



HAL
open science

Modélisation de signaux fortement non stationnaires à phase et à amplitude locales polynomiales.

Meryem Jabloun

► **To cite this version:**

Meryem Jabloun. Modélisation de signaux fortement non stationnaires à phase et à amplitude locales polynomiales.. Traitement du signal et de l'image [eess.SP]. Institut National Polytechnique de Grenoble - INPG, 2007. Français. NNT: . tel-00176079

HAL Id: tel-00176079

<https://theses.hal.science/tel-00176079>

Submitted on 2 Oct 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE

N° attribué par la bibliothèque

THESE

pour obtenir le grade de

DOCTEUR DE L'INPG

Spécialité : « Signal, Image, Parole et Télécoms »

préparée au **GIPSA-lab** Laboratoire de l'image, de la parole, du signal et de
l'automatique de Grenoble : Département des Images et du Signal

dans le cadre de l'**Ecole Doctorale « Electronique, Electrotechnique,
Automatique, Télécommunications et Signal »**

présentée et soutenue publiquement

par

Meryem JABLOUN

10 Juillet 2007

Titre :

**Modélisation de signaux fortement non stationnaires à phase et à amplitude
locales polynomiales.**

Directeurs de thèse : Nadine Martin et Michelle Vieira

JURY

Madame	S. Gentil	Présidente
Monsieur	P. Larzabal	Rapporteur
Monsieur	A. Ferrari	Rapporteur
Monsieur	M. Davy	Examineur
Monsieur	S. Stankovic	Examineur
Madame	N. Martin	Directrice de thèse
Madame	M. Vieira	Co-directrice de thèse

«Il en vient de partout, des spécialistes! On en fabrique en série au Japon! J'ai vu le monde devenir en moins de quelques ans une véritable pétaudière de publications universelles et saugrenues sur ce même sujet rabâché. »

Céline, Voyage au bout de la nuit.

Remerciements

Je tiens tout d'abord à remercier mes deux directrices de thèse Madame Nadine Martin et Madame Michelle Vieira pour leur compétence, leur disponibilité, leur écoute et la qualité de leur encadrement en général. Elles ont su me conseiller et me guider aux moments clefs de ce travail, tout en me laissant la liberté nécessaire à la réalisation de ce projet de longue haleine. Elles ont également su me remonter le moral à certains moments difficiles et m'ont permis de surpasser ma réserve naturelle et de prendre confiance en moi-même.

Je remercie également Monsieur François Léonard, ingénieur à l'Institut de Recherche d'Hydro-Québec avec qui nous avons entretenu une collaboration régulière et enrichissante et qui a été à l'initiative de l'idée de mon sujet de recherche. Je le remercie également pour sa compétence, ses remarques constructives et son amabilité.

Je remercie également l'ensemble des personnes qui m'ont fait l'honneur d'être membres du jury. Je pense tout d'abord à Madame Sylvianne Gentil, qui l'a présidé. Toute ma considération va également aux membres qui ont dû venir de très loin, pour assister à ma soutenance. Je tiens ainsi à remercier les deux rapporteurs, Monsieur Pascal Larzabal et Monsieur André Ferrari, qui ont utilisé leurs grandes compétences au service de l'évaluation de mon travail. Je les remercie pour avoir lu attentivement mon manuscrit et d'en avoir fait une critique très fine et constructive. Ma connaissance va également aux examinateurs, Monsieur Manuel et Monsieur Srdjan Sankovic, qui ont utilisé toutes leurs connaissances dans l'examen attentif de mon travail.

Je tiens également à remercier Monsieur Jean-Marc Chassery, directeur du GIPSA-lab (département DIS), pour m'avoir accueilli dans son laboratoire et conseillé aux moments clés.

Je n'oublie certainement pas tous ceux sans qui mes années de thèse ne seraient aussi bien pas passées. Je pense en particulier aux thésards et anciens thésards avec qui j'ai partagé des bons moments. J'ai une pensée particulière à Frédéric pour m'avoir supporté en tant que co-bureau. Je remercie Cyril pour son amitié. Je remercie Ludovic, Nutapol et Gaël pour leur humour, leur sympathie et pour avoir organisé mon pot de thèse. Je remercie également Fabien, Caroline, Mathieu (s), Grégoire, Moussa, Benoît, Cédric,

Mohammed, Ion, Murtaza, Damien, Pierre, Julien et mes voisins Zilling et Pascal, ... Mes remerciements vont également à ceux avec qui j'ai effectué mes enseignements.

Je remercie enfin mes parents, ma soeur Zouhour et mon frère Karim pour leur soutien sans faille et ce jusqu'à ma soutenance pour laquelle ils ont fait le déplacement depuis la Tunisie.

Je demande à ceux qui j'ai pu oublier dans cette page de bien vouloir m'excuser, mais ils peuvent être sûrs qu'ils sont dans mes pensées.

Table des matières

1	Etat de l'art	13
1.1	Introduction	15
1.2	Représentations temps-fréquence	15
1.2.1	Spectrogramme	16
1.2.2	Distributions de Wigner-Ville et de Wigner-Ville polynomiale	19
1.2.3	Représentations temps-fréquence - Opérateur " <i>warping</i> "	22
1.2.4	Méthode S	23
1.3	Techniques basées sur la modélisation de la phase du signal	24
1.3.1	Techniques basées sur la fonction d'ambiguïté d'ordre supérieur	25
1.3.2	Estimation par la méthode des moindres carrés non-linéaires instantanés	33
1.3.3	Maximum de vraisemblance et échantillonnage pondéré	35
1.4	Autres techniques	38
1.5	Conclusion	39
2	Signal mono-composante non-stationnaire court - Modélisation	41
2.1	Introduction	43
2.2	Modèle du signal	44
2.2.1	Définition du modèle du signal - Unicité	44
2.2.2	Modèle à amplitude et à fréquence polynomiales	45
2.3	Exemples de bases polynomiales	48
2.3.1	Base polynomiale canonique	48
2.3.2	Polynômes de Legendre	48
2.3.3	Base polynomiale orthonormale en temps discret	49
2.4	Méthodes d'estimation	51
2.5	Conclusion	53
3	Signal court mono-composante - Maximum de vraisemblance	55
3.1	Introduction	57
3.2	Estimation par maximum de vraisemblance	57
3.2.1	Définitions et propriétés	57
3.2.2	Lien avec l'estimation par moindres carrés	58
3.2.3	Maximum de vraisemblance compressée	59

3.2.4	Coefficient de corrélation	60
3.3	Optimisation stochastique	61
3.3.1	Méta-heuristiques	63
3.3.2	Recuit simulé - Principe	64
3.4	Recuit simulé - Configuration	65
3.4.1	Initialisation	65
3.4.2	Espace de recherche	67
3.4.3	Choix de la fonction objectif	67
3.4.4	Échantillonnage de la fonction objectif	68
3.4.5	Règle d'acceptation	69
3.4.6	Réglage de la température	71
3.4.7	Choix du rayon du voisinage	72
3.4.8	Conditions d'arrêt de l'algorithme	74
3.4.9	Convergence du recuit simulé	75
3.5	Algorithme du recuit simulé	76
3.6	Conclusions	77
4	Signal court mono-composante - Approche Bayésienne	79
4.1	Introduction	81
4.2	Inférence Bayésienne	82
4.2.1	Généralités	82
4.2.2	Choix d'une loi <i>a priori</i>	83
4.2.3	Estimateurs Bayésiens	85
4.3	Méthodes de Monte Carlo par Chaînes de Markov	87
4.3.1	Introduction	87
4.3.2	Chaînes de Markov : de la théorie aux méthodes MCMC	88
4.3.3	Choix de l'algorithme de simulations MCMC	89
4.4	Algorithme de Metropolis Hastings	89
4.4.1	Schéma itératif d'un algorithme Metropolis Hastings	90
4.4.2	Conditions sur la loi instrumentale	91
4.4.3	Convergence des algorithmes Metropolis-Hastings	92
4.5	Quelques algorithmes Metropolis-Hastings	93
4.5.1	Algorithme Metropolis Hastings à marche aléatoire	93
4.5.2	Algorithme Metropolis Hastings dit une-variable-à-la-fois	94
4.6	Estimation des signaux courts mono-composantes non-stationnaires	95
4.6.1	Positionnement	95
4.6.2	Problématique et inférence Bayésienne	96
4.6.3	Algorithmes Metropolis Hastings une-variable-à-la-fois et à marche aléatoire	98
4.7	Conclusions	100

5	Signal court mono-composante - Analyse et performance	101
5.1	Introduction	103
5.2	Recuit simulé - Réglage	103
5.2.1	Différents jeux de paramètres et compromis	104
5.2.2	Résultats expérimentaux et discussions	106
5.2.3	Efficacité du réglage retenu Reg_0	109
5.3	Performances statistiques du recuit simulé en présence du bruit	110
5.3.1	Bornes de Cramer Rao	110
5.3.2	Signaux à amplitude constante	115
5.3.3	Signaux à modulations d'amplitude et de fréquence non-linéaires	117
5.4	Algorithme MH-MCMC et comparaison avec le recuit simulé	124
5.4.1	Réglage de l'algorithme de MH-MCMC	124
5.4.2	Comparaison MH-MCMC, recuit simulé et HAF	125
5.5	Conclusion	127
6	Signal long non-stationnaire - Approches et analyses	129
6.1	Introduction	131
6.2	Problématique	132
6.3	Segmentation d'un signal non-stationnaire long	133
6.3.1	Segmentation basée sur une Distribution Temps-Fréquence	134
6.3.2	Prélèvement des segments	135
6.3.3	Ajustement de la longueur d'un segment	138
6.4	Processus de fusion des segments	141
6.4.1	Recouvrement unilatéral	142
6.4.2	Recouvrement bilatéral	143
6.4.3	Choix des fenêtres de fusion	143
6.4.4	Arrêt de la fusion	144
6.5	Synthèse de l'algorithme pour estimer des signaux longs non-stationnaires	144
6.6	Analyse des performances	146
6.6.1	Bornes de Cramer-Rao	146
6.6.2	Estimation des signaux à phase et amplitude polynomiales	151
6.6.3	Estimation des signaux ayant un modèle sinusoïdal	155
6.6.4	Application sur un signal réel	158
6.7	Conclusion - Avantages et limitations	160
7	Signal court à multi-composantes non-stationnaires	163
7.1	Introduction	165
7.2	Modélisation et estimation	165
7.2.1	Modélisation	166
7.2.2	Méthodes d'estimation	167
7.3	Etude des performances des algorithmes proposés	170
7.3.1	Bornes de Cramer Rao	170
7.3.2	Application sur des signaux de simulations	171

7.3.3	Application sur des signaux réels	177
7.4	Conclusion	179
Annexes		191
A		193
A.1	Classe de Cohen	193
A.2	Signal réel - Signal analytique	194
A.3	Exemples de modèles du signal	195
A.4	Expressions des polynômes de Legendre Standards	195
A.5	Complément sur le calcul de la base de polynômes discrète et orthonormale	195
A.6	Estimation des paramètres de l'amplitude	196
A.7	Techniques d'optimisation classiques	196
A.8	Marginalisation de la densité <i>a posteriori</i>	197
B		199
B.1	Signaux courts non-stationnaires - Compléments	199

Notations

\mathbb{R}	Ensemble des réels
\mathbb{C}	Ensemble des complexes
\mathbb{Z}	Ensemble des entiers relatifs
\mathbb{N}	Ensemble des entiers naturels
\mathbb{E}	Espérance mathématique
\mathbb{I}	Matrice identité
\cdot^*	Conjugué d'une matrice ou d'un vecteur
\cdot'	Dérivée d'une fonction
$ \cdot $	Module
\langle, \rangle	Produit scalaire continu
$\langle, \rangle_{\mathcal{D}}$	Produit scalaire discret
$e^{\{\cdot\}}$	Fonction exponentielle
j	Le nombre complexe $j^2 = -1$
\top	Transposée d'un vecteur ou d'une matrice
\dagger	Transposée conjuguée d'un vecteur ou d'une matrice
$\delta_{m,h}$	Symbole de Kronecker
C_m^l	Coefficient binomial
$\Re(\cdot)$	Partie réelle d'un nombre complexe
$\Im(\cdot)$	Partie imaginaire d'un nombre complexe
\arg	Argument ou phase
\max	Maximum
\min	Minimum
\tan	Tangente
\det	Déterminant d'une matrice

\mathcal{F}_y	Transformée de Fourier à court terme du signal y
N_{TF}	Taille de la Transformé de Fourier Discrète
F_e	Fréquence d'échantillonnage
$RTF(t, \nu; h)$	Représentation temps t et fréquence ν associée à la fonction de paramétrisation h
\mathcal{S}_y	Spectrogramme du signal y
\mathcal{WV}_y	Distribution de Wigner-Ville du signal y
\mathcal{HAF}_M	Fonction d'ambiguïté d'ordre supérieur HAF
\mathcal{MOS}	Moment instantanée d'ordre Supérieur
ℓ_{MV}	Fonction de vraisemblance
\bullet	Multiplication terme à terme
\mathcal{P}_m	Polynôme de degré m
$Prob$	Probabilité
$\mathcal{P}(\boldsymbol{\theta} \mathbf{y})$	Loi de probabilité <i>a posteriori</i>
$\mathcal{P}(\boldsymbol{\theta})$	Loi de probabilité <i>a priori</i>
$\cdot \sim \mathcal{P}$	Simulé ou généré selon la loi \mathcal{P}
\propto	Proportionnel
$\mathbf{I}_{\boldsymbol{\theta}}$	Matrice d'information de Fisher
dB	Décibels
$Ent[.]$	Partie entière

Liste des abréviations

RTF	Représentation(s) Temps-Fréquence
TF	Transformée de Fourier
TFD	Transformée de Fourier Discrète
DSP	Densité Spectrale de Puissance
HAF	Higher-order Ambiguity Function
MCNI	Méthode des Moindres Carrés Non-linéaires instantanée
RSB	Rapport Signal à Bruit
BCR	Borne de Cramer-Rao
EQM	Erreur(s) quadratique(s) Moyenne(s)
MMSE	Erreur quadratique moyenne minimale
MAP	Maximum <i>a posteriori</i>
MV	Maximum de vraisemblance
MCMC	Monte-Carlo par chaînes de Markov
MH	Metropolis-Hastings

Introduction

Depuis des décennies, le traitement du signal est une discipline qui ne cesse d'évoluer sous tous ses différents axes. Son développement a été accompagné d'une part par l'explosion des capacités des ordinateurs et des processeurs en termes de mémoire et de vitesse de calcul et d'autre part par le perfectionnement du matériel de mesures (le radar, le sonar, les capteurs d'acquisition, *etc*). Seulement, les signaux mesurés par ces appareils demeurent toujours sensibles aux problèmes classiques (la présence du bruit aléatoire, les distortions, les trajets multiples, *etc*). De ce fait, des nouvelles techniques du traitement numérique du signal poursuivent l'amélioration de la qualité et l'augmentation de la quantité d'informations utiles pouvant être extraites du signal étudié.

Le signal peut être d'origine très variée (bio-acoustique, acoustique sous-marine, électrique, mécanique, *etc*). L'évolution temporelle de quelques signaux réels, qui feront l'objet d'une application dans la présente thèse, est représentée sur la figure 1. C'est à partir de ces signaux temporels que nous cherchons à nous informer sur la constitution et le contenu fréquentiel du signal et à caractériser les éventuelles non-stationnarités.

Par le passé, un très grand nombre de travaux était consacré à l'analyse de processus stationnaires. C'est une modélisation du signal sous la forme d'un processus aléatoire dont la fonction de répartition est invariante par translation temporelle. Le signal observé représente alors une réalisation de ce processus. Théoriquement, la stationnarité stricte étant difficile à vérifier, on considère plutôt la stationnarité jusqu'à l'ordre k , c'est-à-dire que les moments et les cumulants jusqu'à l'ordre k inclus sont invariants par translation temporelle [BLP81, LAC97].

Par exemple, un processus dont la moyenne statistique ne dépend pas du temps est dit stationnaire à l'ordre 1. Si sa fonction de corrélation ne dépend que de l'écart entre deux instants de mesure, le processus est dit stationnaire à l'ordre deux. La transformée de Fourier de cette fonction de corrélation n'est rien d'autre que la densité spectrale de puissance, qui constitue un outil classique d'analyse spectrale à l'ordre 2 [Tre76, Kay88]. Pour l'estimer, il existe deux types de méthodes [Kay88] : les méthodes paramétriques (modèle AR, MA et ARMA *Auto-Regressive Moving Average*) et les méthodes non paramétriques (périodogramme classique, moyenné et Blackman-Tukey, estimateur spectral de Capon, *etc*). Par ailleurs, d'autres outils d'analyse spectrale, basés sur les statistiques d'ordre supérieur (le kurtosis spectral par exemple), ont été également développés [Vra03].

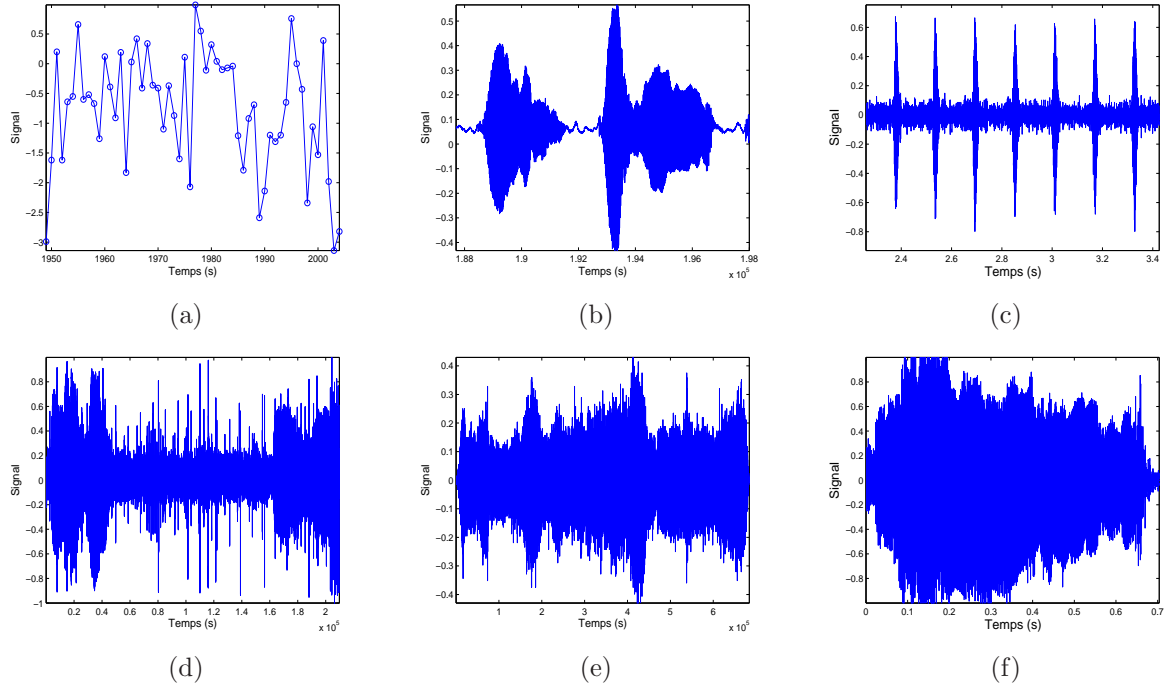


FIG. 1 – Exemples de signaux réels : (a) bilan de masse de glaciers de Sarennes, $Fe = 1 \text{ année}^{-1}$ [Mar77, VVR95, Vin01, GTC05, Des05], (b) chant d’un canari, $Fe = 44.1 \text{ kHz}$, [Ele04, Leo00], (c) son émis par un long vison, $Fe = 11.025 \text{ kHz}$, (d) sifflement d’une baleine (*LongPilotWhale*), $Fe = 44.1 \text{ kHz}$, (e) son émis par un dauphin, $Fe = 44.1 \text{ kHz}$ et (f) son émis par une vache, $Fe = 44.1 \text{ kHz}$.

Cependant, la plupart des signaux réels (cf. figure 1) sont non-stationnaires et une part importante de l’information qu’ils véhiculent est portée par les non-stationnarités. Cette information est souvent nécessaire voire indispensable à la compréhension du phénomène sous-jacent.

Par exemple dans [Gre68, CS95, SFD98, Ele04, ML05], l’évolution temporelle du contenu fréquentiel d’un chant d’oiseau renseigne sur le mécanisme de sa production et permet l’élaboration d’un modèle mécanique pour le syrinx. Le syrinx est un organe vocal spécifique aux oiseaux permettant une modulation non-linéaire du chant. Un autre exemple est le signal représentant le bilan de masses de glaciers [Mar77, VVR95, Vin01, Des05]. L’analyse de ce signal permet non seulement de surveiller l’évolution de la réserve d’eau douce sur la terre, mais aussi d’étudier l’influence du cycle solaire sur cette évolution. Elle permet également d’étudier l’influence de certains changements climatiques tels que l’échauffement de la planète dû à l’effet de serre. Un dernier exemple témoignant de l’importance des informations véhiculées par les non-stationnarités est celui de la caractérisation du mouvement d’une cible *i.e.* son accélération, sa vitesse et son déplacement dans l’analyse de signaux radar [GDL94, VB00, LL05]. Ces informations sont présentes essentiellement dans la distorsion de la phase du signal due à l’écho (effet Doppler).

Pour analyser ces signaux non-stationnaires, des techniques qui tiennent compte de la variation du contenu fréquentiel du signal au cours du temps sont apparues. Citons

par exemple les représentations temps-fréquence telles que le spectrogramme classique, lissé et ré-alloué, la distribution de Wigner-Ville et les classes de Cohen de façon générale [Coh89, Fla93]. Bien que ces techniques dites non-paramétriques, car ne supposant pas de modèle *a priori* pour le signal, permettent de compléter l'analyse classique de Fourier, elles se trouvent souvent limitées

- soit par la résolution temps-fréquence (principe d'incertitude connu également sous le nom de principe de Heisenberg-Gabor),
- soit par l'apparition d'interférences, souvent dépourvues de sens physique, en présence d'une modulation non-linéaire ou de plusieurs composantes dans le signal.

Plus récemment, des méthodes paramétriques adaptées aux signaux non stationnaires modulés non linéairement en fréquence ou/et en amplitude ont été proposées dans [PPF93, FF95, ZGS96, BSG98, FP99, IZ01, ICQ04]. Le modèle du signal étudié, jouant le rôle d'information *a priori*, comporte une phase et une amplitude polynomiales. Les paramètres du modèle sont estimés en utilisant une transformation appelée la fonction d'ambiguïté d'ordre supérieur (*Higher Ambiguity Function*) [PP91]. Ces méthodes sont non seulement sous optimales au sens du maximum de vraisemblance mais nécessitent aussi des approximations d'ordres élevés en présence d'une forte modulation. Elles échouent complètement dans le cas d'un faible rapport signal à bruit (le rapport de la puissance du signal et du bruit).

Dans le contexte de signaux fortement non-stationnaires, à bande spectrale étroite et modulés à la fois en amplitude et en fréquence, l'amélioration des performances en suivi temporel et en résolution fréquentielle est donc un véritable challenge. La clef en est l'adaptation du modèle aux non-stationnarités du signal et la finesse de la description du contenu fréquentiel. Dans cet objectif, nous développons une nouvelle approche basée sur la modélisation locale de l'amplitude et de la fréquence du signal et dont Monsieur François Léonard était à l'initiative de cette idée.

Nous proposons de segmenter le signal temporel sur des durées relativement courtes et adaptées à une modélisation locale efficace, par exemple, un modèle local polynomial de faible degré. Si la modélisation polynomiale de la totalité de la phase instantanée d'un signal a déjà été traitée dans la littérature, l'approximation polynomiale locale à la fois de l'amplitude et de la fréquence de chaque composante du signal est une nouveauté proposée dans cette thèse. Cette approche locale va non seulement permettre une meilleure adaptabilité du modèle à plusieurs types de modulations mais va aussi améliorer nettement les performances d'estimation.

Nous proposons d'estimer les paramètres du chaque modèle local soit par une procédure de maximum de vraisemblance, résolue par une technique de recuit simulé, soit par une estimation bayésienne. Une fois tous les paramètres de modèles locaux déterminés, tous les segments estimés sont fusionnés conduisant ainsi à la reconstruction du signal dans la totalité de sa durée.

Cette thèse est organisée en sept parties. Nous présentons tout d'abord, dans le chapitre

1, l'état de l'art des méthodes et des algorithmes déjà proposés pour traiter les signaux non-stationnaires. Nous rappelons leurs limitations et les différentes améliorations proposées dans la littérature. Dans les parties suivantes, nous proposons une alternative à ces méthodes, qui permet d'obtenir de bien meilleurs résultats et nous illustrons le potentiel de notre approche à s'adapter à plusieurs types de non-stationnarités.

Le chapitre 2 est donc dédié à la description d'un modèle adapté à un signal court non-stationnaire ou à un segment court extrait d'un signal mono-composante long. Nous nous intéressons particulièrement à la modélisation polynomiale et dérivons une base orthonormale polynomiale permettant de réduire le couplage entre les différents paramètres du modèle. Le chapitre 3 et le chapitre 4 présentent les techniques d'estimation utilisées et qui sont respectivement la procédure de maximum de vraisemblance et l'estimation bayésienne. Une analyse comparative de ces deux techniques sera présentée dans le chapitre 5.

L'extension aux signaux mono-composantes de longue durée est présentée dans le chapitre 6. Nous détaillerons la technique de segmentation qui permet d'obtenir des segments courts adaptés à une modélisation locale efficace et cela à partir d'un signal de longue durée. Sur chacun de ces segments, nous appliquerons les techniques présentées dans les chapitres 3 et 4. Ensuite, nous présenterons une technique pour fusionner les segments estimés et par conséquent reconstruire le signal dans sa totalité.

Le chapitre 7 traite les signaux multi-composantes sur un segment de courte durée. La démarche adoptée et les algorithmes proposés dans l'estimation et la reconstruction de ce type de signaux découlent de ceux présentés dans le cas d'un signal mono-composante.

Les approches proposées dans cette thèse sont validées sur des signaux réels illustrant, ainsi, leur caractère générique et la diversité des applications. Dans "Conclusion et Perspectives", nous précisons d'autres applications potentielles de ces approches ainsi que quelques voies de recherche qu'il conviendrait de poursuivre.

Chapitre 1

Etat de l'art

Résumé

Dans ce chapitre, nous présentons l'état de l'art des méthodes et des algorithmes les plus connus et les plus employés dans l'analyse et l'estimation des signaux non-stationnaires. Nous distinguons les approches non-paramétriques de celles basées sur la modélisation des signaux et nous rappelons les avantages et les inconvénients de chacune de ses approches.

Mots clefs

Représentations temps fréquences, spectrogramme, distribution de Wigner-Ville, fonction d'ambiguïté d'ordre supérieur, maximum de vraisemblance, échantillonnage pondéré, opérateur unitaire de déformation.

Sommaire

1.1	Introduction	15
1.2	Représentations temps-fréquence	15
1.2.1	Spectrogramme	16
1.2.1.1	Spectrogramme de puissance	16
1.2.1.2	Spectrogramme de phase	17
1.2.1.3	Spectrogramme ré-alloué	18
1.2.2	Distributions de Wigner-Ville et de Wigner-Ville polynomiale	19
1.2.2.1	Distribution de Wigner-Ville	19
1.2.2.2	Distribution Pseudo-Wigner-Ville	20
1.2.2.3	Distribution de Wigner-Ville polynomiale	21
1.2.3	Représentations temps-fréquence - Opérateur “ <i>warping</i> ”	22
1.2.4	Méthode S	23
1.3	Techniques basées sur la modélisation de la phase du signal	24
1.3.1	Techniques basées sur la fonction d’ambiguïté d’ordre supérieur	25
1.3.1.1	Fonction d’ambiguïté d’ordre supérieur	25
1.3.1.2	Fonction d’ambiguïté d’ordre supérieur multi-retards	30
1.3.1.3	Produit de fonctions d’ambiguïté d’ordre supérieur	31
1.3.1.4	Fonction d’ambiguïté d’ordre supérieur - <i>Warping</i>	31
1.3.2	Estimation par la méthode des moindres carrés non-linéaires instantanés	33
1.3.3	Maximum de vraisemblance et échantillonnage pondéré	35
1.4	Autres techniques	38
1.5	Conclusion	39

1.1 Introduction

L'analyse et l'estimation des signaux non-stationnaires est un problème qui ne cesse de recueillir l'attention des traiteurs des signaux. Dans les techniques et les méthodes qui leur ont été dédiées, nous distinguons principalement deux catégories selon l'attribution ou non d'un modèle *a priori* au signal étudié.

Dans la première catégorie où aucun modèle du signal n'est supposé, figurent les représentations temps-fréquence bilinéaires telles que le spectrogramme, la distribution de Wigner-Ville, la méthode S. La description de ces représentations fera l'objet de la première partie de ce chapitre. Leurs avantages et leurs inconvénients ainsi que les améliorations et les généralisations qui leur étaient apportées y seront également exposées.

La deuxième partie du chapitre sera consacrée aux méthodes basées sur une modélisation paramétrique du signal. Nous nous intéressons en particulier à la procédure de maximum de vraisemblance, à la transformation polynomiale basée sur la fonction d'ambiguïté d'ordre supérieur et à ses dérivés.

Enfin, nous citons quelques récentes techniques telles que le filtrage particulière qui permet le suivi de faible non-stationnarités.

1.2 Représentations temps-fréquence

En fonction de leurs définitions et de leurs propriétés mathématiques [Fla93], les représentations temps-fréquence (RTF) peuvent être linéaires ou bilinéaires.

Historiquement, les RTF linéaires telles que la transformée de Fourier à court terme, la transformée de Gabor et ses variantes [Gab46], ont été les premières méthodes proposées. Insuffisantes dans l'analyse et la caractérisation des signaux non-stationnaires, elles ont cédé la place aux RTF bilinéaires. Ces dernières sont par définition des fonctions conjointes du temps et de la fréquence, ne supposant aucun modèle *a priori* du signal.

Bien que, dans certains cas, elles puissent prendre des valeurs négatives dépourvues de sens physique, les RTF bilinéaires tentent de fournir une distribution énergétique en temps et en fréquence. Cette distribution permet de suivre l'évolution temporelle du contenu fréquentiel du signal. La classe de Cohen [Coh66, Coh89, Coh93] a fourni une forme générale de l'ensemble des RTF bilinéaires covariantes par translation dans le plan temps-fréquence. Elle englobe la définition du spectrogramme, de la distribution de Wigner-Ville, celle de Choi-Williams (cf. annexe A.1). Moyennant la spécification d'une fonction arbitraire de paramétrisation $h(\xi, \tau)$, cette forme générale est définie par :

$$RTF_y(t, \nu; h) = \int_{\mathbb{R}} \int_{\mathbb{R}} \int_{\mathbb{R}} h(\xi, \tau) y\left(\mu + \frac{\tau}{2}\right) y^*\left(\mu - \frac{\tau}{2}\right) e^{-2\pi j \nu \tau} e^{2\pi j \xi (\mu - t)} d\xi d\mu d\tau, \quad (1.1)$$

où le signal d'étude $y(t)$ est considéré analytique.

Dans la suite, quelques RTF seront présentées. Nous rappellerons leurs avantages et leurs inconvénients.

1.2.1 Spectrogramme

Il s'agit d'une RTF basée d'une part sur l'analyse de Fourier à court terme et d'autre part sur l'introduction d'une fenêtre glissante localisant les événements temporels dans le signal. Nous présentons tout d'abord le spectrogramme de puissance (appelé couramment spectrogramme), puis le spectrogramme de phase et nous évoquons à la fin la technique de ré-allocation du spectrogramme.

1.2.1.1 Spectrogramme de puissance

Le spectrogramme est défini en temps continu par :

$$\mathcal{S}_y(t, \nu) = |\mathcal{F}_y(t, \nu)|^2 \quad \text{avec} \quad \mathcal{F}_y(t, \nu) = \int_{\mathbb{R}} h_{T_0}^*(\tau - t) y(\tau) e^{-2\pi j \nu \tau} d\tau, \quad (1.2)$$

où t et ν sont respectivement le temps et la fréquence et $\mathcal{F}_y(t, \nu)$ est la transformée de Fourier à court terme du signal étudié $y(t)$. $h_{T_0}(t)$ est la fenêtre d'analyse de durée T_0 . Il s'agit de calculer le spectre sur chaque portion du signal et de l'associer au temps central de la fenêtre d'analyse. La fonction de paramétrisation permettant de retrouver l'expression du spectrogramme à partir de la forme générale de la classe de Cohen (1.1) est fournie en annexes A.1.

La version en temps discret du spectrogramme s'écrit en utilisant une fenêtre d'analyse $h_{2L}[n]$ définie en temps discret et de longueur $2L$:

$$\mathcal{S}_y[n, k] = |\mathcal{F}_y[n, k]|^2 \quad \text{avec} \quad \mathcal{F}_y[n, k] = \sum_{l=-L+1}^L h_{2L}^*[l - n] y[l] e^{-2\pi j \frac{l}{2L} k}, \quad k \in \mathbb{Z}, \quad (1.3)$$

Les figures 1.1(a) et (b) montrent les spectrogrammes obtenus respectivement dans le cas du chant de canari et du bilan de masses de glaciers de Sarennes (cf. figure 1). Nous constatons l'existence de structures d'énergie maximale dans le plan temps-fréquence, décrivant l'évolution temporelle du contenu fréquentiel. Ces structures représentent les fréquences instantanées qui composent le signal. Le spectrogramme n'introduit pas d'interférences entre les composantes du signal.

Quoiqu'il continue d'être employé comme un outil d'interprétation et d'analyse de signaux [Ele04], le spectrogramme présente certaines limitations. Tout d'abord, l'utilisation de la transformée de Fourier à court terme suppose que le signal est localement stationnaire sur la fenêtre d'analyse $h_{T_0}(t)$, ce qui n'est pas toujours le cas. Ensuite, la résolution temps-fréquence, *i.e.* la qualité de l'image temps-fréquence, dépend du choix de la durée T_0 de la fenêtre d'analyse. Ce choix est d'ailleurs limité par le principe d'incertitude d'Heisenberg-Gabor, défini par $T_0 \Delta \nu \geq \frac{1}{4\pi}$, où $\Delta \nu$ désigne la bande spectrale. Ainsi, le produit "durée-bande" de tout signal est borné inférieurement. Une fenêtre d'analyse ne peut pas être infiniment petite, car elle induit une très bonne localisation en temps et une très mauvaise localisation en fréquence et *vis versa*. Un compromis est donc nécessaire entre la localisation temporelle et celle fréquentielle.

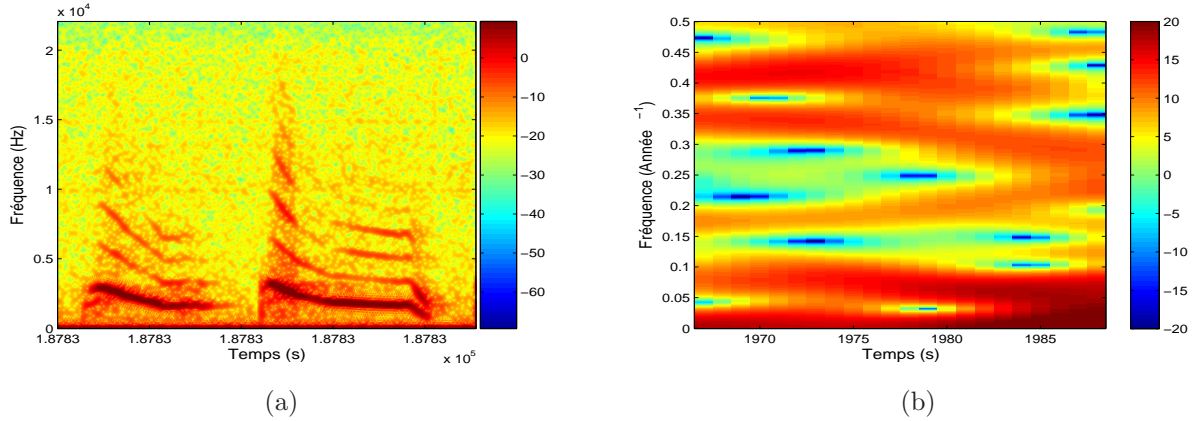


FIG. 1.1 – Spectrogrammes de signaux réels : (a) Spectrogramme de chant d’un canari calculé avec une fenêtre d’analyse glissante de 256 points, un recouvrement de 50% et une transformée de Fourier discrète (TFD) de taille $N_{TF} = 1024$. (b) Spectrogramme de bilan de masses de glaciers de Sarennes, calculé avec $N_{TF} = 1024$, une fenêtre d’analyse de 19 points et un recouvrement de 99%.

Dans le contexte de signaux courts (cf. figure 1.1(b)), la résolution temps-fréquence est un véritable handicap. En effet, la durée du signal est très courte (le nombre d’échantillons étant très petit) et par conséquent la durée de la fenêtre d’analyse l’est aussi. Ceci se traduit par l’obtention d’une mauvaise résolution temps-fréquence avec un étalement des structures dans le plan temps-fréquence. A ce problème, se rajoute celui de la présence du bruit, contribuant davantage à la dégradation de l’estimation des signaux analysés. De plus, le spectrogramme est non inversible, il est nul jusqu’à un instant postérieur (respectivement antérieur) d’une demi-longueur de la fenêtre d’analyse au début (respectivement à la fin) du signal, l’information qui peut être extraite se trouve réduite. Cette situation ne peut s’améliorer qu’en prenant une fenêtre temporelle très courte, ce qui dégrade la résolution fréquentielle.

Plusieurs travaux ont essayé d’améliorer le spectrogramme en intégrant l’information perdue au niveau de la phase de la transformée de Fourier à court terme (1.2). Nous n’évoquerons, dans les deux paragraphes suivants, que le spectrogramme de phase et la technique du spectrogramme ré-alloué.

1.2.1.2 Spectrogramme de phase

Dans [Leo00], une représentation pratique des informations contenues dans la phase de la transformée de Fourier à court terme (1.2) est décrite sous forme d’un spectrogramme de phase, complétant ainsi les informations issues du spectrogramme de puissance.

Le spectrogramme de phase (cf. figure 1.2) consiste à aligner, dans le plan temps-fréquence, la phase de $\mathcal{F}_y(t, \nu)$ selon l’axe des fréquences et à réaliser ensuite un déroulement de cette dernière selon l’axe du temps [Leo00]. Bien que pratique, cette technique présente les mêmes limitations que le spectrogramme de puissance.

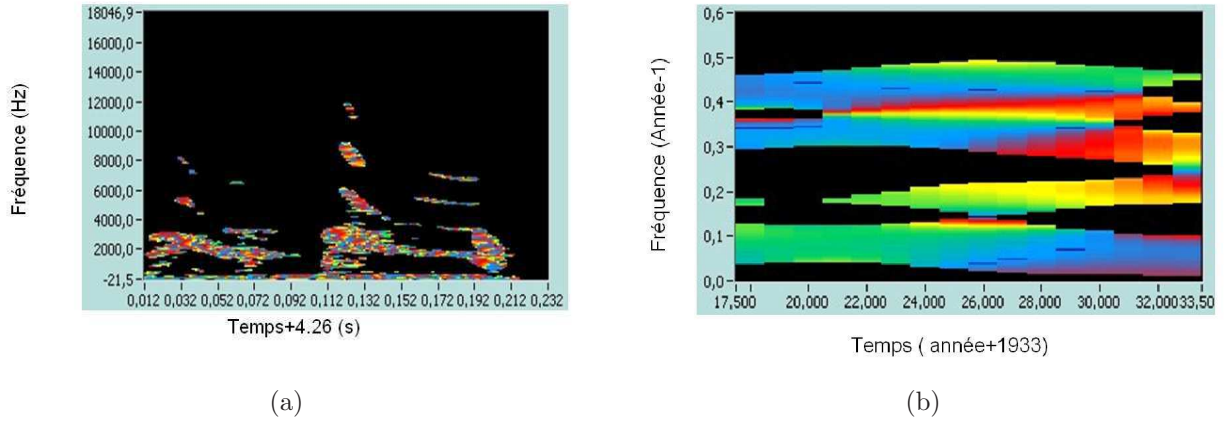


FIG. 1.2 – Spectrogrammes de phase de signaux réels : (a) Celui du chant d'un canari, calculé avec une fenêtre d'analyse de 256 points, un recouvrement de 50% et $N_{TF} = 1024$. (b) Celui du bilan de masses de glaciers de Sarennes calculé avec une fenêtre d'analyse de 24 points, un recouvrement de 99% et $N_{TF} = 1024$. Ces deux spectrogrammes sont calculés en utilisant les programmes fournis par M. François léonard [Leo00].

1.2.1.3 Spectrogramme ré-alloué

La ré-allocation du spectrogramme (de puissance) consiste à recentrer son énergie au voisinage de la loi de la fréquence instantanée [Fla93]. En pratique, la règle de la ré-allocation est définie, à partir du $S_y(t, \nu)$ et de $\mathcal{F}_y(t, \nu)$ (1.2), par :

$$S_y(t, \nu) \longrightarrow S_y(t_a(t, \nu), \nu_a(t, \nu)), \quad (1.4)$$

$$\text{où } t_a(t, \nu) = -\frac{1}{2\pi} \frac{\partial}{\partial \nu} \arg \mathcal{F}_y(t, \nu) \quad \text{et} \quad \nu_a(t, \nu) = \nu + \frac{1}{2\pi} \frac{\partial}{\partial t} \arg \mathcal{F}_y(t, \nu). \quad (1.5)$$

Cette règle est souvent utilisée pour ré-allouer uniquement les points temps-fréquence (t, ν) possédant une énergie supérieure à un seuil. Ce dernier est souvent fixé par l'utilisateur. Sur les figures 1.3(a) et (b), sont illustrés respectivement le spectrogramme ré-alloué de chant de canari et celui du bilan de masses de glacier de Sarennes.

Dans [GM05], la ré-allocation est opérée selon l'axe de la fréquence uniquement (2ème égalité de (1.5)). Pour déterminer les fréquences qui ne sont pas ré-allouées, un consensus noté C a été défini. Pour une même bande spectrale $\Delta\nu$ fixée, C est choisi égal à la valeur médiane de $1/|\nu_a(t, \nu_1) - \nu_a(t, \nu_2)|$, évaluée pour tous canaux fréquentiels ν_1 et ν_2 couverts par $\Delta\nu$ à l'instant t . Ensuite, on détermine $\Delta\nu$ optimale qui maximise ce consensus. Dès lors, les fréquences ν qui vérifient $|\nu_a(t, \nu) - \nu| > C$ ne sont pas ré-allouées et ne contribuent pas dans l'analyse.

Bien que la ré-allocation de spectrogramme améliore l'analyse des signaux bande-étroite non-bruités et ayant des composantes complètement séparées dans le plan temps-fréquence, le problème de la résolution reste inchangé. De plus, la ré-allocation ne fonctionne pas en présence du bruit, de signaux large-bande et de signaux bande-étroite présentant des

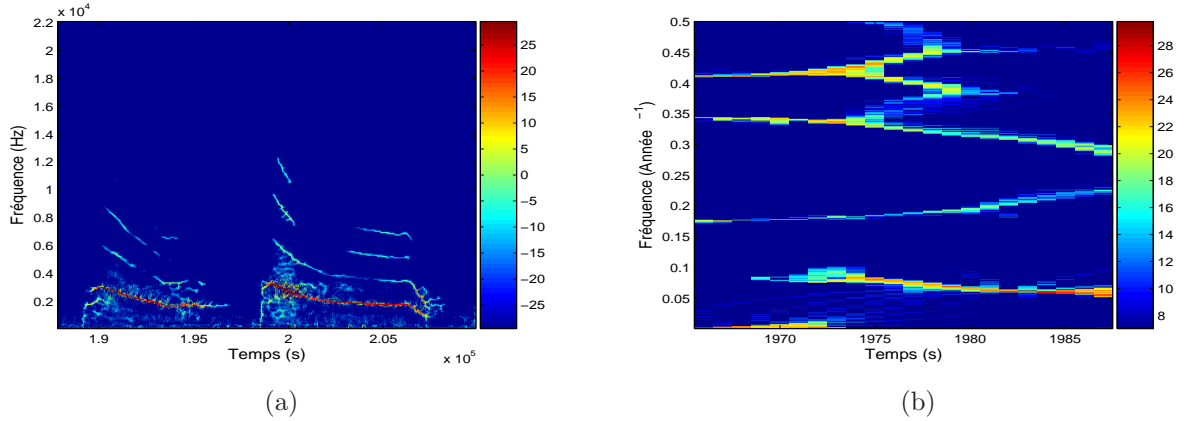


FIG. 1.3 – Spectrogrammes ré-alloués de signaux réels : (a) celui de chant du canari calculé avec une fenêtre d’analyse de 256 points, un recouvrement de 50% et $N_{TF} = 1024$. (b) Celui du bilan de masses de glaciers de Sarennes calculé une fenêtre d’analyse de 24 points, un recouvrement de 99% et $N_{TF} = 1024$.

croisements dans le plan temps-fréquence.

Dans la suite, nous nous intéressons à d’autres représentations temps-fréquence définies à partir de la distribution de Wigner-Ville.

1.2.2 Distributions de Wigner-Ville et de Wigner-Ville polynomiale

Nous rappelons dans un premier temps la définition et les propriétés de la distribution de Wigner-Ville. Nous évoquons les distributions Pseudo-Wigner-Ville et Pseudo-Wigner-Ville polynomiale, qui sont des généralisations de la distribution de Wigner-Ville.

1.2.2.1 Distribution de Wigner-Ville

L’élément central de la classe de Cohen (1.1) est la distribution de Wigner-Ville ($h(\xi, \tau) = 1$) définie¹ pour un signal $y(t)$ par :

$$\mathcal{WV}_y(t, \nu) = \int_{\mathbb{R}} y\left(t + \frac{\tau}{2}\right) y^*\left(t - \frac{\tau}{2}\right) e^{-2\pi j\nu\tau} d\tau. \quad (1.6)$$

Proposée antérieurement par J. Ville [Vil48], cette distribution représente le “spectre instantané” satisfaisant aux contraintes de distributions marginales. Elle ne nécessite pas l’introduction d’une fenêtre extrinsèque au signal comme dans le cas du spectrogramme. Toutefois, on peut l’interpréter comme étant une analyse de Fourier à court terme dans laquelle la fenêtre est adaptée au signal. De ce fait, les supports temporel et fréquentiel du signal analysé sont conservés, l’énergie du signal l’est aussi. De plus, le problème d’incertitude et de compromis temps-fréquence n’existe plus avec cette distribution. L’autre

¹Il s’agit d’une définition symétrique de la distribution de Wigner-Ville. Dans [Coh89], une définition non-symétrique est également proposée : $\mathcal{WV}_y(t, \nu) = \int_{\mathbb{R}} y(t + \tau) y^*(t) e^{-2\pi j\nu\tau} d\tau$.

propriété intéressante que possède la distribution de Wigner-Ville est sa capacité de se localiser et de se concentrer sur la loi de la fréquence instantanée d'un signal mono-composante modulé linéairement en fréquence.

Quand il s'agit de travailler en temps discret, la distribution de Wigner-Ville est définie par :

$$\mathcal{WV}_y[k, \nu] = 2 \sum_{n \in \mathbb{Z}} y[k+n] y^*[k-n] e^{-4\pi j \nu n}. \quad (1.7)$$

Il se trouve que cette expression est périodique en fréquence, de période $\frac{1}{2}$, soit de la moitié de la période de la transformée de Fourier classique. Un repliement du spectre affectera donc la distribution de Wigner-Ville dans les plages de fréquences $[\frac{-1}{2}, \frac{-1}{4}]$ et $[\frac{1}{4}, \frac{1}{2}]$. Un moyen de s'en débarrasser est de sur-échantillonner le signal.

1.2.2.2 Distribution Pseudo-Wigner-Ville

Généralement, pour des commodités de calcul en présence d'un signal à support temporel infini (très grand en pratique), on effectue une opération de lissage fréquentiel de la distribution de Wigner-Ville par l'intermédiaire d'une fenêtre $h_{T_0}(t)$. Il en résulte une distribution appelée Pseudo-Wigner-Ville définie par [Coh89, Fla93] :

$$PWV_y(t, \nu) = \int_{\mathbb{R}} h_{T_0}(\tau) y(t + \frac{\tau}{2}) y^*(t - \frac{\tau}{2}) e^{-2\pi j \nu \tau} d\tau. \quad (1.8)$$

De plus, si on trouve une fenêtre $h_{T_0}(t)$ s'écrivant sous la forme $h(\frac{t}{2}) h^*(\frac{t}{2})$, alors cette distribution garde l'esprit de celle de Wigner-Ville tout en se rapprochant du spectrogramme, *i.e.* d'une analyse glissante de Fourier à court terme. Le lien entre la distribution Pseudo-Wigner-Ville et la transformée de Fourier à court terme (1.2) est le suivant :

$$PWV_y(t, \nu) = 2 \int_{\mathbb{R}} \mathcal{F}_y(t, \nu + \nu_1) \mathcal{F}_y^*(t, \nu - \nu_1) d\nu_1. \quad (1.9)$$

Compte tenu de (1.8), la distribution Pseudo-Wigner-Ville utilise les mêmes fonctions que le spectrogramme, à savoir une fenêtre à court terme glissante et une transformation de Fourier couplée à une opération quadratique, sauf que l'ordre dans lequel ces fonctions interviennent est différent et conduit à des propriétés différentes.

La distribution de Wigner-Ville (ou Pseudo-Wigner-Ville) peut prendre des valeurs négatives, ce qui rend son interprétation comme densité d'énergie difficile. Un autre problème majeur est celui de l'apparition des termes d'interférences externes et internes réduisant la lisibilité de l'image temps-fréquence. Par externe, on désigne les interférences résultantes de l'interaction entre des composantes distinctes du signal et par interne les interférences qui apparaissent en présence d'une modulation non-linéaire de la fréquence. Ces termes d'interférences dus, essentiellement, au caractère quadratique de ces distributions sont illustrés sur la figure 1.4 pour deux signaux. Le premier est formé de deux composantes modulées linéairement en fréquence et le deuxième est un signal mono-composante ayant une modulation de fréquence sinusoidale.

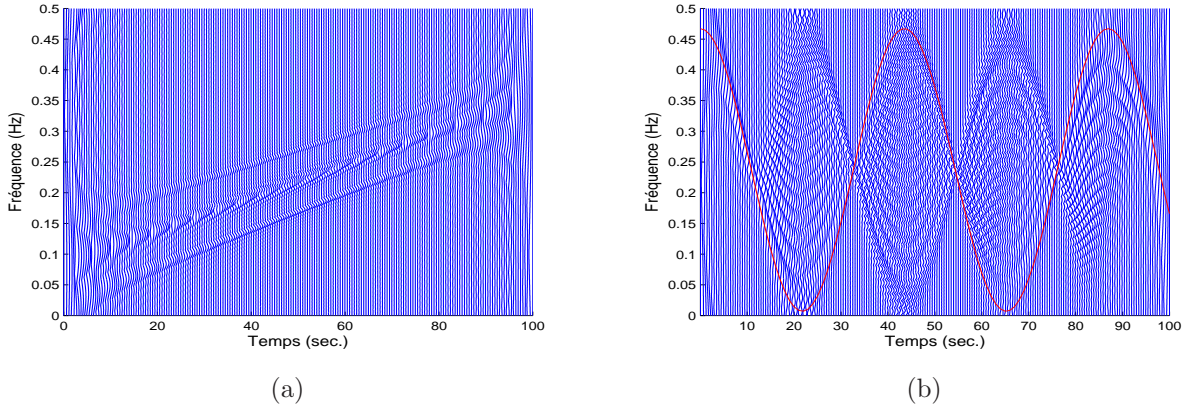


FIG. 1.4 – Distribution de Wigner-Ville : (a) un signal à deux composantes modulées linéairement en fréquence, la raie centrale correspond aux termes d'interférences. (b) Un signal mono-composante ayant une modulation sinusoïdale en fréquence, la courbe réelle (—) représentant cette modulation est également tracée.

1.2.2.3 Distribution de Wigner-Ville polynomiale

Quand la modulation de fréquence du signal devient non-linéaire, une distribution plus adéquate que celle de Wigner-Ville classique est donnée par la distribution de Wigner-Ville polynomiale locale. Cette dernière est définie dans [Kat97, BB99] par :

$$\mathcal{WV}\mathcal{P}_y(t, \boldsymbol{\nu}) = \int_{\mathbb{R}} h_{T_0}(\tau) y(t + \tau) y^*(t - \tau) e^{-j\boldsymbol{\Theta}(\tau, \boldsymbol{\nu})} d\tau, \quad (1.10)$$

où $\boldsymbol{\nu} = [\nu_1, \nu_2, \dots, \nu_r]^T \in \mathbb{R}^r$ et $\boldsymbol{\Theta}(\tau, \boldsymbol{\nu}) = 2\pi (\nu_1 \tau + \nu_2 \tau^3/3! + \dots + \nu_r \tau^{(2r-1)}/(2r-1)!)$. La version de cette distribution en temps discret s'écrit :

$$\mathcal{WV}\mathcal{P}_y[k, \boldsymbol{\nu}] = \sum_{n \in \mathbb{Z}} h_{T_0}[n] y[k + n] y^*[k - n] e^{-j\boldsymbol{\Theta}[n, \boldsymbol{\nu}]}. \quad (1.11)$$

La $\mathcal{WV}\mathcal{P}_y(t, \boldsymbol{\nu})$ a été proposée par V. Katkovnik qui, après avoir généralisé le périodogramme classique au périodogramme local polynomial [Kat95, Kat97, KS98, Kat00], a étendu ce concept à la distribution de Wigner-Ville. La comparaison dans [Kat97] entre $\mathcal{WV}\mathcal{P}_y(t, \boldsymbol{\nu})$ et le périodogramme local polynomial donne des résultats similaires pour des modulations de fréquence linéaires.

Considérons, par exemple, un signal $y(t) = A e^{j\Phi(t)}$ où A est l'amplitude du signal supposée constante et $\Phi(t)$ est sa phase instantanée non-linéaire en fonction du temps. Sa fréquence instantanée, définie par $F(t) = \frac{1}{2\pi} \frac{\partial \Phi(t)}{\partial t}$, est supposée r -dérivable, ayant des dérivées bornées. Sous l'hypothèse d'un bruit additif circulaire, un estimateur de $F(t)$ est exprimé par :

$$\hat{\mathbf{F}}(t) = \arg \max_{\boldsymbol{\nu} \in \mathbb{R}} \mathcal{WV}\mathcal{P}_y(t, \boldsymbol{\nu}). \quad (1.12)$$

Malgré les améliorations apportées par cette généralisation, $\mathcal{WV}\mathcal{P}_y(t, \nu)$ hérite des inconvénients de la distribution de Wigner-Ville classique, à savoir les termes d'interférences et l'apparition des valeurs négatives. De plus, les hypothèses sur la dérivabilité de la fréquence instantanée et l'existence des bornes sur ses dérivées sont assez restrictives. L'application de cette méthode est souvent limitée aux signaux à phase polynomiale de degré faible, inférieur à 3. En pratique, l'implémentation de (1.12) nécessite l'utilisation d'un algorithme d'optimisation récursif Gauss-Newton [Bro70], qui ne garantit pas la convergence en présence de minima locaux.

Différentes améliorations ont été apportées à la distribution de Wigner-Ville polynomiale dans les contextes suivants : signaux multi-composantes [Sta98], signaux noyés dans du bruit additif non-gaussien [Ben04, BS04] .

1.2.3 Représentations temps-fréquence - Opérateur “*warping*”

Partant de la propriété de la distribution de Wigner-Ville, *i.e.* la localisation parfaite sur les modulations de fréquence linéaires des signaux mono-composantes, on a cherché à transformer les RTF existantes en RTF parfaitement localisées sur des modulations non-linéaires. Dans [PKBB93, Bar95a, HPSBB97], les RTF bilinéaires classiques sont associées à un opérateur unitaire de déformation “*warping*”, qui déforme l'axe temporel ou fréquentiel de manière à transformer le signal étudié en un signal stationnaire.

Un opérateur unitaire \mathcal{U}_w relatif à une fonction lisse w (appelée fonction de déformation) est, par définition [Tay58], une transformation linéaire d'un espace de Hilbert vers un autre espace de Hilbert, qui préserve l'énergie et le produit scalaire par isométrie. Cet opérateur peut être défini dans le domaine temporel comme dans celui fréquentiel :

$$\mathcal{U}_w(t; y) = |w'(t)|^{\frac{1}{2}} y(w(t)) \quad \text{ou} \quad \mathcal{U}_w(\nu; \mathcal{F}_y) = |w'(\nu)|^{\frac{1}{2}} \mathcal{F}_y(w(\nu)), \quad (1.13)$$

où w' est la dérivée de w et $\mathcal{F}_y(\nu)$ est la transformée de Fourier de $y(t)$.

La fonction w est choisie de façon à assurer une linéarisation des structures non-linéaires du signal. Dès lors, l'emploi d'une distribution de la classe de Cohen et en particulier celle de Wigner-Ville permet d'obtenir une bonne localisation dans le plan temps-fréquence. Ceci est illustré sur la figure 1.5(b) [Ioa03] dans le cas d'un signal mono-composante.

Il est possible de retourner dans le domaine temporel ou fréquentiel d'origine (cf. figure 1.5(c)) en utilisant l'opérateur de déformation inverse “*unwrapping*”, noté \mathcal{U}_w^{-1} . La nouvelle RTF non-linéaire ainsi obtenue est bien adaptée à la modulation non-linéaire de la fréquence du signal.

Toutefois, cette méthodologie présente un certain nombre de problèmes. D'abord, l'opérateur de déformation w doit être adapté à la classe de la non-linéarité de la structure temps-fréquence du signal analysé. Or, cette information est inconnue dans la plupart des applications. Dans le cas où le signal est formé par des composantes multiples, ayant des structures temps-fréquence différentes, il faut construire plusieurs opérateurs de

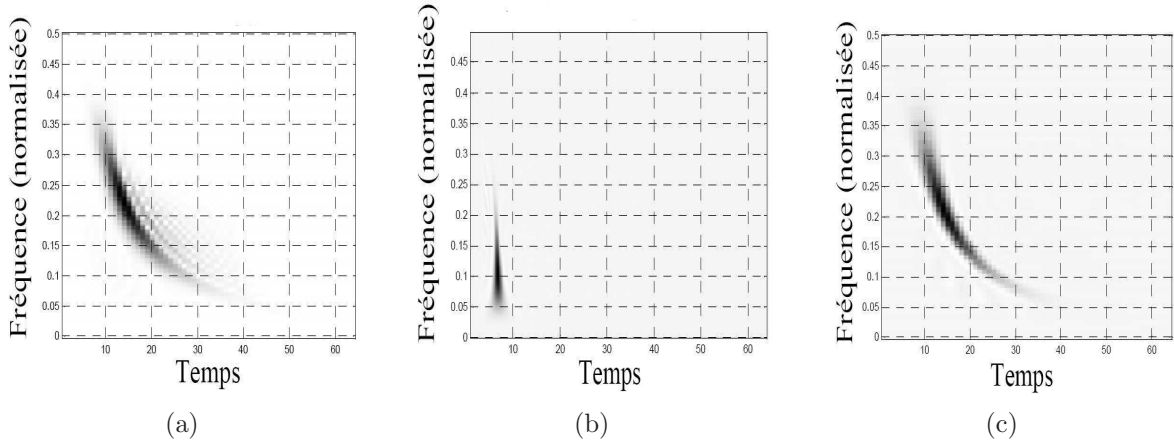


FIG. 1.5 – Application d’un opérateur unitaire de déformation à un signal synthétique ayant une modulation de fréquence hyperbolique [Ioa03] : (a) Distribution de Wigner-Ville du signal avant la déformation. (b) Distribution de Wigner-Ville après la déformation. (c) RTF obtenue après application de l’opérateur de déformation inverse.

déformation. Pour ce faire, les auteurs [PKBB93, Bar95a, HPSBB97, IQ02, Ioa03] proposent d’utiliser un dictionnaire étendu, formé de plusieurs classes de modulations non-linéaires. Ainsi, par une décomposition généralisée du signal selon une démarche itérative basée sur le “*Matching Pursuit*”, on peut obtenir ou définir plusieurs opérateurs de déformation. Le résultat dépend de la taille du dictionnaire contenant les types de modulations, ce qui constitue une limitation essentielle de cette méthode.

1.2.4 Méthode S

La méthode S proposée par Stankovic [Sta94, SIP96, DIS03, SDT05] est une représentation temps fréquence qui se situe entre la distribution de Wigner-Ville et le spectrogramme. En d’autres termes, elle s’approche de la distribution pseudo-Wigner-Ville tout en évitant les termes d’interférences.

Moyennant une fenêtre d’analyse $H(\nu)$ définie en fréquence et en utilisant l’expression de la transformée de Fourier à court terme $\mathcal{F}_y(t, \nu)$ (1.2), la méthode S s’écrit :

$$\mathcal{MS}_y(t, \nu) = 2 \int_{\mathbb{R}} H(\nu_1) \mathcal{F}_y(t, \nu + \nu_1) \mathcal{F}_y^*(t, \nu - \nu_1) d\nu_1. \quad (1.14)$$

En discret, la méthode S est définie comme suit :

$$SM_y[n, k] = \frac{1}{N+1} \sum_{l=-L}^L \mathcal{F}_y[n, k+l] \mathcal{F}_y^*[n, k-l], \quad (1.15)$$

où $N+1$ est le nombre total d’échantillons du signal et $2L+1$ est le nombre de canaux fréquentiels. Notons que le spectrogramme et la distribution de Wigner-Ville sont des cas

marginaux de cette distribution. En effet, le spectrogramme est obtenu avec $L = 0$, tandis que la version discrète de la distribution de Wigner-Ville se retrouve en choisissant $H(\nu_1) = 1$ et $L = \frac{N}{2}$.

En présence du bruit les performances de la méthode S sont nettement meilleures que celles du spectrogramme et de la distribution de son implémentation efficace peut se faire d'une façon peu coûteuse en temps de calcul et en mémoire.

Une propriété très importante de la méthode S est qu'elle fournit une RTF égale à la somme de la distribution de Wigner-Ville de chaque composante d'un signal multi-composantes. Grâce à cette propriété, les termes d'interférences sont réduits dans le cas de composantes linéairement modulées en fréquence.

Lors de l'élimination de ces termes d'interférences et moyennant un choix convenable de $H(\nu)$, il est possible d'obtenir les auto-termes (par opposition aux termes d'interférences), sans que leur aspect soit changé. L'expression de l'estimateur de la fréquence instantanée de la i -ème composante du signal est la suivante [SIP96] :

$$\hat{F}_i(t) = \arg \max_{\nu \in Q_{\nu_i}} \mathcal{SM}(t, \nu), \quad (1.16)$$

où Q_{ν_i} est un intervalle de fréquence situé autour de la fréquence instantanée de la i -ème composante. La moyenne et la variance de cet estimateur sont déterminées dans [SIP96].

D'après l'équation (1.16), cet estimateur ne peut être déterminé que sous l'hypothèse de composantes séparées dans le plan temps-fréquence (sans recouvrement de composantes). Ceci constitue une première limitation de la méthode S. En présence de modulations de fréquence fortement non-linéaires, par exemple polynomiales de degré supérieur à 3, cette méthode présente les mêmes problèmes que la distribution de Wigner-Ville.

Dans le but d'améliorer ses performances, la ré-allocation de la méthode S a été proposée dans [DS99], mais elle s'avère très coûteuse en termes de complexité calculatoire comparée aux autres RTF ré-allouées. De plus, le résultat dépend du niveau du bruit et de la séparation des composantes dans le plan-temps fréquence.

1.3 Techniques basées sur la modélisation de la phase du signal

Dans le contexte des signaux fortement non-stationnaires, plusieurs techniques basées sur un modèle paramétrique du signal ont été proposées, évitant ainsi les limitations de l'analyse par RTF.

Historiquement, un fort intérêt a été porté en particulier à la modélisation polynomiale de la phase instantanée du signal [PPF93, FF95, ZGS96, TVF97, BSG98, IZ01, SK02, ICQ04, O'S04]. Une telle modélisation constitue un moyen satisfaisant pour renseigner sur les non-linéarités présentes dans les signaux rencontrés dans la majorité des applications

[QM91, QDMH92, MKQ93, MMB93, GDL94, GBM⁺96, PMT96, Rod97, CL99, Gir99, DCK00, Vin01].

Les polynômes sont des outils mathématiques, qui présentent beaucoup d'avantages. Ils sont simples à définir et se prêtent facilement à la dérivation, à l'intégration et aux opérations algébriques. Leur implémentation est facile à mettre en oeuvre. De plus, le théorème de Weierstrass assure que chaque fonction continue peut être approchée arbitrairement par un polynôme et justifie, ainsi, l'utilisation intensive de cette modélisation.

Théorème 1 (de Weierstrass) *Soit $f(n)$ une fonction continue sur un intervalle réel \mathcal{J} , alors $\forall \epsilon > 0$, $\exists M_\epsilon \in \mathbb{N}$ et $\exists \mathcal{P}_{M_\epsilon}(f, t)$ un polynôme d'ordre M_ϵ tel que, $\forall t \in \mathcal{J}$ on a $|\mathcal{P}_{M_\epsilon}(f, t) - f(t)| < \epsilon$.*

Dans la suite de ce chapitre, nous considérons les notations suivantes. $b(t)$ est un bruit blanc Gaussien additif et $y(t)$ est le signal bruité. La partie déterministe de ce dernier est formée par la somme de K composantes dont la k -ième est notée $s_k(t)$. Le signal observé $y(t)$ s'écrit sous la forme suivante :

$$y(t) = \sum_{k=1}^K s_k(t) + b(t) \quad \text{avec} \quad s_k(t) = A_k e^{j\Phi_k(t)} \quad \text{et} \quad \Phi_k(t) = \sum_{m=0}^{M_k} \Phi_{k,m} t^m. \quad (1.17)$$

Chaque composante $s_k(t)$ admet une amplitude constante A_k et une phase instantanée $\Phi_k(t)$. Cette dernière est modélisée par un polynôme de degré M_k , dont les coefficients sont notés $\{\Phi_{k,m}\}_{m=0,\dots,M_k}$. Le processus d'estimation consiste, alors, à déterminer chacun des $\Phi_{k,m}$.

Parmi les techniques proposées dans la littérature, nous rappelons, dans un premier temps, celles basées sur la fonction d'ambiguïté d'ordre supérieur (HAF : *Higher-order Ambiguity Function*). Ensuite, nous citons les techniques d'estimation au sens de maximum de vraisemblance et enfin celles utilisant l'échantillonnage pondéré. Les techniques d'estimation basées sur les approches Bayésiennes seront abordées dans le chapitre 4.

1.3.1 Techniques basées sur la fonction d'ambiguïté d'ordre supérieur

Nous présentons dans cette partie la définition de la fonction d'ambiguïté d'ordre supérieur, ainsi que les techniques qui lui sont associées telles que la fonction d'ambiguïté d'ordre supérieur "multi-retards" et le produit de fonctions d'ambiguïté d'ordre supérieur.

1.3.1.1 Fonction d'ambiguïté d'ordre supérieur

Initialement appelée Transformée Polynomiale [Fri93], la HAF est basée sur une différentiation de la phase du signal $y(t)$. Elle représente la transformée de Fourier du moment

instantanée d'ordre supérieur noté \mathcal{MOS} [Fri93] :

$$\mathcal{HAF}_m [y; \alpha, \tau] = \int_{\mathbb{R}} \mathcal{MOS}_m [y(t); \tau] e^{-j\alpha t} dt, \quad (1.18)$$

où m désigne l'ordre de la HAF et τ est appelé retard. Le \mathcal{MOS}_m d'ordre m est défini par la relation suivante² :

$$\mathcal{MOS}_m [y(t); \tau] = \prod_{l=0}^{m-1} [y^{(*l)}(t - l\tau)]^{C_{m-1}^l}. \quad (1.19)$$

Dans l'équation ci-dessus, $C_{m-1}^l = \frac{(m-1)!}{l!(m-l-1)!}$ désigne le coefficient binomial, quant à $y^{(*l)}(t)$ il est défini par :

$$y^{(*l)}(t) = \begin{cases} y(t) & \text{si } l \text{ est pair,} \\ y^*(t) & \text{si } l \text{ est impair.} \end{cases}$$

La version en temps discret de la HAF est définie comme étant la transformée de Fourier discrète de la version discrétisée du \mathcal{MOS} .

Nous notons, sans le démontrer, que le \mathcal{MOS} peut être calculé d'une manière itérative comme suit [Fri93, Por93] :

$$\mathcal{MOS}_m [y(t); \tau] = \mathcal{MOS}_2 [\mathcal{MOS}_{m-1} [y(t); \tau]; \tau]. \quad (1.20)$$

Cas d'un signal mono-composante Considérons le cas d'un signal mono-composante ($K = 1$ dans l'équation (1.17)) dont la phase est un polynôme de degré M .

Dans ce contexte et en l'absence de bruit, la formulation (1.20) présente quelques propriétés intéressantes. Dans ce cas, le $\mathcal{MOS}_m [y(t); \tau]$ d'ordre m est un signal à phase polynomiale de degré $M - m - 1$. Lui appliquer un \mathcal{MOS}_2 d'ordre 2 réduit son degré. En ré-itérant cette procédure, le \mathcal{MOS}_M est alors une composante purement harmonique.

Dans [Por93, Fri93], il est alors établi qu'un signal mono-composante non bruité, dont l'amplitude est constante et la phase est un polynôme de degré M , vérifie la relation suivante :

$$K = 1, \quad \forall \tau > 0 \quad \Phi_{K,M} = \frac{1}{M! \tau^{M-1}} \arg \max_{\alpha} |\mathcal{HAF}_M [y; \alpha, \tau]|. \quad (1.21)$$

Cette propriété intéressante va être l'idée clef pour l'estimation des coefficients de la phase. Une estimation $\hat{\Phi}_{K,M}$ de $\Phi_{K,M}$ est directement déterminée par la position du maximum de la HAF (1.21). L'application de cette dernière au nouveau signal $y(t) e^{-j\hat{\Phi}_{K,M} t^M}$ permet d'estimer le coefficient $\Phi_{K,M-1}$. Les itérations successives de cette procédure fournissent l'estimation de tous les coefficients de la phase. Nous notons que la contrainte "amplitude constante" peut être relaxée contre une hypothèse d'amplitude lentement variable.

²Il s'agit d'une définition symétrique. Il existe également une définition non-symétrique [Ioa03].

Cas d'un signal multi-composantes Dans [BSG98], une analyse utilisant la HAF est présentée dans le cas de signaux multi-composantes. Il était montré que pour de tels signaux (1.17), le MOS d'ordre M_k est un signal à phase polynomiale faisant intervenir des termes dits "propres" et des termes d'interférences. La définition de ces termes est illustrée par l'exemple suivant formé par deux composantes :

$$y(t) = e^{j(\Phi_{1,2} t^2 + \Phi_{1,1} t + \Phi_{1,0})} + e^{j(\Phi_{2,1} t + \Phi_{2,0})}.$$

Pour ce signal, le MOS d'ordre 2 s'exprime par :

$$\begin{aligned} \text{MOS}_2[y(t); \tau] &= e^{j(4\Phi_{1,2}\tau t + 2\Phi_{1,1}\tau)} \\ &+ e^{j(2\Phi_{2,1}\tau)} \\ &+ e^{j(-\Phi_{1,2} t^2 + (\Phi_{1,1} - \Phi_{2,1} + 2\Phi_{1,2}\tau) t + \Phi_{1,0} - \Phi_{2,0} + (\Phi_{1,1} + \Phi_{2,1})\tau + \Phi_{1,2} \tau^2)} \\ &+ e^{j(-\Phi_{1,2} t^2 + (-\Phi_{1,1} + \Phi_{2,1} + 2\Phi_{1,2}\tau) t - \Phi_{1,0} + \Phi_{2,0} + (\Phi_{1,1} + \Phi_{2,1})\tau - \Phi_{1,2} \tau^2)}. \end{aligned} \quad (1.22)$$

Le premier terme (respectivement le deuxième terme) ne fait intervenir que les paramètres de la première composante (respectivement de la deuxième composante), il est dit "propre". La phase des deux derniers termes est un polynôme, dont les coefficients font intervenir les paramètres des deux composantes à la fois. Ces deux termes sont alors appelés "termes d'interférences".

Dans [BSG98], l'influence des termes d'interférences sur le processus d'estimation a été étudiée. Pour un signal vérifiant (1.17) et avec $M_k = M$ pour $k = 1, \dots, K$, les termes propres du MOS d'ordre M sont des harmoniques de pulsation :

$$\forall k = 1, \dots, K, \quad \alpha_k = 2^{M-1} M! \tau^{M-1} \Phi_{k,M}. \quad (1.23)$$

Dans le cas où les coefficients polynomiaux, correspondant au même ordre m (1.17), sont distincts d'une composante à l'autre, les termes d'interférences sont négligeables devant les termes propres. L'équation (1.23) est utilisée pour estimer ces coefficients. En revanche, un problème d'identification se produit [BSG98] dans le cas où deux composantes présentent des coefficients identiques à partir d'un ordre supérieur à 2 et jusqu'à l'ordre M_k . En effet, dans ce cas, les termes d'interférences ne sont plus négligeables et affectent le processus d'estimation en faisant apparaître de faux harmoniques.

Pour réaliser l'estimation des coefficients polynomiaux, deux solutions ont été proposées [BSG98]. Dans la première solution, tous les coefficients d'une même composante doivent être estimés avant de procéder à l'estimation d'autres coefficients correspondant à d'autres composantes. Il est donc nécessaire d'estimer l'amplitude de cette composante pour pouvoir la soustraire. Pour ce faire, on peut utiliser la corrélation avec le signal analysé.

Malheureusement, la présence du bruit dégrade l'estimation de l'amplitude, ce qui affecte l'estimation des coefficients de la phase de la composante suivante. De plus, cette façon de procéder ne fonctionne correctement que si le nombre de composantes est connu, c'est-à-dire on dispose d'une information *a priori* sur le nombre de composantes.

La seconde solution consiste à estimer les coefficients niveau après niveau, c'est-à-dire

on estime en même temps tous les coefficients ayant le même ordre m et correspondant à des composantes distinctes. Dans cette méthode, le nombre de composantes peut être sur-estimé. Puis, par corrélation avec le signal analysé, les composantes qui correspondent à du bruit seront éliminées. Comme la corrélation n'intervient qu'après le processus d'estimation, elle n'influe pas sur le résultat de l'estimation.

Les deux solutions présentent des performances similaires. Le choix entre les deux se fait en fonction de notre connaissance du degré de l'approximation et du nombre de composantes.

Limitations de la HAF Quoique l'algorithme itératif basé sur la HAF est facile à mettre en oeuvre, il présente plusieurs limitations dont les principales sont rappelées.

D'abord, un phénomène de masquage des maximums de la HAF (1.21) se produit en présence d'un faible rapport signal à bruit (RSB). Ces maximums n'étant plus décelables, l'estimation des $\{\Phi_{k,m}\}_{m=0,\dots,M_k}$ est entachée d'erreurs et est de mauvaise qualité.

A cela, s'ajoute le phénomène de propagation d'erreur illustré sur la figure 1.6. Il est dû essentiellement à l'estimation itérative des coefficients. En effet, la détermination du coefficient Φ_{m-1} étant dépendante de celle des Φ_m, \dots, Φ_M , l'erreur d'estimation affectant un de ces derniers s'accumule et s'étend à tous les coefficients d'ordres plus faibles. Dans [FF95, Ioa03], on a quantifié pour un signal mono-composante l'erreur de propagation ϵ_{M-1} obtenue à l'ordre $M-1$, en fonction de ϵ_M déterminée à l'itération précédente :

$$\epsilon_{M-1} = \epsilon_M \frac{M\tau}{(M-1)!} \sum_{m=0}^{M-1} (-1)^m C_{M-1}^m (-m)^{M-1}. \quad (1.24)$$

Cette erreur de propagation s'accroît avec l'augmentation du degré de l'approximation polynomiale.

Par ailleurs, l'estimation de la phase initiale et de l'amplitude des composantes se faisant par corrélation [FF95], la reconstruction du signal est de très mauvaise qualité, l'erreur est d'autant plus importante que le RSB est faible.

Pour les différentes raisons évoquées ci-dessus, l'utilisation de la HAF est souvent restreinte aux signaux présentant un fort RSB et ayant de faibles non-linéarités de la phase, typiquement des modulations linéaires de fréquence. L'échec de l'estimation est presque garanti en présence de fortes modulations de phase nécessitant un degré d'approximation élevé (cf. la figure 1.6).

