dans un espace de représentation donné. Nous commencerons par définir plus précisément cette notion et voir son intérêt pour le cas particulier de la représentation des signaux impulsionnels, puis nous nous intéresserons à son intégration dans des problèmes inverses. Ces problèmes inverses nous amèneront à décrire des algorithmes d'optimisation spécifiques permettant de les résoudre. On présente quelques problèmes inverses linéaires classiques, tels que la restauration des signaux (débruitage) et la séparation de sources (en particulier, les mélanges sous-déterminés, c'est-à-dire lorsqu'il y a plus de sources à estimer que de mélanges observés). A partir de ces exemples, on verra comment une approche par l'utilisation des représentations redondantes, alliée au principe de parcimonie, permet d'attaquer ces problèmes en vue de leur inversion. Les performances de ces méthodes seront illustrées sur les problèmes considérés dans les chapitres qui suivent.

Sommaire

IV.1	Introduction	60
IV.2	Représentation parcimonieuse	60
IV.3	Bases et dictionnaires	61
IV.4	Mesure de la parcimonie des coefficients d'analyse	63
IV.5	Algorithmes de décomposition	66
IV.6	Parcimonie et entropies	68
IV.7	Formulations de la parcimonie dans un espace redondant adapté	69
IV.8	Choix de l'ondelette analysante adaptée au signal vibratoire impulsionnel	70
IV.9	Optimisation de l'ondelette de Morlet pour la détection des impulsions	71
IV.10	Application de la parcimonie dans les problèmes inverses	74
IV.12	Conclusion	78

IV.1 Introduction

Un signal est le support physique d'une information. La facilité de l'extraction de cette information se mesure à sa lisibilité. Afin d'approximer au mieux un signal, on cherche en général un système de représentation idoine. L'idéal serait de connaître l'évolution du contenu fréquentiel au cours du temps. Cependant, les représentations temps-fréquence, se heurtent malheureusement au principe d'incertitude d'Heisenberg qui empêche d'être précis à la fois en temps et en fréquence. Il va falloir donc, faire un choix (difficile !) entre la précision temporelle et fréquentielle. Une approche pourrait être de choisir plusieurs systèmes de représentation différents (échantillons, fréquence, temps-fréquence avec différents choix de précisions en temps et en fréquence ...etc.), puis de rechercher l'information qui nous intéresse effectivement. La difficulté devient : comment chercher et sélectionner ce qu'on veut parmi la masse d'information ? L'objet de ce chapitre est de proposer une méthode assez générale, afin de sélectionner, ou de trouver une représentation parcimonieuse, c'est-à-dire concise, d'un signal dans un ensemble d'informations élémentaires (qu'on appellera atomes). Le cadre mathématique de la représentation parcimonieuse d'un signal dans un dictionnaire de formes d'ondes est présenté en premier lieu, puis les différences entre l'analyse d'un signal et sa synthèse. On effectue aussi un état de l'art des approches les plus populaires.

IV.2 Représentation parcimonieuse

Au sens strict, un signal est dit parcimonieux (sparse) quand il peut être convenablement approximé à l'aide d'un faible nombre de coefficients non-nuls dans une base donnée. Un tel signal est ainsi plus simple à manipuler : la connaissance des valeurs et positions de ces quelques coefficients non-nuls suffit à parfaitement le définir. La recherche de signaux parcimonieux revient alors à chercher des signaux ayant une structure particulière, que l'on sait bien modéliser du point de vue mathématique. L'objet de ce chapitre est de proposer une méthode assez générale, afin de sélectionner, ou de trouver une représentation parcimonieuse, du signal vibratoire impulsionnel, c'est-à-dire une représentation où l'énergie du signal est concentrée sur un faible nombre d'éléments. Ainsi, on cherchera à représenter le signal comme une combinaison linéaire de ces formes d'ondes appelées atomes, d'énergie unité, et qui entrent comme élément constituant d'un signal plus complexe par superposition linéaire. Ces atomes seront choisis dans un ensemble, appelé dictionnaire. Ce dictionnaire peut contenir des exemples d'atomes, ainsi que des formes d'ondes définies par un nombre restreint de paramètres. Par ailleurs, le choix d'une base ou d'un dictionnaire est crucial et sa pertinence se mesure à la capacité qu'elle a à fournir des descripteurs se prêtant bien à des tâches ultérieures. En particulier, il est désirable que l'essentiel de l'énergie du signal soit concentrée sur très peu de coefficients dans le domaine de la transformée. En d'autres termes, le signal peut être expliqué par peu de degrés de liberté. Un signal possédant cette propriété est alors dit parcimonieux dans le domaine de la transformée en question. La parcimonie est une source précieuse d'*a priori* sur les signaux. On verra au Chapitre V qu'il existe des liens étroits entre cet *a priori* et une classe de modèles stochastiques.

De point de vue quantitative, on définit le support ‘supp’ d’un vecteur comme l’ensemble des coefficients non-nuls de ce vecteur, et la pseudo-norme ℓ_0 comme le cardinal ‘card’ du support. En d’autres termes, la pseudo-norme ℓ_0 est le nombre de coefficients non-nuls de ce signal. Pour un vecteur x , on a donc :

$$\|x\|_0 = \text{card}(\text{supp}(x)) \quad (4.1)$$

Un signal est ainsi d’autant plus parcimonieux que sa pseudo-norme ℓ_0 est faible.

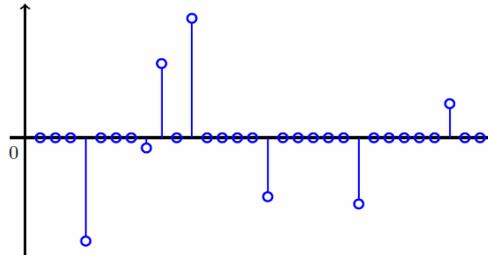


Fig. 4.1 Exemple d’un signal temporel parcimonieux

IV.3 Bases et dictionnaires

Afin d’approximer au mieux un signal, on cherche en général un système de représentation idoine. Considérons un signal $x \in \mathbb{R}^n$. Classiquement, ce signal est représenté dans la base canonique de \mathbb{R}^n comme suit :

$$x = \sum_{i=1}^n x_i e_i \quad (4.2)$$

Cependant, la base canonique n’est pas la seule base de représentation possible : toute autre base Φ de $\mathbb{R}^{n \times p}$ peut décrire x , par une superposition linéaire des atomes ϕ_i selon le modèle génératif linéaire de synthèse par combinaison linéaire de p atomes, tel que :

$$x = \Phi \alpha = \sum_{i=1}^p \alpha_i \phi_i \quad \alpha \in \mathbb{R}^p \quad (4.3)$$

où les α_i sont les coefficients de la décomposition représentant la contribution de l’atome ϕ_i dans x et les ϕ_i sont les colonnes de Φ . Le changement de base peut faciliter l’interprétation des signaux : par exemple, la transformée de Fourier permet une analyse harmonique du signal ; l’énergie est concentrée sur quelques éléments, plutôt que répartie sur l’ensemble. Le cadre général de la représentation des signaux est l’analyse hilbertienne. L’espace de Hilbert \mathcal{H} choisi permet alors définir une base ou un repère dans lequel représenter un signal donné. La définition d’un repère (frame) est rappelée ci-après.

Définition 4.1 : Une famille ϕ_i , $i \in I$, où I est un ensemble d'indices, est un repère (frame) de l'espace \mathcal{H} s'il existe deux constantes A, B , avec $0 < A \leq B < +\infty$ tels que pour tout $x \in \mathcal{H}$:

$$A\|x\|_2^2 \leq \sum_{i \in I} |\langle x, \phi_i \rangle|^2 \leq B\|x\|_2^2 \quad (4.4)$$

Si $A = B$, on dit que le repère est ajustée (tight frame). Plus généralement, on parlera de dictionnaire $\mathcal{D} = \{\phi_i, i \in I\}$ pour un ensemble sur-complet, c'est-à-dire que le nombre de colonnes est supérieur à la longueur de celles-ci, et dont les éléments ϕ_i sont normés. Ces éléments sont aussi appelés atomes du dictionnaire. Dans ce cas là, la décomposition de x n'est plus unique. En dimension infinie, un dictionnaire n'est pas nécessairement un repère : la deuxième partie de l'inégalité de la définition (4.4) n'est pas toujours vérifiée si le dictionnaire est très redondant. On peut, par exemple, construire des dictionnaires par union de bases ou de repères.

IV.3.1 Analyse et synthèse

Étant donné un signal $x \in \mathcal{H}$ et un dictionnaire $\mathcal{D} = \{\phi_i, i \in I\}$, deux questions peuvent se poser. La première consiste à se demander comment x se décompose dans le dictionnaire, c'est-à-dire quelle est l'information apportée par les atomes sur le signal. C'est l'approche *analyse*. L'analyse du signal se fait simplement par le calcul des produits scalaires avec tous les atomes du dictionnaire, et on s'intéresse alors aux coefficients $\{\langle x, \phi_i \rangle\}$. La seconde question consiste à savoir comment représenter x avec les atomes du dictionnaire. C'est l'approche *synthèse*. On cherche les coefficients α_i de manière à représenter x comme une combinaison linéaire des atomes, c'est à dire tel que :

$$x = \Phi\alpha = \sum_{i \in I} \alpha_i \phi_i \quad (4.5)$$

En dimension finie, l'opérateur Φ sera noté comme une matrice dont les colonnes sont formées par les atomes du dictionnaire, et les coefficients α comme un vecteur colonne. Φ est appliquée au vecteur $\alpha = (\alpha_1, \dots, \alpha_{|\mathcal{D}|})^T$ (où $|\mathcal{D}|$ représente le cardinal de l'ensemble \mathcal{D}). Si le dictionnaire choisi est en fait une base orthonormée de \mathcal{H} , alors l'analyse et la synthèse de x sont simplement liées du fait de l'unicité de la décomposition. On a alors :

$$x = \sum_{i \in I} \langle x, \phi_i \rangle \phi_i \quad (4.6)$$

À l'inverse, si \mathcal{D} est redondant, c'est-à-dire qu'on dispose plus d'atomes que nécessaire pour représenter le signal, il existe une infinité de représentations (4.5) possibles. Cependant, dans le cas où $\{\phi_i\}_{i \in I}$ est un repère, une représentation possible du signal est donnée à l'aide du repère dual canonique $\{\hat{\phi}_i\}_{i \in I}$ défini par $\hat{\phi}_i = (\Phi\Phi^*)^{-1}\phi_i$ pour tout i . On a alors:

$$x = \sum_{i \in I} \langle x, \hat{\phi}_i \rangle \phi_i \quad (4.7)$$

Cette représentation n'est pas toujours suffisante, une synthèse plus concise du signal à l'aide des atomes du dictionnaire est souvent plus judicieuse. La question sera alors de savoir quels sont les atomes significatifs, c'est-à-dire porteurs de l'information voulue, du dictionnaire.

IV.3.2 Synthèse et reconstruction exacte

Lorsque le dictionnaire choisi est redondant, il existe une infinité de solutions au problème (4.5). Ce problème, qui consiste à chercher les coefficients de synthèse d'un signal dans un dictionnaire, peut être vu comme un problème inverse mal posé. Avec les notations utilisées ici, le problème consiste à trouver les coefficients $\alpha = \{\alpha_i, i \in I\}$ tels que :

$$x = \Phi\alpha \quad (4.8)$$

Une première idée est de lever l'indétermination en contrôlant l'énergie des coefficients recherchés, on cherche alors à résoudre le problème :

$$\min_{\alpha} \|\alpha\|_2^2 \quad \text{t. q.} \quad x = \Phi\alpha = \sum_{i \in I} \alpha_i \phi_i \quad (4.9)$$

L'avantage de cette formulation est qu'elle admet une solution analytique simple de la forme

$$\hat{\alpha} = \Phi^*(\Phi\Phi^*)^{-1}x \quad (4.10)$$

Si cette solution peut être satisfaisante dans certains cas (c'est ainsi que l'on trouve le repère dual d'un repère donné), elle s'avère insuffisante lorsqu'on cherche à représenter x avec le moins de coefficients possibles. Afin de trouver une solution parcimonieuse au problème (4.5), on va chercher à minimiser une mesure \mathcal{M} de la (diversité), ou de (l'étalement) des coefficients. On pose alors le problème :

$$\min_{\alpha} \mathcal{M}(\alpha) \quad \text{t. q.} \quad x = \Phi\alpha = \sum_{i \in I} \alpha_i \phi_i \quad (4.11)$$

Le problème est donc de trouver une bonne mesure \mathcal{M} de la parcimonie. Le choix de cette mesure sera discuté dans la section suivante et dépend de la définition choisie pour la parcimonie.

IV.4 Mesure de la parcimonie des coefficients d'analyse

S'il est simple de décrire l'idée de la parcimonie, il est moins évident de la quantifier par une méthode de mesure. Le problème de la décomposition parcimonieuse d'un signal $x \in \mathbb{R}^n$ dans un dictionnaire redondant Φ consiste à retrouver les coefficients α_i ayant observé x .

Ceci revient donc à résoudre un système linéaire sous-déterminé. Bien que cette inversion soit notoirement un problème mal posé par les arguments élémentaires de l’algèbre linéaire, la situation change radicalement si le signal à recouvrer est parcimonieux. En effet, si on considère un signal parcimonieux dans Φ , c’est à dire représenté par un vecteur α avec peu de composantes non nulles. On dit que la représentation de x est parcimonieuse si :

$$\left\{ \begin{array}{l} x = \Phi\alpha = \sum_{i \in \text{supp}(\alpha)} \alpha_i \phi_i, \quad \text{avec } k = \text{Card}\{\alpha_i \neq 0 | i = 1, p\} \\ |\text{supp}(\alpha)| = k \ll p \end{array} \right. \quad (4.12)$$

Toutefois, il n’est pas toujours aisé de trouver une base dans laquelle les signaux possèdent naturellement une représentation parcimonieuse. On peut alors ne plus se restreindre à une base et ajouter de la redondance en augmentant le nombre de colonnes de la matrice Φ .

Un signal $x = (x_1, \dots, x_n)$, est strictement ou exactement parcimonieux si la plupart de ses composantes sont nulles ; c.a.d. son support $\text{supp}(x) = \{1 \leq i \leq n \mid x_i \neq 0\}$ est de cardinalité $|\text{supp}(x)| = k \ll n$. Un tel signal est dit k -parcimonieux. Le signal peut ne pas être parcimonieux dans son domaine original, mais plutôt dans le domaine d’une transformée appropriée. C’est typiquement le cas pour un signal sinusoïdal qui est parcimonieux dans le domaine de Fourier.

Remarque 4.1:

L’utilisation d’une transformée orthonormale correspond au cas particulier où les vecteurs ϕ_i sont normés et orthogonaux entre eux. En pratique, rien n’interdit d’utiliser davantage d’atomes qu’il n’y a de degrés de liberté dans \mathbb{R}^n , soit $p > n$. Dans ce cas là, la famille étant liée, plusieurs combinaisons d’atomes sont possibles pour représenter x . Il est alors possible de rajouter une contrainte sur le choix des atomes utilisés. On peut ainsi s’intéresser à trouver quelle est la décomposition la plus parcimonieuse, c’est-à-dire utilisant le moins possible d’atomes de la famille $(d(i))_{i=1, \dots, p}$. En pratique, de nombreux choix de dictionnaires sont envisageables. Les ondelettes redondantes $W \in \mathbb{R}^{n \times p}$, avec $p > n$, peuvent cependant permettre cette décomposition et se montre bénéfique afin d’obtenir une représentation plus parcimonieuse.

Parmi toutes les représentations de x dans Φ , nous voudrions la plus parcimonieuse, autrement dit la solution au problème d’optimisation :

$$(P_0) : \min_{\alpha \in \mathbb{R}^n} \|\alpha\|_0 \quad \text{t. q.} \quad x = \Phi\alpha \quad (4.13)$$

A première vue, cette tâche s’annonce ardue puisque (4.5) est un problème combinatoire NP-difficile ; c’est à dire que dans le cas général, sa complexité est exponentielle en fonction de la taille des vecteurs mis en jeu, ce qui rend les calculs trop longs pour des dimensions de dictionnaires intéressantes. Pour contourner cette difficulté, tout un pan de littérature a vu le jour proposant des approximations par des algorithmes de poursuite gloutons (greedy) comme

le matching poursuit et ses variantes. Cette parcimonie se mesure à l'aide de la pseudo-norme ℓ_0 , qui compte le nombre de coefficients non nuls d'un vecteur :

$$\|x\|_0 = \text{Card}\{\alpha_i \neq 0 | i = 1, p\} \quad (4.14)$$

Cette mesure est appelée abusivement norme ℓ_0 bien qu'il ne s'agisse pas d'une norme puisqu'elle ne respecte pas tous les axiomes d'une norme (la propriété d'homogénéité n'est pas vérifiée). Sachant que la norme ℓ_0 mesure le nombre de termes non-nuls dans la décomposition x . Si un vecteur x est tel que $\|x\|_0 = k$ avec $k \ll P$, on dit que x est k -parcimonieux (k -sparse) il y a k éléments non nuls dans le vecteur x . En minimisant ℓ_0 , on obtient donc la décomposition optimale comprenant un nombre minimal de termes non-nuls dans x . Cependant, si cette norme ℓ_0 semble adaptée à la quantification de la parcimonie d'un signal, le fait que tout élément non nul ait la même importance dans la valeur de la norme peut se révéler un obstacle dans la pratique : la présence de bruit, une approximation due à un effet d'arrondi, peut remettre en cause la valeur et mener à une situation où $\|x\|_0 \cong N$ alors que l'énergie du signal reste concentrée sur un faible nombre de composantes. Pour pallier ce défaut, il est courant de recourir à d'autres mesures par les (quasi) normes ℓ_p .

Définition 4.2.

Soit $x = \{x_i, i \in I\} \in \mathbb{R}^n$, on appelle norme ℓ_p pour $1 \leq p < \infty$ et quasi-norme ℓ_p pour $0 \leq p < 1$ la quantité :

$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} \quad (4.15)$$

On définit la norme ℓ_∞ comme

$$\|x\|_\infty = \sup_{i \in I} |x_i| \quad (4.16)$$

Remarque 4.2. Dans les normes ℓ_p , plus p est grand, plus les gros coefficients prennent de l'importance dans le calcul de la norme. Donc, lorsqu'on cherche à minimiser une quantité du type $\|x\|_p^p$, si $p > 2$ on aura tendance à pénaliser plus fortement les gros coefficients, et donc à favoriser la diversité. À l'inverse, lorsque l'on choisit de minimiser $\|x\|_p^p$ avec $p \leq 2$, ce sont les plus petits coefficients qui seront pénalisés, voire annulés si $p \leq 1$, et la parcimonie est favorisée.

Finalement, suivant le choix de p , la quantité $\|x\|_p^p$ donne une idée de la parcimonie ou de la diversité de décomposition d'un signal dans un dictionnaire donné. Un cas particulier des (quasi) normes ℓ_p est la norme ℓ_0 qui est définie comme :

$$\|x\|_0 = \sum_{i=1}^n |x_i|^0 = \text{Card}\{x_i \neq 0 | i = 1, n\} \quad (4.17)$$

Donoho, dans [51], a proposé de remplacer la norme ℓ_0 par la norme ℓ_1 . Le choix de la norme ℓ_1 permet d'obtenir un problème convexe, et plusieurs algorithmes sont disponibles pour trouver une solution. La section IV.5 reviendra sur les principaux algorithmes permettant de trouver de telles décompositions parcimonieuses.

$$(P_1) : \min_{\alpha \in \mathbb{R}^N} \|\alpha\|_1 \quad \text{t. q.} \quad x = \Phi\alpha \quad (4.18)$$

On se contente donc de travailler avec la norme ℓ_1 , c'est-à-dire la somme des valeurs absolues des coefficients. Ce critère de norme ℓ_1 offre une solution moins optimale que la norme ℓ_0 , mais qui a l'avantage d'être calculable par des techniques de programmation linéaire. Ce problème est connu sous l'appellation (poursuite de base, *Basis Pursuit BP*) et peut être résolu efficacement par des outils de décomposition d'opérateurs monotones maximaux [52]. La régularisation ℓ_1 a été proposée dès la décennie 90 notamment par Chen [52] pour les problèmes inverses linéaires, et formalisée par Donoho et ses collaborateurs [52, 53, 54]. Toutefois, il est clair que les minimisations ℓ_0 et ℓ_1 ne sont en général pas équivalentes, puisque le minimiseur global de BP (si tant est qu'il soit unique) ne coïncide pas toujours avec celui de (4.5). Mais sous des hypothèses appropriées sur Φ et x , BP possède un unique minimiseur global qui de plus coïncide avec celui de (4.5) ; on parle alors d'identifiabilité ℓ_1 . Une littérature très importante s'est penchée sur les conditions suffisantes (parfois même nécessaires) pour assurer l'unicité de la solution de BP, son équivalence avec celle de (4.5) ainsi que sa robustesse aux signaux compressibles. En présence d'un bruit typiquement de variance bornée, le problème (4.5) est modifié en adaptant le terme de fidélité aux données. Des conditions ont été aussi établies assurant la stabilité par minimisation ℓ_1 à ce type de bruits de même que la consistance du support et du signe du vecteur retrouve. Il n'est pas question ici de dresser une revue exhaustive ne serait-ce que parce que celle-ci évolue tellement vite qu'il serait illusoire d'être équitable et complet. Une revue assez complète est dressée dans [55] sur les critères de décomposition parcimonieuse aussi bien par minimisation ℓ_1 que par les algorithmes de poursuite gloutons.

IV.5 Algorithmes de décomposition

La section suivante présente des méthodes permettant d'obtenir une solution α parcimonieuse, au sens de ℓ_0 ou ℓ_1 .

IV.5.1 Poursuite adaptative (Matching Pursuit MP)

Mallat et Zhang [56] ont introduit à la communauté de traitement du signal un algorithme glouton (greedy) de poursuite très similaire à la projection poursuit [57] en statistiques, qu'ils appellent Matching Pursuit, pour approximer la décomposition idéale (P_0). Il s'agit d'un algorithme glouton qui, au lieu de chercher la solution globale recherche successivement la solution optimale d'un sous-problème, à savoir la meilleure approximation à un terme du signal. L'algorithme Matching pursuit consiste à sélectionner itérativement l'atome du dictionnaire le plus corrélé au signal (i.e., celui dont la valeur absolue du produit scalaire avec le signal est maximale, en supposant que les atomes sont normalisés), d'ôter la contribution de

ce vecteur au signal et de recommencer le processus jusqu'à obtention d'un nombre d'éléments égal à la dimension du signal, puis à projeter le signal sur l'orthogonal de l'espace engendré par les atomes identifiés jusque là. L'algorithme s'arrête quand le nombre d'atomes à identifier est atteint ou quand la norme du résidu descend sous un certain seuil. L'implémentation de cet algorithme est très facile, et peut être accélérée, par exemple, par une décomposition QR de la famille des atomes sélectionnés mise à jour à chaque étape de l'algorithme. Ainsi, la décomposition approximative du signal x sur ce dictionnaire Φ s'écrit comme suit :

$$s = \sum_{i=0}^{m-1} \alpha_i \phi_i + R^{(m)} \quad (4.19)$$

où $R^{(m)}$ représente l'erreur résiduelle à l'étape m . En résumé, à partir d'une approximation initiale $x^{(0)}$ de x , et d'un résidu $R^{(0)} = x$, on construit une famille d'approximations pas à pas. A chaque itération k , on recherche dans le dictionnaire l'atome d'indice γk qui a le plus grand produit scalaire avec le résidu $R^{(k-1)}$, et on ajoute sa contribution à l'estimé précédent $x^{(k-1)}$. On obtient ainsi

$$\begin{cases} x^{(k)} = x^{(k-1)} + \alpha_k \phi_{\gamma k} \\ R^{(k)} = R^{(k-1)} - \alpha_k \phi_{\gamma k} \\ \alpha_k = \langle R^{(k-1)}, \phi_{\gamma k} \rangle \end{cases} \quad (4.20)$$

Lorsque le dictionnaire est orthogonal, la méthode fonctionne parfaitement, et si le signal est construit avec seulement n composantes, l'algorithme s'arrête après n itérations. Cependant, bien que le MP fonctionne généralement bien, il existe des exemples [52, 58] de signaux parcimonieux qui la mettent en défaut et lui font produire des vecteurs de représentation non creuse. Cet algorithme de type (glouton) n'est pas optimal, mais il offre de bonnes propriétés de décroissance de l'erreur entre le signal original et le signal parcimonieux. Le principal avantage du Matching pursuit est sa simplicité, et de nombreux résultats théoriques ont été obtenus pour cet algorithme : principalement, sa convergence vers la solution de norme ℓ_0 ou ℓ_1 sous des hypothèses bien précises sur le dictionnaire et le signal. On pourra remarquer que cet algorithme ne fait appel à aucune mesure de parcimonie. Plusieurs variations de cet algorithme existent, dont le *Matching Pursuit orthogonal* et le *Stagewise Matching Pursuit*.

IV.5.2 Orthonormal Matching Pursuit (OMP)

Un raffinement de Matching pursuit [59] consiste à rajouter une étape d'orthogonalisation pour obtenir à chaque itération m la meilleure approximation avec les atomes déjà choisis,

$$\min_{\{\alpha_i, i \in [1, m]\}} \left\| x - \sum_{i=1}^m \alpha_i \phi_{\gamma i} \right\|_2^2 \quad (4.21)$$

en prenant en compte les m termes qui ont été trouvés par l'algorithme Matching pursuit. On construit ainsi une solution pour laquelle le résidu est orthogonal à tous les vecteurs utilisés.

IV.5.3 Basis Pursuit (BP)

Face à la difficulté que peuvent avoir les algorithmes gloutons de poursuite avec certains signaux ayant des composantes dans des bases différentes, pour lesquels l'algorithme perd beaucoup à corriger les erreurs dues aux approximations de ses premiers termes, Chen, Donoho et Saunders [52, 60] ont développé une méthode appelée Basis Pursuit qui résout le problème (P_1) . Il s'agit d'un problème convexe, et la norme ℓ_1 n'étant pas différentiable en 0, elle favorise la parcimonie de la solution. Soit $\lambda > 0$, on note $(P_{1,\lambda})$ le problème régularisé associé :

$$(P_{1,\lambda}) : \min_{\alpha \in \mathbb{R}^N} \frac{1}{2} \|x - \Phi\alpha\|_2^2 + \lambda \|\alpha\|_1 \quad (4.22)$$

Ce problème est soluble par des éléments d'optimisation convexe. La complexité de l'algorithme est importante, la solution est en général satisfaisante, sans toutefois atteindre le résultat qui aurait été obtenu avec la norme ℓ_0 .

IV.6 Parcimonie et entropies

Pour un vecteur $x \in \mathbb{R}^n$ que l'on normalise en norme ℓ_1 , chaque élément $|x_i|$ peut s'interpréter comme une probabilité d'utiliser cet $i^{\text{ème}}$ atome de la base \mathbb{R}^n . L'entropie permet de caractériser la distribution d'une loi de probabilité : plus elle est grande, plus il y a de désordre. Si le support de la loi de probabilité est borné, la distribution qui maximise l'entropie est la distribution uniforme. Inversement, si l'entropie est faible, il y a peu de désordre : la distribution est concentrée sur quelques états. Les entropies de Rényi et de Shannon sont des mesures de cette parcimonie.

IV.6.1 Entropie de Rényi

Soit une variable aléatoire discrète X à N états de probabilités respectives $(p_k)_{k \in [1,N]}$, l'entropie de Rényi R_α de X pour $\alpha > 0$ est définie comme :

$$R_\alpha(X) = \frac{1}{1-\alpha} \log \left(\sum_{i=1}^N p_i^\alpha \right) \quad (4.23)$$

Cependant, si on s'intéresse au logarithme de la norme ℓ_p

$$\log \|x\|_p = \log \left(\sum_{i=1}^N |x_i|^p \right)^{\frac{1}{p}} = \frac{1}{p} \log \left(\sum_{i=1}^N |x_i|^p \right) \quad (4.24)$$

On retrouve l'expression de l'entropie de Rényi R_p , au facteur multiplicatif $\frac{p}{1-\alpha}$ près. Ce facteur est strictement positif si $0 < p < 1$. De fait, minimiser la norme ℓ_p est analogue au fait de minimiser l'entropie de Rényi de la répartition de l'amplitude des coefficients sur les atomes du dictionnaire.

Cela donne une autre explication au fait que la norme ℓ_p avec $0 < p < 1$ constitue un indicateur de parcimonie.

IV.6.2 Entropie de Shannon

L'entropie de Shannon est l'un des critères de parcimonie le bien adoptée [61]. Ainsi, les coefficients ondelettes correspondants à l'entropie de Shannon minimum sont considérés comme le résultat le moins dense (plus parcimonieux). L'entropie de Shannon d'une variable aléatoire discrète X prenant les valeurs $(x_k)_{k \in [1, N]}$ et de probabilités respectives $(p_k)_{k \in [1, N]}$ ($\sum_{k=1}^N p_k = 1$) est notée $H(x)$ et est définie par:

$$H(X) = \sum_{k=1}^N -p_k \log p_k \quad \text{avec} \quad p_k = \alpha_i / \sum_j \alpha_j \quad (4.25)$$

$H(X)$ est donc définie comme espérance mathématique, ou valeur attendue, de la quantité d'information. La quantité d'information étant définie par $\log p_k$. Cette définition de l'entropie comme mesure de la quantité d'information produite par un processus discret aléatoire vérifie les propriétés suivantes:

Soit $H(x) = H_n(p_1, \dots, p_n)$:

1. L'entropie est une fonction continue des probabilités de chaque occurrence. C'est-à-dire que pour tout n , $H_n(p_1, \dots, p_n)$ est une fonction continue de chaque occurrence.
2. L'entropie d'une variable aléatoire uniforme augmente avec le nombre des occurrences. C'est-à-dire que pour tout n , $H_n(p_1, \dots, p_n) < H_{n+1}(p_1, \dots, p_n, p_{n+1})$.
3. L'entropie d'une variable aléatoire est égale à la somme pondérée des entropies des sous choix possibles le cas échéant. L'entropie doit donc être une somme pondérée des probabilités de ses occurrences. En particulier, une probabilité nulle ne contribue pas à l'entropie. Pour tout n , on a $H_{n+1}(p_1, \dots, p_n, 0) = H_n(p_1, \dots, p_n)$.
4. Remarquons que les conditions d'existence de l'entropie (à savoir que: $\sum_{k=1}^N p_k = 1$ et $0 \leq p_k \leq 1$) donnent l'encadrement : $0 \leq H(X) \leq \log N$.

IV.7 Formulations de la parcimonie dans un espace redondant adapté

D'après ce qui a été dit précédemment, si on veut que le signal que l'on observe soit bien représenté, il faut un dictionnaire qui leur soit adapté. On veut donc mettre dans le dictionnaire toutes les formes ou structures qui apparaissent dans nos données. Cependant, les signaux réels étant très variés il faudrait construire un dictionnaire extrêmement long et le calcul de la représentation d'un signal dans un tel dictionnaire serait difficile en un temps raisonnable. Il est donc préférable d'avoir un dictionnaire sur lequel on sache appliquer une transformée (i.e. Φ^T) rapide, et qui permette aussi une synthèse rapide (i.e. Φ) d'un signal à partir de ses coefficients. De nombreuses transformées ont été développées qui permettent d'approximer les signaux, comme les ondelettes. Utiliser des transformées présentant de la

redondance peut cependant être très bénéfique, comme indiqué par Coifman et Donoho dans le cas d'ondelettes redondantes [62]. Les ondelettes redondantes $W \in \mathbb{R}^{n \times p}$, avec $p > n$, forment des dictionnaires à transformée et reconstruction rapide et se montre appropriées afin d'obtenir une représentation plus parcimonieuse. Les translatées modulées de la fenêtre d'analyse générant une famille d'atomes $\Phi = W^T$, forment des dictionnaires à transformée et reconstruction rapide. Par la suite, nous supposons que la transformée représentée par la matrice W est une transformée en ondelettes. La parcimonie induite par les transformées en ondelettes permet de discriminer les coefficients du signal de ceux du bruit additif blanc Gaussien. En effet, ce type de bruit se distribue sur les petits coefficients et affecte très peu les grands coefficients. Récemment, plusieurs auteurs d'horizons divers ont argumenté en faveur de l'utilisation des représentations redondantes. Ces représentations, en offrant un plus large panel d'atomes, permettent une description encore plus parcimonieuse de signaux complexes et une plus grande flexibilité pour s'adapter à leur contenu morphologique. Dans le cas des ondelettes redondantes, cela permet aussi, d'obtenir l'invariance par translation de la transformée. Ces études, citées ci-dessus, ont montré que la formulation en analyse avait tendance à être plus flexible et mieux adaptée à des signaux naturels, qui ne peuvent généralement pas être synthétisés avec un faible nombre d'atomes.