Une autre limitation de la HAF découle directement de la modélisation considérée (1.17). En effet, la restriction de type Nyquist, nécessaire pour respecter Shannon, contraint la pulsation α_k (1.23) à vérifier $|\alpha_k| < \pi$. Ceci se traduit par une contrainte sur le coefficient d'ordre M :

$$|\Phi_{k,M}| < \frac{\pi}{2^{M-1} M! \tau^{M-1}}. \quad (1.25)$$

Pour des degrés d'approximations élevées, $M \geq 7$, les ordres de grandeurs de $\Phi_{k,M}$ sont très faibles, inférieurs à 10^{-6} . L'erreur d'estimation est, alors, du même ordre de grandeur

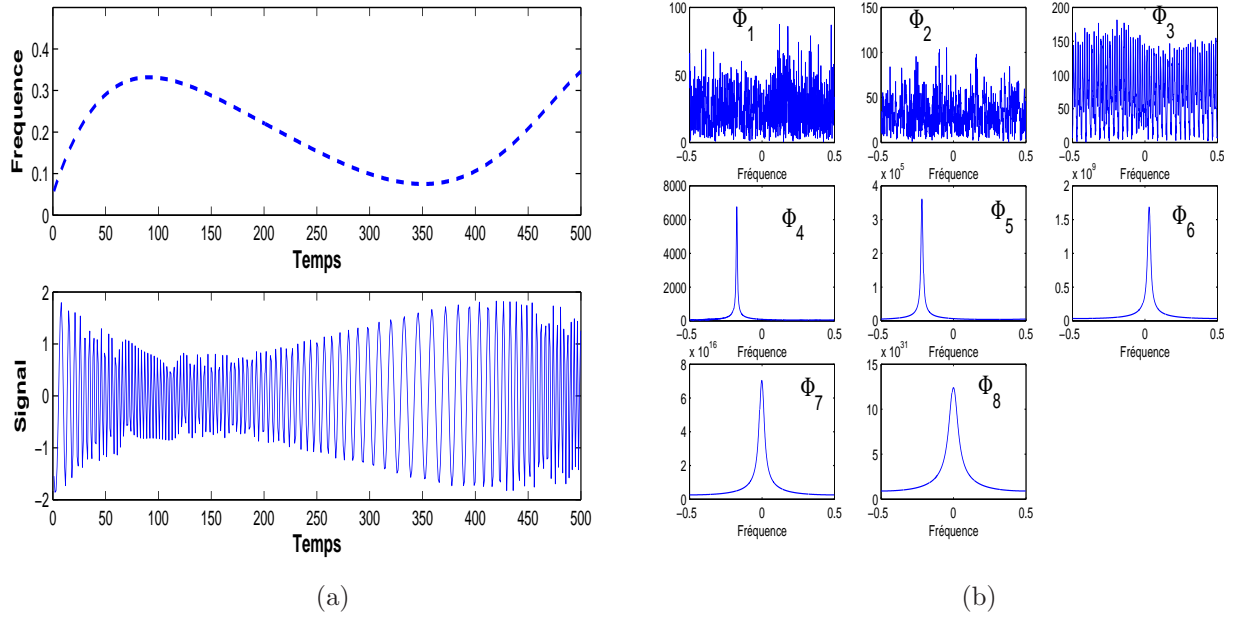


FIG. 1.6 – (a) Signal mono-composante et fortement non-stationnaire ayant une modulation polynomiale de la fréquence et de l'amplitude de degrés respectifs 7 et 3, la phase est de degré 8. (b) Propagation d'erreur de l'ordre le plus élevé au coefficient d'ordre inférieur avec la méthode de la HAF. Le niveau de l'erreur accumulée après l'estimation de Φ_4 affecte considérablement l'estimation celle de Φ_3 . La notation par rapport à K est supprimée.

que les $\Phi_{k,M}$.

Les performances de la HAF dépendent également du choix du retard τ , qui suit des règles empiriques. Dans [BSG98], bien qu'une étude décrit un choix de τ optimal du point de vue de la précision de l'estimation, aucun résultat n'a été rigoureusement démontré de manière analytique. L'optimalité est à considérer au sens de la résolution fréquentielle et conduit à la règle suivante :

- la valeur de τ retenue pour les cas $M = 2$ ou 3 est égale à $\frac{N+1}{2M}$, où $N + 1$ est le nombre d'échantillons,
- tandis que, pour le cas $M > 4$, le choix est $\tau = \frac{N+1}{M}$ [PP91].

Ces valeurs particulières de τ , ne sont pas toujours utilisables quand on travaille en temps discret et on se contente plutôt de leurs approximations.

La règle définie ci-dessus ré-affirme que la robustesse de la méthode dépend du nombre d'échantillons, puisqu'il intervient dans la détermination de τ . Quand N devient faible, on obtient une faible résolution fréquentielle avec la transformée de Fourier de \mathcal{MOS} (1.26), ce qui dégrade la qualité et la précision de l'estimation des coefficients de la phase.

Enfin, nous rappelons que les termes d'interférences, qui apparaissent lors de l'estimation de signaux multi-composantes, constituent une limitation principale de la technique de la HAF. Ces termes sont à l'origine de problèmes d'identification des composantes et

engendrent la perte de certains coefficients durant le processus d'estimation.

Plusieurs techniques ont été proposées et permettent de réduire les limitations de la HAF et d'améliorer les performances en présence d'un faible RSB. Dans [FP99], les auteurs proposent d'utiliser une RTF sans termes d'interférences pour initialiser l'algorithme itératif de la HAF. Cette RTF utilise des informations entropiques (*Minimum Cross Entropy*) et elle est évaluée de façon itérative à partir d'une distribution temps-fréquence positive, le spectrogramme par exemple [FP99]. Les composantes du signal sont estimées et éliminées l'une après l'autre. L'amélioration apportée par cette technique reste très modeste.

Dans [Bar93, Bar95b, BPS96, BMS97], l'idée clef est d'utiliser plusieurs retards au lieu d'un seul, tandis que dans [BSG98, SB99] c'est le concept de produit HAF qui est introduit. Dans [Ioa03, IQ03, ICQ04], on a combiné un opérateur unitaire de déformation (1.13) et la technique HAF. Dans, la suite, nous détaillons chacune de ces méthodes.

1.3.1.2 Fonction d'ambiguïté d'ordre supérieur multi-retards

Dans le contexte de signaux multi-composantes, l'utilisation de la HAF avec des retards multiples s'avère très utile [Bar95b, BPS96, BMS97]. Une meilleure robustesse vis-à-vis du bruit est obtenue et les termes d'interférences sont réduits. La HAF multi-retards, notée $\mathcal{HAF}_{\mathcal{ML}}$ est la transformée de Fourier du moment d'ordre supérieur multi-retards $\mathcal{MOS}_{\mathcal{ML}}$:

$$\mathcal{HAF}_{\mathcal{ML},m}[y; \alpha, \boldsymbol{\tau}_{M-1}] = \int_{\mathbb{R}} \mathcal{MOS}_{\mathcal{ML},m}[y(t); \boldsymbol{\tau}_{M-1}] e^{-j\alpha t} dt. \quad (1.26)$$

Le $\mathcal{MOS}_{\mathcal{ML}}$ est défini d'une manière séquentielle pour un signal $y(t)$, dont les composantes possèdent tous une phase polynomiale de degré M :

$$\begin{aligned} \mathcal{MOS}_{\mathcal{ML},1}[y(t)] &= y(t), \\ \mathcal{MOS}_{\mathcal{ML},2}[y(t); \boldsymbol{\tau}_1] &= \mathcal{MOS}_{\mathcal{ML},1}[y(t + \tau_1)] \mathcal{MOS}_{\mathcal{ML},1}^*[y(t - \tau_1)], \\ &\vdots \\ \mathcal{MOS}_{\mathcal{ML},M}[y(t); \boldsymbol{\tau}_{M-1}] &= \mathcal{MOS}_{\mathcal{ML},M-1}[y(t + \tau_{M-1}); \boldsymbol{\tau}_{M-2}] \mathcal{MOS}_{\mathcal{ML},M-1}^*[y(t - \tau_{M-1}); \boldsymbol{\tau}_{M-2}]. \end{aligned} \quad (1.27)$$

où $\boldsymbol{\tau}_m = [\tau_1, \tau_2, \dots, \tau_m]^\top$. Le symbole $^\top$ désigne la transposée d'un vecteur ou d'une matrice. L'expression de la HAF classique se retrouve facilement en prenant tous les retards égaux. Pour un signal multi-composantes (1.17), il est établi dans [BSG98] que les termes propres du \mathcal{MOS} d'ordre M_k sont des harmoniques de fréquence :

$$f_k = 2^{M-1} M! \left(\prod_{m=1}^{M-1} \tau_m \right) \Phi_{k,M}, \quad k = 1, \dots, K. \quad (1.28)$$

L'ensemble des retards est choisi de façon à optimiser la résolution fréquentielle et de minimiser l'apparition de faux *maxima* spectraux. Ces faux *maxima* sont fréquents avec la HAF

classique et qui sont à dus aux termes d'interférences. Il se fait d'une manière empirique en utilisant différents jeux de retards. Le choix des différents retards et le phénomène de propagation d'erreur restent les limitations principales de la HAF multi-retards.

1.3.1.3 Produit de fonctions d'ambiguïté d'ordre supérieur

Dans le but d'éliminer les termes d'interférences (1.22), la technique produit HAF (PHAF) a été développée dans [BSG98]. Elle consiste à multiplier plusieurs HAF multi-retards (1.26), qui ont subi une transformation d'échelle :

$$\mathcal{PHAF}[y; \alpha, \boldsymbol{\tau}_{M-1}^L] = \prod_{l=1}^L \mathcal{HAF}_{\mathcal{ML}, M} \left[y; \frac{\prod_{m=1}^{M-1} \tau_m^{(l)}}{\prod_{m=1}^{M-1} \tau_m^{(1)}} \alpha, \boldsymbol{\tau}_{M-1}^{(l)} \right], \quad (1.29)$$

où $\boldsymbol{\tau}_{M-1}^L = \left\{ \boldsymbol{\tau}_{M-1}^{(1)}, \boldsymbol{\tau}_{M-1}^{(L)}, \dots, \boldsymbol{\tau}_{M-1}^{(L)} \right\}$ est un ensemble de L vecteurs retards distincts notés $\boldsymbol{\tau}_{M-1}^{(l)} = \left[\tau_1^{(l)}, \tau_2^{(l)}, \dots, \tau_{M-1}^{(l)} \right]^T$.

Bien qu'elle soit définie à partir de la HAF multi-retards, le PHAF est plus robuste au bruit grâce à l'utilisation de la transformation d'échelle d'une part et du produit de la HAF multi-retards d'autre part. Dans [SB99], ses performances statistiques ont été étudiées et elles montrent une amélioration du contraste entre les différentes composantes, qui se traduit par la réduction des termes d'interférences.

Toutefois, la transformation d'échelle utilisée dans (1.29) occasionnent l'apparition des anomalies en présence des fortes non-linéarités, ce qui restreint l'application de cette technique aux modulations de fréquences linéaires. De plus, il n'est pas toujours évident de trouver la meilleure combinaison de retards qui optimisera la résolution fréquentielle de la PHAF.

1.3.1.4 Fonction d'ambiguïté d'ordre supérieur - Warping

Dans [Ioa03, IQ03, ICQ04], on a développé deux méthodes, chacune d'elles utilise un opérateur unitaire de déformation (1.13) et la HAF pour estimer les coefficients de la phase d'un signal mono-composante. Ces deux méthodes sont décrites dans les deux paragraphes suivants.

Déformation de l'axe fréquentiel de la HAF Par définition, un opérateur unitaire de déformation (1.13) transforme une structure, qui est une fonction non-linéaire d'un paramètre, en une autre structure linéaire équivalente. Or, connaissant la loi de dépendance entre les pulsations correspondant aux valeurs maximales de la HAF et les retards (1.21), à savoir $\alpha_M(\tau) = M! \tau^{M-1} \Phi_{K, M}$, on peut considérer la fonction de déformation suivante :

$$w(\tau) = \tau_w = \tau^{\frac{1}{M-1}}. \quad (1.30)$$

L'opérateur de déformation, qui lui est associé, est noté \mathcal{U}_w et est défini par (1.13). L'application de ce dernier au \mathcal{MOS}_M (1.19) permet d'obtenir le \mathcal{MOS}_{M_w} :

$$\mathcal{MOS}_{M_w}[y(t); \tau_w] = \mathcal{U}_w(\tau; \mathcal{MOS}_M) = |w'(\tau)|^{\frac{1}{2}} \mathcal{MOS}_M[y(t); w(\tau)]. \quad (1.31)$$

La HAF résultante après l'application de l'opérateur de déformation est la suivante :

$$\begin{aligned} \mathcal{HAF}_{M_w}[y; \tilde{\alpha}, \tau_w] &= \int_{\mathbb{R}} \mathcal{MOS}_{M_w}[y(t); \tau_w] e^{-j\tilde{\alpha}t} dt, \\ \text{selon la règle } \mathcal{HAF}_M[y; \alpha, \tau] &\xrightarrow{(1.30)} \mathcal{HAF}_{M_w}[y; \tilde{\alpha}, \tau_w]. \end{aligned} \quad (1.32)$$

Les *maxima* de la \mathcal{HAF}_{M_w} sont localisés aux pulsations $\tilde{\alpha}_M = M! \Phi_{K,M}$, ce qui permettra une meilleure précision de l'estimation de $\Phi_{K,M}$. On a proposé d'utiliser l'expression suivante pour déterminer $\Phi_{K,M}$:

$$\hat{\Phi}_{K,M} = \frac{1}{M! \tau^{M-1}} \arg \max_{\alpha} \left| \int_{\mathbb{R}} \mathcal{HAF}_{M_w}[y; \tilde{\alpha}; \tau_w] d\tau_w \right|. \quad (1.33)$$

La sommation porte sur les valeurs de la \mathcal{HAF}_{M_w} dans l'équation (1.33) afin d'augmenter, dans le cas de signaux multi-composantes, le contraste entre les termes à estimer.

L'utilisation de l'opérateur de déformation avec la HAF multi-retards a pour effet le recalage des *maxima* associés à plusieurs ordres. Ceci permet de mieux les distinguer et d'améliorer davantage la précision de l'estimation. La technique résultante est non seulement robuste au bruit mais possède aussi les mêmes avantages que le produit HAF tout en éliminant ses inconvénients liés au choix de retards.

Toutefois, le problème lié à la propagation d'erreur de l'ordre le plus élevé jusqu'aux ordres inférieurs ne disparaît pas avec cette technique. Ce qui empêche l'obtention de bons résultats dans le cas de signaux fortement non-stationnaires nécessitant un degré d'approximation élevé. Dans [Ioa03], on a proposé une alternative qui utilise un opérateur de déformation afin de réduire le phénomène de propagation d'erreur.

Déformation de l'axe temporel de la HAF Admettons qu'on dispose d'une estimation $\hat{\Phi}_{K,M}$ du coefficient $\Phi_{K,M}$ (1.17), obtenue par la HAF multi-retards. On peut alors définir une fonction de déformation de l'axe temporel $w_M(t)$:

$$t_{w_M} = w_M(t) = \left(\frac{t}{\hat{\Phi}_{K,M}} \right)^{\frac{1}{M}}, \quad (1.34)$$

qui sera associée à l'opérateur de déformation \mathcal{U}_{w_M} . L'application de ce dernier au signal $s(t)$ mono-composante ($K = 1$) (1.17) :

$$s_{w_M}(t_{w_M}) = \mathcal{U}_{w_M}(t, s(t)) = |w'_M(t)|^{\frac{1}{2}} s(w_M(t))$$

permet d'obtenir la relation suivante :

$$s_{w_M}(t_{w_M}) \propto e^{j \sum_{m=0}^{M-1} \Phi_{K,m}(t_{w_M})} e^{j \frac{\Phi_{K,m}}{\Phi_{K,m}} t}. \quad (1.35)$$

On obtient ainsi un signal à phase polynomiale de degré $M - 1$, fonction de la variable t_{w_M} . Le deuxième facteur dans (1.35) est un résidu négligeable. La ré-application de nouveau de la HAF multi-retards au signal $s_{w_M}(t_{w_M})$ fournit une estimation de ϕ_{M-1} . La ré-itération du procédé décrit ci-dessus permet de réduire progressivement le degré de la phase du signal et par la même occasion minimiser l'erreur de propagation [Ioa03].

Nous avons cité dans cette partie les principales méthodes généralisant et améliorant la HAF. Il existe d'autres méthodes développées à partir de la HAF pour estimer des signaux à amplitude aléatoire [IZ01, MZ02] ou polynomiale [ZGS96], des signaux ayant des composantes hyperboliques et exponentielles [GF99].

Cependant, nous pouvons conclure que les performances de toutes les méthodes basées sur la HAF sont limitées par les termes d'interférences et la propagation d'erreur. De ce fait, le degré de la modélisation polynomiale de la phase est toujours limité à 3 ou 4. Ces techniques étant sous optimales (par comparaison avec la procédure de maximum de vraisemblance), certains auteurs suggèrent de les utiliser pour initialiser une procédure de maximum de vraisemblance.

1.3.2 Estimation par la méthode des moindres carrés non-linéaires instantanés

Dans [Ang00], on propose une généralisation de l'estimation par la méthode des moindres carrés aux signaux mono-composantes fortement non-stationnaires et discrets. Cette généralisation est appelée estimation par la méthode des moindres carrés non-linéaires instantanée, notée MCNI.

Considérons un signal bruité $y[n]$ mono-composante dont la partie déterministe est notée $A s[n, \boldsymbol{\theta}]$. Elle est le produit d'une amplitude constante A et d'un signal $s[n, \boldsymbol{\theta}]$. Ce dernier a une amplitude égale à 1 et il dépend non-linéairement d'un paramètre ou d'un vecteur $\boldsymbol{\theta}$:

$$y[n] = A s[n, \boldsymbol{\theta}] + b[n], \quad \text{pour } 0 \leq n \leq N - 1. \quad (1.36)$$

$b[n]$ est un bruit blanc additif Gaussien. Nous rappelons d'abord l'estimation par la méthode des moindres carrés classique de $\boldsymbol{\theta}$ avant d'exposer celle utilisant la MCNI [Ang00].

Méthode des Moindres carrés classique Pour un vecteur de paramètres $\boldsymbol{\theta}$ donné, l'estimateur de l'amplitude par la méthode des Moindres carrés s'écrit :

$$\hat{A} = \arg \min_A \sum_{n=0}^{N-1} |y[n] - A s[n, \boldsymbol{\theta}]|^2 \Leftrightarrow \hat{A} = \frac{\sum_{n=0}^{N-1} \Re\{y[n] s[n, \boldsymbol{\theta}]^*\}}{\sum_{n=0}^{N-1} |s[n, \boldsymbol{\theta}]|^2}, \quad (1.37)$$

où \Re désigne la partie réelle. De même, pour une estimation de l'amplitude \hat{A} donnée, un estimateur de $\boldsymbol{\theta}$ par la méthode des moindres carrés est :

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \sum_{n=0}^{N-1} \left| y[n] - \hat{A} s[n, \boldsymbol{\theta}] \right|^2. \quad (1.38)$$

Après plusieurs itérations successives alternant la résolution des équations (1.37) et (1.38), on devrait obtenir une estimation de A et de $\boldsymbol{\theta}$. Le problème c'est que souvent la fonction à minimiser dans (1.38) est non-linéaire par rapport au vecteur $\boldsymbol{\theta}$. L'utilisation des algorithmes d'optimisation classiques [Bro70] ne garantissent pas la convergence, ce qui la rend inutilisable dans le contexte des signaux fortement non-stationnaires dont la phase est polynomiale. Dans la suite, on introduit la version instantanée et non-linéaire des moindres carrés.

Méthode des Moindres carrés instantanée Dans [Ang00], l'auteur a proposé un algorithme de "recherche aveugle". L'idée clef est d'estimer sur des fenêtres glissantes l'amplitude et la phase du signal. Cet algorithme alterne les étapes suivantes :

- D'abord, on se fixe un instant n , qui correspond à une observation $y[n]$ (1.36).
- Ensuite, on sélectionne une suite croissante de séquences d'échantillons centrées autour de n . Une séquence d'observations de longueur $N_i = 2L_i + 1$ et centrée autour de n est définie par $\mathbf{y}_{N_i} = [y[n - L_i], y[n - L_i], \dots, y[n], \dots, y[n + L_i]]^T$. Les séquences obtenues vérifient $\mathbf{y}_{N_1} \subset \mathbf{y}_{N_2} \subset \dots \subset \mathbf{y}_{N_p}$ et sont de longueurs respectives $N_1 < N_2 < \dots < N_p < N$.
- Sur chaque séquence \mathbf{y}_{N_i} , on détermine une estimation de A et $\boldsymbol{\theta}$ par la méthode des moindres carrés classique. Cette estimation est ré-utilisée par la suite pour initialiser la méthode des moindres carrés classique sur la séquence suivante $\mathbf{y}_{N_{i+1}}$. A la fin de cette étape *i.e.* $i = p$, on obtient une estimation de l'amplitude et du paramètre $\boldsymbol{\theta}$ relatifs à l'instant n .

Les estimations par la méthode des moindres carrés classiques de A et de $\boldsymbol{\theta}$, calculées sur la séquence \mathbf{y}_{N_i} centrée sur l'instant n , sont obtenues *via* les équations suivantes :

$$\hat{A}_n = \arg \min_A \sum_{l=-L}^L \left| y[n+l] - A s[n+l, \hat{\boldsymbol{\theta}}_{n-1}] \right|^2 \quad \text{et} \quad \hat{\boldsymbol{\theta}}_n = \arg \min_{\boldsymbol{\theta}} \sum_{l=-L}^L \left| y[n+l] - \hat{A}_n s[n+l, \boldsymbol{\theta}] \right|^2. \quad (1.39)$$

Ces étapes sont reproduites pour tout n allant de 0 à $N - 1$. La dépendance de ces estimations par rapport au temps n est à l'origine de l'appellation MCNI.

Cette façon de procéder "localement" sur des séquences croissantes, on l'explique par le fait qu'avec les techniques classiques d'optimisation [Bro70], il est plus simple de trouver le minimum global d'une fonction non-linéaire sur des fenêtres petites que sur des grandes. En effet, quand la longueur devient grande, le nombre de *minima* locaux augmente et la convergence vers le minimum global n'est plus assurée avec ces techniques. Or, comme il est préférable d'utiliser de grandes fenêtres pour améliorer les performances statistiques de

l'estimateur employé, un compromis sur leurs longueurs est donc nécessaire pour obtenir la meilleure estimation. L'application du MCNI à un signal à phase polynomiale de degré inférieur ou égal à 4, permet d'obtenir une meilleure performance par rapport à la HAF et le PHAF comme l'illustre la figure 1.7 [Ang00].

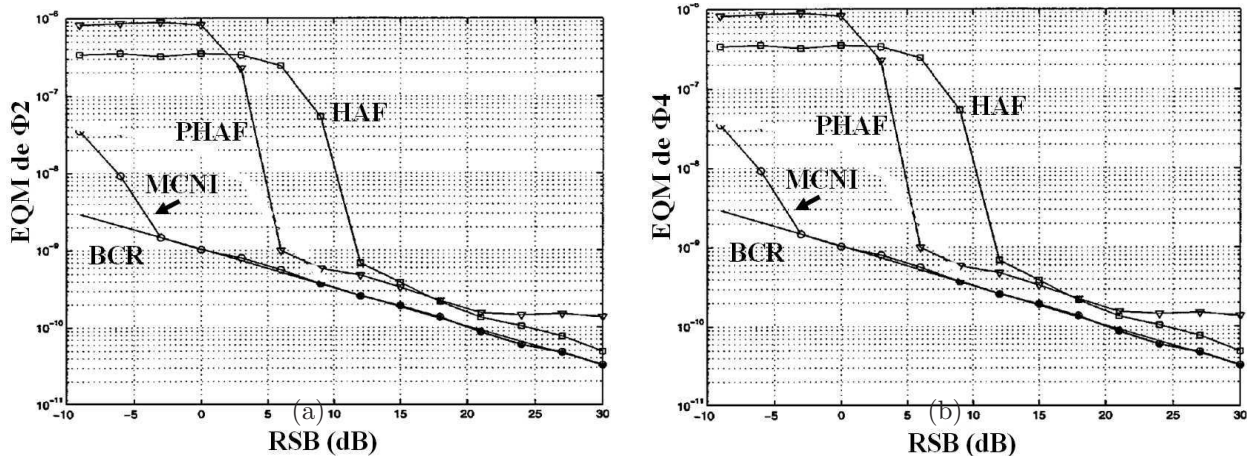


FIG. 1.7 – Erreur Quadratique Moyenne (EQM) dans l'estimation de $\Phi_{1,2}$ et $\Phi_{1,4}$ d'un signal mono-composante à phase polynomiale de degré 4 et à amplitude constante; le nombre d'échantillons est égal à 200 : comparaison entre la technique $(-o-)$ MCNI, $(-\diamond-)$ le HAF et $(-\triangle-)$ le PHAF. Ces figures sont extraites de [Ang00].

La méthode MCNI peut être interprétée comme étant un compromis entre l'optimalité et la complexité de calcul. En effet, en présence d'un bruit gaussien et en utilisant une fenêtre englobant toute les observations, elle fournit théoriquement une estimation au sens du maximum de vraisemblance. Toutefois, outre le fait qu'en pratique la minimisation de (1.39) sur la totalité de la durée temporelle du signal reste très compliquée et la convergence de la MCNI n'est pas garanti, il est difficile de programmer cette méthode. De plus, l'étendre aux signaux multi-composantes est très complexe.

Dans la suite, nous présentons une méthode différente, basée sur la maximisation de la vraisemblance qui est donc équivalente à la méthode des Moindres carrés en présence du bruit gaussien, mais qui réalise l'estimation différemment.

1.3.3 Maximum de vraisemblance et échantillonnage pondéré

Dans [SK02], le signal considéré est formé de composantes multiples modulées linéairement en fréquence. La phase de chaque composante est modélisée par un polynôme de degré 2. Le but de la méthode proposée est d'estimer les paramètres de chaque phase en maximisant la fonction de vraisemblance. Cette dernière étant non-linéaire, sa maximisation directe est très compliquée. Pour résoudre ce problème, une technique de Monte Carlo basée sur l'échantillonnage pondéré "*Importance Sampling*" a été employé [SK02].

Fonction de vraisemblance Le modèle du signal défini en temps discret et son écriture sous forme matricielle sont :

$$y[n] = \sum_{k=1}^K A_k e^{2\pi j(\Phi_{k,1} n + \Phi_{k,2} n^2)} + b[n] \quad \text{et} \quad \mathbf{y} = \mathbf{H} \mathbf{A} + \mathbf{b}. \quad (1.40)$$

K est le nombre de composantes qui sont modulées linéairement et $b[n]$ est un bruit blanc Gaussien de variance σ^2 . Les autres notations sont définies par :

$$\begin{aligned} \mathbf{y} &= [y[0], y[1], \dots, y[N-1]]^\top, & \mathbf{b} &= [b[0], b[1], \dots, b[N-1]]^\top, \\ \mathbf{A} &= [A_1, A_2, \dots, A_K]^\top, & \mathbf{H} &= [\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_K], \\ \mathbf{H}_k &= \left[e^{2\pi j(\Phi_{k,1} 0 + \Phi_{k,2} 0^2)}, \dots, e^{2\pi j(\Phi_{k,1} (N-1) + \Phi_{k,2} (N-1)^2)} \right]^\top. \end{aligned}$$

Soient $\Phi_1 = [\Phi_{1,1}, \dots, \Phi_{K,1}]^\top$ et $\Phi_2 = [\Phi_{1,2}, \dots, \Phi_{K,2}]^\top$ les deux vecteurs correspondant aux paramètres de phases qu'on souhaite estimer. En tenant compte de la nature Gaussienne du bruit blanc $b[n]$, la fonction de vraisemblance qui garantit l'optimalité des estimateurs est définie par :

$$\ell_{MV}(\mathbf{y} | \mathbf{A}, \Phi_1, \Phi_2, \sigma^2) = \frac{1}{\pi^N \sigma^{2N}} e^{-\frac{1}{\sigma^2} (\mathbf{y} - \mathbf{H} \mathbf{A})^\dagger (\mathbf{y} - \mathbf{H} \mathbf{A})}. \quad (1.41)$$

Le symbole \dagger désigne la transposée conjuguée.

Estimateurs au sens du maximum de vraisemblance Les estimateurs de \mathbf{A} , Φ_1 et Φ_2 sont déterminés, sous l'hypothèse du bruit blanc Gaussien, par :

$$\{\hat{\mathbf{A}}, \hat{\Phi}_1, \hat{\Phi}_2\} = \arg \max_{\mathbf{A}, \Phi_1, \Phi_2} \ell_{MV}(\mathbf{y} | \mathbf{A}, \Phi_1, \Phi_2, \sigma^2) = \arg \min_{\mathbf{A}, \Phi_1, \Phi_2} (\mathbf{y} - \mathbf{H} \mathbf{y})^\dagger (\mathbf{y} - \mathbf{H} \mathbf{y}). \quad (1.42)$$

Les estimateurs ainsi obtenus sont donc équivalents à ceux obtenus par la méthode des moindres carrés (cf. paragraphe 1.3.2).

Afin de résoudre le problème d'estimation (1.42), ce dernier a été scindé en deux sous problèmes indépendants [SK02] :

- Dans le premier, l'estimation du vecteur amplitude \mathbf{A} se fait séparément de l'estimation de Φ_1 et Φ_2 . Comme la fonction de vraisemblance est linéaire par rapport aux paramètres d'amplitude, une estimation $\hat{\mathbf{A}}$ de \mathbf{A} est obtenue analytiquement à Φ_1 et Φ_2 fixés :

$$\hat{\mathbf{A}} = (\mathbf{H}^\dagger \mathbf{H})^{-1} \mathbf{H}^\dagger \mathbf{y}. \quad (1.43)$$

- Dans le deuxième, on remplace dans (1.42) \mathbf{A} par l'expression de $\hat{\mathbf{A}}$ (1.43), ce qui permet de transformer le problème d'optimisation (1.42) :

$$\{\hat{\Phi}_1, \hat{\Phi}_2\} = \arg \max_{\Phi_1, \Phi_2} \ell_{MVC}(\hat{\Phi}_1, \hat{\Phi}_2) \quad \text{avec} \quad \ell_{MVC}(\hat{\Phi}_1, \hat{\Phi}_2) = \mathbf{y}^\dagger \mathbf{H} (\mathbf{H}^\dagger \mathbf{H})^{-1} \mathbf{H}^\dagger \mathbf{y}. \quad (1.44)$$

La fonction $\ell_{MVC}(\hat{\Phi}_1, \hat{\Phi}_2)$ est appelée “fonction de vraisemblance compressée”. Elle est non-linéaire et multi-dimensionnelle par rapport aux paramètres de phases et sa maximisation par une méthode classique, comme Gauss Newton [Bro70] ou par une recherche à travers une grille de solutions, est très difficile.

Toutefois, il a été démontré que si une fonction multi-dimensionnelle possède plusieurs *maxima* locaux mais n’a qu’un seul maximum global, alors une expression approchée de ce dernier peut être obtenue [Pin62].

Dans le cas de $\ell_{MVC}(\hat{\Phi}_1, \hat{\Phi}_2)$, l’expression de son maximum global s’écrit pour i égal à 1 ou 2 :

$$\hat{\Phi}_i = \lim_{\rho \rightarrow +\infty} \int \int \Phi_i \ell_{MVCN}(\Phi_1, \Phi_2) d\Phi_1 d\Phi_2 \quad \text{avec} \quad \ell_{MVCN}(\Phi_1, \Phi_2) = \frac{e^{\rho \ell_{MVC}(\Phi_1, \Phi_2)}}{\int \int e^{\rho \ell_{MVC}(\Phi_1, \Phi_2)} d\Phi_1 d\Phi_2}. \quad (1.45)$$

La limite indique que si cette quantité est atteinte pour une certaine valeur $\rho \in \mathbb{R}_+$, alors elle l’est pour toutes les autres valeurs supérieures de ρ .

Etant donné que la matrice $(\mathbf{H}^\dagger \mathbf{H})^{-1}$ est une matrice définie positive, la fonction de vraisemblance compressée et normalisée, ℓ_{MVCN} , peut être considérée comme une densité de probabilité conjointe de deux vecteurs aléatoires Φ_1 et Φ_2 . Dès lors, les techniques de simulations de Monte Carlo [BS94, JJDS02, Rob06] peuvent être employées et permettront d’éviter l’évaluation compliquée de l’intégrale multi-dimensionnelle (1.45).

Echantillonnage pondéré Dans [SK02], on a employé l’échantillonnage pondéré qui est une technique non itérative où les paramètres du modèle sont assimilés à des variables aléatoires [BS94, JJDS02, Rob06]. Cette technique est basée sur le principe suivant :

$$\int h(x)P(x)dx = \int h(x)\frac{P(x)}{g(x)}g(x)dx = \mathbb{E}_{g(x)} \left[h(x)\frac{P(x)}{g(x)} \right], \quad (1.46)$$

$P(x)$ étant une densité de probabilité de la variable aléatoire x . $g(x)$ est une fonction choisie strictement positive et satisfaisant tous les critères d’une densité de probabilité de x . Elle doit être également une fonction dont il est simple de générer des échantillons. L’intégrale (1.46) peut s’exprimer alors sous la forme d’une espérance mathématique, notée \mathbb{E} , de la quantité $h(x)\frac{P(x)}{g(x)}$. Cette dernière est alors approchée par :

$$\mathbb{E}_{g(x)} \left[h(x)\frac{P(x)}{g(x)} \right] \approx \frac{1}{R} \sum_{r=1}^R h(x_r)\frac{P(x_r)}{g(x_r)}. \quad (1.47)$$

où x_r est une réalisation de x distribuée selon la loi $g(x)$ et R est le nombre total de réalisations. De préférence, $g(x)$ doit ressembler à $P(x)$ pour réduire la variance de l’estimateur défini par (1.47).

Dans le cas du modèle (1.40), l’échantillonnage pondéré revient à écrire pour i égal à 1

ou 2 :

$$\hat{\Phi}_i = \frac{1}{R} \sum_{r=1}^R \Phi_{i_r} \frac{\ell_{MVGN}(\Phi_{1_r}, \Phi_{2_r})}{g_{Norm}(\Phi_{1_r}, \Phi_{2_r})}, \quad (1.48)$$

si on considère g la fonction échantillonnage suivante :

$$g(\Phi_1, \Phi_2) \propto e^{\mathbf{y}^\dagger (\mathbf{H} \frac{\mathbb{I}}{N} \mathbf{H}^\dagger) \mathbf{y}} \quad \text{et} \quad g_{Norm}(\Phi_1, \Phi_2) = \frac{g(\Phi_1, \Phi_2)}{\int \int g(\Phi_1, \Phi_2) d\Phi_1 d\Phi_2}, \quad (1.49)$$

\mathbb{I} est la matrice identité. En comparant (1.44) et (1.49), on constate que le choix de g revient à considérer $(\mathbf{H}^\dagger \mathbf{H})^{-1}$ égale à l'identité. Ce choix se justifie grâce à la propriété d'orthogonalité des fonctions exponentielles complexes. En effet, pour plusieurs valeurs différentes de Φ_1 et Φ_2 , la quantité $(\mathbf{H}^\dagger \mathbf{H})^{-1}$ s'annule. Cela revient à supposer que les générations des candidats Φ_1 et Φ_2 sont indépendantes et il faudrait en tenir compte lors de l'évaluation numérique.

La maximisation de la vraisemblance en utilisant l'échantillonnage pondéré a permis d'estimer des modulations linéaires proches mais séparées dans le plan temps-fréquence, en présence d'un faible *RSB*. Le résultat obtenu est optimal au sens du maximum de vraisemblance. De plus, la convergence de cette méthode non-itérative est prouvée.

En revanche, le fait que dans [SK02] les auteurs ont utilisé un grille pour évaluer g et (1.48), rend difficile la généralisation à des degrés d'approximation polynomiale de la phase supérieurs à 2. Il en est de même si les phases initiales des composantes ne sont pas nulles comme le laisser supposer (1.40). De plus, le paramétrage de ρ (1.45) est très compliqué.

Nous pensons que l'extension de cette méthode à l'estimation des modulations sinusoïdales, où le nombre de paramètres à estimer est réduit, peut être envisageable sous réserve que la matrice $(\mathbf{H}^\dagger \mathbf{H})^{-1}$ ne devienne singulière.

1.4 Autres techniques

Il existe d'autres techniques différentes de celles détaillées dans ce chapitre. Nous nous contentons ici d'énumérer quelques unes sans rentrer dans les détails.

Décomposition modale empirique Elle considère les signaux à l'échelle de leurs oscillations locales, sans que celles-ci soient nécessairement harmoniques au sens de Fourier [HSL⁺98, FP03, BCSG04].

Autrement dit, chaque signal est considéré comme étant la somme d'oscillations rapides et d'oscillations lentes déterminées en itérant les étapes suivantes : on identifie les *maxima* et les *minima* locaux du signal ; puis on en déduit une enveloppe supérieure et une enveloppe inférieure par interpolation (spline cubique) ; on calcule ensuite la moyenne des enveloppes et on la soustrait au signal. La décomposition modale empirique réalise de manière automatique une décomposition en sous-bandes très proche de ce que donnerait

une analyse multi-résolution de type ondelettes.

Malheureusement, avec cette méthode, un signal sinusoïdal échantillonné n'est pas une fréquence modale intrinsèque. En effet, d'autres fréquences apparaissent lors de sa décomposition. De plus, si par exemple le signal est composé de deux harmoniques de fréquences proches, la décomposition fournit un seul signal modulé en amplitude et en fréquence. En présence d'un faible RSB, cette méthode ne fonctionne pas.

Méthodes bayésiennes ce sont des méthodes d'estimation de signaux à phase polynomiale basées sur des approches Bayésiennes [TVF97]. Elles utilisent les techniques de simulations de Monte Carlo par Chaînes de Markov. Ces techniques seront détaillées dans le chapitre 4.

Filtrage particulaire C'est une nouvelle technique permettant le suivi en temps réel de l'évolution de plusieurs fréquences instantanées du signal. Ce dernier est alors décomposé, en ligne, en une somme de composantes fréquentielles [CPD03, DDI05].

Pour réaliser le suivi, cette technique utilise les méthodes de Monte Carlo séquentielles, qui combinent la puissance des méthodes d'échantillonnage de Monte Carlo avec l'inférence Bayésienne. Le filtrage particulaire permet de prendre en compte les non-linéarités du modèle d'observation et les changements de dimension de l'état, *i.e.* l'apparition (naissance) ou la disparition (mort) d'une composante. Ainsi, il fournit une estimation de nombre de composantes présentes dans le signal.

Le filtrage particulaire a été appliqué à des signaux modulés linéairement en fréquence et a fourni de bons résultats. Toutefois, l'extension de cette technique aux signaux modulés non-linéairement à la fois en amplitude et en fréquence nécessite la détermination d'une densité de probabilité de simulation très compliquée.

Nous citons également la technique "bootstrap" [ZA97, Abu05] qui fournit des intervalles de confiance des paramètres estimés, et la technique du filtrage de Kalman [Kaa98, Ben04] qui ne permet pas le suivi de fortes modulations de fréquence (variations rapides) en présence d'un faible RSB.

1.5 Conclusion

Nous avons détaillé dans ce chapitre quelques méthodes telles que les RTF, la HAF, le PHAF et le maximum de vraisemblance, qui étaient proposées dans la littérature dans le but d'analyser et d'estimer les modulations d'amplitude et de fréquence. Nous avons rappelé leurs limitations en présence de signaux fortement non-stationnaires.

Dans les chapitres suivants, nous proposons une nouvelle approche basée sur la modélisation et l'estimation locale de la variation des amplitudes et des fréquences instantanées d'un signal fortement non-stationnaires.

Chapitre 2

Signal mono-composante non-stationnaire court - Modélisation

Résumé

Dans ce chapitre, nous définissons le modèle attribué à un signal mono-composante, non-stationnaire et court. Ce modèle sera le point de départ des méthodes que nous développons dans la suite de cette thèse.

Nous proposons de modéliser l'amplitude et la fréquence instantanées du signal court par des fonctions polynomiales, représentées dans une base. Nous discutons le choix de la base polynomiale et son influence sur le processus d'estimation.

Vu qu'une base polynomiale orthogonale contribue à la réduction du couplage entre les paramètres du modèle, mais que l'orthogonalité n'est pas conservée par les bases définies en temps continu lors du passage en temps discret, nous avons calculé une base polynomiale discrète en utilisant la procédure de Gram-Schmidt.

Mots clefs

Signaux courts non-stationnaires, modélisation polynomiale, estimation paramétrique.

Sommaire

2.1	Introduction	43
2.2	Modèle du signal	44
2.2.1	Définition du modèle du signal - Unicité	44
2.2.2	Modèle à amplitude et à fréquence polynomiales	45
2.2.2.1	Formulation	46
2.2.2.2	Positionnement et nouveauté dans le modèle	47
2.3	Exemples de bases polynomiales	48
2.3.1	Base polynomiale canonique	48
2.3.2	Polynômes de Legendre	48
2.3.3	Base polynomiale orthonormale en temps discret	49
2.4	Méthodes d'estimation	51
2.5	Conclusion	53

2.1 Introduction

Dans ce chapitre, nous étudions des signaux non-stationnaires courts mono-composantes et nous nous intéressons, en particulier, à ceux modulés non-linéairement à la fois en amplitude et en fréquence. Le terme “court” désigne des signaux ne possédant qu’un faible nombre d’échantillons temporels.

Un tel signal de 51 échantillons est tracé sur la figure 2.1(a). Son spectrogramme, représenté sur la figure 2.1(b), a une résolution temps-fréquence médiocre à cause du faible nombre d’échantillons. Quant à la distribution de Wigner-Ville, illustrée sur la figure 2.1(c), on voit apparaître les termes d’interférences dûs à la présence de modulations non-linéaires. Les deux figures 2.1(b) et (c) montrent une mauvaise localisation de la modulation de fréquence au début et à la fin du temps.

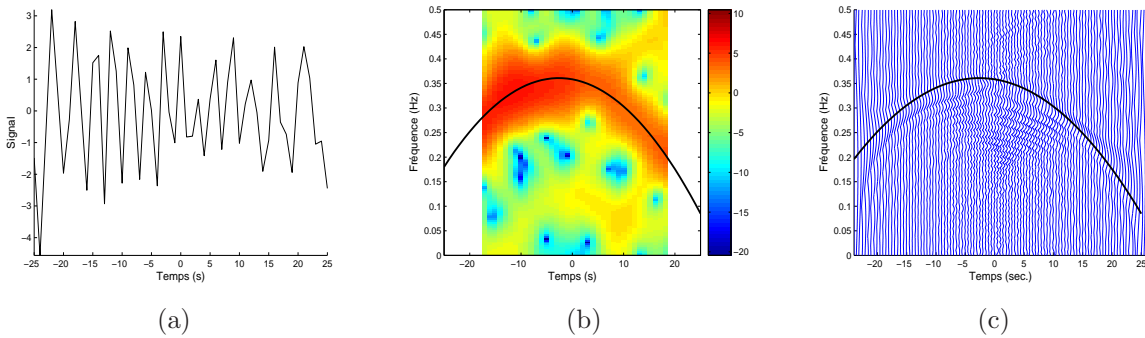


FIG. 2.1 – (a) Signal simulé ayant une modulation quadratique en amplitude et en fréquence et formé de 51 échantillons. La fréquence d’échantillonnage est égale à 1 Hz et le $RSB = 10\text{ dB}$. Sa fréquence instantanée (—) est superposée avec : (b) le spectrogramme calculé avec $N_{TF} = 256$, un recouvrement de 99% et une fenêtre de Hanning de 16 points, (c) la distribution de Wigner-Ville.

Les RTF présentant des limitations dans le contexte des signaux courts et non-stationnaires, l’analyse paramétrique s’avère alors plus adaptée. Dans cet objectif, une structure prédéterminée, jouant le rôle d’une information *a priori*, est attribuée au signal. Chacune des modulations de l’amplitude, de la fréquence et de la phase est également modélisée par une fonction qui varie au cours du temps. Il en résulte donc un modèle du signal.

Plusieurs choix de modèle sont possibles mais, dans ce chapitre, nous étudions et discutons celui basé sur l’approximation polynomiale de l’amplitude et de la fréquence instantanées. La représentation de cette approximation dans une base garantit l’unicité de la décomposition de l’amplitude et de la fréquence. Nous présentons quelques bases polynomiales.

Les techniques envisagées dans l’estimation des paramètres du modèle, c’est-à-dire la maximisation de la fonction de vraisemblance et l’estimation Bayésienne, sont décrites succinctement. Elles seront abordées et analysées plus en détails dans les chapitres suivants 3 et 4.

2.2 Modèle du signal

Soit $s[n]$ un signal complexe, court, non-stationnaire et mono-composante, dont l'amplitude $A[n]$, la fréquence $F[n]$ et la phase $\Phi[n]$ sont des fonctions variant suivant un temps discret, noté n . Sur la figure 2.2, ces fonctions sont représentées pour le signal tracé sur la figure 2.1(a). Le temps n est référencé par rapport au centre de la fenêtre d'observation.

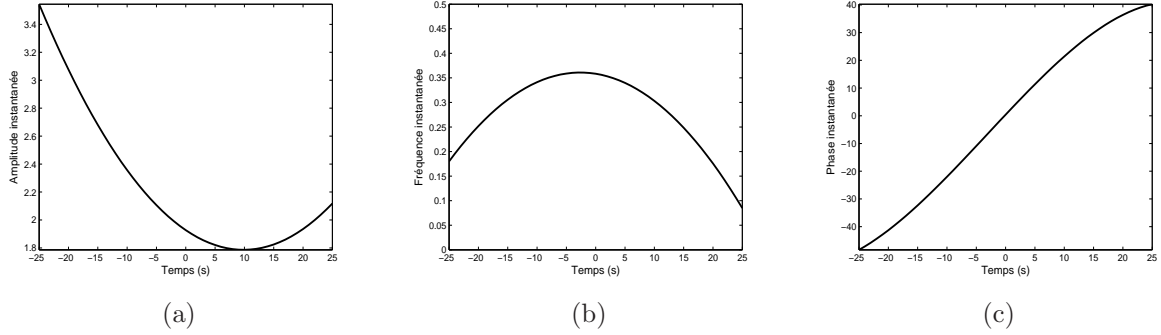


FIG. 2.2 – (a) Amplitude instantanée du signal tracé sur la figure 2.1(a), (b) sa fréquence instantanée et (c) sa phase instantanée.

Dans la suite, nous considérons un signal $s[n]$ ayant un faible nombre d'échantillons temporels, noté $N + 1$ et compris entre 15 et 61. N sera considéré pair pour simplifier les calculs et l'implémentation des algorithmes. Nous supposons également que $s[n]$ est noyé dans du bruit $b[n]$ complexe, additif, Gaussien, centré et de variance inconnue σ_b^2 . Les observations $y[n]$ sont, alors, modélisées par :

$$y[n] = s[n] + b[n] \quad \text{pour} \quad \frac{-N}{2} \leq n \leq \frac{N}{2}. \quad (2.1)$$

Dans les paragraphes suivants, nous allons définir une structure pour le signal $s[n]$ et attribuer un modèle à chacune des fonctions $A[n]$, $F[n]$ et $\Phi[n]$, tout en justifiant nos choix. Le cas des signaux réels sera abordé en annexes A.2.

2.2.1 Définition du modèle du signal - Unicité

Une représentation complexe du signal $s[n]$, dont l'amplitude et la phase instantanées sont notées respectivement $A[n]$ et $\Phi[n]$, peut être définie par :

$$s[n] = A[n] e^{j \Phi[n]}. \quad (2.2)$$

Dans [CLV99], deux définitions possibles de $A[n]$ et $\Phi[n]$ sont discutées.

Première définition Elle est par convention la plus employée et identifie la phase et l'amplitude comme étant, respectivement, l'argument et le module du signal complexe :

$$\tan(\Phi[n]) = \frac{\Im \{s[n]\}}{\Re \{s[n]\}} \quad \text{et} \quad A[n] = \sqrt{\Re \{s[n]\}^2 + \Im \{s[n]\}^2} = |s[n]|, \quad (2.3)$$

où \Re et \Im désignent, respectivement, la partie réelle et la partie imaginaire. La fréquence instantanée est définie comme étant la dérivée de la phase normalisée par $\frac{1}{2\pi}$.

L'équation (2.3) détermine la phase instantanée du signal à *modulo* 2π près et impose que l'amplitude soit positive. Si cette dernière s'annule pour un ensemble fini d'instant, la phase (2.3) peut subir des sauts et, par conséquent, la fréquence peut présenter des discontinuités en ces instants.

Deuxième définition Elle est moins utilisée mais elle est parfois plus adaptée à certains contextes (signaux radio). Elle impose que la fréquence instantanée soit bornée mais elle n'impose aucune restriction sur le signe de l'amplitude définie par :

$$A[n] = s[n] e^{-j \Phi[n]} = \Re \{s[n]\} \cos(\Phi[n]) + \Im \{s[n]\} \sin(\Phi[n]), \quad (2.4)$$

où la phase $\Phi[n]$ est déterminée par :

$$\Phi[n] = \arctan \left(\frac{\Im \{s[n]\}}{\Re \{s[n]\}} \right) \in \left[-\frac{\pi}{2}, \frac{\pi}{2} \right]. \quad (2.5)$$

Problème d'ambiguïté et unicité du modèle Dans les deux définitions, un problème d'ambiguïté peut surgir [CLV99] si le signal complexe s'annule pour un nombre fini d'instant. En effet, il est difficile de savoir si l'amplitude et la phase présente des discontinuités pour ces instants. Par conséquent, il n'y a aucune garantie sur l'unicité du modèle (2.2).

Dans [CLV99], des considérations physiques permettent d'éviter ce problème d'ambiguïté et de garantir l'unicité du modèle (2.2) :

- Dans la première définition, il suffit d'imposer que l'amplitude soit strictement positive et que la phase ne présente pas de discontinuités. Si, de plus, le signal temporel ne présente pas de ruptures, on peut imposer la continuité de l'amplitude.
- Dans la seconde définition, il suffit d'imposer que la fréquence et l'amplitude soient tous les deux continues et ne présentent pas de singularités.

Dans la suite de la thèse, nous allons adopter la première définition, en imposant les contraintes garantissant l'unicité du modèle (2.2) [CLV99, RB74], à savoir une amplitude strictement positive, une phase continue et différentiable et une fréquence sans discontinuités vérifiant le théorème de Shannon et l'équation suivante :

$$0 < F[n] < \frac{F_e}{2}, \quad (2.6)$$

où F_e est la fréquence d'échantillonnage du signal.

2.2.2 Modèle à amplitude et à fréquence polynomiales

Dans la littérature, plusieurs fonctions peuvent représenter l'amplitude et la phase instantanées d'un signal non-stationnaire : polynomiales, sinusoidales, puissance non entière,

aléatoire, exponentielle, *etc* (cf. annexe A.3). En particulier, la modélisation polynomiale, motivée par le théorème de Weierstrass (Théorème 1), a été très utilisée (cf. paragraphe 1.3).

Dans le cadre des signaux courts et non-stationnaires, l'approximation polynomiale nous semble intéressante et parfaitement adaptée à ce contexte pour les raisons suivantes :

- D'abord, elle permet de caractériser le degré de la non-linéarité de la modulation de l'amplitude et de la fréquence.
- Deuxièmement, étant donné la taille des signaux courts considérés ($N + 1$ compris entre 15 à 61 échantillons), le nombre des paramètres du modèle ne devrait pas être élevé sous peine de se retrouver dans un problème sous-déterminé. Une approximation polynomiale de degré faible sera donc adaptée à ce type de signaux et fournira un modèle dont le nombre de paramètres est relativement réduit.

Nous nous limitons, le plus souvent à un degré d'approximation égal à 3. L'effet de cette limitation sera analysé dans le cas d'une forte non-stationnarité dans le chapitre 5.

Dans la suite, nous présentons le modèle polynomial défini en temps discret et discutons le choix de la base polynomiale de représentation.

2.2.2.1 Formulation

Considérons un ensemble de polynômes défini en temps discret par $\{g_m[n]\}_{0 \leq m \leq M}$, où m désigne le degré du polynôme $g_m[n]$ et $M \in \mathbb{N}$.

Ces polynômes forment une base de l'espace vectoriel $\mathbb{R}_M[n]$ de polynômes de degré inférieur ou égal à M , à coefficients dans \mathbb{R} et définis en temps discret $n = -\frac{N}{2}, -\frac{N}{2} + 1, \dots, \frac{N}{2}$. Cet espace est muni du produit scalaire discret suivant :

$$\langle g_m, g_h \rangle_{\mathcal{D}} = \sum_{n=-\frac{N}{2}}^{\frac{N}{2}} g_m[n] g_h[n] \quad \forall \{m, h\} \in \{0, 1, \dots, M\}^2. \quad (2.7)$$

Nous modélisons l'amplitude $A[n]$ et la fréquence instantanée $F[n]$ du signal $s[n]$ par :

$$\begin{aligned} A[n] &= \sum_{m=0}^{M_A} a_m g_m[n], \\ F[n] &= \sum_{m=0}^{M_F} f_m g_m[n], \end{aligned} \quad \text{pour } -\frac{N}{2} \leq n \leq \frac{N}{2}. \quad (2.8)$$

M_A et M_F désignent les degrés respectifs de l'approximation polynomiale de $A[n]$ et de celle de $F[n]$, et vérifient les relations suivantes :

$$0 \leq M_A \leq M \quad \text{et} \quad 0 \leq M_F \leq M. \quad (2.9)$$

Quant à a_m et f_m , ils représentent respectivement les paramètres de l'amplitude et de la fréquence ; leur unicité est garantie par l'utilisation de la base polynomiale.

La phase instantanée $\Phi[n]$ du signal est déterminée par une intégration numérique de

la fréquence. Nous proposons de l'écrire sous la forme suivante :

$$\Phi[n] = \varphi_0 + 2\pi \left(\sum_{k=\frac{-N}{2}}^n F[k] - \sum_{k=\frac{-N}{2}}^0 F[k] \right), \quad (2.10)$$

où φ_0 est la phase initiale référencée au centre de la fenêtre d'observation ($\varphi_0 = \Phi[0]$).

Les équations (2.8) et (2.10) permettent de décrire analytiquement la non-linéarité de l'amplitude, de la phase et de la fréquence et de caractériser, ainsi, la non-stationnarité du signal. Notons que le modèle, défini par ces équations, est linéaire par rapport aux paramètres de l'amplitude $\{a_0, a_1, \dots, a_{M_A}\}$. En revanche, il est fortement non-linéaire par rapport à ceux de la fréquence $\{f_0, f_1, \dots, f_{M_F}\}$ et de la phase initiale φ_0 .

Processus d'estimation Le processus d'estimation consiste à déterminer le vecteur de paramètres $\boldsymbol{\theta}$, de dimension $M_A + M_F + 3$ et défini par :

$$\begin{aligned} \boldsymbol{\theta} &= [\theta_0, \theta_1, \dots, \dots, \theta_{M_A+M_F+2}]^\top, \\ &= [\varphi_0, a_0, \dots, a_{M_A}, f_0, \dots, f_{M_F}]^\top, \end{aligned} \quad (2.11)$$

sous les contraintes suivantes :

$$\begin{aligned} A[n] > 0, \quad 0 < F[n] < \frac{F_e}{2}, \quad N + 1 > M_A + M_F + 3 \\ \text{et } \Phi[n] \text{ ne présentant pas de discontinuités.} \end{aligned} \quad (2.12)$$

2.2.2.2 Positionnement et nouveauté dans le modèle

Le modèle (2.8) et (2.10) diffère en trois points de ceux proposés dans la littérature. En effet, dans [PPF93, FF95, ZGS96, BSG98, IZ01, SK02, ICQ04], la modélisation polynomiale est appliquée à la phase. Ensuite, une estimation de la fréquence est obtenue par une dérivation de la phase. Dès lors, une erreur d'estimation sur le m -ième coefficient de la phase entraîne une erreur m -fois plus grande sur le coefficient d'ordre $m - 1$ de la fréquence. Dans le contexte des signaux courts, la dégradation de l'estimation peut être importante ; nous avons donc décidé de modéliser directement la fréquence (2.8) au lieu de la phase.

Contrairement à [PPF93, FF95, ZGS96, BSG98, IZ01, SK02, ICQ04] où la phase initiale φ_0 est supposée nulle dans tous les exemples pratiques présentés, nous considérons φ_0 à valeur dans $[-\pi, \pi[$ et nous l'estimons afin de reconstruire le signal et sa phase instantanée. De plus, dans l'expression (2.10) de la phase instantanée, nous faisons correspondre l'origine des phases avec le centre de la fenêtre d'observation (*i.e.* $n = 0$). Or, d'après [RB74], cette répartition symétrique des échantillons de part et d'autres du zéro (la référence du temps et l'origine des phases), contribue à minimiser l'erreur de l'estimation.

Enfin, l'amplitude est considérée constante dans [PPF93, FF95, BSG98, ICQ04], elle

est représentée par une variable aléatoire dans [IZ01], et elle est modélisée par une fonction linéaire du temps dans [FP99]. Afin de mieux caractériser la non-stationnarité du signal analysé, nous avons opté pour un modèle non-linéaire de l'amplitude instantanée, permettant de tenir compte de ses variations.

2.3 Exemples de bases polynomiales

Dans la littérature, il existe une diversité de polynômes formant une base [AS64]. Nous citons, par exemple, les polynômes canoniques, Legendre, Bernstein, Chebyshev, Jacobi.

Parmi ces bases polynomiales, nous rappelons celles les plus employées pour représenter les modulations polynomiales de fréquence (2.8), à savoir la base polynomiale canonique et celle de Legendre. Ces deux dernières étant définies en temps continu, nous explicitons leurs expressions que nous discrétisons par la suite.

Soit alors l'espace vectoriel $\mathbb{R}_M[t]$ de polynômes $\mathcal{P}_m(t)$ définis, en temps continu, sur l'intervalle $[T_0, T_1]$, à coefficients dans \mathbb{R} et de degré m inférieur ou égal à M . Cet espace vectoriel est muni du produit scalaire suivant :

$$\langle \mathcal{P}_m, \mathcal{P}_h \rangle = \int_{T_0}^{T_1} \mathcal{P}_h(t) \mathcal{P}_m(t) dt \quad \forall \{m, h\} \in \{0, 1, \dots, M\}^2. \quad (2.13)$$

2.3.1 Base polynomiale canonique

La base canonique, notée $\{\mathcal{P}_m^{Can}(t)\}_{0 \leq m \leq M}$, est une base polynomiale standard de $\mathbb{R}_M[t]$ qui n'est ni orthogonale, ni normalisée. Elle est explicitée par :

$$\mathcal{P}_m^{Can}(t) = t^m \quad \text{pour } t \in \mathbb{R} \text{ et } 0 \leq m \leq M.$$

En temps discret, $\mathcal{P}_m^{Can}(t)$ s'écrit :

$$\mathcal{P}_m^{Can}[n] = n^m \quad \text{pour } n \in \mathbb{Z} \text{ et } 0 \leq m \leq M. \quad (2.14)$$

2.3.2 Polynômes de Legendre

Le polynôme de Legendre standard de degré m , noté $\mathcal{P}_m^{Leg-S}(t)$, est défini sur l'intervalle $[-1, 1]$ à partir de la formule de Rodrigues [AS64] :

$$\mathcal{P}_m^{Leg-S}(t) = \frac{1}{2^m m!} \frac{d^m (t-1)^m}{d t^m} \quad \text{pour } t \in [-1, 1]. \quad (2.15)$$

Ces polynômes (2.15) sont orthogonaux pour le produit scalaire (2.13) puisqu'ils vérifient :

$$\langle \mathcal{P}_m^{Leg-S}, \mathcal{P}_h^{Leg-S} \rangle = \int_{-1}^1 \mathcal{P}_m^{Leg-S}(t) \mathcal{P}_h^{Leg-S}(t) dt = \frac{2}{2m+1} \delta_{m,h}, \quad (2.16)$$

où $\delta_{m,h}$ est le symbole de Kronecker. Leurs expressions sont explicitées dans l'annexe A.4 pour des degrés inférieurs à 5. L'extension de leur définition à un intervalle quelconque $[T_0, T_1]$, est obtenue par :

$$\mathcal{P}_m^{Leg}(t) = \sqrt{\frac{2}{T_1 - T_0}} \mathcal{P}_m^{Leg-S} \left(\frac{2t - T_1 - T_0}{T_1 - T_0} \right) \quad \text{pour } t \in [T_0, T_1]. \quad (2.17)$$

Outre la propriété de l'orthogonalité (2.13), les $\mathcal{P}_m^{Leg}(t)$ possèdent la caractéristique suivante : la m -ième dérivée du polynôme de Legendre (2.15) est égale à $\frac{1}{2}(m+1) \mathcal{P}_{m-1}^{J(1,1)}(t)$, où $\mathcal{P}_m^{J(\alpha,\beta)}(t)$ désigne le polynôme de Jacobi de degré m et de paramètres $\alpha = \beta$ [Dav75].

Ces deux propriétés ont été exploitées dans [PPF93], où la phase instantanée du signal était décomposée sur la base de Legendre (2.17). Ce qui a permis de simplifier d'une part la dérivation de la phase et donc le calcul de la fréquence instantanée, et d'autre part le calcul des bornes minimales de la variance de l'erreur d'estimation.

Par discrétisation des polynômes de Legendre (2.17), on obtient les polynômes suivants, pour $n = \frac{-N}{2}, \frac{-N}{2} + 1, \dots, \frac{N}{2}$ et pour des degrés allant de 0 à 5 :

$$\begin{aligned} \mathcal{P}_0^{Leg-D}[n] &= \frac{\sqrt{2}}{\sqrt{N}}, & \mathcal{P}_3^{Leg-D}[n] &= \frac{1}{\sqrt{2N}} \left(5 \left(\frac{2n}{N} \right)^3 - \frac{6n}{N} \right), \\ \mathcal{P}_1^{Leg-D}[n] &= \frac{2\sqrt{2}}{N\sqrt{N}} n, & \mathcal{P}_4^{Leg-D}[n] &= \frac{1}{4\sqrt{2N}} \left(35 \left(\frac{2n}{N} \right)^4 - 30 \left(\frac{2n}{N} \right)^2 + 3 \right), \\ \mathcal{P}_2^{Leg-D}[n] &= \frac{1}{\sqrt{2N}} \left(3 \left(\frac{2n}{N} \right)^2 - 1 \right), & \mathcal{P}_5^{Leg-D}[n] &= \frac{1}{4\sqrt{2N}} \left(63 \left(\frac{2n}{N} \right)^5 - 70 \left(\frac{2n}{N} \right)^3 + 15 \left(\frac{2n}{N} - 1 \right) \right). \end{aligned} \quad (2.18)$$

Lors du passage en temps discret, les polynômes de Legendre discrétisés $\{\mathcal{P}_m^{Leg-D}[n]\}$ (2.18) ne vérifient plus la propriété de l'orthogonalité pour le produit scalaire défini par (2.7). Par exemple, pour $\mathcal{P}_0^{Leg-D}[n]$ et $\mathcal{P}_2^{Leg-D}[n]$, on obtient la quantité suivante :

$$\langle \mathcal{P}_0^{Leg-D}, \mathcal{P}_2^{Leg-D} \rangle_{\mathcal{D}} = 2 \frac{N+1}{N^2}. \quad (2.19)$$

Cette erreur (2.19) due à la discrétisation est d'autant plus importante que les signaux sont courts c'est-à-dire N faible. Pour un nombre d'échantillons décroissant de 61 à 15, cette erreur croît de 0.0333 à 0.1531 (cf. figure 2.3(a)).

Comme, dans [FF95], il était démontré qu'une base orthogonale permet de découpler les paramètres du modèle d'un signal mono-composante¹ (2.2), et vu que les polynômes de Legendre discrétisés ne sont pas orthogonaux (2.19), nous avons décidé de calculer une base polynomiale orthonormale discrète. Ceci fera l'objet du prochain paragraphe.

2.3.3 Base polynomiale orthonormale en temps discret

A fin de déterminer une base polynomiale, discrète et orthonormale, notée $\{\mathcal{P}_m^{Orth}[n]\}$, nous avons utilisé la procédure de Gram-Schmidt, ce qui revient à résoudre l'équation :

$$\langle \mathcal{P}_m^{Orth}, \mathcal{P}_h^{Orth} \rangle_{\mathcal{D}} = \delta_{m,h} \quad (m, h) \in \mathbb{N}^2. \quad (2.20)$$

¹Cette propriété est très importante, car contribue dans l'amélioration de l'estimation. Elle sera détaillée dans le chapitre 5.

Les cinq premiers polynômes ainsi obtenus et définis pour $n = \frac{-N}{2}, \frac{-N}{2} + 1, \dots, \frac{N}{2}$ sont :

$$\begin{aligned} \mathcal{P}_0^{Orth}[n] &= \frac{1}{\sqrt{N+1}}, & \mathcal{P}_3^{Orth}[n] &= \frac{1}{\sqrt{\sum k^6 - \frac{(\sum k^4)^2}{\sum k^2}}} \left(n^3 - \frac{\sum k^4}{\sum k^2} n \right), \\ \mathcal{P}_1^{Orth}[n] &= \frac{2\sqrt{3}}{\sqrt{N(N+1)(N+2)}} n, & \mathcal{P}_4^{Orth}[n] &= c(\gamma n^4 + \lambda n^2 + 1), \\ \mathcal{P}_2^{Orth}[n] &= \frac{6\sqrt{5}}{\sqrt{(N+3)(N+2)(N+1)N(N-1)}} \left(n^2 - \frac{N(N+2)}{12} \right), & \mathcal{P}_5^{Orth}[n] &= d(\mu n^5 + \nu n^3 + n). \end{aligned} \quad (2.21)$$

Les étapes de calcul et les expressions de c , d , γ , λ , μ et ν sont fournies en annexes A.5. Le symbole \sum désigne la somme discrète de $k = -\frac{N}{2}, \dots, \frac{N}{2}$, qui ne dépend que de N . Ces polynômes sont tracés sur la figure 2.3(b).

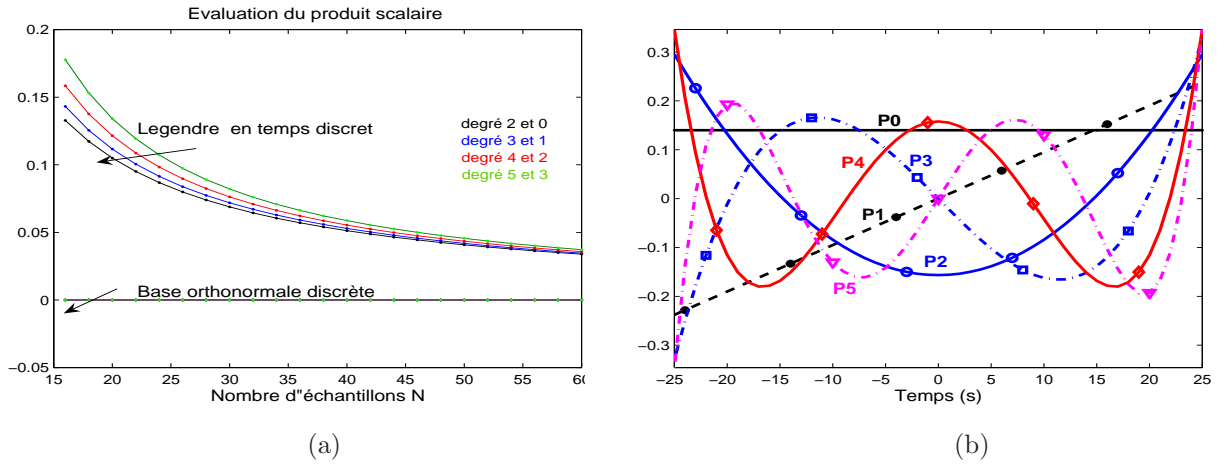


FIG. 2.3 – (a) Evaluation du produit scalaire (2.7) en fonction du nombre d'échantillons, pour les polynômes de Legendre discrétisés (2.18) et pour la base polynomiale discrète et orthonormale (2.21). (b) Représentation temporelle des polynômes de la base orthonormale discrète (2.21).

La normalisation de la base sert à unifier l'ordre de grandeur des paramètres. Ceci est illustré par l'exemple suivant : soit $F[n]$ un polynôme de degré égal à 3, défini pour $n = -\frac{N}{2}, \dots, \frac{N}{2}$. $F[n]$ est décomposé, en premier temps, sur la base canonique (2.14) et en second temps sur la base orthonormale (2.21) :

$$F[n] = 0.1401 + 00.0121 n + 1.0539 \cdot 10^{-4} n^2 - 8.1435 \cdot 10^{-6} n^3, \quad (2.22)$$

$$F[n] = 1.1635 \mathcal{P}_0^{Orth}[n] + 0.9427 \mathcal{P}_1^{Orth}[n] + 0.1458 \mathcal{P}_2^{Orth}[n] - 0.1454 \mathcal{P}_3^{Orth}[n]. \quad (2.23)$$

Les coefficients obtenus dans l'équation (2.23) sont du même ordre de grandeur, comparés à ceux de l'équation (2.22). D'ailleurs, le dernier coefficient de l'équation (2.22) *i.e.* ($-8.1435 \cdot 10^{-6}$) pourrait être du même ordre que les erreurs d'estimation.

Remarque 1 En utilisant cette base (2.21) et en transformant les contraintes sur l'amplitude et la fréquence (2.12) en contraintes sur les paramètres du modèle (2.11), nous pouvons déterminer des bornes supérieures de domaine de définition de ces paramètres.

Par exemple, nous avons calculé les bornes suivantes pour les paramètres de la fréquence :

$$\begin{aligned} |f_0| &\leq \frac{F_e}{2|\beta_{0,0}|} \left(1 + \frac{4|\beta_{2,0}|}{N^2|\beta_{2,2}|}\right), \\ |f_1| &\leq \frac{F_e}{2|\beta_{1,1}|} \left(\frac{1}{N} + \frac{4|\beta_{2,1}|}{N^2|\beta_{2,2}|}\right) \\ |f_2| &\leq \frac{2F_e}{N^2|\beta_{2,2}|}, \end{aligned} \quad (2.24)$$

où les $\beta_{m,h}$ sont définis dans (A.8). Les mêmes expressions sont obtenues pour les paramètres de l'amplitude en substituant $\max_n |s[n]|$ à F_e . De plus, nous imposons que

$$\varphi_0 \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right].$$

Remarque 2 Il existe, par ailleurs, une base discrète polynomiale orthonormale dite base discrète de Legendre [Abu93, Mor69]. Le nom de cette base n'a pas de lien directe avec les polynômes classiques de Legendre (2.15). Le polynôme discret de Legendre de degré m , noté $\{\mathcal{P}_m^{\text{Leg-Orth}}[n]\}$, est défini pour $n = 0, \dots, N+1$ comme suit :

$$\mathcal{P}_m^{\text{Leg-Orth}}[n] = \frac{1}{\mathfrak{c}_m^N} \sum_{l=0}^m (-1)^l (C_m^l) (C_{m+l}^l) \frac{n^{\{l\}}}{(N+1)^{\{l\}}}, \quad (2.25)$$

où $\mathfrak{c}_m^N = \left(\frac{(N+2+m)^{\{m+1\}}}{(2m+1)(N+1)^{\{m\}}}\right)^{\frac{1}{2}}$ est une constante de normalisation, $C_m^l = \frac{m!}{l!(m-l)!}$ est le coefficient binomial et enfin $n^{\{l\}} = n(n-1)\dots(n-l+1)$.

Etant donné la complexité de calcul des coefficients de ces polynômes (2.25), un algorithme rapide a été proposé dans [Abu93] pour les évaluer numériquement.

Pour des degrés faibles, la base discrète de Legendre (2.25) coïncide avec la base orthonormale discrète (2.21) que nous avons calculée et cela à une constante multiplicative près :

$$\mathcal{P}_m^{\text{Orth}}[n] = (-1)^m \mathcal{P}_m^{\text{Leg-Orth}}\left[n + \frac{N}{2}\right], \quad \text{pour } n = -\frac{N}{2}, \dots, \frac{N}{2}. \quad (2.26)$$

En vue d'analyser l'influence de l'orthogonalité et de la normalisation sur le processus de l'estimation (2.11), nous présenterons dans le chapitre 5 une comparaison entre les bases polynomiales suivantes : canonique (2.14), Legendre discrétisée (2.18) et base orthonormale (2.21).

Dans la partie suivante de ce chapitre, nous nous intéressons aux méthodes d'estimation que nous allons utiliser pour déterminer les paramètres du modèle du signal (2.11).

2.4 Méthodes d'estimation

Considérons les notations suivantes :

$$\mathbf{Y} = \left[Y \left[\frac{-N}{2} \right], \dots, Y \left[\frac{N}{2} \right] \right]^T \quad \text{le vecteur aléatoire observé} \quad (2.27)$$

$$\text{et } \mathbf{y} = \left[y \left[\frac{-N}{2} \right], \dots, y \left[\frac{N}{2} \right] \right]^T \quad \text{une réalisation de } \mathbf{Y}. \quad (2.28)$$

Rappelons que, dans l'équation (2.1), $y[n]$ est modélisé par la somme d'une partie déterministe, qui correspond au signal $s[n]$, et d'une réalisation du bruit aléatoire $b[n]$.

Nous cherchons à déterminer une estimation $\hat{\theta} = \hat{\theta}(\mathbf{Y})$ du vecteur des paramètres θ (2.11) afin de caractériser le modèle du signal $s[n]$. Pour décrire l'espace de ces paramètres, deux différentes stratégies (cf. figure 2.4) sont possibles [Tre68].

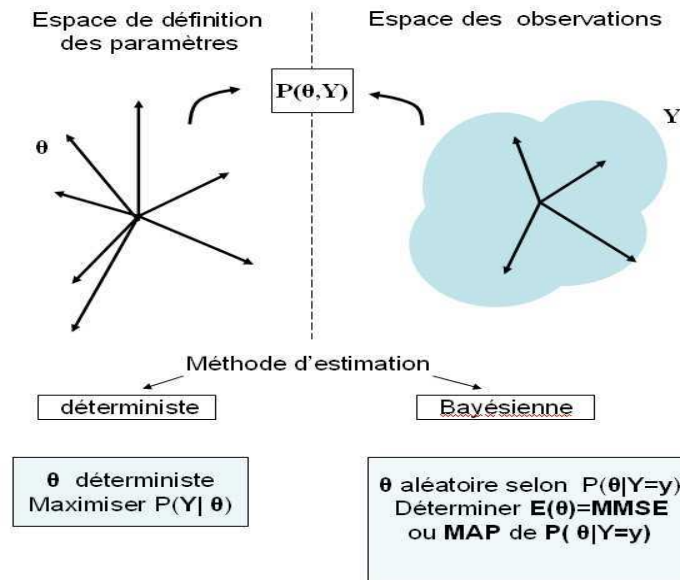


FIG. 2.4 – Deux approches possibles pour l'estimation des paramètres du modèle. Dans la première, θ (2.11) est déterministe et estimé en maximisant la fonction de vraisemblance. Dans la deuxième, θ est considéré aléatoire et un estimateur est déterminé en utilisant la densité de probabilité *a posteriori*.

Dans la première stratégie, les paramètres du modèle sont considérés comme étant des quantités inconnues déterministes. La qualité de l'estimation est alors évaluée d'après le biais de l'estimateur et la variance de l'erreur d'estimation ; le mieux serait d'avoir un biais nul et une variance très faible. Nous avons choisi d'utiliser la procédure d'estimation du maximum de vraisemblance qui fera l'objet du chapitre 3.

La seconde stratégie est une approche Bayésienne, où les paramètres sont assimilés à des variables aléatoires régies par une densité de probabilité. Pour établir la procédure d'estimation, nous avons besoin de connaître la densité *a priori* des paramètres, de déterminer la vraisemblance et de définir une fonction de coût. Un risque Bayésien est alors déterminé et les estimateurs sont ceux qui le minimisent. Il existe deux types d'estimateurs Bayésiens : l'erreur quadratique moyenne minimale MMSE "*Minimum Mean Squares Error*" et le maximum *a posteriori* MAP.

Le MMSE, appelé aussi la moyenne conditionnelle, correspond à la moyenne de la densité *a posteriori* tandis que le MAP représente son maximum et correspond au mode de celle-ci. Dans [Tre68], pour une grande classe de fonctions de coût, l'estimateur optimal coïncide avec le MMSE si la densité *a posteriori* est unimodale et symétrique par rapport à sa valeur moyenne. On verra dans le chapitre 4 que la difficulté essentielle, dans cette

stratégie, est l'évaluation de la densité *a posteriori*.

2.5 Conclusion

Dans ce chapitre, nous avons présenté le modèle adopté dans le contexte d'un signal discret, court et non-stationnaire. Quoique, dans la littérature, la modélisation polynomiale a été déjà appliquée à l'amplitude et à la phase instantanées, le modèle que nous avons proposé reprend cette idée mais l'exploite différemment. En effet, pour bien caractériser des non-stationnarités présentes sur un nombre d'échantillons faible, il est important de soigner le modèle.

Dans ce but, la modélisation polynomiale est appliquée directement à la fréquence instantanée du signal au lieu de la phase, cela permet de réduire l'erreur d'estimation de la fréquence instantanée. En ce qui concerne l'amplitude instantanée, même si dans la littérature elle a été modélisée théoriquement par des fonctions polynomiales, elle était supposée en pratique soit constante soit lentement variable. Nous nous affranchissons de cette supposition et les fortes variations de l'amplitude, nécessitant un degré d'approximation élevé, seront également analysées.

La phase initiale n'est pas considérée égale à zéro et figure parmi l'ensemble des paramètres à estimer. Elle est, également, référencée par rapport au centre de la fenêtre d'observation afin de réduire l'erreur de l'estimation. Nous avons également calculée une nouvelle base polynomiale qui a l'avantage d'être orthonormale. Cette propriété permet de réduire le découplage entre les paramètres du modèle et d'améliorer le résultat de l'estimation [FF95].

La méthode d'estimation étant aussi un élément déterminant dans le processus d'estimation, nous avons proposé deux approches différentes : la première est déterministe basée sur la maximisation de la vraisemblance et la deuxième est une approche Bayésienne. Ces deux approches seront détaillées et comparées dans les chapitres 3 et 4.

Le modèle présenté dans ce chapitre s'inscrit dans deux démarches plus générales. La première consiste à appliquer localement ce modèle pour estimer des signaux fortement non-stationnaires longs (par opposition à court, *i.e.* un grand nombre d'échantillons temporels). Nous procédons, tout d'abord, par l'extraction et l'estimation de segments courts. Le choix de ces derniers doit permettre d'une part l'adaptation du modèle à la non-stationnarité locale et l'amélioration de sa finesse et d'autre part la réduction du degré de l'approximation de la modulation de l'amplitude et de la fréquence. Ces segments seront ensuite fusionnés pour reconstruire le signal long dans sa totalité. Cette démarche est présentée et analysée dans le chapitre 6.

La deuxième démarche est l'estimation des composantes les plus énergétiques des signaux courts multi-composantes. De tels signaux sont le plus souvent rencontrés dans des milieux naturels, comme par exemple, le signal représentant le bilan de masses de glaciers de Sarennes (cf. figure 1(a)) et dont on ne dispose que de 55 échantillons correspondant aux mesures annuelles de 1949 à 2003. Cette démarche fera l'objet du chapitre 7.

Chapitre 3

Signal mono-composante non-stationnaire court - Estimation par maximum de vraisemblance

Résumé

Dans ce chapitre, nous développons l'approche déterministe dans l'estimation des paramètres du modèle, attribué à un signal court, non-stationnaire et mono-composantes. Les estimateurs des paramètres sont déterminés au sens du maximum de vraisemblance.

La maximisation directe de la vraisemblance étant très difficile, pour ce faire nous proposons un algorithme basé sur une technique d'optimisation stochastique, appelée le recuit simulé. Les différentes étapes requises dans la conception de cet algorithme seront détaillées.

Mots clefs

Maximum de vraisemblance, optimisation stochastique, recuit simulé.

Sommaire

3.1	Introduction	57
3.2	Estimation par maximum de vraisemblance	57
3.2.1	Définitions et propriétés	57
3.2.2	Lien avec l'estimation par moindres carrés	58
3.2.3	Maximum de vraisemblance compressée	59
3.2.4	Coefficient de corrélation	60
3.3	Optimisation stochastique	61
3.3.1	Méta-heuristiques	63
3.3.2	Recuit simulé - Principe	64
3.4	Recuit simulé - Configuration	65
3.4.1	Initialisation	65
3.4.2	Espace de recherche	67
3.4.3	Choix de la fonction objectif	67
3.4.4	Échantillonnage de la fonction objectif	68
3.4.4.1	Paramètres concernés	68
3.4.4.2	Mécanisme de perturbation	68
3.4.5	Règle d'acceptation	69
3.4.6	Réglage de la température	71
3.4.7	Choix du rayon du voisinage	72
3.4.8	Conditions d'arrêt de l'algorithme	74
3.4.8.1	Variance du bruit connue	74
3.4.8.2	Variance du bruit inconnue	75
3.4.9	Convergence du recuit simulé	75
3.5	Algorithme du recuit simulé	76
3.6	Conclusions	77

3.1 Introduction

Dans le chapitre précédent, nous avons proposé de modéliser par des polynômes (2.8) l'amplitude et la fréquence instantanées d'un signal court, mono-composante non-stationnaire. Dans ce chapitre, nous procédons à l'estimation du vecteur des paramètres $\boldsymbol{\theta}$ (2.11) de ce modèle par maximum de vraisemblance. En se basant sur une technique d'optimisation stochastique, appelée le recuit simulé, nous proposons un algorithme permettant de maximiser la fonction de vraisemblance sous les contraintes (2.12). Cette fonction étant multi-dimensionnelle et non-linéaire, sa maximisation directe est très difficile.

Ce chapitre se décompose en deux parties principales. La première est consacrée à la procédure du maximum de vraisemblance tandis que la deuxième se focalise sur la description et l'analyse des différents éléments intervenant dans la technique du recuit simulé. Nous présenterons, à l'issue de ces deux parties, l'algorithme conçu pour être utilisé dans l'estimation de $\boldsymbol{\theta}$.

3.2 Estimation par maximum de vraisemblance

L'idée fondamentale de l'estimation par maximum de vraisemblance, comme son nom l'indique, est de trouver un estimateur $\hat{\boldsymbol{\theta}}$ (2.11), qui maximise la densité de probabilité d'obtenir la réalisation $\mathbf{Y} = \mathbf{y}$ (2.28) conditionnellement à $\boldsymbol{\theta}$. Cette densité est appelée la vraisemblance et est notée $Prob(\mathbf{Y} = \mathbf{y} | \boldsymbol{\theta})$.

Dans la suite, nous rappelons l'expression et les propriétés de l'estimateur du maximum de vraisemblance (MV) ainsi que son lien avec l'estimateur par la méthode des moindres carrés.

3.2.1 Définitions et propriétés

Sous l'hypothèse de bruit blanc Gaussien centré et sachant que le modèle du signal est défini par (2.1), (2.2), (2.8) et (2.10), la fonction de vraisemblance, notée $\ell_{MV}(\boldsymbol{\theta})$, s'écrit [Pic96] :

$$\ell_{MV}(\boldsymbol{\theta}) = Prob(\mathbf{Y} = \mathbf{y} | \boldsymbol{\theta}) = \pi^{-(N+1)} |\det(\Gamma_{\mathbf{Y}})| \exp\left(-(\mathbf{y} - \mathbf{s})^\dagger \Gamma_{\mathbf{Y}}^{-1} (\mathbf{y} - \mathbf{s})\right). \quad (3.1)$$

$\Gamma_{\mathbf{Y}} = \mathbb{E}[\mathbf{Y} \mathbf{Y}^H]$ désigne la matrice de covariance des observations (2.27) et \mathbf{y} est donné par (2.28). Le vecteur \mathbf{s} est défini par :

$$\mathbf{s} = \left[s \begin{bmatrix} -N \\ 2 \end{bmatrix}, \dots, s \begin{bmatrix} N \\ 2 \end{bmatrix} \right]^\top \quad \text{la partie déterministe de } \mathbf{y}. \quad (3.2)$$

Le bruit étant supposé blanc de variance σ_b^2 , $\Gamma_{\mathbf{Y}}$ se réduit à $\frac{1}{\sigma_b^2} \mathbb{I}_{N+1}$. Nous obtenons alors l'expression suivante :

$$\ell_{MV}(\boldsymbol{\theta}) = (\pi \sigma_b^2)^{-(N+1)} \exp\left(-\frac{1}{\sigma_b^2} \sum_{n=-\frac{N}{2}}^{\frac{N}{2}} |y[n] - s[n]|^2\right). \quad (3.3)$$

Estimateur du maximum de vraisemblance Cet estimateur est noté $\hat{\boldsymbol{\theta}}_{MV}$ et est défini par :

$$\hat{\boldsymbol{\theta}}_{MV} = \arg \max_{\boldsymbol{\theta} \in \mathbb{R}^{M_A+M_F+3}} \ell_{MV}(\boldsymbol{\theta}) \quad (3.4)$$

Propriétés de l'estimateur du maximum de vraisemblance L'estimateur MV possède un certain nombre de propriétés asymptotiques intéressantes [Tre68], qui sont :

- l'efficacité : s'il existe un estimateur efficace et sans biais, il sera donné par l'estimateur obtenu par la méthode du maximum de vraisemblance,
- la consistance : un estimateur est consistant s'il converge en loi de probabilité vers la vraie valeur du paramètre quand le nombre d'observations tend vers l'infini,
- la normalité asymptotique : l'estimateur suit asymptotiquement une loi de probabilité Gaussienne,
- l'invariance : terme par lequel nous signifions l'invariance à la re-paramétrisation du modèle.

Toutefois, dans le cadre des signaux courts *i.e.* le nombre d'échantillons est fini, les propriétés de cet estimateur peuvent être très différentes de celles asymptotiques. En effet, bien qu'il soit convergent, l'estimateur MV devient biaisé pour N faible. De plus, en pratique les propriétés des échantillons de taille fini sont souvent inconnus, il est difficile de décider comment se fier aux propriétés asymptotiques de cet estimateur.

Pour notre part, nous avons décidé que la qualité de l'estimateur MV, $\hat{\boldsymbol{\theta}}_{MV}$ (3.4), sera évaluée par comparaison aux bornes minimales d'estimation et par la détermination du biais.

3.2.2 Lien avec l'estimation par moindres carrés

Généralement, pour éviter des problèmes de calculs dus à la manipulation des nombres à virgule flottante sur les ordinateurs, il est d'usage de maximiser le logarithme de la fonction de vraisemblance au lieu de la fonction elle-même. Bien évidemment, cela ne modifie pas l'estimateur puisque la fonction logarithmique est strictement croissante. Ce qui nous ramène donc à l'équation suivante :

$$\hat{\boldsymbol{\theta}}_{MV} = \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^{M_A+M_F+3}} \sum_{n=-\frac{N}{2}}^{\frac{N}{2}} |y[n] - s[n]|^2. \quad (3.5)$$

Vu que le bruit est blanc et Gaussien, on retrouve donc l'équivalence de la procédure de maximum de vraisemblance avec l'estimation par la méthode des moindres carrés (MC) [Tre68] :

$$\hat{\boldsymbol{\theta}}_{MV} = \hat{\boldsymbol{\theta}}_{MC} = \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^{M_A+M_F+3}} \ell_{MC}(\boldsymbol{\theta}) \quad \text{avec} \quad \ell_{MC}(\boldsymbol{\theta}) = \sum_{n=-\frac{N}{2}}^{\frac{N}{2}} |y[n] - s[n]|^2. \quad (3.6)$$

Allure de la fonction à minimiser $\ell_{MC}(\boldsymbol{\theta})$ Afin d’avoir une idée de l’allure de la fonction $\ell_{MC}(\boldsymbol{\theta})$, fonction à minimiser dans $\mathbb{R}^{M_A+M_F+3}$, nous l’avons tracée sur la figure 3.1. Le signal $s[n]$ utilisé est celui présenté sur la figure 2.1. Ce signal ayant une modulation polynomiale quadratique en amplitude et en fréquence ($M_A = M_F = 2$), nous avons fixé tous ses paramètres à leurs vraies valeurs sauf f_0 et f_2 , que nous faisons varier.

On peut distinguer sur la figure 3.1, les “paysages d’énergie” que comportent la fonction $\ell_{MC}(\boldsymbol{\theta})$: il y a un grand nombre de “vallées” plus ou moins profondes et plus ou moins proches correspondant aux minima locaux. Le nombre de ces vallées croît et leur rapprochement s’accroît avec l’augmentation du nombre de paramètres du modèle. Autrement dit, la non-linéarité de $\ell_{MC}(\boldsymbol{\theta})$ croît avec les degrés M_A et M_F . Le choix de la méthode d’optimisation doit donc être adapté à notre problème et il sera discuté dans le deuxième partie de ce chapitre (cf. paragraphe 3.3).

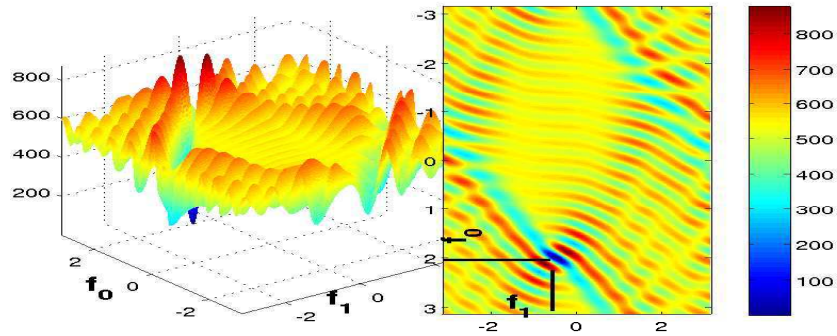


FIG. 3.1 – Illustration de la non-linéarité de $\ell_{MC}(\boldsymbol{\theta})$ et de la présence de plusieurs minima locaux. Tous les paramètres sont fixés sauf f_0 et f_2 : $\{\varphi_0 = 0.4; a_0 = 16; a_1 = 2; a_2 = 3; f_1 = 0.5\}$.

L’estimation conjointe des paramètres selon (3.6) ne suppose aucune information *a priori* sur le découplage des paramètres du modèle. Dans certains cas, il est possible de disposer d’une telle information *a priori*, qui pourra réduire la complexité de la fonction à minimiser (3.6). Deux exemples sont présentés dans la suite.

3.2.3 Maximum de vraisemblance compressée

Dans [FF95], il est démontré que les paramètres d’amplitude sont découplés de ceux de la fréquence et de la phase, le problème d’estimation peut se scinder donc en deux sous-problèmes (cf. chapitre 1.3.3 et chapitre 5). Ceci permet de réduire le nombre de paramètres à estimer en même temps.

Premier sous-problème Les estimations des paramètres de la phase et de la fréquence $\{\hat{\varphi}_0, \hat{f}_0, \dots, \hat{f}_{M_F}\}$ sont déterminées indépendamment de celles des paramètres d’amplitude. On procède de la manière suivante :

- D’abord, une approximation de l’amplitude instantanée, notée $\hat{A}[n]$, est évaluée pour $n = -\frac{N}{2}, \dots, \frac{N}{2}$ en fixant arbitrairement les valeurs des paramètres d’amplitude.
- Puis, on détermine une estimation de $\{\hat{\varphi}_0, \hat{f}_0, \dots, \hat{f}_{M_F}\}$ en maximisant la fonction de vraisemblance compressée “*Compressed Likelihood*” :

$$\{\hat{\varphi}_0, \hat{f}_0, \dots, \hat{f}_{M_F}\} = \arg \min_{\{\varphi_0, f_0, \dots, f_{M_F}\} \in \mathbb{R}^{M_F+2}} \ell_{MC}(\boldsymbol{\theta} \mid \hat{A}[n]). \quad (3.7)$$

Deuxième sous-problème Les estimations de $\{\hat{\varphi}_0, \hat{f}_0, \dots, \hat{f}_{M_F}\}$, obtenues dans le premier sous-problème, sont réutilisées pour évaluer analytiquement les nouvelles valeurs des paramètres d’amplitude. Ces dernières, notées $\{\hat{a}_0, \dots, \hat{a}_{M_A}\}$ sont fournies par :

$$\hat{a}_m = \Re \{ \mathbf{y}^\dagger \mathbf{D} \mathbf{G}_m \} \quad \text{pour } 0 \leq m \leq M_A, \quad (3.8)$$

où \mathbf{y} est défini par (2.28) et $\mathbf{D} = \begin{pmatrix} e^{j\Phi[-\frac{N}{2}]} & & & \\ & e^{j\Phi[-\frac{N}{2}+1]} & & \\ & & \ddots & \\ & & & e^{j\Phi[\frac{N}{2}]} \end{pmatrix}$ est une matrice dia-

gonale. $\mathbf{G}_m = [g_m[-\frac{N}{2}], \dots, g_m[\frac{N}{2}]]^\top$, où $\{g_m[-\frac{N}{2}]\}$ est la base utilisée dans le modèle du signal (2.8) et elle est orthonormale polynomiale (2.21). La démonstration est fournie en annexes A.6.

Dans le cas, où la base est quelconque (non-orthonormale), les expressions des paramètres d’amplitude sont données dans [FF95].

3.2.4 Coefficient de corrélation

Dans le cas où les paramètres d’amplitude sont découplés de ceux de la fréquence et de la phase, on peut remplacer la minimisation de la fonction de vraisemblance compressée (3.7) par la maximisation du coefficient de corrélation entre le signal observé $y[n]$ et le modèle $s[n]$. Ce coefficient de corrélation $\gamma_{y,s}(\boldsymbol{\theta})$ est défini par :

$$\gamma_{y,s}(\boldsymbol{\theta}) = \Re \left\{ \frac{\mathbf{y}^\dagger \mathbf{s}}{|\mathbf{y}| |\mathbf{s}|} \right\}. \quad (3.9)$$

$$(3.10)$$

Cela revient donc à (cf. annexes A.6) :

$$\{\hat{\varphi}_0, \hat{f}_0, \dots, \hat{f}_{M_F}\} = \arg \max_{\{\varphi_0, f_0, \dots, f_{M_F}\} \in \mathbb{R}^{M_F+2}} \gamma_{y,s}(\boldsymbol{\theta}). \quad (3.11)$$

Les paramètres de l’amplitude sont déterminés par (3.8) dans le cas où la base polynomiale est orthonormale et discrète. La maximisation du coefficient de corrélation permet

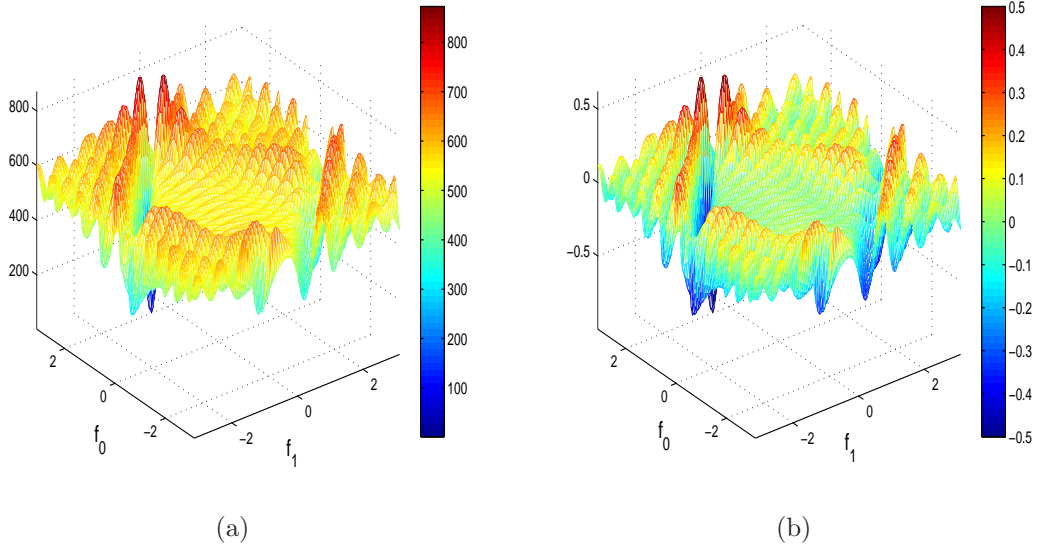


FIG. 3.2 – (a) $\ell_{MC}(\boldsymbol{\theta} | \hat{A}[n])$ fonction des moindres carrés (3.6) déjà tracée sur la figure 3.1 comparée à (b) au coefficient de corrélation (3.11), c’est l’opposé de $\gamma_{y,s}(\boldsymbol{\theta})$ qui est tracé.

de réduire la profondeur des vallées comme le montre la figure 3.2, ce qui facilite leur exploration et par conséquent le processus d’optimisation. En effet, le coefficient de corrélation prend des valeurs dans $[-1, 1]$ quel que soit le signal étudié.

Un cas particulier, où l’utilisation du coefficient de corrélation (3.9) est plus intéressante que la fonction des Moindres Carrés est le cas où l’amplitude du signal est constante. cette amplitude sera estimée directement par ($|\mathbf{s}| = N + 1$) :

$$\hat{a}_0 = \Re \left\{ \frac{\mathbf{y}^\dagger \mathbf{s}}{|\mathbf{y}| \sqrt{N+1}} \right\}. \quad (3.12)$$

3.3 Optimisation stochastique

La fonction à minimiser sera appelée dans cette partie fonction “objectif”. Selon cette dernière et les contraintes du problème, le choix de la méthode d’optimisation est déterminant pour la qualité et la précision de l’estimation. En effet, l’obtention de la solution optimale ou uniquement d’une approximation de celle-ci dépend de l’efficacité de cette méthode et doit correspondre à un temps de calcul raisonnable.

Sur la figure 3.3, est illustrée une classification générale des méthodes d’optimisation en variables continues pour une fonction mono-objectif (*i.e.* une seule fonction est à minimiser). Dans notre problème, l’espace de définition des solutions *i.e.* les paramètres du modèle est continu (non-discret) et inclus dans $\mathbb{R}^{M_A+M_F+3}$.

Dans notre cas, outre l’absence de convexité stricte et la présence de plusieurs minima locaux de la fonction $\ell_{MC}(\boldsymbol{\theta})$ ou $\gamma_{y,s}(\boldsymbol{\theta})$ (cf. figures 3.1 et 3.2), il existe des corrélations entre les paramètres à estimer φ_0, f_0, f_1 et f_2 d’une part et a_0, a_1 et a_2 d’autre part. Ces

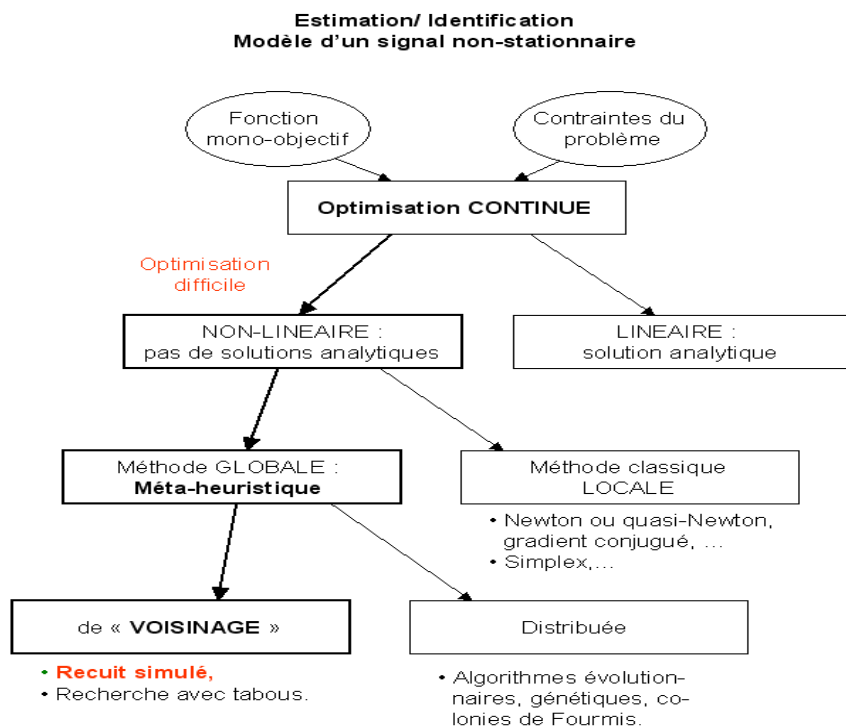


FIG. 3.3 – Classification générale des méthodes d’optimisation en variables continues pour des fonctions mono-objectifs. Le terme “globale” désigne les méthodes convergeant vers un minimum global et “locale” celles stagnant sur un minimum local. Les méthodes “de Voisinage” font référence à celles qui font progresser une solution à la fois et les méthodes “distribuées” désignent celles qui manipulent en parallèle toute une population de solutions.

corrélations sont dues aux contraintes imposées sur le modèle (2.12), c’est-à-dire l’amplitude est strictement positive, la fréquence est positive et respectant le théorème de Shannon.

Dès lors, l’obtention d’une expression analytique des estimateurs n’est pas possible. Nous sommes donc confrontés à un problème difficile, où les techniques d’optimisation classiques [Kay88, Bro70], telles que la descente du gradient, l’algorithme de Gauss-Newton, la méthode de BFGS (Broyden-Fletcher-Goldfarb-Shanno), *etc* [Bro70], ne garantissent pas la convergence vers un minimum global. Nous avons envisagé alors d’utiliser des méthodes d’optimisation stochastique et numérique, appelées les méta-heuristiques. Nous nous intéresserons, en particulier, à la méthode du recuit simulé “*Simulated Annealing*”.

3.3.1 Méta-heuristiques

Les méthodes méta-heuristiques, apparues dans les années 1980, sont des heuristiques¹ génériques, c'est-à-dire extensibles et/ou adaptables à un grand nombre de problèmes, sans modifications ou changements profonds dans la structure principale de l'algorithme.

Ce sont des méthodes itératives, faciles à implémenter et reposent sur un échantillonnage aléatoire de la fonction objectif sans avoir recours à l'information du gradient. Leur intérêt principal, par rapport aux méthodes d'optimisation classiques, réside essentiellement dans leur capacité de s'extraire d'un minimum local. En effet, au cours des itérations, une dégradation de la fonction objectif est possible de temps en temps, *i.e.* on accepte une solution qui augmente cette fonction. Ainsi elles peuvent échapper aux "puits" d'énergie (cf. figure 3.1) de la fonction objectif. Bien sûr, un processus de contrôle de dégradations est prévu pour éviter la divergence.

Parmi les méthodes méta-heuristiques [DPST03], figurent les méthodes "de voisinage", telles que le recuit simulé et la recherche avec tabous, et les méthodes "distribuées" telles que les algorithmes génétiques, évolutifs, colonies de fourmis, *etc.* Les premières font progresser une solution à la fois tandis que les secondes manipulent en parallèle toute une population de solutions.

Il faut souligner que ces méta-heuristiques ont été développées, au départ, pour des problèmes d'optimisation en variables discrètes. Dans la littérature, la plupart des méthodes proposées, qui réalisent l'adaptation (discret \rightarrow continu) pour une méta-heuristique, ne s'appliquent pas aux autres méta-heuristiques. De plus, la comparaison entre ces méthodes n'est pas évidente ; aucune méthode ne surpasse ses concurrentes [DPST03]. Cela est dû à certaines difficultés spécifiques au cas continu, nous citons par exemple :

- la sélection de point de départ de l'algorithme (le point d'initialisation),
- le nombre d'itérations et par conséquent le temps d'exécution,
- la définition de l'approche du point optimal (l'erreur finale).

Ces difficultés s'accroissent avec la non-linéarité de la fonction objectif et la présence des corrélations entre les variables. L'efficacité de la méta-heuristique dépend aussi de l'espace de recherche qui peut varier d'une variable à l'autre. De ce fait, peu d'études sur l'efficacité et la convergence ont été établies dans le cadre de l'optimisation en variables continues *i.e.* le cas qui nous intéresse.

Le libre choix de la méta-heuristique, sa conception et son réglage sont donc laissés à l'utilisateur. Ce dernier doit alors l'adapter aux enjeux de son problème, pour atteindre l'optimum avec une précision acceptable en un temps de calcul raisonnable. Nous avons opté pour la méthode du recuit simulé. La description de ses étapes fait l'objet du paragraphe suivant.

¹Ce sont des méthodes approchées simples, rapides et employant des processus aléatoires dans le but d'optimiser un problème sans calcul du gradient.

3.3.2 Recuit simulé - Principe

En 1983, une nouvelle méthode d'optimisation itérative “le recuit simulé”, qui évite les minima locaux, a été publiée par Kirkpatrick et al dans [KGV83]. Cette méthode s'appuie d'une part sur l'algorithme de Metropolis [MRR⁺53], qui permet de décrire l'évolution d'un système thermodynamique et s'inspire d'autre part d'un processus utilisé en métallurgie. Ce dernier alterne des cycles de refroidissement lent et de réchauffage. Techniquement un matériau, ayant initialement une énergie élevée, est refroidi en marquant des paliers de température. Ces paliers doivent durer “suffisamment” pour que l'équilibre thermodynamique soit établi. Entre-temps, si des défauts locaux apparaissent suite à un refroidissement rapide, ils seront éliminés par un réchauffement local (recuit). Par analogie avec ce processus physique (cf. Tab. 3.1), la fonction objectif devient l'énergie du système et la température un paramètre fictif de contrôle.

Problème d'optimisation	Système physique
Fonction objectif $\ell_{MC}(\boldsymbol{\theta})$	Energie libre des particules
Paramètres du modèle du signal $\boldsymbol{\theta}$	Coordonnées des particules
Trouver le minimum global	Trouver les états à bas niveau d'énergie
Commande : Température	

TAB. 3.1 – Analogie entre le problème d'optimisation et le système physique dans le fonctionnement du recuit simulé.

Depuis son apparition, le recuit simulé a suscité de nombreux et différents aspects d'études. Par exemple, dans [LA87, Ing93, HG89], la convergence théorique de cette méthode vers un optimum global a été établie sous certaines conditions. Dans [SD89, Sia94, Her00], les réglages des éléments intervenant dans le fonctionnement du recuit simulé sont analysés. Plusieurs variantes de la méthode ainsi que des extensions à différents problèmes y sont également proposées. Toutefois, la majorité de ces études ont été consacrées à des problèmes combinatoires discrets et peu d'études ont été faites dans le cas d'optimisation à variables continues [VL84, Mic91, Loc00, MMM06]. La gestion et le contrôle du processus d'optimisation sont complètement différents dans les deux cas.

Les principales étapes dans le recuit simulé sont représentées d'une façon générale sur la figure 3.4. Elles incluent :

- une étape d'initialisation au cours de laquelle une solution initiale $\boldsymbol{\theta}^{(0)}$ est déterminée. La notation $^{(0)}$ désigne la valeur de l'initialisation et $^{(t)}$ la valeur à l'itération t .
- une étape itérative pour générer des candidats $\boldsymbol{\theta}^{(t)}$ (solution potentielle). Cette étape est contrôlée par la température et la règle d'acceptation de Metropolis,
- une étape de contrôle basée sur la vérification d'un critère d'arrêt.

Dans la suite, nous détaillerons ces étapes et les éléments nécessaires à leur fonctionnement.

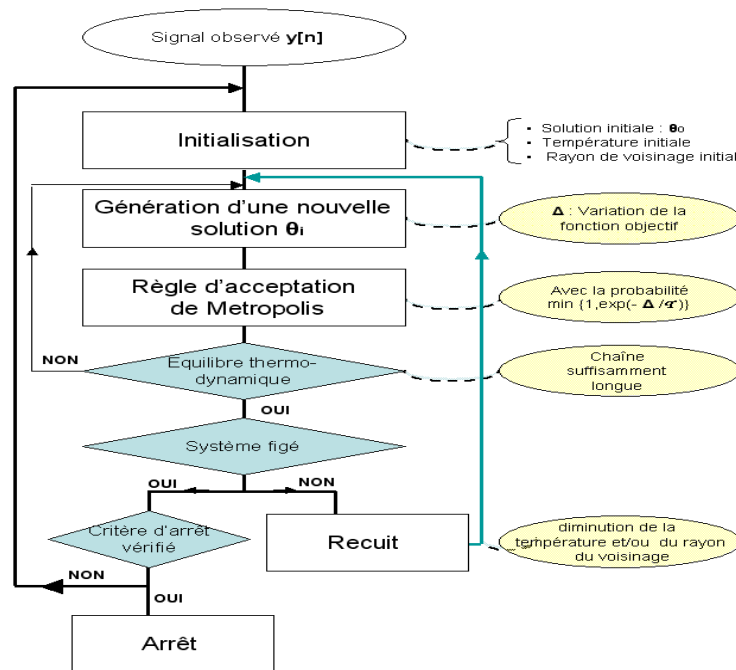


FIG. 3.4 – Organigramme du recuit simulé.

3.4 Recuit simulé - Configuration

Dans cette partie, nous explicitons d'une part les éléments nécessaires au fonctionnement du recuit simulé et nous précisons le choix que nous avons adopté pour résoudre notre problème (3.4).

3.4.1 Initialisation

Une bonne initialisation $\theta^{(0)}$ du vecteur des paramètres du modèle θ permet de réduire le temps de calcul et de simplifier le réglage des éléments intervenant dans le fonctionnement de l'algorithme du recuit simulé. Nous proposons deux méthodes d'initialisation différentes.

Initialisation par transformée de Fourier Pour initialiser les paramètres a_0 , f_0 et φ_0 (2.11), nous pouvons utiliser la transformée de Fourier discrète du signal observé $y[n]$, soit donc :

$$f_0^{(0)} = \arg \max_{\nu} |\mathcal{F}_y(\nu)|, \quad a_0^{(0)} = |\mathcal{F}_y(f_0^{(0)})|, \quad \varphi_0^{(0)} = \arg \mathcal{F}_y(f_0^{(0)}), \quad (3.13)$$

$$\text{avec } |\mathcal{F}_y(\nu)| = \sum_{n=-\frac{N}{2}}^{\frac{N}{2}} y[n] e^{-2\pi j\nu n}. \quad (3.14)$$

Quant aux paramètres restants, ils sont initialisés avec la valeur zéro :

$$\{a_1^{(0)}, a_2^{(0)}, \dots, a_{M_A}^{(0)}, f_1^{(0)}, f_2^{(0)}, \dots, f_{M_F}^{(0)}\} = \{0, \dots, 0\}. \quad (3.15)$$

Dans le but d'illustrer cette façon d'initialiser l'algorithme du recuit simulé, nous avons considéré un signal $s[n]$ mono-composante ayant une modulation quadratique ($M_A = M_F = 2$) et formé de 33 échantillons. Les vraies valeurs de ces paramètres s'écrivent dans la base polynomiale orthonormale et discrète (2.21) :

$$\boldsymbol{\theta} = [\varphi_0 = 0.4, a_0 = 16, a_1 = -3, a_2 = 2, f_0 = 2, f_1 = -0.2, f_2 = -0.5]^\top$$

Ce signal est noyé dans du bruit blanc Gaussien avec un RSB égal à -5 dB.

Sur la figure 3.5(a), le module de $\mathcal{F}_y(\nu)$ est tracé. $a_0^{(0)}$, $f_0^{(0)}$ et $\varphi_0^{(0)}$ correspondent respectivement à l'amplitude, la fréquence et la phase du pic maximal. Pour cet exemple, l'initialisation par la transformée de Fourier fournit l'initialisation suivante :

$$\boldsymbol{\theta}^{(0)} = [2.2022, 15.8312, 0, 0, 0.4181, 0, 0]^\top,$$

ce qui correspond à un écart relatif de 1% sur la vraie valeur de a_0 et de 80% sur f_0 .

Initialisation par spectrogramme Une deuxième façon d'initialiser l'algorithme du recuit simulé peut se faire à l'aide du spectrogramme, $\mathcal{S}_y[n, \nu]$ donné par (1.3). A partir de la trajectoire fréquentielle, définie par les points d'énergie maximale dans le plan temps-fréquence, on obtient des approximations grossières de la fréquence et de l'amplitude du signal :

$$F^{(0)}[n] = \arg \max_{\nu} |\mathcal{S}_y[n, \nu]| \quad \text{et} \quad A^{(0)}[n] = |\mathcal{S}_y[n, F^{(0)}[n]]|^{\frac{1}{2}}, \quad (3.16)$$

La décomposition de $F^{(0)}[n]$ et de $A^{(0)}[n]$ sur la base polynomiale orthonormale discrète $\{g_m[n]\}$ (2.21), fournit l'initialisation des paramètres de l'amplitude et de la fréquence :

$$a_m^{(0)} = \langle A^{(0)}[n], g_m[n] \rangle_{\mathcal{D}}, \quad f_m^{(0)} = \langle F^{(0)}[n], g_m[n] \rangle_{\mathcal{D}}, \quad (3.17)$$

Quant à la phase initiale, elle sera prise égale à l'argument du point ayant l'énergie maximale à l'instant zéro dans le plan temps-fréquence :

$$\varphi_0^{(0)} = \arg \mathcal{F}_y[0, \nu_{max}] \quad \text{avec} \quad \nu_{max} = \arg \max_{\nu} |\mathcal{S}_y[0, \nu]| \quad (3.18)$$

Cette façon d'initialiser est illustrée par la figure 3.5(b) pour le même signal que la figure 3.5(a), ce qui conduit au résultat suivant :

$$\boldsymbol{\theta}^{(0)} = [-1.8864, 13.5072, -2.5042, 0.6101, 1.4032, -0.0167, -0.2773]^\top.$$

Nous avons relevé un écart relatif moyen de 23% sur les paramètres de l'amplitude et de 35% sur ceux de la fréquence, soit une meilleure initialisation des paramètres de la fréquence que la méthode d'initialisation par transformée de Fourier.

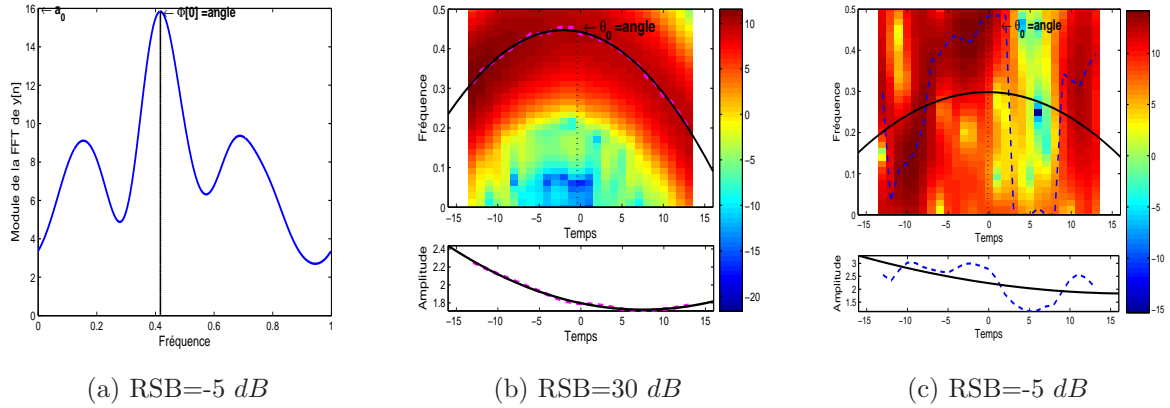


FIG. 3.5 – Initialisation des paramètres d’un signal ayant une modulation quadratique en amplitude et en fréquence pour deux valeurs de RSB : $-5 dB$ et $30 dB$. (a) Module de $|\mathcal{F}_y(\nu)|$ du signal observé, calculée à l’aide d’une fenêtre de Hanning. (b) Fréquence (respectivement amplitude) instantanée (—) et son initialisation (--) $F^{(0)}[n]$ (respectivement $A^{(0)}[n]$) en utilisant le spectrogramme. Ce dernier est calculé avec $N_{FT} = 256$, une fenêtre de Hanning de 8 points et un recouvrement de 99%.

Dans la littérature [DPST03], on propose également d’initialiser les paramètres d’une façon aléatoire. Sauf que dans notre problème, il faut s’assurer que les contraintes ($A[n] > 0$ et $0 < F[n] < \frac{F_c}{2}$) soient respectées. L’influence de l’initialisation sera étudiée dans le chapitre 5, à l’issue duquel nous retenons la méthode qui contribue le mieux dans l’amélioration de l’estimation de θ .

3.4.2 Espace de recherche

D’une façon générale, l’espace de recherche des paramètres dépend non seulement des domaines de définition de ces derniers mais aussi des contraintes du problème étudié. L’exploration de sa structure, nécessaire à la recherche de l’optimum global, est souvent très complexe. Afin de la simplifier, on peut élargir l’espace de définition des paramètres, quitte à tenir compte des contraintes uniquement lors du calcul de la fonction objectif, ou/et introduire une distorsion de cette dernière.

En particulier et dans notre problème, où il s’agit de minimiser la fonction $\ell_{MC}(\theta)$ sous les contraintes $A[n] > 0$ et $0 < F[n] < \frac{F_c}{2}$, nous avons choisi de considérer l’espace de recherche égal à $\mathbb{R}^{M_A+M_F+3}$.

3.4.3 Choix de la fonction objectif

Le choix de la fonction objectif dépend de l’application envisagée. En effet, si nous ne disposons pas d’information *a priori* sur le découplage des paramètres et nous souhaitons faire une estimation conjointe de ces derniers, alors le recuit simulé peut être employé pour minimiser la fonction objectif $\ell_{MC}(\theta)$ (3.6).

En revanche, si les paramètres de l’amplitude sont découplés de ceux de la fréquence et

de la phase initiale, nous avons le choix entre minimiser $\ell_{MC}(\boldsymbol{\theta})$ ou maximiser le coefficient de corrélation $\gamma_{y,s}(\boldsymbol{\theta})$ (3.9). Dans le deuxième choix, le recuit simulé sera employé uniquement dans la recherche de $\{\varphi_0, f_0, \dots, f_{M_F}\}$, tandis que $\{a_0, \dots, a_{M_A}\}$ seront estimés analytiquement (3.8).

3.4.4 Échantillonnage de la fonction objectif

La recherche de l'optimum de la fonction objectif nécessite l'exploration de l'espace de recherche. Dans les problèmes d'optimisation en variables discrètes [DPST03], cette exploration est basée sur la notion de "voisin de la solution courante". Étant donné que nous avons un problème d'optimisation en variables continues, cette notion n'est plus possible à cause de la densité de l'espace de recherche ; dans ce cas, on parle plutôt de notion "de voisinage".

Il est donc nécessaire de définir une hyper-sphère [NGH⁺01] centrée sur la solution courante. Le volume et le rayon de cette hyper-sphère constituent, respectivement, l'étendue et le rayon du voisinage de cette solution. La sélection de la nouvelle solution se fait ensuite, dans ce voisinage, au moyen d'un mécanisme de perturbations. Ce mécanisme consiste à opérer des changements élémentaires sur la solution courante, dans des directions et à des distances pouvant occasionner une dégradation de la fonction objectif. On parle alors de l'échantillonnage de la fonction objectif avec un pas égal au rayon du voisinage. Dès lors, ce rayon devient un paramètre de contrôle supplémentaire, mais souvent crucial, dans la réussite de la méthode de recuit simulé. Son choix reste étroitement lié à celui du mécanisme de perturbation.

3.4.4.1 Paramètres concernés

Avant de choisir le mécanisme de perturbation, la question est quels vont être les paramètres concernés. Plusieurs propositions sont possibles :

- tous les paramètres du modèle en même temps,
- un seul paramètre à la fois,
- un sous ensemble de paramètres à la fois.

Dans les deux dernières propositions, le choix à chaque fois du paramètre, ou le sous ensemble de paramètres, à modifier peut se faire d'une façon déterministe ou aléatoire parmi l'ensemble des paramètres du modèle. Quand ces derniers sont corrélés, il est souvent conseillé de faire le choix aléatoirement.

3.4.4.2 Mécanisme de perturbation

Soit $\theta_i^{(t)}$ une solution courante à l'itération t du paramètre θ_i (2.11), $i \in [0, M_A + M_F + 2]$. Elle est sujet à une perturbation $\delta\theta_i$ *a priori* différente pour chaque paramètre. Un nouveau candidat $\theta_{i,c}$ est généré par :

$$\theta_{i,c} = \theta_i^{(t)} + \delta\theta_i, \quad \boldsymbol{\theta}_c = [\theta_{0,c}, \dots, \theta_{M_A + M_F + 2,c}]^T \quad \text{et} \quad \boldsymbol{\Delta\theta} = [\delta\theta_0, \dots, \delta\theta_{M_A + M_F + 2}]^T \quad (3.19)$$

Plusieurs façons de générer la perturbation $\delta\theta_i$ sont possibles. Dans le chapitre 6 du [DPST03], si le domaine de définition du paramètre θ_i est “peu étendu”, $\delta\theta_i$ est choisie aléatoirement selon une loi uniforme. Dans le cas contraire, une loi logarithmique est utilisée. Dans [Loc01], la perturbation est choisie uniforme sur l’intersection entre le vrai domaine de définition de θ_i et son voisinage d’exploration (Volume de l’hyper-sphère). Dans [MC97], la perturbation est aléatoire (isotrope) générée selon une loi gaussienne ou de Cauchy. Sa variance est proportionnelle à la température (paramètre de contrôle du recuit simulé Tab. 3.1).

Dans le cadre de notre problème, nous avons choisi de générer $\delta\theta_i$ à l’itération t selon une loi normale centrée et de variance $\sigma_i^2(t)$, c’est à dire

$$\delta\theta_i \sim \mathcal{N}(0, \sigma_i^2(t)), \quad (3.20)$$

où $\sigma_i(t)$ est le rayon de voisinage de la solution courante $\theta_i^{(t)}$ à l’itération t . Le choix de ce dernier est discuté dans le paragraphe 3.4.7.

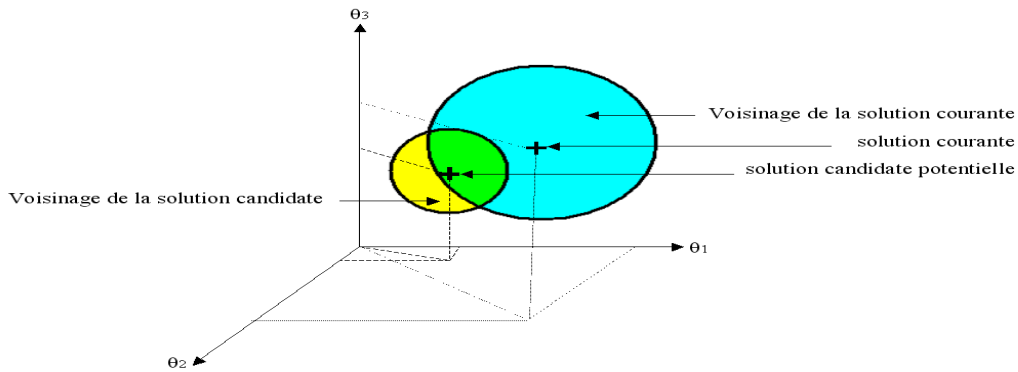


FIG. 3.6 – Illustration du mécanisme de perturbation.

3.4.5 Règle d’acceptation

Pour le recuit simulé, l’acceptation des nouvelles solutions doit se faire occasionnellement et sous le contrôle de la température. La règle la plus utilisée est celle de Metropolis [MRR⁺53]. Elle s’appuie directement sur la physique statique, qui stipule que :

Si l’équilibre thermodynamique est atteint à une température \mathcal{T} , alors la probabilité pour qu’un système physique possède une énergie E est proportionnelle au facteur de Boltzman $e^{\frac{E}{K_b \mathcal{T}}}$, K_b étant la constante de Boltzman à la température \mathcal{T} .

En pratique, cette règle est réalisée de la façon suivante. Partant d’une solution courante $\theta_i^{(t)}$ à l’itération t , à qui on a fait subir des perturbations aléatoires pour obtenir un nouveau candidat $\theta_{i,c}$:

- si ce dernier a pour effet de diminuer la fonction objectif ($\Delta < 0$, cf. figure 3.4), il sera accepté.
- si au contraire il augmente la fonction objectif ($\Delta > 0$), il sera tout de même accepté avec la probabilité d'acceptation $e^{-\frac{\Delta}{\mathcal{T}}}$, \mathcal{T} étant la température (cf. figure 3.7). Autrement dit, on tire au hasard un nombre réel u selon une loi de probabilité uniforme \mathcal{U} sur $[0, 1]$. Si u est inférieur à $e^{-\frac{\Delta}{\mathcal{T}}}$, le nouveau candidat est alors accepté. Il sera rejeté dans le cas contraire :

$$\begin{aligned} u \sim \mathcal{U}[0, 1]; \quad u \leq e^{-\frac{\Delta}{\mathcal{T}}} &\Rightarrow \theta_i^{(t+1)} = \theta_{i,c}, \\ u > e^{-\frac{\Delta}{\mathcal{T}}} &\Rightarrow \theta_i^{(t+1)} = \theta_i^{(t)} \end{aligned} \quad (3.21)$$

Pour une température \mathcal{T} élevée, on a $e^{-\frac{\Delta}{\mathcal{T}}} \approx 1$. Le recuit simulé est alors équivalent à une marche aléatoire dans l'espace de recherche ; c'est-à-dire on accepte de façon quasi-systématique tous les nouveaux candidats générés selon l'équation (3.19), même les très mauvais.

À faible température, les nouveaux candidats augmentant la fonction objectif sont sévèrement rejetés puisque $e^{-\frac{\Delta}{\mathcal{T}}} \approx 0$. Le recuit simulé se résume dans ce cas à une amélioration itérative classique.

À température moyenne, la règle d'acceptation de Metropolis (3.21) offre au recuit simulé une chance de s'extraire des minima locaux.

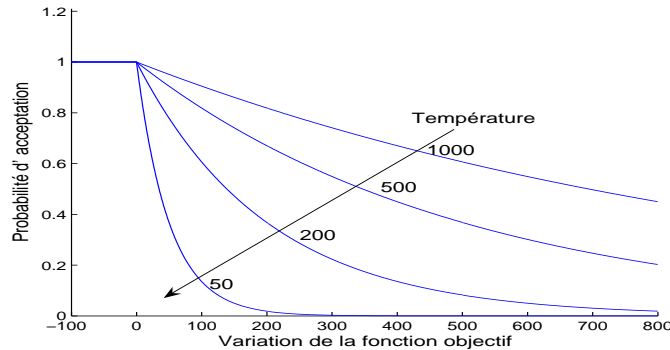


FIG. 3.7 – Probabilité d'acceptation de Metropolis en fonction de la variation Δ de la fonction objectif et du paramètre de contrôle, la température \mathcal{T} .

À une température fixée (palier de température), l'itération de cette règle permet d'engendrer une séquence de solutions (où chacune d'elles dépend uniquement de celle qui la précède immédiatement), qui constitue une chaîne de Markov. Théoriquement, si la longueur de cette chaîne est infinie, le système tend vers l'équilibre thermodynamique. Pratiquement, avec une chaîne "suffisamment longue," on suppose que cet équilibre est établi. Ensuite, on fait baisser la température et on construit une nouvelle chaîne de Markov.

Par ailleurs, d'autres variantes de cette règle d'acceptation [SD89], que nous n'explorons pas ici, ont été proposées pour réduire le temps de calcul quand la température devient très faible et *a fortiori* le taux d'acceptation, ce dernier étant défini comme suit :

Définition 1 *Etant donné une solution courante $\theta_i^{(t)}$ et un palier de température fixé \mathcal{T} , le taux d'acceptation, noté $TA_i(\mathcal{T})$, est le rapport entre le nombre de solutions acceptées, sur le nombre total de candidats générés $\theta_{i,c}$.*

3.4.6 Réglage de la température

Le contrôle de la température est un point très subtil car il doit éviter de se retrouver dans l'un des cas suivants :

- si la décroissance de la température est rapide, l'établissement de l'équilibre thermodynamique de la chaîne de Markov engendrée n'a pas le temps de s'établir.
- si l'optimum global est encore très éloigné et que la température devient très faible, cela entraîne un faible taux d'acceptation (cf. figure 3.8(a)) pouvant conduire à une solution non optimale.
- si l'optimum global est assez proche et que la température est assez élevée, le taux d'acceptation l'est aussi (cf. figure 3.8 (b)). L'algorithme continue alors d'accepter des solutions augmentant la fonction objectif, contribuant ainsi à une perte de temps avant de se figer sur la solution optimale.

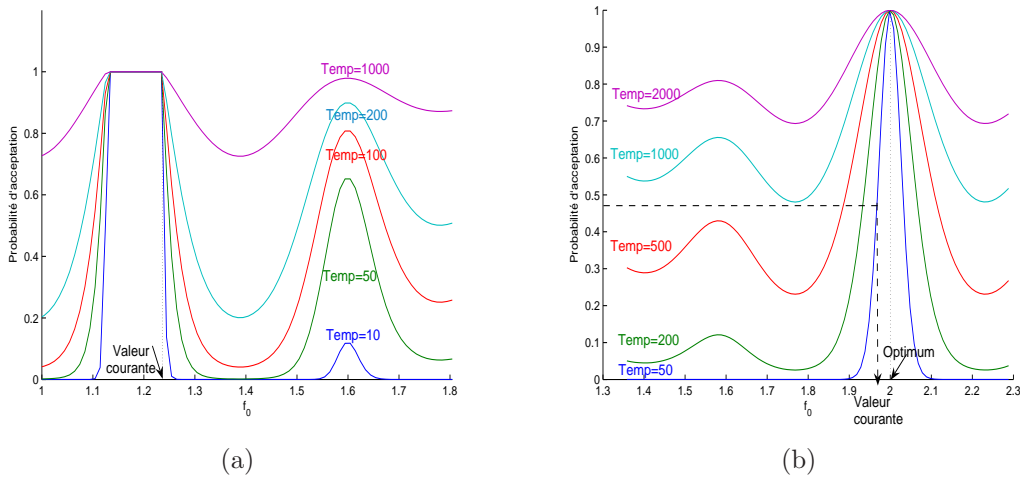


FIG. 3.8 – Importance du réglage de la température : (a) la température est très basse et l'optimum est très loin, on s'interdit de franchir des zones permettant l'accès à ce dernier car la probabilité d'acceptation est quasi-nulle. (b) le voisinage immédiat de l'optimum est atteint et la température est assez élevée, des zones sont explorées inutilement car la probabilité d'acceptation est encore assez grande. Le signal considéré est celui de la figure 3.1. Tous les paramètres du signal sont fixés à leurs vraies valeurs sauf le paramètre f_0 , qu'on fait varier. L'optimum de la fonction $\ell_{MC}(\theta)$ correspond à $f_0 = 2$.

Ces cas montrent l'existence d'une forte corrélation entre la durée minimale d'un palier de température et le taux de décroissance de celle-ci *i.e.* changements de palier. Une

bonne initialisation et un choix convenable de loi de décroissance de la température permettent de remédier à ces trois cas. Il existe plusieurs façons de réduire la température [DPST03, HRA86].

Toutefois en l'absence de résultats théoriques généraux, nous avons considéré dans la suite la loi géométrique : la température au palier numéro $t + 1$ est exprimée en fonction de la température du t -ième palier par :

$$\mathcal{T}^{(t+1)} = (1 - \epsilon_{\mathcal{T}})\mathcal{T}^{(t)}, \quad \epsilon_{\mathcal{T}} \in [0, 1]. \quad (3.22)$$

La température initiale $\mathcal{T}^{(0)}$ est choisie empiriquement. Toutefois $\epsilon_{\mathcal{T}}$ et $\mathcal{T}^{(0)}$ doivent être réglés de façon à obtenir une bonne solution, en un temps de calcul polynomial (raisonnable) et éviter les trois cas cités précédemment. Des exemples de valeurs de $\epsilon_{\mathcal{T}}$ et $\mathcal{T}^{(0)}$ seront donnés lors des applications sur des signaux.

3.4.7 Choix du rayon du voisinage

Au cours des itérations, le rayon du voisinage $\sigma_i^{(t)}$ (3.20) peut être fixe (plus ou moins grand) ou variable ayant une limite supérieure $\sigma_i^{(0)} = \sigma_i$ [DPST03]. Dans ce dernier cas, dit topologie adaptative, le rayon de voisinage $\sigma_i^{(t)}$ est réduit au cours des itérations en fonction de la décroissance de la température \mathcal{T} et du taux d'acceptation $TA_i(\mathcal{T})$ (cf. définition 1), afin de maintenir l'efficacité de l'optimisation. Toutefois, il faut éviter de se piéger dans l'une des situations suivantes :

- l'algorithme du recuit simulé risque de se figer si les trois conditions suivantes sont réunies en même temps : le voisinage de l'optimum global n'est pas encore atteint, $\sigma_i^{(t)}$ est petite ou le devient au cours des itérations (*i.e.* on bouge au ralenti) et la température devient très faible. En effet, la probabilité d'acceptation quasi-nulle combinée avec un voisinage d'exploration très réduit empêcheront l'algorithme de visiter le voisinage du minimum global et par conséquent stagne sur un minimum local (cf. Fig 3.9(a)).
- D'un autre côté, si $\sigma_i^{(t)}$ est encore grand alors qu'on est proche de l'optimum global, l'algorithme va continuer à accepter des mauvais candidats autour de minimum global (cf. figure 3.9(b)).

Le choix de l'initialisation σ_i et de la façon de décroître $\sigma_i^{(t)}$ étant très important, nous avons choisi, dans notre problème d'initialiser σ_i par la borne supérieure de domaine de définition de θ_i , qui est calculée dans le chapitre précédent (2.24) pour chaque paramètre du modèle.

En ce qui concerne la décroissance de $\sigma_i^{(t)}$, il existe plusieurs façons de le faire :

- Dans le chapitre 6 [DPST03] sur l'optimisation en variables continues, si le taux d'acceptation $TA_i(\mathcal{T})$ est supérieur à 20% au cours d'un palier de température, alors $\sigma_i^{(t)}$ est doublé, si ce taux est inférieur à 5%, alors il est divisé par deux. Les auteurs

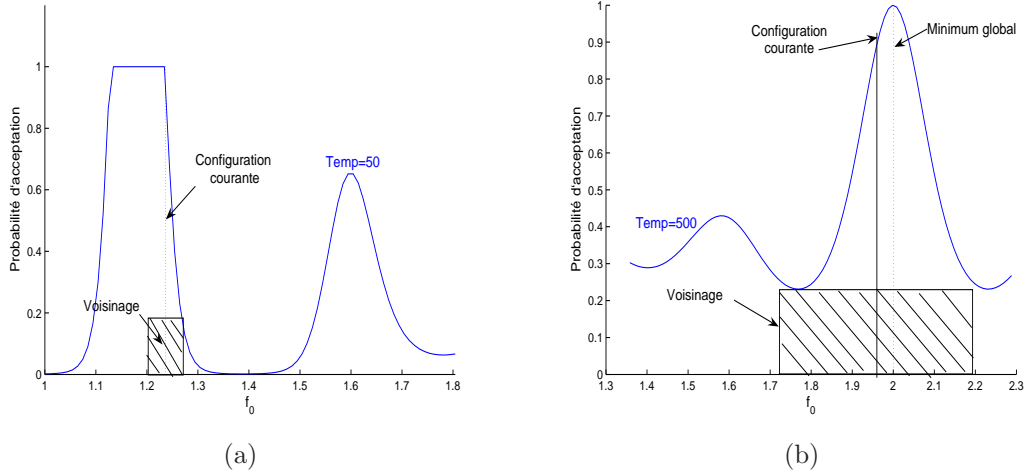


FIG. 3.9 – Importance du réglage de la variance de l'exploration de l'espace de configuration. (a) Cas d'une faible variance et d'un minimum global éloigné : risque de stagnation sur un minimum local. (b) Cas d'une variance élevée au voisinage de minimum global : perte de temps.

suggèrent que les paramètres soient modifiés (perturbés) par groupes de $\cong \frac{M_A + M_F + 3}{3}$ paramètres choisis aléatoirement.

- Dans [MMM06, GFR94], la décroissance de $\sigma_i^{(t)}$ doit équilibrer le taux d'acceptation à 50%. Si ce taux est supérieur à plus de 60%, alors $\sigma_i^{(t)}$ est augmenté, si il est moins de 40%, $\sigma_i^{(t)}$ est diminué. Les auteurs ne précisent pas la façon dont il procèdent pour ajuster le rayon de voisinage, mais dans leurs simulations, il s'agit souvent d'une multiplication ou une réduction par un facteur 2.

Nous avons choisi, la façon la plus simple, d'ajuster $\sigma_i^{(t)}$ au cours des itérations, c'est-à-dire la loi de décroissance suivante :

$$\sigma_i^{2^{(t+1)}} = (1 - \epsilon_{\sigma_i}) \sigma_i^{2^{(t)}}, \quad \text{où } \epsilon_{\sigma_i} = 0.97 \quad (3.23)$$

Néanmoins, nous proposons que cette décroissance ne soit pas exécutée d'une façon régulière, *i.e.* à chaque itération, mais plutôt à des instants choisis aléatoirement selon une loi de Bernoulli \mathcal{B} de paramètre ρ :

$$\text{Générer } \xi \sim \mathcal{B}(\rho), \quad \text{si } \xi = 1 \quad \text{alors diminuer le pas du voisinage.} \quad (3.24)$$

Le paramètre ρ est fixé empiriquement de façon à ce que le résultat final approche au mieux la solution optimale.

Dans la suite, nous noterons σ le vecteur

$$\sigma = [\sigma_0, \dots, \sigma_{M_A + M_F + 2}]^T$$

et sa valeur à l'itération t :

$$\sigma^{(t)} = [\sigma_0^{(t)}, \dots, \sigma_{M_A + M_F + 2}^{(t)}]^T. \quad (3.25)$$

3.4.8 Conditions d'arrêt de l'algorithme

L'étape d'arrêt de l'algorithme est une étape cruciale dans le recuit simulé. En effet, l'arrêt prématuré de l'algorithme peut impliquer une perte d'optimalité, de même un arrêt tardif lui laisse le temps d'explorer des zones de l'espace de recherche inutilement.

Dans la suite, nous proposons un critère d'arrêt adapté à notre problème et basé sur les caractéristiques statistiques du bruit $b[n]$ (2.1). Ce critère est développé pour deux cas différents. Dans le premier cas, la variance du bruit est supposée connue ; ce cas est purement théorique car, en pratique, on ne connaît que rarement la variance du bruit sur des données réelles. Toutefois, il va nous permettre de mieux régler les éléments intervenant dans le recuit simulé, tel que le nombre d'itérations, la décroissance de la température, *etc.* Le deuxième cas est celui général où la variance du bruit est supposée être inconnue.

3.4.8.1 Variance du bruit connue

Soit l'énergie du bruit :

$$\mathcal{E}_N = \sum_{n=-\frac{N}{2}}^{\frac{N}{2}} |b[n]|^2 \sim \sigma_b^2 \chi_{2(N+1)}^2. \quad (3.26)$$

où $\chi_{2(N+1)}^2$ est la loi de χ^2 à $2(N+1)$ degré de liberté. La moyenne E_N et la variance Σ_N^2 de la variable \mathcal{E}_N sont :

$$E_N = \mathbb{E}[\mathcal{E}_N] = \sigma_b^2(N+1), \quad \text{et} \quad \Sigma_N^2 = \sigma_b^4(N+1). \quad (3.27)$$

Sur la figure 3.10, sont représentées certaines densités de probabilité de \mathcal{E}_N en fonction du nombre d'échantillons $N+1$ et de la variance du bruit σ_b^2 .

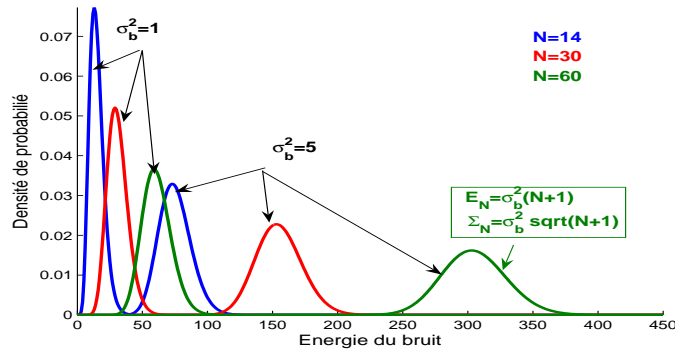


FIG. 3.10 – Loi de distribution de l'énergie $\sum_{n=-\frac{N}{2}}^{\frac{N}{2}} |b[n]|^2$

D'après la loi des grands nombres [Tas89], on admet statistiquement qu'à partir d'un degré de liberté supérieur à 30, la distribution d'une variable aléatoire χ^2 tend empiriquement vers une loi normale gaussienne. Tel est notre cas ($2(N+1) \geq 30$), c'est pourquoi on

a adopté un critère d'arrêt empirique et simple, qui fournit un intervalle de confiance pour l'énergie du signal résidu de l'estimation $\hat{b}[n] = y[n] - \hat{s}[n]$. Ce critère est défini par :

$$\hat{c}_N = \sum_{n=-\frac{N}{2}}^{\frac{N}{2}} \left| \hat{b}[n] \right|^2 \in [E_N - z_\alpha \Sigma_N, E_N + z_\alpha \Sigma_N], \quad (3.28)$$

où z_α est le quantile pour un taux de confiance α et pour une loi normale centrée réduite. Par exemple $z_\alpha = 1.96$ correspond à $\alpha = 5\%$ [Tas89].

3.4.8.2 Variance du bruit inconnue

Dans le cas où σ_b^2 est inconnue, nous utilisons un critère différent basé sur un test de blancheur sur le signal résidu de l'estimation $\hat{b}[n]$:

$$\hat{b}[n] = y[n] - \hat{s}[n] \quad (3.29)$$

où $\hat{s}[n]$ est l'estimation de $s[n]$ (2.1). Ce critère vérifie si le résidu $\hat{b}[n]$ est un processus décorrélé. Théoriquement, la covariance du résidu s'écrit

$$\Gamma[k] = \frac{1}{N+1} \sum_{n=-\frac{N}{2}}^{\frac{N}{2}-k} \Re\{b[n] b[n+k]^*\} \text{ pour } k = 0, \dots, k_0, \quad (3.30)$$

où $k_0 < N$, la covariance normalisée est définie par :

$$\gamma[k] = \frac{\Gamma[k]}{\Gamma[0]} \quad (3.31)$$

Asymptotiquement pour $N \gg k_0$, la covariance normalisée vérifie statistiquement :

$$\gamma[k] \sim \mathcal{N}\left(0, \frac{1}{N+1}\right) \text{ pour } k \neq 0. \quad (3.32)$$

Nous définissons alors un intervalle de confiance à 95% :

$$\gamma[k] \in \left[-\frac{1.95}{\sqrt{N+1}}, \frac{1.95}{\sqrt{N+1}} \right]. \quad (3.33)$$

Si les valeurs prises par $\gamma[k]$ se trouve dans cet intervalle, le résidu sera considéré blanc.

3.4.9 Convergence du recuit simulé

En ce qui concerne la convergence théorique du recuit simulé, il a été établi que ce dernier convergeait en probabilité vers un optimum global sous les conditions suivantes [Mic91, Mit86, HG89] :

- la réversibilité, c'est à dire le retour à une solution de l'itération précédente est permis,

- la connexité de l'espace de recherche, toute solution candidate potentielle peut être atteinte de n'importe quelle autre solution candidate moyennant un nombre fini de changements élémentaires.

Sous ces conditions la convergence est assurée indépendamment de l'état d'initialisation. D'autres conditions [AF87] ont été étudiées en fonction de la probabilité d'acceptation. Il était montré aussi qu'une ré-initialisation occasionnelle de l'algorithme permet d'augmenter la probabilité de convergence.

3.5 Algorithme du recuit simulé

Les principales étapes de l'algorithme du recuit simulé sont décrites dans Tab. 3.2. Il inclut une première étape d'initialisation des paramètres du modèle qu'on cherche à estimer. L'étape 2, qui est une étape itérative à l'issue de laquelle le test de blancheur sur le signal résidu est effectué, consiste d'abord en l'initialisation des paramètres de contrôle du recuit simulé c'est-à-dire la température et le pas de voisinage (étape 2.1). Ensuite, la génération aléatoire d'un candidat correspond à l'étape 2.2-a. L'acceptation de ce candidat est effectuée selon la règle de Metropolis (étape 2.2-b). L'adaptation du pas de voisinage et la décroissance de la température sont réalisées à l'étape 2.2-c. Enfin l'évaluation du signal résidu à l'étape 2.2-d, permet de la validation ou non du critère d'arrêt de l'algorithme.

TAB. 3.2 – Algorithme du recuit simulé pour la maximisation de la vraisemblance

-
1. Initialisation $\boldsymbol{\theta}^{(0)}$,
 - Évaluer $\hat{s}[n]$ pour $\boldsymbol{\theta}^{(0)}$, en utilisant (2.2) et (2.8).
 - Évaluer $\hat{b}[n] = y[n] - \hat{s}[n]$.
 2. Répéter jusqu'à $\hat{b}[n]$ vérifie le critère d'arrêt,
 - 2.1- Initialiser $\boldsymbol{\sigma}_0^{2(0)}$ et $\mathcal{T}^{(0)}$.
 - 2.2- Itérations de $t = 0$ à T
 - a- Générer $\boldsymbol{\theta}_c = \boldsymbol{\theta}^{(t)} + \Delta\boldsymbol{\theta}$ selon la loi gaussienne (3.20).
 - b- Générer $u \sim \mathcal{U}[0, 1]$, si $u < \exp(\frac{\ell_{MC}(\boldsymbol{\theta}^{(t)}) - \ell_{MC}(\boldsymbol{\theta}_c)}{\mathcal{T}})$, alors $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}_c$, sinon $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)}$.
 - c- Générer $\xi \sim \mathcal{B}(\rho)$. Si $\xi = 1$, mettre $\boldsymbol{\sigma}_c^{2(t+1)} = (1 - \epsilon_\sigma) \boldsymbol{\sigma}^{2(t)}$ et $\mathcal{T}^{(t+1)} = (1 - \epsilon_{\mathcal{T}}) \mathcal{T}^{(t)}$.
 - d- Évaluer $\hat{s}[n]$ pour $\boldsymbol{\theta}$. déterminer $\hat{b}[n] = y[n] - \hat{s}[n]$, puis retourner en 2.2.
 - 2.3- Retourner en 2.
-

Dans le Tableau 3.2, $\boldsymbol{\theta}_c$ désigne le vecteur dont les composantes sont des candidats potentiels pour chacun des paramètres et $\Delta\boldsymbol{\theta}$ est le vecteur de perturbation. $\mathcal{B}(\rho)$ représente une loi de Bernoulli, qui permet de gérer la décroissance de la température et du rayon de voisinage de façon aléatoire. Lors des applications sur des signaux, des exemples de valeurs de ρ seront donnés.

3.6 Conclusions

Dans ce chapitre, nous avons présenté la procédure de maximum de vraisemblance qui sera utilisée pour estimer les paramètres du modèle (2.2) et (2.8). La fonction de vraisemblance étant non-linéaire par rapport aux paramètres de la fréquence instantanée et de la phase initiale, elle présente plusieurs minima locaux. Les techniques d'optimisation classiques ne garantissent pas la convergence vers le minimum global, nous avons choisi d'utiliser une technique d'optimisation stochastique : le recuit simulé.

Nous avons donc présenté et détaillé les éléments régissant le fonctionnement du recuit simulé. Dans le Tableau 3.3, nous récapitulons leur rôle, les choix que nous avons retenus et les corrélations pouvant exister entre ces éléments.

Certes l'inconvénient du recuit simulé réside dans le choix et les réglages de tous ces éléments compte tenu des ces corrélations. En effet, il faut maintenir à peu près constante l'efficacité du processus d'optimisation tout au long de la décroissance de la température. Le mode d'évaluation de cette efficacité étant basé sur l'évaluation du taux d'acceptation, la taille du voisinage exploré et la fréquence d'ajustement de rayon du voisinage.

Toutefois, si le réglage est bien fait, la convergence du recuit simulé est assurée et l'obtention d'une solution optimale en un temps de calcul raisonnable est possible. De plus, grâce au développement des outils informatiques et de la puissance de calcul des ordinateurs, l'emploi de cette technique parfois gourmande en temps de calcul n'est plus problématique.

Dans le chapitre 5, nous allons mettre en application l'algorithme du recuit simulé et étudier sa robustesse en fonction des taux et des indices de modulations, du *RSB*, de la taille des échantillons et du degré de l'approximation polynomiale.

Eléments	Rôles	Choix	lié à
Fonction objectif	à minimiser	$\ell_{MC}(\boldsymbol{\theta})$ ou $\gamma_{y,s}(\boldsymbol{\theta})$	$\boldsymbol{\theta}$
Espace de recherche	à être exploré	$\mathbb{R}^{M_A+M_F+3}$	$\ell_{MC}(\boldsymbol{\theta})$
Température	paramètre de contrôle	$\mathcal{T}^{(t+1)} = (1 - \epsilon_{\mathcal{T}})\mathcal{T}^{(t)}$	$\sigma_i^{(t)}$
Rayon du voisinage	efficacité de l'exploration	$\sigma_c^{2(t+1)} = (1 - \epsilon_{\sigma})\sigma_c^{2(t)}$	$\mathcal{T}, \ell_{MC}(\boldsymbol{\theta})$
Mécanisme de perturbation	explorer les solutions	$\theta_{c,i} \sim \mathcal{N}(\theta_i, \sigma_i^{(t)})$	$\sigma_i^{(t)}, \mathcal{T}_i$
Règle d'acceptation	chaîne de Markov équilibrée	Règle de Metropolis	$\mathcal{T}, \ell_{MC}(\boldsymbol{\theta})$

TAB. 3.3 – Eléments intervenants dans le fonctionnement du recuit simulé et le choix que nous avons retenu pour l'appliquer à notre problématique.

Chapitre 4

Signal court mono-composante non-stationnaire - Approche Bayésienne

Résumé

Dans ce chapitre, les paramètres du modèle polynomial d'un signal court non-stationnaire et mono-composante sont estimés par une approche Bayésienne, utilisant les méthodes Monte Carlo par Chaînes de Markov.

Les paramètres du modèle sont donc assimilés à des variables aléatoires, leur densité de probabilité *a priori* est supposée non-informative. Nous déterminons, à une constante de normalisation près, leur densité de probabilité *a posteriori* conditionnellement aux observations. Les paramètres sont ensuite estimés en simulant cette densité *a posteriori* à partir de l'algorithme Metropolis-Hastings à marche aléatoire.

Mots clefs

Inférence Bayésienne, méthodes MCMC, densité de probabilité *a priori*, densité de probabilité *a posteriori*, Metropolis Hastings, estimateurs : MMSE, MAP.

Sommaire

4.1	Introduction	81
4.2	Inférence Bayésienne	82
4.2.1	Généralités	82
4.2.2	Choix d'une loi <i>a priori</i>	83
4.2.2.1	Lois <i>a priori</i> conjuguées	83
4.2.2.2	Lois <i>a priori</i> invariantes	84
4.2.2.3	Loi <i>a priori</i> de Jeffreys	84
4.2.3	Estimateurs Bayésiens	85
4.2.3.1	Estimateur de la moyenne <i>a posteriori</i>	86
4.2.3.2	Estimateur du maximum <i>a posteriori</i>	86
4.3	Méthodes de Monte Carlo par Chaînes de Markov	87
4.3.1	Introduction	87
4.3.2	Chaînes de Markov : de la théorie aux méthodes MCMC	88
4.3.3	Choix de l'algorithme de simulations MCMC	89
4.4	Algorithme de Metropolis Hastings	89
4.4.1	Schéma itératif d'un algorithme Metropolis Hastings	90
4.4.2	Conditions sur la loi instrumentale	91
4.4.3	Convergence des algorithmes Metropolis-Hastings	92
4.5	Quelques algorithmes Metroplis-Hastings	93
4.5.1	Algorithme Metropolis Hastings à marche aléatoire	93
4.5.2	Algorithme Metropolis Hastings dit une-variable-à-la-fois	94
4.6	Estimation des signaux courts mono-composantes non-stationnaires	95
4.6.1	Positionnement	95
4.6.2	Problématique et inférence Bayésienne	96
4.6.3	Algorithmes Metropolis Hastings une-variable-à-la-fois et à marche aléatoire	98
4.7	Conclusions	100

4.1 Introduction

De façon similaire à l'estimateur par maximum de vraisemblance et à d'autres approches proposées dans [Aka83, Bjo90], le formalisme Bayésien permet la mise en oeuvre du principe de vraisemblance. Le choix Bayésien consiste à représenter les paramètres inconnus du problème étudié avec une structure de probabilité et à tenir compte de l'information *a priori* et de son imprécision dans la procédure inférentielle. Les avantages les plus importants de ce formalisme, par comparaison avec les autres approches, sont les suivants :

- La possibilité d'incorporer des informations *a priori* telles que la restriction de l'espace de définition des paramètres,
- le calcul des estimateurs de Bayes est facilité par les méthodes de simulation de Monte Carlo. Ces estimateurs ont l'avantage d'être parfaitement définis pour un nombre faible d'observations. De plus, si ces estimateurs sont uniques alors ils sont admissibles [Rob06].
- les paramètres de nuisance de la fonction de vraisemblance, par exemple la variance du bruit, peuvent être éliminés par marginalisation,

Dans ce chapitre, nous abordons sous une approche Bayésienne l'estimation des paramètres du modèle défini par les équations (2.8) et (2.10) dans le chapitre 2. Nous rappelons que ce modèle décrit un signal mono-composante, non-stationnaire et de courte durée ; son amplitude et sa fréquence instantanées sont modélisées par des polynômes de degrés inférieurs à 3. Les paramètres du modèle vérifient les contraintes (2.12), à savoir l'amplitude instantanée est strictement positive et la fréquence instantanée est positive et limitée supérieurement par la fréquence de Nyquist. Le choix Bayésien est en adéquation avec ce contexte compte tenu :

- des contraintes non-linéaires imposées sur l'espace de définition des paramètres,
- de la présence d'un paramètre de nuisance, la variance du bruit σ_b^2 , dans la fonction de vraisemblance (3.3)
- et surtout du faible nombre d'observations.

Les autres approches [Aka83, Bjo90] sont moins pertinentes dans ce contexte, puisqu'elles nécessitent un nombre d'observations très grand devant le nombre de paramètres du modèle considéré.

Dans un premier temps, nous rappelons les définitions et les notions indispensables à la compréhension du formalisme Bayésien et les estimateurs de Bayes qui découlent de celui-ci. Ensuite, nous exposons quelques méthodes de calcul Bayésien basées sur les méthodes de simulations de Monte Carlo par chaînes de Markov. Enfin, la dernière partie sera consacrée à l'utilisation de l'approche Bayésienne pour estimer des signaux courts mono-composantes et non-stationnaires. Nous nous intéressons, particulièrement, à l'implémentation de l'algorithme de Metropolis-Hastings à marche aléatoire et dit une-variable-à-la-fois, pour réaliser l'estimation des paramètres du modèle.