IV.8 Choix de l'ondelette analysante adaptée au signal vibratoire impulsionnel

Dans cette section, les caractéristiques d'un signal correspondant à la réponse au choc d'un système mécanique sont en premier étudiées dans le but d'adapter l'ondelette appropriée. Un système linéaire soumis à une force d'impact type choc peut être formulé sous forme d'un seul degré de liberté de système de la forme suivante:

$$M \frac{d^2x}{dt^2} + C \frac{dx}{dt} + Kx = F\delta(t) \quad \text{avec} \quad \delta(t) = \begin{cases} 1, & t = b \\ 0, & \text{ailleurs} \end{cases} \quad (4.26)$$

où x représente le déplacement, M la masse concentrée, C le coefficient d'amortissement, K la raideur du système, F est une constante et $\delta(t)$ l'impulsion instantanée. La solution de l'équation (4.26) lorsque le déplacement et la vitesse initiale du système sont zéro peuvent être écrite en tant que [63]:

$$x(t) = \frac{F}{M\omega_d} \exp(-\zeta\omega_n t) \sin(\omega_d t) \quad (4.27)$$

$\omega_n = \sqrt{K/M}$ est la fréquence naturelle non-amortie et $\omega_d = \omega_n \sqrt{1 - \zeta^2}$ la fréquence naturelle amortie. $\zeta = C/2M\omega_n$ est le facteur d'amortissement et $\varphi = \tan^{-1}(\zeta/\sqrt{1 - \zeta^2})$ est l'angle de phase. Il convient de noter que la réponse au choc oscille avec la pulsation propre du système d'un degré de liberté. Toutefois, les machines complexes présentent de nombreux degrés de liberté. Ainsi, un système à plusieurs degrés de liberté est une combinaison linéaire des p systèmes d'un degré de liberté. Cette généralisation impose que la réponse au choc oscille avec les p fréquences. La réponse d'un système p degrés est [64]:

$$x_p(t) = \sum_{k=1}^p A_k \exp(-\zeta_k \omega_{n_k} t) \sin(\omega_{d_k} t) \quad (4.28)$$

ζ_k où est le facteur d'amortissement, ω_{n_k} est la fréquence naturelle non amortie et ω_{d_k} est la fréquence naturelle amortie associés au k ième Mode. Comme indiqué dans l'équation (4.27), la réponse impulsive provoquée par l'impact due au choc externe est caractérisée par une oscillation à amplitude décroissante. Ainsi, selon le mécanisme d'adaptation de la transformée en ondelettes, la fonction d'ondelette de Morlet serait le choix approprié à l'extraction de ces types de fonctions. En raison de son expression mathématique (fonction cosinus avec décroissance exponentielle des deux cotés), l'ondelette de Morlet est très semblable à une impulsion et pourrait être très efficace pour l'extraction des impulsions périodiques noyées dans le bruit et de fournir une représentation très parcimonieuse du signal. Dans le domaine temporel, l'ondelette de Morlet peut être exprimée sous la forme [65]:

$$\psi(t) = \exp(-t^2/2) \exp(j\omega_0 t) \quad (4.29)$$

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) = \exp[-(t-b)^2/2a^2] \exp[j\omega_0(t-b)/a] \quad (4.30)$$

On observe que la fonction décroît exponentiellement sur les deux côtés. La transformée de Fourier de $\psi(t)$ est donnée par:

$$\Psi(\omega) = \Psi^*(\omega) = \sqrt{2\pi} \exp[-(\omega - \omega_0)^2/2] \quad (4.31)$$

$$\Psi_{a,b}(\omega) = \sqrt{2\pi} \exp[-(a\omega - \omega_0)^2/2] \exp(j\omega b) \quad (4.32)$$

IV.9 Optimisation de l'ondelette analysante de Morlet pour la détection des impulsions

Comme représenté dans l'équation (4.31), l'ondelette de Morlet dans le domaine fréquentiel est une fonction réelle ($\Psi(\omega) = \Psi^*(\omega)$). Cependant, cette fonction d'ondelettes n'est pas une ondelette admissible, i.e. $\int_{-\infty}^{\infty} \psi(t) dt \neq 0$. Dans la pratique, lorsque le paramètre ω_0 est suffisamment grand ($\omega_0 > 5$) rend la quantité $\int_{-\infty}^{\infty} \psi(t) dt \approx 0$. Elle est assez petite pour que la fonction de Morlet soit applicable pour l'analyse en ondelettes. Toutefois, lorsque la fréquence analysée est importante, le paramètre de dilatation devient faible et le spectre de la fonction d'ondelette de Morlet est large, ceci conduit à une mauvaise résolution temps-fréquence. Pour palier ce problème, une alternative consiste à modifier la fonction ondelette de Morlet en introduisant un paramètre β qui contrôle la forme de l'ondelette de base [66]. Ce paramètre permet l'optimisation de la résolution temporelle et fréquentielle de l'ondelette de Morlet. L'ondelette de Morlet modifiée utilisée dans ce travail est définie comme suit :

$$\psi(t) = \exp(-\beta^2 t^2/2) [\exp(j\omega_0 t) - \exp[-(\omega_0)^2/2]] \quad (4.33)$$

L'expression temporelle de l'ondelette de Morlet ainsi définie, peut être transformé dans le domaine fréquentiel en appliquant la transformée de Fourier:

$$\Psi(\omega) = \Psi^*(\omega) = \sqrt{2\pi}(\exp[-(\omega - \omega_0)^2/2\beta^2] - \exp[-(\omega^2 + \omega_0^2)/2\beta^2]) \quad (4.34)$$

L'équation (4.34) indique que $\Psi(\omega)$ est une réponse fréquentielle impulsionnelle avec une fréquence centrale arbitraire ω_0 . Ainsi, $\omega = 0$ conduit à la réponse en fréquentielle $\Psi(\omega) = 0$ et implique par conséquent que $\int_{-\infty}^{\infty} \psi(t)dt = 0$. De ce fait, l'ondelette mère modifiée $\psi(t)$ satisfait les conditions d'admissibilité [67]. En appliquant la formule de Parseval, la transformée en ondelettes d'un signal $x(t)$ en utilisant l'ondelette de Morlet modifiée prend la forme alternative suivante [61]:

$$W_\psi(a, b) = \frac{\sqrt{a}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} X(\omega)(\exp[-(a\omega - \omega_0)^2/2\beta^2] - \exp[-(a^2\omega^2 + \omega_0^2)/2\beta^2])\exp(j\omega b)d\omega \quad (4.35)$$

Il est montré à partir de l'équation (4.33), que la forme de l'ondelette de base est contrôlée par le paramètre β . D'une part, lorsque β est infini, l'ondelette de Morlet devient une fonction de Dirac avec la meilleure résolution temporelle. D'autre part, lorsque β prend une valeur nulle, l'ondelette de Morlet devient une fonction cosinus avec la résolution la plus fine de fréquence. Par conséquent, il existe toujours une valeur optimale du paramètre β conduisant à la meilleure résolution temps-fréquence pour un certain signal localisé dans le domaine temps-fréquence. Il convient à souligner que cette fonction ondelette de Morlet modifiée offre un meilleur compromis en termes de localisation en temps et en fréquence pour un signal impulsionnel, que la fonction ondelette de Morlet fait traditionnellement. Le paramètre β idéal est celui qui correspond au plus petit nombre de coefficients d'ondelettes dominants. L'entropie de Shannon décrite précédemment est utilisée comme critère de parcimonie le bien adoptée. Ainsi, les coefficients ondelettes correspondants à l'entropie de Shannon minimum sont considérés comme le résultat le moins dense (parcimonieux) et le facteur de forme β correspondant est adopté comme le paramètre optimal. L'entropie de Shannon est définie par l'équation suivante [68] :

$$H(X) = \sum_{k=1}^N -p_k \log p_k \quad \text{avec} \quad p_k = \alpha_i / \sum_j \alpha_j \quad (4.36)$$

où α_i sont des coefficients d'ondelettes. Le processus suivant décrit la procédure d'optimisation du facteur de forme β : i) varier β de 0,1 à 20 et calculer l'entropie E_n des coefficients correspondants. ii) trouver l'entropie de Shannon minimale E_n . iii) le facteur de forme β optimale correspondant à cette entropie minimale est retenu. Par conséquent, la fonction d'ondelette est définie par la substitution de la valeur β dans l'équation (4.33). La figure (4.2b) montre l'évolution de l'entropie de Shannon des coefficients d'ondelettes en fonction de la variation du facteur β suivant le processus proposé.

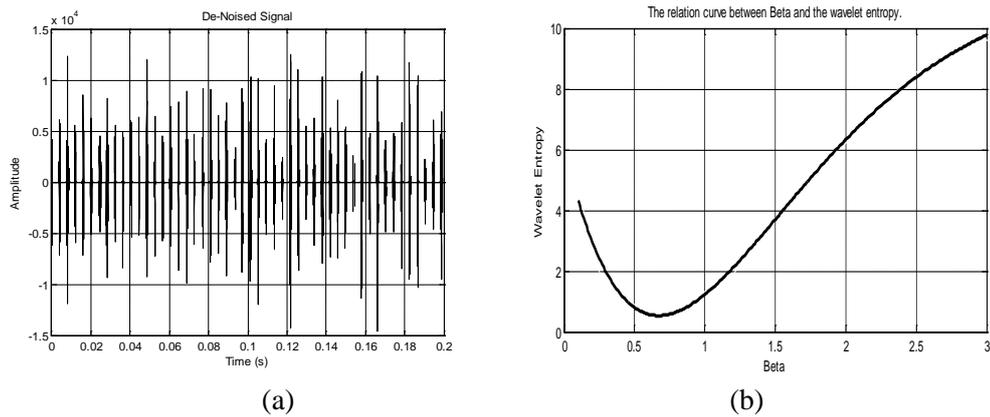


Fig. 4.2 (a) Signal test (b) Relation entre l'entropie de Shannon de coefficients et facteur de forme β

La figure (4.2b) montre que la valeur $\beta \cong 0.7$ du facteur de forme conduit à une valeur minimale de l'entropie des coefficients d'ondelettes. Ainsi, $\beta = 0.7$ est jugée la valeur optimale utilisée dans le présent travail. De plus, la figure (4.3) représente les scalogrammes de la transformée en ondelettes continue basé sur l'ondelette de Morlet modifiée du signal de test de la figure (4.2a) avec différentes valeurs de β respectivement : 0.1, 0.3, 0.7 et 1.0.

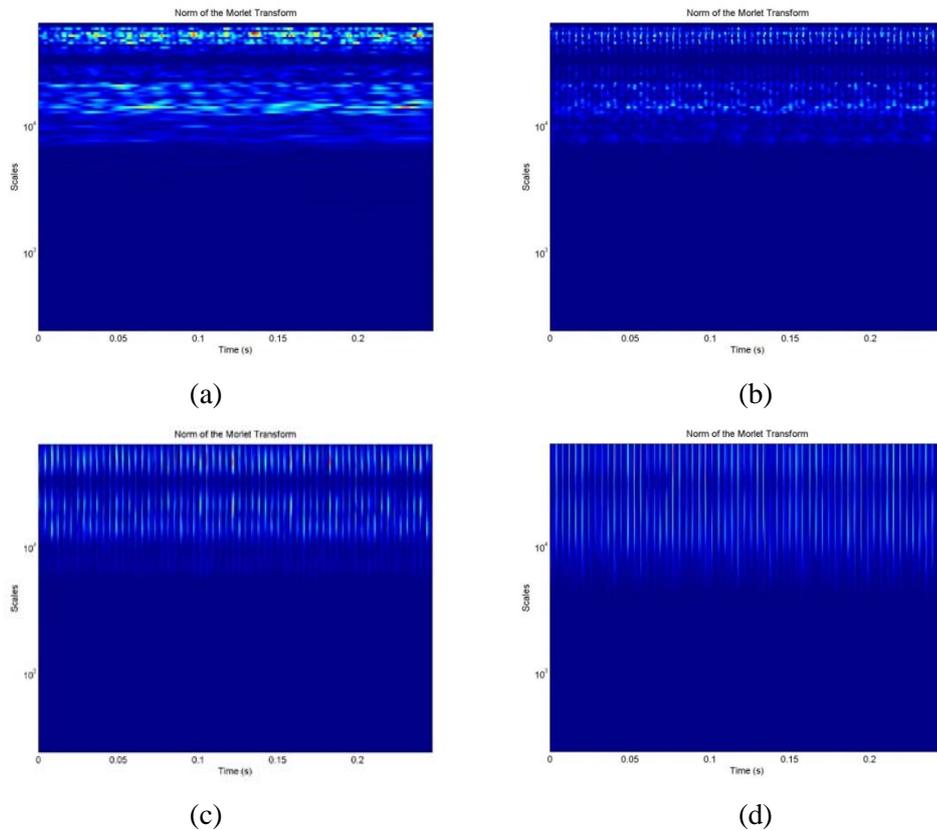


Fig. 4.3 CWT du signal test; (a): $\beta = 0.1$, (b): $\beta = 0.3$, (c): $\beta = 0.7$, (d): $\beta = 1.0$,

D'après la figure (4.3c), on peut voir que la CWT correspondante à la valeur optimale de $\beta = 0.7$ sépare efficacement les composants impulsionsnels. De plus, la figure (4.5b) qui représente le scalogramme relatif aux impulsions bruitées du signal de la figure (4.4b) présente clairement, une assez bonne séparation des composantes proches des impulsions et montre une localisation fréquentielle optimale, comparativement à la représentation de la figure (4.5a) du même signal décomposé en utilisant l'ondelette de Morlet standard. Par conséquent, il peut être conclu que le procédé basé sur l'entropie des coefficients d'ondelettes est très efficace pour la sélection de la fonction d'ondelette optimale.

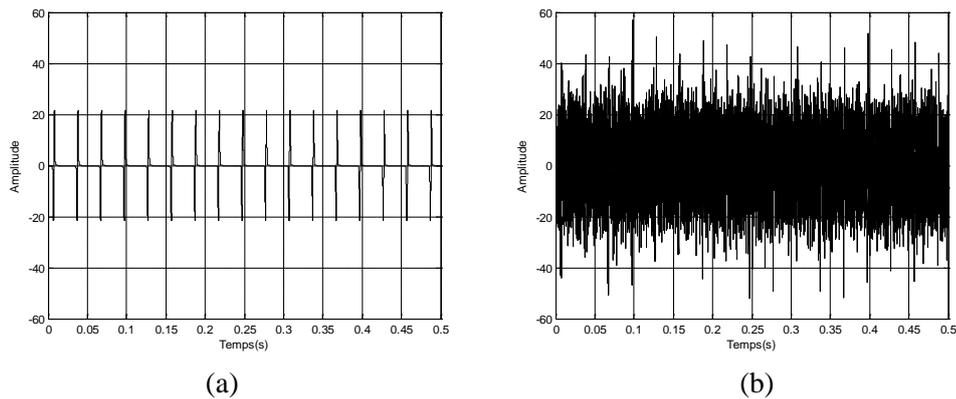


Fig. 4.4 (a): Impulsions simulées (b) Impulsions simulées avec un bruit lourd

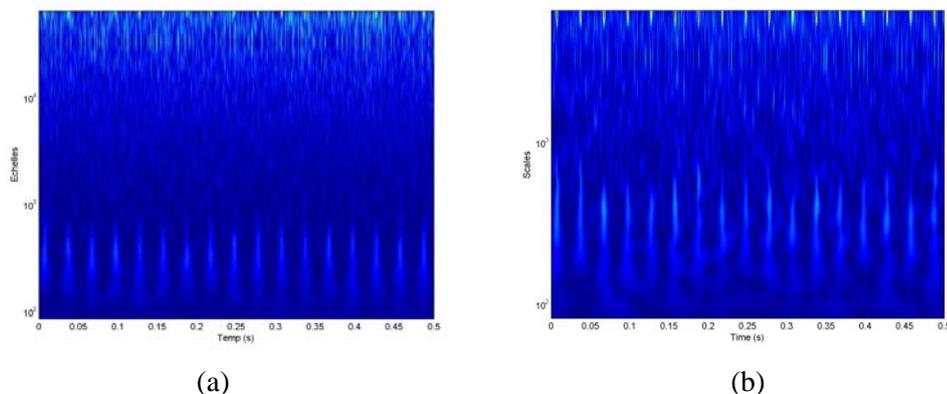


Fig. 4.5 Scalogramme relatif à la représentation des impulsions bruitées (a) en utilisant l'ondelette de Morlet standard (b) en utilisant l'ondelette de Morlet optimisée ($\beta = 0.7$)

IV.10 Application de la parcimonie dans les problèmes inverses

La simulation des mesures à partir des grandeurs physiques est un problème direct. Un problème direct établit une description mathématique des données observées connaissant l'objet et le mécanisme d'observation. Ce modèle doit être assez précis pour fournir une description correcte du phénomène physique d'observation, et assez simple cependant pour se prêter à un traitement numérique ultérieur. Cependant, remonter à ces grandeurs physiques à partir des mesures est loin d'être aussi simple. Un problème inverse consiste à estimer un ou plusieurs signaux, liés physiquement à partir d'observations, souvent incomplètes et bruitées. Ces problèmes inverses en présence de bruit sont ubiquistes en traitement du signal et un

domaine de recherche prolifique tant la littérature sur le sujet est abondante. Or, la caractéristique commune de ces problèmes de reconstruction ou de restauration de signal est qu'ils sont souvent mal-posés ou mal-conditionnés. Le problème inverse peut en effet avoir plus d'une solution, ou être très instable par rapport au bruit, aux incertitudes expérimentales, etc. Dans cette section, essentiellement bibliographique, on présente une vue générale des problèmes inverses, du concept de parcimonie, et de la façon dont des *a priori* parcimonieux peuvent être utilisés pour la résolution de problèmes inverses. On s'intéresse ici aux problèmes inverses linéaires, qu'on modélise par le modèle qui suit :

Soit $y \in \mathbb{R}^M$ et $s \in \mathbb{R}^N$. Soit $H : \mathbb{R}^M \rightarrow \mathbb{R}^N$ un opérateur linéaire, le modèle de dégradation se simplifie alors à :

$$y = Hs + b, \quad b \sim N(0, \sigma^2) \quad (4.37)$$

où $b \in \mathbb{R}^M$ représente un bruit additif Gaussien. Dans ce modèle, les observations sont représentées par y . Dans beaucoup de problèmes, l'opérateur H peut être modélisé (c'est le problème direct), et l'on cherche à estimer s . Cependant, H n'est pas toujours connu, et peut même devenir l'objet d'intérêt, connaissant s et y . Lorsque ni H ni s ne sont connus, on parle de problème aveugle. Dans la majorité des cas, on suppose l'opérateur H linéaire. Enfin, on distinguera aussi deux cas importants : le cas sur-déterminé, lorsque $M > N$ et le cas sous-déterminé lorsque $M < N$. En raison de la perte d'information induite par l'action de H , le problème de restauration de s à partir de y est mal posé au sens de Hadamard. Un cadre pratique pour aborder les problèmes inverses consiste à le poser sous forme d'un problème de régularisation pénalisée:

$$\tilde{s} = \operatorname{argmin} \mathcal{L}(y, H, s) + P(s) \quad (4.38)$$

où \mathcal{L} est un terme d'attache aux données modélisant l'écart entre les observations y et les signaux s à travers l'opérateur H . Il est courant d'utiliser la distance ℓ_2 . P est un terme de régularisation (ou pénalisation). Ce terme a pour but d'apporter une information *a priori* sur le signal recherché, indépendamment des observations. C'est ce terme qui est utilisé pour privilégier la parcimonie du signal recherché. Le fait que y puisse être représenté de manière parcimonieuse dans un dictionnaire donné est une information sur sa morphologie qu'il peut alors être utile de prendre en compte. L'utilisation de la parcimonie pour les résoudre a d'abord été amorcée par les méthodes de seuillage dans les bases d'ondelettes [69]. La régularisation par des pénalités favorisant la parcimonie est encore plus récente. En effet, il est pratique de modéliser le signal d'intérêt s comme une somme d'atomes élémentaires choisis dans un dictionnaire tel que : $s = \Phi\alpha$. Le problème d'optimisation devient alors :

$$\tilde{\alpha} = \operatorname{argmin} \mathcal{L}(y, H, \Phi\alpha) + P(\alpha) \quad (4.39)$$

Le choix du dictionnaire varie selon le signal. On suppose ici que la solution $s = \Phi\alpha$ est parcimonieuse, pour les signaux impulsionnels, les transformées en ondelette sont bien adaptées.

Ce problème est équivalent à la résolution du Lagrangien, donnant la version régularisée suivante [70] :

$$\min_{\alpha \in \mathbb{R}^N} \frac{1}{2\sigma^2} \|y - H\Phi\alpha\|_2^2 + \lambda M(\alpha) \quad (4.40)$$

ou $M(\alpha)$ est une pénalité favorisant la parcimonie. La démarche classique consiste donc à incorporer des connaissances supplémentaires permettant de réduire l'espace des solutions candidates. Ces informations *a priori* traduisent par exemple des propriétés de régularité de la solution (approche variationnelle), ou son appartenance à un modèle stochastique (approche bayésienne). Le paramètre λ est généralement difficile à choisir et dépend de b de façon non triviale. Dans le cadre de cette thèse on s'intéresse à deux problèmes inverses particuliers: la restauration des signaux (débruitage) et la séparation de sources. À travers ces deux exemples, on montrera comment les approches par optimisation avec un choix de parcimonie peuvent être utilisées avec succès. On formule pour chacun de ces deux problèmes, le choix du problème d'optimisation utilisé pour résoudre ces problèmes inverses.

IV.11 Parcimonie et restauration des signaux (débruitage)

Dans le cadre de la restauration d'un signal, une écriture générique possible de l'équation de dégradation, qui relie l'observation signal y au signal source recherché s , est :

$$y = s \odot b \quad (4.41)$$

où b est le bruit reflétant les fluctuations aléatoires non mesurées contaminant le signal ou une perturbation déterministe représentant l'imperfection du modèle, \odot est toute composition à deux arguments (e.g. $+$ pour un bruit additif, \times pour un bruit multiplicatif, etc.). Ce problème est sans doute l'un des plus anciens à la fois en traitement du signal et de l'image, mais aussi en statistiques. Par ailleurs, les débruiteurs basés sur un seuillage dans le domaine des représentations parcimonieuses ont clairement fait leurs preuves aussi bien en termes de performances pratiques que théoriques. En particulier, il est établi que la qualité de l'estimation par seuillage des coefficients d'une transformée est étroitement liée à celle de la compressibilité du signal d'intérêt dans cette transformée. Dans le cas d'un bruit additif Gaussien, le problème (4.41) devient :

$$y = s + b \quad b \sim N(0, \sigma^2) \quad (4.42)$$

La variance σ^2 est : soit connue soit estimée à partir des observations en utilisant des estimateurs robustes tels que la déviation médiane absolue (MAD) [20, 70]. On verra dans la section (V.4.1) dans un contexte Bayésien, que les signaux observés sont contaminés par du bruit blanc Gaussien centré et d'écart-type σ . La distribution d'une telle variable aléatoire $b \in \mathbb{R}^N$ est donnée par [71]:

$$f(b) = \frac{1}{\sqrt{(2\pi\sigma^2)^n}} \exp\left(-\frac{1}{2\sigma^2} \|b\|_2^2\right) \quad (4.43)$$

Une des particularités du bruit blanc Gaussien est que son énergie moyenne se répartit uniformément dans toutes les directions de l'espace, n'en privilégiant donc aucune. À l'inverse quand le signal recherché est parcimonieux, son énergie est concentrée sur quelques grands coefficients non nuls. On peut donc espérer séparer le signal du bruit en recherchant ces quelques grands coefficients. Cela revient à résoudre le problème suivant, qui recherche un signal s aussi parcimonieux que possible tout en restant à une distance de y inférieure à ζ . Le choix de ζ dépend du niveau de bruit. Afin de minimiser l'erreur entre y et s , on cherche à minimiser une mesure de distorsion, notée \mathcal{L} :

$$\tilde{s} = \underset{s}{\operatorname{argmin}} \mathcal{L}(y - s) \quad (4.44)$$

Comme pour la section précédente, on cherche à représenter s dans un dictionnaire adapté. Ainsi, le problème (4.42) s'écrit de manière équivalente :

$$\tilde{\alpha} = \underset{\alpha}{\operatorname{argmin}} \mathcal{L}(y - \Phi\alpha) \quad (4.45)$$

L'estimation de s étant donnée par $\tilde{s} = \sum_{i \in I} \tilde{\alpha}_i \phi_i$. Dans ce travail, on utilisera pour \mathcal{L} , en général, une norme ℓ_2 . La minimisation de cette norme nous impose de chercher un signal (assez proche) de l'observation y , dans le sens où l'énergie de l'erreur doit être faible. Le problème (4.45) admettant une infinité de solutions sur les coefficients α , on doit le régulariser en fonction du type de solution que l'on souhaite obtenir. Une des premières régularisations proposées est celle de Tikhonov [70], le problème se posant alors comme suit :

$$\min_{\alpha} \|y - \Phi\alpha\|_2^2 + \lambda \|\alpha\|_2^2 \quad (4.46)$$

où λ est un paramètre de Lagrange que l'on peut régler. Cette régularisation très classique permet de contrôler l'énergie de la solution, et d'éviter ainsi qu'elle n'explose. Comme pour le cas de la reconstruction exacte, ce problème admet une solution analytique simple :

$$\hat{\alpha} = \Phi^*(\Phi\Phi^* + \lambda I_d)^{-1}y \quad (4.47)$$

Où I_d est l'opérateur d'identité. Dans le cas où Φ est orthogonale la solution s'obtient simplement par une pondération de tous les coefficients avec le même poids. En notant $\bar{y} = \Phi^*y$:

$$\hat{\alpha}_i = \frac{\bar{y}_i}{1 + \lambda} \quad (4.48)$$

Il apparaît clairement que la solution obtenue n'est pas parcimonieuse dans le sens où aucun coefficient n'est mis à zéro. Comme lorsqu'on cherche une décomposition exacte, il convient de régulariser le problème avec une mesure \mathcal{M} favorisant une solution parcimonieuse :

$$\min_{\alpha} \|y - \Phi\alpha\|_2^2 + \lambda \mathcal{M}(\alpha) \quad (4.49)$$

Il existe deux formulations équivalentes au problème d'optimisation

$$\min_{\alpha} \mathcal{M}(\alpha) \quad \text{t. q.} \quad \|y - \Phi\alpha\|_2^2 \leq \zeta \quad (4.50)$$

Ou bien :

$$\min_{\alpha} \|y - \Phi\alpha\|_2^2 \quad \text{t. q.} \quad \mathcal{M}(\alpha) \leq m \quad (4.51)$$

Suivant les cas, on utilisera l'une ou l'autre des formulations : obtenir une décomposition parcimonieuse au détriment de l'erreur de reconstruction ou vice-versa. Si, pour un même problème, ces deux formulations sont équivalentes, le choix de l'*a priori* des paramètres ne se fait pas de la même manière. Dans la formulation (4.50), on contrôle directement le bruit $\|y - \Phi\alpha\|_2^2 \leq \zeta$, ce qui peut être avantageux dans un problème de débruitage.

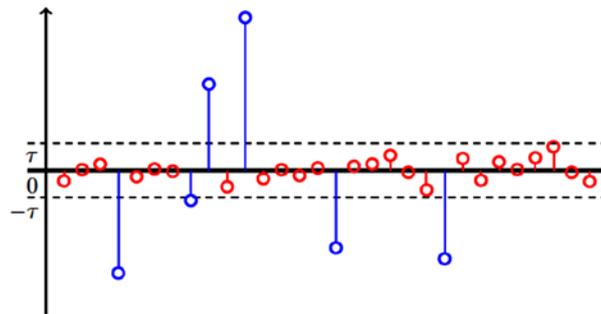


Fig. 4.6 Signal temporel parcimonieux seuillé

IV.12 Conclusion

L'objectif de ce chapitre de thèse était de montrer la pertinence des modèles parcimonieux pour traiter des problèmes inverses en vibration, notamment en tirant parti de la nature des signaux à reconstruire. Les contributions, décrites dans ce document se basent sur ce concept. Quelques définitions ont été données, l'importance de la redondance a été mise en exergue, ainsi que le lien avec l'analyse harmonique computationnelle. Cet objectif est, nous le pensons, rempli. En effet, un modèle parcimonieux a été mis à contribution pour la modélisation des problèmes inverses à étudier notamment dans le cas de débruitage du signal vibratoire impulsionnel et de localisation des sources vibratoires bruitées.

Chapitre V

Approche Bayésienne pour le débruitage des signaux de vibration fortement bruités

Chapitre 5

Approche Bayésienne pour le débruitage des signaux de vibration fortement bruités

Résumé

L'obtention d'information à partir de mesures corrompues par un bruit reste un problème ouvert en traitement du signal. Dans ce chapitre, nous présentons les principales méthodes développées à ce jour s'appuyant sur les représentations dans le domaine des ondelettes. Nous distinguons deux approches : une approche classique de type seuillage de coefficients dite estimation non-paramétrique et une approche Bayésienne basée sur un modèle statistique des coefficients. Un aperçu global sur les différentes méthodes du problème de la régression non-paramétrique est présenté en premier. En effet, il suffit de seuiller les coefficients du signal observée et d'inverser la transformée pour obtenir une estimée du signal utile. Dans le contexte Bayésien et afin de reconstruire le signal non bruité, il est nécessaire d'imposer des contraintes et particulièrement celle concernant le renforcement de la parcimonie sur la solution reconstruite, ce qui consiste à choisir un modèle a priori adapté aux coefficients d'un signal impulsif. Cependant, la distribution du signal n'est pas toujours connue et elle s'avère assez souvent difficile à estimer en pratique. Nous nous penchons alors, dans ce chapitre, sur la modélisation statistique des signaux et nous décrivons les propriétés statistiques permettant de les caractériser dans l'espace des transformées d'ondelettes.

Sommaire

V.1	Introduction	81
V.2	Estimation non-paramétrique (débruitage)	82
V.3	Approche classique sans <i>a priori</i>	83
	V.3.1 Choix du seuil	87
	V.3.2 Estimation du niveau de bruit	88
V.4	Formulation Bayésienne du problème d'identification	89
	V.4.1 Choix de la fonction de vraisemblance	90
	V.4.2 Choix de la distribution de probabilité <i>a priori</i> de parcimonie	91
V.5	Etude expérimentale	93
V.6	Conclusion	99

V.1 Introduction

En traitement de signal, le problème de l'élimination du bruit gaussien des signaux de vibration est un problème inverse intéressant à étudier dans ce chapitre à travers une représentation parcimonieuse. En effet, les données observées sont très souvent composées de l'information utile (signal) corrompue par une perturbation généralement indépendante. La perturbation (bruit) peut être interne au système physique de traitement du signal ou externe, due aux interactions entre le signal et les éléments de son milieu de propagation. Les causes engendrant une perte d'information sont d'origine stochastique puisque les observations sont des mesures soumises à des fluctuations aléatoires non mesurables que l'on regroupe sous la dénomination de bruit. La distribution du signal n'est pas toujours connue et elle s'avère assez souvent difficile à estimer en pratique. Il en est de même concernant la nature de la perturbation : sa distribution n'est pas connue et est assez souvent difficile à modéliser. Face à cette situation, il est donc nécessaire de développer des méthodes à coût minimum en termes de connaissances disponibles sur la nature des composantes du signal. On parle alors de décision et d'estimation (plus généralement de statistique) non-paramétriques au sens où l'on ne tient pas compte de connaissances *a priori* sur les distributions des signaux constituant l'observation. Par ailleurs, les transformées parcimonieuses présentent de très bonnes propriétés mathématiques : l'énergie de signaux réguliers ou réguliers par morceaux se répartit sur un petit nombre de coefficients de grandes amplitudes, ce qui offre ainsi un cadre naturel non linéaire pour estimer le signal utile. Cette capacité de compacter l'énergie de signaux tandis que la contribution du bruit demeure inchangée rend possible et performante l'estimation non-paramétrique par simple seuillage des coefficients du signal observé dans une base d'ondelettes. C'est ainsi que depuis les travaux pionniers de Donoho et Johnstone, sur la construction des seuils appropriés, divers procédés sont élaborés pour le seuillage des coefficients d'ondelettes [18-20], entre autres). Cette démarche repose sur le choix d'une valeur de seuil tel que: Minimax, VisuShrink et SureShrink. Cependant, le seuil VisuShrink, dite 'seuil universel' utilise une seule valeur seuil commune pour tous les niveaux (échelles) des coefficients d'ondelettes. La limitation du seuil VisuShrink est de ne pas considérer la minimisation de l'erreur quadratique moyenne (MSE), ce ceci conduit à la suppression excessive des coefficients d'ondelettes dont certains sont porteur d'informations utiles. Par conséquent le signal récupéré se trouve trop adouci. Par ailleurs, le SureShrink est un seuil adaptatif relatif au niveau de décomposition en ondelettes est lié à l'estimation du risque sans biais de Stein (SURE) de plus, il minimise l'erreur quadratique moyenne. Néanmoins, les performances de ces méthodes de débruitage traditionnelles basées sur la décomposition en ondelettes sont grandement affectées par les niveaux d'énergie relative aux coefficients du signal et au niveau du bruit considéré. Ces techniques n'utilisent aucune information concernant le signal à identifier. Lorsqu'il s'agit de signaux lisses, des résultats satisfaisants peuvent être obtenus généralement par la manipulation du seuil. La raison sous-jacente est que les signaux lisses sont caractérisés dans la base des ondelettes par un petit nombre de grands coefficients, par contre, les composants transitoires qui varient rapidement sont traités comme du bruit. Ces techniques de débruitage classiques ne conviennent pas à l'analyse du signal de vibration de type choc qui émane des roulements ou des engrenages défectueux car

les impulsions d'intérêt synonyme de défaillance de ces organes ne sont pas lisses. Par conséquent, il est beaucoup plus difficile à débruiter des signaux impulsionnels où les coefficients d'ondelettes ne sont pas si concentrés. Dans ce contexte, il semble naturel de pousser les coefficients de signaux vers un ensemble parcimonieux, ayant le moins possible d'éléments non nuls, tout en étant compatible avec les données. En termes probabilistes, ceci correspond à décrire les coefficients d'ondelettes par une fonction de densité fortement non gaussienne (aussi appelé super-gaussienne) présentant un pic marqué à zéro. Par conséquent, un critère de débruitage naturelle peut résulter de cette différence statistique entre les coefficients du signal et le bruit. Dans ce travail, nous décrivons une méthode alternative de renforcement de la parcimonie basée sur la modélisation probabiliste de l'*a priori* dans une approche Bayésienne. La distribution de l'*a priori* dans l'approche Bayésienne exprime la connaissance préalable sur les coefficients, sous la forme d'une distribution de probabilité. Étant donné un modèle approprié de la distribution de l'*a priori* des coefficients, le signal d'intérêt peut alors être estimé par les calculs comme étant un problème inverse Bayésien. La distribution de l'*a priori* est d'une importance primordiale dans la construction d'un système d'estimation appropriée. Le choix simple serait un modèle gaussien de moyenne nulle. Toutefois ce modèle est très limité, et ne renforce pas la parcimonie en aucune manière particulièrement utile. De ce fait, une distribution étroite de l'*a priori* à queue lourde est introduite dans ce travail afin de considérer la forte non-gaussianité des coefficients liés au signal impulsionnel. L'inférence Bayésienne concerne l'estimation du maximum *a posteriori* (MAP) des paramètres à partir des données observées. Cette estimation nécessite l'adoption d'un modèle probabiliste paramétrique des données sous la forme d'une distribution de probabilité conditionnée par la connaissance des paramètres inconnus. La connaissance antérieure est quantifiée par la distribution de l'*a priori*. Le recours à l'utilisation de l'approche Bayésienne est motivée par le fait que l'approche constitue un moyen plus facile pour traduire les connaissances antérieures concernant les données et en particulier la parcimonie des coefficients d'intérêt et un meilleur moyen de prise en compte des caractéristiques du bruit.