4.2 Inférence Bayésienne

4.2.1 Généralités

Soit \mathbf{y} un vecteur d'observations suivant une loi de probabilité paramétrique connue $\text{Prob}(\mathbf{Y} = \mathbf{y}|\boldsymbol{\theta})$ où $\boldsymbol{\theta}$ représente le vecteur des paramètres inconnus. Pour simplifier les notations, cette probabilité sera désignée par $p(\mathbf{y}|\boldsymbol{\theta})$. Ce modèle peut être représenté par :

$$\mathbf{y} \sim p(\mathbf{y}|\boldsymbol{\theta}).$$

La version Bayésienne du principe de vraisemblance implique que l'inférence sur $\boldsymbol{\theta}$ dépend entièrement de la loi *a posteriori* $\mathcal{P}(\boldsymbol{\theta}|\mathbf{y})$ même si $\boldsymbol{\theta}$ n'est pas nécessairement conçu comme vecteur (ou variable) aléatoire. Cette loi *a posteriori* peut être construite conditionnellement aux observations quand une loi *a priori* $\mathcal{P}(\boldsymbol{\theta})$, exprimant les connaissances *a priori* sur $\boldsymbol{\theta}$ avant les observations, est disponible.

En particulier, considérons un ensemble d'échantillons aléatoires i.i.d. $\mathbf{y} = [y_1, \dots, y_N]^T$, dont la fonction de vraisemblance s'écrit $\ell(\mathbf{y}, \boldsymbol{\theta}) = p(\mathbf{y}|\boldsymbol{\theta}) = \prod_{n=1}^N p(y_n|\boldsymbol{\theta})$. En utilisant la formule de Bayes, nous obtenons la densité *a posteriori* de ces paramètres conditionnellement aux observations :

$$\mathcal{P}(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\theta}) \mathcal{P}(\boldsymbol{\theta})}{p(\mathbf{y})}, \quad (4.1)$$

où le dénominateur ne dépend pas de $\boldsymbol{\theta}$. $p(\mathbf{y})$ est une constante d'intégration appelée évidence Bayésienne et elle est définie par :

$$P(\mathbf{y}) = \int_{\boldsymbol{\theta} \in \Theta} p(\mathbf{y}|\boldsymbol{\theta}) \mathcal{P}(\boldsymbol{\theta}) d\boldsymbol{\theta}, \quad (4.2)$$

où Θ est l'espace de définition de $\boldsymbol{\theta}$. Nous écrivons donc

$$\mathcal{P}(\boldsymbol{\theta}|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\theta}) \mathcal{P}(\boldsymbol{\theta}). \quad (4.3)$$

Au niveau des concepts, le formalisme Bayésien (4.1) est simple et flexible. Le problème pratique est :

- d'abord de spécifier la loi *a priori*,
- ensuite de caractériser la vraisemblance,
- et enfin de calculer la loi *a posteriori*.

Nous verrons, dans la suite, que le choix de la loi *a priori* est l'étape la plus difficile et la plus importante dans la détermination de l'inférence Bayésienne. Une fois cette loi connue, l'équation (4.1) permet d'exprimer la loi *a posteriori*.

Bien souvent, la détermination analytique exacte de la constante d'intégration peut conduire à des calculs compliqués surtout quand $\boldsymbol{\theta}$ est un vecteur multidimensionnel, nous présentons alors des méthodes de calcul Bayésien permettant de contourner ce problème. Ces méthodes seront aussi employées dans l'évaluation des quantités *a posteriori* telles

que la moyenne, le mode, la variance et la médiane *a posteriori*, qui sont des estimateurs potentiels de θ . Le choix entre ces estimateurs *a posteriori* est impossible sans critères de coût, ce dernier point sera aussi discuté.

4.2.2 Choix d'une loi *a priori*

La loi *a priori* traduit les connaissances sur les paramètres du problème avant l'observation des données, ainsi que l'incertitude liée à ces informations. Sa détermination est fondamentale dans l'analyse Bayésienne.

Cependant, en pratique, les informations disponibles ne sont pas assez précises pour spécifier exactement cette loi. De plus, quand l'espace de définition des paramètres n'est pas dénombrable, il est généralement très difficile de proposer une forme exacte de la vraie loi *a priori*. On se contente alors d'une approximation dont le choix peut se révéler délicat et affecter le résultat de l'inférence Bayésienne, surtout en présence d'un nombre d'observations faible. Quand ce nombre d'observations tend vers l'infini, la plupart des lois *a priori* mènent asymptotiquement à la même inférence, celle-ci devient alors équivalente à celle fondée sur la fonction de vraisemblance (cf. chapitre 3).

Le plus souvent, le choix des lois *a priori* se fait selon leur caractère informatif ou non-informatif [Rob06]. Nous présentons celles les plus utilisées dans le cas d'une connaissance *a priori* vague. Leurs avantages et leurs inconvénients seront aussi décrits. Une vue plus complète sur ces lois est présentée dans [Vie99, JJDS02, Rob06].

4.2.2.1 Lois *a priori* conjuguées

Ce sont des lois *a priori* informatives et paramétrées dont on exige un apport d'informations subjectives le plus limité possible [Rob06]. Leurs paramètres sont souvent appelés hyper-paramètres.

Définition 2 Une famille \mathcal{F} de distributions de probabilité sur Θ est dite conjuguée par une fonction de vraisemblance $p(\mathbf{y}|\theta)$ si, pour toute loi *a priori* $\mathcal{P}(\theta) \in \mathcal{F}$, la distribution *a posteriori* $\mathcal{P}(\theta|\mathbf{y})$ appartient également à \mathcal{F} .

Des exemples classiques sont les lois *a priori* conjuguées naturelles des familles exponentielles. Celles-ci incluent les lois de probabilité usuelles telles que la loi normale, Gamma, etc. Elles sont très commodes dans les calculs analytiques mais le développement des méthodes numériques réduit cet avantage au profit de la pertinence de la loi *a priori*.

Leur inconvénient essentiel est qu'elles obéissent à des contraintes techniques de calculs plutôt qu'à des impératifs d'adéquation à l'information *a priori* disponible. En effet, une partie de cette information *a priori* peut être ignorée si elle n'est pas complètement compatible avec la structure de la loi *a priori* conjuguée choisie. En l'absence d'information *a priori*, les hyper-paramètres de ces lois ne peuvent être déterminés qu'arbitrairement, ce qui rajoute de la complexité à la loi du modèle. Par ailleurs, ces lois ne sont pas forcément

les plus robustes en présence d'un nombre d'observations faible. Dans ce cas, l'impact du choix de la loi *a priori* peut être considérable sur la distribution *a posteriori* [Rob06].

4.2.2.2 Lois *a priori* invariantes

Il s'agit de lois *a priori* non-informatives, où l'absence d'information *a priori* se traduit par une propriété d'invariance par rapport à une famille de transformations de Θ . Deux exemples couramment employés sont cités ci-dessous :

Exemple 1 La famille des lois invariantes par translation, *i.e.* la loi *a priori* doit vérifier $\mathcal{P}(\boldsymbol{\theta}) = \mathcal{P}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$ pour tout $\boldsymbol{\theta}_0 \in \Theta$. Par conséquent, la loi *a priori* solution de cette équation est la loi uniforme sur Θ , définie par :

$$\mathcal{P}(\boldsymbol{\theta}) = \mathbf{c}, \quad (4.4)$$

où \mathbf{c} est un vecteur à valeurs constantes. $\boldsymbol{\theta}$ est interprété comme étant un “vecteur de position”. Lorsque Θ n'est pas dénombrable, par exemple \mathbb{R} sans restriction, la loi *a priori* obtenue est dite impropre ou généralisée [Rob06] car elle ne conduit pas à une mesure de probabilité finie :

$$\int_{\Theta} \mathcal{P}(\boldsymbol{\theta}) d\boldsymbol{\theta} = \infty. \quad (4.5)$$

Néanmoins, l'emploi de ce type de lois *impropres* se justifie, d'une part, par les bonnes performances en général des estimateurs issus de leur utilisation et, d'autre part, par le fait que tant que la distribution *a posteriori* reste définie, les méthodes Bayésiennes s'appliquent toujours [Rob06].

Exemple 2 La famille des lois invariantes par changement d'échelle, ayant la forme $\mathcal{P}(\boldsymbol{\theta}) = \frac{1}{c} \mathcal{P}(\frac{\boldsymbol{\theta}}{c})$, où c est une constante strictement positive. La loi *a priori* solution de cette équation s'écrit alors

$$\mathcal{P}(\boldsymbol{\theta}) \propto \frac{1}{\boldsymbol{\theta}}. \quad (4.6)$$

Ces lois invariantes non-informatives n'expriment pas une ignorance *totale* sur les paramètres étudiés mais elles constituent des lois de références, auxquelles on aura recours en l'absence d'information *a priori*. A cet égard, certaines lois non-informatives peuvent être plus efficaces que d'autres pourtant informatives.

4.2.2.3 Loi *a priori* de Jeffreys

Les lois *a priori* de Jeffreys sont fondées sur l'information de Fisher [Jef61, Rob06] et elles sont déterminées par

$$\mathcal{P}(\boldsymbol{\theta}) \propto |\det \mathbf{I}_{\boldsymbol{\theta}}|^{\frac{1}{2}}. \quad (4.7)$$

\mathbf{I}_θ est la matrice d'information de Fisher définie à partir de la vraisemblance $p(\mathbf{y}|\theta)$ par :

$$[\mathbf{I}_\theta]_{i,j} = \mathbb{E} \left[\frac{\partial \ln p(\mathbf{y}|\theta)}{\partial \theta_i} \frac{\partial \ln p(\mathbf{y}|\theta)}{\partial \theta_j} \right]. \quad (4.8)$$

Sous certaines hypothèses de régularités, l'équation (4.8) peut s'écrire :

$$[\mathbf{I}_\theta]_{i,j} = -\mathbb{E} \left[\frac{\partial^2 \ln p(\mathbf{y}|\theta)}{\partial \theta_i \partial \theta_j} \right]. \quad (4.9)$$

La loi de Jeffreys ainsi définie n'est pas invariante au sens de l'invariance par une famille de transformations mais par rapport au choix de la paramétrisation, c'est-à-dire, on peut considérer le modèle paramétré soit par le vecteur θ , soit par $\lambda = h(\theta)$, l'image de θ par une fonction bijective h .

Par construction (4.7), la loi de Jeffreys est non-informative puisqu'elle favorise les vecteurs θ correspondants à une grande quantité d'information, donc, ceux qui sont naturellement les plus probables. A vrai dire, la matrice de Fisher mesure, par définition, la quantité d'information apportée par les observations et caractérise la capacité du modèle à discriminer entre θ et $\theta + \delta\theta$.

Bien que pratique, cette loi n'a de sens que si le vecteur de paramètres θ est réel. Elle est fréquemment impropre (4.5) et dépend du nombre d'observations. Il est également déconseillé de l'employer dans les cas multidimensionnel en raison des incohérences qui peuvent apparaître [Rob06]. D'autres inconvénients de la loi de Jeffreys sont cités dans [Rob06].

A l'issue de cette représentation succincte de quelques lois *a priori*, nous rappelons tout de même que l'assignation de ces dernières dépend des données du problème et qu'il n'y a pas un choix dominant d'une loi *a priori* par rapport à un autre. De toute façon, dans l'inférence Bayésienne, il est rare de pouvoir déterminer exactement la loi *a priori* correspondante au problème étudié. Dans la plupart des cas, une certaine imprécision demeure.

Cependant, il est important de toujours s'assurer que l'impact de loi *a priori* choisie ne soit pas prédominant dans l'évaluation de la loi *a posteriori* (4.3), ce qui revient à effectuer une analyse de robustesse [Vie99, Rob06]. Une fois la loi *a priori* fixée, nous nous préoccupons des estimateurs Bayésiens potentiels du vecteur de paramètres θ . Le paragraphe suivant rappelle certains de ces estimateurs et énumère leurs propriétés.

4.2.3 Estimateurs Bayésiens

Il est possible d'associer un estimateur Bayésien $\hat{\theta}$ et une évaluation de la précision de l'estimation à une loi *a posteriori* quasi-disponible à une constante de normalisation près (4.3). Un estimateur Bayésien est toujours associé à un coût \mathbf{C} et à une loi *a priori*. Il est

défini comme l'estimateur qui minimise le risque de Bayes (coût moyen) qui s'écrit :

$$\mathcal{R} = \int \left[\int \mathcal{C}(\epsilon_\theta) p(\mathbf{y}|\boldsymbol{\theta}) d\mathbf{y} \right] \mathcal{P}(\boldsymbol{\theta}) d\boldsymbol{\theta}. \quad (4.10)$$

$\epsilon_\theta = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$ représente l'erreur d'estimation pour un vecteur d'observations \mathbf{y} . Une discussion plus complète sur le concept de ce risque et sur les différents estimateurs paramétriques Bayésiens est présentée dans [Vie99, Rob06, JJDS02]. Nous présentons, ici, les estimateurs Bayésiens les plus utilisés à savoir la moyenne *a posteriori* et l'estimateur du maximum *a posteriori*.

4.2.3.1 Estimateur de la moyenne *a posteriori*

Proposition 1 Soit $\Theta \in \mathbb{R}^d$, où d est la dimension de vecteur $\boldsymbol{\theta}$. L'estimateur de Bayes associé à la loi a priori $\mathcal{P}(\boldsymbol{\theta})$ et au coût quadratique, $\epsilon_\theta^\dagger Q \epsilon_\theta$, est la moyenne *a posteriori* $\hat{\boldsymbol{\theta}} = \mathbb{E}[\boldsymbol{\theta}|\mathbf{y}]$, pour toute matrice Q symétrique définie positive.

Cet estimateur, noté couramment MMSE (*Minimum Mean Squared Error*), est explicité par :

$$\hat{\boldsymbol{\theta}}_{MMSE} = \frac{\int_{\Theta} \boldsymbol{\theta} p(\mathbf{y}|\boldsymbol{\theta}) \mathcal{P}(\boldsymbol{\theta}) d\boldsymbol{\theta}}{\int_{\Theta} p(\mathbf{y}|\boldsymbol{\theta}) \mathcal{P}(\boldsymbol{\theta}) d\boldsymbol{\theta}}. \quad (4.11)$$

Le coût quadratique est simple et particulièrement intéressant quand l'espace des paramètres est borné. Cependant, ce coût pénalise fortement les grandes erreurs et l'estimateur qui lui est associé n'est pas invariant. De plus, il nécessite une intégration globale, complexe et difficile à déterminer analytiquement.

4.2.3.2 Estimateur du maximum *a posteriori*

Considérons la fonction de coût 0 – 1, qui vaut $\mathcal{C}(\epsilon_\theta) = 0$ si $|\epsilon_\theta| \leq \epsilon_l$ et vaut $\mathcal{C}(\epsilon_\theta) = 1$ dans le cas contraire. ϵ_l est une constante fixée par l'utilisateur.

Proposition 2 L'estimateur de Bayes associé à la loi a priori $\mathcal{P}(\boldsymbol{\theta})$ et au coût 0 – 1, est l'estimateur du maximum *a posteriori*, noté MAP.

L'estimateur MAP constitue le mode de la densité *a posteriori* :

$$\hat{\boldsymbol{\theta}}_{MAP} = \arg \max_{\boldsymbol{\theta} \in \Theta} \mathcal{P}(\boldsymbol{\theta}|\mathbf{y}). \quad (4.12)$$

Son évaluation nécessite uniquement la résolution d'un problème d'optimisation et ne requiert pas l'évaluation de la constante d'intégration (4.2). L'équation (4.12) se réduit alors à :

$$\hat{\boldsymbol{\theta}}_{MAP} = \arg \max_{\boldsymbol{\theta} \in \Theta} p(\mathbf{y}|\boldsymbol{\theta}) \mathcal{P}(\boldsymbol{\theta}). \quad (4.13)$$

Cet estimateur vérifie le principe d'invariance et possède les propriétés d'optimalité asymptotiques de l'estimateur du maximum de vraisemblance classique (cf. chapitre 3). En présence d'un bruit gaussien et d'une densité *a priori* uniforme des paramètres, le MAP est équivalent à l'estimateur du maximum de vraisemblance mais il a l'avantage d'être parfaitement défini pour un nombre faible d'observations. De plus, si la densité *a posteriori* est unimodale et symétrique, le MAP coïncide aussi avec le MMSE.

La précision de l'estimation et les performances des estimateurs MMSE et MAP peuvent être caractérisées en utilisant l'erreur quadratique moyenne *a posteriori* dans un cas unidimensionnel et la matrice de covariance dans le cas multidimensionnel. Il existe, aussi, des bornes Bayésiennes minimales d'estimation comme celles de Cramer-Rao utilisées dans l'évaluation des estimateurs non-biaisés [Tre68, BSET97, CFL04, RFL05].

Des difficultés peuvent apparaître lors de l'implémentation pratique de ces estimateurs et de la loi *a posteriori* correspondante, surtout si le nombre de paramètres à estimer est très élevé. Par exemple, l'expression de la loi *a posteriori* peut ne pas être simple voire non-explicite, le calcul des estimateurs peut faire intervenir des intégrations multidimensionnelles, *etc.* Hormis la puissance de calcul considérable nécessaire, les méthodes d'intégration numériques deviennent très vite instables, donc peu fiables, lorsque les dimensions du problème augmente. Au delà de la dimension 3, il est souvent préférable d'utiliser des méthodes de simulation.

La partie suivante est consacrée à des méthodes de simulation puissantes basées sur les méthodes de Monte-Carlo par chaînes de Markov.

4.3 Méthodes de Monte Carlo par Chaînes de Markov

4.3.1 Introduction

Les méthodes de simulation statistique sont apparues pendant les années 1940 dans le domaine de la physique. Elles permettent de résoudre des problèmes d'optimisation en exploitant les propriétés probabilistes du problème considéré. Depuis 1979 et grâce au développement des outils informatiques, les domaines d'application de ces méthodes ne cessent de s'étendre jusqu'à atteindre le domaine du traitement du signal en 1990. Les méthodes de simulation de Monte-Carlo se basent essentiellement sur le théorème ergodique, qui stipule le résultat ci-dessous.

Proposition 3 Soient des échantillons *i.i.d.* $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_K)$ d'un vecteur aléatoire $\boldsymbol{\theta}$, générés selon la loi *a posteriori* $\mathcal{P}(\boldsymbol{\theta}|\mathbf{y})$, et une fonction continue h . Par la loi des grands nombres, on a alors

$$\frac{1}{K} \lim_{K \rightarrow \infty} \sum_{k=1}^K h(\boldsymbol{\theta}_k) = \mathbb{E}[h(\boldsymbol{\theta})|\mathbf{y}].$$

Ces échantillons peuvent être utilisés pour calculer les moments, les cumulants, des estimateurs, *etc.* L'indépendance des échantillons n'est pas fondamentale si on s'intéresse, principalement, aux caractéristiques de la loi *a posteriori* [Rob06].

Malheureusement, la génération directe d'échantillons selon la loi *a posteriori*, *i.e.* sa simulation directe, n'est pas facile par les algorithmes usuels. C'est pourquoi nous faisons appel aux algorithmes basés sur les méthodes de Monte-Carlo par chaînes de Markov (MCMC). Bien sûr, il existe d'autres méthodes de simulation de Monte-Carlo mais les méthodes MCMC ont l'avantage d'être générales et convergentes, du moins sur le plan théorique. Elles sont performantes et efficaces dans la simulation des densités *a posteriori* complexes, non-standards et multidimensionnelles, dont l'expression est connue à une constante de normalisation près.

Dans la suite, nous rappelons d'une façon succincte les notions les plus importantes pour la compréhension des chaînes de Markov et des méthodes MCMC. Ensuite, nous présentons les principales caractéristiques des algorithmes MCMC et précisément celles des algorithmes de Metropolis-Hastings. Ces derniers nous paraissent les plus appropriés au contexte des signaux courts non-stationnaires. Leur implémentation, relative à ce contexte, fera l'objet de la dernière partie de ce chapitre. Par ailleurs, nous renvoyons le lecteur à [Rob96, Rob06] pour des exposés plus complets et commentés sur les méthodes MCMC.

4.3.2 Chaînes de Markov : de la théorie aux méthodes MCMC

Une chaîne de Markov est une séquence de variables aléatoires $\{\boldsymbol{\theta}^{(t)}\}_{t \in \mathbb{N}}$ produite par une densité conditionnelle $\mathcal{K}(\boldsymbol{\theta}^{(t-1)}, \boldsymbol{\theta}^{(t)})$, appelée noyau de transition, selon le schéma suivant :

$$\boldsymbol{\theta}^{(t)} \sim \mathcal{K}(\boldsymbol{\theta}^{(t-1)}, \boldsymbol{\theta}^{(t)}) = p(\boldsymbol{\theta}^{(t)} | \boldsymbol{\theta}^{(t-1)}). \quad (4.14)$$

Ce noyau de transition détermine complètement les propriétés de la chaîne générée. Si cette dernière est irréductible, apériodique et réversible alors elle admet une unique distribution limite, qu'on appelle distribution invariante ou stationnaire. Dans ce cas, en partant d'un candidat initial choisi arbitrairement et après une certaine période transitoire appelée souvent période de chauffage, la chaîne générée selon le schéma (4.14) ne dépend plus du candidat initial ; la t -ième itération converge vers la distribution invariante quand t tend vers l'infini. Pour de plus amples détails sur la théorie des chaînes de Markov, le lecteur peut se référer à [Gre68, MT93, Tie94, CG95, Rob96].

Cette théorie a été le point de départ pour l'élaboration de plusieurs algorithmes MCMC, permettant d'obtenir des échantillons aléatoires distribués suivant une loi d'intérêt sans la simuler directement ; cette loi peut être disponible uniquement à une constante multiplicative près.

Définition 3 *On appelle méthode MCMC une méthode permettant de générer effectivement une chaîne de Markov ergodique (positive, récurrente et apériodique donc irréductible), de loi invariante la loi de probabilité d'intérêt.*

Les chaînes de Markov produites par les algorithmes MCMC bénéficient donc, par construction, de propriétés de stabilité forte à savoir l'existence de la distribution limite et invariante

\mathcal{P} , telle que $\boldsymbol{\theta}^{(t)} \sim \mathcal{P}$, si $\boldsymbol{\theta}^{(t-1)} \sim \mathcal{P}$ pour t grand.

Grâce à cette stabilité, la plupart des chaînes de Markov générées avec les algorithmes MCMC sont récurrentes au sens de Harris, *i.e.* la probabilité d'un nombre infini de visites dans un ensemble arbitraire de Θ vaut 1 [Rob96, JJDS02]. Cela assure, d'une part, que la chaîne générée a les mêmes propriétés quelque soit le vecteur initial $\boldsymbol{\theta}^{(0)}$ et garantit, d'autre part, qu'elle converge vers sa distribution invariante quel que soit l'initialisation. Une conséquence importante de cette convergence est que le théorème ergodique (Proposition 3) est vérifié. Des formes plus fortes de convergence des chaînes de Markov issues des simulations par les méthodes MCMC peuvent également être établies comme la convergence géométrique et celle uniforme [Rob96].

Malheureusement, le noyau de transition $\mathcal{K}(\boldsymbol{\theta}^{(t-1)}, \boldsymbol{\theta}^{(t)})$ étant souvent inconnu, on procède en pratique de la manière suivante. À partir d'un candidat initial $\boldsymbol{\theta}^{(0)}$, on génère d'une manière récursive une chaîne $\boldsymbol{\theta}^{(t)}$ en utilisant une loi instrumentale facile à simuler (appelée également loi génératrice de candidats) et une règle d'acceptation/rejet. Cette loi instrumentale doit garantir la convergence vers la loi d'intérêt et la réversibilité de la chaîne. Après un nombre d'itérations très élevé, on peut considérer que la chaîne générée a atteint sa loi limite, les échantillons obtenus à partir de ce stade sont donc indépendants de l'initialisation et sont distribués asymptotiquement suivant la loi d'intérêt. Notons, tout de même, que ces échantillons ne sont pas i.i.d. suivant cette loi. Pour obtenir de tels échantillons, il faudrait simuler plusieurs chaînes de Markov en utilisant des initialisations différentes.

4.3.3 Choix de l'algorithme de simulations MCMC

Dans la littérature, on peut trouver des algorithmes très variés basés sur les méthodes MCMC, du fait de leur mise en oeuvre simple et de leur adaptabilité. Nous citons par exemple, les MCMC à sauts réversibles, les MCMC séquentiels [Vie99, AD99, Sén02, DMD05].

Cependant, dans ce chapitre, nous ne nous intéressons qu'aux algorithmes Métropolis-Hastings (MH) ne requérant qu'une connaissance limitée de la loi d'intérêt. Ces algorithmes sont suffisants dans le cas de notre problème. Nous présentons leur schéma classique ainsi que quelques exemples de leur implémentation, en vue de les utiliser dans le cadre de l'estimation d'un signal court non-stationnaire.

4.4 Algorithme de Metropolis Hastings

Les algorithmes de Metropolis-Hastings (MH) sont des méthodes MCMC qui ont été développées par Metropolis et al. [MRR⁺53] en 1953 dans un contexte d'optimisation sur un espace discret. Ils ont été généralisés à la simulation statistique par Hastings en 1970. Dans [Gey92, CG95, Vie99, Rob96, Rob06], une description détaillée de différentes versions des algorithmes MH est fournie.

4.4.1 Schéma itératif d'un algorithme Metropolis Hastings

L'algorithme de MH repose sur l'utilisation d'une densité instrumentale conditionnelle $q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ vérifiant l'équation

$$\int_{\Theta} q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) d\boldsymbol{\theta} = 1.$$

Cette densité régit la génération des éléments successifs de la chaîne de Markov et permet à cette dernière de bouger dans l'espace de définition des paramètres. Elle doit être rapidement simulable et soit symétrique (par rapport à l'élément courant), soit disponible analytiquement à une constante de normalisation près. Le choix de la densité instrumentale est déterminant dans l'algorithme et il sera discuté dans le paragraphe suivant.

Le schéma itératif [Gre68, Rob96] de l'algorithme MH peut se résumer en deux étapes principales :

1. Partant d'un état initial arbitraire $\boldsymbol{\theta}^{(0)}$, générer un candidat $\boldsymbol{\theta}_c$ selon la loi instrumentale $q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$, où $\boldsymbol{\theta}^{(t)}$ est l'élément courant de la chaîne de Markov :

$$\boldsymbol{\theta}_c \sim q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}).$$

2. Evaluer la probabilité d'acceptation $\alpha(\boldsymbol{\theta}_c, \boldsymbol{\theta}^{(t)})$ définie par [CG95, Rob96]

$$\alpha(\boldsymbol{\theta}_c, \boldsymbol{\theta}^{(t)}) = \min \{ \rho(\boldsymbol{\theta}_c, \boldsymbol{\theta}^{(t)}), 1 \} \quad (4.15)$$

avec

$$\rho(\boldsymbol{\theta}_c, \boldsymbol{\theta}^{(t)}) = \frac{\mathcal{P}(\boldsymbol{\theta}_c|\mathbf{y}) q(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}_c)}{\mathcal{P}(\boldsymbol{\theta}^{(t)}|\mathbf{y}) q(\boldsymbol{\theta}_c|\boldsymbol{\theta}^{(t)})}. \quad (4.16)$$

Le candidat généré est, accepté selon la règle de Metropolis, c'est-à-dire :

$$\text{Générer } u \sim \mathcal{U}[0, 1], \quad \begin{array}{l} \text{si } u \leq \alpha(\boldsymbol{\theta}_c, \boldsymbol{\theta}^{(t)}) \text{ alors } \boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}_c \\ \text{sinon} \quad \boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)}, \end{array} \quad (4.17)$$

\mathcal{U} étant la loi uniforme sur l'intervalle $[0, 1]$.

L'équation (4.16) suppose la non nullité de $\mathcal{P}(\boldsymbol{\theta}^{(t)}|\mathbf{y})$; il convient toujours d'initialiser la chaîne par un candidat $\boldsymbol{\theta}^{(0)}$ vérifiant $\mathcal{P}(\boldsymbol{\theta}^{(0)}|\mathbf{y}) > 0$. Ceci implique que $\mathcal{P}(\boldsymbol{\theta}^{(t)}|\mathbf{y})$ est strictement positive pour tout $t \in \mathbb{N}$ puisque, compte tenu de la règle d'acceptation utilisée (4.17), les candidats vérifiant $\mathcal{P}(\boldsymbol{\theta}^{(t)}|\mathbf{y}) = 0$ sont systématiquement rejetés par l'algorithme. Pour le reste, les candidats générés sont acceptés comme éléments successifs de la chaîne de Markov avec une probabilité $\alpha(\boldsymbol{\theta}_c, \boldsymbol{\theta}^{(t)})$, leur rejet ne fait pas bouger la chaîne.

La définition de la probabilité d'acceptation suivant l'équation (4.15) satisfait la réversibilité du noyau de transition de la chaîne de Markov générée [CG95]. Ce noyau transition est défini par :

$$\mathcal{K}_{MH}(\boldsymbol{\theta}_c, \boldsymbol{\theta}^{(t)}) \equiv q(\boldsymbol{\theta}_c|\boldsymbol{\theta}^{(t)}) \alpha(\boldsymbol{\theta}_c, \boldsymbol{\theta}^{(t)}) \quad \text{pour } \boldsymbol{\theta}_c \neq \boldsymbol{\theta}^{(t)}. \quad (4.18)$$

Cette définition étant incomplète dans la mesure où il faut considérer aussi la probabilité non nulle de rester figer dans le même état et qui est égale à $1 - \int q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) \alpha(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t)}) d\boldsymbol{\theta}$. L'expression complète du noyau de transition de la chaîne de MH s'écrit alors [Rob06]

$$\mathcal{K}_{MH}(d\boldsymbol{\theta}_c, \boldsymbol{\theta}^{(t)}) = q(\boldsymbol{\theta}_c|\boldsymbol{\theta}^{(t)}) \alpha(\boldsymbol{\theta}_c, \boldsymbol{\theta}^{(t)}) d\boldsymbol{\theta}_c + \left[1 - \int q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) \alpha(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t)}) d\boldsymbol{\theta} \right] \delta(\boldsymbol{\theta}^{(t)} - d\boldsymbol{\theta}_c). \quad (4.19)$$

La réversibilité de ce noyau de transition assure que la chaîne de Markov, générée par l'algorithme de MH, admet une loi invariante égale à la loi d'intérêt $\mathcal{P}(\boldsymbol{\theta}^{(t)}|\mathbf{y})$. Notons au passage que le calcul de $\alpha(\boldsymbol{\theta}_c, \boldsymbol{\theta}^{(t)})$ (4.15) ne dépend que du rapport de vraisemblance et ne requiert pas la détermination des constantes de normalisation (4.2) [Rob96]. De plus, dans le cas où la densité instrumentale est symétrique par rapport au candidat courant, l'équation (4.16) se simplifie et devient

$$\rho(\boldsymbol{\theta}_c, \boldsymbol{\theta}^{(t)}) = \frac{\mathcal{P}(\boldsymbol{\theta}_c|\mathbf{y})}{\mathcal{P}(\boldsymbol{\theta}^{(t)}|\mathbf{y})}. \quad (4.20)$$

Dans le paragraphe suivant, nous exposons les contraintes qui doivent être vérifiées les lois instrumentales et nous citons quelques exemples de celles-ci.

4.4.2 Conditions sur la loi instrumentale

Pour que la loi d'intérêt $\mathcal{P}(\boldsymbol{\theta}|\mathbf{y})$ soit effectivement la loi invariante de la chaîne de Markov, produite par (4.17), il paraît intuitif que des contraintes soient imposées sur la loi instrumentale $q(\boldsymbol{\theta}_c|\boldsymbol{\theta}^{(t)})$ et sur son support. En effet, si le support de la loi d'intérêt, qui doit être connexe sous peine d'invalider l'algorithme de MH, n'est pas totalement couvert par la densité instrumentale, on pourra s'attendre à ce que la loi invariante de la chaîne générée soit différente de celle d'intérêt ; certaines régions risquent de ne jamais être visitées. Cependant, la proposition suivante assure l'invariance quelle que soit la loi instrumentale utilisée, ce qui confère une universalité à l'algorithme de MH.

Proposition 4 *Pour toute loi instrumentale $q(\boldsymbol{\theta}_c|\boldsymbol{\theta}^{(t)})$, la loi correspondant $\mathcal{P}(\boldsymbol{\theta}|\mathbf{y})$ est une loi invariante de la chaîne $(\boldsymbol{\theta}^{(t)})$, produite selon le schéma (4.17) [Rob96].*

Toutefois, l'irréductibilité de la chaîne générée découle de la condition suivante sur le support de $q(\boldsymbol{\theta}_c|\boldsymbol{\theta}^{(t)})$.

Proposition 5 *Lorsque le support de $q(\boldsymbol{\theta}_c|\boldsymbol{\theta}^{(t)})$ est inclus dans le support de $\mathcal{P}(\boldsymbol{\theta}|\mathbf{y})$, la chaîne de Markov $(\boldsymbol{\theta}^{(t)})$ produite par la règle de Metropolis (4.17) est irréductible, ergodique et de distribution invariante $\mathcal{P}(\boldsymbol{\theta}|\mathbf{y})$. [Rob96].*

En théorie, cette condition n'est pas nécessaire pour assurer la validité de l'algorithme de MH (cf. proposition 4), mais elle garantit qu'il soit véritablement un algorithme MCMC (cf. définition 3). En pratique, quand le nombre d'itérations tend vers l'infini, la convergence est assurée si les densités instrumentales choisies ont le même support que la densité

d'intérêt [CG95].

Deux grandes classes de lois instrumentales sont souvent identifiées : les lois indépendantes et les lois symétriques ou à marche aléatoire. La première classe est une généralisation des méthodes d'acceptations/rejets [Rob96]. La loi instrumentale $q(\boldsymbol{\theta}_c|\boldsymbol{\theta}^{(t)})$ est indépendante de $\boldsymbol{\theta}^{(t)}$,

$$q(\boldsymbol{\theta}_c|\boldsymbol{\theta}^{(t)}) = h(\boldsymbol{\theta}_c). \quad (4.21)$$

La seconde classe est basée sur une idée naturelle qui consiste à perturber aléatoirement le candidat tout en restant au voisinage de celui-ci. Elle implique l'utilisation des densités de la forme

$$q(\boldsymbol{\theta}_c|\boldsymbol{\theta}^{(t)}) = h(|\boldsymbol{\theta}_c - \boldsymbol{\theta}^{(t)}|), \quad (4.22)$$

où $\boldsymbol{\theta} -_c - \boldsymbol{\theta}^{(t)}$ est un vecteur aléatoire appelé, souvent, incrément.

Ces deux classes conduisent à une simplification de l'expression de la probabilité d'acceptation (4.15) et (4.20). Elles donnent souvent des bonnes propriétés de convergence sous des conditions de régularité de la loi cible. Le choix de la densité h demeure, néanmoins, contraignant dans le cas des algorithmes MH indépendants [Rob06].

4.4.3 Convergence des algorithmes Metropolis-Hastings

Dans le paragraphe 4.3.2, nous avons évoqué quelques conditions pour assurer la convergence des méthodes MCMC en général, telles que la réversibilité, l'irréductibilité et le caractère apériodique de la chaîne de Markov générée. Certes, la convergence théorique de l'algorithme de MH est assurée sous ces conditions, cependant ces garanties théoriques ne sont pas toujours reproduites par des bonnes performances en pratique. Ces garanties ne permettent de déterminer :

- ni le taux (vitesse) de convergence,
- ni la durée de la période transitoire T_0 correspondant en pratique au nombre d'échantillons à ignorer, T_0 , avant de considérer que la convergence est atteinte,
- ni la longueur totale T de la chaîne à générer [GR92, CG95].

D'autre part, selon le choix de la loi instrumentale, la convergence de l'algorithme MH peut prendre en pratique des centaines de milliers d'itérations. Ce choix est décisif et une complication supplémentaire dans la mise en oeuvre de l'algorithme.

Un contrôle de convergence lors de l'implémentation est donc indispensable. De nombreuses méthodes de contrôle ont été proposées dans la littérature [Rob96, JJDS02, Rob06]. Parmi elles figurent les méthodes graphiques, de discrétisation, de comparaison inter-chaînes. D'autres règles empiriques pour calculer une première approximation du nombre d'itérations nécessaire sont fournies dans [RGG95, Rob96].

Egalement, un indicateur empirique de la bonne ou de la mauvaise adéquation de la loi instrumentale est le taux d'acceptation. En effet, s'il est trop faible, il témoigne d'une faible intersection entre les supports de la loi d'intérêt et de la densité instrumentale. Si au contraire le taux d'acceptation est trop élevé, ceci peut signifier que les extrémités du support de la loi d'intérêt ne sont presque jamais visitées.

Dans le premier cas, on dit que la chaîne générée "mélange" lentement, voire même ne

mélange pas suffisamment dans certains cas, tandis que dans le deuxième cas on dit qu'elle mélange rapidement. Le mélange se réfère à la dépendance entre le candidat courant et le suivant, cette dépendance peut être mesurée par exemple à l'aide de la corrélation. La rapidité du mélange indique la décroissance de la dépendance entre les échantillons au cours des itérations. Il peut arriver que la chaîne converge vers sa loi invariante tout en mélangeant lentement.

Dans cette thèse, nous allons considérer la réalisation d'une seule longue chaîne générée par un algorithme MH (conseillé dans [Gey92, Tie94]) et nous reviendrons sur le critère de convergence et l'utilisation du taux d'acceptation comme indicateur dans la dernière partie de ce chapitre. Dans la partie suivante, nous mettons en avant quelques exemples des algorithmes MH en vue de les réutiliser dans l'estimation paramétrique des signaux non-stationnaires.

4.5 Quelques algorithmes Metropolis-Hastings

Une classification complète de toutes les versions de l'algorithme MH est impossible vue l'universalité de la méthode. La version MH à marche aléatoire et celle dite MH une-variable-à-la-fois sont parmi les plus employées. Ces deux versions ainsi que leurs implémentations seront explicitées dans la suite.

4.5.1 Algorithme Metropolis Hastings à marche aléatoire

L'algorithme MH à marche aléatoire utilise une densité instrumentale à marche aléatoire introduite précédemment par l'équation (4.22). Cette densité peut être choisie parmi les densités standards symétriques (centrées en zéro) et facilement simulables, comme les lois normales ou uniformes, et ce d'une façon quasi indépendante de la loi d'intérêt. La densité instrumentale doit seulement être réglée en fonction du taux d'acceptation moyen de l'algorithme, par exemple, en bien calibrant la matrice de covariance du vecteur des paramètres $\boldsymbol{\theta}$ dans le cas d'une loi normale ou en choisissant un support convenable s'il s'agit d'une loi uniforme. Les étapes de l'algorithme MH à marche aléatoire sont les suivantes :

Procédure MH(\mathcal{P}, q)

1. Générer $\boldsymbol{\theta}_c \sim q(\boldsymbol{\theta}_c | \boldsymbol{\theta}^{(t)}) = h(|\boldsymbol{\theta}_c - \boldsymbol{\theta}^{(t)}|)$,
2. Evaluer $\alpha(\boldsymbol{\theta}_c, \boldsymbol{\theta}^{(t)}) = \min \left\{ 1, \frac{\mathcal{P}(\boldsymbol{\theta}_c | \mathbf{y})}{\mathcal{P}(\boldsymbol{\theta}^{(t)} | \mathbf{y})} \right\}$,
3. Générer $u \sim \mathcal{U}[0, 1]$,
4. Si $u \leq \alpha(\boldsymbol{\theta}_c, \boldsymbol{\theta}^{(t)})$ alors $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}_c$ sinon $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)}$.

Sur le plan théorique, il a été démontré dans [RT96, Rob96] que la convergence géométrique d'une chaîne de Markov générée par un algorithme MH à marche aléatoire est assurée sous certaines propriétés de régularité de la loi d'intérêt. Pourtant, cet algorithme

offre peu de garantie de convergence si le support de la loi d'intérêt n'est pas compact car la chaîne générée n'est pas uniformément ergodique [Rob96]. Dans le contexte de cette thèse, le domaine de définition de chacun des paramètres du modèle (2.3) étant un compact, les convergences théoriques uniforme et géométrique sont donc assurées pour une chaîne générée par un algorithme MH à marche aléatoire.

Nous signalons que la différence essentielle entre l'algorithme MH à marche aléatoire et la technique d'optimisation stochastique : le recuit simulé est que le premier forme une seule chaîne de Markov homogène et que le second produit une chaîne de Markov hétérogène en raison de la décroissance de la température au cours des itérations (cf. chapitre 3)[Rob96]. La théorie de convergence présentée dans ce chapitre et développée dans [Rob96] ne convient pas pour traiter les chaînes hétérogènes comme celle générée par le recuit simulé.

4.5.2 Algorithme Metropolis Hastings dit une-variable-à-la-fois

Même si l'algorithme MH est théoriquement assez satisfaisant du point de vue convergence et implémentation, les simulations de chaînes de Markov nécessitent en pratique la génération d'un très grand nombre de candidats. Cela est d'autant plus vrai et compliqué que le problème étudié est de grande dimension. C'est, par exemple, le cas de cette thèse puisque le vecteur des paramètres $\boldsymbol{\theta}$ est non seulement de dimension élevée $M_A + M_F + 3$ (2.11) mais aussi ses coefficients sont d'ordre de grandeurs variées. Dans ces cas, il est plutôt conseillé d'utiliser l'algorithme MH un-bloc-à-la-fois. On fait évoluer uniquement et en même temps un sous ensemble, formé par d paramètres du vecteur $\boldsymbol{\theta}$. Pour cela, on utilise d étapes de type MH et, donc, d densités instrumentales. Quand l'ensemble est réduit à un paramètre ($d = 1$), l'algorithme est dit MH une-variable-à-la-fois.

L'idée de l'algorithme MH un-bloc-à-la-fois s'appuie sur le fait, qu'il est plus simple de trouver plusieurs densités instrumentales conditionnelles et, donc, plusieurs noyaux de transition qui convergent vers leurs densités conditionnelles respectives, que de trouver un seul noyau de transition conjoint qui converge vers la distribution d'intérêt conjointe de tous les paramètres [CG95]. La densité invariante de la chaîne de Markov générée, néanmoins, reste non modifiée et correspond à la loi d'intérêt.

Etant donné $\boldsymbol{\theta}^{(t)} = [\theta_0^{(t)}, \dots, \theta_d^{(t)}]^\top$ un élément courant d'une chaîne Markov, générée par un algorithme MH une-variable-à-la-fois, on obtient l'élément $\boldsymbol{\theta}^{(t+1)}$ avec la procédure suivante :

Procédure MH($\mathcal{P}, q_0, q_1, \dots, q_d$)

0. $MH(\mathcal{P}(\theta_0 \mid \theta_{1,\dots,d}^{(t)}, \mathbf{y}), q_0)$
- ⋮
- i . $MH(\mathcal{P}(\theta_i \mid \theta_{0,\dots,i-1}^{(t+1)}, \theta_{i+1,\dots,d}^{(t)}, \mathbf{y}), q_i)$

$$\vdots$$

$$d. MH\left(\mathcal{P}(\theta_d \mid \theta_{0,\dots,d-1}^{(t+1)}, \mathbf{y}), q_d\right),$$

où $\theta_{0,\dots,i}^{(t)} = \{\theta_0^{(t)}, \dots, \theta_i^{(t)}\}$. La notation utilise l'indice (i) pour le i -ième paramètre et l'exposant (t) pour désigner l'itération t .

Chaque étape de cet algorithme simule une chaîne de Markov dont la densité invariante est la densité conditionnelle $\mathcal{P}(\theta_i \mid \theta_{0,\dots,i-1}^{(t+1)}, \theta_{i+1,\dots,d}^{(t)}, \mathbf{y})$. Elle utilise, pour ce fait, l'algorithme MH décrit par les équations (4.15), (4.16) et (4.17), et ce avec une densité instrumentale q_i pour chacun des paramètres θ_i . Bien entendu, l'emploi d'une densité instrumentale à marche aléatoire est possible et conduit à un algorithme dit MH à marche aléatoire et une-variable-à-la-fois.

Après avoir présenté les principales notions de la théorie des chaînes de Markov et des algorithmes de simulation par les méthodes MCMC, notamment celui de Metropolis Hastings, passons maintenant à leur application à l'estimation du modèle paramétrique adopté dans le cas des signaux non-stationnaires courts. Nous nous sommes particulièrement intéressés à l'implémentation de l'algorithme MH à marche aléatoire et une-variable-à-la-fois pour réaliser l'estimation des paramètres du modèle.

4.6 Estimation des signaux courts mono-composantes non-stationnaires

Dans cette partie, nous commençons par situer notre application par rapport à ce qui était déjà avancé dans la littérature. Ensuite, nous définissons l'inférence Bayésienne dans le cadre de cette application. Enfin, l'algorithme utilisé dans la simulation de la densité *a posteriori* de tous les paramètres du modèle conditionnellement aux observations sera explicité.

4.6.1 Positionnement

Dans la littérature, les approches Bayésiennes ont été largement développées et utilisées pour estimer des modèles paramétriques de signaux. Dans [TVF97, Vie99, TVG99], l'inférence Bayésienne est appliquée à un signal ayant une phase polynomiale et une amplitude constante, qui est éliminée par marginalisation. La densité de probabilité conjointe des coefficients de la phase a été simulée en utilisant les méthodes MCMC via un algorithme de MH à marche aléatoire, une-variable-à-la-fois et à saut réversible. Cette technique généralise les algorithmes MH en introduisant un mouvement entre les espaces des paramètres de dimensions différentes [Gre68] permettant, ainsi, d'estimer le degré de l'approximation polynomiale de la phase par exemple. L'étude des performances de l'algorithme a été faite pour un signal relativement long et ayant une modulation de fréquence de faible degré

(limité à 3); les résultats sont relativement satisfaisants.

Dans [AD99], le signal considéré est formé par des harmoniques pures dont on cherche à déterminer les fréquences. L'amplitude est constante et est éliminée par marginalisation. La technique utilisée est aussi un algorithme MH à saut réversible mais les densités instrumentales utilisées sont des lois conjuguées. Dans [DMD05], afin d'effectuer un suivi en temps réel de signaux, formés de plusieurs composantes ayant des fréquences instantanées quasi-linéaires et des amplitudes constantes, la technique proposée combine le spectrogramme et le filtrage particulière qui est une méthode MCMC séquentielles. Les performances dépendent de la résolution du spectrogramme et pour l'instant ne sont pas assez concluantes en présence de fortes modulations d'amplitude et de fréquence.

Notre contribution consiste à étendre l'application des méthodes MCMC à l'estimation d'un signal dont la fréquence et l'amplitude instantanées sont approchées par des polynômes de degré inférieur ou égal à 3. Les paramètres d'amplitude sont à déterminer. L'analyse des performances sera exclusivement faite dans le cadre des signaux courts (une trentaine d'échantillons est disponible). A notre connaissance, ce problème n'a pas encore été traité dans la littérature.

Dans ce contexte, il est important d'estimer chacun des paramètres de la fréquence instantanée, de l'amplitude instantanée et la phase initiale, ce qui sous-entend qu'il n'est pas envisageable d'éliminer un de ces paramètres par marginalisation. Ceci s'explique en partie par le fait que cette étude s'inscrit dans un cadre plus général, qui est l'estimation d'un signal long fortement non-stationnaire (cf. chapitre 6), nous prélevons des segments courts sur ce dernier, nous les estimons puis fusionnons pour reconstruire le signal long dans sa totalité.

Nous envisageons d'utiliser un algorithme MH à marche aléatoire et une-variable-à-la-fois afin de simuler la densité *a posteriori* des paramètres du modèle conditionnellement aux observations. Ce choix nous paraît en adéquation avec notre problématique car le vecteur des paramètres est multidimensionnel, à valeurs multivariées. Le degré de l'approximation polynomiale étant fixé au préalable, un algorithme à saut réversible n'est pas nécessaire.

4.6.2 Problématique et inférence Bayésienne

Soit $\mathbf{y} = [y[-\frac{N}{2}], \dots, y[\frac{N}{2}]]^T$ (2.28) un vecteur d'observations provenant d'un signal $s[n]$, court, non-stationnaire et noyé dans un bruit additif, blanc et Gaussien. Les équations (2.2) et (2.3) définissent le modèle, adapté à ce signal, dont les paramètres inconnus $\boldsymbol{\theta}$ sont la phase initiale et les coefficients de l'approximation polynomiale de l'amplitude et de la fréquence instantanées (2.11). Dans l'inférence Bayésienne, se rajoute un paramètre supplémentaire de nuisance σ_b^2 , la variance du bruit.

Supposons que \mathbf{y} est formé par des échantillons aléatoires i.i.d. selon une densité de probabilité, la vraisemblance $p(\mathbf{y}|\boldsymbol{\theta})$. Cette densité est connue conditionnellement à tous les paramètres. Cherchons à déterminer la loi *a posteriori* de ceux-ci conditionnellement aux observations.

Dans un premier temps, il faut définir la densité *a priori* en fonction des connaissances disponibles sur les paramètres du modèle $\boldsymbol{\theta}$ et sur la variance du bruit σ_b^2 . Pour ce faire, on

peut supposer l'indépendance entre les paramètres afin de simplifier l'inférence Bayésienne. La densité *a priori* conjointe de tous les paramètres s'écrit alors :

$$\mathcal{P}(\boldsymbol{\theta}, \sigma^2) = \mathcal{P}(\boldsymbol{\theta}) \mathcal{P}(\sigma^2) = \mathcal{P}(\sigma^2) \prod_{i=0}^{M_A+M_F+2} \mathcal{P}(\theta_i). \quad (4.23)$$

Nous avons décidé d'attribuer des densités *a priori* non-informatives aux paramètres du modèle : soit une densité *a priori* uniforme sur chaque domaine de définition de chacun des paramètres du vecteur $\boldsymbol{\theta}$ et une densité *a priori* de Jeffreys pour σ_b^2 . Le choix de la densité uniforme pour les paramètres de l'amplitude, de la fréquence et de la phase du signal traduit le fait qu'un décalage opéré sur l'un d'eux ne change pas la connaissance *a priori* dont on dispose sur ce dernier. Ainsi, ces paramètres sont considérés comme étant des paramètres de position. En revanche, σ_b^2 étant un paramètre multiplicatif, on préfère utiliser la loi de Jeffreys.

La densité *a priori* conjointe de $\boldsymbol{\theta}$ et de σ_b^2 s'écrit, donc, en respectant le domaine de définition de chacun des paramètres θ_i , comme suit :

$$\mathcal{P}(\boldsymbol{\theta}, \sigma_b^2) \propto \frac{1}{\sigma_b^2}; \quad (4.24)$$

Compte tenu des équations (3.3) et (4.24), l'expression de la distribution *a posteriori* à une constante de normalisation près est donnée par

$$\mathcal{P}(\boldsymbol{\theta}, \sigma_b^2 | \mathbf{y}) \propto \frac{\pi^{-(N+1)}}{\sigma_b^{2(N+2)}} \exp \left[-\frac{1}{\sigma_b^2} \sum_{n=-\frac{N}{2}}^{\frac{N}{2}} |y[n] - s[n]|^2 \right]. \quad (4.25)$$

La marginalisation (intégration) par rapport à σ_b^2 conduit à

$$p(\boldsymbol{\theta} | \mathbf{y}) \propto \left[\sum_{n=-\frac{N}{2}}^{\frac{N}{2}} |y[n] - s[n]|^2 \right]^{-(N+1)}. \quad (4.26)$$

La démonstration est présentée dans l'annexe A.8. Notons que d'après l'équation (4.26), le support de la densité *a posteriori* de $\boldsymbol{\theta}$ est lié au nombre d'échantillons $N + 1$. En effet, plus N est grand, plus les informations apportées par les observations sont précises et par conséquent le support de la densité est étroit.

L'obtention de la constante de normalisation, *i.e.* l'évidence Bayésienne de cette densité *a posteriori* (4.2), nécessite une intégration multidimensionnelle. Cette loi de probabilité étant non-standard, non linéaire et multidimensionnelle, la forme analytique exacte de l'évidence ne peut pas être obtenue par les méthodes classiques d'intégration, c'est également le cas du calcul des estimateurs MMSE (4.11) et MAP (4.13).

Nous utilisons alors les algorithmes MH pour simuler une série d'échantillons $\boldsymbol{\theta}^{(t)}$ générés

aléatoirement selon la loi de probabilité d'intérêt (4.26). Les T_0 premiers échantillons sur les T générés de la chaîne de Markov sont éliminés ; T_0 et T sont déterminés expérimentalement. Les estimateurs Bayésiens sont approchés, en utilisant le reste des échantillons de la chaîne, par les expressions suivantes :

$$\hat{\boldsymbol{\theta}}_{MMSE} = \frac{1}{T - T_0} \sum_{t=T_0+1}^T \boldsymbol{\theta}^{(t)} \quad (4.27)$$

$$\hat{\boldsymbol{\theta}}_{MAP} = \arg \max_{\boldsymbol{\theta}^{(t)} / t \in [T_0+1, T]} \mathcal{P}(\boldsymbol{\theta}^{(t)} | \mathbf{y}) \quad (4.28)$$

Dans le paragraphe suivant, nous détaillons l'algorithme MH utilisé pour réaliser l'estimation Bayésienne. Le choix des différents paramètres de réglage de cet algorithme seront également justifiés.

4.6.3 Algorithmes Metropolis Hastings une-variable-à-la-fois et à marche aléatoire

Dans notre cas, nous avons choisi d'implémenter un algorithme MH ne requérant qu'une connaissance limitée de la loi d'intérêt. L'algorithme est à marche aléatoire utilisant une mise à jour une-variable-à-la-fois. Nous rappelons que ce choix de mise à jour simplifie la recherche du noyau de transition, tout en garantissant que la loi d'intérêt soit la loi invariante de la chaîne générée (cf. paragraphe 4.5.2). D'autre part, il est parfaitement adapté à notre problème vu que le vecteur des paramètres est multidimensionnel et ses composantes peuvent prendre des valeurs multivariées.

Les éléments requis dans l'implémentation de cet algorithme sont détaillés dans la suite. Nous continuons d'utiliser l'indice (i) pour indiquer le i -ième paramètre et l'exposant (t) pour désigner l'itération t de l'algorithme.

Pour chacune des composantes du vecteur $\boldsymbol{\theta}$, une densité instrumentale symétrique de loi normale est assignée. Elle décrit une marche aléatoire de variance $\sigma_{\mathcal{M}\mathcal{A},i}^2$ autour du candidat $\theta_i^{(t)}$ accepté à l'itération t :

$$\theta_{c,i} | \theta_i^{(t)} \sim \mathcal{N}(\theta_i^{(t)}, \sigma_{\mathcal{M}\mathcal{A},i}^2). \quad (4.29)$$

Le choix de la loi normale pour la marche aléatoire est très important dans notre problème puisque la localisation précise de la loi d'intérêt n'est pas nécessaire. Nous traduisons notre ignorance de cette localisation en supposant que le nouveau candidat a plus de probabilité de se trouver dans le voisinage du candidat courant qu'ailleurs. La loi normale est un compromis entre des lois très sélectives telle que celle de Jeffreys et des lois non-informatives comme celle uniforme ayant un support large.

L'algorithme MH génère le nouveau candidat $\theta_{c,i}$ avec une probabilité d'acceptation donnée par l'équation (4.20). Cette dernière s'écrit, pour un algorithme une-variable-à-la-fois et à marche aléatoire, comme suit :

$$\alpha_i(\beta_i, \theta_i^{(t-1)}) = \min \left\{ 1, \rho_i^{(t)} \right\}, \quad (4.30)$$

avec

$$\rho_i^{(t)} = \frac{\mathcal{P}\left(\theta_{c,i} \mid \theta_{0,\dots,i-1}^{(t)}, \theta_{i+1,\dots,M_A+M_F+2}^{(t-1)}, \mathbf{y}\right)}{\mathcal{P}\left(\theta_i^{(t-1)} \mid \theta_{0,\dots,i-1}^{(t)}, \theta_{i+1,\dots,M_A+M_F+2}^{(t-1)}, \mathbf{y}\right)}, \quad (4.31)$$

où $\theta_{0,\dots,i-1}^{(t)} = \left(\theta_0^{(t)}, \theta_1^{(t)}, \dots, \theta_{i-1}^{(t)}\right)$.

Passons maintenant au choix de l'étalement de ces densités (l'écart-type dans le cas d'une loi normale : $\sigma_{\mathcal{MA},i}^2$). Nous rappelons qu'il est très important car il affecte la chaîne générée, essentiellement, en deux points : le probabilité d'acceptation des candidats proposés et la région couverte par l'échantillonnage de la loi *a posteriori*. Des lois instrumentales trop dispersées génèrent des candidats souvent refusés. La chaîne peut rester longtemps bloquée sur un candidat, dépensant ainsi un nombre d'itérations très important inutilement. Inversement, avec des densités instrumentales peu étendues, la chaîne peut bouger très lentement pour parcourir tout l'espace de recherche. Les régions à faible probabilité vont être sous échantillonnées tandis que trop de candidats seront acceptés dans les régions à forte probabilité. L'effet est donc le même *i.e.* un nombre d'itérations très élevé est nécessaire pour atteindre la convergence.

Pour un algorithme MH à marche aléatoire où la densité instrumentale est une loi normale, il est recommandé que l'étalement soit choisi de façon à obtenir un taux d'acceptation de l'ordre de 0.45 à 0.5 pour un problème unidimensionnel [Rob96, RGG95, GRS96, CG95, Chi96]. Quant au nombre d'itérations T_0 , nécessaire pour atteindre la loi d'intérêt, il est propre à chaque problème étudié et il sera déterminé expérimentalement dans notre application.

La réalisation pratique de l'algorithme MH est facile à mettre en oeuvre. La table 4.1 présente l'implémentation de l'algorithme que nous utilisons dans la simulation de la densité *a posteriori* (4.26). La chaîne de Markov produite converge asymptotiquement vers cette loi d'intérêt après une période de chauffage de longueur T_0 itérations. Dans le calcul des estimateurs MMSE (4.27) et MAP (4.28), nous ignorons les T_0 échantillons produits par cette chaîne et nous conservons les $(T - T_0)$ suivants, T étant la longueur totale de la chaîne générée.

L'initialisation de l'algorithme et plus précisément des paramètres peut se faire de la même manière, que pour l'algorithme du recuit simulé, décrite dans le chapitre 3, en utilisant :

- soit la FFT du signal bruité (3.13),
- soit le spectrogramme de ce dernier et en décomposant ensuite la fréquence et l'amplitude instantanées sur la base polynomiale considérée (3.17),
- soit en générant aléatoirement des valeurs de paramètres qui respectent les contraintes du problème (2.12), *i.e.* l'amplitude instantanée doit être positive et la fréquence instantanée positive et respecte le théorème de Shannon.

Toutefois, il faut s'assurer que les valeurs utilisées dans l'initialisation correspondent à une probabilité *a posteriori* strictement positive $\mathcal{P}(\boldsymbol{\theta}^{(0)}|\mathbf{y}) > 0$.

TAB. 4.1 – Algorithme MH à marche aléatoire et une-variable-à-la-fois

-
1. Initialisation des paramètres, $\boldsymbol{\theta}^{(0)} = [\theta_0^{(0)}, \theta_1^{(0)}, \dots, \theta_{M_A+M_F+2}^{(0)}]^\top$
 2. Itération de $t = 1$ à T
 - 2.1) Itération de $i = 0$ à $M_A + M_F + 2$
 - 2.1.1) Mise à jour
 - a- Générer un candidat $\theta_{c,i} \sim \mathcal{N}(\theta_i^{(t-1)}, \sigma_{\mathcal{M.A.},i}^2)$.
 - b- Evaluer $\alpha_i(\theta_{c,i}, \theta_i^{(t-1)})$ comme indiqué par (4.30).
 - c- Générer $u \sim \mathcal{U}[0, 1]$, si $u \leq \alpha_i(\theta_{c,i}, \theta_i^{(t-1)})$ faire une mise à jour $\theta_i^{(t)} = \theta_{c,i}$ sinon $\theta_i^{(t)} = \theta_i^{(t-1)}$.
 - 2.1.2) $i = i + 1$. retour à l'étape 2.1.
 - 2.2) $\boldsymbol{\theta}^{(t)} = [\theta_0^{(t)}, \dots, \theta_{M_A+M_F+2}^{(t)}]^\top$, $t = t + 1$, retour à l'étape 2.
 3. Evaluer l'estimateur : soit MMSE $\hat{\boldsymbol{\theta}}_{MMSE}$ (4.27), soit $\hat{\boldsymbol{\theta}}_{MAP}$ (4.28).
-

4.7 Conclusions

Dans ce chapitre, nous avons présenté une approche Bayésienne pour réaliser l'estimation du modèle non-linéaire d'un signal court non-stationnaire et mono-composante. Après avoir attribué des lois *a priori* non-informatives aux paramètres du modèle, nous avons déterminé à une constante de normalisation près l'expression de la loi *a posteriori* conjointe de ces paramètres conditionnellement aux observations (4.26).

Afin de simuler cette densité *a posteriori*, nous avons opté pour l'utilisation d'un algorithme MH à marche aléatoire et une-variable-à-la-fois. Ce choix est adapté à un vecteur de paramètres multidimensionnel, à valeurs multivariées. En effet, comme tous les algorithmes de MH, l'algorithme utilisé a l'avantage de n'avoir besoin que d'une connaissance limitée de la densité de probabilité à simuler et son implémentation est simple. De plus, la mise à jour des paramètres un par un facilite le choix de la densité génératrice de candidats.

Notre contribution a consisté essentiellement à étendre l'approche Bayésienne et l'application de l'algorithme MH à l'estimation de modèle de signaux courts dont l'amplitude et de la fréquence instantanées sont tous les deux polynomiales. A notre connaissance, dans la littérature ce problème n'a pas été considéré, même si des problèmes similaires ont déjà été abordés, tels que l'estimation des signaux relativement longs à phase polynomiale, à faible modulation de fréquence et à amplitude constante.

Dans le chapitre suivant, nous présentons une comparaison entre l'algorithme MH présenté dans ce chapitre (cf. Tab. 4.1), celui du recuit simulé introduit dans le chapitre 3, et celui de la méthode HAF (cf. chapitre 1). Les performances statistiques et les résultats obtenus par ces trois algorithmes sont analysés, exclusivement, dans le cadre d'un signal court non-stationnaire.

Chapitre 5

Signal court mono-composante non-stationnaire - Analyse et performance

Résumé

Ce chapitre est dédié à l'analyse statistique des performances des algorithmes du recuit simulé et celui du Metropolis Hastings, que nous avons proposés dans les chapitres 3 et 4. Ces deux algorithmes sont appliqués sur des signaux courts non-stationnaires et mono-composantes. Afin d'évaluer leur robustesse, des bornes minimales sur l'erreur d'estimation sont déterminées. Les deux algorithmes proposés sont ensuite comparés à la technique basée sur la méthode HAF. Une application sur un signal réel est également présentée.

Mots clefs

Recuit simulé, algorithme MH-MCMC, HAF, bornes de Cramer-Rao.

Sommaire

5.1	Introduction	103
5.2	Recuit simulé - Réglage	103
5.2.1	Différents jeux de paramètres et compromis	104
5.2.2	Résultats expérimentaux et discussions	106
5.2.2.1	Choix de la fonction objectif	106
5.2.2.2	Initialisation des paramètres du modèle	106
5.2.2.3	Choix de la température initiale	107
5.2.2.4	Rapidité de la décroissance de la température et du rayon du voisinage	108
5.2.2.5	Nombres de paramètres à perturber	108
5.2.3	Efficacité du réglage retenu Rego	109
5.3	Performances statistiques du recuit simulé en présence du bruit	110
5.3.1	Bornes de Cramer Rao	110
5.3.1.1	Notations	111
5.3.1.2	Expressions	111
5.3.1.3	Bornes de Cramer Rao adaptées au modèle du signal court	112
5.3.1.4	Bornes de Cramer-Rao - Découplage des paramètres de la phase	113
5.3.1.5	Bornes de Cramer-Rao - Fréquence et amplitude instantanées	114
5.3.2	Signaux à amplitude constante	115
5.3.2.1	Comparaison avec la HAF	115
5.3.2.2	Comparaison avec les Bornes de Cramer Rao découplées	116
5.3.2.3	Estimation de la fréquence instantanée	117
5.3.3	Signaux à modulations d'amplitude et de fréquence non-linéaires	117
5.3.3.1	Comparaison entre différentes bases polynomiales	118
5.3.3.2	Influence du taux de modulation de l'amplitude	120
5.3.3.3	Effet de la limitation de degré d'approximation	122
5.3.3.4	Conclusion	122
5.4	Algorithme MH-MCMC et comparaison avec le recuit simulé	124
5.4.1	Réglage de l'algorithme de MH-MCMC	124
5.4.2	Comparaison MH-MCMC, recuit simulé et HAF	125
5.5	Conclusion	127

5.1 Introduction

Dans les chapitres 3 et 4, nous avons proposé deux approches afin d'estimer les paramètres du modèle d'un signal court, non-stationnaire et mono-composante, ainsi que les algorithmes qui en découlent. L'objectif de ce chapitre est d'analyser les performances et la robustesse de ces derniers.

Pour chacun des algorithmes proposés, nous avons envisagé et testé plusieurs réglages des paramètres de contrôle. Cela nous permet d'établir les compromis nécessaires à l'obtention des estimations efficaces des paramètres du modèle. Etant donnée la difficulté de mesurer théoriquement et exactement l'influence de ces paramètres de contrôle sur le comportement des algorithmes, des simulations de Monte-Carlo nous permettent d'évaluer statistiquement l'erreur de l'estimation et ce pour les différents réglages.

Nous présentons ensuite une comparaison entre les algorithmes proposés et la méthode itérative basée sur la HAF. Les résultats obtenus sont comparés à des bornes minimales d'estimation que nous avons calculées et ce afin d'évaluer statistiquement la robustesse de chacune de ces méthodes. Les algorithmes proposés sont ensuite appliqués à un signal réel.

5.2 Recuit simulé - Réglage

Cette partie est dédiée à l'étude du comportement du recuit simulé lors de la minimisation de la fonction des moindres carrés (3.6) et ce dans le contexte de signaux courts ayant une amplitude et une fréquence polynomiales.

Le nombre de paramètres de contrôle de l'algorithme du recuit simulé proposé étant assez élevé (cf. chapitre 3), nous souhaitons obtenir le réglage pouvant fournir la meilleure estimation possible. De plus, nous visons l'automatisation du réglage de façon à ce qu'il soit adaptable et applicable dans la majorité des modulations polynomiales de degré limité à 3 et de très courte durée. N'oublions pas que, l'un des buts est de pouvoir ré-utiliser cet algorithme dans l'estimation des signaux courts extraits d'une façon automatique d'un signal long.

Nous élaborons et testons des différents réglages des paramètres de contrôle de l'algorithme du recuit simulé. Rappelons que ces derniers sont principalement :

- le choix de la fonction objectif : la fonction des moindres carrés (3.6) ou le coefficient de corrélation (3.9),
- $\boldsymbol{\theta}^{(0)}$ l'initialisation des paramètres du modèle (3.13) et (3.17),
- $\boldsymbol{\sigma}$ le rayon du voisinage (3.20) permettant l'exploration de l'espace de définition des paramètres du modèle,
- $\mathcal{T}^{(0)}$ la température initiale (3.21),
- ϵ_σ et $\epsilon_{\mathcal{T}}$ les paramètres régulateurs de la décroissance du rayon du voisinage (3.22) et celle de la température (3.23) au cours de l'itération t :

$$\boldsymbol{\sigma}^{2(t+1)} = (1 - \epsilon_\sigma)\boldsymbol{\sigma}^{2(t)} \quad \text{et} \quad \mathcal{T}^{(t+1)} = (1 - \epsilon_{\mathcal{T}})\mathcal{T}^{(t)}. \quad (5.1)$$

- le paramètre ρ (3.24) permet de choisir aléatoirement les instants où \mathcal{T} et σ^2 sont réduits.
- le nombre de paramètres du modèle à perturber, *i.e.* à faire évoluer lors d'une même itération.

Dans un premier temps, nous spécifions les réglages des différents paramètres de contrôle qui influencent le comportement de cet algorithme. Ensuite, nous montrons l'influence du choix de la base polynomiale sur la robustesse de ce dernier. Enfin, nous comparons les erreurs quadratiques moyennes d'estimation obtenues aux bornes de Cramer-Rao que nous avons calculées.

5.2.1 Différents jeux de paramètres et compromis

Dans cette partie, est mise en avant l'efficacité de l'algorithme du recuit simulé (sa capacité à trouver une valeur approximative de l'optimum global), plutôt que sa rapidité à la trouver (le nombre d'itérations nécessaire). Le jeu de réglages que nous avons élaboré est présenté dans le tableau 5.1 et permet de mettre en évidence la difficulté d'obtenir un réglage automatique. Seul le paramètre ρ est fixé égal à 0.01 et ne varie pas dans tous ces réglages.

TAB. 5.1 – Jeu de réglages empiriques des différents paramètres de contrôle de l'algorithme du recuit simulé.

	Reg ₀	Reg _{F.O.}	Reg _{$\theta^{(0)}$}	Reg _{$\sigma \nearrow$}	Reg _{$\sigma \searrow$}	Reg _{$\mathcal{T} \nearrow$}	Reg _{$\mathcal{T} \searrow$}	Reg _{$\epsilon_{\mathcal{T}} \searrow$}	Reg _{$\epsilon_{\sigma} \searrow$}	Reg _{$\epsilon \searrow$}	Reg _{N.P. \searrow}
F.O.	$\ell_{MC}(\boldsymbol{\theta})$ (3.3)	$\gamma_{y,s}(\boldsymbol{\theta})$ (3.9)									
$\theta^{(0)}$	I.FFT (3.13)		I.SPEC (3.18)								
$\sigma^{(0)}$	V.B. (2.26)			5 V.B.	$\frac{1}{5}$ V.B.						
$\mathcal{T}^{(0)}$	1					100	0.01				
ϵ_{σ}	0.99							0.75		0.75	
$\epsilon_{\mathcal{T}}$	0.99								0.75	0.75	
N.P.	7										3

F.O. = Fonction objectif, V.B. = Vecteur formé par les Bornes supérieures du domaine de définition (2.24) des composantes du vecteur $\boldsymbol{\theta}$, I.FFT = Initialisation par FFT (3.13), I.SPEC. = Initialisation par spectrogramme(3.17), N.P. = nombre de paramètres du modèle choisi au hasard pour être perturbé en même temps.

Nous avons établi d'une façon empirique un réglage automatique, noté Reg₀, en se basant sur des signaux dont les paramètres du modèle (2.11) sont générés aléatoirement. Dans ce réglage, la fonction objectif est la fonction des moindres carrés. L'initialisation des paramètres du modèle se fait par la transformée de Fourier tandis que le pas de voisinage

est initialisé par les bornes supérieures du domaine de définition des paramètres du modèle (2.24). Tous les paramètres sont perturbés lors d'une itération.

Nous comparons ce réglage à d'autres réglages où nous faisons varier un seul paramètre de contrôle à la fois celui indiqué explicitement dans le tableau 5.1.

Le jeu de réglages est testé d'abord sur deux signaux non bruités ayant chacun 33 échantillons et une modulation quadratique ($M_A = M_F = 2$) en amplitude et en fréquence (cf. figure 5.13). Leurs paramètres sont représentés dans la base orthonormale (2.21) par :

$$\begin{aligned} \text{Signal 1 : } \boldsymbol{\theta}_{s_1} &= [a_0 = 16, \quad a_1 = -3.0, \quad a_2 = 2.0, \quad \varphi_0 = 0.4, \quad f_0 = 2.0, \quad f_1 = -0.2, \quad f_2 = -0.5]^\top, \\ \text{Signal 2 : } \boldsymbol{\theta}_{s_2} &= [a_0 = 16, \quad a_1 = 10.4, \quad a_2 = 4.5, \quad \varphi_0 = 0.4, \quad f_0 = 1.5, \quad f_1 = -0.2, \quad f_2 = -0.5]^\top. \end{aligned} \quad (5.2)$$

Ces signaux ont le même indice de modulation fréquentiel¹ (67%) mais un taux de modulation d'amplitude² différent, soit 50% pour Signal 1 et 87% pour Signal 2.

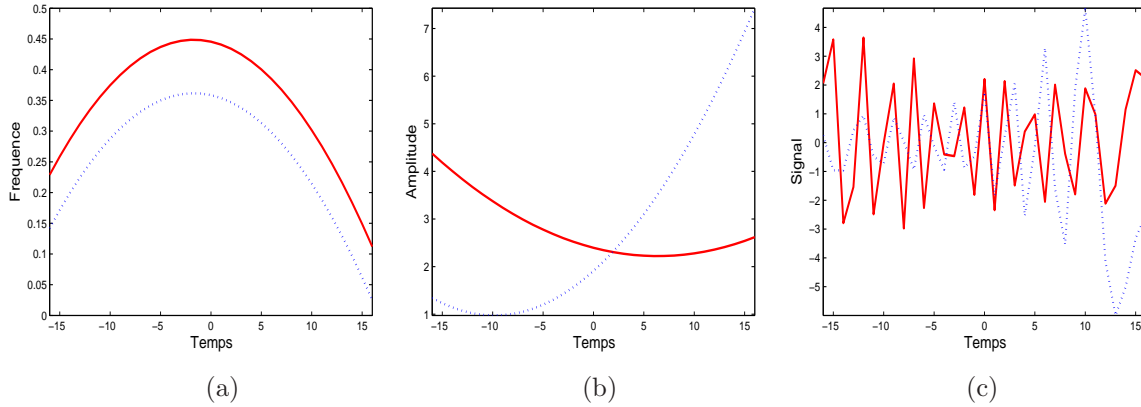


FIG. 5.1 — (—) Signal 1 et (...) Signal 2 : deux signaux simulés ayant une modulation d'amplitude et de fréquence quadratique, le taux de modulation d'amplitude est différent pour chaque signal. (a) Fréquences instantanées, (b) Amplitudes instantanées et (c) Signaux temporels.

Pour chaque signal, 50 réalisations d'estimation sont effectuées en initialisant le générateur des nombres pseudo-aléatoires différemment à chaque essai, ceci permet d'éviter des problèmes dus à la génération des nombres aléatoires. Le nombre d'itérations noté "Itération" et l'erreur moyenne notée "Erreur" sont obtenus en moyennant les résultats sur les 50 réalisations et ils sont rassemblés dans les tableaux 5.2, 5.3 et 5.4; les écarts-types sont indiqués entre parenthèses.

En l'absence de bruit, un test est considéré réussi si l'inégalité suivante est vérifiée :

$$|\ell(\boldsymbol{\theta}^*) - \ell(\hat{\boldsymbol{\theta}})| < \epsilon_1, \quad (5.3)$$

¹L'indice de modulation est défini comme étant l'excursion en fréquence normalisée par la largeur de la bande de fréquence tolérée. Dans le modèle que nous avons adopté, la largeur de cette bande vaut $\frac{1}{2}$ en travaillant en fréquence réduite.

²Le taux de modulation est défini comme étant égal à la variation max-min de l'amplitude normalisée par sa valeur moyenne.

où $\boldsymbol{\theta}^*$ désigne l'optimum vrai de la fonction objectif $\ell(\boldsymbol{\theta})$ et $\hat{\boldsymbol{\theta}}$ est la valeur obtenue par l'estimateur. ϵ_1 est un paramètre de précision que nous avons choisi égal à 10^{-4} , un ordre de grandeur permettant expérimentalement d'obtenir une estimation satisfaisante des paramètres du modèle en l'absence de bruit. Etant donné que les signaux ne sont pas bruités, on a les fonctions objectif $\ell_{MC}(\boldsymbol{\theta}^*) = \gamma_{y,s}(\boldsymbol{\theta}^*) = 0$ et donc cela revient à tester que la précision sur la fonction objectif en son optimum reste supérieur à ϵ_1 . Nous comptabilisons le pourcentage d'échec, noté (% Echec), de l'algorithme à trouver le bon optimum.

5.2.2 Résultats expérimentaux et discussions

5.2.2.1 Choix de la fonction objectif

Nous étudions le comportement de l'algorithme pour deux fonctions objectifs différentes : la fonction des moindres carrés $\ell_{MC}(\boldsymbol{\theta})$ (3.6) et le coefficient de corrélation $\gamma_{y,s}(\boldsymbol{\theta})$ (3.9). Le réglage des autres éléments du recuit simulé est le même pour les deux fonctions objectifs.

Puisque la maximisation de $\gamma_{y,s}(\boldsymbol{\theta})$ ne concerne pas les paramètres d'amplitude (3.9), ces derniers sont fixés, uniquement dans ce paragraphe, à leurs vraies valeurs. Les résultats mentionnés dans le tableau le tableau 5.2, montrent des performances similaires avec un nombre d'itérations relativement faible en utilisant le coefficient du corrélation.

TAB. 5.2 – Influence du choix de la fonction objectif sur l'algorithme du recuit simulé - Application sur les signaux définis par (5.2) avec des paramètres d'amplitude connus.

	Rego $\ell_{MC}(\boldsymbol{\theta})$			Reg $F.O.$ $\gamma_{y,s}(\boldsymbol{\theta})$			
	Itération	Erreur	% Echec	Itération	Erreur	Corrélation	% Echec
Signal 1	124840 (3888)	7.91 10^{-5} (1.72 10^{-6})	0	92439 (4669)	4.24 10^{-6} (1.82 10^{-7})	1 (2.2 10^{-9})	0
Signal 2	123940 (4994)	7.71 10^{-5} (1.70 10^{-5})	0	92477 (5129)	3.42 10^{-6} (1.07 10^{-6})	1 (2.3 10^{-9})	0

Cependant, en pratique on ne connaît pas les vraies valeurs des paramètres d'amplitude. Pour obtenir une bonne précision des paramètres du modèle en utilisant le coefficient de corrélation comme fonction objectif, il faut alterner plusieurs fois les deux étapes suivantes : la réévaluation des paramètres (3.8) et l'estimation des paramètres de la phase et de la fréquence avec l'algorithme du recuit simulé. En revanche, en utilisant la fonction des moindres carrés comme fonction objectif, tous les paramètres sont estimés en même temps, c'est pour cette raison qu'on va continuer de l'employer dans la suite.

5.2.2.2 Initialisation des paramètres du modèle

Le tableau 5.3 compare les résultats obtenus en initialisant les paramètres du modèle, soit par la transformée de Fourier (3.13), soit par le spectrogramme (3.17). Nous avons obtenu un nombre d'itérations et une erreur finale du même ordre de grandeur dans les deux cas. De plus, le nombre d'échec vaut zéro dans les deux cas.

TAB. 5.3 – Influence de l’initialisation des paramètres du signal sur l’algorithme du recuit simulé - Application aux signaux définis par (5.2).

	Reg ₀ I.FFT			Reg _{θ(0)} I.Spect		
	Itération	Erreur	% Echec	Itération	Erreur	% Echec
Signal 1	53775 (3950)	9.92 10 ⁻⁵ (1.82 10 ⁻⁶)	0	54176 (3149)	9.92 10 ⁻⁵ (1.25 10 ⁻⁶)	0
Signal 2	45001 (2612)	8.90 10 ⁻⁵ (1.04 10 ⁻⁶)	0	44694 (2370)	9.02 10 ⁻⁵ (8.37 10 ⁻⁶)	0

De façon générale, nous avons constaté que l’utilisation de l’une ou de l’autre initialisation conduit à une bonne approximation des paramètres à estimer et que le réglage Reg₀ convient à ces deux façons d’initialisation. Pour des raisons de simplicité, nous allons poursuivre cette étude en utilisant une initialisation à partir de la transformé de Fourier. Les remarques restent valables pour l’initialisation à partir du spectrogramme.

Choix du rayon du voisinage Le tableau 5.4 illustre l’influence du choix initial du pas du voisinage. Nous constatons essentiellement qu’un rayon de voisinage très grand ou très petit peut occasionner des échecs de l’algorithme à trouver le bon optimum. Le choix du rayon du voisinage selon le réglage Reg₀ (cf. Tab. 5.1) permet de trouver un compromis entre les deux autres réglages du rayon du voisinage et cela pour les deux signaux analysés ici. Dans le paragraphe 5.2.3, nous testons ce réglage sur des signaux dont les paramètres du modèle, le degré de la modulation et la longueur du signal seront générés aléatoirement.

TAB. 5.4 – Influence de l’initialisation du rayon du voisinage sur l’algorithme du recuit simulé - Application aux signaux définis par (5.2).

	Reg ₀			Reg _{σ↗}			Reg _{σ↘}		
	Itération	Erreur	% Echec	Itération	Erreur	% Echec	Itération	Erreur	% Echec
Signal 1	53775 (3950)	9.92 10 ⁻⁵ (1.82 10 ⁻⁶)	0	74347 (14346)	2.35 10 ⁻⁴ (3.02 10 ⁻⁴)	20	44893 (2623)	9.84 10 ⁻⁵ (2.37 10 ⁻⁶)	0
Signal 2	45001 (2612)	8.90 10 ⁻⁵ (1.04 10 ⁻⁶)	0	50329 (2488)	9.00 10 ⁻⁵ (8.55 10 ⁻⁶)	0	42555 (11153)	0.53 (2.15)	3

5.2.2.3 Choix de la température initiale

La température initiale conditionne également le comportement de l’algorithme comme l’illustre le tableau 5.5. Une très grande ou une très faible température initiale provoque un échec de l’algorithme dans sa recherche de l’optimum global. Elle ralentit son exécution en occasionnant une augmentation du nombre d’itérations. Le choix de la température initiale retenu dans le réglage Reg₀ est un compromis qui dépend de la précision finale voulue, de la rapidité de la décroissance de la température (3.23) et de celle du rayon du voisinage (3.22). Le choix de l’un de ces derniers conditionne le choix de la valeur de température initiale.

TAB. 5.5 – Influence de la température initiale sur le comportement de l’algorithme du recuit simulé - Application aux signaux définis par (5.2).

	Reg ₀				Reg _{ε_T↗}			Reg _{ε_T↘}		
	Itération	Erreur	% Echec		Itération	Erreur	% Echec	Itération	Erreur	% Echec
Signal 1	53775 (3950)	9.92 10 ⁻⁵ (1.82 10 ⁻⁶)	0		87432 (3770)	12.81 (10.16)	50	66972 (14501)	8.54 10 ⁻⁴ (3.66 10 ⁻³)	17
Signal 2	45001 (2612)	8.90 10 ⁻⁵ (1.04 10 ⁻⁶)	0		89287 (2904)	10.60 (6.52)	50	46387 (6874)	0.18 (1.28)	1

5.2.2.4 Rapidité de la décroissance de la température et du rayon du voisinage

Concernant le choix de ϵ_T et ϵ_σ les paramètres régulateurs de la décroissance respectivement de la température et du rayon du voisinage (5.1), quatre réglages différents sont présentés : Reg₀, Reg_{ε_T↘}, Reg_{ε_σ↘} et Reg_{ε↘}.

Dans le réglage noté Reg_{ε_T↘} (respectivement Reg_{ε_σ↘}) la valeur de ϵ_T (respectivement ϵ_σ) est faible comparée à celle considérée dans le réglage Reg₀. Dans le dernier réglage Reg_{ε↘}, les paramètres ϵ_T et ϵ_σ sont tous les deux diminués par rapport à ceux considérés dans Reg₀.

Le tableau 5.6 illustre le fait que la décroissance trop rapide de la température et/ou du rayon du voisinage provoque la stagnation de l’algorithme sur un optimum local et ne permet pas à ce dernier d’explorer d’autres solutions. Toutefois, même si la décroissance de la température et du rayon du voisinage est modérée, il faut s’assurer que la température initiale ne soit ni trop faible ni trop élevée, sous peine de se piéger rapidement dans un optimum local.

TAB. 5.6 – Influence de la raison de la décroissance de la température et/ou de celle du pas du voisinage (ϵ_T et ϵ_σ) sur l’algorithme du recuit simulé - Application aux signaux définis par (5.2).

	Signal 1			Signal 2		
	Itération	Erreur	% Echec	Itération	Erreur	% Echec
Reg ₀	53775 (3950)	9.92 10 ⁻⁵ (1.82 10 ⁻⁶)	0	45001 (2612)	8.90 10 ⁻⁵ (1.04 10 ⁻⁶)	0
Reg _{ε_T↘}	70014 (14121)	4.44 10 ⁻⁴ (8.66 10 ⁻⁴)	20	45102 (2679)	8.94 10 ⁻⁵ (9.86 10 ⁻⁶)	0
Reg _{ε_σ↘}	9341 (766)	7.20 (3.56)	50	9909 (897)	9.18 (4.66)	50
Reg _{ε↘}	9362 (1051)	5.33 (4.29)	50	9492 (919)	5.59 (4.16)	50

5.2.2.5 Nombres de paramètres à perturber

Le nombre de paramètres du modèle, à perturber en même temps, influence également le comportement du recuit simulé comme l’illustre le tableau 5.7. Dans ce tableau, nous comparons le réglage Reg₀, où les 7 paramètres du modèle sont perturbés en même temps à chaque itération, et le réglage Reg_{N.P.} où uniquement 3 paramètres choisis aléatoirement parmi les 7 sont perturbés.

Les deux réglages proposés permettent de retrouver avec une bonne précision l’optimum global. Le nombre d’itérations est plus petit dans le cas du réglage Reg_{N.P.}. Cependant, même si ce dernier réglage a donné des résultats satisfaisants dans cet exemple, rien n’est

garanti quand le nombre total des paramètres du modèle devient plus grand. En effet, *a priori* le meilleur rapport entre “nombre de paramètres à faire évoluer” et “nombre total des paramètres” est difficile à établir pour tous les signaux ayant notre modèle. Pour cette raison, nous allons continuer à perturber tous les paramètres en même temps.

TAB. 5.7 – Influence du nombre de paramètres à faire évoluer en même temps sur l’algorithme du recuit simulé - Application aux signaux définis par (5.2).

	Signal 1			Signal 2		
	Itération	Erreur	% Echec	Itération	Erreur	% Echec
Reg ₀	53775 (3950)	9.92 10 ⁻⁵ (1.82 10 ⁻⁶)	0	45001 (2612)	8.90 10 ⁻⁵ (1.04 10 ⁻⁵)	0
Reg _{N.P.}	38910 (2284)	8.94 10 ⁻⁵ (9.28 10 ⁻⁶)	0	41573 (2545)	9.17 10 ⁻⁵ (8.43 10 ⁻⁶)	0

5.2.3 Efficacité du réglage retenu Reg₀

Avec les différents résultats proposés précédemment, nous avons illustré l’importance des différents réglages dans le déroulement de l’algorithme du recuit simulé et par la même occasion la difficulté de trouver un réglage automatique adapté à tout signal court et ayant une modulation d’amplitude et de fréquence polynomiale de degré inférieur à 3. Le réglage Reg₀ (cf. Tab. 5.1) a été déterminé dans ce but.

Le tableau 5.8 illustre les résultats obtenus à partir de 100 réalisations différentes de signaux courts modélisés selon (2.8) et non-bruités. Nous générons aléatoirement les paramètres du modèle en respectant les contraintes (2.12). Les degrés de l’approximation polynomiale varient entre 0 et 3 et la longueur du signal est comprise entre 15 et 65. De façon générale ces signaux ont des taux et des indices de modulations pouvant varier entre 0% et 100%. On constate l’absence d’échec dans la détermination des paramètres de leur modèle.

TAB. 5.8 – Efficacité du réglage retenu Reg₀ - Application aux signaux modélisés par (2.8) dont les paramètres sont générés aléatoirement.

Paramètres du modèle générés aléatoirement		
Itération	Erreur	% Echec
127350 (15887)	8.80 10 ⁻⁵ (5.42 10 ⁻⁶)	0

Dans le tableau 5.9, nous étudions le nombre d’itérations et l’erreur finale obtenus en utilisant l’algorithme proposé avec le réglage retenu et ce pour estimer Signal 1 et Signal 2, qui sont noyés dans du bruit additif blanc et Gaussien (5.2). 500 réalisations du bruit sont effectuées pour chaque RSB et pour chaque signal. On considère que l’échec de l’algorithme s’est produit dans le cas où l’équation (3.28) n’est pas vérifiée.

Le tableau 5.9 résume les résultats obtenus pour des RSB variant de 0 dB à 25 dB. Le nombre d’itérations moyen nécessaire pour atteindre l’optimum global de la fonction de vraisemblance varie peu en fonction du RSB. Quant à la capacité de retrouver le minimum global, elle reste inchangée pour des RSB situés entre 5 dB à 25 dB.

TAB. 5.9 – Efficacité du réglage Reg_0 en présence du bruit - Application aux signaux définis par (5.2).

RSB (dB)	σ_b^2	Signal 1			Signal 2		
		Itération	Erreur	% Echec	Itération	Erreur	% Echec
25	0.03	55274 (2560)	0.90 (0.003)	0	50202 (6529)	1.07 (0.006)	0
20	0.11	59927 (9927)	1.55 (0.04)	0	48921 (9406)	2.03 (0.04)	0
15	0.37	57290 (7276)	2.34 (0.14)	0	46476 (14829)	3.34 (0.23)	0
10	1.15	53585 (9854)	4.85 (0.50)	0	45963 (4532)	6.39 (1.02)	0
5	2.37	62035 (16695)	8.20 (1.33)	0	34514 (6759)	9.75 (1.32)	0
0	5.37	66663 (22410)	16.8 (1.71)	1	42414 (9309)	19.75 (1.52)	3

La partie suivante est dédiée à l'analyse statistique de la robustesse de l'algorithme du recuit simulé avec le réglage Reg_0 en présence du bruit.

5.3 Performances statistiques du recuit simulé en présence du bruit

Afin d'évaluer la robustesse de l'algorithme du recuit simulé proposé, nous calculons les bornes de Cramer Rao qui sont des bornes minimales d'estimation. Ces dernières sont ensuite comparées aux erreurs quadratiques moyennes d'estimation. L'étude sera établie en fonction du rapport signal à bruit et pour les trois bases polynomiales définies par les équations (2.14), (2.18) et (2.21). Nous mettrons ainsi en évidence la contribution de la base orthonormale discrète, que nous avons calculée, dans l'amélioration des résultats.

5.3.1 Bornes de Cramer Rao

Considérons le modèle d'un signal mono-composante (2.2) noyé dans un bruit additif Gaussien (2.1). La fréquence et l'amplitude instantanées de ce signal sont approchées par une combinaison linéaire de polynômes (2.8). Nous allons déterminer les bornes de Cramer-Rao correspondant à ce modèle. Ces bornes représentent les bornes asymptotiques minimales d'estimation pour un estimateur de maximum de vraisemblance non-biaisé. Dans certains cas, elles peuvent se révéler très optimistes car elle ne tient pas compte des contraintes des problèmes étudiés.

Dans [FF95], les auteurs ont exprimé d'une part les bornes de Cramer-Rao de l'amplitude et la phase instantanées et d'autre part les bornes de Cramer-Rao des paramètres de leurs modèles polynomiaux. Le premier type de bornes correspond à des fonctions variant dans le temps et permet d'avoir une idée sur l'évolution temporelle de l'erreur minimale d'estimation. En effet, lorsque l'amplitude est une fonction du temps, le rapport signal à bruit instantané varie au cours du temps et peut induire des différences dans les performances de l'algorithme. En se basant sur les travaux présentés dans [FF95], nous allons déterminer d'abord les bornes de Cramer-Rao pour les paramètres du modèle adopté (2.8) et ensuite celles pour l'amplitude et la fréquence instantanées.

5.3.1.1 Notations

Considérons dans la suite les notations suivantes :

- le vecteur temps noté $\mathbf{n} = [-\frac{N}{2}, \dots, \frac{N}{2}]^\top$,
- deux vecteurs à valeurs réelles

$$\mathbf{a} = [a_0, \dots, a_{M_A}]^\top \quad \text{et} \quad \boldsymbol{\varphi} = [\varphi_0, \dots, \varphi_{M_F+1}]^\top \quad (5.4)$$

formés respectivement par les paramètres de la décomposition de l'amplitude instantanée sur la base polynomiale $(g_i[n])_{i=0, \dots, M_A}$ et les paramètres de la décomposition de la phase instantanée sur une autre base $(\mu_i[n])_{i=0, \dots, M_F+1}$, non forcément différente de la première. M_A et $M_F + 1$ sont respectivement les degrés de l'approximation polynomiale de l'amplitude et de la phase instantanées. On notera ici $\boldsymbol{\theta}$ le vecteur de paramètres $[a_0, \dots, a_{M_A}, \varphi_0, \dots, \varphi_{M_F+1}]^\top$.

- Pour une fonction scalaire donnée $h[n]$ quelconque, $h[\mathbf{n}]$ indique le vecteur colonne constitué des valeurs de $h[n]$ prises en $n = -\frac{N}{2}, \dots, \frac{N}{2}$. Par exemple, nous notons les vecteurs temporels de la phase instantanée $\Phi[n]$ et ceux des bases polynomiales considérées $g_i[n]$ et $\mu_i[n]$ par $\Phi[\mathbf{n}]$, $g_i[\mathbf{n}]$ et $\mu_i[\mathbf{n}]$.

A l'aide de ces notations, l'équation des observations (2.1) est ré-écrite sous une forme matricielle

$$\mathbf{y} = \boldsymbol{\Psi} \mathbf{a} + \mathbf{b}, \quad (5.5)$$

où $\boldsymbol{\Psi}$ est la matrice de taille $(N+1) \times (M_A+1)$ définie par

$$\boldsymbol{\Psi} = [g_0[\mathbf{n}] \bullet e^{j\Phi[\mathbf{n}]}, \dots, g_{M_A}[\mathbf{n}] \bullet e^{j\Phi[\mathbf{n}]}]. \quad (5.6)$$

$\mathbf{b} = [b[-\frac{N}{2}], \dots, b[\frac{N}{2}]]^\dagger$ désigne le vecteur temporel représentant le bruit. Le symbole (\bullet) dénote la multiplication terme à terme entre des vecteurs de même dimension. Notons qu'avec cette écriture matricielle (5.5), \mathbf{y} est une fonction linéaire par rapport aux paramètres de l'amplitude et non-linéaire par rapport à ceux de la phase. La log-vraisemblance (3.3) s'exprime alors par :

$$\Lambda = -(N+1) \ln \left(\frac{\pi}{\sigma_b^2} \right) - \frac{1}{\sigma_b^2} |\mathbf{y} - \boldsymbol{\Psi} \mathbf{a}|^2. \quad (5.7)$$

5.3.1.2 Expressions

Le calcul des bornes de Cramer-Rao nécessite la détermination de la matrice de Fisher précédemment définie par (4.9) et notée \mathbf{I}_θ .

- En dérivant partiellement la log-vraisemblance Λ , on obtient l'expression suivante [FF95] :

$$\frac{\partial \Lambda}{\partial \varphi_i} = \frac{1}{\sigma_b^2} \left\{ \mathbf{a}^\dagger \left(\frac{\partial \boldsymbol{\Psi}}{\partial \varphi_i} \right)^\dagger (\mathbf{y} - \boldsymbol{\Psi} \mathbf{a}) + (\mathbf{y} - \boldsymbol{\Psi} \mathbf{a})^\dagger \frac{\partial \boldsymbol{\Psi}}{\partial \varphi_i} \mathbf{a} \right\}, \quad (5.8)$$

$$= \frac{2}{\sigma_b^2} \Re \left\{ \mathbf{a}^\dagger \left(\frac{\partial \boldsymbol{\Psi}}{\partial \varphi_i} \right)^\dagger (\mathbf{y} - \boldsymbol{\Psi} \mathbf{a}) \right\}. \quad (5.9)$$

Les coefficients de la matrice de Fisher correspondant aux paramètres de la phase sont les suivants :

$$-\mathbb{E} \left[\frac{\partial^2 \Lambda}{\partial \varphi_i \partial \varphi_l} \right] = \frac{2}{\sigma_b^2} \Re \left\{ \mathbf{a}^\dagger \left(\frac{\partial \Psi}{\partial \varphi_i} \right)^\dagger \frac{\partial \Psi}{\partial \varphi_l} \mathbf{a} \right\}. \quad (5.10)$$

Or, étant donné que $\frac{\partial \Psi}{\partial \varphi_i} \mathbf{a} = j \mu_i[\mathbf{n}] \bullet s[\mathbf{n}]$, on peut écrire l'équation (5.10) sous la forme suivante :

$$-\mathbb{E} \left[\frac{\partial^2 \Lambda}{\partial \varphi_i \partial \varphi_l} \right] = \frac{2}{\sigma_b^2} \Re \{ \tilde{\Psi}^\dagger \tilde{\Psi} \}, \quad (5.11)$$

où

$$\tilde{\Psi} = j [\mu_0[\mathbf{n}] \bullet s[\mathbf{n}], \dots, \mu_{M_F+1}[\mathbf{n}] \bullet s[\mathbf{n}]] \quad (5.12)$$

est une matrice de taille $(N+1) \times (M_F+2)$.

- La dérivation de la log-vraisemblance par rapport aux paramètres de l'amplitude conduit à l'équation suivante :

$$-\mathbb{E} \left[\frac{\partial^2 \Lambda}{\partial a_i \partial a_l} \right] = \frac{2}{\sigma_b^2} \Re \{ \Psi^\dagger \Psi \}. \quad (5.13)$$

- La dérivation partielle par rapport aux paramètres de l'amplitude et de la phase conjointement mène aux résultats suivants :

$$-\mathbb{E} \left[\frac{\partial^2 \Lambda}{\partial a_i \partial \varphi_l} \right] = \frac{2}{\sigma_b^2} \Re \{ \Psi^\dagger \tilde{\Psi} \}. \quad (5.14)$$

Finalement, la matrice d'information de Fisher s'écrit [FF95]

$$\mathbf{I}_\theta = \frac{2}{\sigma_b^2} \Re \left\{ \begin{bmatrix} \Psi^\dagger \Psi & \Psi^\dagger \tilde{\Psi} \\ \tilde{\Psi}^\dagger \Psi & \tilde{\Psi}^\dagger \tilde{\Psi} \end{bmatrix} \right\}. \quad (5.15)$$

Les bornes de Cramer-Rao du vecteur des paramètres θ sont données par les éléments diagonaux de l'inverse de la matrice d'information de Fisher.

5.3.1.3 Bornes de Cramer Rao adaptées au modèle du signal court

Dans le contexte des signaux courts, puisque la modélisation polynomiale est appliquée à la fréquence instantanée au lieu de la phase, nous adaptons les expressions des bornes de Cramer-Rao formulées précédemment [FF95]. A partir de l'expression de l'approximation polynomiale de la fréquence (2.8) et du modèle de la phase (2.10), on déduit l'approximation polynomiale de la phase :

$$\begin{aligned} \Phi[n] &= \varphi_0 + 2\pi \left(\sum_{k=-\frac{N}{2}}^n F[k] - \sum_{k=-\frac{N}{2}}^0 F[k] \right), \\ &= \varphi_0 + 2\pi \left(\sum_{k=-\frac{N}{2}}^n \sum_{i=0}^{M_F} f_i g_i[k] - \sum_{k=-\frac{N}{2}}^0 \sum_{i=0}^{M_F} f_i g_i[k] \right), \\ &= \varphi_0 + 2\pi \sum_{i=0}^{M_F} f_i \left(\sum_{k=-\frac{N}{2}}^n g_i[k] - \sum_{k=-\frac{N}{2}}^0 g_i[k] \right), \\ &= \sum_{i=0}^{M_F+1} \varphi_i \mu_i[n]. \end{aligned} \quad (5.16)$$

Posons d'une part $\mu_i[n] = \sum_{k=-\frac{N}{2}}^n g_{i-1}[k] - \sum_{k=-\frac{N}{2}}^0 g_{i-1}[k]$ pour $1 \leq i \leq M_F + 1$ et $\mu_0[n] = 1$ pour $n = -\frac{N}{2}, \dots, \frac{N}{2}$ et d'autre part $\boldsymbol{\theta} = [a_0, \dots, a_{M_A}, \varphi_0, f_0, \dots, f_{M_F}]^T$. La matrice d'information de Fisher pour le vecteur des paramètres v est directement fournie par l'équation (5.15).

Compte tenu de (5.6), (5.12), (5.15) et (5.16), la matrice d'information de Fisher ne dépend que de la base polynomiale $\{g_i[n]\}$, de la forme temporelle du signal $s[n]$ et de celle de la phase instantanée $\Phi[n]$. De plus, dans le cas d'un signal mono-composante, $\boldsymbol{\Psi}^\dagger \tilde{\boldsymbol{\Psi}} = j \left[\sum_{n=-\frac{N}{2}}^{\frac{N}{2}} g_i[n] \mu_l[n] |s[n]| \right]_{i,l}$ est une matrice purement imaginaire ; la matrice d'information de Fisher est par conséquent une matrice diagonale par blocs. Les paramètres d'amplitude sont alors découplés de ceux de la fréquence et de la phase initiale.

Les bornes de Cramer-Rao des paramètres d'amplitude ne dépendent que de la base polynomiale utilisée. Dans le cas particulier où la base est orthogonale, les paramètres d'amplitude sont découplés entre eux, ce qui contribue davantage à l'amélioration de l'estimation. En effet, une telle base transforme la matrice $\boldsymbol{\Psi}^\dagger \boldsymbol{\Psi}$ en une matrice diagonale :

$$\boldsymbol{\Psi}^\dagger \boldsymbol{\Psi} = \text{diag} \left\{ \sum_{n=-\frac{N}{2}}^{\frac{N}{2}} g_0[n]^2, \dots, \sum_{n=-\frac{N}{2}}^{\frac{N}{2}} g_{M_A}[n]^2 \right\}. \quad (5.17)$$

$$\boldsymbol{\Psi}^\dagger \boldsymbol{\Psi} = \mathbb{I}_{M_A+1} = \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix} \text{ si la base est orthonormale.} \quad (5.18)$$

En ce qui concerne les paramètres de la phase, leurs bornes de Cramer-Rao dépendent de l'amplitude instantanée et de la base de fonctions choisies pour la modélisation mais elles ne dépendent pas explicitement de la phase instantanée ni des paramètres du modèle de l'amplitude.

5.3.1.4 Bornes de Cramer-Rao - Découplage des paramètres de la phase

Le couplage entre les paramètres de la phase est représenté par les coefficients sous diagonaux du bloc $\tilde{\boldsymbol{\Psi}}^\dagger \tilde{\boldsymbol{\Psi}}$ de la matrice de Fisher (5.15). Pour que ce couplage ne gêne pas l'interprétation des résultats, nous avons décidé de diagonaliser ce bloc. Etant donné que $\tilde{\boldsymbol{\Psi}}^\dagger \tilde{\boldsymbol{\Psi}}$ est une matrice réelle et symétrique, elle est diagonalisable et s'écrit sous la forme suivante :

$$\mathcal{D} = \mathbf{V}^\dagger \tilde{\boldsymbol{\Psi}}^\dagger \tilde{\boldsymbol{\Psi}} \mathbf{V}; \quad (5.19)$$

où \mathcal{D} est une matrice diagonale et \mathbf{V} est une matrice orthogonale dont les colonnes sont les vecteurs propres de la matrice $\tilde{\boldsymbol{\Psi}}^\dagger \tilde{\boldsymbol{\Psi}}$.

Les paramètres de la phase $\boldsymbol{\varphi} = [\varphi_0, f_0, \dots, f_{M_F}]^T$ réécrits dans la base formée par ces

vecteurs propres sont exprimés par $\mathbf{V}\boldsymbol{\varphi}$ (une combinaison linéaire des coefficients de $\boldsymbol{\varphi}$). Notons que la matrice \mathbf{V} dépend du signal considéré et plus précisément de son amplitude, puisque $\tilde{\boldsymbol{\Psi}}^\dagger \tilde{\boldsymbol{\Psi}}$ l'est aussi (cf. équation (5.12)).

Les bornes découplées formées par les élément diagonaux de \mathbf{D} seront évidemment comparées aux erreurs d'estimation du vecteur paramètres $\mathbf{V}\boldsymbol{\varphi}$.

5.3.1.5 Bornes de Cramer-Rao - Fréquence et amplitude instantanées

Dans [FF95], les auteurs ont déterminé les bornes minimales d'estimation de la phase, de la fréquence et de l'amplitude instantanées en utilisant la forme généralisée des bornes de Cramer-Rao. Cette dernière consiste en la proposition suivante.

Proposition 6 *Soit h une fonction continue et différentiable de vecteur $\boldsymbol{\theta}$, les bornes de Cramer-Rao du vecteur $\boldsymbol{\beta} = h(\boldsymbol{\theta})$ sont reliées aux bornes de Cramer-Rao du vecteur $\boldsymbol{\theta}$ [FF95] :*

$$BCR(\boldsymbol{\beta}) = \frac{\partial \mathbf{h}^\top}{\partial \boldsymbol{\theta}} BCR(\boldsymbol{\theta}) \frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}}$$

Appliquons cette proposition à l'amplitude, la phase et la fréquence instantanées sachant que, d'une part, nous avons

$$\begin{aligned} \forall i \in [0, M_A] \quad & \frac{\partial A[n]}{\partial a_i} = g_i[n], \\ \forall i \in [0, M_F + 1] \quad & \frac{\partial \Phi[n]}{\partial \varphi_i} = \mu_i[n], \\ \forall i \in [1, M_F + 1] \quad & \frac{\partial F[n]}{\partial \varphi_i} = \frac{\partial F[n]}{\partial f_{i-1}} = g_{i-1}[n] \quad \text{et} \quad \frac{\partial F[n]}{\partial \varphi_0} = 0, \end{aligned}$$

et que, d'autre part, nous avons

$$BCR([a_0, \dots, a_{M_A}]^\top) = \frac{\sigma_b^2}{2} (\boldsymbol{\Psi}^\dagger \boldsymbol{\Psi})^{-1}, \quad \text{et} \quad BCR([\varphi_0, f_0, \dots, f_{M_F}]^\top) = \frac{\sigma_b^2}{2} (\tilde{\boldsymbol{\Psi}}^\dagger \tilde{\boldsymbol{\Psi}})^{-1}.$$

Nous obtenons alors les expressions des bornes de Cramer-Rao pour l'amplitude et la fréquence instantanées :

$$BCR(A[n]) = \frac{\sigma_b^2}{2} \mathbf{g}^\dagger \{ \boldsymbol{\Psi}^\dagger \boldsymbol{\Psi} \}^{-1} \mathbf{g} \quad BCR(F[n]) = \frac{\sigma_b^2}{2} \mathbf{h}^\dagger \{ \tilde{\boldsymbol{\Psi}}^\dagger \tilde{\boldsymbol{\Psi}} \}^{-1} \mathbf{h}, \quad (5.20)$$

où $\mathbf{g} = [g_0[n], g_1[n], \dots, g_{M_A}[n]]^\top$ et $\mathbf{h} = [0, g_0[n], g_1[n], \dots, g_{M_F}[n]]^\top$.

Nous insistons sur l'intérêt de ces bornes (5.20), qui sont des fonctions variant dans le temps. Vu que l'amplitude évolue au cours du temps et par conséquent le RSB n'est pas constant, des différences dans les performances de l'estimation peuvent apparaître au cours du temps. La comparaison des erreurs quadratiques moyennes obtenues dans l'estimation de l'amplitude et de la fréquence instantanées avec ces bornes (5.20) fournit une idée sur l'évolution temporelle des performances.

Dans la suite, nous étudions les performances de l'algorithme du recuit simulé proposé en l'appliquant à des signaux simulés et bruités. Nous commençons par l'étude d'un cas classique d'un signal non-stationnaire à amplitude constante et ce afin d'illustrer la qualité de l'estimation en présence d'un RSB constant. Nous comparons les résultats obtenus avec les BCR classiques (cf. équations (5.15) et (5.17)) et celles découplées (5.19).

Ensuite, nous nous focalisons sur des signaux courts ayant des modulations non-linéaires à la fois en amplitude et en fréquence. Notons que, dans ce cas, le RSB est une fonction de temps puisque l'amplitude instantanée l'est aussi. Nous analysons d'une part l'influence de la base polynomiale et du taux de modulation sur la précision de l'estimation et d'autre part l'effet d'une limitation du degré de l'approximation.

5.3.2 Signaux à amplitude constante

Considérons deux signaux simulés, notés Signal 3 et Signal 4, dont les paramètres s'écrivent dans la base polynomiale orthonormale discrète (2.21) respectivement :

$$\begin{aligned} \text{Signal 3 : } \boldsymbol{\theta}_{s_3} &= [a_0 = 16, \theta_0 = 0.4, f_0 = 2.0, f_1 = -0.2, f_2 = -0.5]^\top, \\ \text{Signal 4 : } \boldsymbol{\theta}_{s_4} &= [a_0 = 16, \theta_0 = 0.4, f_0 = 1.5, f_1 = -0.2, f_2 = -0.5]^\top. \end{aligned} \quad (5.21)$$

Ces deux signaux ont une amplitude constante et une modulation quadratique de fréquence, l'indice de modulation étant le même pour les deux signaux. La modulation de fréquence du Signal 3 (respectivement Signal 4) est la même que Signal 1 (5.2) (respectivement Signal 2) tracée sur la figure 5.13. L'étude présentée dans ce paragraphe est basée sur 100 simulations du bruit pour chaque RSB, ce dernier variant de 0 à 25 *dB*.

5.3.2.1 Comparaison avec la HAF

Sur les figures 5.2 (a) et (b) sont représentés les erreurs quadratiques moyennes (EQM) et le biais obtenus dans l'estimation du paramètre de fréquence f_1 et de la phase initiale θ_0 dans le cas du Signal 3 et Signal 4 (5.21). Pour les autres paramètres du modèle, des résultats similaires sont fournis en annexe B.1.

Résultats La figure 5.2(a) montre que l'algorithme du recuit est plus robuste au bruit que la technique de la HAF. Pour un RSB inférieur à 10 *dB*, nous obtenons un gain de 20 *dB* sur la précision de l'estimation. De plus, contrairement à la HAF qui échoue complètement dans l'estimation de la phase initiale, l'algorithme proposé arrive à bien estimer θ_0 et réalise des bonnes performances : l'EQM est proche de la BCR et le biais est relativement faible.

Toutefois, ces courbes montrent que pour un RSB situé entre 5 et 15 *dB*, l'EQM obtenue avec le recuit simulé peut descendre en-dessous de la BCR jusqu'à atteindre 20 *dB* (cf. également figures B.1 et B.2 présentées en annexes); cette tendance se confirme sur les deux signaux simulés. L'EQM repasse de nouveau au dessus de la BCR pour des RSB situés en dessous de 5 *dB*. Pour pouvoir expliquer ces résultats, nous les comparons dans le paragraphe suivant avec les BCR découplées.

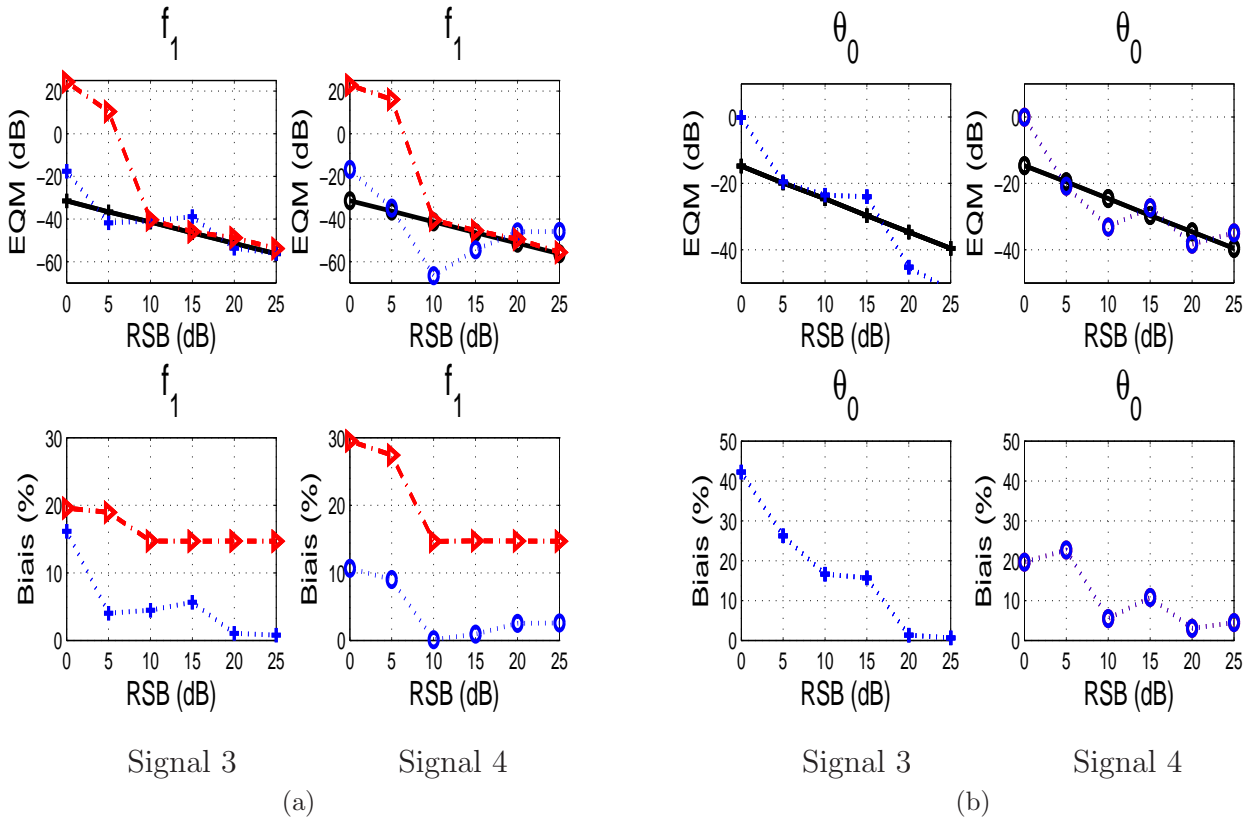


FIG. 5.2 – EQM et biais de l'estimation du paramètre de la fréquence f_1 et de la phase initiale θ_0 en fonction du RSB. Résultats obtenus pour Signal 3 (colonne gauche) de la figure (a) et de la figure (b) : (..) Recuit simulé, (-.Δ-) HAF et (-+) BCR (5.15). Résultats obtenus pour Signal 4 (colonne droite) de la figure (a) et de la figure (b) (..) Recuit simulé, (-.Δ-) HAF et (-o) BCR.

5.3.2.2 Comparaison avec les Bornes de Cramer Rao découplées

Sur la figure 5.3, nous avons tracé les BCR découplées (5.19) et les EQM des estimations des paramètres de la fréquence instantanée et de la phase initiale découplés ($\mathbf{V}\varphi$).

La figure 5.3(a) montre des résultats classiques *i.e.* l'EQM est située au dessus de la BCR découplée et s'en écarte au fur et à mesure que le RSB devient faible. Le couplage entre les paramètres explique les performances présentées dans le paragraphe précédent.

Rappelons que les BCR, présentées dans ce chapitre, sont des bornes asymptotiques déterminées pour des estimateurs non-biaisés. Or, nous notons que les estimateurs deviennent très biaisés pour des faibles RSB inférieurs à 5 dB (cf. figure 5.3(b)). Dans le contexte des signaux courts, ces bornes sont très optimistes mais nous continuons de les utiliser pour avoir une idée sur les performances de l'algorithme proposé.

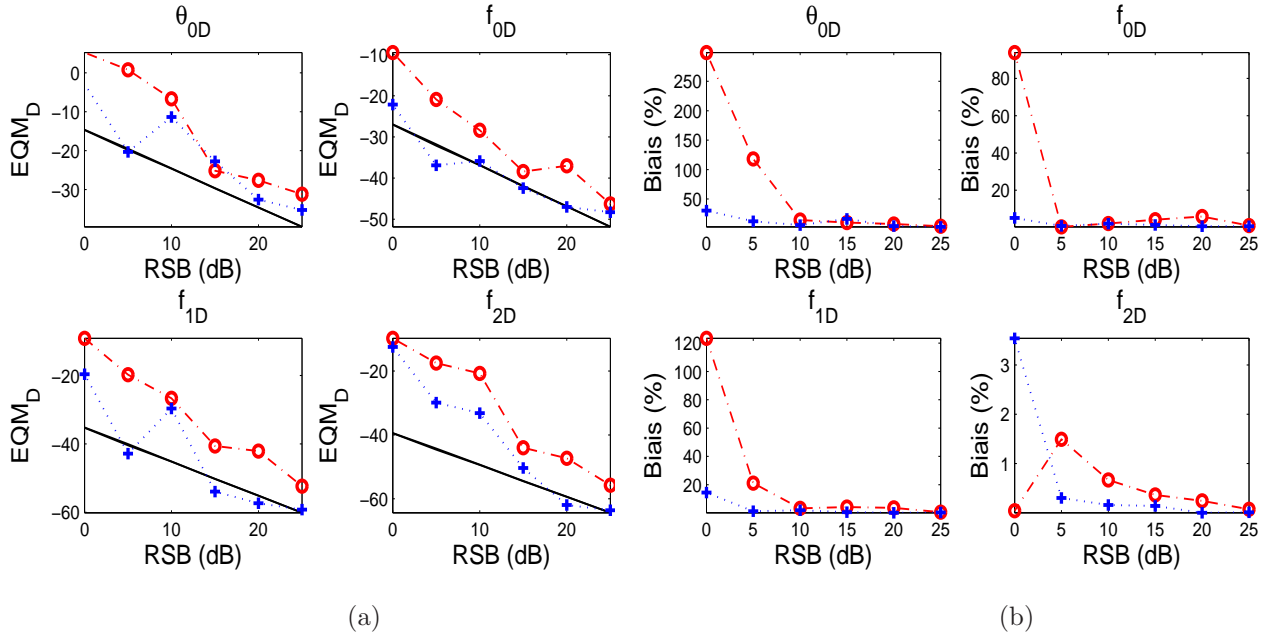


FIG. 5.3 – (a) EQM de l’estimation des paramètres de la fréquence instantanée et de la phase initiale découplés en fonction du RSB : **Signal 3** EQM (..+) et BCR (-+) (5.19) et **Signal 4** EQM (- -o) et BCR (-o). (b) Biases de l’estimation de ces paramètres.

5.3.2.3 Estimation de la fréquence instantanée

La figure 5.4 représente les EQM obtenues dans l’estimation de la fréquence instantanée et le biais en fonction du temps pour Signal 3 et Signal 4 et cela pour un RSB égal à 15 dB. On voit que les performances de l’estimation varient au cours du temps. L’EQM la plus faible et le biais le plus élevé sont situés sur les bords de la fenêtre temporelle.

5.3.3 Signaux à modulations d’amplitude et de fréquence non-linéaires

Dans cette partie, nous étudions des signaux courts ayant une modulation d’amplitude et de fréquence non-linéaires. Dans ce contexte, l’amplitude n’étant pas constante, nous définissons d’abord le RSB. Ensuite, en comparant des bases polynomiales différentes, nous mettons en évidence la contribution du découplage des paramètres d’amplitude dans l’amélioration des résultats. Nous appliquons à la fin l’algorithme proposé sur des signaux à amplitude et à phase non-polynomiale. Ceci servira à illustrer l’effet de la limitation du degré de l’approximation sur le résultat de l’estimation.

Définition du RSB L’amplitude du signal étant variable dans le temps, nous définissons le RSB du signal comme étant égal à : l’énergie d’un signal à amplitude constante (équi-

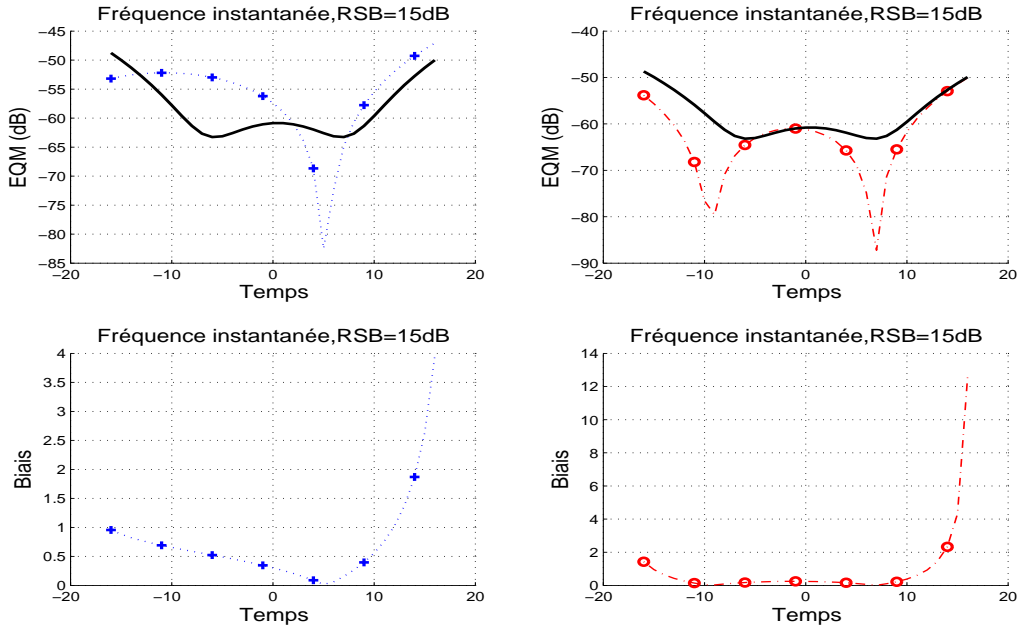


FIG. 5.4 – EQM de la fréquence instantanée et BCR (—) (5.15) en fonction du temps pour un RSB 15 dB : Signal 3 (...+) et Signal 4 (-o-).

valente à celle du signal à amplitude non-constante) normalisée par la variance du bruit :

$$RSB = 10 \log_{10} \left(\frac{\sum_{n=-\frac{N}{2}}^{\frac{N}{2}} |s[n]|^2}{\sigma_b^2 (N+1)} \right). \quad (5.22)$$

Ce RSB, qu'on appellera dans la suite RSB moyen, est différent du RSB instantané défini par

$$RSB_{inst} = 10 \log_{10} \left(\frac{A[n]^2}{\sigma_b^2} \right).$$

Sur la figure 5.5 (respectivement la figure 5.6), cette différence est illustrée dans le cas du Signal 1 (respectivement Signal 2) déjà représentés sur la figure 5.13. Notons que le RSB instantané peut être très faible par rapport au RSB moyen. Pour cette raison, nous représenterons dans ce paragraphe les EQM d'estimation en fonction de la variance du bruit σ_b^2 au lieu du RSB, comme il est classiquement fait.

5.3.3.1 Comparaison entre différentes bases polynomiales

D'après les expressions des bornes de Cramer-Rao (5.17), l'utilisation d'une base polynomiale orthogonale permet de découpler les paramètres de l'amplitude, ce qui laisse présumer une amélioration des résultats de l'estimation.

L'analyse théorique de l'influence de la base choisie est un peu plus délicate puisque l'algorithme proposé comporte des étapes aléatoires dont, par exemple, le choix de l'évolution

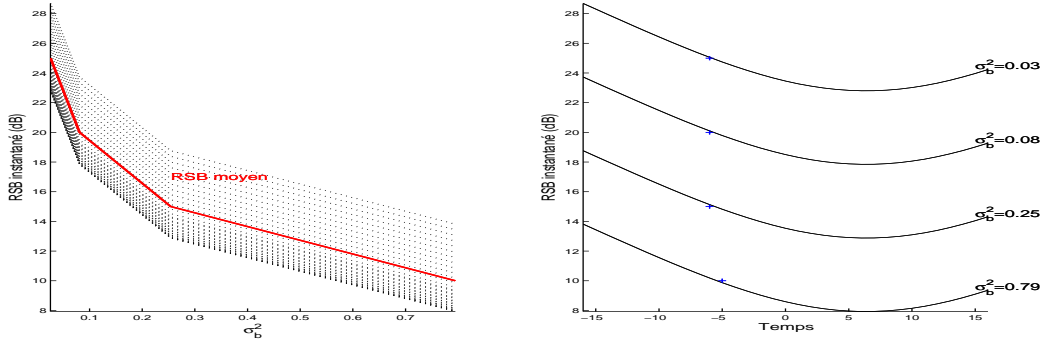


FIG. 5.5 – Variation du RSB instantané en fonction du temps et de la variance du bruit σ_b^2 dans le cas du Signal 1 (5.2), le RSB moyen est indiqué en rouge.

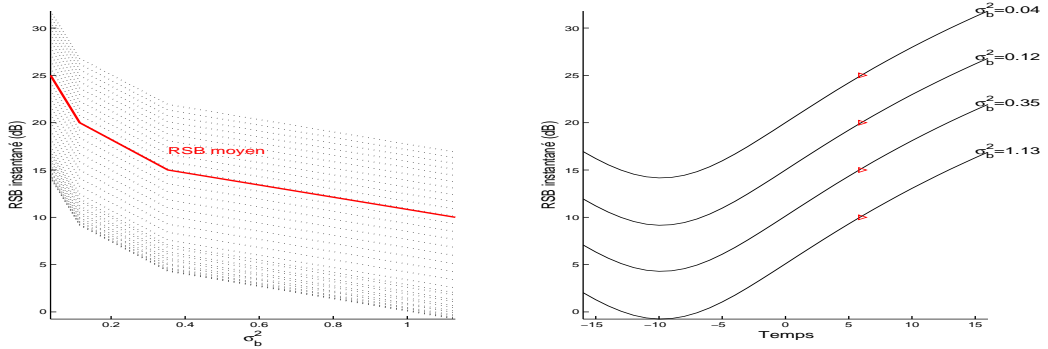


FIG. 5.6 – Variation du RSB instantané en fonction du temps et de la variance du bruit σ_b^2 dans le cas du Signal 2 (5.2), le RSB moyen est indiqué en rouge.

des paramètres du modèle au cours des itérations. Nous présentons comme alternative une étude statistique basée sur des simulations de Monte-Carlo, où trois bases différentes sont comparées : la base canonique (2.14), la base de Legendre discrétisée (2.25) et la base discrète orthonormale (2.21).

100 réalisations bruitées du Signal 1 (5.2) sont effectuées pour chaque valeur de la variance σ_b^2 , les paramètres de ce signal étant ré-exprimés préalablement dans les trois bases citées ci-dessus. Les résultats sont représentés sur les figures 5.7, 5.8 et 5.9 respectivement pour le paramètre de l'amplitude a_2 , la phase initiale θ_0 et le paramètre de la fréquence f_0 . Des résultats similaires sont obtenus pour les autres paramètres du modèle et sont reportés sur les figures B.3 à B.6 fournies en annexes. L'EQM, le biais et les bornes de Cramer-Rao (5.15) sont tracés en fonction de la variance du bruit σ_b^2 .

En comparant les différentes courbes obtenues avec les trois bases utilisées, nous constatons que :

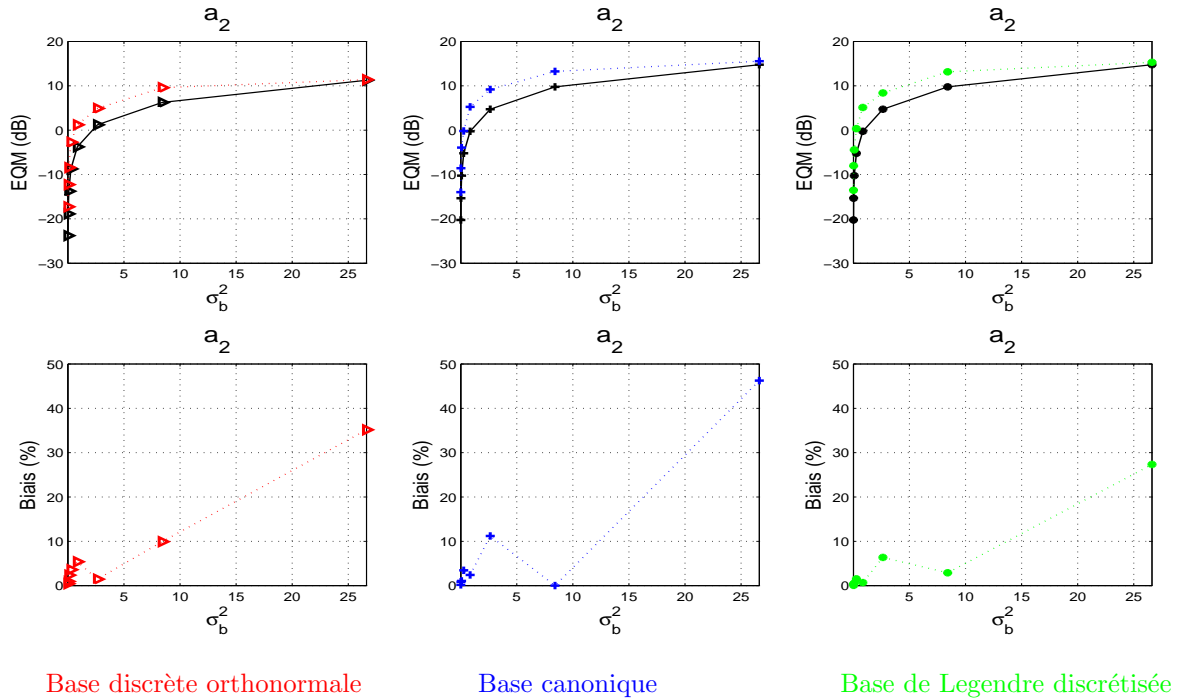


FIG. 5.7 – EQM et biais (%) de a_2 du Signal 1 (5.2) obtenus en utilisant le recuit simulé : base discrète orthonormale (Δ), la base canonique ($+$), base de Legendre discrétisée ($*$) et BCR (5.15) (—).

- pour chacun des paramètres, les courbes présentent pratiquement le même écart entre l'EQM et la courbe de Cramer-Rao correspondante,
- de façon générale l'estimation des paramètres présente le faible biais en utilisant la base orthonormale discrète,
- pour les paramètres d'amplitude a_0 et a_2 (respectivement f_0 et f_2) l'EQM est plus faible en utilisant la base orthonormale discrète, on gagne 5 dB (respectivement jusqu'à 8 dB) par rapport aux autres bases. En effet, ces paramètres sont découplés dans cette base.

L'amélioration ainsi apportée par la base orthonormale justifie son utilisation dans la suite de la thèse.

5.3.3.2 Influence du taux de modulation de l'amplitude

Dans ce paragraphe, nous comparons les résultats pour Signal 1 et Signal 2 (5.2) qui ont un taux de modulation différent pour deux RSB moyens 15 et 5 dB. Sur la figure 5.10, nous avons tracé les EQM obtenues pour la modulation de fréquence et de l'amplitude. Nous constatons une différence dans les performances en fonction du temps surtout pour un RSB moyen égal à 5 dB. Cela est dû essentiellement aux variations du RSB instantané. En effet, comme on peut le constater sur la figure 5.6, ce RSB instantané est très faible du côté gauche de la fenêtre temporelle.

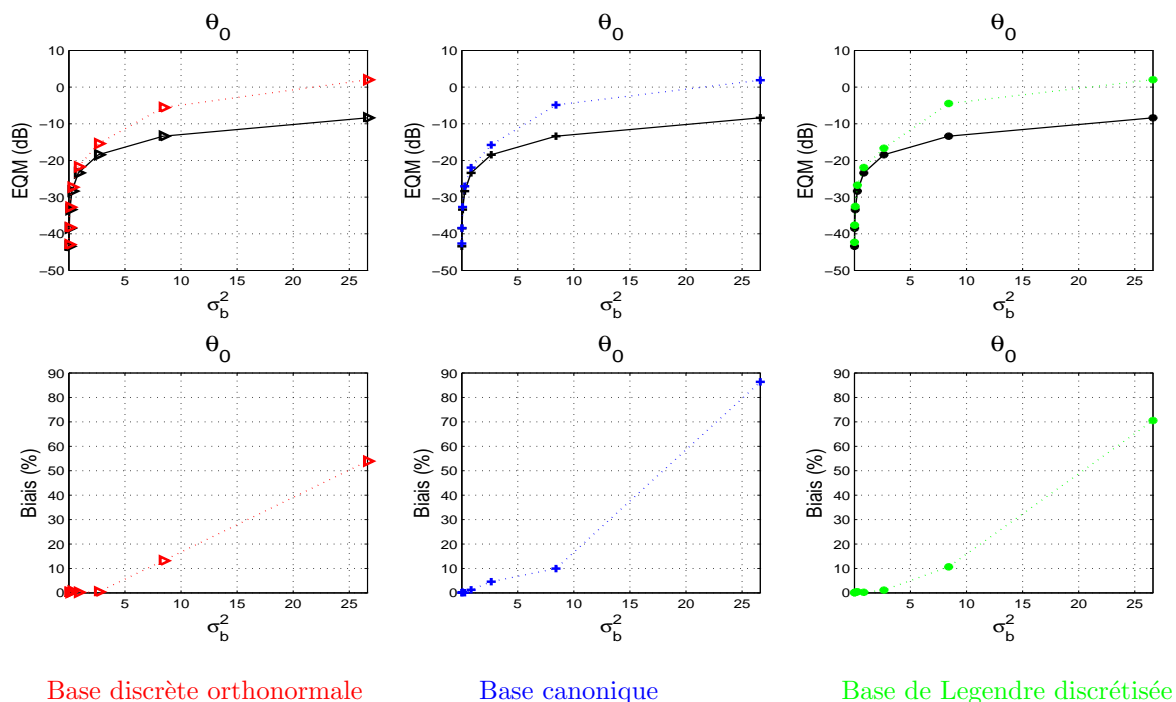


FIG. 5.8 – EQM et biais (%) de θ_0 du Signal 1 (5.2) obtenus en utilisant le recuit simulé : base discrète orthonormale (\triangle), la base canonique ($+$), base de Legendre discrétisée ($*$) et BCR (5.15) (—).

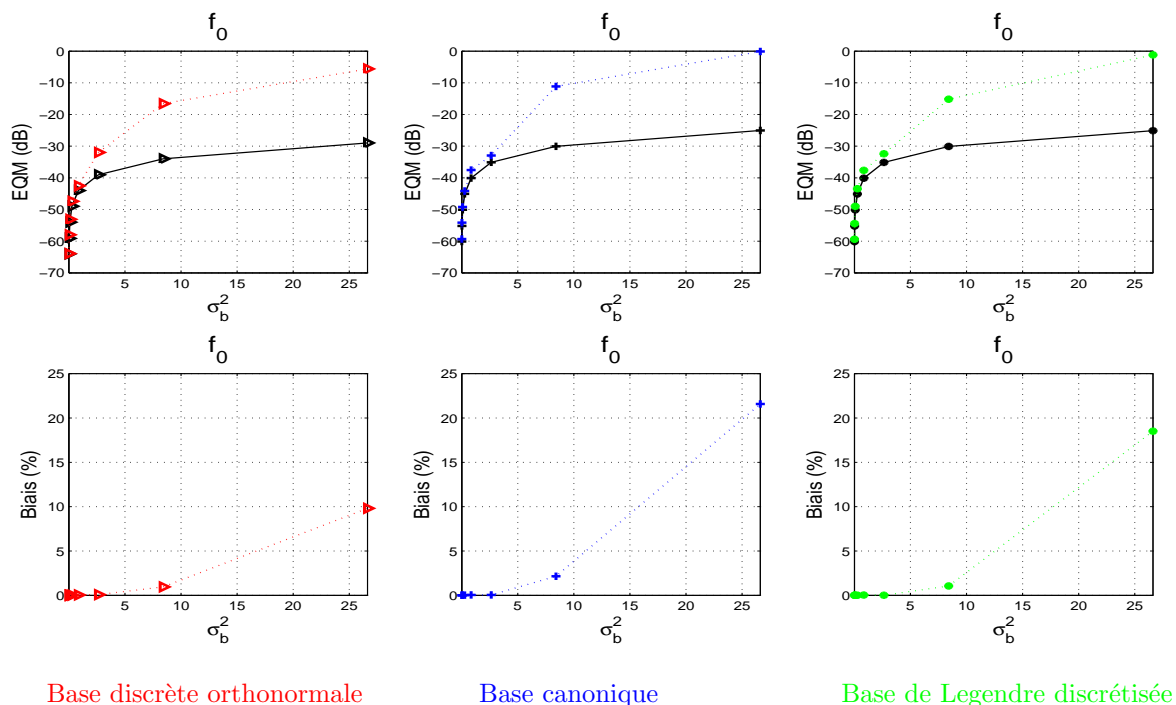


FIG. 5.9 – EQM et biais (%) de f_0 du Signal 1 (5.2) obtenus en utilisant le recuit simulé : base discrète orthonormale (\triangle), base canonique ($+$), base de Legendre discrétisée ($*$) et BCR (5.15) (—).

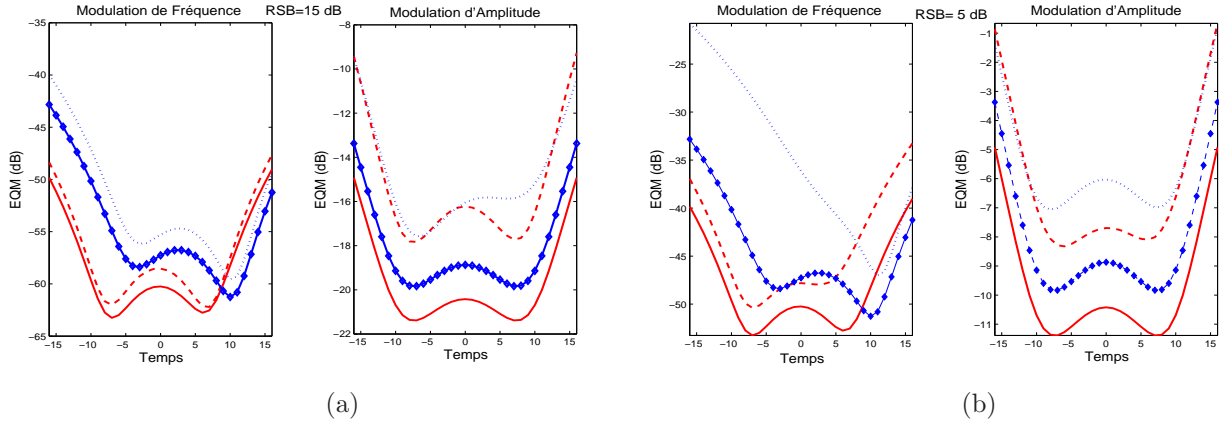


FIG. 5.10 – Influence du taux de modulation sur l'estimation des fréquences des amplitudes instantanées : (a) un RSB 15 dB et (b) 5 dB dans le cas du Signal 1 ((- -) EQM et (-) BCR) et Signal 2 ((...) EQM et (-◇-) BCR).

5.3.3.3 Effet de la limitation de degré d'approximation

Dans ce paragraphe, le but est d'illustrer la robustesse du modèle et de l'estimation par maximum de vraisemblance et l'algorithme du recuit simulé sur des signaux ne suivant pas le modèle (2.8). Nous considérons alors deux signaux longs mono-composantes de 257 échantillons ayant des modulations de natures différentes. La fréquence et l'amplitude instantanées du premier signal sont polynomiales de degré 5 chacun. Le deuxième signal simulé a une modulation en amplitude et en fréquence sinusoidale, dont les expressions sont :

$$\begin{aligned} F[n] &= 0.25 + 0.1 \cos\left(\frac{2\pi n}{256}\right) \\ A[n] &= 5 + 5 \sin\left(\frac{\pi n}{256}\right) \end{aligned} \quad \text{pour } 0 \leq n \leq 256. \quad (5.23)$$

Sur chacun de ces signaux, on prélève deux segments courts sur lesquels le modèle (2.8) sera appliqué. Les EQM comparées aux BCR correspondant (5.20) ainsi que les biais obtenus dans l'estimation de ces segments sont reportés sur la figure 5.11 pour le premier signal et sur la figure 5.12 pour le deuxième signal.

Nous obtenons sur chacun de ces segments une EQM moyenne très proche de la BCR et un biais faible pour l'amplitude et la fréquence. Les performances diffèrent d'un segment à l'autre en raison de la variation du RSB local. Par exemple, sur la figure 5.12, ce RSB est égal à 9.6 dB sur le segment (1) et égal à 16 dB sur le segment (2).

5.3.3.4 Conclusion

Dans cette partie, nous avons analysé la robustesse de l'algorithme proposé du recuit simulé pour réaliser la maximisation de la fonction de la vraisemblance et ainsi obtenir des estimateurs optimaux des paramètres du modèle adopté pour les signaux courts à amplitude et fréquence instantanées non-linéaires. Nous avons obtenu une très bonne robustesse sur des signaux courts en dépit d'un taux de modulation élevé et indépendamment de la nature de modulation. Les performances sont proches des BCR calculées. Nous avons conclu

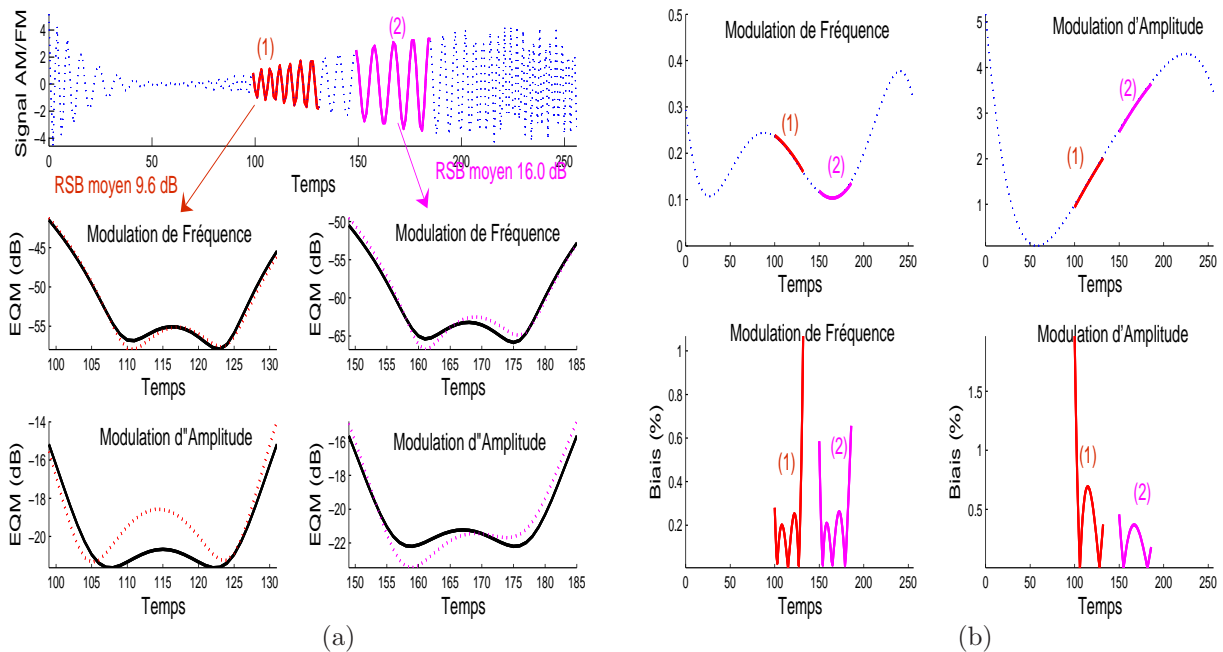


FIG. 5.11 – Modélisation et estimation de deux segments prélevés sur un signal ayant une modulation polynomiale de degré 5 en fréquence et en amplitude : (a) le signal et les segments extraits (1) et (2), (...) EQM , (—) BCR. (b) (...) courbes originales, (—) segments estimés et (—) biais d'estimation.

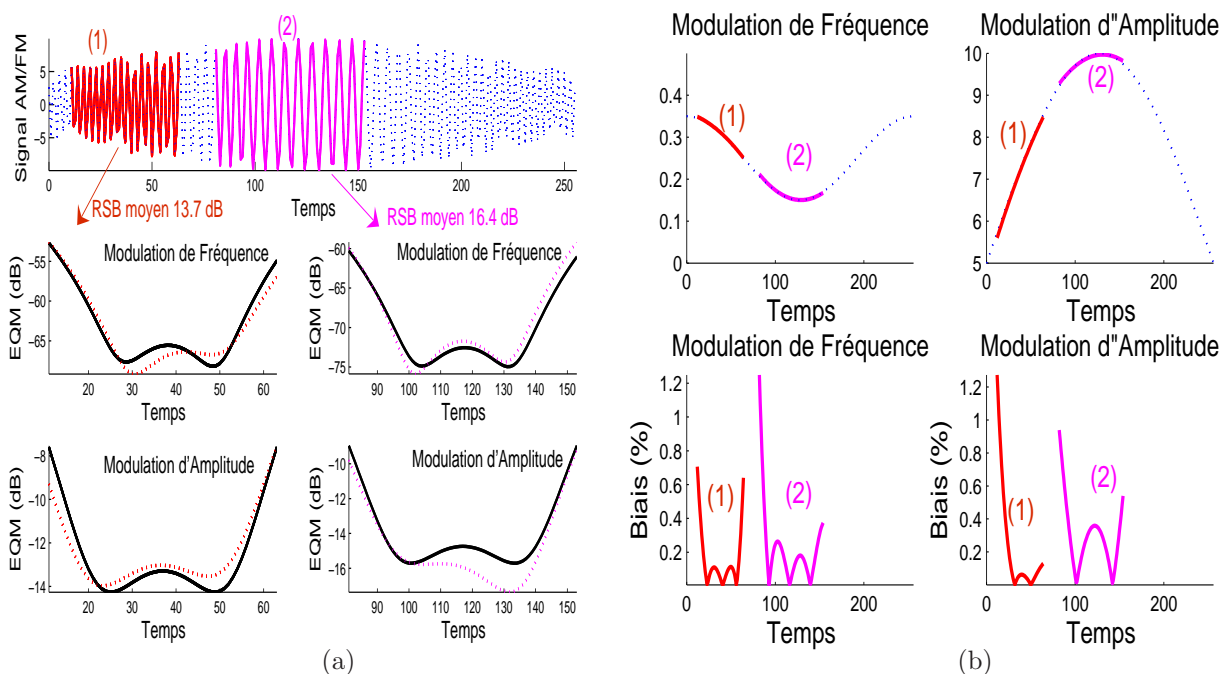


FIG. 5.12 – Modélisation et estimation de deux segments prélevés sur un signal ayant une modulation sinusoïdale en fréquence et en amplitude : (a) le signal et les segments extraits (1) et (2), (...) EQM , (—) BCR. (b) (...) courbes originales, (—) segments estimés et (—) biais d'estimation.

qu'il ne faut pas représenter les résultats en fonction de RSB mais plutôt en fonction, de la variance du bruit, le premier étant variable au cours du temps et la variance étant la même quelle que soit la réalisation. Toutefois, pour rester cohérent avec ce qui est présenté dans la littérature, nous continuons à afficher le résultat en fonction du RSB moyen.

Dans la suite, nous allons comparé l'algorithme du recuit simulé avec les résultats que fournit l'algorithme MH-MCMC proposé.

5.4 Algorithme MH-MCMC et comparaison avec le recuit simulé

5.4.1 Réglage du l'algorithme de MH-MCMC

Dans le but d'estimer les paramètres du modèle (2.11) d'un signal court mono-composante et non-stationnaire par une approche Bayésienne, nous avons proposé dans le chapitre 4 d'utiliser un algorithme de MH-MCMC.

Contrairement au recuit simulé, le nombre de paramètres à régler dans l'algorithme MH-MCMC est réduit. Effectivement, sont à régler le pas de la marche aléatoire (4.29) pour assurer que le domaine de définition des paramètres du modèle du signal soit complètement couvert par l'échantillonnage et réduire le temps de calcul. Rappelons qu'un pas trop large ou très faible ralentit la convergence et qu'il est préférable d'obtenir un taux d'acceptation de l'ordre de 0.45 à 0.5. Il faut aussi choisir d'une façon expérimentale la longueur de la période de chauffage et la longueur totale de la chaîne générée. Nous avons testé plusieurs réglages empiriques sur Signal 1 et Signal 2 (5.2) et nous avons opté pour les valeurs suivantes : la période de chauffage est égale à $T_0 = 9000$, la longueur de la chaîne générée est de $T = 10000$ et le pas est pris égal à 10% des bornes supérieures calculées pour les paramètres (2.24).

La figure 5.13 illustre une chaîne mélangée obtenu pour le paramètre a_1 de l'amplitude

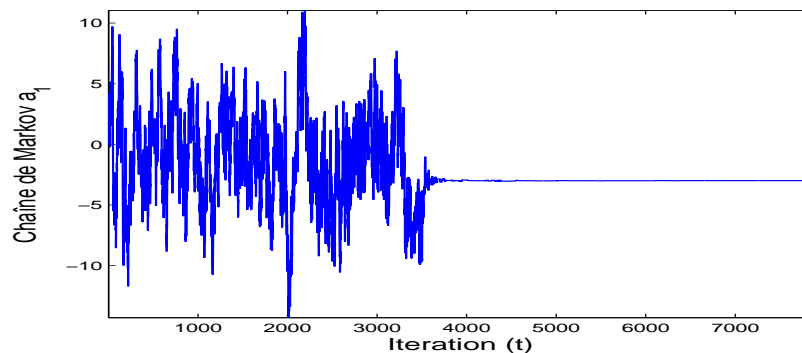


FIG. 5.13 – Chaînes de Markov générées par l'algorithme MH-MCMC pour le paramètre $a_1 = -3$ de l'amplitude de Signal 1 (5.2).

de Signal 1 (5.2). Le taux d'acceptation est de 0.4432.

Toutefois un inconvénient de cet algorithme est le temps nécessaire de calcul. Il est en moyenne 6 fois plus élevé comparé au temps de calcul que met le recuit simulé et 600 fois plus long que la HAF. Un compromis entre la précision souhaitée et le temps de calcul est nécessaire.

5.4.2 Comparaison MH-MCMC, recuit simulé et HAF

Nous considérons dans cette partie Signal 1 (5.2) et nous évaluons l'EQM en fonction du RSB pour effectuer une comparaison entre l'algorithme du recuit simulé et celui de MH. Les deux algorithmes proposés utilisent la base orthonormale.

Le résultat est illustré sur la figure 5.14 où les EQM obtenues dans l'estimation de la fréquence et de l'amplitude instantanées sont comparées pour un RSB égal à 10 dB puis à 0 dB. Nous avons également tracé les bornes de BCR (5.15).

Rappelons que les résultats du MH-MCMC devrait être comparés à des bornes minimales d'estimation Bayésienne [BSET97]. Ces dernières n'ont pas encore été établies dans littérature pour un modèle à amplitude et à fréquence polynomiale.

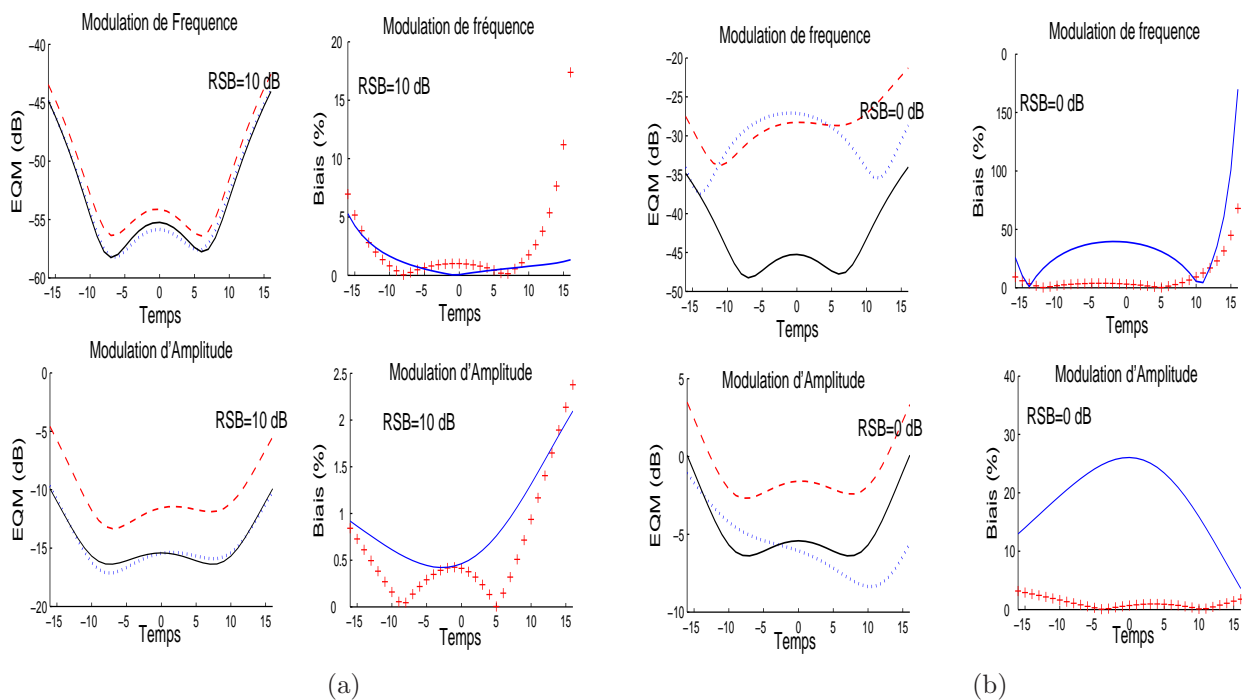


FIG. 5.14 – Comparaison entre le MH-MCMC (...) EQM et (—) Biais, le recuit simulé (-.-) EQM et (++++) Biais et BCR (5.20) (—) appliqué au Signal 1 pour un RSB égal à 10 et 0 dB.

Sur la figure 5.14, l'EQM obtenue par un estimateur MMSE en utilisant l'algorithme MH-MCMC proposé est beaucoup plus faible que celle fournie par le recuit simulé. Cependant le biais obtenu par l'algorithme MH-MCMC est très élevé comparé au recuit simulé.

Dans le Tableau 5.10, nous avons reporté la moyenne, le biais et l'écart type des paramètres f_0 , f_2 et a_2 pour un RSB égal à 5 dB en utilisant ces deux algorithmes et en les comparant à la technique de la HAF. Dans le Tableau 5.11, le RSB moyen évalué à partir du signal résidu en utilisant ces trois méthodes est indiqué pour un RSB vrai de 10 dB puis de 0 dB.

TAB. 5.10 – Comparaison entre recuit simulé, MH-MCMC et HAF : Moyenne, biais et écart-type des estimateurs de f_0 , f_2 et a_2 pour un RSB égal à 5 dB.