V.2 Estimation non-paramétrique

Le problème de l'estimation non paramétrique (débruitage) connu sous le nom de 'régression non-paramétrique' en statistiques mathématiques, a constitué un outil fondamental au cours des vingt dernières années dans le domaine du traitement statistique du signal, et est toujours un domaine actif de recherche. L'objectif consiste à recouvrer le signal d'intérêt mettant en avant différents types d'estimateurs linéaires comme une fonction inconnue corrompue par du bruit, sans spécification d'un modèle explicite sur la fonction à retrouver. Les techniques de régression non-paramétrique (ou débruitage) offrent un panel d'outils simples d'emploi permettant de récupérer une structure, sans imposer un modèle paramétrique au préalable. Au cours des années 80 et 90, des centaines d'articles sont parus, solution à la question de la régression non-paramétrique. Parmi les plus populaires se trouvent notamment ceux reposant sur l'emploi de séries orthogonales ou de splines [72, 73].

Cependant ces différents types d'estimateurs souffrent d'un comportement peu satisfaisant dans le cas d'échantillons de taille finie et aussi de beaucoup de problèmes lorsque la fonction est irrégulière ou avec des singularités. Bien que certains aient atteint des performances optimales du point de vue asymptotique, leur implémentation s'avère encore complexe, et leurs domaines d'application souvent limités à des fonctions présentant certaines propriétés de régularité. Durant les années 90, le domaine de la régression non-paramétrique a été dominé par deux types d'estimateurs non linéaires introduits par Donoho & Johnstone [19], à savoir la contraction par ondelettes (wavelet shrinkage) et le seuillage par ondelettes (wavelet thresholding). Ces estimateurs, grâce aux efforts menés par Mallat pour aller de l'analyse multi-résolution par ondelettes à la théorie des bancs de filtres, sont facilement mis en application par des algorithmes rapides, et sont ainsi très attrayants dans leur mise en œuvre [24]. En statistiques mathématiques, de nouveaux travaux théoriques ont par ailleurs montré l'optimalité de la régression non-paramétrique dans le domaine des ondelettes au sein d'espaces fonctionnels beaucoup plus larges que ceux utilisés précédemment. En d'autres termes, les différentes classes de fonctions envisagées couvrent aussi bien celles de Sobolev que celles des classes de fonctions irrégulières comme celles appartenant aux espaces de Besov. Cet accroissement du domaine d'application de ces différents estimateurs se révèle être d'un grand intérêt dans le cadre d'analyse de signaux réels, comme dans le cas de la parole, d'électrocardiogrammes ou des signaux sismiques [74]. Cependant, l'analyse d'un point de vue asymptotique ne permet pas de prédire le comportement d'un estimateur pour un échantillon de taille finie et ainsi de déterminer la taille minimale de l'échantillon permettant une estimation statistique. L'analyse du risque exact réalisée par Marron [75] et appliquée aux deux types de seuillage par ondelettes introduits par Donoho & Johnstone [20], à savoir le seuillage minimax et le seuillage universel, a permis la détermination du domaine d'application des estimateurs employant la transformée en ondelettes. Bruce & Gao [76] et Antoniadis [77] ont également fourni des outils permettant l'analyse du comportement de ces estimateurs en présence d'échantillons de taille finie. Il a été montré que ces types d'estimateurs sont asymptotiquement presque optimaux ou optimaux au sens minimax alors que les estimateurs linéaires traditionnels sont sous-optimaux en ce qui concerne les classes de fonctions appartenant aux espaces de Besov [78]. Depuis la parution des articles fondateurs par Donoho & Johnstone [19, 20, 79], la littérature de traitement du signal et d'image a assisté à une abondance de papiers appliquant ou proposant des modifications de l'algorithme original dans des problèmes d'estimation et/ou restauration des signaux et d'images. Différentes alternatives au seuillage par ondelettes ont été développées. Ainsi, Donoho & Johnstone [19] proposèrent l'estimateur SURE basé sur l'estimation du risque sans biais de Stein. Abramovich & Benjamini [80, 81] considérèrent le seuillage sous l'angle d'une procédure de test d'hypothèses multiples. Il a été montré que ce type d'approche offrait un gain de performances au sens de l'erreur quadratique moyenne.

V.3 Approche classique sans *a priori*

Cette approche non-paramétrique est motivée par le fait que dans la plupart des situations rencontrées en pratique, les distributions des signaux constituant l'observation ne sont pas

connues et sont souvent difficiles à estimer avec précision. Lorsqu'on ne dispose pas d'information sur la distribution du signal que l'on cherche à estimer, il est tout de même possible d'estimer ce signal de manière satisfaisante si l'on dispose d'une base qui permet de le distinguer (discriminer) du bruit. L'idée est la suivante. Si nous disposons d'une base orthonormale capable de 'concentrer' l'énergie de ce signal sur un petit nombre de coefficients de grandes amplitudes, alors la projection de l'observation sur une telle base sera assez sensiblement composée d'une petite classe de coefficients relativement grands (donc contenant principalement l'information sur le signal) et une grande classe de petits coefficients dont la nature est presque essentiellement de bruit. L'estimation consisterait alors à récupérer et éventuellement à traiter les grands coefficients, après avoir forcé à zéro, les petits coefficients. Une telle estimation est donc réalisable avec une fonction dite de seuillage car elle va dépendre d'un seuil en dessous duquel, tout coefficient sera forcé à zéro. On dit d'une base capable de concentrer une grande partie de l'énergie d'un signal sur un petit nombre de coefficients de grandes amplitudes qu'elle effectue une représentation parcimonieuse (sparse) du signal. Cette définition est cependant très peu précise car les termes 'petits' et 'grands' ne sont pas clairement définis. Nous parlerons donc plutôt de représentation parcimonieuse au sens large. Nous dirons alors d'une base qu'elle réalise une représentation parcimonieuse au sens strict pour un signal donné si l'énergie de ce signal est 'presque entièrement' contenue dans un petit nombre de coefficients de grandes amplitudes. Dans ce dernier cas de parcimonie, presque tous coefficients qualifiés de 'petits' sont en fait quasiment nuls ou de très petites amplitudes, et ne contiennent donc pas d'information significative sur le signal. Dans le cas d'une représentation parcimonieuse au sens large, mais non strictement parcimonieuse, il peut être souvent utile de traiter également les petits coefficients. Dans ce cas, nous considérerons, non pas une fonction de seuillage, mais une fonction dite d'atténuation en ce sens qu'elle ne force pas systématiquement à zéro les petits coefficients, mais leur impose une plus grande atténuation que les grands coefficients. Et puisque la fonction de seuillage peut être vue comme étant une fonction d'atténuation dégénérée (qui opère une atténuation brutale par mise à zéro des petits coefficients et une autre éventuelle atténuation des grands coefficients), nous garderons la terminologie 'fonction d'atténuation' pour l'estimation des coefficients via la représentation parcimonieuse. Si la fonction d'atténuation inclue un seuillage, nous parlerons alors de 'fonction de seuillage'. Pour réaliser l'estimation non-paramétrique décrite plus haut, il faut donc choisir une représentation parcimonieuse et une fonction d'atténuation appropriée. En ce qui concerne le choix de la représentation, on sait que les bases d'ondelettes satisfont au critère de parcimonie de la représentation pour les signaux réguliers ou réguliers par morceaux. De plus, il a été montré dans [19] que de simples fonctions de seuillage comme le seuillage *dur* (forcer à zéro ou conserver) et le seuillage *doux* (forcer à zéro ou atténuer d'une valeur égale au seuil) permettent d'obtenir de résultats très satisfaisants. Dans le modèle de perturbation considéré seule la variable aléatoire y est observée. Le modèle sous-jacent concerne un signal bruité :

$$y = s + \varepsilon \tag{5.1}$$

Dans ce modèle, s est le signal d'origine et ε est le bruit reflétant les fluctuations aléatoires contaminant le signal utile, ces échantillons indépendants sont considéré une variable gaussienne de moyenne nulle et de variance σ_ε^2 ; c'est-à-dire $\varepsilon \sim \mathcal{N}(0, \sigma^2 I)$, avec I désignant une matrice identité de taille appropriée. L'objectif est de recouvrer l'information s contenue dans le signal bruité y sans pour autant faire d'hypothèses sur une structure paramétrique de s . Étant donné le vecteur y qui représente les valeurs échantillonnées du signal observé, la décomposition de y dans un dictionnaire donnée est $D = \Phi^T y$ où Φ est une matrice de taille $n \times p$. Dans le cas de la transformée d'ondelettes discrète orthogonale, la matrice $\Phi = W$ où W est une matrice orthogonale associée à la base orthonormée choisie, et D un vecteur comprenant d'une part les coefficients d'approximation et d'autre part les coefficients de détail de la transformée en ondelettes discrète (*DWT*). Du fait de l'orthogonalité de la matrice W , la transformée en ondelettes discrète inverse (*IDWT*) est donnée par $y = W^T D$. La *DWT* ainsi que sa transformée inverse peuvent être implémentées à l'aide de l'algorithme pyramidal proposé par Mallat [24] employant un banc de filtres miroirs en quadrature. Du fait du caractère creux (propriété de parcimonie, (sparse en anglais) dans le domaine de la transformée en ondelettes, nous pouvons supposer de façon intuitive que seuls quelques coefficients de détail d_{ij} ayant une valeur suffisamment élevée contribuent au signal s à recouvrer, alors que les coefficients de faibles valeurs sont essentiellement, dus au bruit qui contamine de façon uniforme tous les coefficients. Il est également recommandé de conserver les coefficients d'approximation s_{mn} . Ces derniers, relatifs aux composantes de basses fréquences, sont essentiellement caractéristiques du signal original. Les coefficients ainsi mises à jour, peuvent être utilisés pour reconstruire le signal original sans les effets du bruit. Le débruitage du signal y concerne le calcul de l'estimateur \hat{s} du signal d'origine sans bruit s . Les trois étapes suivantes résument le processus complet de débruitage dans le domaine des transformées en ondelettes :

$$\begin{cases} Y = W(y) \\ Z = D(Y, \lambda) \\ \hat{s} = W^{-1}(Z) \end{cases} \quad (5.2)$$

où $W(\cdot)$ la transformée en ondelettes et $W^{-1}(\cdot)$ transformée inverse, $D(\cdot, \lambda)$ désigne l'opérateur de débruitage et λ est le seuil associé. Il est implicite dans la stratégie que seul un petit nombre de grands coefficients ondelettes sont associés avec le signal original, et que leur identification et l'élimination de tous les autres coefficients permettront une reconstruction parfaite du signal sous-jacent. Plusieurs méthodes utilisent cette idée et l'implémente de manière différentes. Dans la méthode linéaire chaque coefficient d'ondelette est affecté par un seuil linéaire particulièrement associée au niveau de résolution du coefficient. Le seuillage linéaire est suffisant seulement pour des signaux homogènes avec un niveau de régularité important. Lorsque les conditions d'homogénéité et de régularité ne sont pas remplies, les méthodes de seuillage non linéaires des coefficients d'ondelettes sont plus appropriées. Donoho & Johnstone ont proposé un estimateur non-linéaire reposant sur la conservation en intégralité des coefficients d'approximation et sur une sélection judicieuse des coefficients de détail $d_{j,k}$ [19, 20, 79]. Ces auteurs suggèrent l'extraction des coefficients de détail significatifs par comparaison de ces derniers avec un paramètre de seuillage $\lambda > 0$ dont le

choix est décrit ultérieurement. Les fonctions de seuillage résultantes, se déclinent sous deux formes, seuillage dur (figure 5.1 a) et seuillage doux (figure 5.1 b). Le débruitage impliquant un seuil par lequel les coefficients en dessous d'une valeur (i.e. le seuil) spécifiée sont mis à zéro est appelé le seuillage dur.

$$\delta_{\lambda}^{hard}(d_{j,k}) = \begin{cases} d_{j,k} & |d_{j,k}| > \lambda \\ 0 & otherwise \end{cases} \quad (5.3)$$

Alternativement, dans le seuillage doux tous les coefficients sont réduits par la valeur du seuil.

$$\delta_{\lambda}^{soft}(d_{j,k}) = \text{sign}(d_{j,k}) \max(0, |d_{j,k}| - \lambda) \quad (5.4)$$

où $d_{j,k}$ sont les coefficients de détail du signal analysé. Le seuillage dur (fonction discontinue) correspond à une loi du type 'kill or keep'. En d'autres termes, soit le coefficient est conservé, soit il est mis à zéro. Le seuillage doux (fonction continue) correspond à un opérateur de contraction de type 'kill or shrink'. En d'autres termes, soit le coefficient se voit retrancher la valeur du seuil, soit il est mis à zéro. En terme de risque d'estimation, le seuillage dur aboutit à une variance plus importante de la fonction estimée que celle obtenue par seuillage doux avec un seuil identique. Du fait de sa discontinuité, il présente en outre une sensibilité importante vis-à-vis des faibles variations des données. La réduction des coefficients de grande valeur, dans le cas du seuillage doux entraîne un biais d'estimation plus élevé.

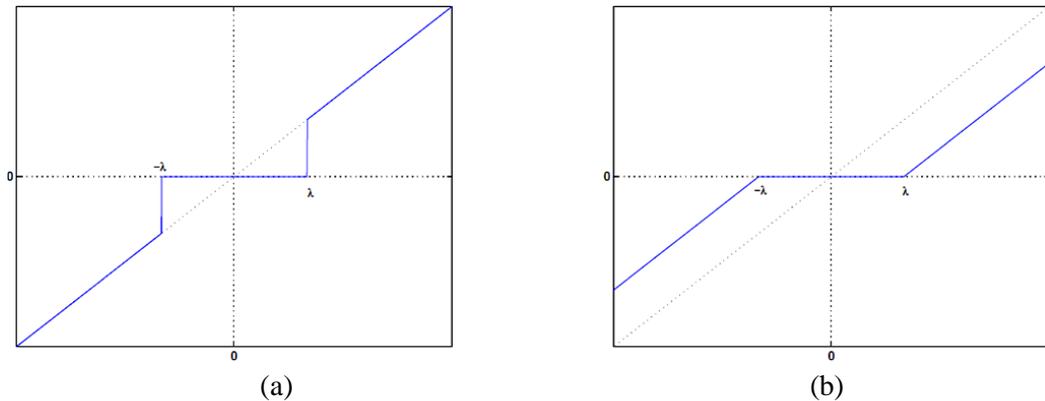


Fig. 5.1 Estimateurs par seuillage d'ondelettes. (a) Seuillage dur (b) Seuillage doux

Les propriétés et les limites de ces deux stratégies ont été rapportées par plusieurs auteurs. D'une part, le seuillage dur maintient l'échelle du signal mais introduit de petits changements après la reconstruction en raison d'une discontinuité dans les coefficients d'ondelettes. D'autre part, le seuillage doux élimine cette discontinuité et les signaux reconstruits se trouvent plus doux et avec des amplitudes légèrement diminuées. Le problème apparent dans l'application des méthodes de seuillage des coefficients d'ondelettes est la manière de sélectionner une valeur appropriée pour le seuil λ .

V.3.1 Choix du seuil

Depuis le travail pionnier de Donoho et Johnstone, beaucoup de recherches dans la définition des niveaux de seuil et de leurs types ont été menées. De manière générale, deux stratégies de seuillage peuvent être identifiées: la stratégie avec des seuils globaux et la stratégie avec des seuils dépendants du niveau de décomposition. Dans la stratégie avec des seuils globaux implique la sélection d'une valeur $\lambda > 0$, appliquée à tous les coefficients d'ondelettes. En revanche, dans la stratégie avec des seuils dépendants du niveau de décomposition implique qu'un seuil $\lambda_j > 0$ est appliqué à chaque niveau de résolution. Les algorithmes de seuillage, les plus connus en littérature sont : Minimax, le seuil universel, SURE et le seuil hybride ou heuristique SURE.

a. Seuil minimax

Donoho & Johnstone ont proposé le seuillage minimax qui applique un seuil optimal (au sens minimax) [20]. La valeur du seuil minimax λ^M telle que définie dans l'équation (5.5) correspond à un seuil optimal obtenu en minimisant le risque associé à la fonction d'estimation. Le seuil proposé dépend de la taille de l'échantillon et prend également, en compte le niveau de bruit dans le signal. En d'autres termes, le seuil minimax est défini de la manière suivante :

$$\lambda^M = \tilde{\sigma}_\varepsilon \lambda_N^* \quad (5.5)$$

où, $\tilde{\sigma}_\varepsilon$ définit la variance du bruit présent dans le signal et λ_n^* est donné par l'équation suivante:

$$\lambda_n^* = \inf_{\lambda} \sup_{d_{ij}} \left\{ \frac{E(\delta_\lambda)^2}{N^{-1} + R_{oracle}(d_{ij})} \right\} \quad (5.6)$$

avec $R_{oracle}(d_{ij})$ est un opérateur appelé oracle utilisée pour comptabiliser le risque associé à la modification de la valeur d'un coefficient d'ondelette donné. Les notions d'oracles et de risque d'oracle sont détaillées dans [20]. Deux oracles sont considérés, la projection diagonale linéaire DLP (Diagonal Linear Projection) faisant intervenir un opérateur diagonal réalisant une projection linéaire et aboutissant à une décision de type 'kill or keep', et le DLS (Diagonal Linear Shrinker), oracle définissant le facteur d'atténuation à appliquer à chacun des coefficients de détail d'ondelettes. Les risques idéales pour ces oracles sont données par:

$$R_{oracle}^{DLP}(d) = \min(d^2, 1) \quad \text{and} \quad R_{oracle}^{DLS}(d) = \frac{d^2}{d^2 + 1} \quad (5.7)$$

b. Seuil universel

Une alternative à l'utilisation du seuil minimax a été proposée par Donoho & Johnstone [20]. Elle repose sur l'utilisation d'une valeur de seuil universel :

$$\lambda_{univ} = \tilde{\sigma}_\varepsilon \sqrt{2 \log(N)} \quad (5.8)$$

où N est la taille du signal en nombre d'échantillons. Le seuillage universel appelé aussi VisuShrink [19, 20], est substantiellement plus important que celui obtenu au sens minimax, mais s'avère aisément implémentable. Il est de loin le plus répandu dans la communauté de traitement du signal et des images. La limitation du seuil VisuShrink est de ne pas considérer la minimisation de l'erreur quadratique moyenne (MSE), ce ceci conduit à la suppression excessive des coefficients d'ondelettes dont certains sont porteur d'informations utiles. Par conséquent le signal récupéré se trouve trop adoucie. Ainsi, un nombre plus limité de coefficients est employé lors de la reconstruction ce qui a pour effet de lisser le signal en sortie par rapport au cas minimax d'où un biais d'estimation plus élevé.

c. SureShrink

Dans [19], Donoho & Johnstone ont proposé l'estimateur SureShrink où un seuil est défini pour les coefficients d'ondelettes de chaque échelle de décomposition. L'idée est basée sur la minimisation de l'estimation sans biais du risque de Stein 'SURE' (Stein's Unbiased Risk Estimator) [82]. Le seuil λ_j est défini en minimisant SURE sur l'ensemble des niveaux. Un seuil est attribué à chaque niveau de résolution dyadique. Une expression de λ_j est donnée par l'expression suivante :

$$\lambda_j^{\text{SURE}} = \underset{0 \leq \lambda \leq \sqrt{2 \log(N)}}{\operatorname{argmin}} \operatorname{SURE} \left(\lambda, \frac{d_{j,k}}{\tilde{\sigma}_\varepsilon} \right) \quad (5.9)$$

Le SureShrink est un seuil adaptatif relatif au niveau de décomposition en ondelettes et il minimise l'erreur quadratique moyenne. L'inconvénient de cet estimateur réside dans les situations extrêmement parcimonieuses des coefficients d'ondelettes. Dans ce cas, l'adoption du seuil universel $\lambda = \lambda_{univ} = \sigma \sqrt{2 \log(N)}$ pour tous les niveaux et l'appliquée à tous les coefficients d'ondelettes donne de bons résultats avec pratiquement pas de complexité de calcul. Alors, si la plupart des coefficients d'ondelettes sont des zéros, le seuil universel est préféré. Cependant, si le rapport signal sur bruit est très faible l'estimation SURE est utilisée. Pour éviter ce problème, Donoho & Johnstone ont proposé une autre alternative au seuillage SureShrink, dite le mode hybride du seuillage SURE. Afin de détecter cette situation, Donoho et Johnstone ont proposé une mesure de la parcimonie de la représentation en ondelettes afin de détermine à chaque niveau de résolution le seuil qui doit être appliqué *i.e.* le seuil universel ou le seuil SURE. En termes mathématiques, le seuillage du mode hybride s'écrit sous la forme suivante :

$$\lambda_j^{\text{HSURE}} = \begin{cases} \tilde{\sigma}_\varepsilon \sqrt{2 \log(2^j)} & \text{si } \sum_{m,n}^{2^j-1} (d_{m,n}^j)^2 \leq \tilde{\sigma}_\varepsilon^2 2^j (2^j + j^{3/2}) \\ \lambda_j^{\text{SURE}} & \text{sinon} \end{cases} \quad (5.10)$$

V.3.2 Estimation du niveau de bruit

Dans la littérature, la majorité des méthodes de débruitage n'aborde que le cas de bruit blanc gaussien, plus simple à traiter, bien que, en situation de données réelles, il ne soit pas

spécialement facile d'estimer le niveau de bruit σ_ε . Avec une transformée discrète orthogonale en ondelettes, le bruit blanc se décompose en série de coefficients aléatoires normaux centrés et décorrélés. En utilisant des arguments de la statistique robuste, Donoho & Johnstone ont proposé une estimation de σ_ε dans le domaine des ondelettes en ne considérant que les coefficients de détail de l'échelle de décomposition la plus fine. Le choix de l'échelle la plus fine repose sur l'hypothèse que les coefficients en ondelettes correspondants sont en grande majorité dus au bruit blanc. Une estimée de $\tilde{\sigma}_\varepsilon$ est alors obtenue par un résultat classique en statistique robuste :

$$\tilde{\sigma}_\varepsilon = \frac{MAD|d_{L-1,k}|}{0.6745} \quad k = 0,1, \dots, 2^{L-1} - 1 \quad (5.11)$$

où L désigne le nombre de niveaux de décomposition. MAD est la valeur médiane absolue des coefficients de l'échelle la plus fine $d_{L-1,j}$. Le facteur 0.6745 est choisi après une calibration avec une distribution gaussienne. Cet estimateur très robuste est également très populaire pour le débruitage multi-échelle.

V.4 Formulation Bayésienne du problème d'identification

L'approche bayésienne est proposée dans cette section, dans le cadre de la restauration du signal source vibratoire. L'inférence bayésienne, offre un cadre mathématique rigoureux permettant de combiner des informations probabilistes afin de renforcer la parcimonie des coefficients. Les estimateurs bayésiens s'avèrent plus efficaces que ceux présentés précédemment [77,83, 84, 85]. Dans le cadre d'une approche bayésienne, on suppose que la distribution des coefficients d'ondelettes est imposée comme *a priori*. La distribution de l'*a priori* dans l'approche bayésienne exprime la connaissance préalable sur les coefficients, sous la forme d'une distribution de probabilité. Mallat dans [40] a été le premier à proposer les distributions gaussiennes généralisées (GGD). Elles sont très communément adoptées comme *a priori* à des fins d'estimation et de compression dans le domaine des ondelettes. Cependant, à cause de sa décroissance exponentielle rapide, l'*a priori* GGD se révèle incapable de modéliser correctement les queues de distribution relevées, comportement typique des distributions des coefficients d'ondelettes. De plus, aussi bien pour l'*a priori* GGD que bon nombre d'*a priori* dans la littérature, aucune forme analytique n'a été proposée pour l'estimateur bayésien. Ce qui nécessite une implémentation numérique des intégrales, impliquées dans les estimateurs, particulièrement instable (notamment à cause des bornes d'intégration infinies) et coûteuse en temps. Formellement, la solution du problème inverse bayésien correspond à l'estimation du *maximum a posteriori*. Tout d'abord, nous supposons que seule la variable aléatoire y est observée selon le modèle de l'équation (5.1). Comme mentionné précédemment, la parcimonie est une propriété qui peut être décrite soit directement pour le signal lui-même ou après une transformation, par exemple sur les coefficients de la projection du signal sur un ensemble de fonctions. En conséquence, dans le domaine des transformées en ondelettes, notée $W(\cdot)$, le modèle de l'équation (5.1) peut être décrite par une observation analogue comme suit:

$$\omega = W(y) = W(s) + W(n) = \theta + b \quad (5.12)$$

Où ω sont les coefficients bruités et $b = W(n) \sim \mathcal{N}(0, \sigma^2 I)$; le bruit est blanc et gaussien. L'inférence Bayésienne concerne l'évaluation de la densité de probabilité *a posteriori* des sources θ à partir des données observées ω . Elle nécessite un modèle probabiliste paramétrique des données sous la forme d'une distribution de probabilité conditionnée par la connaissance de la distribution des sources $p(\omega/\theta)$. L'état antérieur de la connaissance de θ est quantifiée par la distribution *a priori* $p(\theta)$. Notre état des connaissances sur θ après avoir observé ω est quantifiée par la distribution *a posteriori* $p(\theta/\omega)$. Ces fonctions sont liées par la règle de Bayes:

$$p(\theta/\omega) = \frac{p(\theta, \omega)}{p(\omega)} = \frac{p(\omega/\theta)p(\theta)}{p(\omega)} \quad (5.13)$$

Nous référons à $p(\theta, \omega)$ la distribution conjointe des sources et des données, ou, de manière plus concise, la distribution conjointe et $p(\theta/\omega)$ la distribution de θ sachant les valeurs observées.

- $p(\theta/\omega)$ est la distribution de probabilité *a posteriori*, définissant la probabilité d'observer le champ excitateur θ étant donnée un champ vibratoire mesuré ω ;
- $p(\omega/\theta)$ est la fonction de vraisemblance, définissant la probabilité d'observer le champ vibratoire mesuré ω étant donné un champ excitateur θ ;
- $p(\theta)$ est la distribution de probabilité *a priori*, définissant la probabilité d'observer le champ excitateur θ vibratoire à identifier.

De point de vue Bayésien, la solution du problème de l'équation (5.12), consiste à rechercher la source vibratoire excitatrice $\tilde{\theta}$ la plus probable étant donnée le champ vibratoire mesuré ω , cela revient à trouver un estimateur $\tilde{\theta}$ qui maximise la probabilité *a posteriori* (MAP). Cette estimation nécessite l'adoption d'un modèle probabiliste paramétrique des données sous la forme d'une distribution de probabilité, conditionnés par la connaissance des paramètres inconnus. Mathématiquement on cherche donc l'estimateur du maximum *a posteriori* tel que

$$\tilde{\theta}(\omega) = \arg \max_{\theta} \{p(\theta/\omega)\} \quad (5.14)$$

'argmax' est la valeur de l'argument où la fonction est à son maximum. Pratiquement, maximiser $p(\theta/\omega)$ étant équivalent à maximiser $\log p(\theta/\omega)$ d'où, on peut écrire:

$$\log p(\theta/\omega) = \log p(\omega/\theta) + \log p(\theta) - \log p(\omega) \quad (5.15)$$

V.4.1 Choix de la fonction de vraisemblance

La fonction de vraisemblance traduit l'incertitude associée à la mesure du champ vibratoire ω . Par définition, cette incertitude est principalement liée au bruit de mesure b supposé centré

Gaussien i.e. $\varepsilon \sim \mathcal{N}(0, \sigma^2 I)$ et de variance σ^2 . Par conséquent la fonction de vraisemblance $p(\omega/\theta)$ peut se réécrire sous la forme :

$$p(\omega/\theta) = p(b) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\|\omega - \theta\|^2}{2\sigma^2}\right) \quad (5.16)$$

où $\|x\| = (\mathbf{x}^T \mathbf{x})$ est la norme ℓ_2 du vecteur x .

$$\log p(\theta/\omega) = -\frac{1}{2\sigma^2} \|\omega - \theta\|^2 + \log p(\theta) - \log p(\omega) \quad (5.17)$$

Soit : $f(\theta) = -\log p(\theta)$, alors :

$$J(\theta) = \frac{1}{2\sigma^2} \|\omega - \theta\|^2 + f(\theta) \quad (5.18)$$

Enfin, le problème de maximisation est reformulé en problème de minimisation du critère suivant :

$$\tilde{\theta} = \arg \min_{\theta} \left\{ \frac{1}{2\sigma^2} \|\omega - \theta\|^2 + f(\theta) \right\} \quad (5.19)$$

D'après l'équation (5.16), on remarque que l'on retrouve la méthode de régularisation de Tikhonov où $f(\theta)$ représente le terme de régularisation ou de pénalisation favorisant la parcimonie. L'estimateur MAP de θ est donnée par la relation $g(\omega)$ qui est une fonction de la variance de bruit et de la dérivée de $f(\theta)$ qui est à son tour fonction de la distribution de θ :

$$\tilde{\theta} = g(\omega) = \text{sign}(\omega) \max\left(0, |\omega| - \sigma^2 f'(\theta)\right) \quad (5.20)$$

V.4.2 Choix de la distribution de probabilité *a priori* de parcimonie

La distribution de probabilité *a priori* traduit l'incertitude associée au champ vibratoire non observable relatif aux forces excitatrices ω . Il s'agit d'un modèle d'expertise, qui peut être vu comme une mesure de la connaissance *a priori* que l'on possède sur le champ excitateur ω . La modélisation de l'apriori a le rôle de la traduction de notre connaissance antérieure sur le signal inconnu dans une loi de probabilité. Mallat a été le premier à proposer la distribution gaussienne généralisée (\mathcal{GG}) à des fins de compression. Ensuite, plusieurs estimateurs ont proposé un estimateur bayésien de type MAP basé sur cet *a priori* à des fins d'estimation [86,87]. On rappelle qu'une distribution (\mathcal{GG}) d'échelle γ et de forme β est donnée par l'équation (5.21):

$$p(\theta) = \mathcal{GG}_{\gamma, \beta}(\theta) = \frac{\beta\gamma}{2\Gamma(1/\beta)} \exp\{-(\gamma\theta)^\beta\} \quad (5.21)$$

où β représente le paramètre de forme et γ se rapporte à la variance de la distribution, $\Gamma(\cdot)$ est la fonction Gamma. Par un choix approprié de β il est possible de modéliser les distributions à queue plus ou moins lourds, qui favorisent des solutions très parcimonieuses ou moins parcimonieuses. Deux cas spéciaux peuvent être précisés pour ce modèle : cas d'une

distribution gaussienne (Normal) lorsque $\beta = 2$ et cas de distribution double exponentielle ou distribution de Laplace lorsque $\beta = 1$. Les formes générales de ces distribution d'a priori sont présentés dans la figure (5.2), où les cas: $\beta = 0.5, 1, 1.5$ et $\beta = 4$ sont comparés au cas Gaussien $\beta = 2$. La PDF de la \mathcal{GG} a une décroissance exponentielle rapide vers l'infini pour des valeurs décroissantes de β . Cette famille permet alors, des queues plus étroites que la normale lorsque $\beta < 2$ et moins étroites que la normale lorsque $\beta > 2$. Du fait de ce comportement, l'a priori se révèle incapable de modéliser les queues lourdes de la distribution des coefficients d'ondelettes. Toutefois, le modèle \mathcal{GG} avec un paramètre de forme $0.3 < \beta < 1$ présente un modèle raisonnable pour la distribution des coefficients d'ondelettes [40]. Il peut être appliqué à chaque échelle de la décomposition sauf à l'échelle la plus grossière. Dans le cadre de cette thèse, le modèle de lois leptokurtique à queues lourdes défini par l'équation (5.22) est introduit pour modéliser l'estimateur MAP avec a priori de parcimonie [40]. Ce modèle est utilisé pour décrire la densité des coefficients d'ondelettes des signaux impulsionnels de distribution non-gaussienne et qui sont considérées de densité plus parcimonieuse que la distribution Laplacienne [88]:

$$p(\theta) = \frac{1}{2d} \frac{(\alpha + 2)[\alpha(\alpha + 1)/2]^{(\alpha/2+1)}}{[\sqrt{\alpha(\alpha + 1)/2} + |\theta/d|]^{(\alpha+3)}} \quad (5.22)$$

où d est l'écart-type de l'impulsion à isoler et α est le paramètre de contrôle de la parcimonie de la fonction de densité. Quand $\alpha \rightarrow \infty$, la densité de Laplace est obtenue comme limite. Les paramètres $d, \alpha > 0$ peuvent être obtenus à partir des relations $d = \sqrt{E\{\theta^2\}}$ et $\alpha = \frac{2-k+\sqrt{k(k+4)}}{2k-1}$ avec $k = d^2 p_\theta(0)^2$. La figure (5.3) montre la fonction $p(\theta)$ donnée dans l'équation (5.22) à des fins de comparaison. Pour raison d'illustration, les densités dans les figures sont normalisées à la variance unité.