Valeur vraie		f_0	f_2	a_2
		2	-0.5	2
Recuit Simulé	Moyenne	2.0015	-0.4995	2.0290
	Biais	0.0015	0.0005	0.0290
	Ecart-type	0.0251	0.0351	1.7572
MH-MCMC	Moyenne	2.0034	-0.4915	1.7150
	Biais	0.0034	0.0085	0.2850
	Ecart-type	0.0199	0.0319	1.1865
HAF	Moyenne	0.9433	-0.5621	1.8326
	Biais	1.0567	0.0621	0.1674
	Ecart-type	7.6622	1.6436	1.4775

TAB. 5.11 – Comparaison entre recuit simulé, MH-MCMC et HAF : RSB estimé à partir du signal résidu obtenu par les trois algorithmes.

Valeur vraie		RSB (dB)	
		10	0
Recuit simulé	RSB estimé	9.8421	-0.2383
MH-MCMC	RSB estimé	9.8048	-1.8771
HAF	RSB estimé	9.7519	-7.1053

A l'issue de la comparaison de ces deux tableaux, nous concluons sur le fait que la méthode de la HAF n'est adaptée à l'étude des signaux courts non-stationnaires. Cette méthode donne des résultats complètement erronés pour un faible RSB inférieur à 5 dB. Les estimations des paramètres obtenues en utilisant le recuit simulé sont plus proches de leurs vraies valeurs que les estimations obtenues par les deux autres algorithmes. L'algorithme de MH-MCMC utilisant un estimateur MMSE donne la plus faible EQM en présence d'un faible RSB (inférieur à 5 dB mais les résultats sont très biaisés. L'algorithme du recuit simulé présente donc un compromis entre le biais, l'EQM et le temps de calcul.

Ainsi en vue d'une application dans un contexte plus général qui est celui des signaux longs, nous allons employer l'algorithme du recuit simulé dans la suite de la thèse. De plus, nous avons pu facilement trouvé un réglage empirique permettant d'être adapté au modèle considéré quels que soient ses paramètres, ce qui n'est pas le cas avec le MH-MCMC.

5.5 Conclusion

Dans ce chapitre, nous avons essentiellement présenté une étude des performances des algorithmes proposés *i.e.* le recuit simulé et celui du MH-MCMC. Nous avons comparé les estimateurs obtenus. Il en résulte que l'algorithme du recuit simulé présente un bon compromis en termes d'efficacité, de performances et de temps de calcul. Nous avons présenté un réglage automatique du recuit simulé adapté aux signaux non-stationnaires courts et ayant un degré de modulation inférieur à 3. Ce travail a été publié dans [VLJM04, JVLM04, JMV07a].

Le but est d'essayer maintenant d'étendre l'application de cet algorithme aux signaux longs. L'idée est de les modéliser sur des segments courts de longueur comprise entre 16 et 64 échantillons et de représenter leur modulation avec un polynôme de degré inférieur à 3. Ceci constitue l'objet du chapitre suivant.

Chapitre 6

Signal long non-stationnaire - Approches et analyses

Résumé

Nous proposons une nouvelle approche qui consiste à modéliser et estimer des signaux longs, non-stationnaires et mono-composantes en procédant par morceaux.

D'abord, le signal est découpé en plusieurs segments courts, sur chacun d'eux la fréquence et l'amplitude instantanées sont modélisées localement par des polynômes de faible degré. Ensuite, les paramètres de chaque modèle local sont estimés en maximisant la fonction de vraisemblance locale *via* la technique du recuit simulé. Les bonnes performances de cette technique et sa robustesse en présence du bruit ont été précédemment établies dans le contexte des signaux non-stationnaires courts. Enfin, les différents modèles locaux estimés sont fusionnés afin de reconstituer les modulations de fréquence et d'amplitude, la phase instantanée et le signal long temporel.

La stratégie de segmentation que nous proposons est basée sur un critère énergétique qui lui confère un caractère non-séquentiel. Les premiers segments sélectionnés et traités sont ceux les plus énergétiques, ainsi leur estimation ne sera pas affectée par les erreurs dues aux segments de faibles énergies. Les segments obtenus avec cette stratégie ne sont pas ordonnés dans le temps, n'ont pas la même longueur et peuvent se chevaucher. De plus, le degré du modèle local étant différent d'un segment à l'autre, nous proposons un processus de fusion adapté à ce contexte et nécessaire pour reconstruire le signal long.

Nous évaluons les performances de cette nouvelle approche par morceaux pour l'estimation des signaux longs fortement non-stationnaires et nous la comparons à la méthode de la HAF et aux bornes de Cramer Rao.

Mots clefs

Signaux non-stationnaires longs, segmentation, fusion, modélisation locale polynomiale, fonction de vraisemblance locale, recuit simulé, bornes de Cramer-Rao.

Sommaire

6.1	Introduction	131
6.2	Problématique	132
6.3	Segmentation d'un signal non-stationnaire long	133
6.3.1	Segmentation basée sur une Distribution Temps-Fréquence	134
6.3.1.1	Choix de la Distribution Temps-Fréquence	134
6.3.1.2	Extraction de la trajectoire fréquentielle	135
6.3.2	Prélèvement des segments	135
6.3.2.1	Prélèvement du premier segment	135
6.3.2.2	Prélèvement du segment suivant	136
6.3.2.3	Extraction des segments - Cas général	137
6.3.2.4	Gestion du début et de la fin du signal long	138
6.3.3	Ajustement de la longueur d'un segment	138
6.3.3.1	Procédure d'estimation du modèle local	140
6.4	Processus de fusion des segments	141
6.4.1	Recouvrement unilatéral	142
6.4.2	Recouvrement bilatéral	143
6.4.3	Choix des fenêtres de fusion	143
6.4.4	Arrêt de la fusion	144
6.5	Synthèse de l'algorithme pour estimer des signaux longs non-stationnaires	144
6.6	Analyse des performances	146
6.6.1	Bornes de Cramer-Rao	146
6.6.1.1	BCR adaptées à l'approche par morceaux	147
6.6.1.2	BCR adaptées à un modèle polynomial d'un signal long	149
6.6.2	Estimation des signaux à phase et amplitude polynomiales	151
6.6.2.1	Modulation d'amplitude lentement variable	151
6.6.2.2	Modulation rapide de l'amplitude	153
6.6.3	Estimation des signaux ayant un modèle sinusoïdal	155
6.6.4	Application sur un signal réel	158
6.7	Conclusion - Avantages et limitations	160

6.1 Introduction

Nous avons proposé, dans les chapitres 3 et 4, une approche adaptée à l'estimation des signaux courts non-stationnaires et mono-composantes. L'amplitude et la fréquence instantanées ont été modélisées par des polynômes de degré faible définis dans une base polynomiale orthonormale discrète. L'estimation des paramètres du modèle était réalisée par deux méthodes différentes. La première est basée sur la maximisation de la fonction de vraisemblance à l'aide de la technique du recuit simulé, tandis que la deuxième s'appuie sur une inférence Bayésienne utilisant des méthodes MCMC simulées par un algorithme de Metropolis Hastings. A l'issue du chapitre 5, il était convenu de se servir plutôt de la première méthode, puisqu'elle correspond au meilleur compromis en terme de faibles erreurs et biais d'estimation et en terme de temps de calcul.

La généralisation triviale de cette approche aux signaux longs et fortement non-stationnaires nécessite, non seulement, un degré d'approximation polynomiale très élevé impliquant un très grand nombre des paramètres à estimer, mais aussi, un temps de calcul colossal.

Nous proposons, en contrepartie, une nouvelle approche basée sur l'estimation par morceaux de ce type de signaux. Moyennant une stratégie de segmentation que nous détaillerons, le signal non-stationnaire long est découpé en segments temporels courts et adaptés à l'application locale de l'approche développée pour les signaux non-stationnaires courts. La fréquence et l'amplitude instantanées du signal long sont alors approchées localement, *i.e.* sur chacun de ces segments, par un polynôme de faible degré. L'estimation des paramètres propres à chaque modèle se fait indépendamment des autres modèles locaux et cela en maximisant la fonction de vraisemblance locale *via* la technique du recuit simulé. Les différents modèles locaux estimés sont ensuite fusionnés dans le but de reconstituer le signal long temporel, ainsi que l'intégralité de sa fréquence, de son amplitude et de sa phase instantanées.

Au lieu de désigner les segments d'une manière séquentielle, ordonnée dans le temps, nous les sélectionnons en fonction de leurs contenus énergétiques. Les premiers segments prélevés et estimés sont ceux centrés sur les parties les plus énergétiques du signal long et *a fortiori* celles correspondant à un fort RSB. Les derniers segments traités sont bien entendu ceux les moins énergétiques. Le recouvrement entre les segments est toléré. Nous soulignerons, dans ce chapitre, l'intérêt de cette planification non-séquentielle de la segmentation dans la réduction de l'erreur d'estimation.

Les longueurs des segments sont contraintes par les variations locales de la fréquence et de l'amplitude instantanées de façon à ce qu'un modèle local polynomial de degré faible soit approprié. Par conséquent, ces longueurs varient d'un segment à l'autre. Nous proposons alors un processus de fusion adapté à ce problème et qui tient compte du recouvrement entre les segments.

Ce chapitre se décompose en trois parties principales. Nous allons tout d'abord exposer la procédure de segmentation non-séquentielle que nous proposons. Puis, nous détaillerons le processus de fusion qui lui est adapté. L'algorithme final regroupe donc ces deux procédures et la méthode d'estimation des paramètres des modèles locaux. Ses performances sont ensuite étudiées dans l'estimation des signaux fortement non-stationnaires longs. L'approche proposée sera particulièrement testée et évaluée sur des signaux à phase polynomiale et non-polynomiale. Les résultats obtenus seront comparés à ceux obtenus par la méthode de la HAF et aux bornes de Cramer-Rao appropriées. Nous verrons enfin l'application de l'approche par morceaux ainsi développée dans l'estimation d'un signal réel non-stationnaire décrivant un chant de canari.

6.2 Problématique

Considérons un signal $S[n]$ non-stationnaire, mono-composante et de longue durée noyé dans du bruit blanc, additif et Gaussien $B[n]$. La variance du bruit est inconnue et notée σ_B^2 . Le signal bruité $Y[n]$ s'écrit :

$$Y[n] = S[n] + B[n], \quad (6.1)$$

$$S[n] = A[n] e^{j\Phi[n]} \quad \text{pour } n = 0, 1, \dots, N-1, \quad (6.2)$$

où $A[n]$ et $\Phi[n]$ désignent respectivement l'amplitude et la phase instantanées du signal $S[n]$ et N représente le nombre d'échantillons disponibles. La fréquence instantanée de ce signal est notée $F[n]$ et est définie par :

$$\Phi[n] = 2\pi \sum_{k=0}^n F[k] - \Phi[0] \quad \text{pour } n = 0, 1, \dots, N-1. \quad (6.3)$$

Dans ce chapitre, nous supposons que l'amplitude, la fréquence et la phase instantanées vérifient les conditions suivantes :

- la phase instantanée ne présente pas de discontinuités, *i.e.* absence de sauts de phase,
- l'amplitude instantanée est strictement positive :

$$A[n] > 0, \quad \text{pour } n = 0, 1, \dots, N-1, \quad (6.4)$$

- la fréquence instantanée est strictement positive et inférieure à la moitié de la fréquence d'échantillonnage, notée F_e :

$$0 < F[n] < \frac{F_e}{2}, \quad \text{pour } n = 0, 1, \dots, N-1. \quad (6.5)$$

Ces conditions permettront d'éviter les problèmes d'ambiguïté relative à la définition de la phase du signal $S[n]$ [RB74], comme précédemment expliqué dans le chapitre 2.

Le but de l'approche par morceaux, qui sera développée au cours de ce chapitre, est d'estimer localement $A[n]$, $F[n]$ et $\Phi[n]$ sur des segments temporels courts, puis de fusionner les estimées locales de chacune de ces grandeurs pour déterminer son estimation globale sur $[0, N-1]$. Ainsi, on s'affranchira d'un modèle global défini sur la totalité de l'intervalle $[0, N-1]$ et nécessitant un degré d'approximation très élevé pour traduire d'une façon satisfaisante les variations de ces grandeurs.

Ne pas supposer de modèle *a priori* ni pour $A[n]$, ni pour $F[n]$, ni pour $\Phi[n]$, constitue d'ailleurs une nouveauté supplémentaire de l'approche proposée, qui assurera sa supériorité dans l'estimation des signaux fortement non-stationnaires par rapport à d'autres méthodes paramétriques existantes.

6.3 Segmentation d'un signal non-stationnaire long

Considérons un signal fortement non-stationnaire long $Y[n]$ tel que celui présenté sur la figure 6.1). Nous cherchons à en extraire des segments courts qui conviendront à une modélisation locale polynomiale de faible degré (inférieur ou égal à 3) de la fréquence et de l'amplitude instantanées.

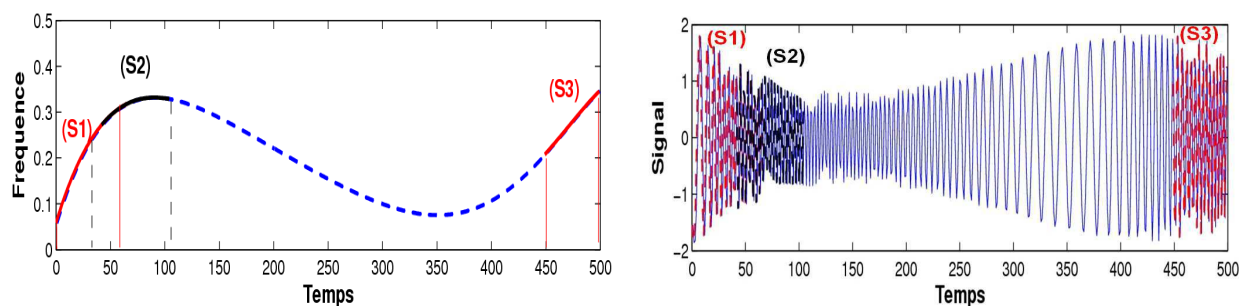


FIG. 6.1 – (a) Exemples de segments courts adaptés à une modélisation locale polynomiale de second degré et extraits à partir d'un signal fortement non-stationnaire ayant une amplitude et une fréquence instantanées polynomiales de degrés respectifs 3 et 7. (S1), (S2) et (S3) sont 3 segments de longueurs respectives 55, 63 et 51 échantillons. (b) La fréquence instantanée.

Ils existent différentes stratégies permettant de prélever des segments à partir d'un signal long, nous citons par exemple :

- a- utiliser une fenêtre glissante de longueur fixe avec ou sans appliquer de recouvrement entre les segments,
- b- utiliser une fenêtre glissante de longueur variable avec ou sans appliquer de recouvrement entre les segments,
- c- utiliser une méthode itérative par "dichotomie", qui consiste à répéter le découpage d'un intervalle en deux parties de même taille, puis de tester si sur les sous-intervalles obtenus $F[n]$ peut être approchée localement par un polynôme de faible degré.

La longueur des segments dans la stratégie (a) étant fixe, $F[n]$ peut présenter une variation locale très importante sur un des segments, exigeant ainsi un degré d'approximation élevé (supérieur à 3). Or, sous-estimer ce degré entraîne forcément une dégradation de l'estimation. Pour pouvoir conserver un modèle de faible degré (contrainte du problème), il serait alors nécessaire de diminuer la longueur du segment, ce qui n'est pas possible avec cette stratégie.

Les stratégies (b) et (c) sont alors plus flexibles, puisque la longueur des segments peut être ajustée pour convenir à une modélisation de degré faible. En revanche, elles présentent chacune un inconvénient majeur. En effet, l'étude présentée dans le chapitre 5 montre que l'erreur d'estimation la plus élevée est obtenue sur les bords des segments tandis qu'elle est plus faible au centre. Dans la stratégie (c), l'absence de recouvrement entraîne donc une moins bonne estimation sur l'ensemble des bords, même si ces derniers appartiennent à des zones à fort RSB. Certes, ce problème est amélioré avec la stratégie (b) grâce à l'application d'un recouvrement, seulement rien ne garantit que les zones les plus énergétiques du signal coïncident avec les centres des segments prélevés. Or, ceci constitue un point essentiel permettant la réduction de l'erreur d'estimation.

Afin d'améliorer la précision de l'estimation, nous proposons et détaillerons dans les paragraphes suivants une stratégie de segmentation non-séquentielle qui tient compte de l'aspect énergétique du signal long.

6.3.1 Segmentation basée sur une Distribution Temps-Fréquence

Nous allons considérer une distribution d'énergie du signal bruité $Y[n]$ définie dans le plan temps-fréquence et notée $DTF_Y[n, \nu]$. Cette DTF va nous fournir une estimation approximative des variations locales de $A[n]$ et de $F[n]$ à partir de laquelle seront déterminées la position temporelle des segments et leurs longueurs.

Avant d'explicitier le déroulement de la stratégie proposée, nous allons discuter le choix de la DTF, essentiel pour la suite de la démarche.

6.3.1.1 Choix de la Distribution Temps-Fréquence

La plupart des DTFs appartenant à la classe de Cohen, telles que la distribution de Wigner et celle de Choi-Williams, ne peuvent pas être interprétées comme une distribution d'énergie. Elles peuvent prendre des valeurs négatives dues probablement aux termes d'interférences qui apparaissent en présence d'une modulation de fréquence non-linéaire (ou d'un signal à composantes multiples, cf. chapitre 1). Le spectrogramme, en revanche, donne une DTF définie-positive. Toutefois, la résolution de ce dernier étant limitée par le principe d'Heisenberg-Gabor, on peut lui préférer parfois le spectrogramme ré-alloué (cf. chapitre 1).

Dans [FP99] est également décrite une procédure itérative permettant la construction d'une nouvelle DTF positive, en partant du spectrogramme et en utilisant le principe de la minimisation de l'entropie mutuelle. La TFD ainsi obtenue reflète mieux la répartition

de la totalité de l'énergie du signal en temps et en fréquence.

De façon générale, n'importe quelle DTF définie positive peut être utilisée dans la stratégie de segmentation que nous développerons. En particulier, celles ne présentant pas de termes d'interférences nous conviendraient parfaitement pour étendre la stratégie proposée à l'estimation des signaux multi-composantes. Néanmoins, ne jugeant pas nécessaire d'avoir une description très précise de la répartition d'énergie du signal étudié, nous nous contentons d'utiliser le spectrogramme en privilégiant la résolution temporelle.

Dans la suite, nous explicitons les principales étapes que comportent la stratégie de segmentation proposée et qui sont :

- Extraction de la trajectoire fréquentielle,
- Prélèvement d'un segment,
- Ajustement de sa longueur.

Rappelons que les techniques d'estimation des paramètres des modèles locaux polynomiaux sont celles proposées dans les chapitres dédiés aux signaux courts (cf. chapitres 3, 4 et 5) et ne seront pas abordées ici.

6.3.1.2 Extraction de la trajectoire fréquentielle

Etant donné que dans ce chapitre le signal étudié $Y[n]$ (6.1) est mono-composante, la $DTF_Y[n, \nu]$ comporte une seule trajectoire fréquentielle. L'extraction de cette trajectoire sera notée $F^{Init}[n]$, qui constituera une estimation approximative de la fréquence instantanée $F[n]$ tandis que son énergie fournira une approximation grossière, notée $A^{Init}[n]$, de l'amplitude instantanée $A[n]$. Les expressions de $F^{Init}[n]$ et $A^{Init}[n]$ sont comme suit :

$$F^{Init}[n] = \arg \max_{\nu} |DTF_Y[n, \nu]|, \quad (6.6)$$

$$A^{Init}[n] = |DTF_Y[n, F^{Init}[n]]|^{\frac{1}{2}}. \quad (6.7)$$

La caractérisation d'un segment, *i.e.* la détermination de sa position temporelle et de sa longueur, dépendra de $F^{Init}[n]$ et $A^{Init}[n]$ et constitue l'objet des paragraphes suivants.

6.3.2 Prélèvement des segments

Nous exposons en premier et à part le traitement du premier segment et du deuxième segment. Puis, une généralisation s'en suivra.

6.3.2.1 Prélèvement du premier segment

Soient n_0 et ν_0 les coordonnées en temps et en fréquence du point le plus énergétique du signal, déterminées par :

$$[n_0, \nu_0] = \arg \max_{n, \nu} |DTF_Y[n, \nu]|. \quad (6.8)$$

Elles vérifient bien entendu la relation $\nu_0 = F^{Init}[n_0]$. Dès lors, la position temporelle du milieu du premier segment prélevé est indiquée par n_0 tandis que sa longueur est calculée par :

$$L_0 = \gamma \text{Ent} [\nu_0^{-1}] + 1, \quad \text{avec} \quad \gamma \in \mathbb{N}^*. \quad (6.9)$$

$\text{Ent}[\cdot]$ désigne la partie entière. L_0 est bien définie grâce à l'hypothèse de la non-nullité de la fréquence instantanée ($\nu_0 \neq 0$) (6.5) et elle est contrainte à être impaire pour pouvoir utiliser les méthodes d'estimation proposées dans le chapitre précédent.

L'équation (6.9) prend en considération le fait qu'au moins une période ($\gamma \geq 1$) ou deux du signal sont nécessaires pour comprendre la variation de l'amplitude et de la fréquence et d'estimer correctement les paramètres du modèle polynomial qui sera appliqué¹.

Le déploiement du segment (cf. figure 6.2(a)) vaut $\frac{L_0}{2}$ en partant du point central n_0 et dans les deux directions des temps croissants et décroissants ; le segment résultant s'écrit alors :

$$\text{Segment}_0 = \left[n_0 - \frac{L_0 - 1}{2}, n_0 + \frac{L_0 - 1}{2} \right]. \quad (6.10)$$

Par souci de calcul et de précision de l'estimation, nous ajoutons quelques contraintes supplémentaires sur la longueur d'un segment, qui se résument à :

- Pour ne pas se retrouver dans un problème d'estimation sous-déterminé, où le nombre des paramètres du modèle qui vaut au maximum $M_A + M_F + 3 = 9$ est supérieur au nombre d'échantillons prélevés, nous imposons une longueur minimale, notée L_{min} , en pratique de l'ordre de quinze,
- les méthodes d'estimation utilisées nécessitant un temps de calcul croissant exponentiellement avec la longueur du segment, cette dernière sera limitée par une valeur maximale, notée L_{max} , ne dépassant pas en pratique la soixantaine.

Afin de mieux satisfaire les contraintes du problème, nous proposons dans le paragraphe 6.3.3 un algorithme permettant d'ajuster la longueur d'un segment et le degré du modèle en fonction des variations locales de la fréquence instantanée. Ceci va d'une part augmenter l'adaptabilité du modèle et d'autre part réduire le nombre total de segments à traiter et par conséquent le temps de calcul.

6.3.2.2 Prélèvement du segment suivant

Le choix du segment suivant va dépendre de l'estimation du segment précédent. Notons alors $\hat{A}_0[n]$, $\hat{F}_0[n]$ et $\hat{\Phi}_0[n]$ les estimées locales respectives de $A_0[n]$, $F_0[n]$ et $\Phi_0[n]$ obtenues sur Segment_0 (6.10) en maximisant la vraisemblance locale par l'algorithme du recuit simulé (cf. chapitre 3). L'estimée locale du signal $S[n]$ (6.1) obtenue également sur Segment_0 est notée $\hat{S}_0[n]$.

¹Dans [Ade97], les auteurs montrent que la fréquence, l'amplitude et la phase d'un signal sinusoïdal (harmonique) peut être estimé à partir d'un faible nombre d'échantillons 3 ou 5.

Le signal résidu, noté $Y_{r_0}[n]$, s'écrit à ce stade :

$$Y_{r_0}[n] = \begin{cases} Y[n] - \hat{S}_0[n] & \text{pour } n \in [n_0 - \frac{L_0-1}{2}, n_0 + \frac{L_0-1}{2}] \\ Y[n] & \text{sinon.} \end{cases} \quad (6.11)$$

Nous calculons une nouvelle $DTF_{Y_{r_0}}[n, \nu]$ en utilisant $Y_{r_0}[n]$ pour localiser le point le plus énergétique suivant (cf. figure 6.2(b)), de coordonnées respectives en temps et en fréquence n_1 et ν_1 :

$$[n_1, \nu_1] = \arg \max_{n, \nu} |DTF_{Y_{r_0}}[n, \nu]|, \quad \text{où la } DTF_{Y_{r_0}} \text{ est celle du signal résidu } Y_{r_0}[n]. \quad (6.12)$$

La longueur du nouveau segment, noté $Segment_1$, est $L_1 = \gamma Ent[\nu_1^{-1}] + 1$ et ce dernier s'écrit $Segment_1 = [n_1 - \frac{L_1-1}{2}, n_1 + \frac{L_1-1}{2}]$.

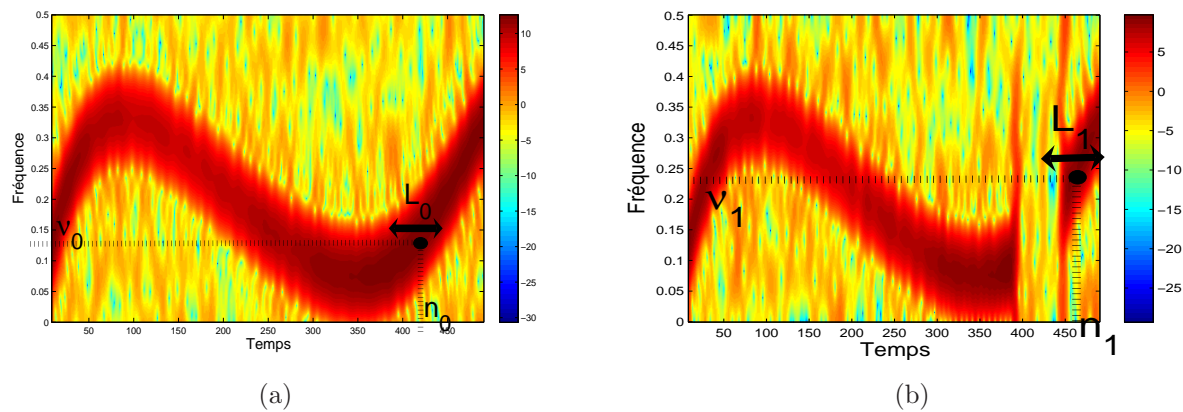


FIG. 6.2 – (a) Prélèvement du 1er segment du signal présenté sur la figure 6.1. (b) Prélèvement du 2ème segment.

De nouveau pour satisfaire les contraintes du problème définies dans le paragraphe précédent, la longueur du nouveau segment pourra subir un ajustement (cf. paragraphe 6.3.3).

6.3.2.3 Extraction des segments - Cas général

Admettons qu'à ce stade, $i - 1$ segments ont été prélevés et estimés. Pour définir la position temporelle du i ème segment, nous avons besoin d'évaluer le signal résidu. Or, dans ce cas deux possibilités peuvent se présenter :

- Soit les segments obtenus jusqu'à l'instant sont à supports disjoints et dans un tel cas, le signal résidu s'écrit :

$$Y_{r_{i-1}}[n] = \begin{cases} Y[n] - \sum_{i=0}^{i-1} \hat{S}_i[n] & \text{pour } n \in Segment_0 \cup \dots \cup Segment_{i-1}, \\ Y[n] & \text{sinon.} \end{cases} \quad (6.13)$$

- Soit il existe un ou plusieurs recouvrement entre les segments obtenus et dans ce cas, il faut se reporter au paragraphe 6.4 pour gérer la fusion des estimées locales de la phase et de l'amplitude. Puis, désignons par $\hat{S}_{f_{i-1}}[n]$ l'estimée provisoire du signal $S[n]$ obtenue sur l'ensemble des segments prélevés après le processus de la fusion à l'itération $i - 1$. Le nouveau signal résidu s'écrit alors :

$$Y_{r_{i-1}}[n] = \begin{cases} Y[n] - \hat{S}_{f_{i-1}}[n] & \text{pour } n \in \text{Segment}_0 \cup \dots \cup \text{Segment}_{i-1}, \\ Y[n] & \text{sinon.} \end{cases} \quad (6.14)$$

Dès alors, les coordonnées en temps et en fréquence du centre de Segment_i , notées n_i et ν_i , sont évaluées à l'aide de la localisation du point le plus énergétique de la nouvelle $DTF_{Y_{r_{i-1}}}[n, \nu]$ de $Y_{r_{i-1}}[n]$. Sa longueur L_i est également déterminée (cf. équations (6.8) et (6.9)).

La procédure décrite dans ce paragraphe est ensuite répétée autant de fois nécessaires pour que la totalité du signal $Y[n]$ soit traitée.

6.3.2.4 Gestion du début et de la fin du signal long

Dans le cas où la distribution temps-fréquence employée n'est pas définie pour le début et la fin du signal, tel est le cas du spectrogramme, la gestion du début et de la fin du signal long doit être faite différemment. Il s'agit de considérer dans ce cas les segments temporels suivants : le segment du début $[0, L_i - 1]$ et celui de la fin $[N - L_i, N - 1]$, qui ont tous les deux une longueur L_i fixée (impair). Par exemple, pour un spectrogramme calculé avec une fenêtre glissante de taille L (pair), N_1 sera pris égal à $L + 1$.

La stratégie de segmentation présentée constitue une des trois étapes fondamentales dans l'estimation des signaux longs non-stationnaires, synthétisées dans le tableau 6.2. Dans la suite, nous détaillons le procédé utilisé pour ajuster un segment prélevé à un modèle de degré faible. Ce procédé va jouer un rôle très important dans la démarche proposée dans ce chapitre.

6.3.3 Ajustement de la longueur d'un segment

Les différentes étapes requises dans l'algorithme d'ajustement de la longueur L_i d'un Segment_i et centré au tour du point énergétique de coordonnées (n_i, ν_i) sont décrites dans le tableau 6.1. Il s'agit, dans un premier temps, de définir une suite de segments croissante $\{\text{Seg}^{(i)}\}$, qu'on initialise avec Segment_i :

$$\text{Segment}_i = \text{Seg}^{(i_0)} \subset \dots \subset \text{Seg}^{(i)} \subset \text{Seg}^{(i+1)} \subset \dots, \quad (6.15)$$

chaque segment étant de longueur $L^{(i)}$:

$$\text{Seg}^{(i)} = \left[n_i - \frac{L^{(i)} - 1}{2}, n_i + \frac{L^{(i)} - 1}{2} \right] \quad (6.16)$$

$$\text{avec } L_i \leq L_{\min} \leq L^{(0)} \leq L^{(i)} \leq \dots \leq L_{\max} \quad \text{et } L^{(i+1)} = L^{(i)} + 2. \quad (6.17)$$

Ensuite, nous déterminons de manière itérative le degré du modèle polynomial. En pratique, on répète les opérations suivantes pour un degré q allant de 0 à 3 :

- a- déterminer un polynôme $\mathcal{P}_{q,i_l}[n]$ de degré q en tant qu'approximation de $F^{Init}[n]$ (6.6) sur $Seg^{(i_l)}$ en utilisant une estimation classique au sens des moindres carrés,
- b- évaluer l'erreur définie par :

$$\mathcal{E} = \sum_{n \in Segment^{(i_l)}} |\mathcal{P}_{q,i_l}[n] - F^{Init}[n]|^2, \quad (6.18)$$

- c- si \mathcal{E} est inférieure à un seuil \mathcal{E}_0 fixé au préalable et déterminé expérimentalement ($\mathcal{E}_0 = 0.1$), alors on augmente la taille du segment, *i.e.* on passe au segment suivant $Seg^{(i_{l+1})}$, et on ré-itére les étapes a, b et c. On s'arrête quand la taille du segment atteindra L_{max} .

A l'issue de ces étapes, nous obtenons la longueur finale de $Segment_i = Seg^{(i_l)}$ et le degré qui minimisent \mathcal{E} .

Une fois $Segment_i$ déterminé, le degré de l'approximation polynomiale de l'amplitude instantanée est évalué de la manière suivante :

- Pour un degré q allant de 0 à 3, déterminer un polynôme $\mathcal{P}_{q,i}[n]$ de degré q en tant qu'approximation de $A^{Init}[n]$ (6.7) sur $Segment_i$ en utilisant une estimation classique au sens des moindres carrés,
- Evaluer l'erreur définie par $\mathcal{E} = \sum_{n \in Segment_i} |\mathcal{P}_{q,i}[n] - A^{Init}[n]|^2$. Si \mathcal{E} est inférieure au seuil $\mathcal{E}_0 = 0.1$, alors on incrémente q .

Le polynôme $\mathcal{P}_{q,i_l}[n]$ de degré q , ayant les coefficients $\mathbf{P} = [p_0, p_1, \dots, p_q]^T$, peut être calculé au sens des moindres carrés, par exemple, par la procédure suivante :

- soit le vecteur $\mathbf{X} = \left[x_{n_0 - \frac{L(i_l)-1}{2}}, \dots, x_{n_0 + \frac{L(i_l)-1}{2}} \right]^T$, formé par les valeurs prises localement par $F^{Init}[n] = x_n$ (ou $A^{Init}[n] = x_n$) sur $Segment^{(i_l)}$,
- soit la matrice de Vandermonde formée par les coefficients $U_{n,l}$ défini par $U_{n,l} = n^{q-l}$ avec $0 \leq l \leq q$,
- l'expression du vecteur \mathbf{P} , calculée au sens des moindres carrés, s'écrit :

$$U \mathbf{p} = \mathbf{X} \quad (6.19)$$

Remarque : il existe une fonction "polyfit" du logiciel "matlab" utilisant cette démarche pour déterminer une approximation polynomiale d'une fonction donnée et à un degré donné au sens des moindres carrés.

Le procédé d'ajustement, décrit dans ce paragraphe, est intéressant pour trois raisons :

- permettre une meilleur adaptabilité du degré de l'approximation polynomiale aux variations locales de la fréquence instantanée sur le segment,

TAB. 6.1 – Algorithme d’ajustement d’un segment à un modèle polynomial de faible degré.

-
- $L_i = \gamma \text{Ent} [\nu_i^{-1}] + 1$, où $\text{Ent}[\cdot]$ est la partie entière et ν_i est définie par (6.8). γ est souvent fixé égal à 2.
 - Soit la suite arithmétique, notée $\{L^{(i)}\}$, de raison 2, de valeur initiale $L^{(0)}$, minorée par L_{min} et majorée par L_{max} .

$$L_{min} \leq L^{(0)} \leq L^{(1)} \leq \dots \leq L^{(i)} \leq \dots \leq L_{max}, \text{ avec } L^{(i+1)} = L^{(i)} + 2.$$

Soit la suite de segments croissante, $\{Seg^{(i)}\}$, initialisée par $Segment_i$ et définie par :

$$Seg^{(i)} = \left[n_i - \frac{L^{(i)} - 1}{2}, n_i + \frac{L^{(i)} - 1}{2} \right], \text{ avec } Seg^{(i)} \subset Seg^{(i+1)}.$$

1. Commencer par $i_l = 1$ et $q = 0$.
2. Approcher $F^{Init}[n]$ restreinte au $Seg^{(i)}$ par un polynôme de degré q , noté $\mathcal{P}_{q,i_l}[n]$, en utilisant une procédure de moindres carrés classique.
3. Evaluer l’erreur quadratique

$$\mathcal{E} = \sum_{n \in Seg^{(i)}} |\mathcal{P}_{q,i_l}[n] - F^{Init}[n]|^2.$$

4. Si $\mathcal{E} \leq \mathcal{E}_0$, incrémenter i_l : $i_l = i_l + 1$, sinon incrémenter q : $q = q + 1$.
5. Arrêter l’algorithme si $q > 3$ ou si $L^{(i)} > L_{max}$. Le segment final obtenu est $Segment_i = Seg^{(i-1)}$ et le degré d’approximation de la fréquence correspondant est $q - 1$.

Remarque : $\mathcal{E}_0 = 0.1$ est déterminée expérimentalement.

- favoriser le recouvrement entre les segments, ainsi l’estimation est améliorée aux bords des segments après la fusion des grandeurs estimées,
- réduire le nombre de segments et par la même occasion le temps de calcul nécessaire à estimer tout le signal long. Par exemple, au lieu de traiter séparément deux segments très courts adaptés tous les deux à un modèle de même degré et qui se chevauchent, le procédé d’ajustement va les regrouper automatiquement en un seul (dans la limite d’une longueur inférieure à L_{max}). Ainsi, un seul modèle ayant le même degré que les deux précédents sera estimé.

Il est tout de même important d’insister sur le fait que, lors de l’ajustement des segments, il faut toujours se servir de la trajectoire fréquentielle, $F^{Init}[n]$ (6.6), obtenue au paragraphe 6.3.2.1 et calculée en utilisant le signal bruité $Y[n]$ et non le signal résidu. Cette remarque est valable pour tous les segments extraits et également dans la recherche des degrés d’approximation appropriés.

6.3.3.1 Procédure d’estimation du modèle local

Une fois l’étape d’ajustement du segment à un modèle de faible degré achevée, l’estimation du modèle est engagée en maximisant la vraisemblance locale par l’algorithme

du recuit simulé comme précédemment décrit dans le chapitre 3. Cette procédure sera appliquée à la partie prélevée du signal bruité $Y[n]$ (6.1) correspondant au segment considéré avec un changement de référence temporelle locale :

$$s[k] = S[n - n_i] \quad \text{pour } k = -\frac{L_i - 1}{2}, \dots, \frac{L_i - 1}{2}, \quad (6.20)$$

$$y[k] = Y[n - n_i]. \quad (6.21)$$

L'initialisation des paramètres du modèle peut se faire en décomposant sur la base polynomiale orthogonale et discrète (2.21) : soient $F^{Init}[n]$ et $A^{Init}[n]$ restreintes localement au segment prélevé, soient leurs approximations polynomiales $\mathcal{P}_{q,i}[n]$ (6.18) déterminées précédemment. Ces dernières permettent uniquement un gain en mémoire de stockage et de calcul.

Dans le cas des segments du début et de la fin du signal long, l'initialisation de l'estimation des paramètres du modèle polynomial se fait par la transformée de Fourier du signal $Y[n]$ restreint à ces deux segments (3.13). D'ailleurs, pour ces deux segments les degrés de l'approximation sont fixés dès le départ à leurs valeurs maximales, *i.e.* 3 que ce soit pour l'approximation de la fréquence ou celle de l'amplitude. Ceci est valable également pour les signaux tels que celui illustré sur la figure 6.3. Sur cette figure, le point énergétique correspondant au centre du segment est situé au bord du spectrogramme.

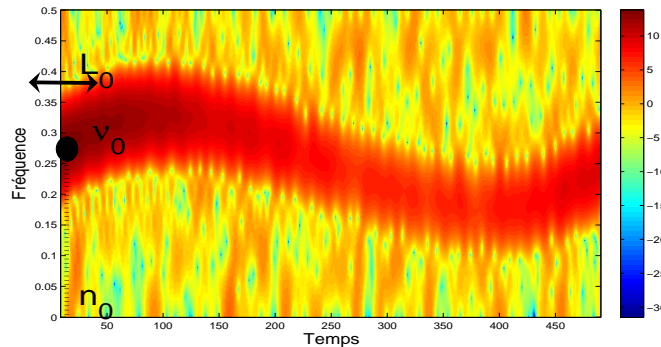


FIG. 6.3 – (a) Prélèvement d'un segment d'un signal ayant une modulation polynomiale d'amplitude de fréquence de degrés respectifs 2 et 3. L'estimation de ce signal sera étudiée dans le paragraphe 6.6.2.1.

La partie suivante de ce chapitre va être consacrée à la gestion du recouvrement des segments estimés, *i.e.* le processus de fusion.

6.4 Processus de fusion des segments

Dès la deuxième itération de la stratégie de segmentation, nous pouvons être confrontés au problème de chevauchement entre les segments. Deux types de recouvrement sont possibles (cf. figure 6.4) : soit un recouvrement d'un seul côté soit des deux côtés.

Pour gérer ces deux situations, nous proposons un traitement adapté et simple : les estimées locales de la phase et de l'amplitude instantanées sont fusionnées en utilisant une moyenne pondérée pour respecter la continuité de la phase. La pondération appliquée tient compte de la répartition de l'erreur d'estimation sur la durée des segments, *i.e.* qu'elle est plus importante au niveau des bords qu'aux centres. Nous explicitons en deux temps le processus de fusion selon le cas de recouvrement.

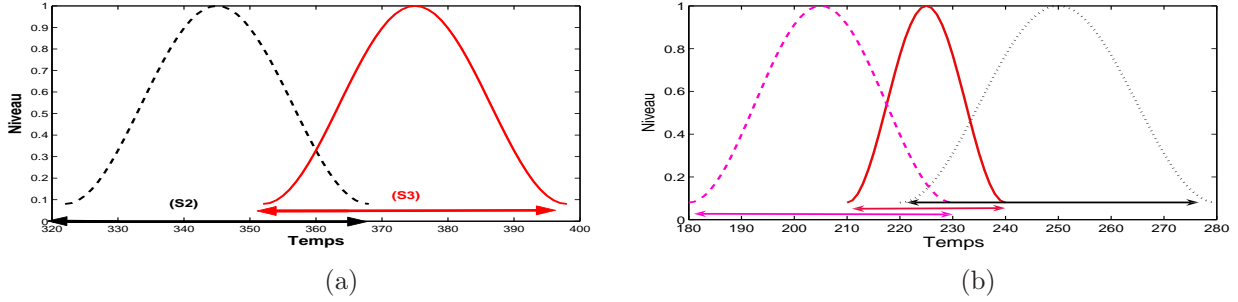


FIG. 6.4 – Deux cas de figures de segments à fusionner : (a) recouvrement d'un seul côté et (b) recouvrement des deux côtés. Les fenêtres de pondération utilisées sont celles de Hamming.

6.4.1 Recouvrement unilatéral

Admettons qu'on dispose de deux segments : un ancien $Segment_p$ prélevé à l'itération p et le nouveau $Segment_i$ avec $p < i$, qui se recouvrent partiellement et considérons les notations suivantes :

$$Seg_{\cap} = Segment_p \cap Segment_i, \quad (6.22)$$

$$Seg_{\cup} = Segment_p \cup Segment_i, \quad (6.23)$$

$$Seg_{i \setminus \cap} = Segment_i \setminus Seg_{\cap}. \quad (6.24)$$

Seg_{\cap} est l'intervalle sur lequel le recouvrement a lieu entre $Segment_i$ et $Segment_p$. Seg_{\cup} désigne l'intervalle composé de l'union des deux segments, il est de longueur inférieure ou égale à la somme de la longueur des deux segments considérés. $Seg_{i \setminus \cap}$ représente la partie du $Segment_i$ où le recouvrement n'a pas eu lieu.

Soient également les estimées locales de la phase et de l'amplitude relatives au $Segment_i$, notées respectivement $\hat{\Phi}_i[n]$ et $\hat{A}_i[n]$. La fusion de chacune de ses dernières et des estimées locales correspondant au $Segment_p$, n'est pas triviale. En effet, l'utilisation d'une simple moyenne arithmétique occasionne l'apparition de sauts de phases et de discontinuités au niveau de l'amplitude. Pour éviter ce problème, nous utilisons des fenêtres de pondération plus lisses (cf. figure 6.4(a)).

Soient $H_i[n]$ (respectivement $H_p[n]$) une fenêtre positive sur le $Segment_i$ (respectivement $Segment_p$) et nulle ailleurs. Le choix de la fenêtre sera discuté dans le paragraphe 6.4.3. La fusion des estimées locales de la phase (notée $\hat{\Phi}_f[n]$) et de l'amplitude (notée

$\hat{A}_f[n]$) est calculée par :

$$\begin{aligned}\hat{\Phi}_f[n] &= \frac{H_i[n]\hat{\Phi}_i[n]+H_p[n]\hat{\Phi}_p[n]}{H_i[n]+H_p[n]}, \quad \forall n \in Seg_\cap \\ &= \hat{\Phi}_i[n], \quad \forall n \in Seg_{i \setminus \cap} \\ &= \hat{\Phi}_p[n], \quad \forall n \in Seg_{p \setminus \cap}.\end{aligned}\tag{6.25}$$

$$\begin{aligned}\hat{A}_f[n] &= \frac{H_i[n]\hat{A}_i[n]+H_p[n]\hat{A}_p[n]}{H_i[n]+H_p[n]}, \quad \forall n \in Seg_\cap \\ &= \hat{A}_i[n], \quad \forall n \in Seg_{i \setminus \cap} \\ &= \hat{A}_p[n], \quad \forall n \in Seg_{p \setminus \cap}.\end{aligned}\tag{6.26}$$

Nous avons, ici, choisi et utilisé la moyenne arithmétique pondérée en dépit de l'existence d'autres moyennes pondérées : géométrique, harmonique *etc*, parce qu'elle est intuitive, simple et adaptée à notre contexte. La seule recommandation est que les fenêtres utilisées ne soient pas nulles au même moment.

6.4.2 Recouvrement bilatéral

Il s'agit d'une généralisation à trois segments : deux anciens notés $Segment_p$ et $Segment_q$ et le nouveau segment $Segment_i$ (cf. figure 6.4(b)). Les trois fenêtres d'analyse qui leur sont associées sont respectivement $H_p[n]$, $H_q[n]$ et $H_i[n]$. Avec les notations suivantes :

$$Seg_\cap = Segment_i \cap Segment_p \cap Segment_q,\tag{6.27}$$

$$Seg_\cup = Segment_i \cup Segment_p \cup Segment_q.\tag{6.28}$$

et en utilisant une moyenne arithmétique pondérée, nous fusionnons les estimées locales de chacune de la phase et de l'amplitude :

$$\hat{\Phi}_f[n] = \frac{H_i[n]\hat{\Phi}_i[n]+H_p[n]\hat{\Phi}_p[n]+H_q[n]\hat{\Phi}_q[n]}{H_i[n]+H_p[n]+H_q[n]}, \quad \forall n \in Seg_\cap\tag{6.29}$$

$$\hat{A}_f[n] = \frac{H_i[n]\hat{A}_i[n]+H_p[n]\hat{A}_p[n]+H_q[n]\hat{A}_q[n]}{H_i[n]+H_p[n]+H_q[n]}, \quad \forall n \in Seg_\cap\tag{6.30}$$

Dans la zone $Seg_\cup \setminus Seg_\cap$, nous avons au plus deux segments qui se recouvrent, nous leur appliquons le calcul décrit dans le paragraphe précédent (cf. l'équation (6.25)).

6.4.3 Choix des fenêtres de fusion

La fenêtre doit attribuer un faible poids aux extrémités du segment, *i.e.* là où l'erreur d'estimation est maximale, et un fort poids au centre du segment puisque l'erreur y est moins importante. Il serait même plus judicieux de déterminer une fenêtre qui épouserait la forme des Bornes Cramer Rao définies par l'équation (6.32), on serait ainsi fidèle aux variations minimales de l'EQM au cours du temps sur le segment considéré. Malheureusement, ce type de fenêtres ne peut pas être envisageable car le calcul des BCR nécessite la connaissance du signal non bruité.

Des formes plus simples des fenêtres de pondération peuvent être employées : Hanning, Hamming, *etc.* Nous avons choisi une fenêtre de Hamming centrée sur le segment, plutôt qu'une fenêtre de Hanning, pour que les estimées aux extrémités d'un segment ne soient pas complètement négligées. La fenêtre de Hamming permet de régler un compromis entre la qualité des estimés locales.

Pour expliquer ce compromis, considérons l'exemple de deux segments dont le premier est de longueur 15 et l'autre est de longueur 61 et qui se recouvrent unilatéralement. Pour un instant n appartenant à la zone du recouvrement, laquelle des deux estimées locales de $\Phi[n]$ obtenues sur ces segments est la plus fiable? Doit-on avantager celle issue du premier segment car on est situé plus près de son centre ou doit-on lui préférer celle issue du deuxième segment qui est plus long et forcément on a plus d'informations? Le choix de la fenêtre de Hamming permet de trouver un compromis.

6.4.4 Arrêt de la fusion

Le processus proposé pour fusionner les segments estimés a pour but de respecter les hypothèses du problème (cf. paragraphe 6.2). En particulier, il permet de conserver la continuité de la phase et de l'amplitude instantanées et cela quel que soit la nature du recouvrement entre les segments. Or, les signaux réels peuvent réellement présenter des discontinuités de la phase, la fusion ne doit pas, donc, avoir lieu dans ce cas. Pour cela, un critère d'arrêt de l'algorithme a été mis en place si l'écart entre les estimées locales de la phase devient très important et dépasse une valeur seuil e_Φ que nous avons fixée expérimentalement égale à $\pi/6$:

$$e_\Phi = \pi/6. \quad (6.31)$$

Nous considérons, dans ce cas, que nous sommes en présence de composantes distinctes du signal réel.

Dans la suite, la stratégie de segmentation, le processus de fusion et l'estimation des modèles polynomiaux locaux sont synthétisés.

6.5 Synthèse de l'algorithme pour estimer des signaux longs non-stationnaires

Dans le tableau 6.2, nous avons récapitulé toutes les étapes intervenant dans le traitement des signaux longs mono-composantes et non-stationnaires. Ces étapes alternent la segmentation, la fusion et l'estimation des modèles locaux. Une illustration du déroulement de cet algorithme est représentée sur les figures 6.5(a) et (b), où le signal fortement non-stationnaire de la figure 6.1 est considéré.

TAB. 6.2 – Algorithme pour estimer les signaux longs non-stationnaires.

1. Calculer $DTF_Y[n, \nu]$ la distribution temps-fréquence de $Y[n]$ et déterminer $F^{Init}[n]$ une approximation grossière de la fréquence instantanée (6.6).
2. Localiser le pic d'énergie en temps et en fréquence respectivement n_0 et ν_0 (6.8).
3. Utiliser la Table 6.1 pour déterminer le segment à traiter et ajuster sa longueur.
4. Estimer l'amplitude et la fréquence instantanées locales en utilisant une approximation polynomiale de degré ≤ 3 . La technique utilisée est celle du maximum de vraisemblance et du recuit simulé présentée dans les chapitres précédents 3 et 5.
5. Si un recouvrement avec un segment précédent existe deux cas sont possibles
 - a- Si la différence des phases locales sur deux segments se chevauchant est inférieure à $e_\Phi = \frac{\pi}{6}$, fusionner les segments en utilisant la stratégie de fusion. (cf. paragraphe 6.4).
 - b- Sinon, retourner à l'étape 8.
6. Retourner à l'étape 8 si toute les observations ont été traitées, sinon aller à l'étape suivante.
7. Calculer $DTF_{Y_r}[n, \nu]$ la distribution temps-fréquence du signal résidu $Y_r[n]$ (6.13) ou (6.14). Retourner à l'étape 2.
8. Arrêter l'algorithme. Le signal estimé de sortie est reconstruit à l'aide de l'équation (6.6) calculée avec les estimées de la phase globale et l'amplitude globale.

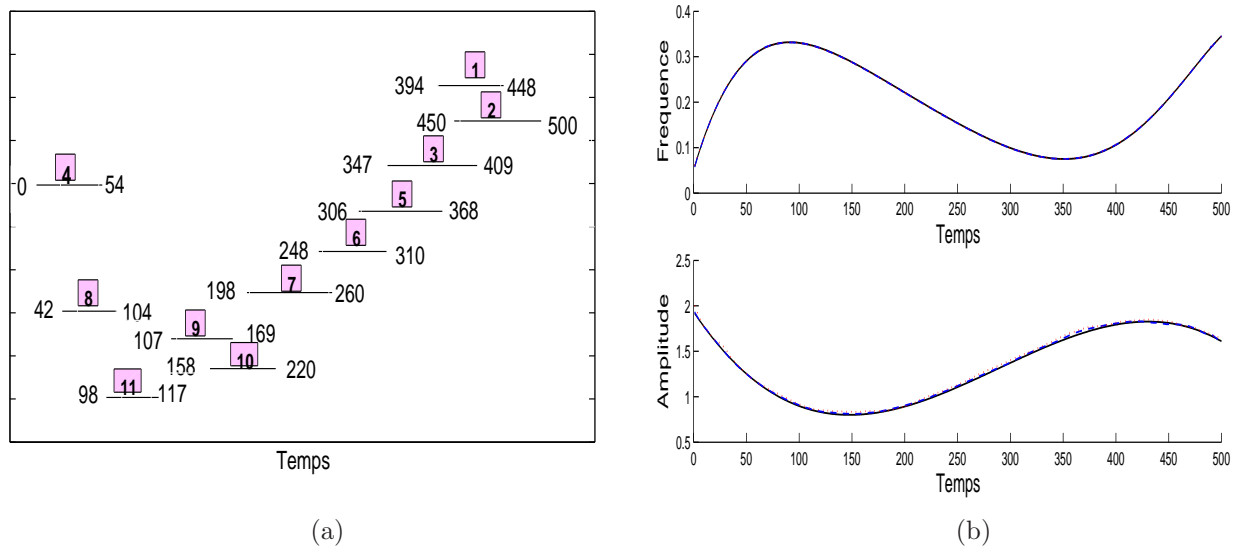


FIG. 6.5 – (a) Illustration de la procédure de la segmentation du signal : le nombre encadré correspond à l'ordre de traitement du segment. (b) Courbes originales (—) de l'amplitude instantanée (polynôme de degré 3) et de la fréquence instantanée (polynôme de degré 7) superposées à leur estimées (...) basées sur une moyenne de 50 réalisations de bruit par des simulations de Monte Carlo pour des RSB moyens de 15 et 7 dB.

Particularités de l'algorithme présenté L'algorithme proposé se distingue des techniques existantes dans la littérature en plusieurs points qui contribuent directement à l'amélioration de la qualité de l'estimation :

- La segmentation non-séquentielle enchaîne l'extraction des segments courts ordonnés en fonction de leur contenu énergétique. L'estimation des segments où le RSB est localement très faible n'affecte pas les autres segments où le RSB est plus fort. Ces derniers sont d'ailleurs estimés en premier.
- La technique d'estimation des modèles locaux (maximum de vraisemblance résolu par la technique du recuit simulé) élimine tout phénomène de propagation d'erreurs (phénomène fréquent avec la méthode de la HAF). En effet, même si la position temporelle et la longueur initiale d'un segment sont déterminées en fonction des segments déjà estimés, l'estimation du modèle local qui lui correspond se fait d'une manière totalement indépendante des autres segments.
- Les centres des segments coïncident avec les points les plus énergétiques, ce qui contribue à la réduction de l'erreur d'estimation.
- Le processus de fusion des segments permet de conserver la continuité de la phase et de l'amplitude instantanées conformément aux hypothèses émises au début de ce chapitre.
- Le signal long reconstruit au final est calculé en utilisant l'équation (5.15) (cf. Tab. 6.2) où la phase et l'amplitude sont celles reconstruites. Le modèle global de départ est, ainsi, conservé et n'a pas été altéré par l'effet de la segmentation et la fusion locale des segments.

Malgré la complexité apparente de cet algorithme, il est facilement implémentable et donne un compromis entre précision de l'estimation et un temps de calcul (CPU) réduit. Dans la suite, son applicabilité sera illustrée sur plusieurs exemples dont un signal réel.

6.6 Analyse des performances

Ayant explicité l'approche par morceaux de l'estimation des signaux longs mono-composantes et fortement non-stationnaires, cette partie est dédiée à l'étude de ses performances. Dans un premier temps, nous définissons les bornes d'estimation adaptées aux estimateurs étudiés. Ces bornes sont des fonctions variant dans le temps. Ensuite, une étude statistique sera menée sur des signaux ayant soit des modulations polynomiales soit des modulations non-polynomiales (sinusoïdales par exemple). Enfin un cas réel sera traité.

6.6.1 Bornes de Cramer-Rao

Dans cette partie, nous nous proposons de déterminer les bornes de Cramer-Rao adaptées à notre modèle d'estimation et qui tiennent compte de la stratégie de la segmentation et de la fusion. Nous commençons par donner l'expression de ces bornes pour un signal non-stationnaire long et mono-composante, en l'absence d'informations *a priori* sur la na-

ture de ses modulations. Ensuite, nous nous intéressons au cas particulier de modulations de fréquence et d'amplitude polynomiales.

6.6.1.1 BCR adaptées à l'approche par morceaux

Etant donnée que, localement, nous employons des modèles polynomiaux de degré inférieur ou égale à 3, nous allons rappeler ici les expressions des BCR obtenues sur un segment noté $Segment_i$ de longueur L_i et de centre n_i (cf. chapitre 3). Ces expressions ont été calculées en utilisant la base discrète orthonormale (2.21), notée :

$$\{g_l[n - n_i]\}_{l=0, \dots, \max(M_{A_i}, M_{F_i})} \quad \text{pour } n \in Segment_i,$$

où M_{A_i} et M_{F_i} représentent les degrés respectifs de l'approximation locale polynomiale de la fréquence instantanée et de celle de l'amplitude instantanée. Les BCR de l'amplitude, de la fréquence et de la phase locales sont alors explicitées par :

$$\begin{aligned} BCR(\hat{A}_i[n]) &= \frac{\sigma_{b_i}^2}{2} \mathbf{g}^\dagger \{ \Psi^\dagger \Psi \}^{-1} \mathbf{g}, & BCR(\hat{F}_i[k]) &= \frac{\sigma_{b_i}^2}{2} \mathbf{h}^\dagger \left\{ \tilde{\Psi}^\dagger \tilde{\Psi} \right\}^{-1} \mathbf{h}, \\ \text{et } BCR(\hat{\Phi}_i[k]) &= \frac{\sigma_{b_i}^2}{2} \boldsymbol{\mu}^\dagger \left\{ \tilde{\Psi}^\dagger \tilde{\Psi} \right\}^{-1} \boldsymbol{\mu}, \end{aligned} \quad (6.32)$$

où $\sigma_{b_i}^2$ est la variance locale du bruit sur le segment considéré et

$$\Psi^\dagger \Psi = \text{diag} \left\{ \sum_{k=-\frac{L_i-1}{2}}^{\frac{L_i-1}{2}} g_0[k]^2, \dots, \sum_{k=-\frac{L_i-1}{2}}^{\frac{L_i-1}{2}} g_{M_{A_i}}[k]^2 \right\}, \quad (6.33)$$

$$\begin{aligned} \tilde{\Psi} &= j [\mu_0[\mathbf{n}] \bullet \mathbf{s}[\mathbf{n}], \dots, \mu_{M_{F_i}+1}[\mathbf{n}] \bullet \mathbf{s}[\mathbf{n}]], & \mathbf{g} &= [g_0[n - n_i], g_1[n - n_i], \dots, g_{M_{A_i}}[n - n_i]]^\top, \\ \mathbf{h} &= [0, g_0[n - n_i], g_1[n - n_i], \dots, g_{M_{F_i}}[n - n_i]]^\top, & \boldsymbol{\mu} &= [\mu_0[n - n_i], \mu_1[n - n_i], \dots, \mu_{M_{F_i}}[n - n_i]]^\top, \end{aligned} \quad (6.34)$$

avec $\mu_l[n - n_i] = 2\pi \left(\sum_{k=-\frac{L_i}{2}}^{n-n_i} g_{l-1}[k] - \sum_{k=-\frac{L_i}{2}}^0 g_{l-1}[k] \right)$ pour $1 \leq l \leq M_{F_i} + 1$ et $\mu_0[n - n_i] = 1$ pour $n \in Segment_i$.

Puisque l'estimation du modèle d'un segment se fait indépendamment des autres, nous supposons l'indépendance des estimateurs issus de segments distincts. Nous supposons également que ces estimateurs ne sont pas biaisés. Dès alors, la proposition suivante peut être utilisée afin de déterminer des bornes d'estimation adaptées à notre contexte :

Proposition 7 Soient \hat{X} et \hat{Y} deux estimateurs indépendants respectivement de deux paramètres X et Y et deux variables réelles déterministes α et β alors

$$\text{variance}(\alpha \hat{X} + \beta \hat{Y}) = \alpha^2 \text{variance}(\hat{X}) + \beta^2 \text{variance}(\hat{Y}) \geq \alpha^2 BCR(X) + \beta^2 BCR(Y)$$

Selon le recouvrement, deux bornes peuvent être calculées.

Cas d'un recouvrement unilatéral Considérons les mêmes notations que celles du paragraphe 6.4. Nous avons :

$$\begin{aligned} \text{variance} \left(\hat{\Phi}_f[n] \right) &= \frac{H_i[n]^2 \text{variance}(\hat{\Phi}_i[n]) + H_p[n]^2 \text{variance}(\hat{\Phi}_p[n])}{(H_i[n] + H_p[n])^2}, \quad \forall n \in \text{Seg}_\cap & (6.35) \\ &= \text{variance} \left(\hat{\Phi}_i[n] \right), \quad \forall n \in \text{Seg}_{i \setminus \cap} \\ &= \text{variance} \left(\hat{\Phi}_p[n] \right), \quad \forall n \in \text{Seg}_{p \setminus \cap}. \end{aligned}$$

$$\begin{aligned} \text{variance} \left(\hat{A}_f[n] \right) &= \frac{H_i[n]^2 \text{variance}(\hat{A}_i[n]) + H_p[n]^2 \text{variance}(\hat{A}_p[n])}{(H_i[n] + H_p[n])^2}, \quad \forall n \in \text{Seg}_\cap & (6.36) \\ &= \text{variance} \left(\hat{A}_i[n] \right), \quad \forall n \in \text{Seg}_{i \setminus \cap} \\ &= \text{variance} \left(\hat{A}_p[n] \right), \quad \forall n \in \text{Seg}_{p \setminus \cap}. \end{aligned}$$

Etant donné que $\text{variance} \left(\hat{\Phi}_i[n] \right) \geq \text{BCR} \left(\hat{\Phi}_i[n] \right)$ et $\text{variance} \left(\hat{A}_i[n] \right) \geq \text{BCR} \left(\hat{A}_i[n] \right)$, nous obtenons de nouvelles bornes d'estimation :

$$\begin{aligned} \text{variance} \left(\hat{\Phi}_f[n] \right) &\geq \frac{H_i[n]^2 \text{BCR}(\hat{\Phi}_i[n]) + H_p[n]^2 \text{BCR}(\hat{\Phi}_p[n])}{(H_i[n] + H_p[n])^2}, \quad \forall n \in \text{Seg}_\cap & (6.37) \\ &\geq \text{BCR} \left(\hat{\Phi}_i[n] \right), \quad \forall n \in \text{Seg}_{i \setminus \cap} \\ &\geq \text{BCR} \left(\hat{\Phi}_p[n] \right), \quad \forall n \in \text{Seg}_{p \setminus \cap}. \end{aligned}$$

$$\begin{aligned} \text{variance} \left(\hat{A}_f[n] \right) &\geq \frac{H_i[n]^2 \text{BCR}(\hat{A}_i[n]) + H_p[n]^2 \text{BCR}(\hat{A}_p[n])}{(H_i[n] + H_p[n])^2}, \quad \forall n \in \text{Seg}_\cap & (6.38) \\ &\geq \text{BCR} \left(\hat{A}_i[n] \right), \quad \forall n \in \text{Seg}_{i \setminus \cap} \\ &\geq \text{BCR} \left(\hat{A}_p[n] \right), \quad \forall n \in \text{Seg}_{p \setminus \cap}. \end{aligned}$$

Nous constatons à partir de l'équation ci-dessus qu'en l'absence de recouvrement ce sont les BCR établies dans le chapitre précédent qui sont considérées.

La variance de l'estimation locale de la fréquence instantanée nécessite un calcul plus compliqué, nous nous contentons d'établir une expression approchée. Nous définissons la fréquence obtenue, après fusion appliquée sur Segment_i et Segment_p , par :

$$\hat{F}_f[n] = \frac{1}{2\pi} \left(\hat{\Phi}_f[n] - \hat{\Phi}_f[n-1] \right). \quad (6.39)$$

Après simplification (6.25), nous obtenons :

$$\begin{aligned} \hat{F}_f[n] &= \frac{H_i[n] \hat{F}_i[n] + H_p[n] \hat{F}_p[n]}{H_i[n] + H_p[n]} + & (6.40) \\ &\frac{\hat{\Phi}_i[n-1] - \hat{\Phi}_p[n-1]}{2\pi} \frac{H_i[n] H_p[n-1] - H_i[n-1] H_p[n]}{(H_i[n] + H_p[n]) (H_i[n-1] + H_p[n-1])} \\ &\text{pour } n \in \text{Seg}_\cap. \end{aligned}$$

Dans le premier terme de l'équation (6.40), on reconnaît une moyenne pondérée des estimées locales des fréquences instantanées. Nous rappelons leur relation avec les estimées locales

de la phase $\hat{\Phi}_i[n] = \varphi_{0_i} + 2\pi \left(\sum_{k=-\frac{L_i}{2}}^{n-n_i} \hat{F}_i[n_i+k] - \sum_{k=-\frac{L_i}{2}}^0 \hat{F}_i[n_i+k] \right)$.

En utilisant l'équation (6.31) définissant l'écart maximum toléré pour réaliser la fusion, qui est égal à $\frac{\pi}{6}$, le second terme est largement inférieur à $\frac{\hat{\Phi}_i[n-1] - \hat{\Phi}_p[n-1]}{2\pi} \leq \frac{1}{12}$. Si de plus on réalise des bonnes estimations de la phase, nous devons théoriquement avoir $\hat{\Phi}_i[n-1] \approx \hat{\Phi}_p[n-1]$.

Admettons qu'on peut donc négliger le second terme, l'expression de la fréquence après fusion est :

$$\hat{F}_f[n] \approx \frac{H_i[n]\hat{F}_i[n] + H_p[n]\hat{F}_p[n]}{H_i[n] + H_p[n]}. \quad (6.41)$$

Une borne d'estimation sera donc :

$$\begin{aligned} \text{variance} \left(\hat{F}_f[n] \right) &\geq \frac{H_i[n]^2 BCR(\hat{F}_i[n]) + H_p[n]^2 BCR(\hat{F}_p[n])}{(H_i[n] + H_p[n])^2}, \quad \forall n \in \text{Seg_INTER} \\ &\geq BCR \left(\hat{F}_i[n] \right), \quad \forall n \in \text{Seg}_{i \setminus \cap} \\ &\geq BCR \left(\hat{F}_p[n] \right), \quad \forall n \in \text{Seg}_{i \setminus \cap}. \end{aligned} \quad (6.42)$$

Les BCRs ainsi obtenus sont illustrées sur la figure 6.6.

Cas d'un recouvrement bilatéral Les expressions obtenues dans le cas d'un recouvrement bilatéral est une généralisation du cas précédent, mais calculées avec les équations (6.29) et (6.30).

6.6.1.2 BCR adaptées à un modèle polynomial d'un signal long

Dans certains contextes, nous pouvons connaître *a priori* la nature de la modulation de l'amplitude et/ou de la fréquence. En particulier, il est facile de déterminer les BCR dans le cas où il s'agit de modulations polynomiales. Ces BCR ne sont qu'une généralisation des expressions (6.32), obtenues dans le chapitre précédent, à des signaux de longue durée.

Soit un signal long non-stationnaire mono-composante, $S[n]$, dont l'amplitude et la phase instantanées sont des fonctions polynomiales de degrés respectifs P et Q, qui s'écrivent :

$$\text{pour } n = 0, \dots, N-1 \quad A[n] = \sum_{p=0}^P \alpha_p n^p > 0 \quad (6.43)$$

$$\Phi[n] = \sum_{q=0}^Q \beta_q n^q \quad (6.44)$$

Les BCR correspondant à ce modèle global polynomial, notées $CRB_{\mathcal{P}}$, sont :

$$\begin{aligned} BCR_{\mathcal{P}}(A[n]) &= \frac{\sigma_b^2}{2} \mathbf{b}^\dagger (\mathbf{A}^\dagger \mathbf{A})^{-1} \mathbf{b}, \\ BCR_{\mathcal{P}}(\Phi[n]) &= \frac{\sigma_b^2}{2} \mathbf{b}^\dagger (\mathbf{B}^\dagger \mathbf{B})^{-1} \mathbf{b}, \quad \text{pour } 0 \leq n \leq N-1. \\ BCR_{\mathcal{P}}(F[n]) &= \frac{\sigma_b^2}{2} \mathbf{h}^\dagger (\mathbf{B}^\dagger \mathbf{B})^{-1} \mathbf{h}, \end{aligned} \quad (6.45)$$

avec

$$\begin{aligned}
 \mathbf{b} &= [1, n, n^2, \dots, n^P]^\top, & \mathbf{h} &= \frac{1}{2\pi} [0, 1, 2n, \dots, Qn^{(Q-1)}]^\top, \\
 \mathbf{A} &= [\mathbf{b}_0 \bullet e^{j\Phi}, \mathbf{b}_1 \bullet e^{j\Phi}, \dots, \mathbf{b}_P \bullet e^{j\Phi}], & \mathbf{B} &= j[\mathbf{b}_0 \bullet \mathbf{S}, \mathbf{b}_1 \bullet \mathbf{S}, \dots, \mathbf{b}_Q \bullet \mathbf{S}], \\
 \mathbf{b}_m &= [0^m, 1^m, 2^m, \dots, (N-1)^m]^\top, & e^{j\Phi} &= [e^{j\Phi[0]}, e^{j\Phi[1]}, \dots, e^{j\Phi[N-1]}]^\top, \\
 & & \mathbf{S} &= [S[0], S[1], \dots, S[N-1]]^\top.
 \end{aligned} \tag{6.46}$$

Nous insistons sur le fait que le calcul de ces bornes ($BCR_{\mathcal{P}}$) nécessite la connaissance *a priori* des degrés de modulations P et Q , sinon il faut les pré-estimer. Une sous ou sur estimation de ces derniers modifie ces BCR (cf. figure 6.6).

Dans tous les cas, ces $BCR_{\mathcal{P}}$ ne sont adaptées à notre contexte puisqu'elles sont, non seulement, déterminées pour des estimateurs non-biaisés mais aussi très optimistes. Ceci s'explique par le fait qu'elles exploitent toute l'information contenue dans le signal, vu que leurs expressions font intervenir la totalité des échantillons disponibles. A cela vient s'ajouter le fait qu'elles ne prennent pas en compte le fait que l'augmentation des degrés de l'approximation polynomiale favorise le risque d'amplifier l'erreur d'estimation. Toutefois, nous comparons, à titre indicatif, nos résultats à ces $BCR_{\mathcal{P}}$, classiquement employées dans la littérature.

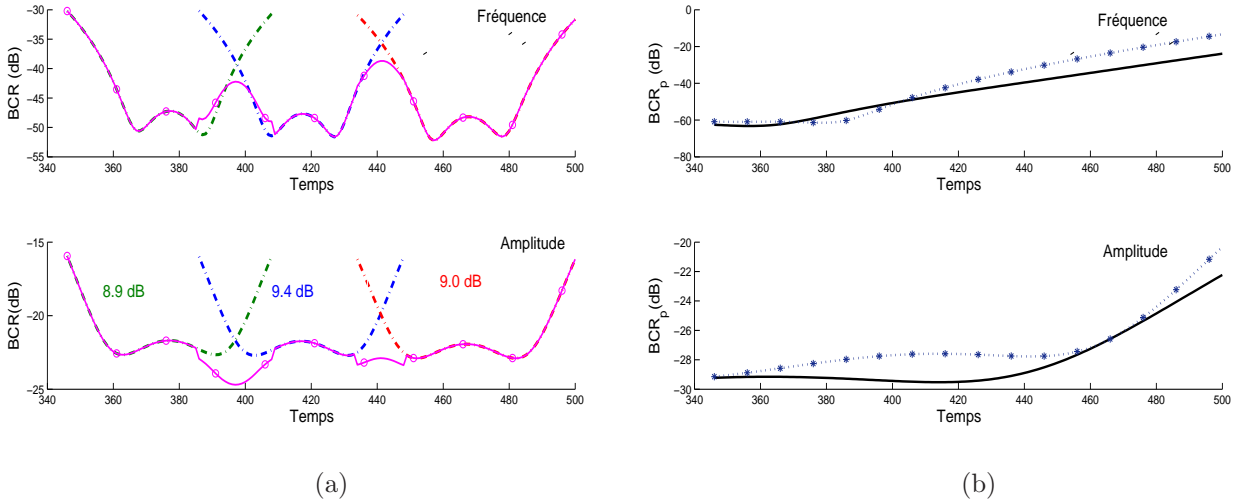


FIG. 6.6 – (a) BCR (6.38) et (6.42) obtenues après la fusion de trois segments qui se recouvrent. (b) Effet d'une sur-estimation de 1 du degré de l'approximation polynomiale de $F[n]$ et $A[n]$ sur les BCR polynomiales (6.45). Le signal utilisé est celui tracé sur la figure 6.1.

Nous allons nous pencher maintenant sur l'application de l'approche par morceaux, proposée dans ce chapitre, sur des signaux fortement non-stationnaires. Nous commençons par évaluer ses performances sur des signaux ayant une modulation de phase polynomiale et nous comparons les résultats obtenus avec la méthode du HAF. Dans une deuxième partie,

nous étudions des signaux ayant des modulations non-polynomiales et nous terminons ce chapitre par une application sur un signal réel terminera cette partie.

6.6.2 Estimation des signaux à phase et amplitude polynomiales

Nous considérons deux cas de signaux à phase et amplitude polynomiale : le premier est caractérisé par une modulation d'amplitude lentement variable (polynomiale de degré 2) et va nous permettre d'établir une comparaison avec la méthode de la HAF, tandis que le second présente un degré de modulation élevé pour l'amplitude comme pour la phase et pour lequel la HAF ne donne pas de résultats.

6.6.2.1 Modulation d'amplitude lentement variable

Considérons un signal (cf. figure 6.7) ayant une amplitude lentement variable, une phase polynomiale de degré 4, un indice de modulation de fréquence de 87%. La fréquence et l'amplitude instantanées de ce signal (cf. figure 6.7(a)) s'écrivent respectivement :

$$\begin{aligned} F[n] &= 0.2483 - 7.6474 \cdot 10^{-4}(n-250) + 1.6796 \cdot 10^{-7}(n-250)^2 + 1.2057 \cdot 10^{-8}(n-250)^3, \\ A[n] &= 0.9747 - 0.0058(n-250) + 3.9447 \cdot 10^{-5}(n-250)^2, \quad \text{pour } 0 \leq n < 500. \end{aligned} \quad (6.47)$$

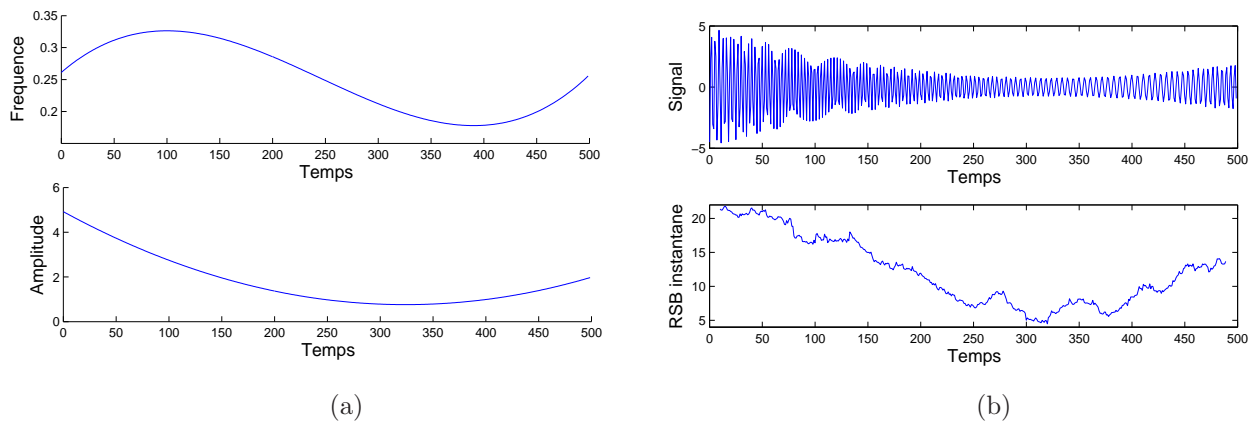


FIG. 6.7 – (a) $F[n]$ et $A[n]$ du signal défini par l'équation (6.47). (b) Signal temporel et RSB instantané en (dB).

Sur la figure 6.7(b) est indiqué le RSB instantané qui varie localement entre 20 et 5 dB pour un RSB moyen égal à 15 dB.

Résultats Nous reportons, sur la figure 6.8, les EQM et les biais d'estimation de $F[n]$ et de $A[n]$ obtenus avec l'algorithme proposé (cf. tableau 6.2) en comparaison avec ceux obtenus avec la méthode de la HAF. Les résultats sont issus de 100 simulations de bruit à un RSB moyen de 15 dB.