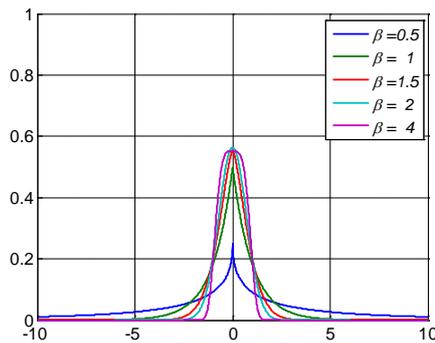


Fig. 5.2 Distributions gaussiennes généralisées

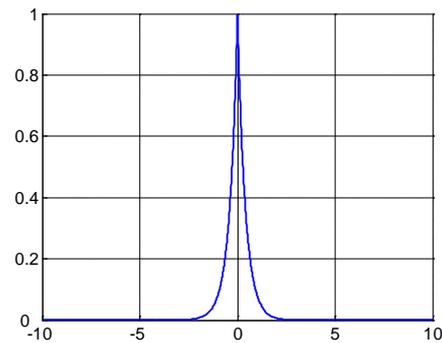


Fig. 5.3 Modèle de densité parcimonieuse ($\alpha = 0.1$)

En considérant la densité représentée par l'équation (5.22) de l'a priori choisis, la fonction non linéaire d'atténuation résultante g prend la forme suivante [89]:

$$g(\omega) = \text{sign}(\omega) \max\left(0, \frac{|\omega| - ad}{2} + \frac{1}{2} \sqrt{(|\omega| + ad)^2 - 4\sigma^2(\alpha + 3)}\right) \quad (5.23)$$

où $a = \sqrt{\alpha(\alpha + 1)/2}$, et $g(\omega)$ est mis à zéro dans le cas où la racine carrée dans l'équation (5.23) est imaginaire. Les résultats de reconstruction après seuillage des coefficients d'ondelettes en utilisant la règle de l'équation (5.23) représentent l'approximation des impulsions. L'effet de la fonction d'atténuation dans l'équation (5.23) est de réduire la valeur absolue de son argument par une certaine quantité paramétrée. Les petits arguments sont ainsi mis à zéro. Cet estimateur non linéaire est à la base de la présente méthode bayésienne.

V.5 Etude expérimentale

V.5.1 Analyse des signaux de synthèse

Dans la section suivante, les résultats obtenus par l'application de l'approche proposée sur un signal test corrompu avec différents niveaux de bruit blanc gaussien sont présentés. Une étude comparative des techniques de débruitage dans le domaine des ondelettes est faite et les résultats obtenus sont analysés. Pour réaliser cette comparaison, un train d'impulsions simulées de 50000 échantillons est utilisé pour étudier la performance de l'approche proposée. Les impulsions simulées considérées sont très proches de ceux générés par les composants mécaniques endommagées et présentent la forme indiquée dans la figure (5.4a). Le bruit additif est supposé être un bruit blanc gaussien. Plusieurs valeurs de RSB (ie rapport signal sur bruit entre signal original et le signal bruité) sont introduites.

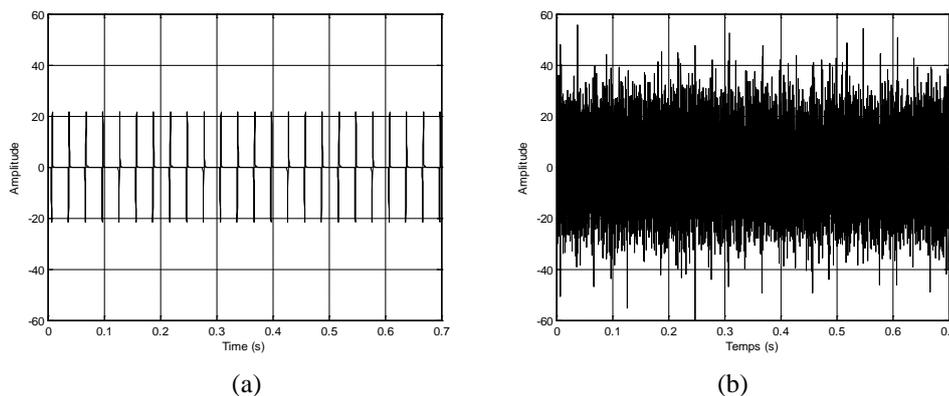


Fig. 5.4 (a): Impulsions simulées (b) Impulsions simulées avec un bruit gaussien lourd

Le signal indiqué sur la figure (5.4b) qui est artificiellement corrompu par un bruit gaussien est utilisé pour tester l'efficacité de la méthode proposée. Le but est d'extraire, à partir des signaux de vibration, les impulsions périodiques faibles noyées dans du bruit de fond lourd (faible niveau de RSB). L'ensemble des techniques basées sur le seuillage des coefficients d'ondelettes en utilisant l'approche traditionnelle telles que, le seuillage SURE, hybride ou heuristique SURE, le seuillage minimax et le seuillage universel sont considérés. Les deux types de seuillage dur et doux sont utilisés dans ces techniques sur les coefficients de la décomposition en ondelette de Daubechies (db20). Le résultat de débruitage du signal au moyen d'estimation du risque non biaisé de Stein (SURE), heuristique SURE, seuils minimax et le seuil universel, respectivement, est présenté dans les figures (5.5 - 5.8).

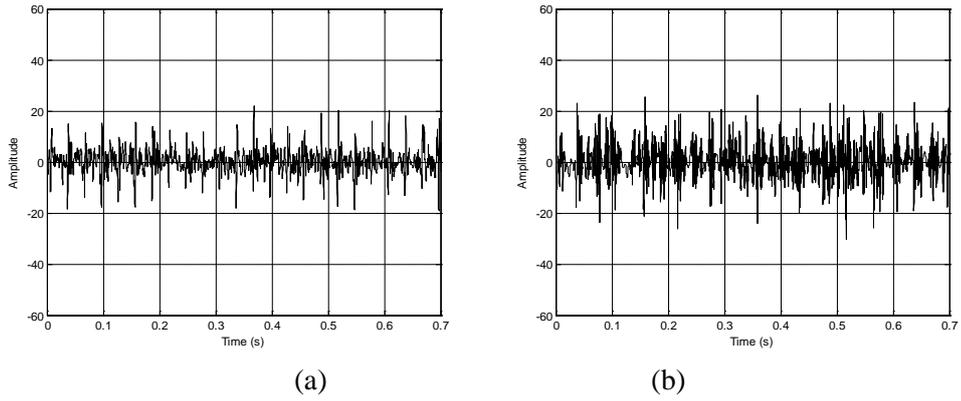


Fig. 5.5 Impulsions débruitées en utilisant (a): Seuillage SURE doux (b): Seuillage SURE dur

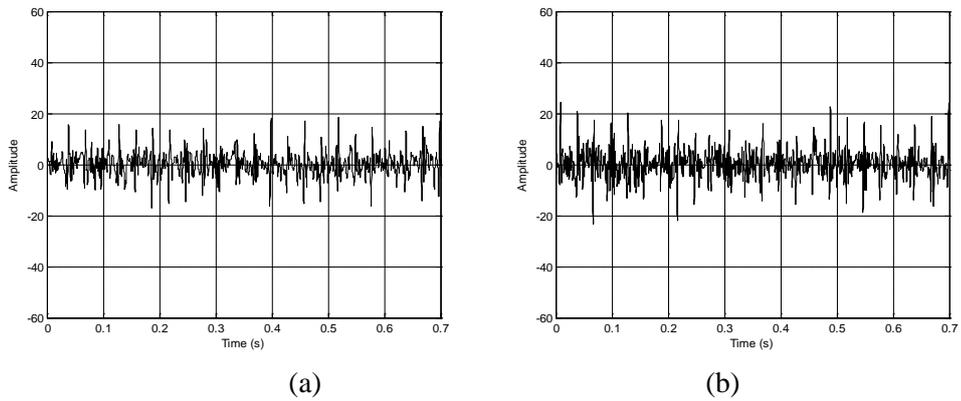


Fig. 5.6 Impulsions débruitées en utilisant (a): Seuillage Heuristique SURE doux (b): Seuillage Heuristique SURE dur

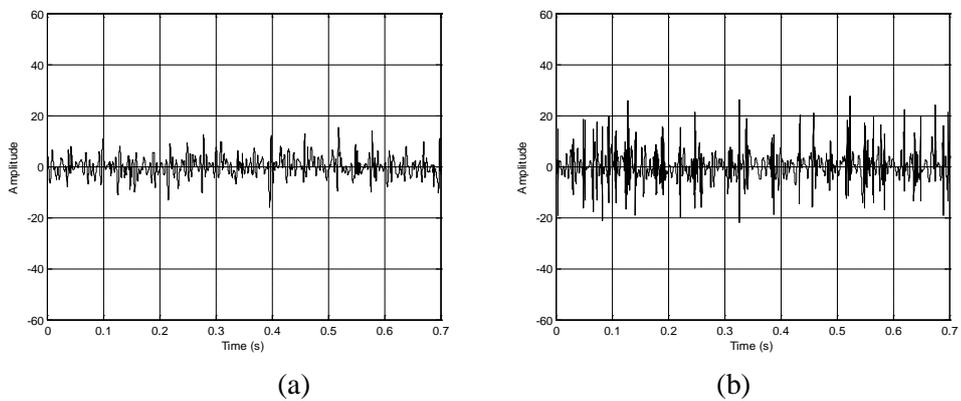


Fig. 5.7 Impulsions débruitées en utilisant (a): Seuillage minimax doux (b): Seuillage minimax dur

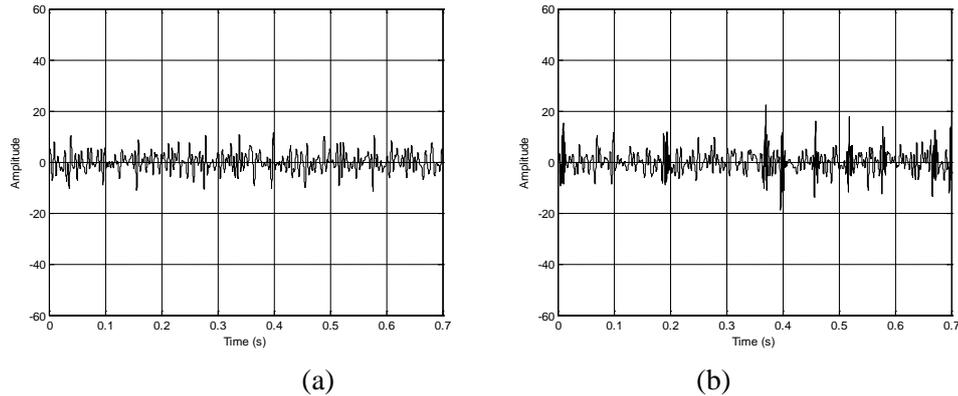


Fig. 5.8 Impulsions débruitées en utilisant (a): Seuillage universel doux (b): Seuillage universel dur

Il peut être observé à partir des résultats de débruitage, qu'aucune stratégie n'a bien montré les impulsions. Bien que plusieurs impulsions sont extraites à l'aide du seuillage SURE et heuristique SURE illustrées dans les figures (5.5) et (5.6), beaucoup de fausses impulsions existent également, ce qui affecterait la reconnaissance des vrais impulsions. Par conséquent, on peut conclure que les deux méthodes basées sur le seuillage SURE et heuristique SURE donnent une performance optimale et les résultats de ces deux méthodes sont en général, très proches les uns des autres. La performance du seuillage SURE et heuristique SURE est meilleure que celle par seuillage Minimax et ou par l'utilisation du seuil universel. Cependant, dans tous les cas, l'isolation des impulsions reste difficile. En revanche, les résultats du signal purifié basé sur l'approche bayésienne sont présentés dans la figure (5.9). On peut observer que toutes les impulsions immergées dans le bruit sont extraites. En d'autres termes, l'approche proposée en utilisant des calculs bayésiens présenté dans la figure (5.9), isole tous les vraies impulsions et aucune fausse impulsion existent dans les résultats.

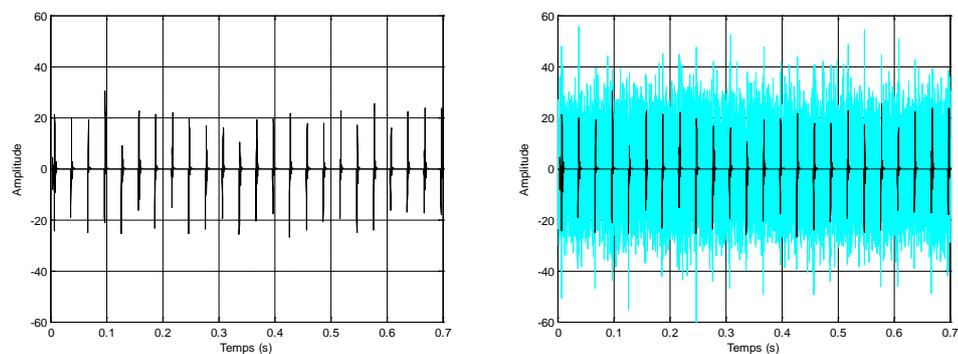


Fig. 5.9 Impulsions débruitées avec l'approche de Bayes et en utilisant l'ondelette de Morlet optimisée

Comme on peut le voir sur la figure (5.9), on peut conclure que, l'approche proposée est très efficace dans l'extraction de faibles impulsions périodiques noyées dans du bruit lourd, et ses performances sont nettement mieux que les méthodes précédentes.

V.5.2 Analyse des signaux réels

Des expériences sont menées pour explorer l'efficacité de l'algorithme de débruitage par l'approche proposé de Bayes et de le comparer avec les techniques de débruitage standard.

Les signaux de vibration collectés à partir d'un banc d'essais ont été analysés pour la détection des défauts de roulements. Les images du banc d'essais utilisé à cette fin et de ses accessoires sont présentées dans la figure (5.10). Généralement, le défaut local se produit sur les pistes des bagues interne ou externe du roulement. Lorsque les billes passent par le défaut, des impulsions apparaîtront due à cet impact.



Fig. 5.10 (a) Photographie du banc d'essais, (b) Banc d'essais et Analyseur double canal (B&K 2133)

Selon la durée de l'intervalle temporelle entre ces impulsions, on peut juger de l'apparition du défaut à l'aide des formules donnant les fréquences caractéristiques. Pour cela, un défaut artificiel (une raie d'environ 0,9 mm de large) est introduit sur la piste de la bague externe du roulement. L'acquisition du signal d'accélération radiale du palier testé est réalisée par un accéléromètre 4371 B&K. Ensuite, le signal est amplifié et filtré dans la gamme de fréquence de 0,2 Hz à 20 kHz par un amplificateur de charge B&K. Un analyseur de spectre double canal de type B&K 2133 est utilisé ensuite pour enregistrer le signal d'acquisition. Les acquisitions sont transférées ensuite vers le PC où des programmes Matlab sont mis en œuvre pour le calcul des coefficients d'ondelettes liés à ce signal. Les fréquences caractéristiques relatives aux défaillances des bagues internes et externes des roulements dans le cas de l'arbre tournant, sont données par les formules suivantes [63]:

$$\text{Défaut de bague interne:} \quad f_i = (1/2)Nf_r \left(1 + \frac{d}{D_p} \cos\phi \right) \quad (5.24)$$

$$\text{Défaut de bague externe:} \quad f_e = (1/2)Nf_r \left(1 - \frac{d}{D_p} \cos\phi \right) \quad (5.25)$$

où f_r est la fréquence de rotation de l'arbre, N est le nombre d'éléments roulants (billes), ϕ représente l'angle de contact, d et D_p sont respectivement, les diamètres interne et primitif du roulement. Les paramètres géométriques et les fréquences caractéristiques correspondantes du palier testé sont listés dans le Tableau 1.

Type de roulement	SNR 1205
Vitesse de rotation (tr/min)	2986
Nombre d'éléments roulants	$N = 12$
Fréquence de rotation	$f_r = 49.77 \text{ Hz}$
Ball Passing Frequency Inner Race - <i>BPFI</i>	$f_i = 353.99 \text{ Hz}, T_i = 0.0028 \text{ sec}$
Ball passing Frequency Outer Race - <i>BPFO</i>	$f_e = 243.16 \text{ Hz}, T_e = 0.0041 \text{ sec}$

Table1. Caractéristiques du palier testé du banc d'essais

Le diagramme temporel et le spectre correspondant du signal d'acquisition sont présentés respectivement, par les figures (5.11a) and (5.11b).

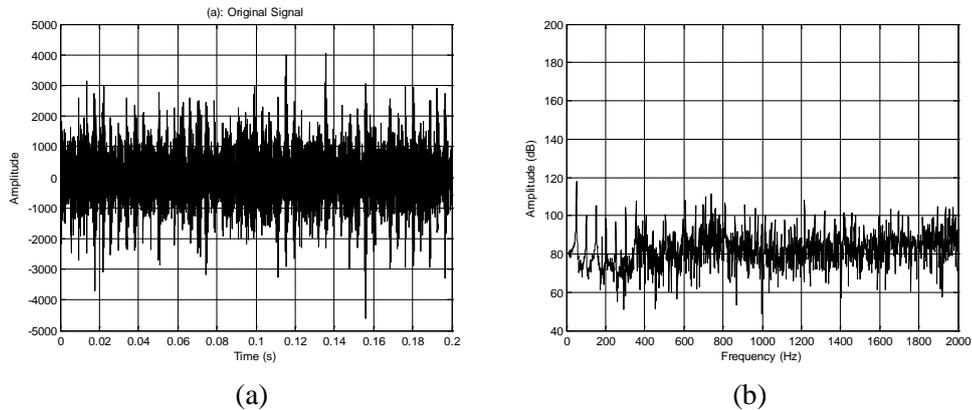


Fig. 5.11 (a): signal temporel avec défaut de roulement (b): Spectre associé du signal d'acquisition,

De la figure (5.11b), on peut observer que l'efficacité de l'analyse spectrale est insuffisante pour effectuer un suivi correct. En effet, les défauts des paliers ne peuvent être diagnostiqués avec certitude car les pics du spectre situés au niveau des fréquences caractéristiques ont des amplitudes comparables à celles relatives à un palier en bon état. En outre, il est à remarquer que le niveau du bruit donne des amplitudes équivalentes à celles des impulsions périodiques. Une analyse comparative des méthodes de débruitage standards utilisées précédemment sont reconduites pour analyser le signal réel présenté dans la figure (5.11a). Les résultats du débruitage du signal par les méthodes (SURE), heuristique SURE, Minimax et par seuillage universel sont présenté par les figures (5.12 - 5.15), respectivement. Les coefficients d'ondelettes sont calculés par la décomposition du signal en utilisant l'ondelette de Daubechies (db20). Cette ondelette est utilisée, car elle offre de meilleures performances comparée à d'autres fonctions ondelettes. Toutefois, les résultats expérimentaux doivent être interprétés comme une indication de la performance des techniques de débruitage proposées.

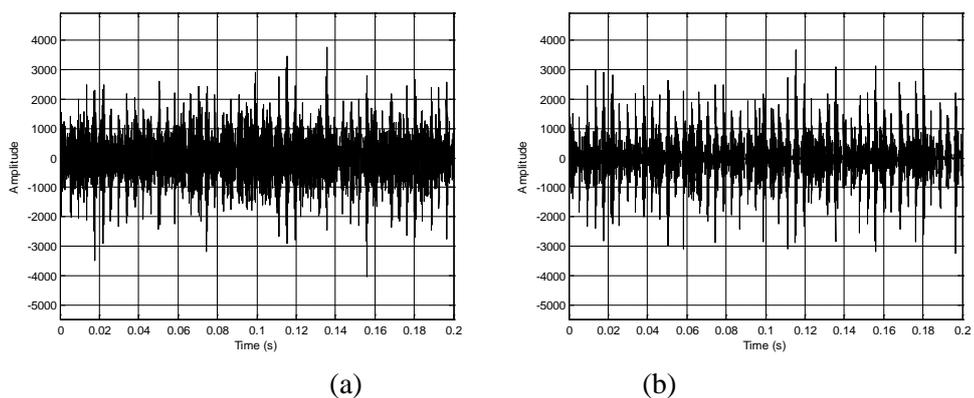


Fig. 5.12 Signal purifié en utilisant (a) Seuillage SURE doux (b) Seuillage SURE dur

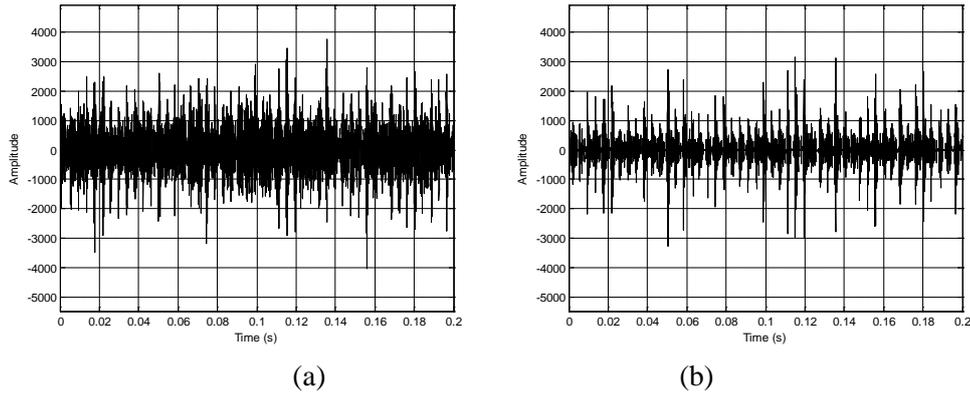


Fig. 5.13 Signal purifié en utilisant (a) Seuillage heuristique SURE doux (b) Seuillage heuristique SURE dur

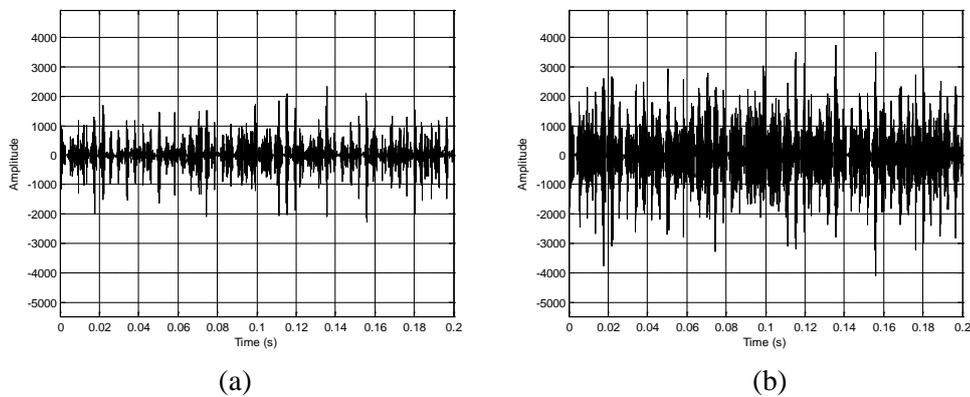


Fig. 5.14 Signal purifié en utilisant (a) Seuillage minimax doux (b) Seuillage minimax dur

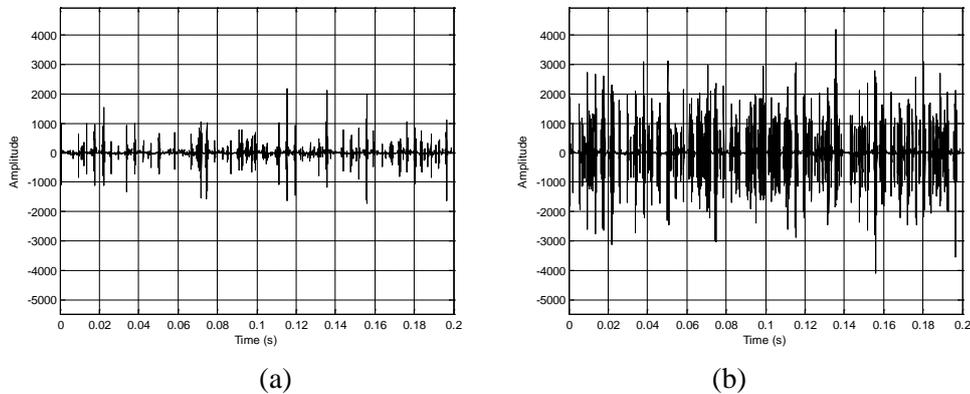


Fig. 5.15 Signal purifié en utilisant (a) Seuillage universel doux (b) Seuillage universel dur

L'analyse des figures (5.12 - 5.15) indique que les impulsions périodiques sont à peine visibles. La réponse du système à cette excitation apparaît comme un train d'impulsions irrégulières et beaucoup de fausses impulsions existent. Ce genre de réponse affecterait certainement la reconnaissance des véritables impulsions. De plus, les intervalles du temps entre les impulsions ne sont pas identiques au cours de l'évolution de défaut. Par conséquent, il est impossible d'isoler le phénomène par l'extraction des caractéristiques du défaut. Cependant, le signal reconstruit par la transformée en ondelettes inverse des coefficients seuillés en utilisant l'approche bayésienne et la fonction ondelette optimale de Morlet est montré par la figure (5.16). On peut remarquer que, lorsque les éléments roulants traversent le

défaut, le diagramme montre un train d'impulsions dont l'amplitude est proportionnelle à la dimension de la zone endommagée et fournit une représentation efficace du défaut.

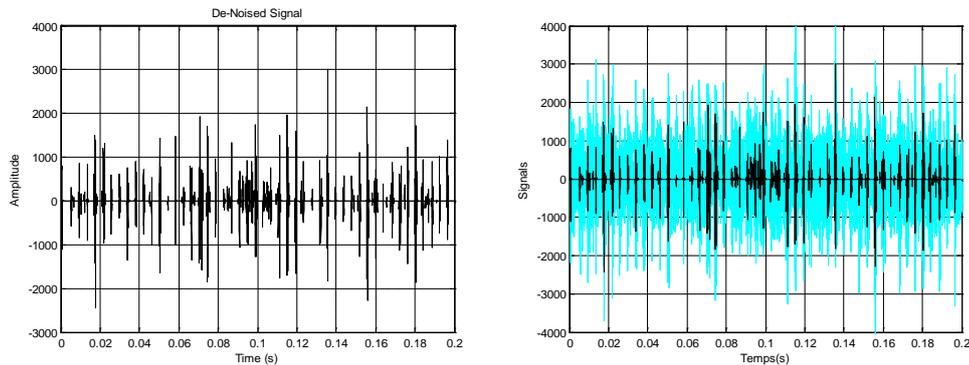


Fig. 5.16 Signal débruité avec l'approche de Bayes et en utilisant l'ondelette de Morlet optimisée

En comparant les résultats présentés par les figures (5.12 - 5.15) et la figure (5.16), on peut affirmer que, l'approche de Bayes donne des résultats bien supérieurs aux méthodes de débruitage classique en utilisant les seuils SURE, minimax et le seuil universels. D'autres observations à partir de la figure (5.16), montrent que la distance mesurée entre deux pics d'impulsions successives des diagrammes présentés représente la période caractéristique du défaut, soit l'inverse de la fréquence caractéristique. Dans le cas de défaut de la bague externe des intervalles quasi-périodiques égaux à 4,1 ms peuvent être trouvés dans la figure (5.16). Ces intervalles quasi-périodiques sont équivalents à l'inverse de la fréquence (BPFO) qui est 243.16 Hz comme indiqué dans le Tableau 1. Par conséquent, il peut être conclu que les impulsions sont causées par le défaut sur la bague externe. Enfin, il est intéressant d'observer que seules les impulsions induites par le défaut sont retenus dans le signal reconstruit (figure (5.16)). Ce fait, démontre l'efficacité de l'algorithme proposé pour l'annulation du bruit, même pour le cas de petits défauts.

V.6 Conclusion

Dans ce chapitre, nous avons analysé l'estimation non-paramétrique de signaux de vibration par le biais d'une transformée parcimonieuse : la transformation adaptée en ondelettes. Dans un premier temps, nous avons étudié le problème du choix du seuil pour distinguer les coefficients de grandes amplitudes, ceux attribués principalement au signal, des petits coefficients ne contenant pas suffisamment d'information sur le signal et pour lesquels le bruit est prédominant. Après un bref rappel théorique sur le principe de cette estimation. Une analyse expérimentale des performances de l'estimation est alors effectuée. En effet, l'analyse de la parcimonie de la transformée en ondelettes sur des signaux synthétiques et des signaux réels révèle que la parcimonie découlant de la représentation en ondelettes de signaux vibratoires impulsionnels est très fortement affectée par le nombre de singularités de tels signaux. De ce fait, le seuillage n'est pas très adéquat pour l'estimation de ce type de signaux présentant un nombre élevé de singularités. Le fait que ces derniers signaux admettent une représentation en ondelettes parcimonieuse au sens large ne justifie pas l'adéquation du seuillage pour l'estimation de ces signaux impulsionnels. Ainsi, nous nous sommes focalisés,

sur l'estimation bayésienne dans le domaine des transformées d'ondelettes parcimonieuses comme solution au problème. Nous nous sommes penché sur la modélisation statistique des signaux afin de définir les propriétés du signal qu'on cherche à recouvrer et pour traduire les connaissances *a priori* qu'on possède sur ce signal. Nous avons donc évoqué quelques propriétés et lois statistiques permettant de caractériser les signaux dans l'espace des transformées. Nous avons aussi exposé les principaux modèles statistiques adaptés à capturer le caractère parcimonieux des coefficients de représentation des signaux. Plus précisément, la densité de probabilité choisie pour *l'a priori* afin de modéliser la notion de rareté a été analytiquement modélisée par le biais d'une distribution leptokurtique. L'approche est basée sur l'optimisation de l'estimateur du maximum *a posteriori* (MAP) de variables aléatoires non gaussiennes corrompus avec du bruit gaussien additif. Ensuite, les performances de l'approche Bayésienne pour le débruitage des vibrations ont été étudiées. Cette analyse des performances est réalisée en comparaison avec les performances des seuils classiques. Il est démontré que l'approche par estimation Bayésienne va au-delà des limites des approches classiques de débruitage. En raison de la nature fortement non-linéaire de l'opérateur d'atténuation, qui tire parti de la structure statistique inhérente de signal devant être identifié, le procédé proposé a été appliquée avec succès pour extraire les caractéristiques impulsifs enfouis dans les signaux bruités. Dans l'ensemble, l'approche décrite combine les propriétés de calcul rapide, de bonnes propriétés théoriques et de bonnes performances dans les simulations et dans la pratique. Le constat majeur de l'ensemble des expériences est que la présente méthode a de meilleurs résultats de débruitage des signaux impulsions comparés à la méthode de régression non paramétrique. Les faibles impulsions cachées dans le bruit sont révélées et renforcées avec succès. Ce résultat permet aux opérateurs de machines non seulement à détecter l'existence de défauts sur des roulements à sa phase initiale, mais aussi à identifier les causes de défauts en utilisant les informations des intervalles de temps qui sont fournis par le signal reconstruit.

Chapitre VI

Séparation aveugle des sources vibratoires dans le cas des mélanges bruités

Chapitre 6

Séparation aveugle des sources vibratoires dans le cas des mélanges bruités

Résumé

La séparation aveugle des sources vibratoires vise à rechercher des signaux sources inconnus et mélangés de manière inconnue au sein de plusieurs observations. Cette approche très générique ne fournit cependant pas nécessairement des résultats exploitables. Il est alors nécessaire d'ajouter des contraintes afin de privilégier la recherche de sources ayant une structure particulière. L'ajout de davantage d'information reste cependant souvent nécessaire afin de pouvoir séparer les sources. Dans ce chapitre, nous nous intéressons à l'intégration du concept de parcimonie qui permet d'améliorer le contraste entre celles-ci en particulier en présence du bruit. Nous montrons aussi, que l'utilisation de la parcimonie dans un espace transformé potentiellement redondant, permet de capturer la structure des signaux naturels de vibrations. Plusieurs ingrédients seront empruntés aux chapitres précédents notamment ceux traitant de la parcimonie des sources et du débruitage. Les sources sont supposées périodiques et corrompues par un bruit lourd (rapport signal sur bruit RSB faible).

Sommaire

VI.1	Introduction	103
VI.2	Modélisation mathématique des mélanges.	104
VI.3	Modèle de mélanges linéaires instantanés	105
VI.4	Principe de résolution de la SAS	107
VI.5	Analyse en Composantes Indépendantes (ACI)	109
VI.6	Critères d'indépendance utilisés en ACI	113
VI.7	Critères de performances de la SAS.	120
VI.8	Séparation des sources dans des mélanges bruités	121
VI.9	Etude expérimentale	132
VI.10	Conclusion	139

VI.1 Introduction

La Séparation Aveugle de Sources (SAS) est un problème général en traitement du signal, dont le principe consiste à estimer un ensemble de signaux inconnus dits 'sources', à partir d'un ensemble de signaux connus dits 'observations'. Ces observations sont généralement des mélanges de ces sources et proviennent de capteurs comme par exemple les sondes et les accéléromètres. Nous pouvons observer sur chaque capteur, la sortie d'un système réalisant le 'mélange' des signaux sources. Vu les nombreuses applications que permet la SAS, son exploitation est devenue primordiale dans de nombreuses disciplines scientifiques telles que les télécommunications, l'acoustique, le génie biomédical, l'astrophysique, la mécanique, ... etc. La nature du mélange et le milieu de propagation de ces sources sont généralement inconnus. Aucune information n'est donc disponible sur les sources ni sur les mélanges. Vu ces ambiguïtés, il est difficile voire impossible de retrouver les sources sans faire quelques hypothèses. La plupart des méthodes de SAS supposent l'indépendance statistique des sources, elles consistent donc à rendre les observations indépendantes au sens statistique. Nous retrouvons dans la littérature différentes méthodes exploitant cette idée dont certaines utilisent les statistiques d'ordre deux, d'autres les statistiques d'ordre supérieur alors que d'autres encore exploitent l'Information Mutuelle (IM). La séparation à l'ordre 2 est essentiellement fondée sur la décorrélation des signaux observés. Sauf dans le cas de signaux sources gaussiens, l'utilisation de la décorrélation n'exploite pas complètement l'indépendance qui est une propriété bien plus forte que la décorrélation. C'est ici qu'apparaissent l'intérêt de l'usage des statistiques d'ordre supérieur en séparation de sources. On obtient ainsi des critères permettant de réaliser la séparation sans information supplémentaire. Ceci conduit à un nouveau concept : l'Analyse en Composantes Indépendantes (ACI) introduite par Jutten et Héroult dans [90], et développée par Comon dans [91]. Cette nouvelle approche vient compléter l'Analyse en Composantes Principales (ACP) fondée sur la seule décorrélation, qui est une méthode très utilisée en statistique. L'ACI est une des voies majeures de la SAS. Son principe, sous l'hypothèse d'indépendance mutuelle des sources, consiste à transformer linéairement le vecteur des signaux observés en un vecteur dont les composantes sont statistiquement indépendantes. En tant que procédé de séparation aveugle de sources typique, l'ACI est connue pour être capable d'extraire efficacement les informations latentes dans des signaux complexes, même lorsque le modèle de mélange des sources est inconnu. L'analyse en composantes indépendantes a été développée dans le but de récupérer les signaux source d'origine, en utilisant des hypothèses minimales sur la matrice de mélange. Ce qui rend l'ACI fondamentalement différente de l'analyse factorielle classique est que les signaux sources sont supposés être non-gaussiens en plus de l'hypothèse de base de leur indépendance. Cela implique que le modèle peut être estimé de façon unique à partir des données, en utilisant des statistiques supplémentaires. Il peut être conclu que la forme élémentaire du modèle de l'ACI peut être comparable à la recherche de la projection adéquate qui maximise la non-gaussienneté des sources. L'ACI a été utilisée avec succès dans la séparation de signal acoustique et le débruitage des images. Récemment, elle a été introduite en analyse des signaux dynamiques des systèmes mécaniques. Les composants indépendants estimés portent l'information directe sur les sources de vibration, et donc, ils peuvent être

utilisés à des fins de diagnostic des machines. Cependant, cette méthode ne peut être appliquée directement lorsqu'il n'y a qu'un seul capteur de collecte des signaux. La méthode ACI nécessite des signaux à partir de capteurs séparés au moins égal au nombre de sources à identifier. Il est bien connu que l'analyse en composantes indépendantes (ACI) est une méthode utile pour la séparation aveugle de sources même si elle a quelques inconvénients. Tout d'abord, l'indice initial des sources est inconnu. C'est-à-dire, l'ACI ne fournit pas d'étiquetage ou de permutation des signaux récupérés à l'égard de leurs sources réelles. Ceci provient du fait que l'indépendance en mathématique est insensible à la permutation des sources. La seconde ambiguïté est que l'échelle réelle des sources ne peut être déterminée. Cela signifie que les signaux récupérés peuvent être une version amplifiée ou atténuée des signaux d'origine. C'est aussi en raison de l'insensibilité de l'indépendance en mathématique au facteur d'échelle. En outre, les composants indépendants pourraient ne pas être constants dans les calculs répétés du fait que l'ACI s'appuie sur des techniques statistiques de traitement du signal. En d'autres termes, les algorithmes fournissent des résultats quelque peu différents à chaque compilation de l'ACI. Enfin, avant d'entamer cette approche, il nous a semblé utile de passer en revue au préalable l'état de l'art du problème de la SAS qui ne peut être exhaustif. Nous rappelons le problème de la SAS et les éléments qui s'y rattachent, la modélisation du mélange, l'objectif et les difficultés de la résolution de ce problème

VI.2 Modélisation mathématique des mélanges

Les signaux sources se propagent dans un environnement appelé communément milieu de propagation. Ils y subissent naturellement des transformations. Les signaux observés (mesurés) sont donc des mélanges plus ou moins complexes des signaux sources. La nature du mélange et le milieu de propagation de ces sources sont généralement inconnus. Aucune information n'est donc disponible sur les sources ni sur les mélanges. Vu ces ambiguïtés, il est difficile voire impossible de retrouver les sources sans faire quelques hypothèses. La plupart des méthodes de SAS supposent l'indépendance statistique des sources, elles consistent donc à rendre les observations indépendantes au sens statistique. Parallèlement, les mélanges peuvent être 'instantanés' ou 'convolutifs'. Ils peuvent être aussi variants ou invariants dans le temps. En général, ce sont les modèles linéaires invariants dans le temps qui sont considérés dans la plupart des applications. Les mélanges linéaires sont utilisés lorsque la relation signaux observés signaux sources est linéaire : on considère que chaque observation est constituée d'une somme pondérée des sources, ou des versions filtrées de celles-ci. Les mélanges non-linéaires sont utilisés si la relation entre les signaux observés et les signaux sources est non linéaire. Le nombre p de sources présentes dans les mélanges joue aussi un rôle important. Il est souvent supposé connu, mais peut aussi, suivant les applications, être estimé en amont (par exemple par Analyse en Composantes Principales, ACP). Suivant la valeur de p , par rapport au nombre d'observations r , on a trois configurations différentes du problème de SAS, qui peuvent nécessiter des méthodes différentes :

- Lorsque p est égal à r , on parle de mélanges déterminés.
- Si p est inférieur à r , on a plus d'observations que de sources pour résoudre le problème, on est donc au moins dans un cas équivalent au précédent, si ce n'est plus avantageux. On parle dans ce cas de mélanges 'surdéterminés'.
- Si p est supérieur à r , on se retrouve avec plus de signaux inconnus que de données, on est alors dans le cas 'sous-déterminé'. La résolution du problème est dans ce cas-là plus difficile et nécessite des contraintes supplémentaires.

Dans tous les cas, le nombre d'inconnues est souvent supérieur aux données, car en plus des sources il y a aussi les paramètres de mélange à estimer. Les méthodes de SAS nécessitent donc des contraintes supplémentaires sur les sources et/ou paramètres de mélange pour diminuer le champ des solutions possibles au problème. Ce sont les différentes natures de contraintes imposées qui ont donné lieu aux différentes familles de méthodes de SAS existantes. Ces contraintes ou hypothèses dépendent bien sûr du type de données que l'on désire traiter et donc de l'application.

VI.3 Modèle de mélange linéaire instantané

C'est le mélange le plus simple et le plus répandu dans la littérature. Le processus de mélange entre les sources et les observations est modélisé par l'équation, en temps continu, sous forme vectorielle par l'équation suivante :

$$x(t) = As(t) \quad (6.1)$$

$s(t) = (s_1(t); \dots ; s_p(t))^T$ est le vecteur des signaux sources que l'on cherche à estimer ; il est composé de p signaux, $x(t) = (x_1(t); \dots ; x_r(t))^T$ est le vecteur des signaux observés ; il est constitué de r composantes et $A \in \mathbb{R}^{r \times p}$ est la matrice de mélange (inconnue) ; elle est composée de coefficients scalaires. On parle ici de mélange 'instantané' car à un instant t de l'observation $x_i(t)$ dépend uniquement des sources à ce même échantillon :

$$x_i(t) = \sum_{j=1}^p a_{ij} s_j(t) \quad (6.3)$$

La version discrète de l'équation (6.3) s'écrit :

$$x(n) = [A(n)]s(n) \quad (6.2)$$

En cas de présence de bruit additif dans l'environnement, le modèle mathématique de mélange relatif à cette situation s'écrit

$$x(t) = As(t) + b(t) \quad (6.4)$$

où $b(t) = (b_1(t); \dots ; b_r(t))^T$ est le vecteur des signaux bruits supposé être centré et dont les composantes sont généralement supposées mutuellement indépendantes, et indépendantes des

signaux sources. Le schéma de modélisation de la relation sources-observations peut être illustré par la figure (6.1).

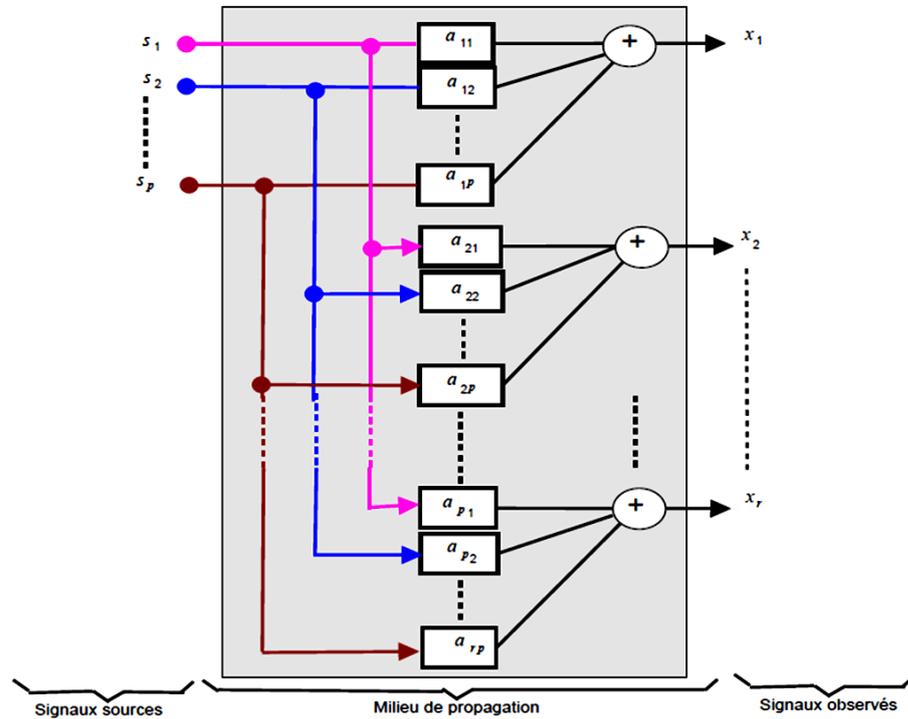


Fig. 6.1 Schéma de modélisation d'un mélange linéaire de sources

Idéalement, le but de la SAS est de retrouver avec exactitude les signaux sources à l'origine des observations que l'on a à disposition. En réalité, le problème posé induit l'estimation des sources avec quelques degrés d'incertitudes, qui dépendent du modèle de mélange et des hypothèses possibles sur les sources (propriétés des sources), ainsi que de la méthode utilisée. On appelle ces incertitudes des 'indéterminations'. En effet, la séparation peut être obtenue avec une infinité de solutions [92]. La contrainte d'indépendance des sorties n'est cependant pas suffisante pour résoudre complètement le problème. Comon présente dans [91] l'ensemble des indéterminations pour le cas linéaire instantané. Il montre que sans condition supplémentaire, les sources ne peuvent être estimées qu'à une permutation et un facteur d'échelle près. Dans la section suivante, nous allons présenter ce problème et son incidence sur la reconstruction des sources.

VI.3.1 L'indétermination de permutation

La première indétermination liée au problème de la SAS est celle de la permutation. En effet, l'ordre de restitution des signaux est arbitraire car toute permutation appliquée sur le vecteur s et sur les colonnes de la matrice A correspondante donne naissance au même vecteur x . Dans ce cas, la relation qui lie les deux vecteurs s et x sous forme vectorielle s'écrit :

$$x(t) = \sum_{j=1}^p a_j s_j(t) \quad \text{où} \quad a_j = (a_{1j}, \dots, a_{rj}(t))^T \quad (6.5)$$

En utilisant la propriété de commutativité de l'opération d'addition, la relation ci-dessus peut s'écrire comme suit :

$$x(t) = \tilde{A}\tilde{s}(t) \quad (6.6)$$

où \tilde{A} et $\tilde{s}(t)$ sont respectivement la nouvelle matrice de mélange et le nouveau vecteur source. Cette permutation peut être modélisée par une matrice de permutation P telle que

$$x(t) = As(t) = (AP)(P^{-1}s(t)) = \tilde{A}\tilde{s}(t) \quad (6.7)$$

Nous aboutirons donc au même vecteur d'observations $x(t)$. Nous concluons que la multiplication à droite de la matrice de mélange A par une matrice de permutation P ne change pas les mélanges, mais implique que l'ordre des sources est modifié.

VI.3.2 L'indétermination du facteur d'échelle

Nous avons vu dans le paragraphe précédent que la permutation des colonnes dans la matrice de mélange et des sources, ne change en rien les observations. Nous allons voir dans ce paragraphe, que la multiplication d'une colonne (de la matrice de mélange) et la division d'une source par un scalaire non nul ne changera pas le vecteur mélange. En effet, pour tout $\alpha_i \neq 0$; on a :

$$x(t) = \sum_{i=1}^p a_i s_i(t) = \sum_{i=1}^p (\alpha_i a_i) \left(\frac{s_i(t)}{\alpha_i} \right) \quad (6.8)$$

Cela montre que l'amplitude des sources est indéterminée. Il est à noter que le principal inconvénient des méthodes de la SAS réside dans l'indétermination liée à la puissance des sources estimées.

VI.4 Principe de résolution de la SAS

Sous l'hypothèse d'indépendance statistique des sources, la résolution du problème de la SAS dans le cas de mélanges linéaires instantanés, consiste à appliquer des transformations aux observations pour obtenir des signaux statistiquement indépendants. Le but est d'estimer les sources s à partir des observations x . L'estimateur s'écrit :

$$y(t) = Wx(t) \quad (6.9)$$

où W est la matrice séparante. Le problème consiste à déterminer cette matrice scalaire W à partir des observations $x(t)$, qui conduit à la 'meilleure' estimation possible des sources $s(t)$. Il existe différentes types de méthodes pour résoudre ce problème réparties en deux grandes familles : les méthodes par bloc et les méthodes adaptatives.

VI.4.1 Méthode de traitement par bloc

Les méthodes de séparation par bloc sont basées sur le schéma de la figure (6.2).

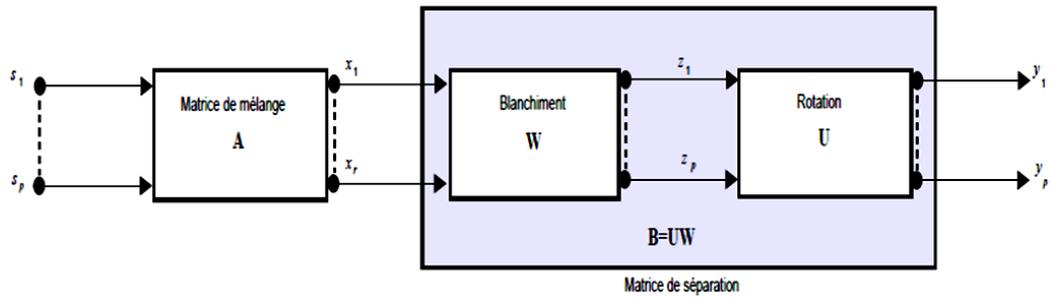


Fig. 6.2 Schéma de la méthode de séparation par bloc

Pour ce modèle nous pouvons distinguer deux approches :

- **Les approches à une seule étape**

Ces approches consistent, à partir des observations, à identifier directement la matrice de mélange ou à trouver une matrice W qui, appliquée au vecteur des observations $x(t) = (x_1(t); \dots ; x_p(t))^T$, permet d'obtenir des signaux $y(t) = (y_1(t); \dots ; y_p(t))^T$ statistiquement indépendants.

- **Les approches à deux étapes**

La première étape consiste à 'Blanchir' (i.e., décorréler) les signaux observés $x(t)$; cela se fait par multiplication de $x(t)$ par la matrice de 'Blanchiment' B : les signaux blanchis (décorrélés) s'écrivent donc :

$$z(t) = Bx(t) \quad (6.10)$$

Diverses méthodes ont été développées pour réaliser cette transformation et l'une des plus populaires est l'utilisation de la décomposition en valeurs propres de la matrice de covariance de x . Le blanchiment des données ne permet évidemment pas d'assurer l'estimation des composantes indépendantes, mais simplifie celle-ci de moitié. Le blanchiment des données n'est pas systématique. Le blanchiment est cependant recommandé puisqu'il permet d'accélérer la convergence des algorithmes. L'algorithme FastICA, qui requiert le blanchiment préalable des données sera présenté à la fin de cette section. La deuxième étape consiste à trouver une matrice unitaire (de rotation) U qui, appliquée aux signaux $z(t)$, rend ces derniers statistiquement indépendants. Cette dernière étape se fait généralement par optimisation d'un critère d'indépendance.

VI.4.2 Méthodes adaptatives

Les méthodes adaptatives de séparation sont généralement schématisées comme illustré sur la figure (6.3).

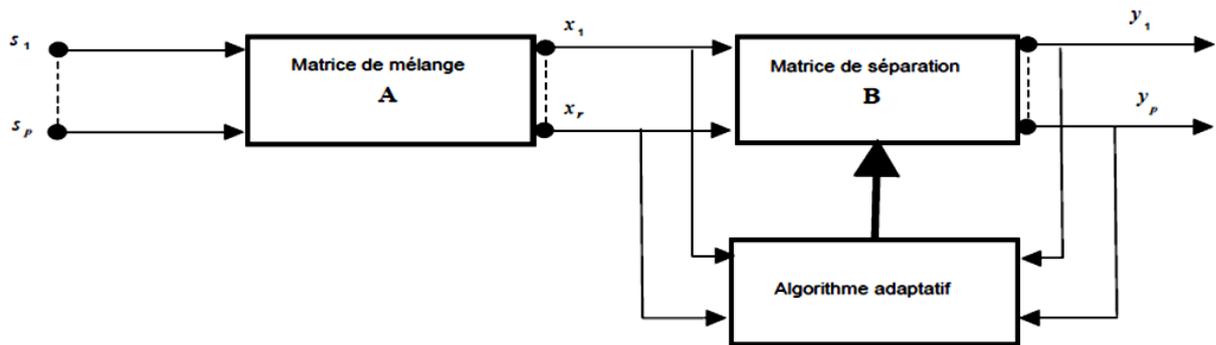


Fig. 6.3 Schéma de la méthode de séparation adaptative

Le principe de ces méthodes consiste à modifier de manière itérative la matrice séparante W pour rendre les signaux, reconstitués à partir des observations, statistiquement indépendants. Ces méthodes ont vu le jour grâce à l'algorithme proposé par Jutten et Héroult [93], présenté à l'aide d'un réseau de neurones récurrent. Il est à noter que le principe général utilisé pour déterminer la matrice séparante W est basé sur l'optimisation d'un critère d'indépendance.

VI.5 Analyse en Composantes Indépendantes (ACI)

L'ACI ou l'ICA (pour Independent Component Analysis, en anglais) est une des voies majeures de la séparation aveugle de sources. Son principe, sous l'hypothèse d'indépendance mutuelle des sources, consiste à transformer linéairement le vecteur des signaux observés en un vecteur dont les composantes sont statistiquement indépendantes, éventuellement à certaines indéterminations près (permutation et échelle) comme nous l'avons mentionné précédemment ; tout en supposant un mélange linéaire des sources au niveau des capteurs et en utilisant des techniques impliquant des statistiques d'ordre supérieur. Il s'avère que, dans un contexte non bruité, cela est équivalent à l'estimation des paramètres du modèle instantané. En ACI, les signaux sources sont supposés aléatoires, mais les signaux vérifiant à la fois les trois propriétés suivantes ne peuvent être séparés à savoir :

- Pour chaque signal source, les variables aléatoires associées aux échantillons du signal sont statistiquement indépendantes,
- Le signal source est 'identiquement distribué' et donc stationnaire,
- La loi de probabilité associée au signal source est gaussienne.

Il a, en effet, été démontré, que séparer des signaux i.i.d. (indépendants et identiquement distribués) gaussiens est impossible [94]. Les méthodes ACI sont donc généralement basées sur la non vérification de l'une de ces propriétés.

Définition 6.1: *Mathématiquement parlant, l'ACI d'un vecteur aléatoire $x \in \mathbb{R}^r$ consiste à estimer le modèle génératif des données $x = As$, avec $A \in \mathbb{R}^{r \times p}$, en déterminant une transformation linéaire $s = Wx$, avec $W \in \mathbb{R}^{p \times r}$, de telle manière que les composantes de*

$s \in \mathbb{R}^p$ soient aussi indépendantes que possible par maximisation d'une fonction $F(s)$ qui est une mesure de l'indépendance statistique.

Plus précisément, Comon a donné dans [91], la définition de l'ACI suivante :

Définition 6.2: L'ACI d'un vecteur aléatoire $x(t) = (x_1(t); \dots ; x_r(t))^T$ est défini comme la donnée d'une matrice de 'séparation' W telle que le vecteur $y(t) = (y_1(t); \dots ; y_p(t))^T = Wx(t)$ soit à composantes indépendantes, dans le sens de la maximisation d'un 'contraste' $C(y_1(t); \dots ; y_p(t))^T$ qui mesure l'indépendance des variables aléatoires $y_1(t); \dots ; y_p(t)$.

De nombreuses méthodes implémentations différentes de l'ACI peuvent être trouvées dans la littérature. Afin de simplifier ce problème, le bruit est omis dans la suite de ce paragraphe puisque le problème d'ACI sans bruit de mesure est assez complexe en lui-même. De plus, ce modèle sans bruit suffit dans de nombreuses applications. Le problème plus exacte de l'ACI avec du bruit dans les mesures sera introduit plus loin. Les sources s_i ne sont pas observées directement. Leurs densités de probabilité $f_i(s_i)$ ne sont évidemment pas connues car il est rare en pratique d'avoir une connaissance étendue de ces fonctions. Il est donc impossible en général d'utiliser la définition de l'indépendance de l'équation (6.13) pour estimer des sources indépendantes. Les méthodes d'ACI reposent donc sur une mesure d'indépendance qui reste à définir. Les différentes approches se différencient par la mesure d'indépendance retenue et par la méthode d'optimisation choisie. Mais toute méthode d'ACI repose sur les mêmes hypothèses et restrictions que nous allons répertorier dans le paragraphe suivant. En substance, l'ACI suppose un ensemble r de signaux mixtes $x = [x_1, x_2, \dots ; x_r]^T$ mesurée dans des lieux différents et composé linéairement de p de sources inconnues et statistiquement indépendants $s = [s_1, s_2, \dots ; s_p]^T$. Ces signaux linéairement mixtes peuvent être décrits comme suit:

$$x_i = \sum_{j=1}^p a_{ij}s_j \quad i = 1, \dots, r; \quad j = 1, \dots, p \quad (6.11)$$

où x_i est le $i^{\text{ème}}$ signal mixte mesurée dans la position du capteur i , s_j est le $j^{\text{ème}}$ signal source, a_{ij} sont les coefficients de mélange. $x(t) = [x_1(t) \dots x_r(t)]^T$ est un vecteur de signaux mélangés, $s(t) = [s_1(t) \dots s_p(t)]^T$ est le vecteur inconnu des signaux sources, A est la matrice de mélange inconnue ayant dimension $(r \times p)$. Le but de l'ACI est de trouver une matrice de séparation $W = [w_1, w_2, \dots ; w_p]^T$, qui est utilisé pour calculer le signal estimée $\hat{s}(t)$ des signaux de sources comme :

$$\hat{s}(t) = y(t) = Wx(t) \quad (6.12)$$

Calculer une matrice de séparation W permet donc d'estimer les sources $y(t) = [y_1(t) \dots y_p(t)]^T$ à partir des signaux mélangés $x(t)$. En d'autres termes, l'ACI peut être définie comme une méthode qui recherche une transformation linéaire qui maximise la

non gaussianité des composants de $s(t)$. Le mode opératoire de l'application de l'analyse en composantes indépendantes est le suivant: Selon le théorème de la limite centrale (*TLC*), la distribution de la somme d'un grand nombre de variables aléatoires indépendantes tend vers une distribution gaussienne. Du fait que les signaux recueillis sont des sommes pondérées des sources indépendantes, les sources à isolées doivent être moins gaussiennes que les signaux recueillis. Ainsi, une mesure d'indépendance est conçu pour classer l'interdépendance entre les éléments de y . Les valeurs des éléments de la matrice W sont estimées de manière itérative jusqu'à ce que l'indépendance maximale de y est atteinte. Parmi les mesures les plus populaires de l'indépendance sont: maximisation de la non-gaussianité, la minimisation de l'information mutuelle, l'estimation du maximum de vraisemblance et (dans des études antérieures) la dé-corrélation non linéaire. Bien que ces mesures soient différentes en principe, parfois, leurs algorithmes sont assez semblables.

VI.5.1 Mesures d'indépendance et algorithmes

L'indépendance entre des composants est estimée par quelques mesures caractéristiques et autour desquelles ont été développés différents algorithmes. Certains ont eu un impact considérable dans la communauté de l'ACI par leur efficacité, leur généralité, et leur vitesse de convergence [95, 96]. Cependant, pour simplifier et permettre de résoudre le problème, on admet quelques hypothèses.

Hypothèses de base

Hypothèse 6.1: Les signaux sources $s_i(t)$ de $s(t)$ sont mutuellement statistiquement indépendants avec une distribution non gaussienne. D'un point de vue mathématique, cela signifie que la densité de probabilité conjointe des p sources peut se factoriser comme le produit de leurs densités marginales

$$f\left(s_1(t - k_1), s_2(t - k_2), \dots, s_p(t - k_p)\right)^T = \prod_{i=1}^p f_i(s_i(t - k_i)) \quad (6.13)$$

pour tout instant t et pour tout décalage k_i , $i = 1, \dots, p$. Cette condition d'indépendance se simplifie dans le cas des mélanges linéaires instantanés en posant $k_i = 0$. Les composants s_i , $i \in \{1, \dots, p\}$, sont supposées statistiquement indépendantes. Cette hypothèse est fondamentale pour garantir l'estimation du modèle direct $x = As$ de l'ACI. Les composants indépendantes doivent avoir des distributions non-gaussiennes. Il s'avère qu'au plus une composante indépendante peut avoir une distribution gaussienne.

Hypothèse 6.2: Dans le cas où les signaux sources sont mutuellement statistiquement indépendants et les échantillons de chaque source indépendants et identiquement distribués (*iid*), on suppose qu'au plus une source est autorisé à suivre une distribution gaussienne. Comon [91], dans le cadre de l'hypothèse 2 et dans le cas d'un mélange linéaire instantané, a montré que les signaux $y = Wx$ sont indépendants si et seulement si la matrice séparante W est de la forme suivante :

$$W = DPA^{-1} \quad (6.14)$$

où D représente une matrice diagonale et P une matrice de permutation. Donc l'indépendance est équivalente à la séparation à un facteur d'échelle et à une permutation près. Dans le cadre de l'hypothèse 2, s'il y a plus d'une source gaussienne, l'indépendance ne conduit pas forcément à la séparation [94].

Hypothèse 6.3: Le nombre de signaux de source doit être inférieurs ou égale au nombre des signaux mixés observés ($p \leq r$). Cette hypothèse est nécessaire dans la plupart des algorithmes existants. Cependant, certaines méthodes traitent le cas où le nombre d'observations est inférieur au nombre de sources ($r < p$). Dans ce cas contraire, le mélange est dit sous-déterminé et le problème n'est pas soluble sans connaissances *a priori* supplémentaires sur les sources puisque même si la matrice de mélange A est connue, elle n'est pas inversible.

Hypothèse 6.4: La plupart des méthodes de SAS supposent que les mélanges sont linéaires. Cette hypothèse est largement considérée dans la plupart des méthodes existantes. Cependant, certains algorithmes traitent le cas où les mélanges sont non-linéaires. Il existe néanmoins moins de méthodes non linéaires que linéaires dans la littérature et cela est essentiellement dû à la complexité des méthodes non linéaires, mais aussi à la difficulté d'appréhender le modèle de mélange auquel on est confronté, suivant l'application.

Par ailleurs, d'autres hypothèses supplémentaires peuvent être faites sur les sources. En général ces hypothèses permettent la conception de nouveaux algorithmes. Les méthodes d'ACI exploitées dans la suite de ce travail divisent le problème de séparation en deux, à savoir une étape d'identification de la matrice de mélange A et une étape d'extraction des sources. Mais d'autres méthodes telles que l'algorithme de Héroult et Jutten [90] mettent à jour les sorties en premier plutôt que les mélanges, mais elles ne seront pas utilisées ici. Dans la suite du travail, nous nous placerons dans le cas plus simple où le nombre de sources est inférieur ou égal au nombre de signaux mélanges observés. La matrice de mélange A doit être inversible c'est-à-dire que ses colonnes doivent être linéairement indépendantes. Les mélanges redondants peuvent être éliminés de l'analyse tout en s'assurant que le nombre de lignes de A reste supérieur ou égal à son nombre de colonnes pour éviter le cas sous-déterminé. Sous ces hypothèses, le modèle de l'ACI présenté dans le paragraphe précédent est identifiable. La matrice de mélange A et le vecteur des sources s sont ainsi estimables sur la seule connaissance du vecteur des observations $x(t)$. Les deux indéterminations inhérentes à la modélisation de l'ACI subsistent. La prise en compte de ces deux indéterminations signifie que les sources indépendantes ne sont extractibles qu'à une matrice $\Delta = DP$ près, où la matrice D est diagonale, d'éléments α_j pour $j \in \{1, \dots, p\}$ et de rang plein, et la matrice P est une matrice de permutation. Des prétraitements sur le vecteur des données, tels que le centrage et le blanchiment, permettent de s'affranchir de certaines indéterminations et de simplifier le problème de séparation.

VI.5.2 Prétraitements

a. Centrage

Afin de simplifier la théorie et les algorithmes d'ACI, les composantes du vecteur des observations x sont forcées à une moyenne nulle sans perte de généralité. Ceci est lié aux cumulants croisés et à leurs estimateurs qui sont d'expression beaucoup plus simple dans le cas de variables aléatoires centrées. Cette étape est réalisée par une simple soustraction de la moyenne des différentes composantes estimée à partir de leurs réalisations. Le vecteur $x(t)$ est ainsi transformé en

$$x^c(t) = x(t) - E\{x(t)\} \quad (6.15)$$

Par cette opération, les composantes indépendantes sont elles-aussi centrées. La matrice de mélange A n'est pas modifiée par cette opération qui peut donc toujours être appliquée aux données sans affecter l'estimation de A .

b. Blanchiment

Cette étape consiste à décorréliser et à imposer une variance unité aux variables du vecteur centré $x(t)$. Cette transformation linéaire est réalisée en multipliant $x(t)$ par une matrice B de dimensions $p \times r$ afin d'obtenir le vecteur :

$$z(t) = Bx(t) \quad (6.16)$$

$z(t)$ est le vecteur blanchi de dimensions $p \times 1$ où les composantes de $z(t)$ sont décorrélées et de variance unité, ce qui signifie que la matrice de covariance de $z(t)$ vaut :

$$R_z = E\{z(t) z^T(t)\} = I_p \quad (6.17)$$

où I_p est la matrice identité de dimensions $p \times p$.

Diverses méthodes ont été développées pour réaliser cette transformation et l'une des plus populaires est l'utilisation de l'ACP. La décomposition en valeurs propres de la matrice de covariance de $x(t)$. Le blanchiment des données ne permet évidemment pas d'assurer l'estimation des composantes indépendantes, mais simplifie celle-ci de moitié. Le blanchiment des données n'est pas systématique. Certains algorithmes d'ACI s'appliquent directement sur les données centrées comme l'algorithme Infomax de Bell et Sejnowski [97]. Le blanchiment est cependant recommandé puisqu'il permet d'accélérer la convergence des algorithmes [98, 99]. L'algorithme FastICA de l'ACI, qui requièrent le blanchiment préalable des données sera présenté à la fin de cette section.

VI.6 Critères d'indépendance utilisés en ACI

Comme nous l'avons mentionné précédemment, la plupart des méthodes de la SAS utilisent l'hypothèse 1. Différents critères d'indépendance peuvent être trouvés dans la littérature. Par

définition, deux variables aléatoires sont dites indépendantes lorsque leur densité de probabilité conjointe est égale au produit de leurs densités marginales. Nous rappelons, dans ce qui suit, quelques critères qui permettent de vérifier l'indépendance des sources et qui sont utilisés en SAS.

VI.6.1 Information Mutuelle (IM)

L'information mutuelle (IM) est un critère permettant la mesure de la dissimilarité entre la densité conjointe d'un ensemble de signaux et le produit des densités marginales de chacune de ses composantes. Pour un vecteur aléatoire $y(t) = (y_1, \dots, y_p)^T$ l'information mutuelle $I(y)$ s'écrit :

$$I(y) = \int_{\mathbb{R}^p} f_y(t) \log \frac{f_y(t)}{\prod_{i=1}^p f_{y_i}(t)} dt \quad (6.18)$$

où f_y est la densité conjointe du vecteur aléatoire y ; et f_{y_i} la densité marginale de la variable aléatoire y_i , $i = 1, \dots, p$. D'autre part, la divergence de Kullback-Leibler modifiée (KL_m) entre les deux densités $\prod_{i=1}^p f_{y_i}$ et f_y est définie par :

$$KL_m \left(\prod_{i=1}^p f_{y_i}, f_y \right) = - \int_{\mathbb{R}^p} f_y(t) \log \frac{\prod_{i=1}^p f_{y_i}(t)}{f_y(t)} dt = \int_{\mathbb{R}^p} f_y(t) \log \frac{f_y(t)}{\prod_{i=1}^p f_{y_i}(t)} dt \quad (6.19)$$

D'où la relation d'égalité entre l'information mutuelle et la divergence de Kullback-Leibler modifiée :

$$I(y) = KL_m \left(\prod_{i=1}^p f_{y_i}, f_y \right) \quad (6.20)$$

Ainsi, en utilisant les propriétés de la fonction logarithme, on peut montrer que l'IM est toujours positive ou nulle. En particulier, elle est nulle $I(y) = 0$ si et seulement si : $f_y = \prod_{i=1}^p f_{y_i}$, i.e., si et seulement si les composantes du vecteur $y = (y_1, \dots, y_p)$ sont statistiquement indépendantes.

Il est démontré également que l'IM est égale à la différence entre la somme des entropies des variables aléatoires y_1, \dots, y_p et l'entropie du vecteur aléatoire y . Rappelons d'abord la définition de l'entropie d'un vecteur aléatoire ou d'une variable aléatoire. Nous allons utiliser la notation \mathbb{E} pour désigner l'espérance mathématique. L'entropie du vecteur aléatoire y s'écrit :

$$H(y) = -\mathbb{E} \left(\log f_y(y) \right) = - \int_{\mathbb{R}^p} f_y(t) \log f_y(t) dt \quad (6.21)$$

L'entropie de la variable aléatoire y_i , $i = 1, \dots, p$, s'écrit :

$$H(y_i) = -\mathbb{E}(\log f_{y_i}(y_i)) = - \int_{\mathbb{R}^p} f_{y_i}(t) \log f_{y_i}(t) dt \quad (6.22)$$

Un calcul simple conduit à la relation suivante :

$$I(y) = \sum_{i=1}^p H(y_i) - H(y) \quad (6.23)$$

Compte tenue de la définition relative à la divergence de Kullback-Leibler modifiée, on peut utiliser l'IM comme critère de la SAS. En effet, en définissant la SAS comme une transformation linéaire inversible du vecteur observé $x(n)$;

$$y(n) = Wx(n) \quad (6.24)$$

où W est une matrice de dimension $p \times p$; la séparation est obtenue par minimisation de $I(y) = I(Wx)$:

$$\min_W I(Wx) \quad (6.25)$$

Si on note $\hat{W} = \operatorname{argmin}_W I(Wx)$ le vecteur : $\hat{y} = \hat{W}x$ est donc de composantes statistiquement indépendantes. En utilisant la relation (6.24), on peut montrer que :

$$H(y) = H(x) + \log |\det(W)| \quad (6.26)$$

Puisque nous avons :

$$f_y(y) = \frac{f_x(x)}{|\det(W)|} \quad (6.27)$$

L'application du logarithme à l'équation (6.27) nous donne

$$\log f_y(y) = \log f_x(x) - \log |\det(W)| \quad (6.28)$$

D'où, en prenant l'espérance, on obtient

$$H(y) = H(x) + \log |\det(W)| \quad (6.29)$$

L'équation (6.23) peut donc s'écrire

$$I(y) = \sum_{i=1}^p H(y_i) - H(x) - \log |\det(W)| \quad (6.30)$$

L'IM a été introduite pour la première fois comme un critère de séparation de sources par Comon [91]. Comme le calcul de la relation (6.18) est impossible car il dépend de densités inconnues ; la séparation de signaux est réalisée en minimisant un 'estimateur' de (6.30).