En analysant ces résultats, nous constatons ce qui suit :

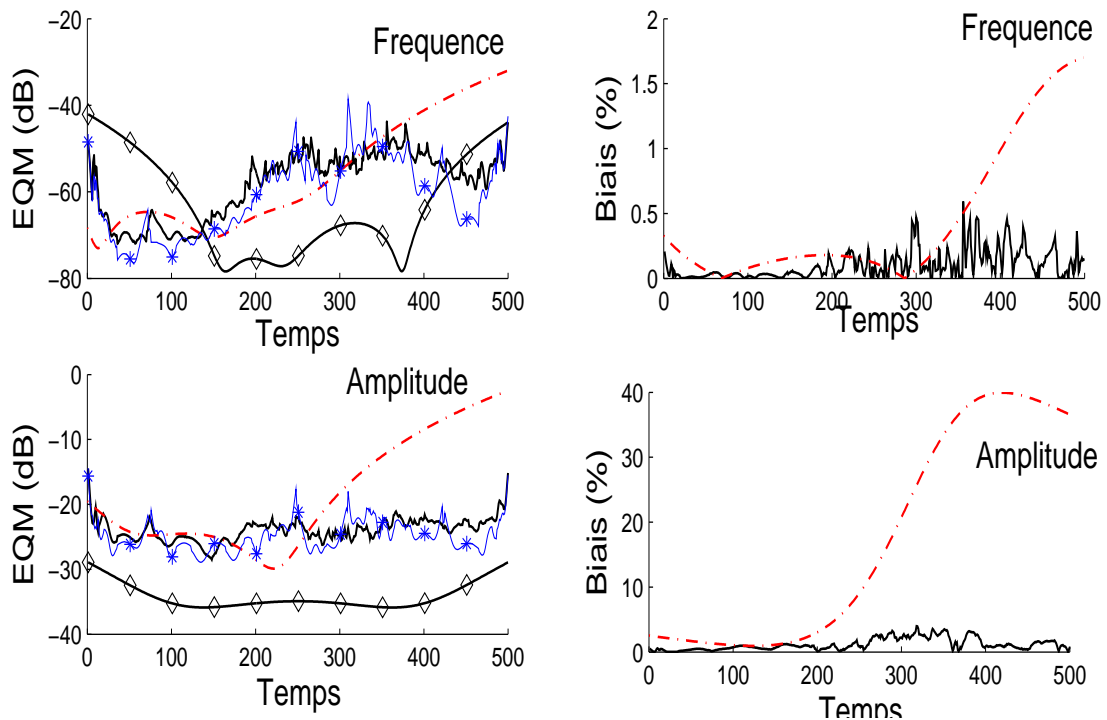


FIG. 6.8 – EQM et biais (%) de l'estimation de $A[n]$ et $F[n]$ du signal (6.47) obtenues, à un RSB moyen égal à 15 dB , en utilisant (-.-) HAF et (—) l'approche locale proposée, (-◇) BCR classiques (6.45) et les (-*) BCR adaptée (6.38) et (6.42).

- une différence des performances se manifeste au cours du temps. En effet, nous nous apercevons que les plus faibles EQM et biais sont relevés pour les temps $n = 0, \dots, 256$, cette observation est valable pour les deux méthodes comparées. Bien que ces deux dernières présentent des performances équivalentes sur cette tranche temporelle, l'écart entre les résultats se creuse sur la seconde tranche temporelle $n = 256, \dots, 500$ et peut atteindre les 20 dB . Pour la méthode de la HAF, l'EQM et le biais croient rapidement au fur et à mesure qu'on avance dans le temps, tandis qu'avec la méthode proposée, la variation de l'EQM et du biais reste modérée.

En consultant la figure 6.7(b), une explication plausible de cette différence de performance se dégage. Dans la première tranche temporelle $n = 0, \dots, 256$, le RSB instantané est relativement important $> 10\text{ dB}$ et sur la deuxième tranche $n = 256, \dots, 500$ il est beaucoup plus faible $< 10\text{ dB}$. Or, comme il était prévisible, les performances de la méthode se dégradent en présence d'un faible RSB. En revanche, celles de la méthode proposée est beaucoup plus robuste aux forts bruits.

- la courbe représentant l'EQM obtenue par la méthode proposée est très proche de celle de la BCR adaptée (6.38) et (6.42), illustrant ainsi la bonne performance de

cette approche. A titre indicatif, la comparaison avec la courbe de BCR classique (6.45), montre qu'on améliore l'estimation aux extrémités temporelles.

Afin de mieux consolider ces remarques, nous avons comparé également l'EQM des paramètres α_1 et β_0 (6.43) de la modulation polynomiale de l'amplitude et de la phase. Ces paramètres sont les moins bien estimés par la méthode de la HAF à cause du phénomène de propagation d'erreurs. Les résultats sont affichés sur la figure 6.9(a) et comparés à une BCR classique correspondant à la diagonale de la matrice

$$\frac{\sigma_b^2}{2} \begin{pmatrix} \mathbf{A}^\dagger \mathbf{A} & 0 \\ 0 & \mathbf{B}^\dagger \mathbf{B} \end{pmatrix}^{-1}.$$

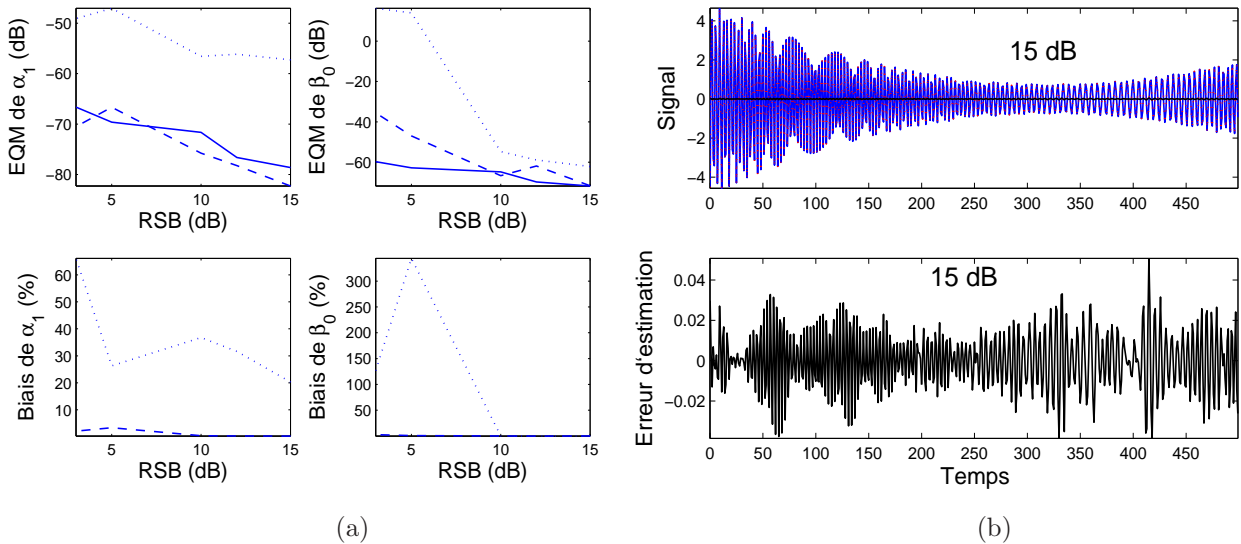


FIG. 6.9 – (a) EQM et biais de l'estimation des paramètres α_1 de $A[n]$ et β_0 de $\Phi[n]$ (6.43) du signal défini par (6.47) en fonction du RSB : (...) HAF, (- -) l'approche proposée et (—) les BCR appropriées. (b) Reconstruction du signal pour un RSB=15 dB : (—) Signal original, (- -) Signal estimé, (—) Erreur d'estimation pour un RSB de 15 dB. Les deux figures ne sont pas à la même échelle.

Là-dessus, il est évident que l'approche proposée, tout en étant proche de la BCR appropriée, apporte une amélioration de plus de 20 dB par rapport à la HAF. Le biais d'estimation est aussi nettement plus faible.

6.6.2.2 Modulation rapide de l'amplitude

Reprenons l'exemple du signal déjà présenté sur la figure 6.1, où $A[n]$ et $F[n]$ sont des polynômes de degrés respectifs 3 et 7, *a fortiori* la phase est donc un polynôme de degré 8. En fait, les coefficients de ces polynômes ont été générés aléatoirement. Nous appliquons, à ce signal, l'approche proposée. Les résultats sont reportés sur la figure 6.11.

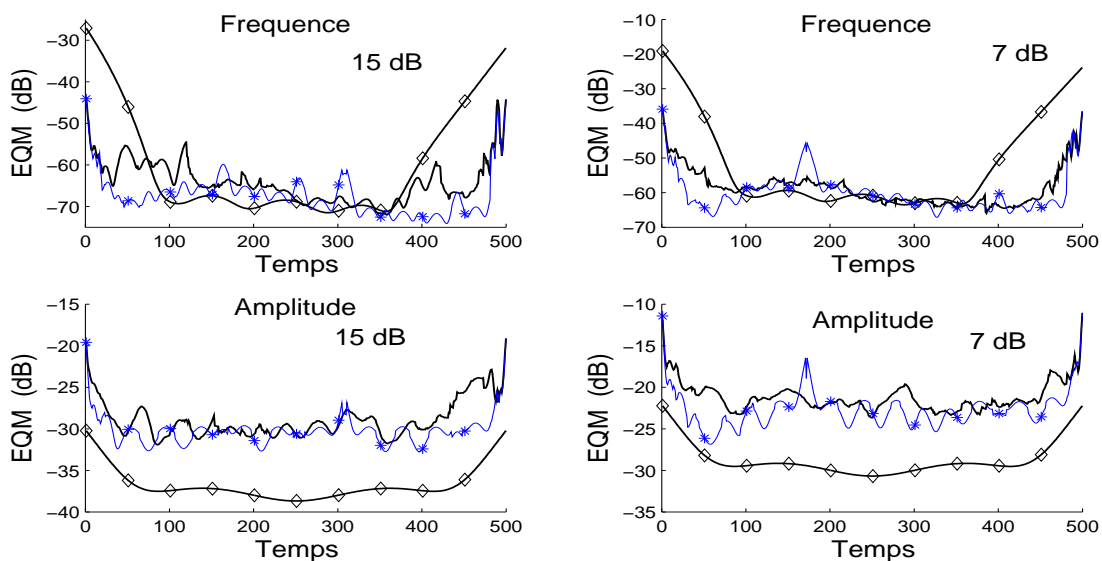


FIG. 6.10 – EQM obtenues dans l'estimation de $A[n]$ (polynôme de degré 3) et $F[n]$ (polynôme de degré 7) du signal représenté sur la figure 6.11 pour un RSB moyen de 15 dB (colonne gauche) et 7 dB (colonne droite) : (—) l'approche locale proposée, (\diamond) BCR classiques (6.45) et ($*$) BCR adaptées (6.38) et (6.42).

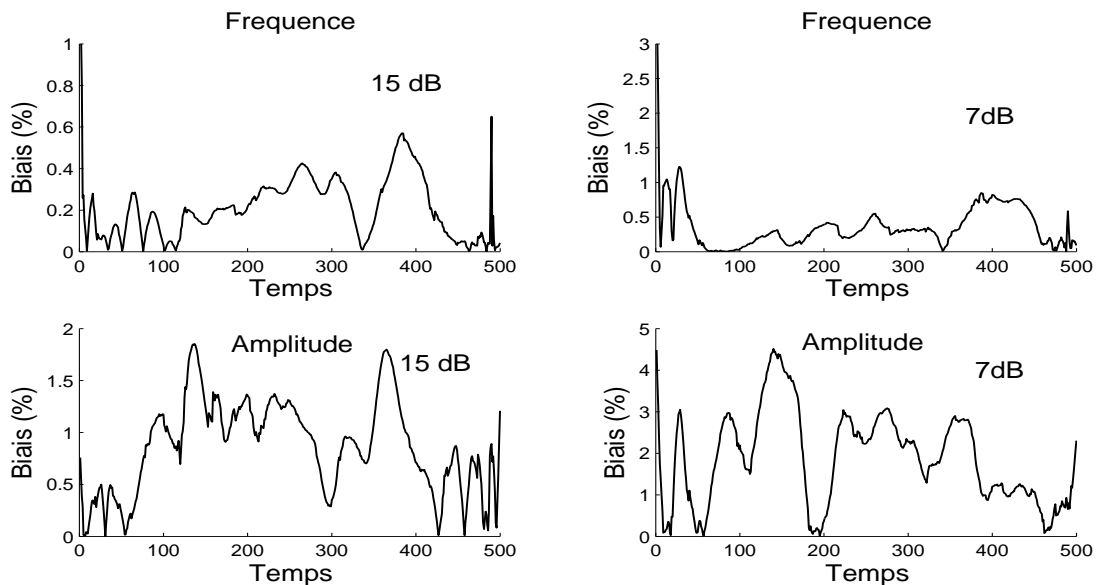


FIG. 6.11 – Biais obtenus dans l'estimation de $A[n]$ (polynôme de degré 3) et $F[n]$ (polynôme de degré 7) du signal représenté sur la figure 6.11 pour un RSB global de 15 dB (colonne gauche) et 7 dB (colonne droite).

Résultats Les EQM et les biais sont tracés respectivement sur les figures 6.10 et 6.11. Ces résultats sont établis à partir de 100 réalisations du bruit pour des RSB moyens de 15 et de 7 *dB*.

En ce qui concerne la fréquence instantanée, l'EQM est très proche de la courbe de BCR adaptées. Elle l'est également par rapport aux BCRS classiques sur la tranche temporelle $n = 100, \dots, 400$ et s'en écarte pour les temps au début et à la fin. En revanche pour l'amplitude instantanée, une différence de 10 *dB* existe entre les deux mais l'EQM reste collées à la BCR adaptée, ce qui témoigne de la robustesse de l'approche. Le résultat des courbes reconstruites de $A[n]$, $F[n]$ et $S[n]$ ont été déjà présenté sur la figure 6.5.

Rappelons que l'estimation de ce signal par la méthode de la HAF ne fonctionne pas et elle était présentée sur la figure 1.6. Sur cette figure, il apparaît que les paramètres β_1 , β_2 et β_3 (notés sur cette figure respectivement Φ_1 , Φ_2 et Φ_3) ne peuvent pas être estimés correctement, conséquence immédiate du phénomène de propagation de l'erreur d'estimation connu avec cette méthode.

6.6.3 Estimation des signaux ayant un modèle sinusoidal

Afin de tester l'algorithme sur un signal ayant une modulation non-polynomiale, l'amplitude, la fréquence et la phase instantanées ne sont pas polynomiales, nous l'appliquons au signal décrit par :

$$\begin{aligned} F[n] &= 0.1011 + 0.0993 \cos(2\pi \cdot 0.3840 n) \\ A[n] &= 0.1485 + 0.0294 \sin(2\pi \cdot 0.2786 n), \text{ pour } 0 \leq n \leq 500. \end{aligned} \quad (6.48)$$

Résultats Sur la figure 6.13, nous pouvons voir les courbes de la modulation de l'amplitude et de la fréquence reconstruites pour un RSB global égal à 15 et à 7 *dB*, tandis que sur la figure 6.12 nous avons tracé l'EQM obtenues dans l'estimation de ces modulations. Les BCR classiques sont calculés en supposant une approximation polynomiale de la phase et de l'amplitude instantanées avec un degré égal à 7. Un total de 22 segments de longueur en moyenne égale à 30 échantillons, ces segments ont été estimés puis fusionnés pour reconstruire le signal dans l'intégralité de sa durée.

Les courbes estimées de la fréquence et de l'amplitude instantanées sont collées aux courbes originales pour 15 et 7 *dB*. Le signal est également bien reconstruit. Sur la figure 6.12, les courbes représentant l'EQM de la fréquence et de l'amplitude instantanées pour 15 *dB* sont très proches des BCR adaptées (6.25) et (6.40). Pour 7 *dB*, on obtient un écart de 10 *dB* entre l'EQM de l'amplitude et la BCR adaptée.

Exemple de signal où l'analyse par RTF ne fonctionne pas Nous présentons ici un autre signal ayant une modulation sinusoidale en amplitude et en fréquence de 1024 échantillons, pour lequel l'analyse par RTF ne fonctionne pas et que la technique proposée

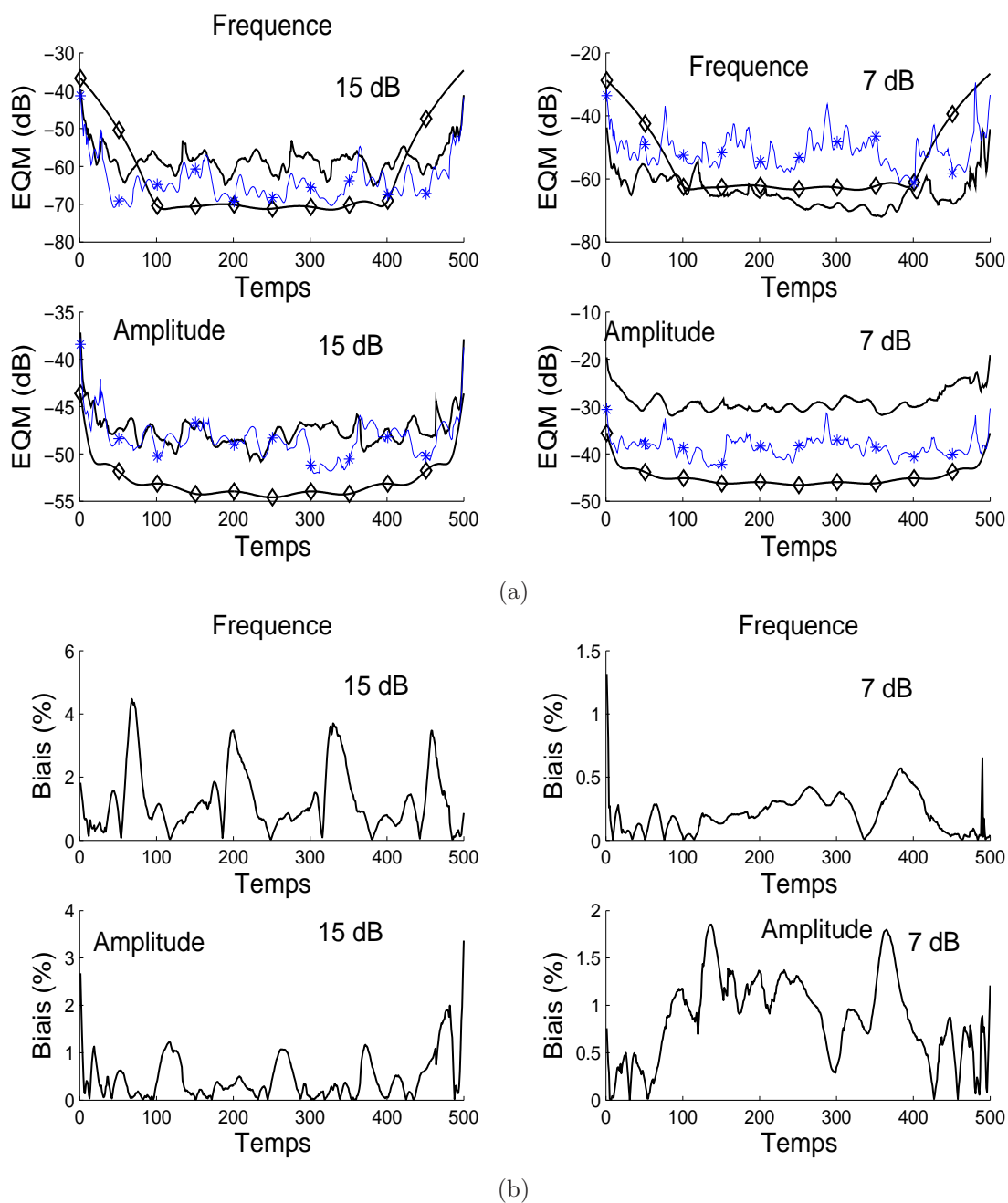


FIG. 6.12 – (a) EQM (—) obtenues par l'estimation du signal modulé sinusoidalement en amplitude et en fréquence. Les $- \diamond$ BCR classiques (6.45) sont calculés pour un RSB moyen égal à 15 dB (colonne gauche) et un RSB moyen égal à 7 dB (colonne droite) en supposant une approximation polynomiale de degré 7 pour la modulation de l'amplitude et pour celle de la fréquence, les $(-*)$ BCRs adaptées (6.38) et (6.42). (b) Biases de l'estimation.

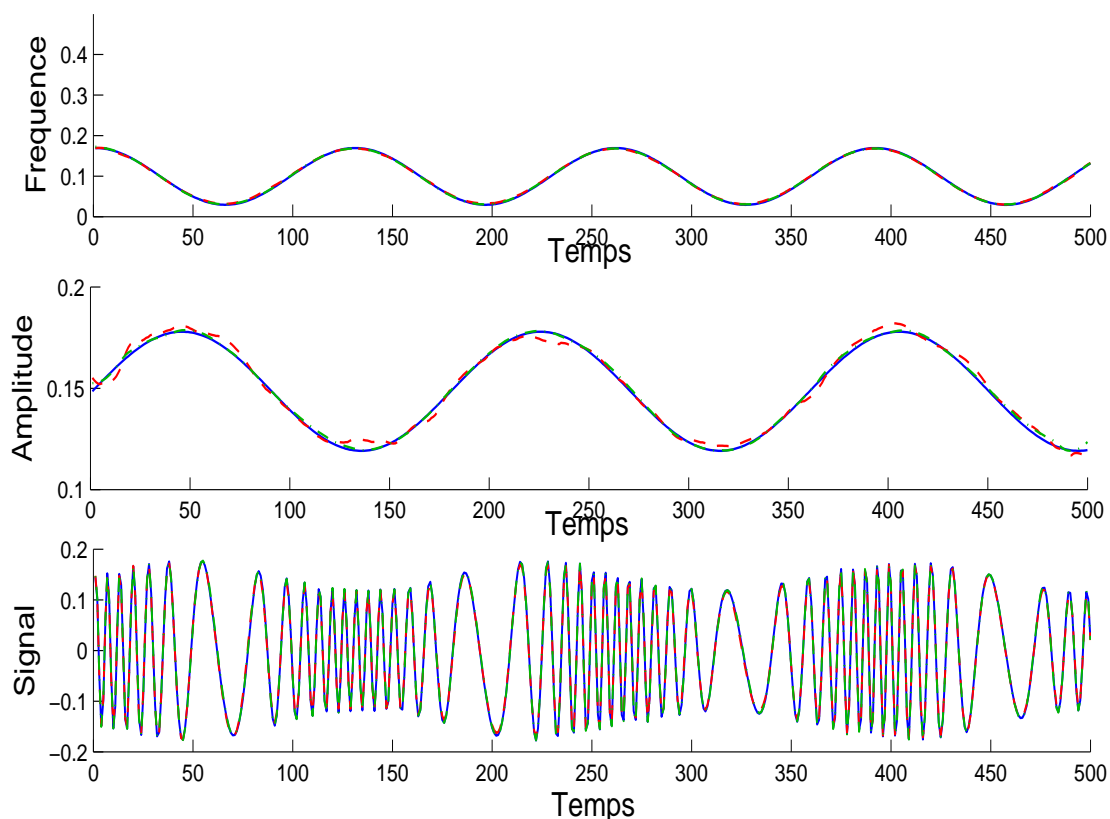


FIG. 6.13 – Les courbes originales (—) de la modulation sinusoidale d’amplitude et de fréquence superposées à leurs estimées par l’approche proposée : (-.-) obtenues à 15 dB et (- -) celles obtenues à 7 dB.

donne de bons résultats. Le RSB moyen est de 15 dB et la fréquence d’échantillonnage est $F_e = 256$. L’amplitude instantanée et la fréquence instantanée sont définies pour $n \in [0 : 1023]$ par :

$$A[n] = 10 + 3 \sin\left(8\pi \frac{n}{F_e}\right), \quad \text{et} \quad F[n] = 50 + 5 \sin\left(16\pi \frac{n}{F_e}\right). \quad (6.49)$$

Sur la figure 6.14(b), nous avons reporté quelques représentations temps fréquence pour le signal défini par (6.49) : le spectrogramme, la distribution de Wigner-Ville et de Wigner-Ville lissée (cf. chapitre 1). Sur cette figure, on voit que l’analyse par ces RTFs est limitée et n’arrive pas à trouver la modulation de fréquence qui correspond à ce signal est non-stationnaire. En revanche, sur la figure 6.14(a) sont représentés les résultats obtenus par l’approche proposée. Les courbes estimées de l’amplitude et de la fréquence sont collées aux courbes originales.

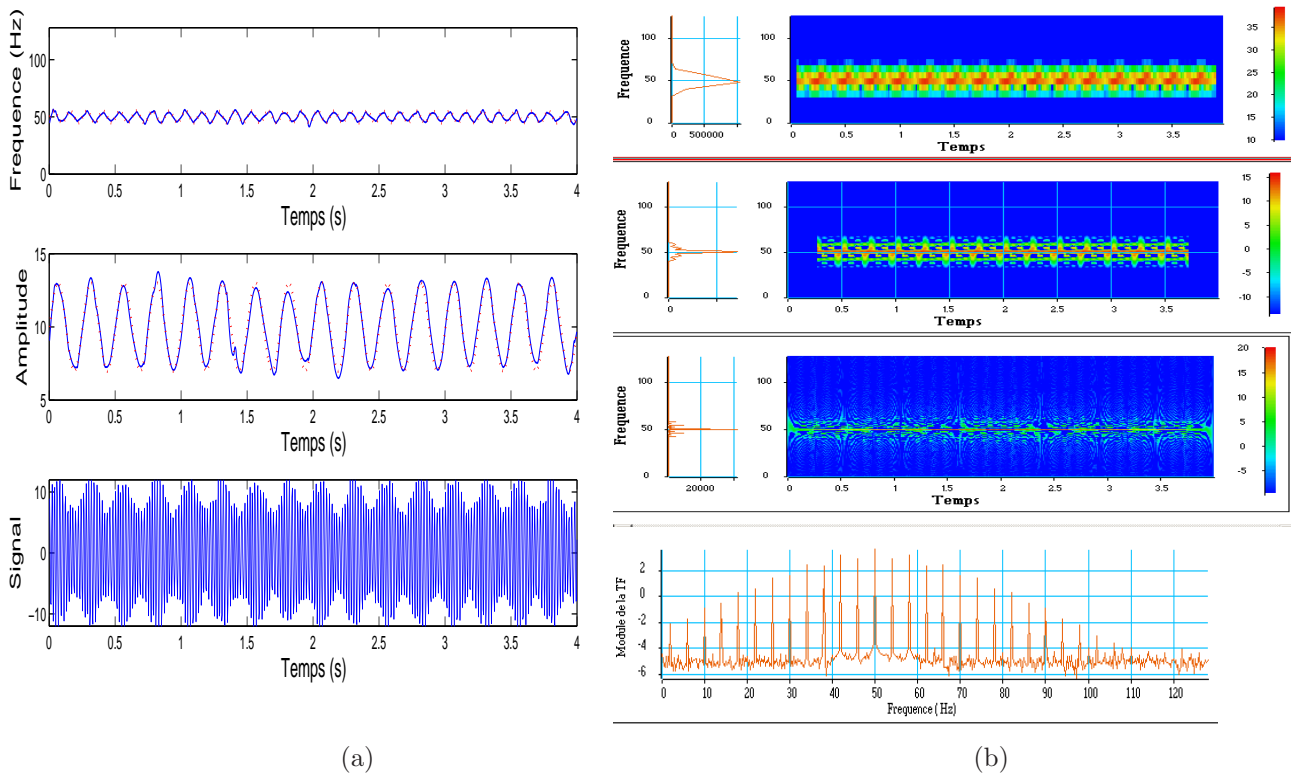


FIG. 6.14 – (a) (–) Les courbes estimées par la méthode proposée superposées avec celles originales obtenue par l’estimation du signal modulé sinusoidalement en amplitude et en fréquence (6.49). (b) De haut en bas : le spectrogramme avec une fenêtre d’analyse de 32 points, la distribution de Wigner-Ville, la distribution de Wigner-Ville lissée [Mar07].

6.6.4 Application sur un signal réel

Dans plusieurs projets de recherche acoustique [Gre68, CS95, SFD98, Ele04, ML05], les scientifiques cherchent à déterminer le mécanisme de production du son et notamment celui des chants d’oiseaux dans le but de constituer un modèle mécanique pouvant générer de tels signaux.

Les chants d’oiseaux sont essentiellement générés par un écoulement d’air dans ce qui est appelé un sac d’air. Le rôle joué par ce dernier est très important dans la vocalisation. Il exerce une différence de pression, contrôlée par cet écoulement d’air, sur les structures des organes vocaux provoquant ainsi des oscillations au niveau du syrinx. Le syrinx est un organe propre aux oiseaux chanteurs et admet des propriétés mécaniques non-linéaires. Grâce aux contractions extrêmement rapides de ses muscles, le syrinx va moduler l’énergie et la fréquence du chant produit (signal) [SFD98, Ele04, ML05]. Dès lors, l’analyse de ces chants à l’aide des techniques temps-fréquence prend tout son intérêt et permet une meilleure identification des prototypes générateurs et des modulateurs utilisés dans le mécanisme de la production du son. L’outil le plus utilisé par les biologistes est

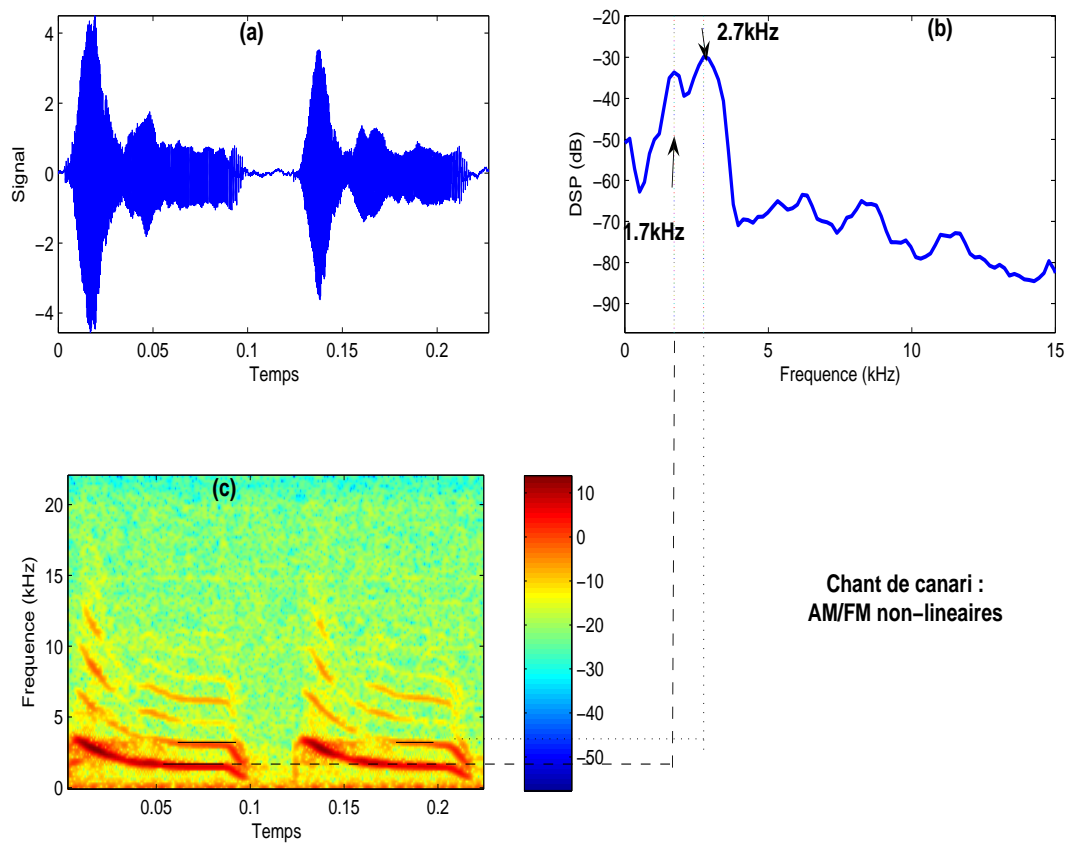


FIG. 6.15 – Chant d’un canari fortement non-stationnaire : (a) Signal temporel, (b) Densité spectrale de puissance et (c) Spectrogramme calculé avec une fenêtre de Hamming glissante de 256 points, $N_{FT} = 256$ points et un recouvrement de 99%.

le spectrogramme et récemment le spectrogramme alloué. Or, ces techniques sont limitées au niveau de la résolution, c'est pourquoi nous nous proposons d'appliquer l'approche par morceaux présentée dans ce chapitre.

Considérons le chant de canari de la figure 1(b) et dont le spectrogramme est reproduit sur la figure 6.15. La fréquence d'échantillonnage est $F_e = 44$ kHz, le RSB est relativement élevé et le nombre total d'échantillons disponibles est 10000.

Sur ce spectrogramme, nous pouvons constater que ce chant est un signal multi-composantes modulées non-linéairement. Les fréquences instantanées de ces multiples composantes semblent décroître exponentiellement au cours du temps et tendent à se superposer dans les intervalles $[0.09s, 0.11s]$ et $[0.20s, 0.22s]$. L'indice de modulation tel que défini (cf. chapitre 5) est supérieur à 4 kHz. Sur la figure 6.15(b), est tracée la densité spectrale de puissance sur laquelle, nous avons repéré la présence des deux harmoniques les plus énergétiques, ce sont les fréquences fondamentales 1.7 kHz et 2.7 kHz correspondant aux filtres adaptés de la trachée et du bec.

Bien que le signal est multi-composantes, nous allons lui appliquer l'approche par morceaux proposée pour estimer uniquement la composante la plus énergétique. Etant donné que le nombre d'échantillons total disponible est important, égal à 10000, nous comptabilisons 300 segments courts traités et fusionnés pour aboutir à l'estimation et la reconstruction globale de la composante la plus énergétique.

Résultats Le résultat de l'estimation de la composante en question et le spectrogramme du signal résidu sont représentés sur les figures 6.17 et 6.17. En utilisant la segmentation non-séquentielle et la stratégie de la fusion pour extraire la composante la plus énergétique du chant de canari, nous avons amélioré la qualité de l'estimation, par comparaison à l'analyse que fournit le spectrogramme. En réappliquant la stratégie proposée de nouveau au signal résidu, nous obtenons l'estimation des composantes suivantes.

6.7 Conclusion - Avantages et limitations

Dans ce chapitre, nous avons généralisé l'approche basée sur la modélisation polynomiale des signaux courts non-stationnaires aux signaux longs non-stationnaires et mono-composantes. La nouvelle approche, dite approche par morceaux, a consisté à segmenter le signal long en des segments courts sur lesquels il est possible de modéliser localement chacune de l'amplitude et de la fréquence instantanée par un polynôme. Les paramètres des modèles locaux sont estimés en maximisant la fonction de vraisemblance par la technique du recuit simulé.

Afin de reconstruire le signal dans sa totalité, les segments estimés sont fusionnés. La stratégie de fusion s'applique sur l'amplitude et la phase en assurant leur continuité.

Cette approche a fourni des bons résultats quand elle a été appliquée aux signaux for-

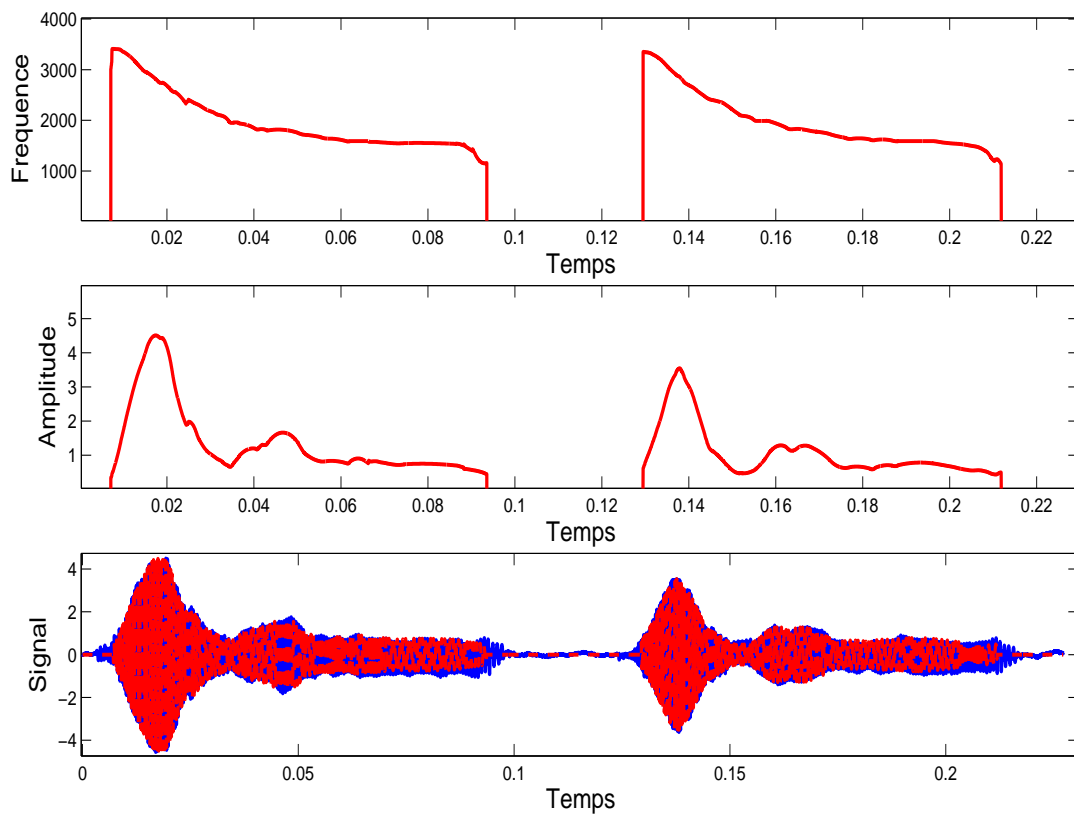


FIG. 6.16 – Chant d’un canari fortement non-stationnaire : estimation de la fréquence et de l’amplitude instantanées de la composante la plus énergétique du signal et reconstruction de cette dernière.

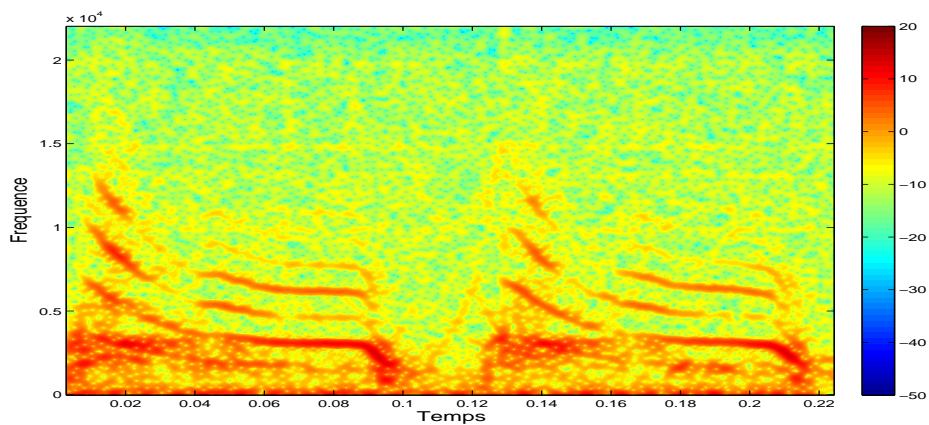


FIG. 6.17 – Signal résidu obtenu après l’estimation de la composante la plus énergétique du signal chant de canari : spectrogramme calculé avec une fenêtre de Hamming glissante de 256 points, $N_{FT} = 256$ et un recouvrement de 99%.

tement non-stationnaires simulés et aux signaux réels. Une amélioration de la précision par rapport à la méthode existante de la HAF, se traduisant par un gain de 10 dB au minimum au niveau de l'estimation de l'amplitude et 20 dB pour celle de la fréquence, a été obtenu grâce à l'approche proposée. Les résultats obtenus étaient également comparés aux bornes de Cramer-Rao.

Cependant, cette approche peut présenter des limitations dans certains cas où la modulation de fréquence devient extrêmement rapide (par exemple des impulsions ultra-rapides). Ce type de modulation nécessite un degré d'approximation polynomial local très élevé supérieur à 3 pour un nombre d'échantillons très petit, nous avons dans ce cas plutôt intérêt à changer de modèle local qu'à augmenter le degré de l'approximation. Hormis ce type de cas, l'approche offre un grand intérêt par rapport aux méthodes existantes, vu qu'on s'est affranchi d'un modèle polynomial global ayant un degré élevé pour approcher le signal entier. Rappelons surtout le fait qu'un degré très élevé implique une limitation sur les coefficients des polynômes utilisés pour pouvoir respecter le théorème de Shannon. Ses coefficients prennent donc des valeurs de plus en plus faibles surtout s'ils sont exprimés dans la base canonique, la seule base utilisable avec la méthode de la HAF. L'approche que nous avons proposée constitue une généralisation et un compromis d'une part entre modélisation polynomiale locale et l'absence d'un modèle polynomial global, et d'autre part une imposition d'un degré d'approximation locale et l'absence d'un degré global. Le travail présenté dans ce chapitre a été publié dans [JVLM05, JMVL07b]

Dans les chapitres suivants, nous allons étendre l'étude aux signaux multi-composantes.

Chapitre 7

Signal court à multi-composantes non-stationnaires

Résumé

L'approche proposée dans l'estimation des signaux courts mono-composantes dans le chapitre 3 est généralisée aux signaux courts multi-composantes sous formes de deux méthodes. Quoique ces deux dernières s'appuient sur le même modèle polynomial de l'amplitude et de la fréquence instantanées de chacune des composantes, elles procèdent différemment pour l'estimation des paramètres de ce modèle.

Dans la première méthode, dite optimale, les paramètres de toutes les composantes sont estimés tous en même temps et cela en maximisant la fonction de vraisemblance *via* la technique du recuit simulé. Quant à la deuxième méthode, dite sous-optimale, uniquement les paramètres d'une seule et même composante sont estimés à la fois. Cette composante est ensuite éliminée du signal étudié avant d'entamer l'estimation de la composante suivante. Ainsi, le signal est reconstruit composante par composante, d'une manière itérative.

Ces deux méthodes proposées sont comparées et de très bonnes performances sont obtenues dans le cas de signaux présentant des croisements de fréquences et ce en dépit du faible nombre d'échantillons (signaux courts).

Mots clefs

Signaux courts multi-composantes non-stationnaires, modélisation polynomiale, fonction de vraisemblance, recuit simulé, bornes de Cramer-Rao, croisement de composantes fréquences.

Sommaire

7.1	Introduction	165
7.2	Modélisation et estimation	165
7.2.1	Modélisation	166
7.2.2	Méthodes d'estimation	167
7.2.2.1	Méthode optimale - Tous les paramètres estimés à la fois	167
7.2.2.2	Méthode sous-optimale - Estimation composante par composante	169
7.3	Etude des performances des algorithmes proposés	170
7.3.1	Bornes de Cramer Rao	170
7.3.2	Application sur des signaux de simulations	171
7.3.2.1	Composantes fréquentielles très proches	171
7.3.2.2	Croisement de composantes fréquentielles	177
7.3.3	Application sur des signaux réels	177
7.3.3.1	Signal de baleine	179
7.3.3.2	Bilan de masses du Glacier de Sarennes	179
7.4	Conclusion	179

7.1 Introduction

Outre leur sensibilité aux bruits et aux fortes modulations, les techniques d'estimation des signaux multi-composantes non-stationnaires connues dans la littérature, telles que HAF et P-HAF (cf. chapitre 1), sont sujets aux apparitions d'interférences. Ces dernières sont dues à la présence de composantes multiples et favorisent la dégradation de l'estimation dans le contexte de signaux courts non-stationnaires.

Motivés par les bonnes performances de l'approche proposée dans les chapitres 3 et 5 et sa robustesse en présence de bruit et de modulations non-linéaires, nous proposons de la généraliser aux signaux courts multi-composantes.

Considérons un modèle du signal où la fréquence et l'amplitude instantanées de chacune des composantes soient approximées par un polynôme de degré faible. Nous proposons d'estimer les paramètres de ce modèle par deux méthodes différentes, ayant chacune ses inconvénients et ses avantages. La première méthode, bien que coûteuse en temps de calcul, est optimale car réalise l'estimation de tous les paramètres à la fois et ce en maximisant la fonction de vraisemblance *via* le recuit simulé. La deuxième méthode est, en revanche, sous-optimale puisqu'elle consiste à estimer d'une manière itérative le signal, composante par composante. Lors d'une itération sont estimés uniquement les paramètres relatifs à une seule composante. Cette dernière est ensuite reconstruite pour être éliminée du signal étudié avant de procéder à l'estimation de la composante suivante.

Ce chapitre est orienté comme suit. Dans un premier temps, nous décrivons le modèle adopté pour un signal court multi-composantes non-stationnaires et nous exposons les deux méthodes utilisées pour l'estimation des paramètres. Dans un deuxième temps, les bornes de Cramer-Rao sont déterminées, puis des simulations de Monte Carlo illustrent les bonnes performances de chacune de ces méthodes en présence du bruit. Nous soulignerons, en particulier, leur aptitude à gérer certains cas délicats tels que la présence de composantes très proches ou se croisant dans le plan temps fréquence. Enfin, les deux approches seront appliquées sur un signal réel.

7.2 Modélisation et estimation

Considérons un signal multi-composantes court, noté $s[n]$, noyé dans du bruit blanc gaussien additif $b[n]$, centré et de variance σ_b^2 , défini comme suit :

$$\begin{aligned} y[n] &= s[n] + b[n], \\ s[n] &= \sum_{i=1}^K s_i[n], \\ s_i[n] &= A_i[n] \exp(j \Phi_i[n]) \end{aligned} \quad \text{pour } \frac{-N}{2} \leq n \leq \frac{N}{2}, \quad (7.1)$$

où $y[n]$ est le signal bruité et $N + 1$ est le nombre d'échantillons disponible considéré impair pour simplifier le calcul. $s[n]$ est formé par la somme de K composantes ($K \geq 2$) dont la i -ème composante est notée $s_i[n]$. L'amplitude et la phase instantanées de cette dernière

sont désignées respectivement par $A_i[n]$ et $\Phi_i[n]$. Quant à sa fréquence instantanée, elle est reliée à $\Phi_i[n]$ par :

$$\Phi_i[n] = \varphi_{i,0} + 2\pi \left(\sum_{k=-\frac{N}{2}}^n F_i[k] - \sum_{k=-\frac{N}{2}}^0 F_i[k] \right), \quad (7.2)$$

où $\varphi_{i,0} = \Phi_i[0] \in [-\pi, \pi[$ représente la phase initiale référencée au centre du segment temporel.

Afin d'éviter le problème dû à l'ambiguïté dans la définition de la phase instantanée (cf. paragraphe 2.2.1) et garantir l'unicité du modèle, nous considérons les hypothèses supposées dans le chapitre 2 valables pour chaque composante, à savoir :

- $\Phi_i[n]$ est continue *i.e.* l'absence de sauts de phase,
- $A_i[n]$ est continue et strictement positive,
- $F_i[n]$ vérifie la relation : $0 < F_i[n] < \frac{F_c}{2}$ quel que soit $n \in [-\frac{N}{2}, \frac{N}{2}]$ et quel que soit $i = 1, \dots, K$.

Nous rappelons dans la suite le modèle polynomial qui sera appliqué à chacune des $A_i[n]$ et $F_i[n]$. $\Phi_i[n]$ sera directement déduite *via* l'équation (7.2).

7.2.1 Modélisation

Considérons un signal $s[n]$ non-stationnaire, défini par (7.1) et disponible sur une courte durée, c'est-à-dire N est faible pouvant prendre une valeur entre la quinzaine et la soixante d'échantillons. Supposons que, pour chaque composante i , la non-linéarité de $F_i[n]$ et de celle de $A_i[n]$ puissent être approximées, sur cette durée, par des polynômes de degrés faibles (inférieurs ou égales à 3).

Soit $\{g_k[n]\}$ une base polynomiale définie pour $n \in [-\frac{N}{2}, \frac{N}{2}]$, où $g_k[n]$ est un polynôme de degré k . Les décompositions sur cette base de la fréquence et de l'amplitude instantanées de la i ème composante s'écrivent respectivement :

$$\begin{aligned} F_i[n] &= \sum_{k=0}^{M_{F_i}} f_{i,k} g_k[n], \\ A_i[n] &= \sum_{k=0}^{M_{A_i}} a_{i,k} g_k[n], \end{aligned} \quad (7.3)$$

où M_{F_i} et M_{A_i} sont leurs degrés respectifs, vérifiant : $0 \leq M_{F_i} \leq 3$ et $0 \leq M_{A_i} \leq 3$.

Nous devons donc estimer $(M_{F_i} + M_{A_i} + 3)$ paramètres pour chaque composante :

- $M_{F_i} + 1$ paramètres de la fréquence : $\{f_{i,0}, f_{i,1}, f_{i,M_{F_i}}\}$,
- $M_{A_i} + 1$ paramètres de l'amplitude $\{a_{i,0}, a_{i,1}, a_{i,M_{A_i}}\}$
- et une phase initiale $\varphi_{i,0}$ (7.2).

Au total, $3K + \sum_{i=1}^K (M_{F_i} + M_{A_i})$ paramètres doivent être déterminés pour caractériser les K composantes et reconstruire le signal $s[n]$.

Afin de simplifier les notations, nous récrivons les paramètres à estimer sous forme vectorielle :

$$\boldsymbol{\theta} = [\boldsymbol{\theta}_1^\top, \dots, \boldsymbol{\theta}_K^\top]^\top, \quad (7.4)$$

$$\text{avec } \boldsymbol{\theta}_i = [\varphi_{i,0}, a_{i,0}, \dots, a_{i,M_{A_i}}, f_{i,0}, \dots, f_{i,M_{F_i}}]^\top. \quad (7.5)$$

Dans le paragraphe suivant, nous proposons et détaillons deux méthodes d'estimation du vecteur $\boldsymbol{\theta}$. Pour des raisons de clarté de raisonnement, le choix de la base polynomiale sera évoqué dans le paragraphe 7.3.1 après la détermination des bornes d'estimation.

7.2.2 Méthodes d'estimation

Nous définissons ici deux méthodes d'estimation découlant directement de celle proposée dans le contexte des signaux mono-composantes (cf. chapitre 3). Leurs avantages et leurs inconvénients seront explicités.

7.2.2.1 Méthode optimale - Tous les paramètres estimés à la fois

Principe Afin de conserver l'optimalité au sens du maximum de vraisemblance, l'estimation de tous les paramètres (exactement $3K + \sum_{i=1}^K (M_{F_i} + M_{A_i})$ paramètres inconnus) est réalisée simultanément en maximisant la fonction du vraisemblance. Etant donnée l'hypothèse du bruit Gaussien, ceci est équivalent à la minimisation de l'erreur des moindres carrés et revient finalement à écrire :

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \ell_{MC}(\boldsymbol{\theta}) \quad (7.6)$$

avec

$$\ell_{MC}(\boldsymbol{\theta}) = \sum_{n=-\frac{N}{2}}^{\frac{N}{2}} |y[n] - s[n]| = \sum_{n=-\frac{N}{2}}^{\frac{N}{2}} \left| y[n] - \sum_{i=1}^K A_i[n] \exp(j \Phi_i[n]) \right| \quad (7.7)$$

Notons que cette fonction $\ell_{MC}(\boldsymbol{\theta})$, fortement non-linéaire par rapport à $\boldsymbol{\theta}$, présente plusieurs minima locaux et qu'une minimisation utilisant les techniques classiques ne garantit pas la convergence vers le minimum global (cf. figure 3.1). Pour cela, nous allons utiliser la technique d'optimisation stochastique basée sur le recuit simulé. L'algorithme utilisé dans l'estimation des paramètres est quasi-identique à celui présenté dans le Tableau 3.2 dans le chapitre 3, sauf que le nombre de paramètres est augmenté. Nous rappelons brièvement les étapes clés de cet algorithme :

- initialisation des paramètres,
- génération aléatoire de candidats $\boldsymbol{\theta}_c$ susceptibles de diminuer la fonction $\ell_{MC}(\boldsymbol{\theta})$ (7.6),
- acceptation ou refus du candidat généré et régulation de la décroissance de la température et du pas du voisinage du recuit simulé (cf. chapitre 3).

Algorithme Dans la pratique pour pouvoir mettre en oeuvre cette méthode, certaines connaissances *a priori* sont requises : le nombre de composantes K , les degrés d'approximation M_{A_i} et M_{F_i} de chaque composante.

Pour simplifier l'algorithme, soit nous fixons le nombre de composantes K , soit on le pré-estime *via* des techniques classiques [FF95]. Nous attribuons, aussi, à chacun des degrés M_{A_i} et M_{F_i} une valeur fixe M choisie inférieure ou égale à 3 :

$$\forall i = 1, \dots, K, \quad M_{A_i} = M_{F_i} = M. \quad (7.8)$$

Dès lors, les étapes de l'algorithme se résument à :

1. l'initialisation de l'algorithme peut se faire :
 - soit d'une façon aléatoire *i.e.* en affectant aux paramètres des valeurs prises au hasard mais respectant les contraintes imposées sur l'amplitude et la fréquence instantanées (cf. paragraphe 7.2),
 - soit d'une manière itérative qui consiste à initialiser les paramètres composante par composante en utilisant la transformée de Fourier du signal résidu selon le schéma indiqué dans le tableau 7.1.

TAB. 7.1 – Méthode optimale - Initialisation des paramètres.

-
- Initialiser $i = 1$ et $y_i[n] = y[n]$.
 - Répéter tant que $i \leq K$,
 - (a) initialiser $\theta_i = \theta_i^{(0)}$ en utilisant la transformée de Fourier (ou le spectrogramme) du signal $y_i[n]$ (cf. équations (3.13) et (3.17)),
 - (b) calculer une approximation initiale de $s_i[n]$, notée $\hat{s}_i[n] = A_i \hat{\Phi}_i[n] \exp j \hat{\Phi}_i[n]$, avec $A_i \hat{\Phi}_i[n]$ et $\hat{\Phi}_i[n]$ obtenues en remplaçant dans les équations (7.1), (7.2) et (7.3) les coefficients par ceux de $\theta_i^{(0)}$.
 - (c) évaluer $y_{i+1}[n] = y_i[n] - \hat{s}_i[n]$, puis incrémenter $i = i + 1$.
-

2. La génération des candidats se fait selon une loi Gaussienne de matrice de covariance diagonale et dont les coefficients non nuls sont déterminés par un vecteur σ^2 appelé rayon du voisinage (cf. paragraphe 3.4.6). Les coefficients de ce vecteur et la température du recuit simulé sont diminués de la même manière que celle décrite dans les paragraphes 3.4.6 et 3.4.8.
3. L'arrêt de l'algorithme est contrôlé par le même critère que celui décrit dans le paragraphe 3.4.7, c'est-à-dire nous vérifions si la puissance du signal résidu normalisée par celle du bruit est une variable de χ^2 de $2(N + 1)$ degrés de liberté.

Avantages et inconvénients Outre la nécessité de connaître le nombre de composantes et de fixer les degrés de l'approximation, cet algorithme nécessite un temps de calcul qui croît exponentiellement en fonction des nombres de composantes. Cependant, les estimateurs des paramètres ainsi obtenus sont optimaux au sens du maximum de vraisemblance

(quoiqu'on rappelle que cette notion a une consonance asymptotique ce qui n'est pas notre cas à cause du faible nombre d'échantillons). Une autre limitation est que le nombre total de paramètres à estimer ne doit pas dépasser le nombre d'échantillons mis à notre disposition. Autrement dit, il faut que :

$$3K + \sum_{i=1}^K (M_{F_i} + M_{A_i}) < N + 1.$$

Toutefois, le plus grand avantage de cet algorithme est essentiellement le fait que l'estimation d'une composante ne dépend pas des autres vu qu'elles sont toutes estimées simultanément. L'erreur d'estimation ne sera pas plus importante pour une composante par rapport à une autre. Cet algorithme sera comparé à celui présenté dans le paragraphe suivant et qui estime le signal composante par composante.

7.2.2.2 Méthode sous-optimale - Estimation composante par composante

Principe L'algorithme itératif présenté ici estime les composantes une par une. A chaque itération, sont déterminés l'amplitude, la fréquence et la phase instantanées d'une seule et même composante et ce en maximisant la fonction de vraisemblance sous contraintes. Cette dernière s'écrit :

$$\hat{\boldsymbol{\theta}}_i = \arg \min_{\boldsymbol{\theta}_i} \ell_{MC} \left(\boldsymbol{\theta}_i | \hat{\boldsymbol{\theta}}_1, \dots, \hat{\boldsymbol{\theta}}_{i-1}, \{ \boldsymbol{\theta}_{i+1} = \dots, \boldsymbol{\theta}_K = 0 \} \right), \quad (7.9)$$

$$\text{avec } \ell_{MC} \left(\boldsymbol{\theta}_i | \hat{\boldsymbol{\theta}}_1, \dots, \hat{\boldsymbol{\theta}}_{i-1}, \{ \boldsymbol{\theta}_{i+1} = \dots, \boldsymbol{\theta}_K = 0 \} \right) = \ell_{MC}(\boldsymbol{\theta}), \quad (7.10)$$

et $\boldsymbol{\theta} = [\hat{\boldsymbol{\theta}}_1^\top, \dots, \hat{\boldsymbol{\theta}}_i^\top, [0, \dots, 0]^\top]^\top$. Les grandeurs $A_i[n]$, $F_i[n]$, $\Phi_i[n]$ et $s_i[n]$ sont ensuite reconstruites en utilisant les équations (7.1), (7.2) et (7.3).

Algorithme Le tableau 7.2 décrit l'algorithme correspondant à cette méthode. L'initialisation des paramètres d'une composante se fait de la même manière que dans le paragraphe 3.4.1 (cf. équations (3.13) ou (3.17)). A chaque itération, uniquement les paramètres de la composante en question sont initialisés. L'algorithme est arrêté si la puissance du signal résidu (le signal obtenu après élimination de toutes les composantes estimées) normalisée par celle du bruit est une variable de χ^2 de $2(N + 1)$ degrés de liberté (cf. paragraphe 3.4.8).

Avantages et inconvénients Contrairement à l'algorithme optimal (cf. paragraphe 7.2.2.1), un phénomène de propagation d'erreur se produit avec l'algorithme sous-optimal puisque l'estimation d'une composante dépend forcément de toutes celles précédentes. L'erreur d'estimation est plus importante sur les dernières composantes estimées que les premières. Toutefois, une bonne estimation du nombre de composantes présentes dans le signal est obtenue avec cet algorithme sans compter le gain au niveau du temps du calcul.

TAB. 7.2 – Méthode sous-optimale - Algorithme.

-
1. Initialisation : $\begin{cases} i = 1, \\ y_i[n] = y[n], \\ \boldsymbol{\theta}_i = \boldsymbol{\theta}_i^{(0)} \text{ (cf. équation (3.13)) et } \boldsymbol{\theta}_l = [0, \dots, 0]^T \text{ pour } l > i. \end{cases}$
 2. Itérations de $i = 1, \dots, K$:
 - Appliquer l'étape [2.2-] (a,b,c et d) de l'algorithme du recuit simulé, présenté dans le tableau 3.2, sur le signal $y_i[n]$ pour obtenir $\hat{\boldsymbol{\theta}}_i$.
 - Reconstruire la i -ème composante $s_i[n]$ en utilisant les équations (7.1), (7.2) et (7.3).
 - Evaluer le signal $y_{i+1}[n] = y_i[n] - s_i[n]$. Si $y_{i+1}[n]$ ne vérifie pas le critère d'arrêt, incrémenter i de 1 et réitérer les étapes 1 et 2.
-

La partie suivante de ce chapitre est dédiée à une étude des performances de ces deux méthodes proposées. Une comparaison avec des bornes de Cramer Rao sera présentée.

7.3 Etude des performances des algorithmes proposés

Cette partie est organisée comme suit : nous commençons par déterminer les bornes de Cramer-Rao qui serviront dans l'analyse de la robustesse des algorithmes proposés. Ces deux derniers sont appliqués sur des signaux simulés. Un cas très intéressant, celui de composantes se croisant dans le plan temps-fréquence sera particulièrement étudié. Ensuite, une application sur des signaux réels sera développée.

7.3.1 Bornes de Cramer Rao

En utilisant les mêmes notations que celles du paragraphe 5.3.1, la Matrice d'Information de Fisher pour des signaux multi-composantes s'écrit [FF95] :

$$\mathbf{I}_{\boldsymbol{\theta}} = \frac{2}{\sigma_b^2} \text{Re} \left\{ \left[\begin{array}{cc} \boldsymbol{\Psi}_i^\dagger \boldsymbol{\Psi}_l & \boldsymbol{\Psi}_i^\dagger \tilde{\boldsymbol{\Psi}}_l \\ \tilde{\boldsymbol{\Psi}}_i^\dagger \boldsymbol{\Psi}_l & \tilde{\boldsymbol{\Psi}}_i^\dagger \tilde{\boldsymbol{\Psi}}_l \end{array} \right]_{\substack{1 \leq i \leq K \\ 1 \leq l \leq K}} \right\}, \quad (7.11)$$

avec i et l sont les indices de la i -ème et la l -ème composantes.

Rappelons d'abord que les termes $\boldsymbol{\Psi}_i^\dagger \tilde{\boldsymbol{\Psi}}_l$ et $\tilde{\boldsymbol{\Psi}}_i^\dagger \boldsymbol{\Psi}_l$ sont nuls pour $i = l$ et que l'utilisation de la base orthonormale discrète (2.25) permet de découpler les paramètres de l'amplitude d'une même composante : $\boldsymbol{\Psi}_i^\dagger \boldsymbol{\Psi}_i$ est la matrice identité.

Dans le cas d'un signal à composantes multiples, il est difficile de trouver une base polynomiale qui assure le découplage des paramètres de toutes les composantes. C'est pourquoi nous continuons d'utiliser la base orthonormale discrète (2.21) puisqu'elle, au moins, garantit le découplage entre les paramètres de l'amplitude de la même composante.

Remarque Si le signal étudié comporte des composantes qui deviennent très proches ou qui se coupent dans le plan temps-fréquence, la matrice \mathbf{I}_θ devient mal conditionnée et tend vers une matrice singulière. Ceci explique la difficulté de l'estimation dans ce cas.

7.3.2 Application sur des signaux de simulations

Nous présentons l'analyse de deux signaux simulés chacun étant constitué de deux composantes à phases polynomiales :

- le premier signal est noté $Signal_1$, ses deux composantes ont des fréquences instantanées très proches mais séparées (cf. figure 7.1 (colonne gauche)).
- le deuxième signal est noté $Signal_2$, les fréquences instantanées de ses deux composantes se coupent (cf. figure 7.1 (colonne droite)).

La fréquence d'échantillonnage est de 1Hz. Ces deux signaux, formés chacun de 33 échantillons, correspondent aux cas (extrêmes) où la matrice d'information de Fisher devient très mal conditionnée. En particulier, l'estimation du $Signal_2$ représente un cas d'étude délicat auquel vient se rajouter le faible nombre d'échantillons et la présence de bruit.

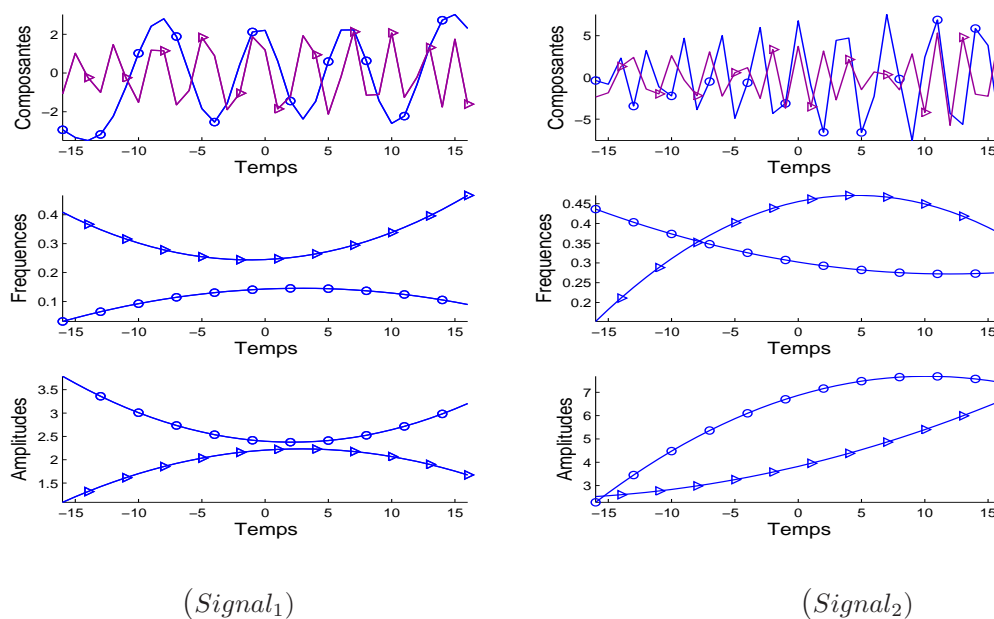


FIG. 7.1 – Deux signaux simulés, chacun d'eux est composé de deux composantes de 33 échantillons. Les amplitudes et les fréquences instantanées de ces dernières sont des polynômes de degré 2 : (- o) l'amplitude et la fréquence de la 1ère composante et (- Δ) l'amplitude et la fréquence de la 2ème composante.

7.3.2.1 Composantes fréquentielles très proches

Sur la figure 7.2, nous avons tracé les estimations de la fréquence et de l'amplitude instantanées du $Signal_1$ obtenues pour un RSB égal à 20 dB et 10 dB superposées avec

les courbes originales et ce en utilisant les deux algorithmes proposés. L'auto-corrélation du signal résidu, obtenu après la reconstruction des deux composantes et leur élimination du $Signal_1$, est représentée sur la figure 7.3 pour les deux techniques utilisées.

Les courbes estimées, ainsi obtenues, sont très proches de celles originales pour les deux RSB, témoignant alors d'une bonne performance des deux algorithmes. Les courbes d'auto-corrélation estimées correspondent aussi à celles réelles *i.e.* l'auto-corrélation du bruit simulé. Ce résultat va être consolidé par l'étude statistique suivante des performances de ces deux algorithmes.

Estimation des paramètres du modèle Pour prouver statistiquement la robustesse de ces deux algorithmes en présence du bruit, nous reportons sur les figures 7.4 à 7.7, les EQM et les biais d'estimation de tous les paramètres des deux composantes formant $Signal_1$ pour un RSB allant de 0 dB à 25 dB.

Rappelons, toutefois, que le RSB variant au cours du temps, une représentation plus judicieuse serait de tracer l'EQM et le biais en fonction de la variance du bruit et non du RSB, comme ce qui est classiquement fait (compte tenu du faible nombre d'échantillons, on est à la limite de la notion statistique du bruit blanc). Nous gardons ici cette représentation classiquement trouvée dans la littérature.

La comparaison de la figure 7.4 et celle 7.5 montre que, pour chaque algorithme, les résultats obtenus pour la première composante sont globalement similaires à ceux de la seconde composante. L'écart de l'EQM de la BCR est quasiment pareil dans les deux cas de figures. Cet écart s'explique en grande partie d'une part par la forte corrélation entre les différents paramètres (les termes en dehors de la diagonale de la matrice d'information de Fisher ne sont pas négligeables) et d'autre part par la forte non-linéarité en fonction des paramètres de fréquence. D'ailleurs, l'effet de seuil (connu dans l'estimation des paramètres de fréquences et de la phase à cause de la non-linéarité de la fonction de vraisemblance par rapport à ces paramètres) s'est produit à partir d'un RSB très inférieur à 20 dB.

La comparaison entre les deux algorithmes montre que l'EQM est plus faible en utilisant celui optimal quand le RSB devient inférieur à 5 dB. Cet algorithme est également celui qui donne les estimateurs les moins biaisés. Il est donc le plus robuste au bruit.

Estimation de l'amplitude et de la fréquence instantanées Sur la figure 7.8, sont tracées les EQM obtenues dans l'estimation des fréquences instantanées $F_1[n]$ et $F_2[n]$ et des amplitudes instantanées $A_1[n]$ et $A_2[n]$ ainsi que les BCR qui leur correspondent. Quant aux biais, ils sont représentés sur la figure 7.9.

Pour un RSB fixé égal à 20 dB, les deux algorithmes présentent des performances similaires au cours du temps que ce soit dans l'estimation des fréquences instantanées

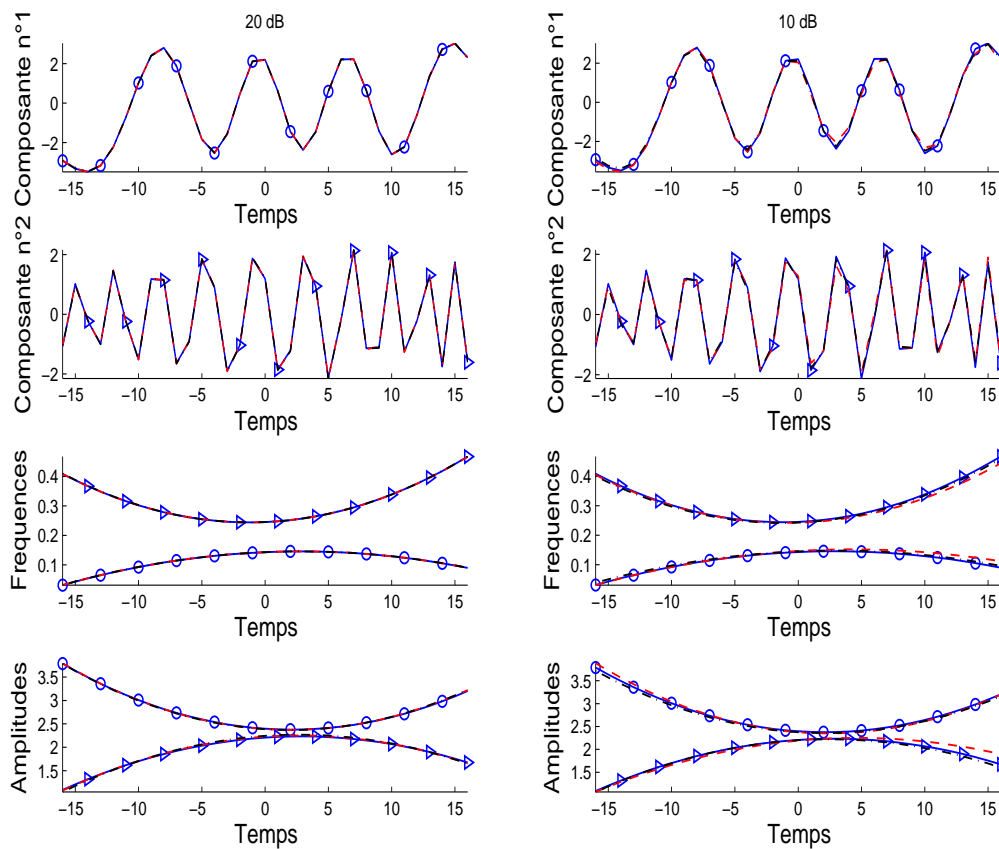


FIG. 7.2 – Estimation des composantes du $Signal_1$ et de leurs amplitudes et fréquences instantanées : les courbes estimées à un RSB de 20 dB (colonne gauche) et 10 dB (colonne droite) en utilisant l’algorithme optimal (- -) et celui sous-optimal (-.) superposées aux courbes originales (-).

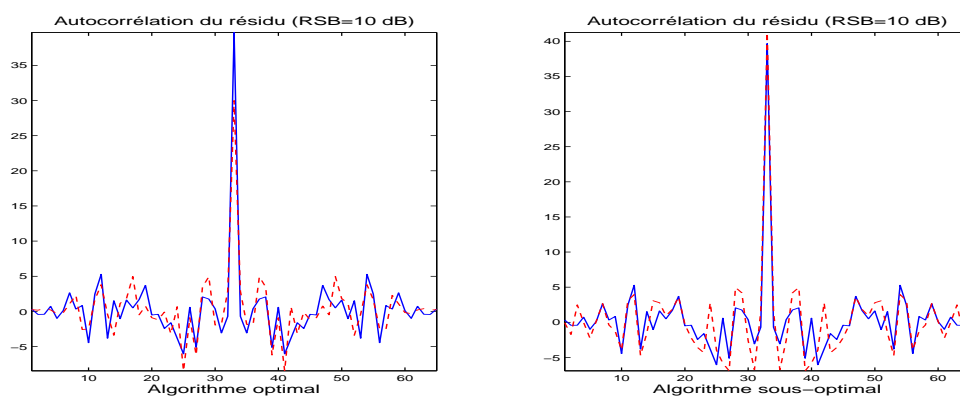


FIG. 7.3 – Auto-corrélation du signal résidu obtenu lors de l’estimation des composantes du $Signal_1$ pour un RSB de 10 dB : les courbes estimées (- -) en utilisant l’algorithme optimal et celui sous-optimal superposées aux courbes originales (-).

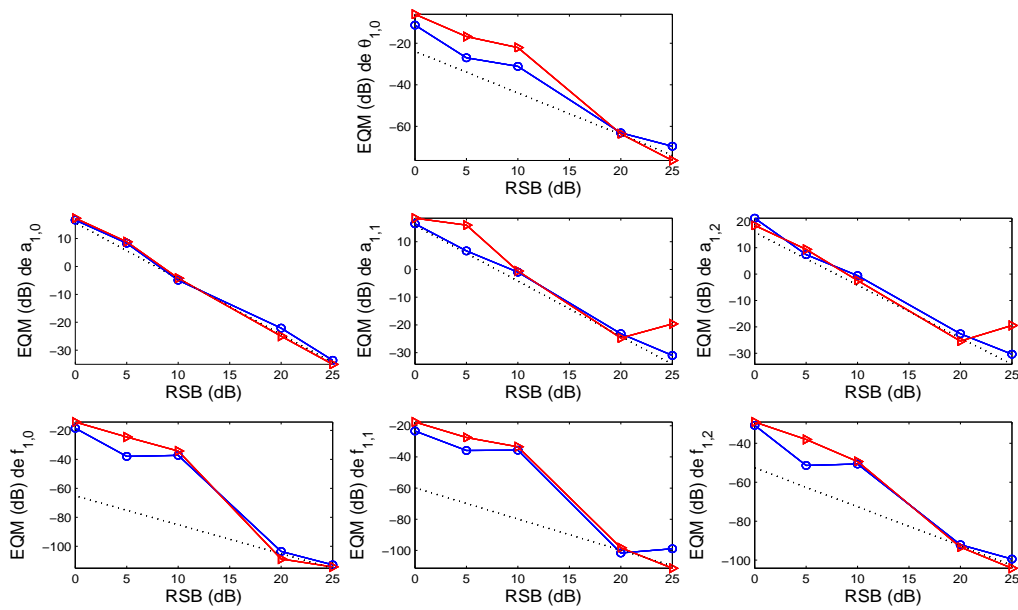


FIG. 7.4 – EQM obtenues dans l'estimation des paramètres de la 1ère composante du $Signal_1$ en utilisant l'algorithme optimal (-o) et celui sous-optimal (- Δ) superposées aux BCR (:).

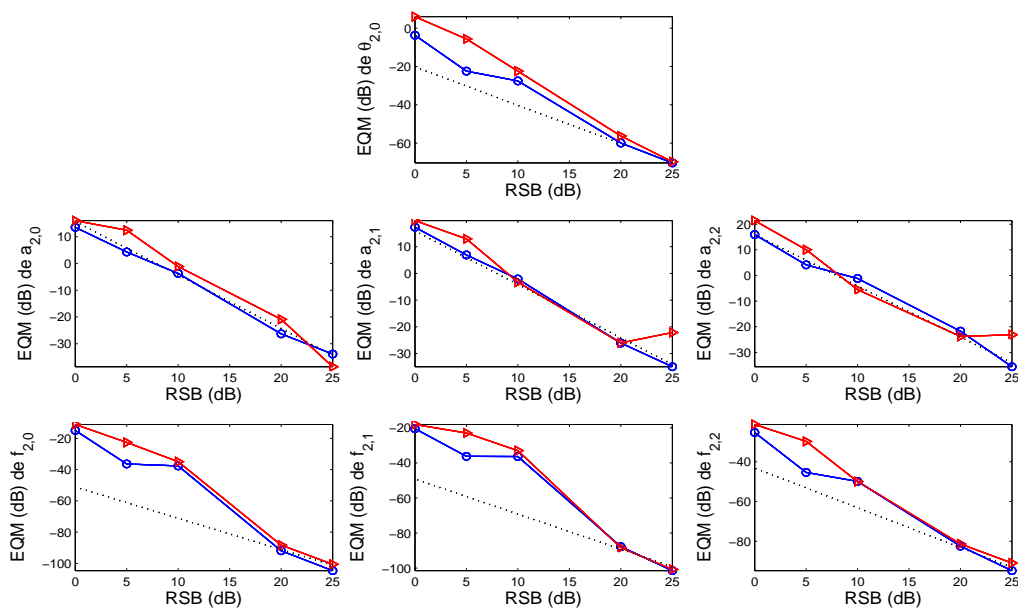


FIG. 7.5 – EQM obtenues dans l'estimation des paramètres de la 2ème composante du $Signal_1$ en utilisant l'algorithme optimal (-o) et celui sous-optimal (- Δ) superposées aux BCR (:).