Ce critère peut être estimé par différentes méthodes, dans la littérature, on trouve : la méthode à noyau, la méthode des espacements, la méthode polynomiale, ... etc. Notons qu'il existe une relation très forte entre l'IM et la méthode d'estimation statistique par maximum de vraisemblance.

VI.6.2 Maximum de Vraisemblance

Une approche très répandue pour résoudre le problème de séparation de sources par l'ACI est obtenue par la maximisation de la vraisemblance. L'objectif est de chercher les paramètres du mélange qui maximisent la probabilité d'occurrence des observations. Dans un mélange linéaire instantané, le vecteur des signaux mélangés est :

$$x(n) = As(n) \quad (6.31)$$

Nous pouvons exprimer la densité de probabilité du vecteur $x(n)$ en fonction des densités de probabilité des sources $s(n)$ et du déterminant de la matrice inverse B de la matrice de mélange A . On a :

$$f_x(x(n)) = |\det(B)| f_s(s(n)) = |\det(B)| \prod_{i=1}^p f_{s_i}(s_i(n)) \quad (6.32)$$

Cette densité peut aussi être donnée en fonction des lignes b_i de la matrice B par la relation :

$$f_x(x(n)) = |\det(B)| f_s(s(n)) = |\det(B)| \prod_{i=1}^p f_{s_i}(b_i x(n)) \quad (6.33)$$

Considérons un ensemble de N échantillons indépendants et de même loi du vecteur x : $x(n), n = 1, \dots, N$. La vraisemblance V de l'obtention de cet ensemble peut s'écrire comme le produit de cette densité évaluée à ces N points :

$$V(B) = \prod_{n=1}^N \left[|\det(B)| \prod_{i=1}^p f_{s_i}(b_i x(n)) \right] \quad (6.34)$$

Il est souvent plus pratique d'utiliser le logarithme de la vraisemblance qui s'écrit

$$\log V(B) = \sum_{n=1}^N \sum_{i=1}^p \log(f_{s_i}(b_i x(n))) + N \log |\det(B)| \quad (6.35)$$

En divisant par le nombre d'échantillons N , nous avons

$$\frac{1}{N} \log V(B) = \sum_{i=1}^p \frac{1}{N} \sum_{n=1}^N \log f_{s_i}(b_i x(n)) + \log |\det(B)| \quad (6.36)$$

D'autre part, par la loi des grands nombres, on a

$$\frac{1}{N} \sum_{n=1}^N \log f_{s_i}(b_i x(n)) \approx \mathbb{E}(\log f_{s_i}(b_i x)) \quad (6.37)$$

Lorsque N est suffisamment grand. D'où

$$\frac{1}{N} \log V(B) \approx \sum_{i=1}^p \mathbb{E}(\log f_{s_i}(b_i x)) + \log |\det(B)| \quad (6.38)$$

Ce qui donne

$$\frac{1}{N} \log V(B) \approx - \sum_{i=1}^p H(b_i x) + \log |\det(B)| = - \sum_{i=1}^p H(y_i) + \log |\det(B)| \quad (6.39)$$

Comme $H(x)$ représente une constante par rapport à la matrice de séparation B , nous avons :

$$I(y) \approx - \frac{1}{N} \log V(B) + C^{\text{te}} \quad (6.40)$$

Le critère de l'IM est égal à une constante près à l'opposé du critère de log-vraisemblance. Gaeta et Lacoume [100, 101] et Cardoso [94] été les premiers à montrer la possibilité d'utiliser le principe du maximum de vraisemblance pour réaliser la séparation de sources dans le cas de mélanges linéaires instantanés ou convolutifs.

VI.6.3 Non-gaussianité et algorithme FastICA

Cette classe de méthodes suppose donc les signaux non gaussiens et peut alors être appliquée à des sources i.i.d., c.à.d. les signaux réunissant les deux premières caractéristiques citées précédemment. Ces méthodes sont les plus utilisées dans la littérature. Elles exploitent les statistiques d'ordre supérieur à deux, et sont donc dites méthodes statistiques d'ordre supérieur. Sans l'hypothèse de non-gaussianité des signaux, les statistiques d'ordre supérieur à deux n'apporteraient aucune information supplémentaire. Le principe est de forcer les signaux de sortie (les sources estimées) à être les plus non-gaussiens possible. Le principe découle du théorème de la limite centrale 'central limite', selon lequel les combinaisons des sources (donc les observations) aléatoires indépendantes et de même densité de probabilité tendent vers des gaussiennes. Ainsi, les composantes du vecteur $x(t)$ des mélanges possèdent une distribution plus proche de la distribution gaussienne que les composantes du vecteur $s(t)$ des sources. L'un des critères les plus utilisés pour mesurer la non-gaussianité est le kurtosis ou auto-cumulant d'ordre 4. On trouve aussi des mesures de non-gaussianité basées sur l'entropie différentielle ou la négentropie. Parmi les algorithmes les plus connus basés sur la maximisation de la non-gaussianité, on trouve FastICA. L'algorithme FastICA est basée sur un schéma itératif de point fixe pour trouver un maximum de la non-gaussianité d'un composant projetée $w_j^T x$. Cet algorithme est proposé et développé par Aapo Hyvarinen et al. [98, 99] et se caractérise par une convergence d'ordre élevé.

La non-gaussianité, qui est la fonction de contraste de l'algorithme FastICA, est mesurée avec l'entropie différentielle (négentropie). La négentropie est un concept hérité de la théorie de l'information qui fournit une mesure de la non-gaussianité d'une variable aléatoire. Sa définition dépend de l'entropie que nous allons définir. Dans son contexte original, l'entropie mesure le degré d'incertitude d'une variable aléatoire. Plus une variable est aléatoire, non prédictible et non structurée, et plus son entropie est grande. Sous forme mathématique, l'entropie d'un vecteur aléatoire s de densité de probabilité $f_s(s)$ est définie par :

$$H(s) = - \int f_s(s) \log f_s(s) ds \quad (6.41)$$

$$H(s) = - \int p(s) \log p(s) ds \quad (6.42)$$

La négentropie a été construite afin de fournir une quantité qui soit nulle pour une variable gaussienne et qui soit non-négative pour tout autre type de variable aléatoire. Elle est définie comme la différence entre l'entropie d'un vecteur aléatoire gaussien u et l'entropie du vecteur considéré s . Le vecteur gaussien u utilisé dans la mesure de la négentropie est contraint à avoir une matrice de covariance identique à celle de s . Sous forme mathématique, la négentropie est définie par :

$$J(s) = H(u) - H(s) \quad (6.43)$$

La négentropie peut être considérée comme la mesure optimale de la non-gaussianité. Cependant, dans sa forme actuelle, la négentropie n'est pas attractive pour l'estimation du modèle de l'ACI puisqu'elle dépend de la densité de probabilité du vecteur $s(t)$ qui est supposé inconnu et dont l'estimation est particulièrement lourde. Ainsi, plusieurs approximations de la négentropie sont développées, tels que celle définis par Hyvärinen [98, 99] en utilisant une fonction non quadratique G et s'écrit dans le cas monodimensionnel comme suit :

$$J(s_j) \approx c \left[(E\{G(s_j)\}) - E[G(u_j)] \right]^2 \quad (6.44)$$

où G peut être toute fonction non-quadratique et c une constante. Les composantes indépendantes sont obtenues par $Wz(t)$ où $z(t)$ est le vecteur des données blanchies et où la matrice $W = [w_1, \dots, w_n]^T$ est estimée de manière à ce que ses lignes w_j^T maximisent la fonction objective suivante :

$$J_G(W) = \sum_{j=1}^p \left[(E\{G(w_j^T z)\}) - E[G(u)] \right]^2 \quad (6.45)$$

La variable u est une variable gaussienne de même variance que $w_j^T z$. Le choix de la fonction G doit obéir à certaines propriétés afin que l'estimateur W soit consistant, qu'il ait une variance asymptotique minimale et qu'il soit robuste.

L'algorithme de point fixe rapide est une méthode de calcul très efficace pour effectuer l'estimation de l'ACI. Il utilise un point fixe pour le processus d'itération qui se trouve plus rapide que les méthodes conventionnelles.

VI.7 Critères de performances de la SAS

Afin d'étudier la qualité de la séparation des algorithmes, il est possible de mesurer la précision de chaque source estimée y_i en fonction de la vraie source s_i par des critères de type erreur quadratique moyenne ou rapport signal sur résidus.

VI.7.1 Erreur quadratique moyenne (EQM)

L'erreur quadratique moyenne (EQM) mesure la moyenne du carré de l'écart entre le signal source s_i et le signal estimé y_i , $i = 1, \dots, p$. Ce terme s'écrit comme suit :

$$EQM_i = \widehat{\mathbb{E}}\{(s_i - y_i)^2\} = \frac{1}{N} \sum_{n=1}^N [s_i(n) - y_i(n)]^2 \quad (6.49)$$

Où $\widehat{\mathbb{E}}$ désigne la moyenne temporelle et N représente le nombre d'échantillons utilisés. La valeur moyenne de l'EQM sur toutes les sorties est :

$$EQM = \frac{1}{p} \sum_{i=1}^p EQM_i = \frac{1}{p} \sum_{i=1}^p \frac{1}{N} \sum_{n=1}^N [s_i(n) - y_i(n)]^2 \quad (6.50)$$

La qualité de séparation s'apprécie naturellement avec une EQM la plus petite possible.

VI.7.2 Rapport signal sur résidus (SNR)

Le rapport signal sur résidus (noté *SNR*) est la mesure de performance la plus répandue dans la séparation de sources. Dans le cas de mélanges instantanés, le rapport signal sur résidus (*SNR*) est défini comme le logarithme du rapport en décibel (*dB*) de la puissance de la source s_i sur celle de l'écart entre la source s_i et son estimée y_i . Il s'écrit donc comme suit (en supposant qu'il n'y a pas de permutation).

$$SNR_i = 10 \log_{10} \frac{\widehat{\mathbb{E}}\{s_i^2\}}{\widehat{\mathbb{E}}\{(s_i - y_i)^2\}}, \quad i = 1, \dots, p \quad (6.51)$$

La qualité de séparation s'apprécie avec une valeur du SNR la plus grande possible. Ce qui signifie qu'il n'y a pas une contribution importante provenant d'autres sources (s_i , $i = 1, \dots, p$) à cette sortie y_i .

VI.8 Séparation des sources dans le cas des mélanges bruités

L'approche en composantes indépendantes (ACI) devenue l'approche de référence en séparation de sources basée sur l'hypothèse principale de l'indépendance des sources ne permet pas de considérer explicitement une éventuelle erreur ou incertitude de modélisation ou un quelconque bruit de mesures. Les méthodes de séparation présentées dans les sections précédentes, liées à cette approche, supposent implicitement que le modèle d'observation (que ce soit un mélange instantané/convolutif, ou linéaire/non-linéaire) est exact et que les données ne sont pas bruitées. En pratique, lors de l'application, ces méthodes se heurtent à des difficultés car souvent la réalité est plus complexe. Il est donc impératif de tenir compte du bruit dans le modèle pour une bonne restauration des sources. Dans la littérature, ce problème malgré son importance a été peu abordé. Par bruit, on entend l'ensemble des incertitudes sur les données (modélisation, discrétisation, quantification, ...etc., et le bruit de mesure proprement dit). Lorsque le mélange est linéaire, on peut alors modéliser les observations par l'équation suivante :

$$x(t) = As(t) + b(t) \quad (6.52)$$

$s(t) = (s_1(t); \dots ; s_p(t))^T$ est le vecteur des signaux sources que l'on cherche à estimer ; il est composé de p signaux.

$x(t) = (x_1(t); \dots ; x_r(t))^T$ est le vecteur des signaux observés; il est constitué de r composantes.

$A \in \mathbb{R}^{r \times p}$ est la matrice de mélange (inconnue) ; elle est composée de coefficients scalaires.

$b(t) = (b_1(t); \dots ; b_r(t))^T$ est le vecteur des signaux bruits supposé être centré et dont les composantes sont généralement supposées mutuellement indépendantes, et indépendantes des signaux sources. L'ajout du bruit entraîne typiquement deux problèmes :

- L'estimation de la matrice de séparation est généralement biaisée
- Même si l'estimation est exacte, $W = A^{-1}$ (pour simplifier, on ne tient pas compte des indéterminations), la restauration des sources est bruitée : $\hat{s}(t) = s(t) + A^{-1}b(t)$.

L'identification de sources vibratoires dans le cas d'un mélange linéaire bruité est connue pour être un problème inverse mal-posé, ce qui signifie que l'existence d'une solution unique et stable n'est pas garantie. Une approche classique pour pallier cette difficulté consiste à inclure dans la formulation du problème inverse de l'information *a priori* sur les sources à identifier pour contraindre l'espace des solutions admissibles. Cette idée est à la base des méthodes de régularisation [70]. Pour le cas général où $r \neq p$, une solution régularisée est définie comme l'optimisation d'un critère à deux parties :

$$\hat{s} = \arg \min_s \{ \|x - As\|_2^2 + \lambda \|s\|_2^2 \} \quad (6.53)$$

Cette régularisation développée depuis le travail pionnier de Tikhonov [102] introduit des termes de régularisation quadratique pour tenir compte de certaines propriétés antérieures de

la solution (régularisé). Depuis, de nombreux différents termes de régularisation ont été proposées. En particulier, à la place de la norme ℓ_2 on a proposé d'utiliser la norme ℓ_0 ou la norme ℓ_1 pour renforcer la parcimonie de la solution [103]. Cependant, en raison du fait que ℓ_0 est non convexe et ℓ_1 est convexe, mais pas continue, l'optimisation d'un critère de ces expressions est plus difficile que dans le cas de la norme ℓ_2 . Pour cette raison, un grand nombre d'ouvrages spécialisés en littérature proposent des algorithmes pour l'optimisation de ces critères. Fait intéressant, la définition de la solution du problème (6.53) comme l'optimisation d'un critère en deux parties peut être assimilée à un maximum d'une solution a posteriori (MAP) dans une approche bayésienne où le premier terme du critère d'optimisation peut être liée à la probabilité de vraisemblance et le second terme à un modèle *a priori* que nous verrons dans la suite. L'approche bayésienne peut aller plus loin que la régularisation au moins dans les aspects suivants:

- Une meilleure prise en compte des caractéristiques du bruit;
- Une façon meilleure et plus facile pour traduire les connaissances antérieures et en particulier la parcimonie;
- Un nouvel outil pour évaluer le paramètre de régularisation et d'optimisation.

VI.8.1 Approche par analyse en composantes parcimonieuses 'SCA'

L'hypothèse de parcimonie suppose que la plupart des coefficients des sources sont nuls ou proches de zéro. La distribution des valeurs des coefficients présente alors un pic étroit en zéro et une queue lourde [104]. Une telle distribution est souvent assimilée à la distribution Laplacienne. L'étape d'estimation des sources consiste à maximiser la parcimonie des sources sous la contrainte d'égalité $x = As$. En pratique on définit la mesure de parcimonie d'un vecteur, par le nombre de composantes non nulles de ce vecteur, c'est à dire sa norme ℓ_0 . Bien que ce critère garantisse l'unicité et l'exactitude de la solution, si les sources ont une certaine parcimonie et sous certaines conditions sur A [105], ce problème d'optimisation est combinatoire, et impraticable en pratique pour des matrices de grandes dimensions. Pour ces raisons, il est utilisé en général une forme relaxée de la mesure de parcimonie qui n'est autre que la norme ℓ_p . Le critère à optimiser est donc le suivant:

$$f(A, s) = \arg \min_{s/x=As} \|s\|_p \quad (6.54)$$

L'optimisation de ce critère, correspond au calcul des coefficients s les plus vraisemblables sachant que $x = As$ et sous l'hypothèse que les sources soient des Gaussiennes Généralisées (\mathcal{GG}) de paramètre p (voir la section V.4.2 pour plus de détails sur le modèle Gaussien Généralisé). Les valeurs de p comprises entre 0 et 1 inclus produisent des solutions parcimonieuses assez semblables, mais les algorithmes mis en jeu ne sont pas les mêmes. Pour des valeurs de p telles que $1 < p \leq 2$, le problème est strictement convexe, et contient par conséquent un minimum global, mais les solutions obtenues ne sont plus vraiment parcimonieuses. Pour des valeurs de p telles que $p < 1$, le problème est non convexe et contient des minimums locaux. Cependant pour une valeur de $p = 1$, le problème reste

convexe et les solutions obtenues sont parcimonieuses, ce qui explique pourquoi ce cas particulier a été beaucoup étudié dans la littérature. Si $p = 2$, la Gaussienne généralisée se ramène à une distribution gaussienne, tandis que pour $p = 1$, la Gaussienne généralisée se ramène à une distribution Laplacienne. Pour tenir compte d'un éventuel bruit additif, ou de l'inexactitude strict de l'hypothèse de parcimonie, le problème se ramène à la forme lagrangienne suivante :

$$\hat{s} = \arg \min_s \|x - As\|_2^2 + \lambda \|s\|_p \quad (6.55)$$

où le paramètre λ détermine le compromis entre la qualité de l'approximation de la reconstruction représentée par le premier terme du critère, et l'hypothèse de parcimonie des sources représentée par le second terme. Pour les signaux qui sont très peu parcimonieux dans le domaine temporel (leur domaine d'origine), il existe des transformations linéaires qui permettent d'obtenir des représentations dans lesquelles les coefficients des sources sont parcimonieux. C'est le cas par exemple des transformées temps-fréquences ou temps-échelle. La transformation étant linéaire, la matrice de mélange garde la même forme dans le domaine transformé que dans le domaine temporel. La SCA se divise alors en quatre étapes :

- Transformation du mélange $x(t)$ dans une représentation parcimonieuse $x(t, f)$;
- Estimation de la matrice de mélange \hat{A} à partir de $x(t, f)$;
- Estimation des sources $\hat{s}(t, f)$ à partir de \hat{A} et $x(t, f)$ par un critère de parcimonie
- Reconstruction des sources dans le domaine temporel $\hat{s}(t)$ en appliquant la transformation inverse.

Enfin l'analyse en composantes parcimonieuses prend comme hypothèse que les sources sont parcimonieuses dans une représentation et propose de diviser le problème de séparation de sources en deux étapes bien distinctes : l'estimation de la matrice de mélange connaissant les observations, puis l'estimation des sources connaissant les observations et la matrice de mélange. L'avantage de cette architecture modulaire séparant clairement une première étape d'estimation de la matrice de mélange et une seconde d'estimation des sources, est qu'il est alors possible de connecter n'importe quelle méthode de la première étape avec n'importe quelle méthode de la seconde.

VI.8.2 Approche bayésienne pour l'identification aveugle des sources bruitées

Une autre approche considérée en séparation de sources est l'approche par inférence bayésienne qui permet entre autre de prendre en compte naturellement toute incertitude ainsi que toute information disponible (ou information imposée: contrainte) afin de définir des solutions au problème mal posé de la séparation de sources bruitées. La formulation du problème de séparation de sources dans un cadre d'estimation bayésienne est assez récente. Les premiers travaux remontent à Roberts [106] et Knuth [107, 108] qui ont présenté des algorithmes dans lesquels, en plus d'un modèle *a priori* sur les signaux sources, des informations sur les coefficients de mélange sont introduites.

L'article de Mohammad-Djafari [109] présente les liens avec les méthodes d'analyse en composantes indépendantes et le cadre général de la méthodologie de séparation de sources par approche bayésienne. En fonction des informations considérées et des modèles probabilistes utilisés, plusieurs orientations ont été suivies. Plus récemment, les travaux de Févotte [110] montrent que la parcimonie est une information supplémentaire qui peut être introduite pour la séparation dans le cas sous-déterminé. Le schéma global de cette démarche repose sur l'attribution de lois de probabilités aux données recherchées. Les trois étapes suivantes résument le processus du calcul Bayésien:

- Ecrire la vraisemblance des inconnues $p(x/s, A)$
- Attribuer des lois *a priori* aux inconnues du problème : $p(s)$ et $p(A)$. Cela permet de décrire la connaissance *a priori* que l'on a du problème, mais aussi de restreindre l'espace des solutions possibles.
- La loi de Bayes permet alors de déduire la loi *a posteriori* $p(s, A/x)$, qui est utilisée pour estimer les inconnues.

Formellement, la solution du problème inverse correspond à calculer des estimateurs en faisant appel à des algorithmes d'optimisation et intégration itératif tel que l'algorithme EM (Expectation Maximization) ou des modélisations par chaînes de Markov cachée. L'avantage principal de l'approche estimation bayésienne est qu'elle offre un cadre théorique favorable pour traiter de tel problèmes car la parcimonie des sources peut être prise en compte explicitement d'une part, et d'autre part, la formulation probabiliste du problème permet d'utiliser des méthodes Monte Carlo par Chaînes de Markov (MCMC) qui offrent une alternative à la résolution analytique des problèmes d'optimisation, notamment dans le cas de densités de probabilité *a posteriori* complexes. L'approche bayésienne offre un cadre mathématique rigoureux permettant de combiner des informations probabilistes et structurelles et peut donner des résultats intéressants mais à l'inconvénient de souvent demander un grand coût calculatoire [111].

VI.8.3 Principe d'estimation bayésienne pour la séparation de sources

L'idée principale de l'approche bayésienne en séparation de sources est de prendre en compte toute information préalablement disponible sur les signaux sources et les coefficients de mélange. Le problème est formulé dans un cadre probabiliste en affectant des densités de probabilité aux différentes variables. Le modèle direct est décrit par la vraisemblance qui résulte de la modélisation du processus d'observation et le modèle inverse est caractérisé par la traduction des informations disponibles sur les signaux sources et les coefficients de mélange par des densités de probabilité *a priori*. Le point fondamental dans l'estimation bayésienne est le théorème de Bayes qui permet de fusionner toutes ces informations afin de déduire une description probabiliste *a posteriori* des variables recherchées.

La formulation bayésienne du problème d'indentification s'appuie sur la règle de Bayes comme suit :

$$p(s, A/x) = \frac{p(x/s, A) p(s, A)}{p(x)} \quad (6.56)$$

où $p(x/s, A)$ est la fonction de vraisemblance, définissant la probabilité d'observer le champ vibratoire mesuré x étant donné le champ excitateur s , $p(s, A)$ est la densité de probabilité conjointe *a priori* de s et A et $p(x)$ est la 'densité de probabilité' des observations, qui peut être vue comme un terme de normalisation. Etant donné que les variables recherchées sont A et s , et afin de simplifier l'écriture en ne gardant que les termes dans lesquels apparaissent les variables d'intérêt, l'équation (6.56) peut être exprimée par une relation de proportionnalité :

$$p(s, A/x) \propto p(x/s, A) p(s, A) \quad (6.57)$$

D'une façon générale, il est totalement justifié de supposer que la matrice de mélange est indépendante des signaux sources et par conséquent on peut écrire

$$p(s, A/x) \propto p(x/s, A) p(s) p(A) \quad (6.58)$$

En notant θ l'ensemble des hyperparamètres du problème, $\theta = \{\theta_e, \theta_s, \theta_a\}$, qui sont les paramètres inconnus des densités de probabilité affectées aux différentes variables du modèle. Ainsi, la relation (6.58) devient :

$$p(s, A/x, \theta) \propto p(x/s, A, \theta_e) p(s/\theta_s) p(A/\theta_a) \quad (6.59)$$

Dans cette approche, la probabilité $p(x/s, A, \theta_e)$ résume nos connaissances sur le bruit et le modèle reliant les données $x(t)$ observé aux inconnues $s(t)$ et le terme *a priori* $p(s/\theta_s)$ résume nos connaissances incomplètes préalable sur les inconnues et le la loi *a posteriori* $p(s, A/x, \theta)$ combine ces deux termes et contient tout notre état des connaissances sur les inconnues s après comptabilisation de l'*a priori* et les données observées. Pour renforcer la parcimonie, il est considéré des modèles probabilistes à queue lourds (heavy tailed) telles que : la distribution gaussienne généralisée, Weibull symétrique, Student-t ou Cauchy, hyperbolique généralisée et Dirichlet ou des modèles mixés telles que : mélange de gaussiennes, Bernoulli-gaussien, Bernoulli-Gamma, mixture de multinomial, etc.) [23]. Selon le modèle *a priori* sélectionnée, les calculs bayésiens (optimisation du maximum conjoint *a posteriori* (JMAP) ou MCMC ou approximation variationnelle de Bayes (VBA) pour la moyenne *a posteriori* (PM) peuvent devenir plus complexe et relativement performants. Comme un exemple simple, lorsque le bruit est supposé gaussien, alors la solution MAP (maximum *a posteriori*) $\hat{s} = \arg \max_s \{p(s, A/x, \theta)\}$ est obtenue comme l'optimisation du critère $J(s) = \|x - As\|_2^2 + \lambda \Omega(s)$ où l'expression de $\Omega(s)$ dépend de la loi *a priori*. Lorsque la connaissance *a priori* est traduit comme une loi de probabilité gaussienne alors $\Omega(s) = \|s\|_2^2$, et quand il est traduit comme une loi de probabilité de Laplace, alors $\Omega(s) = \|s\|_1$ [112].

L'autre intérêt de l'utilisation de l'approche bayésienne par rapport à l'approche de régularisation est d'avoir de nouveaux outils pour gérer les hyper paramètres [113]. Si l'on se place dans la situation où l'on ne connaît pas les valeurs de ces hyperparamètres, il est nécessaire de les estimer également. Il faut alors les prendre en compte dans (6.58) en écrivant

$$p(s, A, \theta/x) \propto p(x/s, A, \theta_e) p(s/\theta_s) p(A/\theta_a) p(\theta) \quad (6.60)$$

C'est principalement à partir de cette densité de probabilité *a posteriori* que les différents estimateurs sont construits. Les estimateurs qui peuvent être utilisés sont le maximum conjoint *a posteriori* (JMAP pour *joint maximum a posteriori*), le maximum marginal *a posteriori* (MMAP pour *marginal maximum a posteriori*) et la moyenne marginale *a posteriori* (MPM pour *marginal posterior mean*). En pratique, le choix d'un estimateur particulier est très lié à la technique de calcul utilisée pour la résolution du problème d'optimisation. Par exemple pour l'estimateur JMAP [114, 23] :

$$(\hat{s}, \hat{\theta}) = \arg \max_{(s, \theta)} \{p(s, A, \theta/x)\} \quad (6.61)$$

VI.8.4 Exemple d'une formulation bayésienne simplifié d'un problème de séparation de sources

Considérons la situation générale où le champ vibratoire $x(t)$, mesuré à la surface d'une structure, est produit par un champ excitateur inconnu $s(t)$. Lorsque la structure est linéaire, son comportement vibratoire est complètement déterminé par la connaissance de sa matrice des fonctions de transfert A , reliant le champ vibratoire $x(t)$ au champ excitateur inconnu $s(t)$. Si l'on suppose, par ailleurs, que le champ vibratoire mesuré est corrompu par un bruit de mesure $b(t)$, alors le champ vibratoire $x(t)$ est obtenu grâce au problème direct suivant :

$$x(t) = As(t) + b(t) \quad (6.62)$$

L'objectif du problème d'identification est de reconstruire le champ excitateur inconnu $s(t)$ (sources) à partir du champ vibratoire $x(t)$ mesuré à la surface de la structure. Le modèle bayésien est un modèle d'incertitude, dans lequel tous les paramètres sont supposés aléatoires. L'incertitude sur chaque paramètre est alors représentée par une distribution de probabilité, qui décrit l'état de connaissance ou l'*a priori* sur ce paramètre. Du point de vue bayésien, cela revient à trouver un estimateur ponctuel de \hat{s} qui soit un mode (i.e. qui maximise) la distribution de probabilité *a posteriori*. Mathématiquement, on cherche donc l'estimateur du maximum *a posteriori* tel que :

$$\hat{s} = \arg \max_s \{p(s/x)\} = \arg \max_s \{p(x/s)p(s)\} \quad (6.63)$$

Pratiquement, il est généralement plus simple de chercher une solution du problème dual, c'est-à-dire :

$$\hat{s} = \arg \min_s \{-\ln p(x/s) - \ln p(s)\} \quad (6.64)$$

La fonction de vraisemblance traduit l'incertitude associée à la mesure du champ vibratoire $x(t)$. Par définition, cette incertitude est principalement liée au bruit de mesure $b(t)$. Par conséquent la fonction de vraisemblance $p(x/s)$ peut se réécrire sous la forme :

$$p(x/s) = p(x - As/b) \quad (6.65)$$

ce qui traduit la probabilité d'obtenir $x - As = 0$ étant donné le bruit de mesure b . Si l'on suppose que le bruit provient de multiples causes indépendantes, alors en vertu du théorème central limite la distribution de probabilité $p(x/s)$ peut être assimilée à une loi normale de moyenne As et de variance σ^2 .

$$p(x/s) \propto \exp - \left[\frac{1}{2\sigma^2} \|x - As\|_2^2 \right] \quad (6.66)$$

La distribution de probabilité *a priori* traduit l'incertitude associée au champ non observable $s(t)$. Il s'agit d'un modèle d'expertise, qui peut être vu comme une mesure de la connaissance à priori que l'on possède sur le champ excitateur $s(t)$. Si l'on suppose qu'une structure est excitée en différentes zones par des excitations non corrélées de nature différentes, alors les champs excitateurs locaux s_i peuvent être considérés comme des vecteurs aléatoires indépendants. De ce fait, la distribution de probabilité *a priori* peut s'écrire sous la forme :

$$p(s) = \prod_{i=1}^N p(s_i) \quad (6.67)$$

où $p(s_i)$ est la distribution locale de probabilité, traduisant la connaissance *a priori* sur la nature des sources dans la zone i . Pour des raisons de simplification pratique, on suppose que les sources excitatrices s_i suivent des lois normales généralisées de moyenne nulle de la forme

$$p(s_i) \propto \exp - \left[\frac{1}{2\sigma_i^2} \|s_i\|_2^2 \right] \quad (6.68)$$

Pour obtenir la forme finale du problème d'identification, il suffit alors, d'introduire les expressions de la fonction de vraisemblance et de la distribution de probabilité *a priori* dans l'équation (6.64). Ce faisant, on obtient :

$$\tilde{s} = \arg \min_s \frac{1}{2} \left[\frac{\|x - As\|_2^2}{\sigma^2} \right] + \frac{1}{2} \sum_{i=1}^N \frac{\|s_i\|_2^2}{\sigma_i^2} \quad (6.69)$$

On remarque que l'on retrouve la méthode de régularisation de Tikhonov lorsqu'on réécrit l'équation précédente de la forme :

$$\hat{s} = \arg \min_s \frac{1}{2} [\|x - As\|_2^2] + \lambda \sum_{i=1}^N \frac{\|s_i\|_2^2}{\sigma_i^2} \quad (6.70)$$

où $\lambda = \sigma^2/2\sigma_i^2$ généralement appelé paramètre de régularisation. La résolution du problème d'optimisation (6.70) consiste à remplacer la minimisation en un processus itératif équivalent. La résolution du problème de séparation de sources par une approche bayésienne nécessite deux étapes importantes dans lesquelles certains choix doivent être faits. La première concerne la description probabiliste des différentes variables. En général les erreurs d'observation sont décrites par un modèle gaussien. Par ailleurs, le point crucial est celui du choix du modèle *a priori* sur la distribution des sources et les coefficients de mélange. Pour une information disponible, plusieurs descriptions probabilistes peuvent être employées par différents statisticiens, d'où le caractère subjectif de cette étape. Néanmoins, la complexité des calculs bayésiens et les prédispositions du statisticien sont des facteurs importants qui favorisent certains modèles par rapport à d'autres. Le deuxième point concerne la définition du risque bayésien dont résulte le critère à optimiser pour l'estimation. En général, cette étape de définition de l'estimateur est très conditionnée par la complexité des calculs nécessaires pour l'optimisation numérique. On en distingue les algorithmes EM et les méthodes MCMC. Les algorithmes EM sont généralement utilisés pour la recherche du maximum a posteriori alors que les algorithmes MC permettent de simuler des Chaînes de Markov ayant comme distribution stationnaire la densité conjointe *a posteriori*. A partir de ces échantillons, des approximations de ces statistiques peuvent être calculées en utilisant la méthode de Monte Carlo. Enfin, dans le cadre de séparation de sources, l'étape cruciale est celle de la recherche de l'information préalable sur les signaux sources et les coefficients de mélange susceptible d'assurer une bonne séparation. Idéalement, cette information doit non seulement assurer l'unicité de la solution mais aussi doit permettre d'obtenir une solution correcte. Suit alors une étape de codage parcimonieux de cette information via un choix de la densité de probabilité *a priori* la plus adéquate.

VI.8.5 Approche basée sur le pré et post-traitement des coefficients

Cette approche combine des opérations de débruitage, avant et après la séparation. Des résultats asymptotiques ont été obtenus qui démontrent une preuve de séparabilité pour les mélanges linéaires. Une première méthode pour effectuer un débruitage est d'utiliser des filtres spectraux : ils traitent les coefficients issus de la transformée de Fourier en fonction de leur fréquence, en ajustant la bande effective au signal recherché de façon à réduire la puissance du bruit, d'où une augmentation du rapport signal à bruit (RSB). Cependant, l'emploi du filtrage suppose que l'on possède une connaissance de la statistique au second ordre du signal. Le débruitage par filtrage, qui repose sur l'idée de traiter les échantillons en fonction de leur fréquence, sera d'autant plus performant que le signal et le bruit occupent des bandes de fréquences distinctes. Or, les mélanges dans notre application occupent une large bande spectrale. En revanche l'approche proposée, issue d'un travail réalisé avec B. Rivet [115] et qui a donné lieu à une publication à ICA'2004, consiste en trois étapes successives :

1. Débruiter le mélange afin d'estimer correctement la matrice de séparation W ; soit \hat{W} un tel estimateur
2. Calculer les sources bruitées : $\hat{s}(t) = \hat{W}As(t) + \hat{W}b(t)$,

3. Débruiter les sources estimées, à partir des signaux bruités, afin d'obtenir : $\hat{s}(t)_{\text{débruitée}} \approx s(t)$ (à une permutation et un facteur échelle près).

L'hypothèse principale pour la mise en œuvre de cette méthode est que le bruit occupe un spectre large, alors que les sources sont à spectre étroit. Dans ce cas, en choisissant correctement les échelles d'ondelettes, on peut réaliser simplement les opérations 1 à 3. Une première étude sur l'influence du débruitage par ondelettes préalablement à la séparation a été menée par Paraschiv-Ionescu et al. [116] qui ont proposé le schéma de débruitage selon la figure (6.4). Ainsi, pour obtenir de bonnes performances dans des environnements fortement bruités, les observations $x(k)$ sont débruitées en appliquant un seuillage doux avant d'appliquer l'algorithme de séparation de sources à \hat{x} :

$$\hat{s}(k) = W \hat{x}(k) \quad (6.71)$$

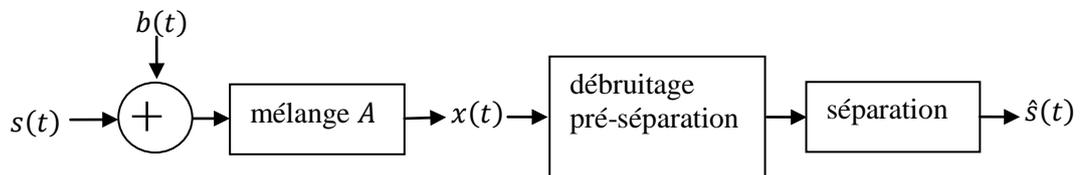


Fig. 6.4 Séparation de sources couplée à un débruitage pré-séparation

Une amélioration de la méthode existante peut être apportée en ajoutant après la séparation un débruitage post-séparation (voir figure 6.5). Les sources sont estimées par $\tilde{s}(k)$ obtenues par débruitage des signaux issues de la séparation $\hat{s}(k)$. Un débruitage classique obtenu en s'appuyant sur la seule variance du bruit σ^2 estimée à partir des coefficients en ondelettes de la première échelle de $\hat{s}(k)$. Cependant, du fait du pré-débruitage ayant été effectué avant la séparation, le bruit n'est plus considéré gaussien [116].

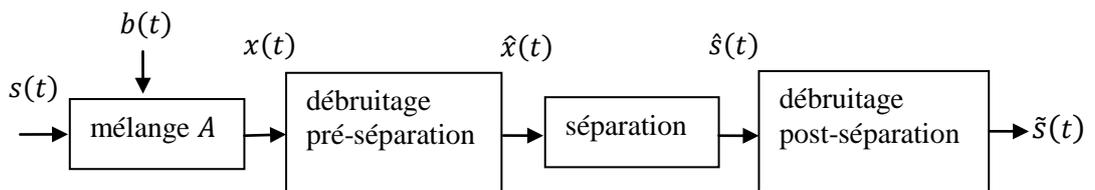


Fig. 6.5 Schéma de SAS avec débruitage pré-séparation et post-séparation

Cette amélioration permet de compléter le débruitage pré-séparation dans le cas où on aurait fait le choix de ne pas trop débruiter les observations $x(k)$ afin de ne pas perdre trop de signal ce qui serait néfaste pour l'estimation de la matrice de séparation W . En effet, un débruitage pré-séparation trop énergique risquerait de modifier la structure du mélange $x(k) = As(k)$ et de nuire à la qualité de l'estimation de la matrice de séparation.

VI.8.6 Nouvelle approche basé sur le débruitage adaptif des coefficients

Le débruitage pré-séparation risque de supprimer le signal utile, synonyme des sources d'excitation faibles et noyées dans le bruit (notamment les détails de celui-ci, correspondants aux petits coefficients de détail d'ondelettes relatifs aux sources). Si cette perte peut ne pas s'avérer gênante pour l'estimation de la matrice de séparation, elle est lourde de conséquence dans l'estimation des sources, car les plus faibles risquent d'être éliminées dans le débruitage pré-séparation. Pour palier cet inconvénient, nous avons proposé une nouvelle approche ciblée de débruitage adapté à la séparation de source. L'objectif est de choisir un débruitage post-séparation adapté à la nature des sources recherchées car on possède bien évidemment plus d'information sur les sources recherchées que sur la nature des mélanges observés. Les sources sont alors estimées par $\tilde{s}(k) = Wx(k)$ puis épurées par un débruitage post-séparation. L'algorithme donné illustré par la figure (6.7) en résume les principales étapes. Par cette méthode nous obtenons alors :

$$\tilde{s}(k) = Wx(k) = W_{\tilde{s}}(s(k) + b(k)) \quad (6.72)$$

Ce qui donne

$$\tilde{s}(k) = W_{\tilde{s}} A s(k) + W_{\tilde{s}} b(k) \quad (6.73)$$

Cette méthode présente l'avantage de pouvoir adapter le débruitage post-séparation à la séparation de sources comme suit : A partir des observations des mélanges bruités, on estime la matrice de séparation \hat{W} considéré par cette démarche comme le codage parcimonieux des coefficients du signal d'observation brute. Cette matrice de codage parcimonieux \hat{W} tel que : $\tilde{s}(k) = Wx(k)$ est estimée par l'algorithme de point fixe FastICA [99]. L'algorithme FastICA est considéré comme la projection du signal d'observation dans espace orthogonal qui maximise la non-gaussiannité des sources vibratoires. Ensuite l'estimation finale de $\hat{s}(k)$ est réalisée par un débruitage post-séparation adaptif dans le domaine des ondelettes en utilisant l'approche bayésienne et en renforçant la parcimonie des sources estimées par la considération d'une distribution Laplacienne appropriée.

VI.8.7 Connexion avec l'analyse en composantes indépendantes

L'algorithme FastICA (Fast Independent Component Analysis) intégré dans l'approche adaptif de séparation des sources vibratoire bruitées est le plus couramment utilisé pour la séparation des mélanges instantanés de sources et très connu pour son excellente performance et sa rapidité de traitement des données. FastICA met en évidence des liens forts entre l'analyse en composantes indépendantes et la poursuite de projection (*Basis Pursuit*). Cette dernière est une méthode d'analyse des données qui relève des statistiques. Elle a été initialement proposée par Jérôme H. Friedman and John Tukey [117] en 1974. Elle consiste à rechercher un sous espace dans lequel un indice de projection est maximisé. L'idée est alors de trouver des directions intéressantes pour la représentation des données ; c.à.d. exprimer les données selon un premier axe dans lequel les données sont bien représentées puis faire de même avec le reste des données sur un nouvel axe et poursuivre itérativement. Fast-ICA est

un algorithme très performant de maximisation de la fonction de contraste pour les sources non gaussiennes [118]. Il est basé sur le principe de l'algorithme d'apprentissage itératif de type point fixe (Fixed point algorithm). L'algorithme d'apprentissage du réseau cherche à trouver une direction qui est en fait un vecteur w tel que la projection $w^T x$ maximise l'aspect non gaussien. Si on considère que chaque sortie du réseau est donnée sous forme d'une fonction de type $G(w^T x)_i$ tel que G_i est une fonction scalaire non linéaire. La convergence dans FastICA signifie que les valeurs de w (anciennes et nouvelles) vont dans la même direction, c'est-à-dire que la différence entre ces deux valeurs est au-dessous du critère de convergence entre deux itérations. Les composantes indépendantes peuvent être extraites une à une. Ainsi, on obtient un gain considérable en performance si l'on a seulement besoin de quelques composantes indépendantes à extraire. La maximization de la non-gaussanité dans l'algorithme fastICA, est à la base de l'idée du codage parcimonieux qui consiste à trouver la matrice de poids W de sorte que les composants sources s_i sont plus parcimonieuses que possible. Nous rappelons qu'une variable aléatoire s_i de moyenne nulle est appelé parcimonieuse quand sa fonction de densité de probabilité présente pic prononcé à zéro et a queue lourde. La parcimonie est équivalente alors à la super-gaussianité ou à la leptokurticité (Kurtosis positif) [89]. Le codage parcimonieux est étroitement liée à l'analyse en composantes indépendantes (ICA) [88, 89]. Dans le modèle de données utilisé dans ICA, on considère que $x(t)$ est une transformation linéaire de composants indépendants $x = AS$. En Inversant la relation, on obtient $s = Wx$, W étant l'inverse de A . En outre, il a été prouvé que l'estimation du modèle de données de l'ACI peut être réduite à la recherche d'une direction de dé-corrélation dans lequel les composants sont non-gaussiens le plus possible. Si les composants indépendants sont rares, plus précisément, super-gaussien, cela équivaut à la recherche de projections non corrélées qui ont des distributions parcimonieuses. Ainsi, l'estimation du modèle ICA pour des données du mélange est à peu près équivalent à un codage parcimonieux où les composants sont contraints à être dé-corrélés. Cette connexion à l'ICA montre aussi clairement que le codage parcimonieux peut être considéré comme une méthode de réduction de la redondance, qui était en effet l'un des principaux objectifs de codage parcimonieux en premier lieu [89]. En fait, le codage parcimonieux peut être considéré comme un moyen d'adaptation au principe de détermination d'une base orthogonale de représentation des composants significatifs seulement. Ensuite, l'approche bayésienne (méthode statistique) est introduite afin d'éliminer le bruit additif gaussien après le codage parcimonieux de composants en utilisant l'algorithme FastICA. Intuitivement, parce que quelques-uns des coefficients sont actifs simultanément après l'application du codage parcimonieux i.e ICA, on peut supposer que les autres coefficients de petites valeurs absolues sont purement bruit qui devront être mises à zéro, tout en conservant les grands composants. L'avantage de ce procédé est la possibilité d'intégration de la parcimonie des données par l'adoption d'un *a priori* leptokurtique comme démontré précédemment dans l'approche bayésien en chapitre V (voir la section V.4.2 pour plus de détails sur le modèle leptokurtique).

Le diagramme suivant résume les étapes principales de l'approche adaptative adoptée pour l'identification des sources vibratoires relatives à la détection des défauts combinés.

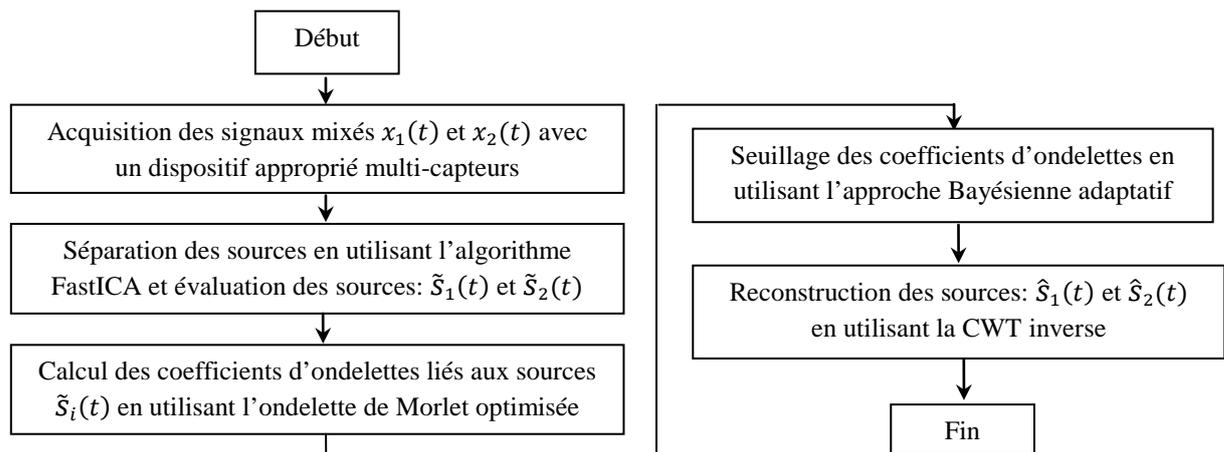


Fig. 6.7 Diagramme relatif à l'approche adaptatif de séparation des sources vibratoires bruitées

VI.9 Etude expérimentale

Le schéma fonctionnel de la figure (6.7), explique la démarche générale adoptée pour la séparation aveugles des composants vibratoire bruités et mixés. Nous supposons que les mélanges sont corrompus par un bruit additif Gaussien iid et que les débruitages sont opérés individuellement sur chaque signal (source estimée). Les signaux de vibrations utilisés pour la détection des défauts sont recueillis à partir du banc d'essai dont les photos de cette l'installation et ses accessoires sont présentés aux figures (6.8) et (6.9). Deux défauts locaux ont été introduits sur la piste de la bague extérieure pour le premier roulement et sur la piste de la bague intérieure pour le second roulement. Lorsque les billes traversent chaque défaut, des impulsions régulières apparaîtront. Selon la durée des impulsions, on peut juger de l'apparition du défaut à l'aide des fréquences caractéristiques.



Fig. 6.8 (a) Photo du banc d'essais avec deux capteurs d'acquisition (b) Banc d'essais et Analyseur double canal (B&K 2133)

Pour révéler l'information des sources inconnues à partir des signaux de vibration mixés, nous effectuons une séparation simultanée des sources indépendantes qui sont cachés dans les signaux recueillis fournies par deux capteurs. Par ailleurs, pour simuler le manque d'accessibilité aux roulements des machines complexes, deux accéléromètres B&K 4381 ont été utilisés pour acquérir des signaux d'accélération en même temps à partir du châssis du banc d'essai, comme indiqué dans la figure (6.8).

Le signal recueilli par chaque capteur est en fait, la combinaison de l'énergie des vibrations produites simultanément par les défauts des deux paliers en plus de bruit. La dissipation de l'énergie de vibration par voie de transmission complique encore davantage la situation. Les signaux sont enregistrés simultanément par l'analyseur de fréquence à double canal B&K 2133 où ils ont été amplifiés et filtrés par un filtre passe-bande dans la gamme de fréquence de 0,2 Hz à 20 kHz. Ensuite, les acquisitions sont transférées vers le PC où des programmes en Matlab ont été mis en œuvre pour exécuter la séparation aveugle des composants vibratoire. Les fréquences caractéristiques correspondantes aux roulements testés sont listées dans le tableau 1. Les amplitudes temporelles et spectrales des signaux acquis à partir des roulements défaillants avec la présence d'écaillage sur les pistes interne et externe sont respectivement, présentées par les figures (6.9 et 6.10).

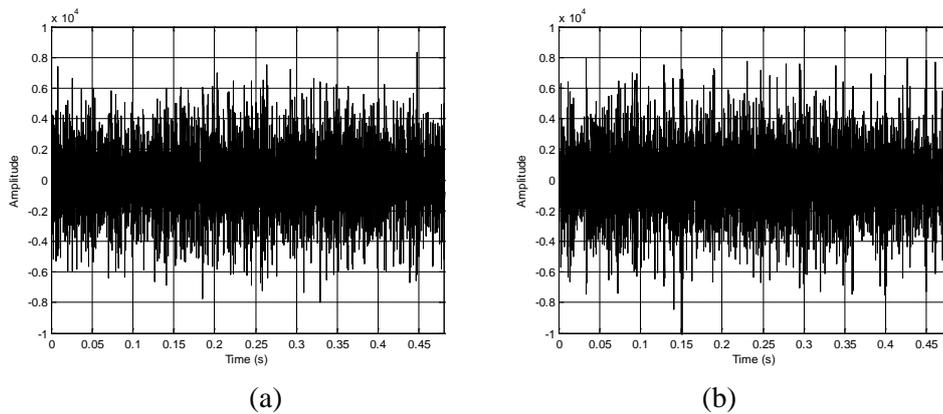


Fig. 6.9 Signaux temporels mixés du banc d'essais avec deux défauts de roulements (a) Capteur en position 1 (b) Capteur en position 2

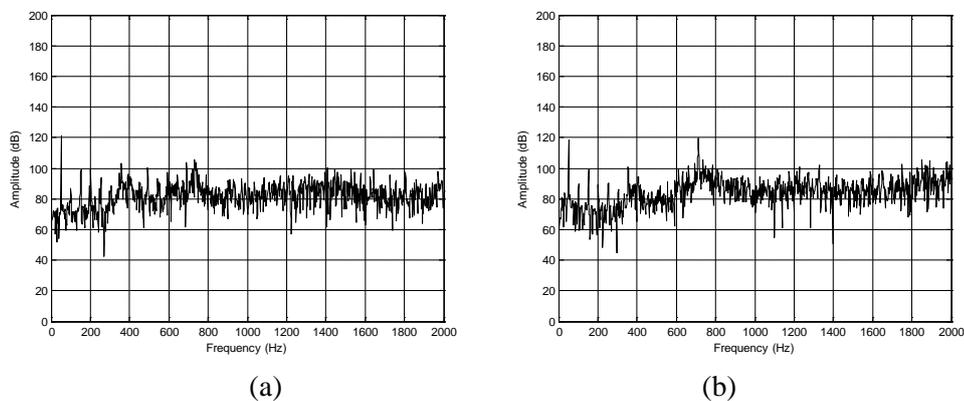


Fig. 6.10 Spectres de puissance associés (a) Spectre du signal 1 (b) Spectre du signal 2

Les diagrammes temporels et les spectres FFT ne présentent pas de symptômes clairs correspondants aux défauts simulés. En fait, les défauts des roulements ne peuvent être diagnostiqués avec certitude car les spectres fournissent des pics situés aux fréquences caractéristiques, dont les amplitudes sont comparables à celles des roulements relativement en bon état. En outre, il est remarqué aussi que le niveau de bruit relativement important masque les impulsions périodiques d'intérêt. De ce fait, la méthode d'analyse en composants indépendants (ACI) est appliquée pour séparer les signaux sources à partir des signaux de mesure obtenus par les capteurs. L'analyse est effectuée en utilisant l'algorithme FastICA.

D'après les résultats représentés sur la figure (6.11) on peut remarquer qu'il est difficile de trouver des impulsions périodiques dans les signaux.

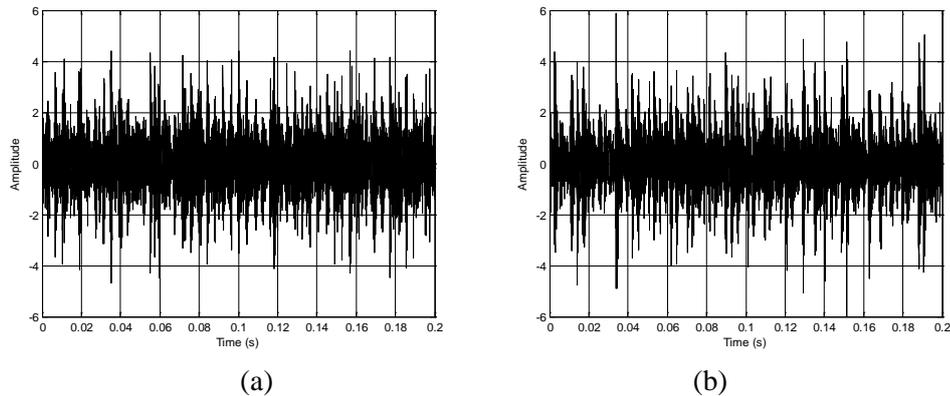


Fig. 6.11 Signaux sources séparés en utilisant l'algorithme FastICA (a) Signal source 1 (b) Signal source 2

Cependant, on constate qu'après la séparation de sources, l'intégralité du signal informatif est gardée. En effet, même si la séparation a été mal estimée, l'intégralité du signal est encore présente. Il est possible, par ailleurs, de choisir un débruitage post-séparation adapté à la nature des sources recherchées. Un certain nombre de pics aléatoires sans périodicité fixe existent comme indiqué sur la figure (6.11). Cela est dû au fait que le banc d'essais tests comme d'autres machines; contient un grand nombre de processus aléatoires. Aussi, le modèle ACI classique exige que le nombre de capteurs doit être plus grand que le nombre de sources à séparer. Cela implique qu'un certain nombre d'accéléromètres sont nécessaires pour séparer toutes les sources. Ceci, n'est pas pratique pour la conduite d'une surveillance en ligne des machines en raison de leurs coûts élevés. De ce fait les systèmes multi-capteurs sont à éviter. Néanmoins, théoriquement l'ACI se concentre sur la séparation des grands composants dominants en maximisant leur indépendance. De plus, chaque composant indépendant séparé doit contenir au moins un composant correspondant à l'une des sources physiques. Pour les signaux de vibration du banc d'essai, ce composant ne peut être que l'une des sources impulsives liées à un défaut de pistes des roulements en plus du bruit de fond. Par conséquent, un tel signal aura besoin d'un débruitage approprié dans le domaine temps-fréquence. Par conséquent, les méthodes de débruitage standard sont effectuées en premier lieu pour analyser les signaux séparés présentés sur la figure (6.11) liées aux défauts des pistes des bagues interne et externe des roulements. Les signaux débruitées par les méthodes (SURE), heuristique SURE, Minimax et la méthode au seuil universel, respectivement, sont présentés sur les figures (6.12 - 6.19). Sur les mêmes figures les deux types de seuillage dure et doux sont appliqués sur les coefficients de décomposition en ondelettes de Daubechies (db20). La décomposition basée sur l'utilisation de l'ondelette mère de Daubechies de niveau 20 est utilisée car elle offre de meilleures performances comparée à d'autres ondelettes. Toutefois, il est rappelé que les résultats expérimentaux doivent être interprétés comme une indication de performance des techniques de débruitage proposées.

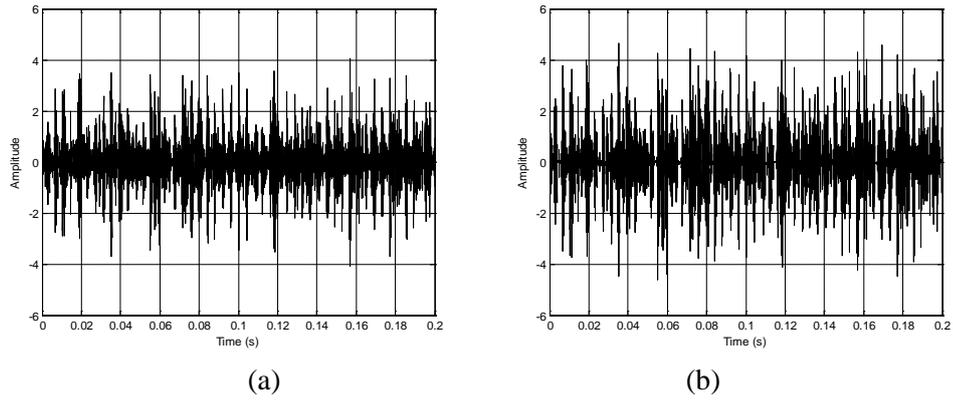


Fig. 6.12 Signal source 1 débruité en utilisant (a) Seuillage SURE doux (b) Seuillage SURE dur

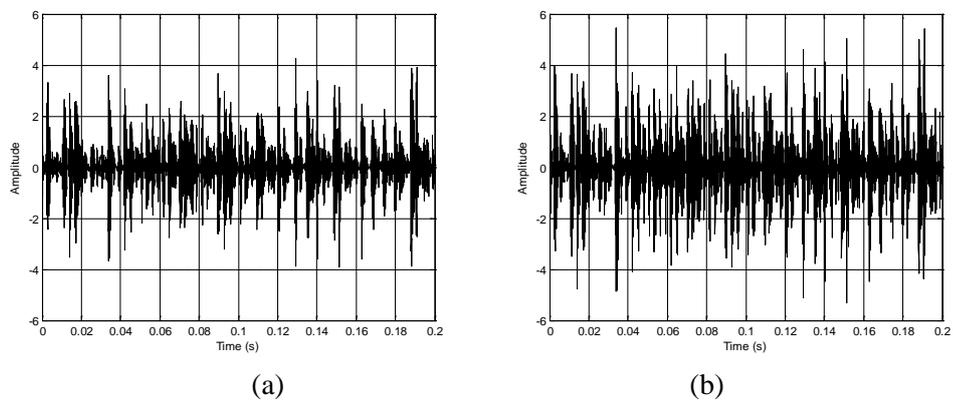


Fig. 6.13 Signal source 2 débruité en utilisant (a) Seuillage SURE doux (b) Seuillage SURE dur

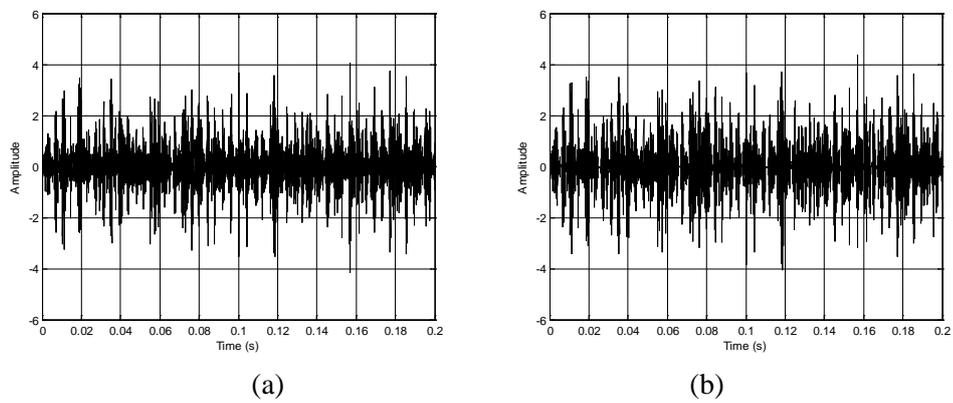
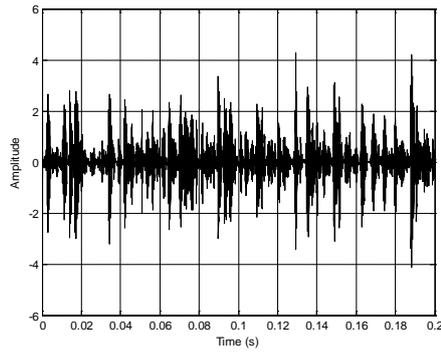
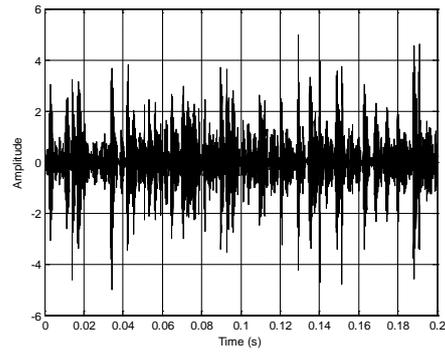


Fig. 6.14 Signal source 1 débruité en utilisant (a) Seuillage heuristique SURE doux (b) Seuillage heuristique SURE dur

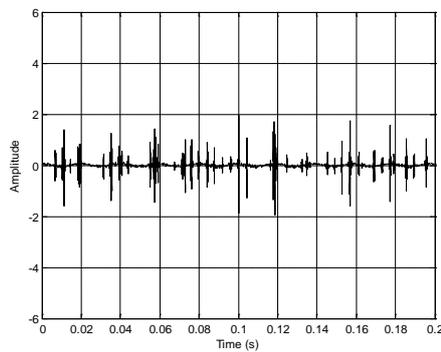


(c)

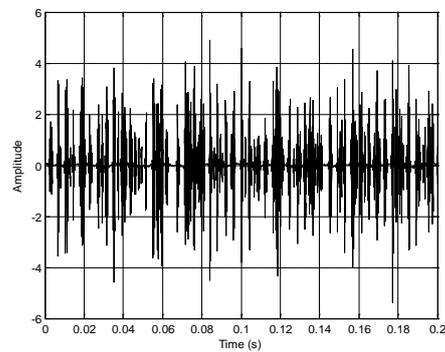


(d)

Fig. 6.15 Signal source 2 purifié en utilisant (a) Seuillage heuristique SURE doux (b) Seuillage heuristique SURE dur

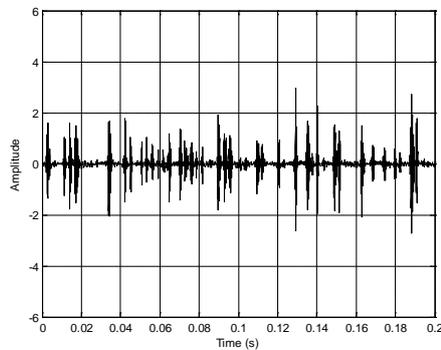


(a)

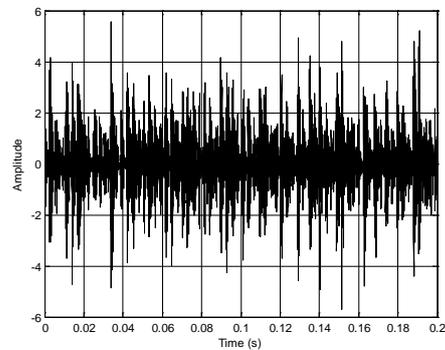


(b)

Fig. 6.16 Signal source 1 débruité en utilisant (a) Seuillage minimax doux (b) Seuillage minimax dur



(a)



(b)

Fig. 6.17 Signal source 2 débruité en utilisant (a) Seuillage minimax doux (b) Seuillage minimax dur

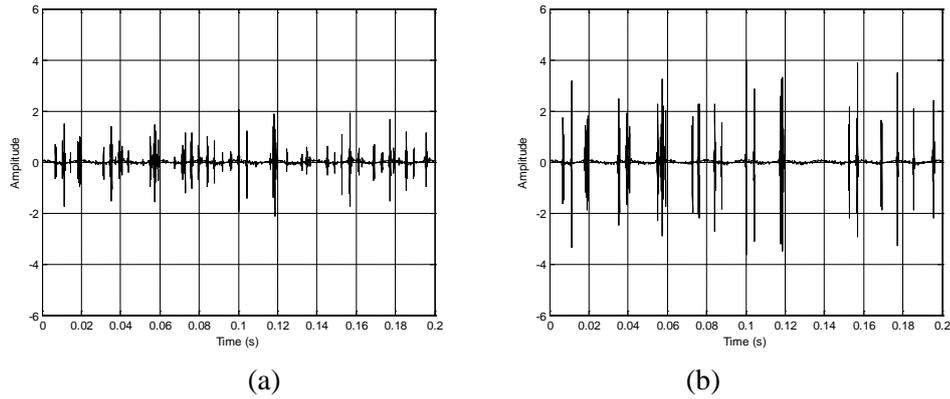


Fig. 6.18 Signal source 1 débruité en utilisant (a) Seuillage universel doux (b) Seuillage universel dur

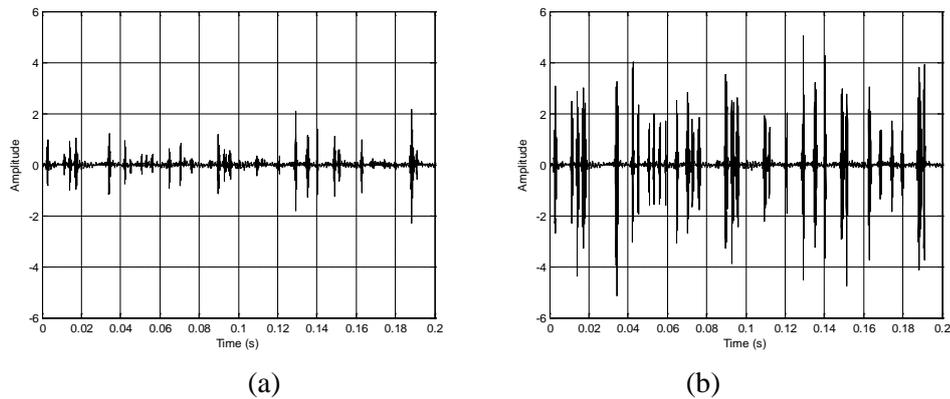


Fig. 6.19 Signal source 2 débruité en utilisant (a) Seuillage universel doux (b) Seuillage universel dur

Comme le montre les figures (6.12 - 6.19), l'observation générale montre que les incidences périodiques sont à peine visibles dans tous les cas. La réponse du système à cette excitation apparaît comme un ensemble d'impulsions irrégulières avec la présence de beaucoup de fausses impulsions. Ce genre de réponse affecterait certainement la reconnaissance des impulsions d'intérêt. Les intervalles irréguliers des amplitudes du signal présenté dans chaque figure ne sont pas capables d'isoler le phénomène par extraction de période de défaut caractéristique. Toutefois, des impulsions régulièrement espacées distinctes peuvent être observées sur les figures (6.20 et 6.21). Ces impulsions correspondent au signal reconstruit à travers la transformée en ondelettes inverse des coefficients atténués par l'approche de Bayes et en utilisant l'ondelette optimisée de Morlet. On peut remarquer que, lorsque les éléments roulants traversent les défauts, les diagrammes montrent un train d'impulsions dont l'amplitude est proportionnelle à la dimension de la zone endommagée.

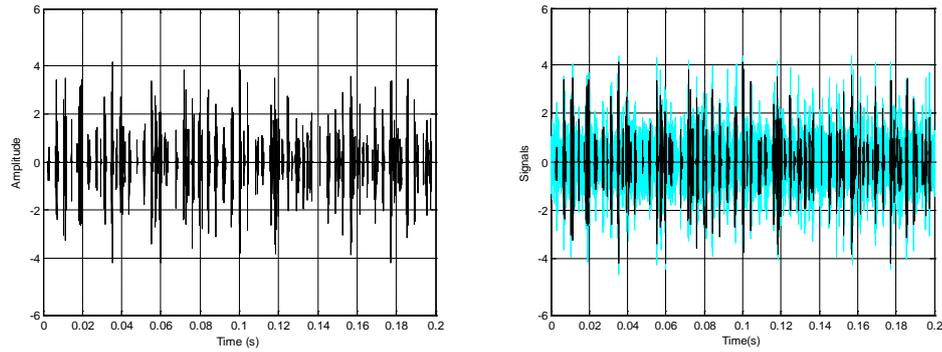


Fig. 6.20 Résultat de débruitage du signal source 1 relatif à un défaut de roulement sur la piste externe (BPFO) en utilisant l'approche Bayésienne

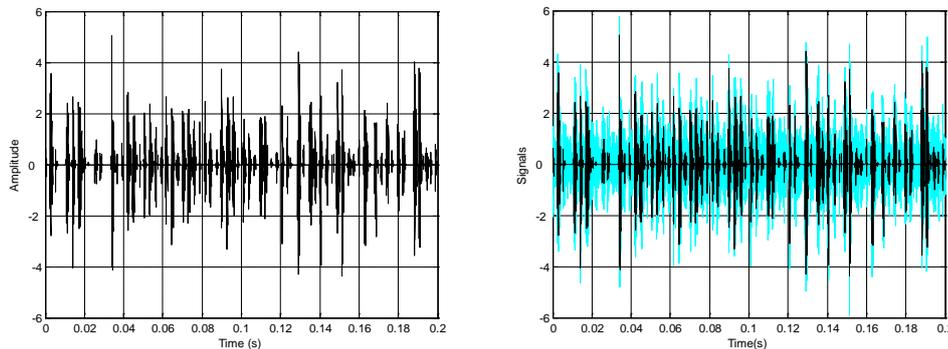


Fig. 6.21 Résultat de débruitage du signal source 2 relatif à un défaut de roulement sur la piste interne (BPFI) en utilisant l'approche Bayésienne

En comparant les résultats des figures (6.12 - 6.19) et des figures (6.20 et 6.21), on peut affirmer que, l'approche bayésienne réalise des résultats bien supérieurs comparée aux méthodes SURE, minimax et au seuil universel. D'autres observations sur les figures (6.20 et 6.21) montrent que la distance mesurée entre deux crêtes successives des impulsions présentées sur les diagrammes représente la période des défauts caractéristiques, à savoir l'inverse des fréquences caractéristiques. Dans la figure (6.20), les intervalles quasi-périodiques égaux à 4,1 ms peuvent être retrouvés, ainsi, dans la figure (6.21) les intervalles périodiques égaux à 2,8 ms peuvent être distingués. Ces intervalles quasi-périodiques sont respectivement équivalents à l'inverse de la fréquence caractéristique d'un défaut de bague externe (BPFO, figure (6.20) et d'un défaut de bague intérieure (BPFI, figure 6.21), qui sont: 243,16 Hz et 353,99 Hz, comme indiqué dans le tableau 1. Enfin, il est intéressant d'observer que seules les impulsions correspondantes aux défauts induites sont retenues dans le signal reconstruit (figures (6.20 et 6.21)). Ce fait démontre l'efficacité de l'algorithme proposé à éliminer le bruit. L'approche proposée de séparation des sources bruitées en combinant l'algorithme FastICA et le procédé de débruitage par la méthode de Bayes, a montré que les sources inconnues mixtes peuvent être récupérées avec succès sur la base de leur indépendance statistique en maximisant la non-gaussianité, même à des niveaux d'énergie plutôt faibles.

VI.10 Conclusion

Dans ce chapitre qui ne peut être exhaustif, nous avons présenté le problème de la SAS et les éléments qui s'y rattachent. Nous avons commencé par donner quelques exemples d'application de la SAS dans divers domaines. Après avoir répertorié les modèles de mélanges, nous avons rappelé différentes méthodes pour résoudre le problème de la SAS, dans le cas d'un mélange instantané, réparties en deux familles : les méthodes par bloc et les méthodes adaptatives. Nous avons expliqué que la solution fournie par les algorithmes de SAS n'est pas unique car les sources sont estimées à une permutation et à un facteur d'échelle près (dans le cas instantané). Nous avons rappelé le principe de l'analyse en composantes indépendantes, des hypothèses de base nécessaires à la séparation de sources, et plusieurs critères permettant de réaliser la séparation. Parallèlement, dans l'approche proposée, la méthode générique de l'analyse en composantes indépendantes (ACI) par le biais de l'algorithme FastICA est utilisée pour la sélection du codage parcimonieux et l'estimation des composants optimaux indépendants (ICs). L'idée principale retenue dans notre approche est que pour assurer la séparabilité des sources, il est fondamental de disposer de mesures de diversité ou de contraste entre elles (indépendance). Toutefois, les sources à rechercher ne sont cependant pas nécessairement décorrélées. Un critère supplémentaire est cependant nécessaire afin de distinguer ces sources les unes des autres. La parcimonie peut apporter ce critère et fournit ainsi une information structurelle. En effet, les signaux sont plus parcimonieux dans les espaces qui capturent bien leur morphologie, notamment dans le cas sous-déterminé. Les sources peuvent alors être identifiées en recherchant leurs quelques grands coefficients. Ainsi, les capacités de la méthode de débruitage bayésien qui exploite les propriétés statistiques des composants indépendants à débruitées d'une manière plus réaliste sont exploitées. Les expériences ont montré la performance de séparation par rapport à l'application directe de l'ACI sur le signal et les algorithmes de débruitage standard. Ceci est dû à la nature fortement non-linéaire de l'opérateur d'atténuation qui tire profit de la structure statistique inhérente du signal à identifier. À l'issue de l'utilisation de cette approche sur des roulements défailants, les signatures faibles cachées dans le bruit ont été révélées et améliorées avec succès.

Conclusion Générale

Conclusion générale

Dans ce travail de thèse, nous avons résolu le problème du manque de résolution dans la représentation temps-fréquence par l'amélioration de la parcimonie des signaux vibratoires de type impulsions. L'objectif était de montrer la pertinence des modèles parcimonieux pour traiter des problèmes inverses, notamment en tirant parti de la nature des signaux à reconstruire. Cet objectif est, nous le pensons, rempli. En effet, un modèle parcimonieux a été mis à contribution pour la modélisation des problèmes inverses étudiés notamment dans le cas de débruitage et de localisation des sources vibratoires. Au-delà de ces premiers résultats, cette méthode paraît être un bon point de départ pour résoudre d'autres problèmes inverses. La méthode de restauration des signaux et de localisation de sources développée ici adopte un point de vue différent des autres méthodes déjà existantes. Le mariage fructueux entre le modèle a priori pour la représentation parcimonieuse d'une part, et de la théorie de l'estimation Bayésienne, nous a permis d'atteindre l'objectif fixé. Il est démontré que l'approche d'estimation Bayésienne va au-delà des limites des méthodes classiques de débruitage dans le domaine des ondelettes. Dans le contexte Bayésien, la notion de parcimonie est naturellement modélisée par une distribution a priori appropriée pour les coefficients d'ondelettes. Pour formuler cette contrainte, nous avons proposé le modèle statistique adapté 'leptokurtique' pour capter le caractère parcimonieux des coefficients de représentation de ces signaux. En raison de la nature fortement non-linéaire de l'opérateur de réduction qui tire profit de la structure statistique inhérente du signal à identifier, la méthode proposée a été appliquée avec succès pour extraire les composants impulsifs noyés dans les signaux bruités. Dans l'ensemble, l'approche décrite combine les propriétés de calcul rapide, de bonnes propriétés théoriques et de bonnes performances dans les simulations et dans la pratique. Le constat majeur de l'ensemble des expériences est que la méthode actuelle a fait de meilleurs résultats lors du débruitage des signaux impulsifs. La faible signature cachée dans le bruit est révélée et renforcée avec succès. Ce résultat permet aux opérateurs de machines non seulement la détection de l'existence de défauts sur des paliers à son stade initial, mais aussi la détermination des causes de défauts en utilisant les informations des intervalles de temps qui est fournie par le signal reconstruit. La seconde contribution de ce travail de recherche a porté sur la localisation des sources vibratoires bruitées. Les méthodes classiques de séparation de sources répandues en littérature, supposent implicitement que le modèle d'observation est exact et que les données ne sont pas bruitées. En pratique, les mélanges sont généralement corrompus avec des bruits corrélés et les méthodes SAS perdent dans ce cas, de leur efficacité. En effet, après l'application directe de l'algorithme de l'ACI décrit dans ce document, sur des signaux mélangés, les impulsions synonymes des sources vibratoires restent difficiles à identifier. Ainsi, pour trouver l'information complète des sources et de révéler l'état des composants clés et leurs influences, il a été donc impératif de tenir compte du bruit dans le modèle. Une nouvelle approche basée sur l'amélioration de la parcimonie des composants indépendants a été proposée dans le présent document. Dans l'approche proposée l'ACI est utilisée pour la sélection de la transformation parcimonieuse et l'estimation des composants optimaux indépendants (CIs). Par conséquent, l'objectif de maximiser la non-gaussianité par l'ACI peut être reformulé comme extraire le codage

parcimonieux des composants du signal. En effet, l'ACI est utilisée pour extraire les composants d'influences des signaux mixés. Il peut être conclu que la forme élémentaire du modèle de l'ACI peut être comparable à la recherche de la projection adéquate qui maximise la non-gaussianité des sources. Cela implique que le modèle peut être estimé de façon unique à partir des données, en utilisant des statistiques supplémentaires. L'idée derrière l'extraction de tel codage parcimonieux est que celui-ci se prête bien au procédé d'inférences par l'approche Bayésien afin de calculer l'estimateur de ce composant d'influence i.e. la source. De plus, l'approche Bayésienne peut être considérée comme un moyen de séparer deux sources (signal et bruit) en les représentant sur la même base où l'un est fortement non-gaussien (le signal) et l'autre reste gaussien (le bruit). Par conséquent, un critère de débruitage naturelle peut résulter de cette différence statistique entre les coefficients du signal et le bruit, car il ya déjà une information préalable sur les coefficients qui sont supposés être parcimonieux. Généralement, on possède plus d'information sur les sources recherchées que sur la nature des mélanges observés. Cette méthode présente l'avantage de pouvoir adapter le débruitage post-séparation, opérés individuellement sur chaque signal, à la nature des sources recherchées. L'efficacité de la séparation des sources vibratoires bruitées et les capacités de l'approche proposée basée sur l'amélioration de la parcimonie en utilisant une fonction analysante optimisé sont mises en évidence par une étude expérimentale. Les expériences ont montré la performance de séparation compétitive par rapport à l'application directe de l'ACI sur le signal et les algorithmes de débruitage standard.

6. Bibliographie

- [1] M. Boufenar, S. Rechak et M. Rezig, "Time-frequency analysis techniques review and their application on roller bearings prognostics", *Condition Monitoring of Machinery in Non-Stationary Operations*, Springer, pp. 239-246, 2012.
- [2] K. Javed, "A Robust & Reliable Data-driven Prognostics Approach Based on Extreme Learning Machine and Fuzzy Clustering", Thèse de Doctorat, École doctorale sciences pour l'ingénieur et microtechniques, Université de France – Comté, 2014.
- [3] H. I. Choi et W. J. Williams, "Improved Time-Frequency Representation of multi-components signals using exponential kernels", *IEEE Trans. Acoustic., Speech, Signal Processing ASSP*, Vol. 37, pp. 862-871, 1989.
- [4] B.D. Forrester, "Use of the Wigner-Ville distribution in helicopter fault detection", *Proc. ASSPAS9*, pp. 78-82, 1989.
- [5] P.D. Mcfadden et W. Wang, "time-Frequency Domain Analysis of Vibration Signals for Machinery Diagnostics", Department of Engineering Science, Oxford University, 1990.
- [6] W.J. Wang et P.D. McFadden, "Early Detection of Gear Failure by vibration Analysis, Calculation of the Frequency Distribution", *Mechanical Systems and Signal Processing*, Vol. 7(3), (1993).
- [7] R. Rohrbaugh, "Application of time-Frequency Analysis to Machines Condition Assessment", *Soc. Proceedings of the 27^{iem} Asilomar Conference on Signal, Systems & Computer*, 1993.
- [8] R.A. Rohrbaugh et L. Cohen, "Time-Frequency Analysis of a Cam Operated Pump. Life Extension of Aging Machinery and Structures", *Proc. of the 49^{iem} Meet. of MFPT Soc., Vibration Inst.: Virginia Beach*, Vol. 49 (349), 1995.
- [9] P. Loughlin, J. Pitton et E. Atlas, "Construction of Positive time-Frequency Distribution", *IEEE Trans. Sig. Proc.*, Vol. 42 (2697), 1994.
- [10] L. Atlas, G. Bernard et S. B. Narayanan, "Applications of Time-frequency Analysis to Signal From manufacturing and Machine monitoring Sensors", *Proceedings of the EEE*, Vol. 84(9) 13-19, 1996.
- [11] M. Hongbin, "Application of Wavelet Analysis to Detection of Damages in Rolling Element Bearings", *Proceeding of the International Conference on Structural Dynamics. Vibrations, Noise and Control*, pp. 1334-1339, 1995.
- [12] W. J. Wang et P. D. McFadden, "Application of Wavelets to Early Gear Damaged Detection", *Mechanical Systems and Signal Processing*, 497-507, 1995.
- [13] K. Belaid et A. Miloudi, "Detection of gear defects by resonance demodulation detected by wavelet transform and comparison with the kurtogram", *21ème Congrès Français de Mécanique Bordeaux, France*, 2013.
- [14] G. Dalpiaz, et A. Rivola, "Fault detection and diagnostics in cam mechanisms", *2nd International Symposium on Acoustical and Vibratory Surveillance Methods and Diagnostic Techniques*, pp. 327-339, 1995.
- [15] N. Wiener, "Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications", the technology Press. of the MIT, 2004.

- [16] Y. Ephraïm et D. Malah, “Speech enhancement using a minimum mean-square error log-spectral amplitude estimator”, *IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP.*, Vol. 33(2), pp. 443–445, 1985.
- [17] C. Stein, “Estimation of the mean of a multivariate normal distribution”. *Ann. Statist.*, Vol. 9, pp. 1135–1151, 1982.
- [18] D.L. Donoho, et I.M. Johnstone, “Minimax estimation via wavelet shrinkage”, *The Annals of Statistics*, Vol. 26, pp. 879-921, 1998.
- [19] D.L. Donoho et I.M. Johnstone, “Adapting to unknown smoothness via wavelet shrinkage”, *Journal of the American Statistical Association*, Vol. 90, pp. 1200-1224, 1995.
- [20] D.L. Donoho et I.M. Johnstone, “Ideal spatial adaptation by wavelet shrinkage”, *Biometrika*, pp. 425-455, 1994.
- [21] I. M. Johnstone, et B. W. Silverman, “Empirical Bayes selection of wavelet thresholds”, *The Annals of Statistics*, Vol. 33, pp. 1700–1752, 2005.
- [22] M.H. Neumann, et , R.V. Sachs, “Wavelet thresholding: beyond the gaussian i.i.d. situation”, *Lecture Notes in Statistics, Springer*, Vol. 103, pp. 281–300, 1995.
- [23] A. M. Djafari, “Bayesian approach with prior models which enforce sparsity in signal and image processing”, *Journal on Advances in Signal Processing EURASIP*, Vol. 52, 2012.
- [24] S. G. Mallat, “A theory for multiresolution signal decomposition : The wavelet representation”, *IEEE trans. PAMI*, Vol. 11(7), pp. 674–693, 1989.
- [25] C. Jutten, et J. Herault, “Blind separation of sources, Part I: an adaptive algorithm based on neuromimetic architecture”, *Signal Processing*, Vol. 24(1), pp. 1–10, 1991.
- [26] D. T. Pham, P. Garat et C. Jutten, “Separation of a mixture of independent sources through a maximum likelihood approach”, *In Proc. EUSIPCO*, pp. 771-774, 1992.
- [27] W. Cheng, Z. Zhang, S. Lee et Z. He, “Source contribution evaluation of mechanical vibration signals via enhanced independent component analysis”, *Journal of Manufacturing Science and Engineering*, Vol. 134, 2012.
- [28] M. El Rhabi, H. Fenniri, G. Gelle et G. Delaunay, “Blind separation of rotating machine signals using penalized mutual information and minimal distortion principle”, *Mechanical Systems and Signal Processing*, pp. 1282-1292, 2005.
- [29] R.B. Randall, J. Antoni et S. Chobsaard, “The relationship between spectral correlation and envelope analysis for cyclostationary machine signals, application to ball bearing diagnostics”, *Mechanical Systems and Signal Processing*, Vol. 15(5), pp. 945-962, 2001.
- [30] N. Bouguerriou, M. Haritopoulos, C. Capdessus et L. Allam, “Novel cyclostationarity-based blind source separation algorithm using second order statistical properties: Theory and application to the bearing defect diagnosis”, *Mechanical Systems and Signal Processing*, Vol. 19(6), pp. 1260-1281, 2005.
- [31] A. Paraschiv-Ionescu, C. Jutten, K. Aminian, B. Najafi et Ph. Robert, “Source separation in strong noisy mixtures : a study of wavelet de-noising pre-processing”, *In ICASSP’2002, Orlando, May 2002*.
- [32] D. Gabor, “Theory of communication”, *Journal of the IEE*, Vol. 93, pp.429–457,1946.

- [33] M. V. Wickerhauser, “Adapted Wavelet Analysis from Theory to Software”, AK Peters, Boston, MA, USA, 1994.
- [34] W. Kozek. “Matched Weyl–Heisenberg Expansions of Nonstationary Environments”. PhD thesis, NuHAG, University of Vienna, 1996.
- [35] L. Cohen, “Time-frequency analysis”, Prentice Hall, 1995.
- [36] B. Bulke Hubbard, “Ondes et Ondelettes, la Saga d’un Outil Mathématique, Pour la Science”, Diffusion Belin, Paris, 1995.
- [37] S. Mallat, “A Wavelet Tour of Signal Processing (Second Edition)”, Academic Press, 1998.
- [38] Y. Bentaleb, “Analyse par ondelettes des signaux sismiques : Applications aux ondes de surface”, thèse PhD, Université Mohammed V – Agdal, 2007.
- [39] C. E. Heil et D. F. Walnut, “Continuous and Discrete Wavelet Transforms”, SIAM Rev., Vol. 31, pp. 628-666.
- [40] S. G. Mallat, “A theory for multiresolution signal decomposition : the wavelets representation”, IEEE Trans. Patt. Anal. Mach., Vol.11 (7), pp. 674-693, 1989.
- [41] Y. Meyer, “Ondelettes, fonctions splines et analyse graduée”, In Cahiers du Ceremade, number 8703, Université Paris-Dauphine, 1987.
- [42] D. Esteban et C. Galand, “Application of quadrature mirror filters to split-band voice coding schemes”, Proc. IEEE Int. Conf. Acoust. Signal Speech Process, pp. 191-195, Hartford, USA, 1977.
- [43] J. Feauveau, ‘Filtres miroirs conjugués : une théorie pour les filtres miroirs en quadrature et l’analyse multirésolution en ondelettes’, Comptes-Rendus de l’Académie des Sciences de Paris, Théorie des signaux, Vol. 309, pp.853-856, 1989.
- [44] P. Burt et E. Adelson, ‘The Laplacian pyramid as a compact image code’, IEEE Transactions on Communications, Vol. 31(4), pp.532-540, 1983.
- [45] A. Cohen, ‘Ondelettes et traitement numérique du signal’, Thèse de doctorat, Université Paris IX, 1992.
- [46] S. Mallat, “Multiresolution approximation and wavelet orthonormal bases of L^2 ”, Trans. Amer. Math. Soc., Vol. 315, pp. 69-87, 1989.
- [47] D. Esteban et C. Galand, “Application of quadrature mirror filters to split-band voice coding schemes”, Proc. IEEE Int. Conf. Acoust. Signal Speech Process, pp. 191-195. Hartford, CT, USA, 1977.
- [48] J. Feauveau, ‘Filtres miroirs conjugués : une théorie pour les filtres miroirs en quadrature et l’analyse multirésolution en ondelettes’, Comptes-Rendus de l’Académie des Sciences de Paris, Théorie des signaux, Vol. 309, pp.853-856, 1989.
- [49] I. Daubechies, “Orthonormal bases of compactly supported wavelets”, Communication on Pure and App. Math., Vol. 41, pp.909-996, 1988.
- [50] D.L. Donoho, “Compressed sensing”, IEEE Transactions on Information Theory, Vol. 52(4), pp. 1289–1306, 2006.
- [51] D.L. Donoho et X. Huo, “Uncertainty principles and ideal atomic decomposition”, IEEE Transactions on Information Theory, Vol. 47(7), pp. 2845–2862, 2001.
- [52] S. Chen, L. Donoho et D. A. Saunders, “Atomic decomposition by basis pursuit”, SIAM Journal on Scientific Computing, Vol. 20(1), pp. 33–61, 1998.

- [53] D. Donoho, I. Johnstone, J. Hoch et A. Stern, “Maximum entropy and the nearly black object”, *J. R. Stat. Soc. B*, Vol. 54(1), pp.41–81, 1992.
- [54] D. Donoho et B. F. Logan, “Signal recovery and the large sieve”, *SIAM J. Appl. Math*, Vol. 52, pp. 577–591, 1992.
- [55] A.M. Bruckstein, D.L. Donoho et M. Elad, “From sparse solutions of systems of equations to sparse modeling of signals and images”, *SIAM Review*, Vol. 51(1), pp. 34–81, 2009.
- [56] S. Mallat et Z. Zhang, “Matching pursuits with time-frequency dictionaries”, *IEEE Transactions on Signal Processing*, Vol. 41(12), pp. 3397–3415, 1993.
- [57] J.H. Friedman et W. Stuetzle, “Projection pursuit regression”, *J. Amer. Statist. Asso.*, vol. 76, pp. 817–823, 1981.
- [58] R.A. DeVore et V.N. Temlyakov, “Some remarks on greedy algorithms”, *Advances in Computational Mathematics*, Vol. 5(2-3), pp. 173–187, 1996.
- [59] Y.C. Pati, R. Rezaifar et P.S. Krishnaprasad, “orthogonal Matching Pursuit”, *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*, 1993.
- [60] S. Chen, “Basis Pursuit”, PhD thesis, Stanford University, Nov. 1995.
- [61] M. Boufenar, S. Rechak, “Vibration impulses denoising based on a Bayesian approach and signal sparsity enhancement”, *The Mediterranean Journal of Measurement and Control*, Vol. 10 (3), pp. 273-283, 2014.
- [62] R. R. Coifman et D. L. Donoho, “Translation-invariant denoising in *Lecture Notes in Statistics : Wavelets and Statistics*”, New York : Springer-Verlag, pp. 125–150, 1995.
- [63] W. X. Yang et X. M. Ren, “Detecting impulses in mechanical signals by wavelets”, *EURASIP Journal on Applied Signal Processing*, Vol. 8, 2004, pp. 1156-1162.
- [64] J. Lardies et S. Gouttebroze, “Identification of modal parameters using the wavelet transforms”, *International Journal of Mechanical Sciences*, Vol. 44, 2002, pp. 2263–2283.
- [65] J. D. Wu et J. C. Chen, “Continuous wavelet transform technique for fault signal diagnosis of internal combustion engines”, *NDT&E International Journal*, Vol. 39, 2006, pp. 304-311.
- [66] Y. Sheen et C. K. Hung, “Constructing a wavelet-based envelope function for vibration signal analysis”, *Journal of Mechanical Systems and Signal Processing*, Vol. 18, 2004, pp. 119-126.
- [67] H. Qiu, J. Lee, J. Lin et G. Yu, “Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics”, *Journal of sound and vibration*, Vol. 289, 2006, pp. 1066-1090
- [68] M. A. Jafarizadeh, R. Hassannejad, M. M. Etefagh et S. Chitsaz, “Asynchronous input gear damage diagnosis using time averaging and wavelet filtering, *Journal of Mechanical Systems and Signal Processing*, Vol. 22, 2008, pp. 172-201.
- [69] M. Kowalski, K. Siedenburger et M. Dörfler, “Une approche par la synthèse et la parcimonie de quelques problèmes inverses en audio”, *IEEE transactions on signal processing*, Vol. 61(10), pp. 2498–2511, 2014.
- [70] M. Kowalski, “approximation des signaux : approches variationnelles et modèles aléatoires”, Thèse de Doctorat, université de Provence U.F.R. M.I.M., 2008.

- [71] M. Aharon, M. Elad et M. Bruckstein, “An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation”, *IEEE Transactions on Signal Processing*, Vol. 54 (11), pp. 4311-4322, 2006.
- [72] E. I. Green et B. W. Silverman, “Nonparametric Regression and Generalised Linear Models”, London : Chapman & Hall, 1994.
- [73] R. L. Eubank, “Nonparametric Regression and Spline Smoothing”, New York : Marcel dekker, 2nd edition. (1999).
- [74] W.Härdle, G. Kerkycharian, D. Pickard et A.Tsybakov, “Wavelets, approximation and statistical applications”, *Lecture Note en Statistics* 129, 1998.
- [75] J. S. Marron, S. Adak, I. M. Johnstone, M. H. Neumann et P. Patil, “Exact risk analysis of wavelet regression”, *Journal of Computational and Graphical Statistics*, Vol. 7(3), pp. 278–309, 1998.
- [76] A. G. Bruce et H. Y. Gao, “Understanding waveshrink : variance and bias estimation. *Biometrika*, Vol. 83, pp. 727–745, 1996.
- [77] A. Antoniadis, J. Bigot et T. Sapatinas, “Wavelet estimators in nonparametric regression: A comparative simulation study”, *Journal of Statistical Software*, Vol. 6(6), pp. 1–83, 2001.
- [78] F. Abramovich, Y. Benjamini, D. Donoho et I. Johnstone, “Adapting to unknown sparsity by controlling the false discovery rate”, *Computational Statistics & Data Analysis*, Vol. 22, 351–361, 2000.
- [79] D. Donoho, I. M. Johnstone, G. Kerkycharian et D. Picard, “Wavelet shrinkage : Asymptopia”, *J. R. Statist. Soc. B.*, 57(2), 301–337(1995).
- [80] F. Abramovich et Y. Benjamini, “Thresholding of wavelet coefficients as multiple hypotheses testing procedure”, In Antoniadis, A. and Oppenheim, G., editors, *Wavelets and Statistics*, vol. 103 of *Lecture Notes in Statistics*, pp. 5–14, 1995.
- [81] F. Abramovich et Y. Benjamini, “Adaptive thresholding of wavelet coefficients. *Computational Statistics & Data Analysis*”, Vol. 22, pp. 351–361, 1996.
- [82] C. Stein, “Estimation of the mean of a multivariate normal distribution”, *Annals of Statistics*, Vol. 9, pp. 1135–1151, 1982.
- [83] E. P. Simoncelli et E. H. Adelson, “Noise removal via Bayesian wavelet coding”, *IEEE Sig Proc Society*, In *Third Int’l Conf on Image Proc*, Vol.1, pp. 379–382, 1996.
- [84] F. Abramovich, T. Sapatinas et B. Silverman, “Wavelet thresholding via a bayesian approach”, *J. R. Statist. Soc. B*, 60, pp. 725–749, 1998.
- [85] I. Johnstone et B. Silverman, “Empirical Bayes approaches to mixture problems and wavelet regression”, *Technical report*, Department of Mathematics, University of Bristol, UK, 1998.
- [86] P. Moulin et J. Liu, “Analysis of multiresolution image denoising schemes using generalized gaussian and complexity priors”. *IEEE Transactions on Information Theory*, Vol. 45(3), pp. 909–919, 1999.
- [87] A. M. Djafari, “Bayesian approach with prior models which enforce sparsity in signal and image processing”, *EURASIP Journal on Advances in Signal Processing*, 2012:52, 2012.

- [88] A. Hyvärinen, “Sparse Code Shrinkage: De-noising of Non-Gaussian Data by Maximum Likelihood estimation”, *Neural Computation*, pp. 1739–1768, 1999.
- [89] A. Hyvärinen, P. Hoyer et E. Oja, “Sparse Code Shrinkage: De-nosing by Nonlinear Maximum Likelihood Estimation,” Helsinki University of Technology, Laboratory of Computer and Information Science. 1998.
- [90] C. Jutten et J. Herault, “Blind separation of sources, part I : An adaptive algorithm based on neuromimetic architecture”, *Signal Processing*, Vol. 24 (1), pp. 1-10, 1991.
- [91] P. Comon, “Independent Component Analysis, a New Concept” *Signal Processing*, Vol. 36 (3), pp. 287-314, 1994.
- [92] R. Tong et Y. Soon, “Indeterminacy and identifiability of blind identification”, *IEEE Trans. CS*, Vol. 38, pp. 499-509, 1991.
- [93] C. Jutten, “Calcul neuromimétique et traitement du signal. analyse en composantes indépendantes”. Thèse de Doctorat d’Etat, INPG, Grenoble, France, 1987.
- [94] J. F. Cardoso, “The three easy routes to independent component analysis, contrasts and geometry”, In *Proceedings of ICA*, pp. 1–6, San Diego, USA, 2001.
- [95] A. Hyvärinen, et E. Oja, “Independent component analysis: algorithms and applications”, *Neural Networks*, 13(4–5), 411–430, (2000).
- [96] A. K. Barros, J. Carlos et D. Erdogmus, “Independent component analysis and blind source separation”, *Signal Processing*, Vol. 87(8), 1817–1818, (2007).
- [97] A. J. Bell et T. J. Sejnowski, “An information-maximization approach to blind separation and blind deconvolution”, *Neural Computation*, Vol. 7(6), pp. 1129-1159, 1995.
- [98] A. Hyvärinen et E. A. Oja, “fast fixed-point algorithm for independent component analysis”, *Neural Computation*, Vol. 9(7), 1483–1492, (1997).
- [99] A. Hyvärinen, “Fast and robust fixed-point algorithms for independent component analysis”, *IEEE Trans. Neural Network*, Vol. 10(3), 626–634, (1999).
- [100] M. Gaeta et J. L. Lacoume, “Estimateurs du maximum de vraisemblance étendus à la séparation de sources non gaussiennes”, *Traitement du Signal*, Vol. 7(5), pp. 419-434, 1990.
- [101] M. Gaeta et J. L. Lacoume, “Source separation without a priori knowledge : the maximum likelihood solution”, In *EUSIPCO*, pp. 621-624, 1990.
- [102] A. Tikhonov, “Solution of incorrectly formulated problems and the regularization method”, In *Soviet Math. Dokl.*, Vol. 5, pp. 1624–1627, 1963.
- [103] I. Daubechies et M. Defrise, “An iterative thresholding algorithm for linear inverse problems with a sparsity constraint”. *Comm Pure Appl Math*. Vol. 57, pp. 1413–1457, 2004.
- [104] P. Grady, A. Pearlmutter et T. Rickard, “Survey of sparse and non-sparse methods in source separation”, *International Journal of Imaging Systems and Technology*, Vol. 15(1), pp. 18-33, 2005.
- [105] P. Georgiev, F. Theis et A. Cichocki, “Sparse component analysis and blind source separation of underdetermined mixtures”, *IEEE Transactions on Neural Networks*, Vol. 16(4), pp. 992-996, 2005.

- [106] S.J. Roberts, "Independent component analysis : Source assessment and separation, a Bayesian approach", IEEE Proceedings on Vision, Image and Signal Processing, Vol. 145(3), pp. 149–154, 1998.
- [107] K. Knuth, "Bayesian source separation and localization". In A. Mohamm-Djafari, éditeur : proceedings of International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MaxEnt'98). American Institute of Physics, 1998.
- [108] K. Knuth, "A Bayesian approach to source separation. In J. Cardoso, C. Jutten et P. Loubaton, éditeurs : Proceedings of International Workshop on Independent Component Analysis and Signal Separation (ICA'99), pp. 283–288, France, 1999.
- [109] A. Mohammad-Djafari, "A Bayesian approach to source separation", In proceedings of International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MaxEnt'99), Vol. 567, pp. 221–244, 1999.
- [110] C. Févotte, S.J. Godsill et P.J. Wolfe, "Bayesian approach for blind separation of underdetermined mixtures of sparse sources", In proceedings of 5th International Conference on Independent Component Analysis and Blind Signal Separation, 2004.
- [111] A. Doucet, S.J. Godzill et C.P. Robert, "Marginal maximum a posteriori estimation using Markov chain Monte Carlo", Statistics and Computing, Vol. 12, pp. 77–84, 2002.
- [112] N. Polson et J. Scott, "Shrink globally, act locally: sparse Bayesian regularization and prediction", Bayesian Stat. Vol. (9), 1–24 (2010).
- [113] A. Mohammad-Djafari, "On the estimation of hyperparameters in Bayesian approach of solving inverse problems", in Proc IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP-93, vol. 5, pp. 495–498.
- [114] A Mohammad-Djafari, "Joint estimation of parameters and hyperparameters in a Bayesian approach of solving inverse problems", in IEEE Int Conf on Image Processing (ICIP), IEEE ICIP 96, vol. II., pp. 473–477, 1996.
- [115] B. Rivet, V. Vigneron, A. Paraschiv-Ionescu et C. Jutten, "Independent Component Analysis and Blind Signal Separation", volume LNCS 3195 of Lectures Notes in Computer Science, chapter Wavelet de-noising for Blind Source Separation in Noisy Mixtures, pp. 263–271. Springer, 2004.
- [116] A. Paraschiv-Ionescu, C. Jutten, K. Aminian, B. Najafi et P. Robert, "Source separation in strong noisy mixtures: a study of wavelet de-noising pre-processing", In ICASSP'2002, Orlando, May 2002.
- [117] J. H. Friedman et J. W. Tukey, "A Projection Pursuit Algorithm for Exploratory Data Analysis", IEEE Transactions on Computers, C-23, pp. 881-890, 1974.
- [118] A. Hyvärinen, "New approximations of differential entropy for independent component analysis and projection pursuit", Advances in Neural Information Processing Systems, Vol. 10, pp. 273-279. MIT Press, 1998.