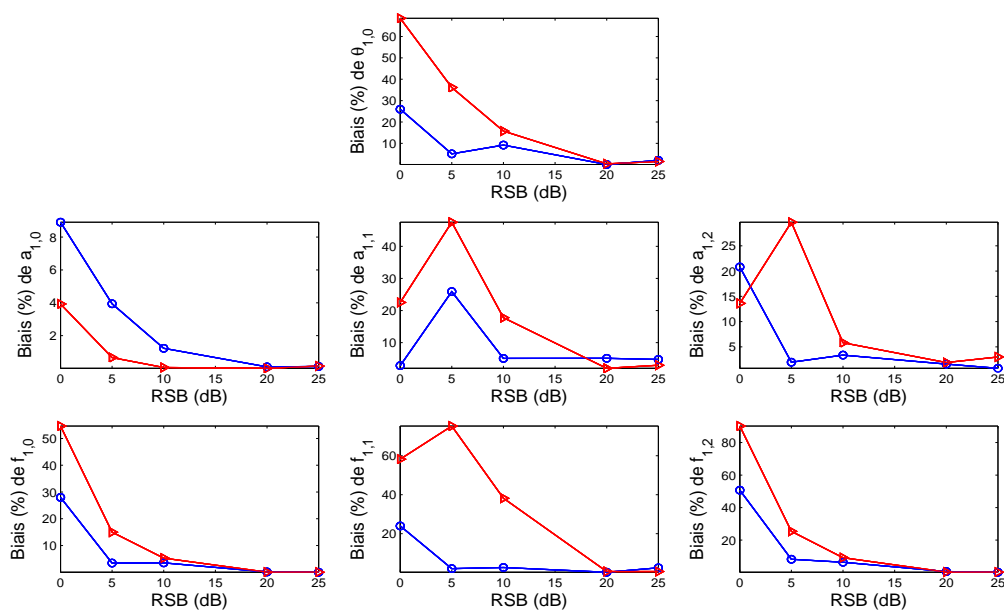


FIG. 7.6 – Biais obtenus dans l'estimation des paramètres de la 1ère composante du $Signal_1$ en utilisant l'algorithme optimal (-o) et celui sous-optimal (- Δ) superposées aux BCR (:).

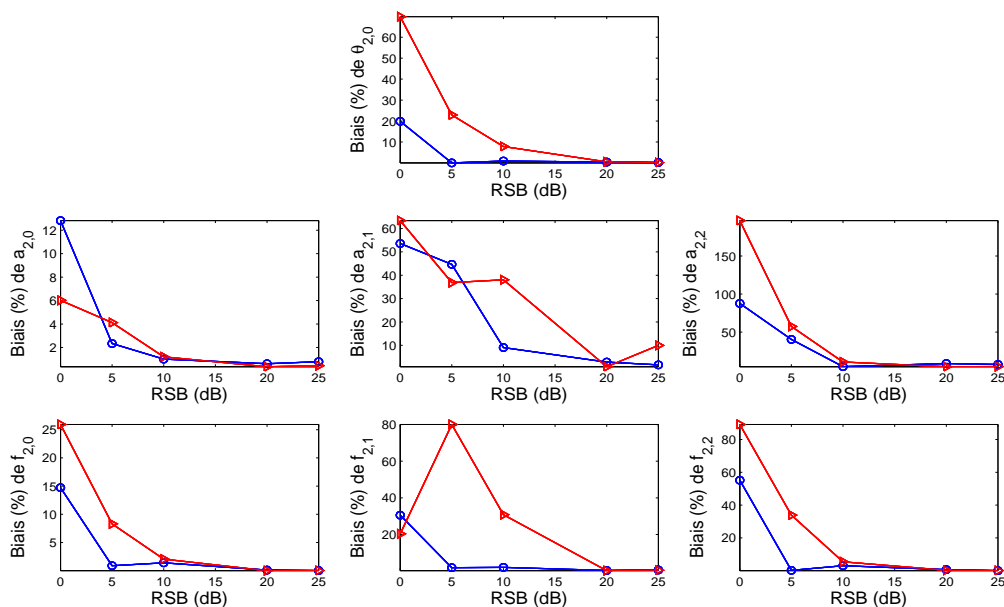


FIG. 7.7 – Biais obtenus dans l'estimation des paramètres de la 2ème composante du $Signal_1$ en utilisant l'algorithme optimal (-o) et celui sous-optimal (- Δ) superposées aux BCR (:).

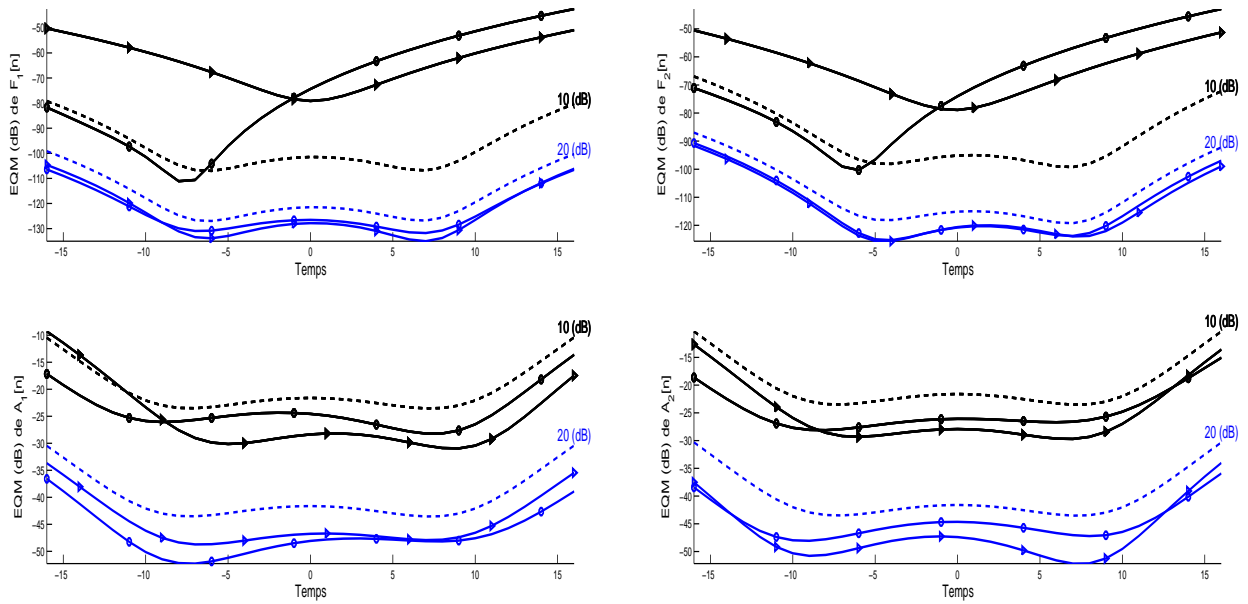


FIG. 7.8 – EQM obtenues dans l'estimation de $F_1[n]$ et $A_1[n]$ (respectivement $F_2[n]$ et $A_2[n]$) du $Signal_1$ en utilisant l'algorithme optimal (-o) et celui sous-optimal (- Δ) superposées aux BCR (...) pour des RSB de 20 et 10 dB : colonne gauche (respectivement colonne droite).

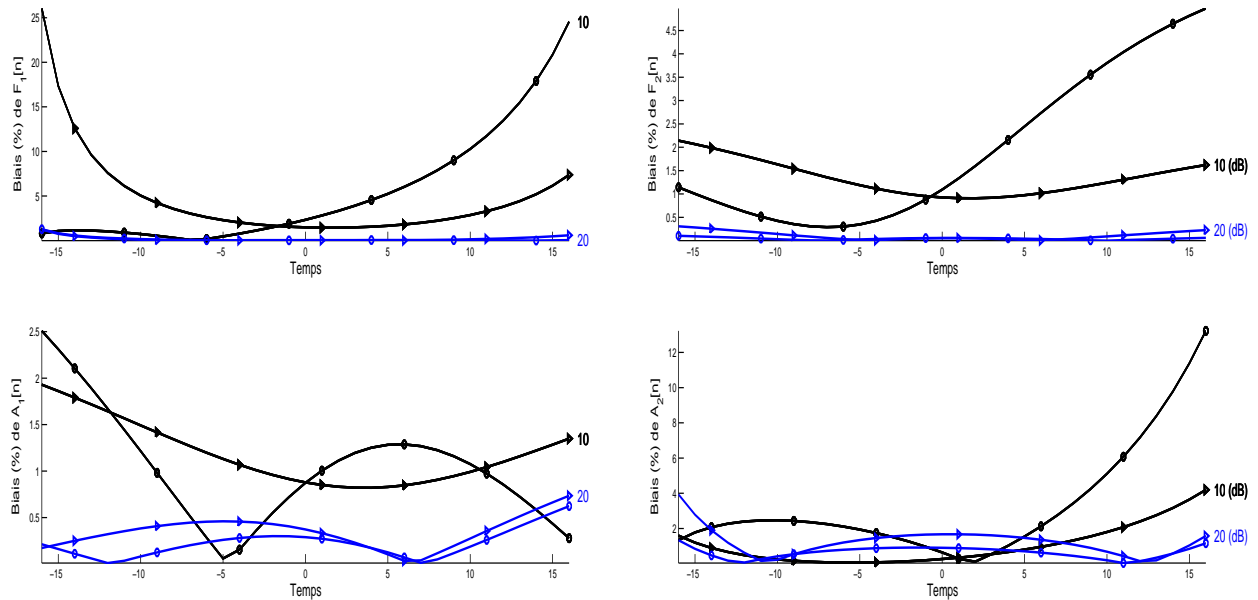


FIG. 7.9 – Biais obtenus dans l'estimation de $F_1[n]$ et $A_1[n]$ (respectivement $F_2[n]$ et $A_2[n]$) du $Signal_1$ en utilisant l'algorithme optimal (-o) et celui sous-optimal (- Δ) pour des RSB de 20 et 10 dB : colonne gauche (respectivement colonne droite).

des deux composantes ou de leurs amplitudes. Cette observation reste valable pour les amplitudes quand le RSB est fixé à 10 dB . En revanche, l'EQM obtenues dans l'estimation de la fréquence instantanée diffèrent d'une méthode à l'autre.

En effet, pour l'algorithme optimal l'EQM reste du même ordre de grandeur au cours du temps, nous voulons dire par cela qu'elle suit plus au moins la forme de la courbe théorique de la BCR tout en s'écartant de celle-ci en moyenne de plus de 20 dB pour les deux composantes. Quant aux performances de l'algorithme sous-optimal, elles présentent plus de variations au cours du temps. Pour un RSB de 10 dB , l'EQM de la fréquence est la plus faible au début des temps, *i.e.* là où l'amplitude de la première composante est plus importante, puis cette EQM se dégrade au fur et à mesure que le temps augmente. Etant donné que l'estimation de la seconde composante dépend de la première, le même scénario se reproduit pour l'EQM de $F_2[n]$, c'est-à-dire elle est faible au début du temps et plus importante à la fin des temps.

Notons au passage que l'algorithme sous-optimal va privilégier souvent les composantes les plus énergétiques en premier lieu. Toutefois, si plusieurs composantes semblent avoir la même énergie, elles peuvent être estimées dans n'importe quel ordre. En conclusion, nous pouvons dire que cet algorithme représente un bon compromis en terme d'EQM, de biais et de temps de calcul.

7.3.2.2 Croisement de composantes fréquentielles

Le deuxième cas d'étude présenté dans cette partie concerne le $Signal_2$, dont les composantes se croisent dans le plan temps-fréquence. L'estimation de ses deux composantes, de leurs fréquences et de leurs amplitudes est reportée sur la figure 7.10 pour deux RSB égaux à 20 dB et 10 dB et ce en utilisant les deux algorithmes proposés. Nous traçons également l'auto-corrélation du signal résidu sur la figure 7.11.

Ces courbes estimées, très proches des courbes originales, montrent que les deux algorithmes fournissent des bonnes estimations de la fréquence instantanée $F_1[n]$ et $F_2[n]$ pour un RSB de 20 dB et 10 dB , malgré la présence de croisement de fréquence. Ceci s'explique en partie par le respect des deux algorithmes de la condition de continuité de la phase. L'estimation des amplitudes instantanées $A_1[n]$ et $A_2[n]$ est très bonne pour un RSB de 20 dB mais celle de $A_2[n]$ se dégrade pour un RSB de 10 dB . A vrai dire cette deuxième composante a une puissance du même ordre de grandeur que le bruit.

On retrouve également que l'algorithme sous-optimal donne de bons résultats dans ce problème d'estimation délicat en raison du croisement des fréquences.

7.3.3 Application sur des signaux réels

Deux signaux réels sont considérés : un signal de baleine et le signal représentant le bilan de masses de glacier de Sarennes (cf. figures 1(a) et (d)) et auxquels nous appliquons les deux algorithmes proposés dans ce chapitre.

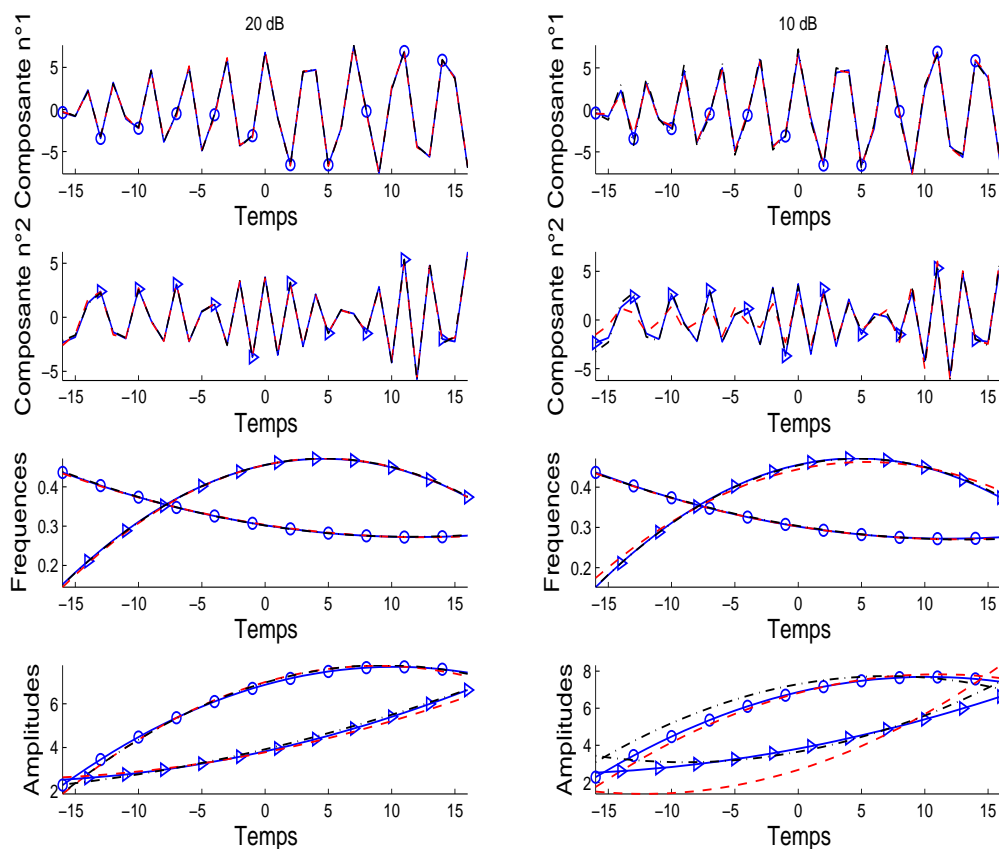


FIG. 7.10 – Estimation des composantes du $Signal_2$ et de leurs amplitudes et fréquences instantanées : les courbes estimées à un RSB de 20 dB (colonne gauche) et 10 dB (colonne droite) en utilisant l'algorithme optimal (- -) et celui sous-optimal (-.) superposées aux courbes originales (-).

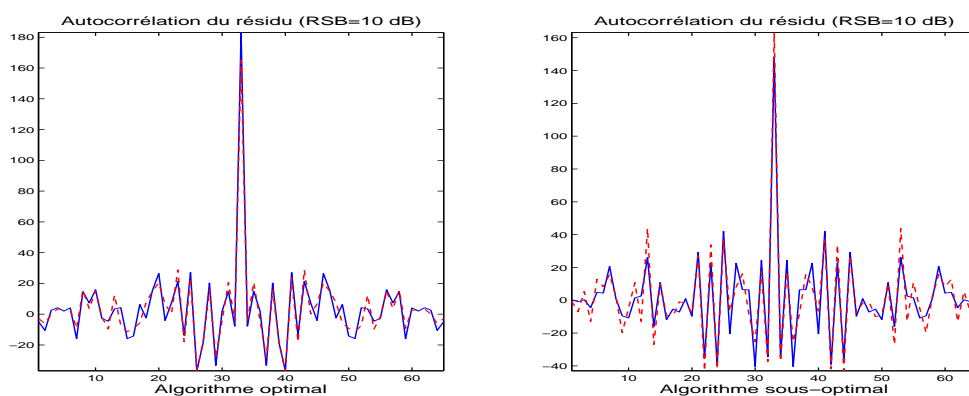


FIG. 7.11 – Auto-corrélation du signal résidu obtenu lors de l'estimation des composantes du $Signal_2$ pour un RSB de 10 dB : les courbes estimées (- -) en utilisant l'algorithme optimal et celui sous-optimal superposées aux courbes originales (-).

7.3.3.1 Signal de baleine

Il s'agit d'un segment prélevé sur un signal multi-composantes émis par une baleine et qui comporte 31 échantillons (L'origine des temps est référencée au centre du segment). Sur la figure 7.12(a), nous avons superposé le résultat de l'estimation des fréquences instantanées et le spectrogramme de ce signal en fréquence réduite. Nous pouvons distinguer trois composantes dont une est basse fréquence et possède une énergie similaire que celle haute fréquence. La composante de fréquence intermédiaire est la plus énergétique.

Les estimations fournies par les deux algorithmes sont très proches. Sur la figure 7.12(b), nous pouvons voir l'estimation des trois amplitudes instantanées tandis que sur la figure 7.12(c) sont reportés la somme des trois composantes reconstruites par les deux algorithmes et le signal résidu. L'auto-corrélation de ce dernier est représentée sur la figure 7.12. Le RSB est estimé à 14 *dB*.

Sachant que l'hypothèse du bruit blanc n'est pas vraie en milieu sous-marin, les résultats obtenus restent relativement corrects.

7.3.3.2 Bilan de masses du Glacier de Sarennes

Vu que ce signal court (cf. chapitre 1) comporte des composantes qui commencent et finissent à des instants différents et que nous n'avons pas encore développés de techniques maîtrisant la gestion du début et de la fin des composantes, nous allons estimer le signal sur une tranche de temps allant de l'année 1966 à 2004. Sur cette tranche temporelle, les composantes sont présentes simultanément. Le résultat de l'estimation est présenté sur la figure 7.13.

Trois composantes ont également été estimées en utilisant l'algorithme sous-optimal. On voit sur le spectrogramme du signal $s[n] - s_1[n]$ que la première composante estimée n'apparaît plus sur le spectrogramme.

7.4 Conclusion

Ce chapitre avait pour but de généraliser les méthodes proposées dans les chapitres précédents aux signaux multi-composantes courts non-stationnaires, dont les amplitudes et les fréquences sont des fonctions non-linéaires du temps. Les deux algorithmes proposés (celui dit optimal et l'autre sous-optimal) ont donné de très bons résultats souvent très proches mais celui sous-optimal est le plus avantageux en temps de calcul, au moins trois fois plus rapide.

La comparaison avec les BCR montre une bonne performance dans l'estimation des paramètres d'amplitude. L'analyse de deux signaux réels a validé l'applicabilité de ces deux algorithmes. Ce travail a été également publié dans [JMVL05b, JMVL05a].

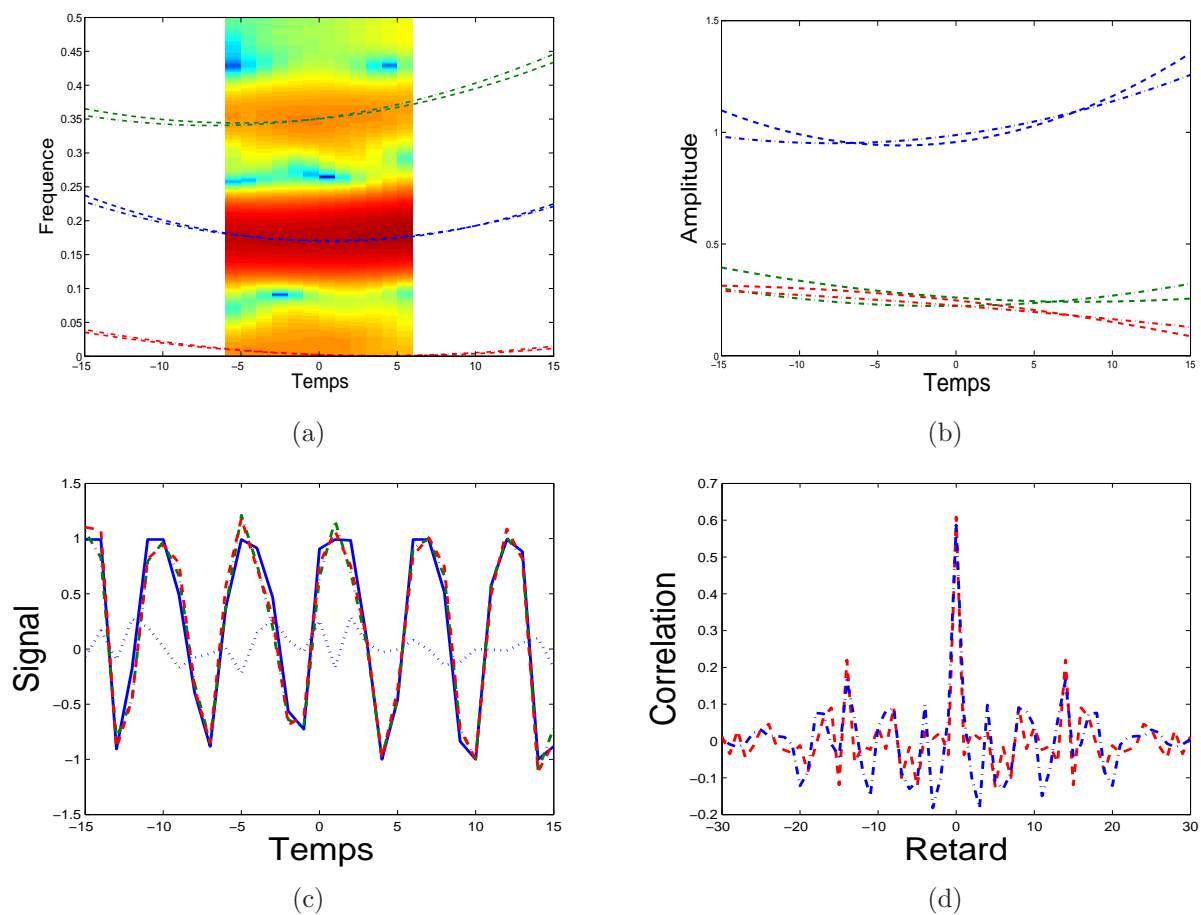


FIG. 7.12 – Signal réel de baleine : (—) algorithme optimal et (---) algorithme sous-optimal : (a) Les trois fréquences instantanées estimées sont superposées au spectrogramme calculé avec une fenêtre glissante de 12 points et $N_{FT} = 128$. (b) Les trois amplitudes instantanées estimées par les deux algorithmes. (c) Signal estimé et reconstruit par les deux méthodes superposé au (—) signal réel et au (...) signal résidu. (d) Auto-corrélation du signal résidu obtenu par les deux algorithmes.

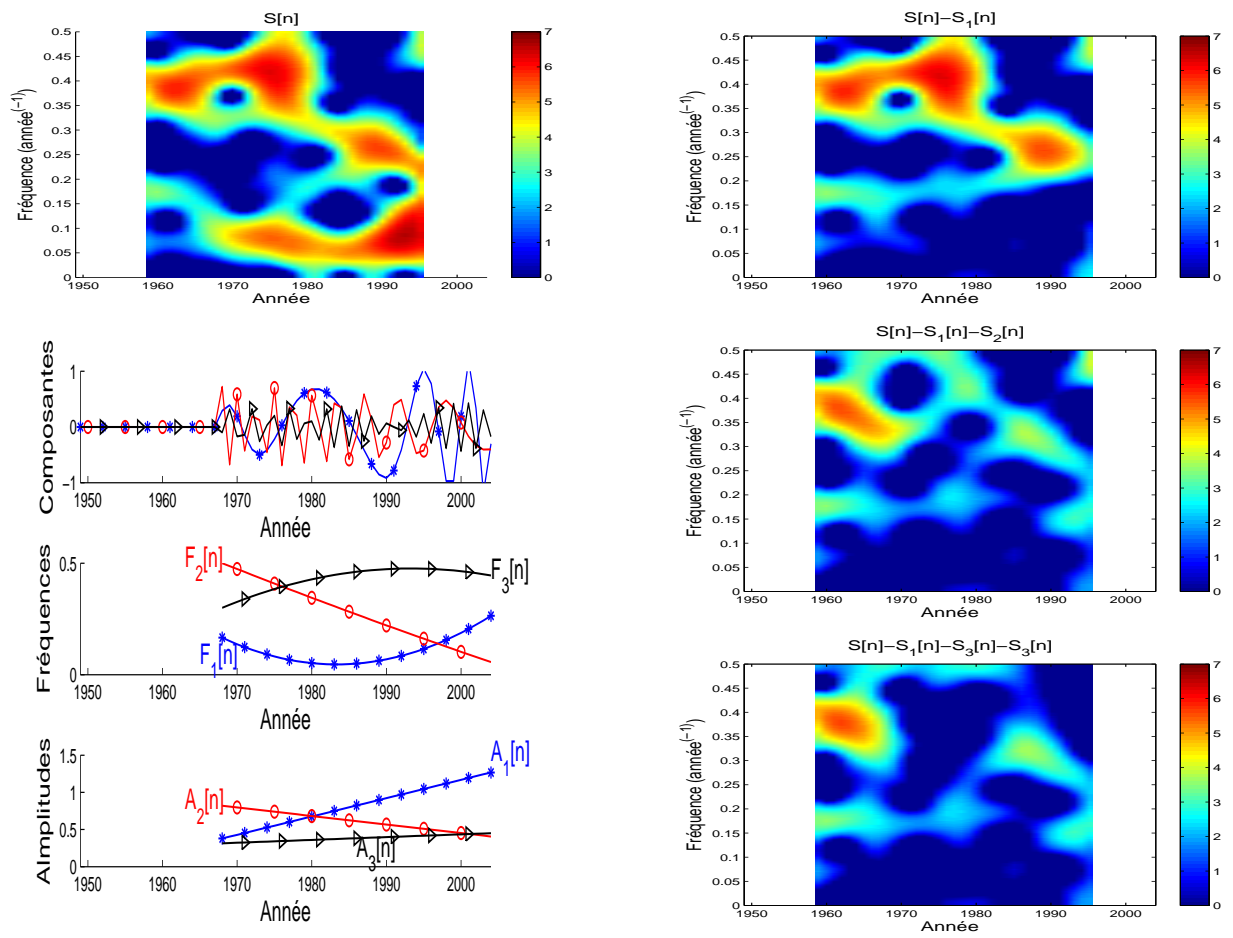


FIG. 7.13 – Estimation des composantes du signal représentant l'évolution du bilan de masses du Glacier de Sarennes depuis l'année 1966 et jusqu'à l'année 2000.

Conclusion et perspectives

Nous nous sommes intéressés dans cette thèse à l'analyse, la modélisation et l'estimation de signaux modulés à la fois en amplitude et en fréquence dans un contexte fortement non stationnaire et bruité.

Nous avons abordé, dans un premier temps, l'estimation des signaux mono-composantes courts et non-stationnaires. Notre approche paramétrique de ce problème a consisté à appliquer un modèle polynomial à l'amplitude et à la fréquence instantanées du signal étudié, ce modèle étant représenté dans une base polynomiale. Nous avons proposé et utilisé une base discrète et orthonormale permettant le découplage des paramètres de l'amplitude et ainsi une amélioration de l'estimation.

Nous avons proposé deux types de méthodes, que nous avons automatisées, pour réaliser l'estimation des paramètres du modèle :

- Dans la première méthode, la fonction de vraisemblance est maximisée en utilisant une technique d'optimisation stochastique, appelée recuit simulé.
- La seconde méthode est une approche bayésienne, dans laquelle la densité *a posteriori* conjointe des paramètres du modèle est déterminée.

Ces deux méthodes ont été comparées entre elles d'une part et avec la méthode basée sur la fonction d'ambiguïté d'ordre supérieur d'autre part. Cette comparaison montre que la méthode du maximum de vraisemblance constitue un bon compromis en termes de faibles erreurs quadratiques moyennes, de faibles biais et un temps de calcul réduit. Cette méthode a constitué, ensuite, une base sur laquelle nous avons développé une nouvelle technique pour estimer des signaux mono-composantes non-stationnaires de longue durée.

L'originalité de cette nouvelle technique est sa capacité à estimer différents types de non-stationnarités en les considérant et modélisant localement sur des segments courts.

Nous avons d'abord élaboré une stratégie de segmentation permettant d'obtenir des segments courts, sur lesquelles l'amplitude et la fréquence instantanées se prêtent à une modélisation locale polynomiale de faible degré. Les modèles locaux ainsi obtenus, et estimés à l'aide du recuit simulé en maximisant la fonction de vraisemblance locale, sont ensuite fusionnés pour reconstruire le signal, son amplitude et sa fréquence sur la totalité de leur durée. Pour ce fait, nous avons proposé une stratégie de fusion adaptée.

Cette technique proposée est performante, robuste aux variations du rapport signal à bruit et très simple à implémenter. Elle est flexible et adaptée à l'estimation de fortes non-stationnarités puisqu'elle permet de suivre, avec précision, les variations locales de la fréquence et de l'amplitude instantanées. Elle nous affranchit, ainsi, de l'emploi d'un modèle polynomial global nécessitant un ordre d'estimation très élevé pour tenir compte de la non-stationnarité. Le phénomène de propagation d'erreurs connu avec la méthode de la HAF est éliminé.

La troisième partie du travail présenté dans cette thèse a été consacré aux signaux multi-composantes courts. Deux méthodes ont été également proposées.

La première est une généralisation de celle employée pour estimer les signaux mono-composantes courts, *i.e.* les paramètres de toutes les composantes sont tous estimés simultanément, à l'aide de la maximisation de la fonction de vraisemblance en utilisant le recuit simulé. Bien qu'elle soit optimale et fournisse de très bons résultats, cette technique est très coûteuse en temps de calcul en raison du grand nombre de paramètres à estimer en même temps.

La deuxième procède d'une manière itérative, de façon à reconstruire le signal composante par composante. A chaque itération, seuls les paramètres d'une même composante sont estimés. Chaque composante traitée est éliminée du signal étudié avant d'entamer l'estimation de la composante suivante.

Ces deux méthodes sont performantes en présence d'un faible RSB. Elles produisent de bons résultats d'estimation et gèrent bien les cas où des composantes présentent des fréquences instantanées très proches ou qui se croisent dans le plan temps-fréquence.

Les algorithmes développés dans cette thèse ont été testés sur des signaux réels fortement non-stationnaires : un signal de canari de longue durée et un signal court représentant le bilan de masses de Glacier de Sarennes.

A la suite de ce travail de thèse, les perspectives sont de continuer la généralisation aux signaux longs et multi-composantes. Une façon de procéder à cette extension a été suggérée dans le chapitre 6 : la composante la plus énergétique du signal est déterminée itérativement, puis éliminée du signal étudié pour traiter la composante suivante. Dans le contexte des signaux courts, vu le faible nombre d'échantillons et la nature biaisé des estimateurs, il serait intéressant de comparer avec des bornes d'estimation plus adaptées à ce contexte, par exemple les bornes de Barankin [CFL04]. Il serait également intéressant de tester d'autres modèles différents : sinusoïdal, splines. D'autres types de bruit peuvent être étudiés : colorés, impulsions *etc.* Une comparaison avec les nouvelles techniques qui viennent d'apparaître récemment et qui ne supposent pas de modèle polynomial [DMD05] serait très intéressante.

Table des figures

1	Exemples de signaux réels : (a) bilan de masse de glaciers de Sarennes, $Fe = 1 \text{ année}^{-1}$ [Mar77, VVR95, Vin01, GTC05, Des05], (b) chant d'un canari, $Fe = 44.1 \text{ kHz}$, [Ele04, Leo00], (c) son émis par un long vison, $Fe = 11.025 \text{ kHz}$, (d) sifflement d'une baleine (<i>LongPilotWhale</i>), $Fe = 44.1 \text{ kHz}$, (e) son émis par un dauphin, $Fe = 44.1 \text{ kHz}$ et (f) son émis par une vache, $Fe = 44.1 \text{ kHz}$	10
1.1	Spectrogrammes de signaux réels : (a) Spectrogramme de chant d'un canari calculé avec une fenêtre d'analyse glissante de 256 points, un recouvrement de 50% et une transformée de Fourier discrète (TFD) de taille $N_{TF} = 1024$. (b) Spectrogramme de bilan de masses de glaciers de Sarennes, calculé avec $N_{TF} = 1024$, une fenêtre d'analyse de 19 points et un recouvrement de 99%.	17
1.2	Spectrogrammes de phase de signaux réels : (a) Celui du chant d'un canari, calculé avec une fenêtre d'analyse de 256 points, un recouvrement de 50% et $N_{TF} = 1024$. (b) Celui du bilan de masses de glaciers de Sarennes calculé avec une fenêtre d'analyse de 24 points, un recouvrement de 99% et $N_{TF} = 1024$. Ces deux spectrogrammes sont calculés en utilisant les programmes fournis par M. François léonard [Leo00].	18
1.3	Spectrogrammes ré-alloués de signaux réels : (a) celui de chant du canari calculé avec une fenêtre d'analyse de 256 points, un recouvrement de 50% et $N_{TF} = 1024$. (b) Celui du bilan de masses de glaciers de Sarennes calculé une fenêtre d'analyse de 24 points, un recouvrement de 99% et $N_{TF} = 1024$	19
1.4	Distribution de Wigner-Ville : (a) un signal à deux composantes modulées linéairement en fréquence, la raie centrale correspond aux termes d'interférences. (b) Un signal mono-composante ayant une modulation sinusoidale en fréquence, la courbe réelle (—) représentant cette modulation est également tracée.	21
1.5	Application d'un opérateur unitaire de déformation à un signal synthétique ayant une modulation de fréquence hyperbolique [Ioa03] : (a) Distribution de Wigner-Ville du signal avant la déformation. (b) Distribution de Wigner-Ville après la déformation. (c) RTF obtenue après application de l'opérateur de déformation inverse.	23
1.6	(a) Signal mono-composante et fortement non-stationnaire ayant une modulation polynomiale de la fréquence et de l'amplitude de degrés respectifs 7 et 3, la phase est de degré 8. (b) Propagation d'erreur de l'ordre le plus élevé au coefficient d'ordre inférieur avec la méthode de la HAF. Le niveau de l'erreur accumulée après l'estimation de Φ_4 affecte considérablement l'estimation celle de Φ_3 . La notation par rapport à K est supprimée.	29

1.7	Erreur Quadratique Moyenne (EQM) dans l'estimation de $\Phi_{1,2}$ et $\Phi_{1,4}$ d'un signal mono-composante à phase polynomiale de degré 4 et à amplitude constante ; le nombre d'échantillons est égal à 200 : comparaison entre la technique (-o-) MCNI, (-◇-) le HAF et (-△-) le PHAF. Ces figures sont extraites de [Ang00].	35
2.1	(a) Signal simulé ayant une modulation quadratique en amplitude et en fréquence et formé de 51 échantillons. La fréquence d'échantillonnage est égale à 1 Hz et le $RSB = 10$ dB. Sa fréquence instantanée (—) est superposée avec : (b) le spectrogramme calculé avec $N_{TF} = 256$, un recouvrement de 99% et une fenêtre de Hanning de 16 points, (c) la distribution de Wigner-Ville.	43
2.2	(a) Amplitude instantanée du signal tracé sur la figure 2.1(a), (b) sa fréquence instantanée et (c) sa phase instantanée.	44
2.3	(a) Evaluation du produit scalaire (2.7) en fonction du nombre d'échantillons, pour les polynômes de Legendre discrétisés (2.18) et pour la base polynomiale discrète et orthonormale (2.21). (b) Représentation temporelle des polynômes de la base orthonormale discrète (2.21).	50
2.4	Deux approches possibles pour l'estimation des paramètres du modèle. Dans la première, θ (2.11) est déterministe et estimé en maximisant la fonction de vraisemblance. Dans la deuxième, θ est considéré aléatoire et un estimateur est déterminé en utilisant la densité de probabilité <i>a posteriori</i>	52
3.1	Illustration de la non-linéarité de $\ell_{MC}(\theta)$ et de la présence de plusieurs minima locaux. Tous les paramètres sont fixés sauf f_0 et $f_2 : \{\varphi_0 = 0.4; a_0 = 16; a_1 = 2; a_2 = 3; f_1 = 0.5\}$	59
3.2	(a) $\ell_{MC}(\theta \hat{A}[n])$ fonction des moindres carrés (3.6) déjà tracée sur la figure 3.1 comparée à (b) au coefficient de corrélation (3.11), c'est l'opposé de $\gamma_{y,s}(\theta)$ qui est tracé.	61
3.3	Classification générale des méthodes d'optimisation en variables continues pour des fonctions mono-objectifs. Le terme "globale" désigne les méthodes convergeant vers un minimum global et "locale" celles stagnant sur un minimum local. Les méthodes "de Voisinage" font référence à celles qui font progresser une solution à la fois et les méthodes "distribuées" désignent celles qui manipulent en parallèle toute une population de solutions.	62
3.4	Organigramme du recuit simulé.	65
3.5	Initialisation des paramètres d'un signal ayant une modulation quadratique en amplitude et en fréquence pour deux valeurs de $RSB : -5$ dB et 30 dB. (a) Module de $ \mathcal{F}_y(\nu) $ du signal observé, calculée à l'aide d'une fenêtre de Hanning. (b) Fréquence (respectivement amplitude) instantanée (—) et son initialisation (- -) $F^{(0)}[n]$ (respectivement $A^{(0)}[n]$) en utilisant le spectrogramme. Ce dernier est calculé avec $N_{FT} = 256$, une fenêtre de Hanning de 8 points et un recouvrement de 99%.	67
3.6	Illustration du mécanisme de perturbation.	69
3.7	Probabilité d'acceptation de Metropolis en fonction de la variation Δ de la fonction objectif et du paramètre de contrôle, la température \mathcal{T}	70

3.8 Importance du réglage de la température : (a) la température est très basse est l'optimum est très loin, on s'interdit de franchir des zones permettant l'accès à ce dernier car la probabilité d'acceptation est quasi-nulle. (b) le voisinage immédiat de l'optimum est atteint et la température est assez élevée, des zones sont explorées inutilement car la probabilité d'acceptation est encore assez grande. Le signal considéré est celui de la figure 3.1. Tous les paramètres du signal sont fixés à leurs vraies valeurs sauf le paramètre f_0 , qu'on fait varier. L'optimum de la fonction $\ell_{MC}(\theta)$ correspond à $f_0 = 2$ 71

3.9 Importance du réglage de la variance de l'exploration de l'espace de configuration. (a) Cas d'une faible variance et d'un minimum global éloigné : risque de stagnation sur un minimum local. (b) Cas d'une variance élevée au voisinage de minimum global : perte de temps. 73

3.10 Loi de distribution de l'énergie $\sum_{n=-\frac{N}{2}}^{\frac{N}{2}} |b[n]|^2$ 74

5.1 (—) Signal 1 et (...) Signal 2 : deux signaux simulés ayant une modulation d'amplitude et de fréquence quadratique, le taux de modulation d'amplitude est différent pour chaque signal. (a) Fréquences instantanées, (b) Amplitudes instantanées et (c) Signaux temporels. 105

5.2 EQM et biais de l'estimation du paramètre de la fréquence f_1 et de la phase initiale θ_0 en fonction du RSB. Résultats obtenus pour Signal 3 (colonne gauche) de la figure (a) et de la figure (b) : (..+) Recuit simulé, (-.Δ-) HAF et (-+) BCR (5.15). Résultats obtenus pour Signal 4 (colonne droite) de la figure (a) et de la figure (b) (..o) Recuit simulé, (-.Δ-) HAF et (-o) BCR. 116

5.3 (a) EQM de l'estimation des paramètres de la fréquence instantanée et de la phase initiale découplés en fonction du RSB : Signal 3 EQM (..+) et BCR (-+) (5.19) et Signal 4 EQM (-o) et BCR (-o). (b) Biais de l'estimation de ces paramètres. 117

5.4 EQM de la fréquence instantanée et BCR (—) (5.15) en fonction du temps pour un RSB 15 dB : Signal 3 (..+) et Signal 4 (-o-). 118

5.5 Variation du RSB instantané en fonction du temps est de la variance du bruit σ_b^2 dans le cas du Signal 1 (5.2), le RSB moyen est indiqué en rouge. 119

5.6 Variation du RSB instantané en fonction du temps est de la variance du bruit σ_b^2 dans le cas du Signal 2 (5.2), le RSB moyen est indiqué en rouge. 119

5.7 EQM et biais (%) de a_2 du Signal 1 (5.2) obtenus en utilisant le recuit simulé : base discrète orthonormale (..Δ), la base canonique (..+), base de Legendre discrétisée (..*) et BCR (5.15) (—). 120

5.8 EQM et biais (%) de θ_0 du Signal 1 (5.2) obtenus en utilisant le recuit simulé : base discrète orthonormale (..Δ), la base canonique (..+), base de Legendre discrétisée (..*) et BCR (5.15) (—). 121

5.9 EQM et biais (%) de f_0 du Signal 1 (5.2) obtenus en utilisant le recuit simulé : base discrète orthonormale (..Δ), base canonique (..+), base de Legendre discrétisée (..*) et BCR (5.15) (—). 121

5.10 Influence du taux de modulation sur l'estimation des fréquences des amplitudes instantanées : (a) un RSB 15 dB et (b) 5 dB dans le cas du Signal 1 ((- -) EQM et (-) BCR) et Signal 2 ((...) EQM et (-o-) BCR). 122

5.11	Modélisation et estimation de deux segments prélevés sur un signal ayant une modulation polynomiale de degré 5 en fréquence et en amplitude : (a) le signal et les segments extraits (1) et (2), (...) EQM , (—) BCR. (b) (...) courbes originales, (—) segments estimés et (—) biais d'estimation.	123
5.12	Modélisation et estimation de deux segments prélevés sur un signal ayant une modulation sinusoïdale en fréquence et en amplitude : (a) le signal et les segments extraits (1) et (2), (...) EQM , (—) BCR. (b) (...) courbes originales, (—) segments estimés et (—) biais d'estimation.	123
5.13	Chaînes de Markov générées par l'algorithme MH-MCMC pour le paramètre $a_1 = -3$ de l'amplitude de Signal 1 (5.2).	124
5.14	Comparaison entre le MH-MCMC (...) EQM et (—) Biais, le recuit simulé (-.-) EQM et (+++) Biais et BCR (5.20) (—) appliqué au Signal 1 pour un RSB égal à 10 et 0 dB.	125
6.1	(a) Exemples de segments courts adaptés à une modèle local polynomial de second degré et extraits à partir d'un signal fortement non-stationnaire ayant une amplitude et une fréquence instantanées polynomiales de degrés respectifs 3 et 7. (S1), (S2) et (S3) sont 3 segments de longueurs respectives 55, 63 et 51 échantillons. (b) La fréquence instantanée.	133
6.2	(a) Prélèvement du 1er segment du signal présenté sur la figure 6.1. (b) Prélèvement du 2ème segment.	137
6.3	(a) Prélèvement d'un segment d'un signal ayant une modulation polynomiale d'amplitude de fréquence de degrés respectifs 2 et 3. L'estimation de ce signal sera étudiée dans le paragraphe 6.6.2.1.	141
6.4	Deux cas de figures de segments à fusionner : (a) recouvrement d'un seul côté et (b) recouvrement des deux côtés. Les fenêtres de pondération utilisées sont celles de Hamming.	142
6.5	(a) Illustration de la procédure de la segmentation du signal : le nombre encadré correspond à l'ordre de traitement du segment. (b) Courbes originales (—) de l'amplitude instantanée (polynôme de degré 3) et de la fréquence instantanée (polynôme de degré 7) superposées à leur estimées (...) basées sur une moyenne de 50 réalisations de bruit par des simulations de Monte Carlo pour des RSB moyens de 15 et 7 dB.	145
6.6	(a) BCR (6.38) et (6.42) obtenues après la fusion de trois segments qui se recouvrent. (b) Effet d'une sur-estimation de 1 du degré de l'approximation polynomiale de $F[n]$ et $A[n]$ sur les BCR polynomiales (6.45). Le signal utilisé est celui tracé sur la figure 6.1.	150
6.7	(a) $F[n]$ et $A[n]$ du signal défini par l'équation (6.47). (b) Signal temporel et RSB instantané en (dB).	151
6.8	EQM et biais (%) de l'estimation de $A[n]$ et $F[n]$ du signal (6.47) obtenues, à un RSB moyen égal à 15 dB, en utilisant (-.-) HAF et (—) l'approche locale proposée, (-◊) BCR classiques (6.45) et les (-*) BCR adaptée (6.38) et (6.42).	152
6.9	(a) EQM et biais de l'estimation des paramètres α_1 de $A[n]$ et β_0 de $\Phi[n]$ (6.43) du signal défini par (6.47) en fonction du RSB : (...) HAF, (- -) l'approche proposée et (—) les BCR appropriées. (b) Reconstruction du signal pour un RSB=15 dB : (—) Signal original , (- -) Signal estimé , (—) Erreur d'estimation pour un RSB de 15 dB. Les deux figures ne sont pas à la même échelle.	153

6.10	EQM obtenues dans l'estimation de $A[n]$ (polynôme de degré 3) et $F[n]$ (polynôme de degré 7) du signal représenté sur la figure 6.11 pour un RSB moyen de 15 dB (colonne gauche) et 7 dB (colonne droite) : (—) l'approche locale proposée, ($-\diamond$) BCR classiques (6.45) et ($-\ast$) BCR adaptées (6.38) et (6.42).	154
6.11	Biais obtenus dans l'estimation de $A[n]$ (polynôme de degré 3) et $F[n]$ (polynôme de degré 7) du signal représenté sur la figure 6.11 pour un RSB global de 15 dB (colonne gauche) et 7 dB (colonne droite).	154
6.12	(a) EQM (- -) obtenues par l'estimation du signal modulé sinusoidalement en amplitude et en fréquence. Les $-\diamond$) BCR classiques (6.45) sont calculés pour un RSB moyen égal à 15 dB (colonne gauche) et un RSB moyen égal à 7 dB (colonne droite) en supposant une approximation polynomiale de degré 7 pour la modulation de l'amplitude et pour celle de la fréquence, les ($-\ast$) BCRs adaptées (6.38) et (6.42). (b) Biais de l'estimation.	156
6.13	Les courbes originales (—) de la modulation sinusoidale d'amplitude et de fréquence superposées à leurs estimées par l'approche proposée : (-.-) obtenues à 15 dB et (- -) celles obtenues à 7 dB.	157
6.14	(a) (- -) Les courbes estimées par la méthode proposée superposées avec celles originales obtenue par l'estimation du signal modulé sinusoidalement en amplitude et en fréquence (6.49). (b) De haut en bas : le spectrogramme avec une fenêtre d'analyse de 32 points, la distribution de Wigner-Ville, la distribution de Wigner-Ville lissée [Mar07].	158
6.15	Chant d'un canari fortement non-stationnaire : (a) Signal temporel, (b) Densité spectrale de puissance et (c) Spectrogramme calculé avec une fenêtre de Hamming glissante de 256 points, $N_{FT} = 256$ points et un recouvrement de 99%.	159
6.16	Chant d'un canari fortement non-stationnaire : estimation de la fréquence et de l'amplitude instantanées de la composante la plus énergétique du signal et reconstruction de cette dernière.	161
6.17	Signal résidu obtenu après l'estimation de la composante la plus énergétique du signal chant de canari : spectrogramme calculé avec une fenêtre de Hamming glissante de 256 points, $N_{FT} = 256$ et un recouvrement de 99%.	161
7.1	Deux signaux simulés, chacun d'eux est composé de deux composantes de 33 échantillons. Les amplitudes et les fréquences instantanées de ces dernières sont des polynômes de degré 2 : (- o) l'amplitude et la fréquence de la 1ère composante et (- Δ) l'amplitude et la fréquence de la 2ème composante.	171
7.2	Estimation des composantes du $Signal_1$ et de leurs amplitudes et fréquences instantanées : les courbes estimées à un RSB de 20 dB (colonne gauche) et 10 dB (colonne droite) en utilisant l'algorithme optimal (- -) et celui sous-optimal (-.) superposées aux courbes originales (-).	173
7.3	Auto-corrélation du signal résidu obtenu lors de l'estimation des composantes du $Signal_1$ pour un RSB de 10 dB : les courbes estimées (- -) en utilisant l'algorithme optimal et celui sous-optimal superposées aux courbes originales (-).	173
7.4	EQM obtenues dans l'estimation des paramètres de la 1ère composante du $Signal_1$ en utilisant l'algorithme optimal (-o) et celui sous-optimal (- Δ) superposées aux BCR (:).	174

7.5	EQM obtenues dans l'estimation des paramètres de la 2ème composante du <i>Signal</i> ₁ en utilisant l'algorithme optimal (-o) et celui sous-optimal (-Δ) superposées aux BCR (:).	174
7.6	Biais obtenus dans l'estimation des paramètres de la 1ère composante du <i>Signal</i> ₁ en utilisant l'algorithme optimal (-o) et celui sous-optimal (-Δ) superposées aux BCR (:).	175
7.7	Biais obtenus dans l'estimation des paramètres de la 2ème composante du <i>Signal</i> ₁ en utilisant l'algorithme optimal (-o) et celui sous-optimal (-Δ) superposées aux BCR (:).	175
7.8	EQM obtenues dans l'estimation de $F_1[n]$ et $A_1[n]$ (respectivement $F_2[n]$ et $A_2[n]$) du <i>Signal</i> ₁ en utilisant l'algorithme optimal (-o) et celui sous-optimal (-Δ) superposées aux BCR (...) pour des RSB de 20 et 10 dB : colonne gauche (respectivement colonne droite).	176
7.9	Biais obtenus dans l'estimation de $F_1[n]$ et $A_1[n]$ (respectivement $F_2[n]$ et $A_2[n]$) du <i>Signal</i> ₁ en utilisant l'algorithme optimal (-o) et celui sous-optimal (-Δ) pour des RSB de 20 et 10 dB : colonne gauche (respectivement colonne droite).	176
7.10	Estimation des composantes du <i>Signal</i> ₂ et de leurs amplitudes et fréquences instantanées : les courbes estimées à un RSB de 20 dB (colonne gauche) et 10 dB (colonne droite) en utilisant l'algorithme optimal (- -) et celui sous-optimal (-.) superposées aux courbes originales (-).	178
7.11	Auto-corrélation du signal résidu obtenu lors de l'estimation des composantes du <i>Signal</i> ₂ pour un RSB de 10 dB : les courbes estimées (- -) en utilisant l'algorithme optimal et celui sous-optimal superposées aux courbes originales (-).	178
7.12	Signal réel de baleine : (- -) algorithme optimal et (-.) algorithme sous-optimal : (a) Les trois fréquences instantanées estimées sont superposées au spectrogramme calculé avec une fenêtre glissante de 12 points et $N_{FT} = 128$. (b) Les trois amplitudes instantanées estimées par les deux algorithmes. (c) Signal estimé et reconstruit par les deux méthodes superposé au (—) signal réel et au (...) signal résidu. (d) Auto-corrélation du signal résidu obtenu par les deux algorithmes.	180
7.13	Estimation des composantes du signal représentant l'évolution du bilan de masses du Glacier de Sarennes depuis l'année 1966 et jusqu'à l'année 2000.	181
B.1	EQM et biais obtenus dans l'estimation du paramètre de la fréquence f_0 en fonction du RSB : Signal 3 EQM (..+) et BCR (-+) (5.15) et Signal 4 EQM (- -o) et BCR (-o).	199
B.2	BCR (5.15), EQM et biais de l'estimation du paramètre f_2 de la fréquence instantanée en fonction du RSB : Signal 3 EQM (..+) et BCR (-+) et Signal 4 EQM (- -o) et BCR (-o).	200
B.3	EQM et biais (%) de a_0 du Signal 1 obtenus en utilisant la base discrète orthonormale (..Δ), la base canonique (..+) et la base de Legendre discrétisée (..*) comparés avec BCR correspondants (—).	200
B.4	EQM et biais (%) de a_1 du Signal 1 obtenus en utilisant la base discrète orthonormale (..Δ), la base canonique (..+) et la base de Legendre discrétisée (..*) comparés avec BCR correspondants (—).	201
B.5	EQM et biais (%) de f_1 du Signal 1 obtenus en utilisant la base discrète orthonormale (..Δ), la base canonique (..+) et la base de Legendre discrétisée (..*) comparés avec BCR correspondants (—).	201

B.6	EQM et biais (%) de f_2 du Signal 1 obtenus en utilisant la base discrète orthonormale (Δ), la base canonique (+) et la base de Legendre discrétisée (*) comparés avec BCR correspondants (—).	202
-----	---	-----

Annexe A

A.1 Classe de Cohen

Quelques exemples de représentations temps-fréquence $RTF(t, \nu; h)$ et les fonctions de paramétrisation, notées $h(\xi, \tau)$, qui leur sont associées *via* l'équation

$$RTF(t, \nu; h) = \int_{\mathbb{R}} \int_{\mathbb{R}} \int_{\mathbb{R}} e^{2\pi j \xi (s-t)} h(\xi, \tau) x\left(s + \frac{\tau}{2}\right) x^*\left(s - \frac{\tau}{2}\right) e^{-2\pi j \nu \tau} d\xi ds d\tau, \quad (\text{A.1})$$

sont exposés dans le tableau ci-dessous.

Nom	Expression	$h(\xi, \tau)$
Wigner-Ville	$\int_{\mathbb{R}} x\left(t + \frac{\tau}{2}\right) x^*\left(t - \frac{\tau}{2}\right) e^{-2\pi j \nu \tau} d\tau$	1
S-Wigner	$\int_{\mathbb{R}} x\left(t - \left(s - \frac{1}{2}\right)\tau\right) x^*\left(t - \left(s + \frac{1}{2}\right)\tau\right) e^{-2\pi j \nu \tau} d\tau$	$e^{2\pi j s \xi \tau}$
Choi-Williams	$\int_{\mathbb{R}} \int_{\mathbb{R}} \frac{\sigma}{ \tau } e^{-2\sigma^2 \frac{(s-t)^2}{\tau^2}} x\left(s + \frac{\tau}{2}\right) x^*\left(s - \frac{\tau}{2}\right) e^{-2\pi j \nu \tau} ds d\tau$	$e^{-\frac{(\pi \xi \tau)^2}{2\sigma^2}}$
Séparable	$\int_{\mathbb{R}} \int_{\mathbb{R}} g_1(\tau) g_2(s-t) x\left(s + \frac{\tau}{2}\right) x^*\left(s - \frac{\tau}{2}\right) e^{-2\pi j \nu \tau} ds d\tau$	$g_1(\tau) G_2(\xi)$
Spectrogramme	$\left \int_{\mathbb{R}} x(s) g^*(s-t) e^{-2\pi j \nu s} ds \right ^2$	$A_g^*(\xi, \tau)$
Rihaczek	$x(t) X^*(\nu) e^{-2\pi j \nu t}$	$e^{\pi j \xi \tau}$

$x(t)$ est le signal analysé et $X(\nu)$ est sa transformée de Fourier. $g(t)$, $g_1(t)$ et $g_2(t)$ sont des fenêtres d'analyse de transformées de Fourier respectives $G(\nu)$, $G_1(\nu)$ et $G_2(\nu)$.

$A_g^*(\xi, \tau)$ est défini par :

$$A_g^*(\xi, \tau) = \int_{\mathbb{R}} g\left(s + \frac{\tau}{2}\right) g^*\left(s - \frac{\tau}{2}\right) e^{2\pi j \xi s} ds. \quad (\text{A.2})$$

A.2 Signal réel - Signal analytique

D'après [Gab46, Vil58], pour définir une fréquence instantanée d'un signal réel, il faut lui associer un signal complexe. Pour un signal réel $s_r[n]$, un signal complexe peut être défini par exemple par :

$$s[n] = s_r[n] + j \mathcal{T}(s_r[n]) = s_r[n] + j s_i[n] = A[n] e^{j \Phi[n]}, \quad (\text{A.3})$$

où $\mathcal{T}(\cdot)$ est une transformation quelconque non-nulle. $s_i[n]$, $A[n]$ et $\Phi[n]$ représentent, respectivement, la partie imaginaire du signal complexe obtenu, son amplitude et sa phase. Il est montré dans [Vak96, CLV99], qu'on retrouve la définition de Gabor du signal analytique dont la partie réelle est $s_r[n]$ et la partie imaginaire est la transformée de Hilbert de $s_r[n]$:

- la continuité de l'amplitude du signal complexe, *i.e.* tout changement affectant le signal réel $s_r[n]$ provoque une variation au niveau de l'amplitude $A[n]$.
- la phase du signal complexe doit être inchangée suite à la multiplication du signal réel par une constante positive réelle.
- l'identification des harmoniques pures : si le signal réel est de la forme $A_0 \cos(2\pi w_0 n + \Phi_0)$, alors l'amplitude est donnée par $A[n] = A_0$, la phase $\Phi[n] = 2\pi w_0 n + \Phi_0$ et la fréquence $F[n] = w_0$.

Bien que cette définition permet de spécifier une représentation unique complexe pour le signal réel, *via* la transformée de Hilbert, un problème d'ambiguïté dans la détermination de l'amplitude et de la phase instantanées du signal complexe peut surgir dans les situations où le module du signal complexe s'annule pour un ensemble d'instantanés (cf. paragraphe 2.2.1).

Pour illustrer ce problème d'ambiguïté, un exemple d'un signal composé de deux harmoniques pures est considéré [CLV99] :

$$s_r[n] = \frac{1}{2} (\cos(2\pi w_1 n) + \cos(2\pi w_2 n)) = \cos\left(2\pi \frac{w_1 - w_2}{2} n\right) \cos\left(2\pi \frac{w_1 + w_2}{2} n\right). \quad (\text{A.4})$$

Le signal analytique, qui lui est associé, est le suivant :

$$s[n] = \cos\left(2\pi \frac{w_1 - w_2}{2} n\right) \cos\left(2\pi \frac{w_1 + w_2}{2} n\right) + j \cos\left(2\pi \frac{w_1 - w_2}{2} n\right) \sin\left(2\pi \frac{w_1 + w_2}{2} n\right) \quad (\text{A.5})$$

$$= \cos\left(2\pi \frac{w_1 - w_2}{2} n\right) e^{2\pi j \frac{w_1 + w_2}{2} n}. \quad (\text{A.6})$$

Une des possibilités de définir l'amplitude du signal est $A[n] = |\cos(2\pi \frac{w_1 - w_2}{2} n)|$ et la phase $\Phi[n] = 2\pi \frac{w_1 + w_2}{2} n + \frac{\pi}{2} (1 - \text{sgn}(\cos(2\pi \frac{w_1 - w_2}{2} n)))$. Ces deux expressions sont celles utilisées classiquement et correspondent, respectivement, au module du signal et à son argument. La fréquence instantanée est la dérivée de cette phase.

On peut, également, adopter la définition suivante pour l'amplitude [CLV99] : $A[n] = \cos(2\pi \frac{w_1 - w_2}{2} n)$. La phase est alors exprimée par $\Phi[n] = 2\pi \frac{w_1 + w_2}{2} n$.

A.3 Exemples de modèles du signal

En dehors du modèle à amplitude et à phase polynomiales (cf. paragraphe 2.2.2), plusieurs autres modèles de signaux non-stationnaires sont possibles. Parmi lesquels, on cite

- Modèle exponentielle complexe [GF99], dont les paramètres de la phase sont complexes $s[n] = e^{\sum_{m=1}^M c_m n^m}$, où c_m sont des coefficients complexes.
- Modèle à amplitude aléatoire et à phase polynomiale [MZ02]

$$s[n] = (\psi_n + \xi_n) e^{j \sum_{m=1}^M \Phi_m n^m}, \quad \psi_n = \sum_{m=0}^{M_\psi} b_m \gamma_m \left[\frac{n}{N+1} \right].$$

avec $N + 1$ est le nombre d'échantillons du signal et ξ est un processus réel centré aléatoire stationnaire indépendant du bruit. Sa moyenne est modélisée par la deuxième équation, où les b_k sont des réels inconnus et les γ sont des fonctions réelles continues bornées. Modèle où le bruit additif sur le signal est transformé en un bruit additif sur la phase instantanée du signal [Kaa98].

A.4 Expressions des polynômes de Legendre Standards

Sur l'intervalle $[-1,1]$, les expressions explicites, des polynômes de Legendre standards, pour un degré allant de 1 à 5, sont obtenues directement à partir de l'équation (2.15) et elles sont données par :

$$\begin{aligned} \mathcal{P}_0^{Leg-S}(t) &= 1, & \mathcal{P}_2^{Leg-S}(t) &= \frac{3t^2-1}{2}, & \mathcal{P}_4^{Leg-S}(t) &= \frac{35t^4-30t^2+3}{8}, \\ \mathcal{P}_1^{Leg-S}(t) &= t, & \mathcal{P}_3^{Leg-S}(t) &= \frac{5t^3-3t}{2}, & \mathcal{P}_5^{Leg-S}(t) &= \frac{63t^5-70t^3+15t}{8}. \end{aligned} \quad (\text{A.7})$$

A.5 Complément sur le calcul de la base de polynômes discrète et orthonormale

Ecrivons $\mathcal{P}_m^{Orth}[n]$ (2.21) sous la forme suivante :

$$\mathcal{P}_m^{Orth}[n] = \beta_{m,0}n^0 + \beta_{m,1}n^1 + \dots + \beta_{m,m-1}n^{m-1} + \beta_{m,m}n^m, \quad \text{pour } \frac{-N}{2} \leq n \leq \frac{N}{2}. \quad (\text{A.8})$$

En remplaçant (A.8) dans l'équation (2.20) qui correspond à la procédure de Gram-Schmidt, nous obtenons les expressions des coefficients $b_{m,i}$. Les expressions des cinq premiers polynômes sont reportées dans (2.21), où

$$\begin{aligned} \lambda &= \frac{(N+1) \sum k^6 - \sum k^2 \sum k^4}{(\sum k^4)^2 - \sum k^2 \sum k^6}, \\ \gamma &= -\frac{\lambda \sum k^2 + (N+1)}{\sum k^4}, \\ c &= (\gamma^2 \sum k^8 + 2\lambda\gamma \sum k^6 + (\lambda^2 + 2\gamma) \sum k^4 + 2\lambda \sum k^2 + N + 1)^{-\frac{1}{2}}; \end{aligned} \quad (\text{A.9})$$

et

$$\begin{aligned}
\nu &= \frac{\sum k^2 \sum k^8 - \sum k^4 \sum k^6}{(\sum k^6)^2 - \sum k^4 \sum k^8}, \\
\mu &= \frac{\nu \sum k^4 + \sum k^2}{\sum k^6}, \\
d &= (\mu^2 \sum k^{10} + 2\mu\nu \sum k^8 + (\nu^2 + 2\mu) \sum k^6 + 2\nu \sum k^4 + \sum k^2)^{-\frac{1}{2}}.
\end{aligned} \tag{A.10}$$

Le symbole \sum dénote la somme discrète de $k = \frac{-N}{2}$ à $\frac{N}{2}$.

A.6 Estimation des paramètres de l'amplitude

Considérons les mêmes notations que celles utilisées dans le paragraphe 3.2.3, à savoir \mathbf{y} , \mathbf{s} , \mathbf{D} , $\{\mathbf{G}_m\}$ base orthonormale et enfin les paramètres d'amplitude $\{\hat{a}_0, \dots, \hat{a}_{M_A}\}$.

Nous considérons les écritures matricielles de ces différentes quantités et de celle de la fonction des moindres carrés (3.9) :

$$\mathbf{s} = \mathbf{D} \sum_{m=0}^{M_A} a_m \mathbf{G}_m, \tag{A.11}$$

$$\ell_{MC}(\boldsymbol{\theta}) = (\mathbf{y} - \mathbf{s})^\dagger (\mathbf{y} - \mathbf{s}) \tag{A.12}$$

$$= \mathbf{y}^\dagger \mathbf{y} + \mathbf{s}^\dagger \mathbf{s} + 2\Re \{ \mathbf{y}^\dagger \mathbf{s} \}, \tag{A.13}$$

$$= \sum_{m=0}^{M_A} a_m^2 + 2 \sum_{m=0}^{M_A} a_m \Re \{ \mathbf{y}^\dagger \mathbf{D} \mathbf{G}_m \}. \tag{A.14}$$

La minimisation de cette dernière équation (A.14) par rapport à a_m , et en considérant la phase $\Phi[n]$ connue, fournit l'équation (3.8).

En revanche, la minimisation (A.13) par rapport aux paramètres de la phase, à amplitude constante, est équivalente à la maximisation de coefficient de corrélation (3.11), puisque $\mathbf{y}^\dagger \mathbf{y} + \mathbf{s}^\dagger \mathbf{s}$ est une quantité constant par rapport aux paramètres de la phase.

A.7 Techniques d'optimisation classiques

Descente du gradient La descente du gradient est un algorithme d'optimisation, qui cherche à approcher le minimum locale d'une fonction, par étape. La fonction à minimiser doit être définie et différentiable sur un voisinage du minimum, qu'on cherche à déterminer. A chacune étape, cet algorithme emploie la valeur et la direction du gradient (ou de son approximation), calculé en un point, pour déterminer le point suivant. Ses inconvénients sont essentiellement

1. L'algorithme peut effectuer un nombre d'itérations conséquent, avant de converger vers le minimum locale.
2. La convergence n'est pas assurée.

Broyden-Fletcher-Goldfarb-Shanno La méthode de BFGS est utilisée dans les problèmes d'optimisation non-linéaires sans contraintes. L'idée de base découle de la descente du gradient. Il s'agit de reconstruire la matrice du Hessien (relatif à la dérivée seconde de la fonction à minimiser), en analysant des différents vecteurs de gradient et en choisissant la meilleure direction. Puisque la fonction, à minimiser, est localement approchée par une fonction quadratique autour de minimum, la matrice du Hessien ne sera pas calculée à chaque étape de l'algorithme. La recherche du minimum se fera ensuite par la méthode de quasi-Newton [Bro70].

Gauss Newton Cet algorithme itératif est employé dans la résolution des problèmes de moindres carrés non linéaires. Il s'agit d'une modification de l'algorithme de Newton, où on ne se sert pas de la dérivée seconde de la fonction à minimiser, et où le pas décroissant vaut $\frac{1}{i}$, i désigne la i -ème itérations.

A.8 Marginalisation de la densité *a posteriori*

Partant de l'équation 4.25, et intégrons cette expression par rapport à σ_b^2 :

$$p(\boldsymbol{\theta}|\mathbf{y}) = \int_0^{+\infty} \mathcal{P}(\boldsymbol{\theta}, \sigma_b^2|\mathbf{y}) d\sigma_b^2 \propto \int_0^{+\infty} \frac{\pi^{-(N+1)}}{\sigma_b^{2(N+2)}} \exp\left[-\frac{1}{\sigma_b^2} \sum_{n=-\frac{N}{2}}^{\frac{N}{2}} |y[n] - s[n]|^2\right] d\sigma_b^2.$$

En utilisant, l'égalité suivante :

$$\int_0^{+\infty} x^{\beta-1} e^{-Qx} dx = \frac{\Gamma(\beta)}{Q^\beta},$$

et en posant $Q = \sum_{n=-\frac{N}{2}}^{\frac{N}{2}} |y[n] - s[n]|^2$ et $x = \frac{1}{\sigma_b^2}$, on obtient la relation suivante :

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto \int_0^{+\infty} x^N e^{-Qx} dx = \frac{\Gamma(N+1)}{Q^{N+1}}.$$

Ce qui revient à écrire

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto \left(\sum_{n=-\frac{N}{2}}^{\frac{N}{2}} |y[n] - s[n]|^2 \right)^{-(N+1)}.$$

Annexe B

B.1 Signaux courts non-stationnaires - Compléments

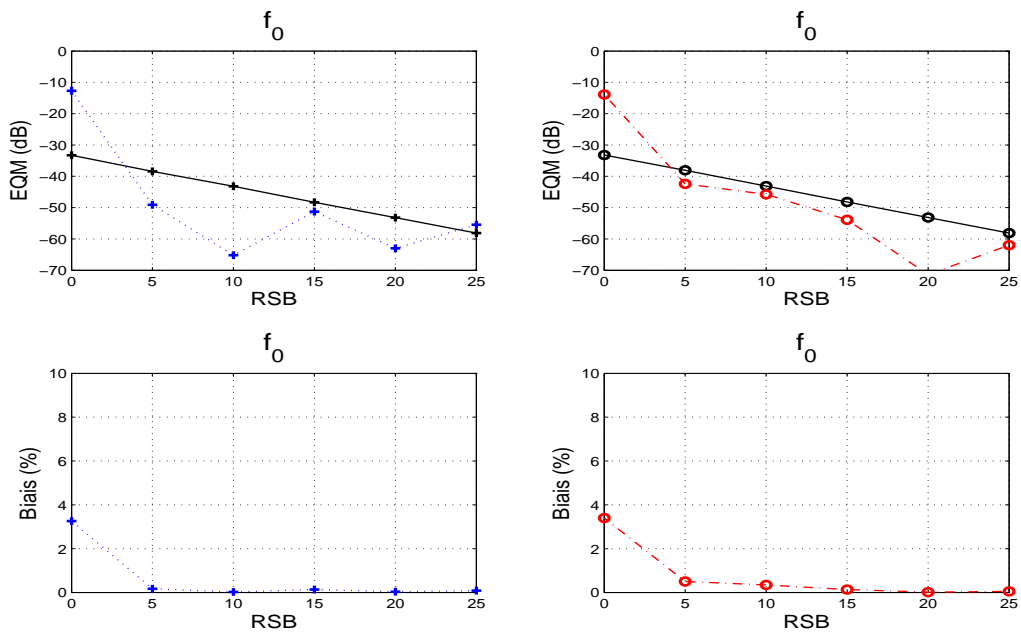


FIG. B.1 – EQM et biais obtenus dans l'estimation du paramètre de la fréquence f_0 en fonction du RSB : **Signal 3** EQM (..+) et BCR (-+) (5.15) et **Signal 4** EQM (- -o) et BCR (-o).

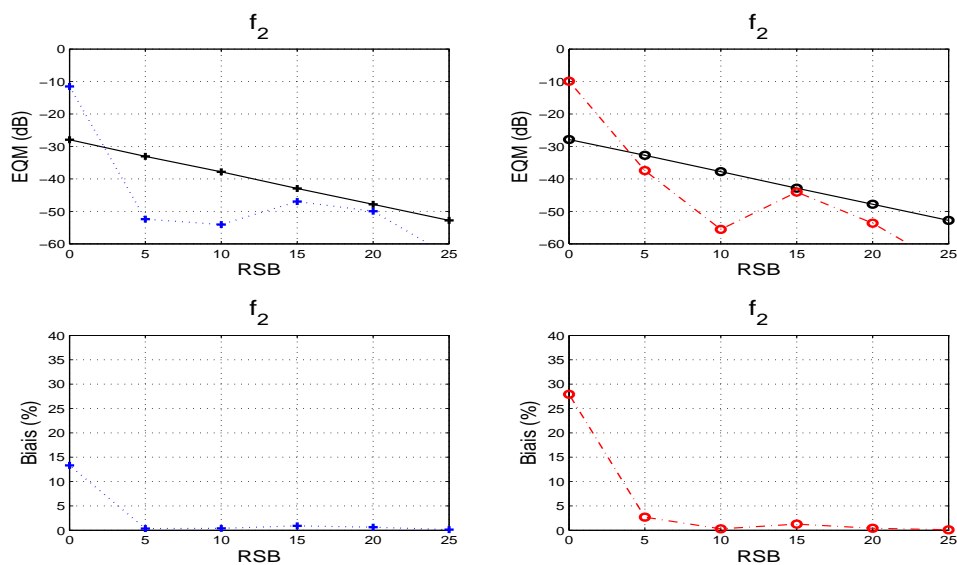
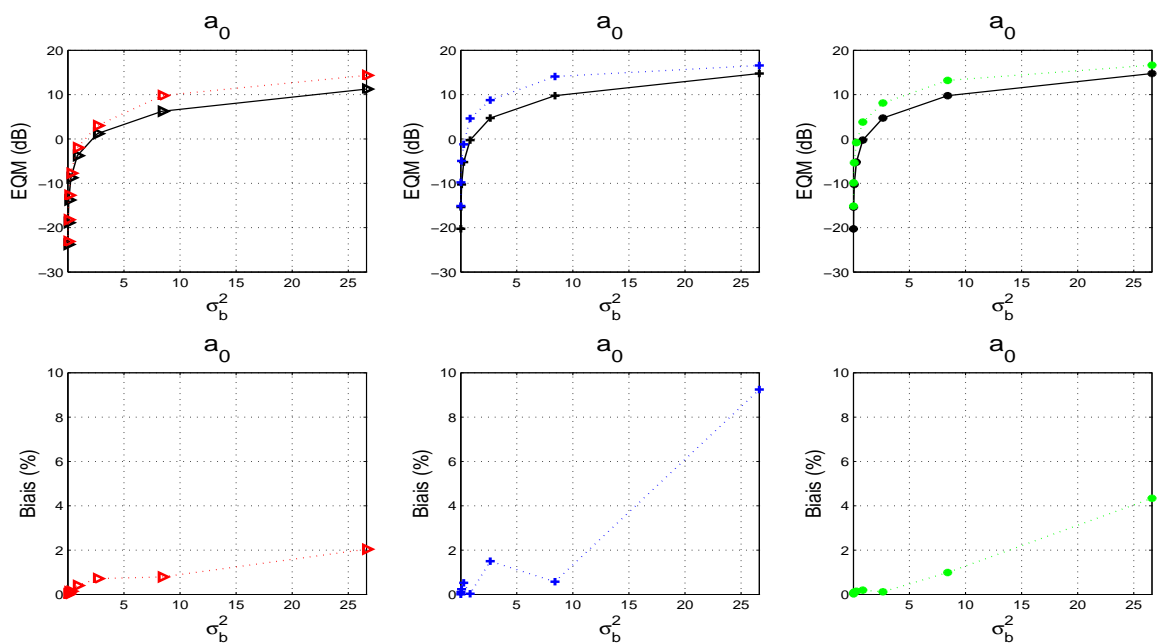


FIG. B.2 – BCR (5.15), EQM et biais de l'estimation du paramètre f_2 de la fréquence instantanée en fonction du RSB : Signal 3 EQM (..+) et BCR (-+) et Signal 4 EQM (-o) et BCR (-o).



Base discrète orthonormale

Base canonique

Base de Legendre discrétisée

FIG. B.3 – EQM et biais (%) de a_0 du Signal 1 obtenus en utilisant la base discrète orthonormale (.. Δ), la base canonique (..+) et la base de Legendre discrétisée (..*) comparés avec BCR correspondants (—).

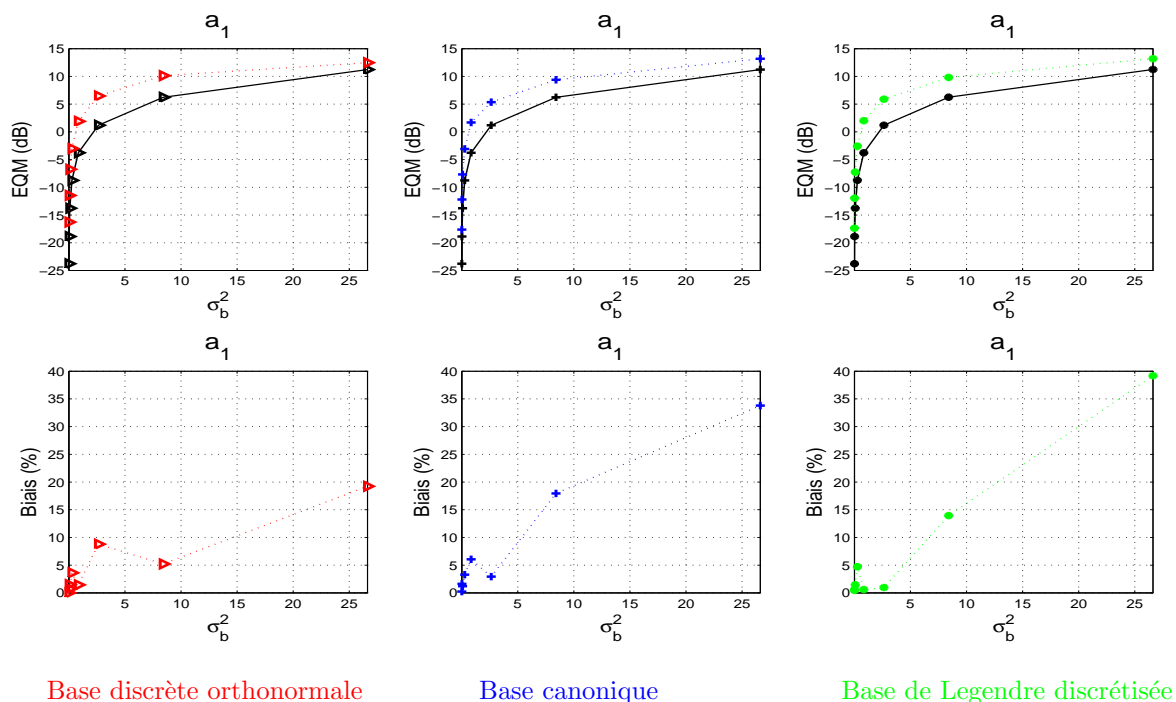


FIG. B.4 – EQM et biais (%) de a_1 du Signal 1 obtenus en utilisant la base discrète orthonormale (Δ), la base canonique (+) et la base de Legendre discrétisée (*) comparés avec BCR correspondants (—).

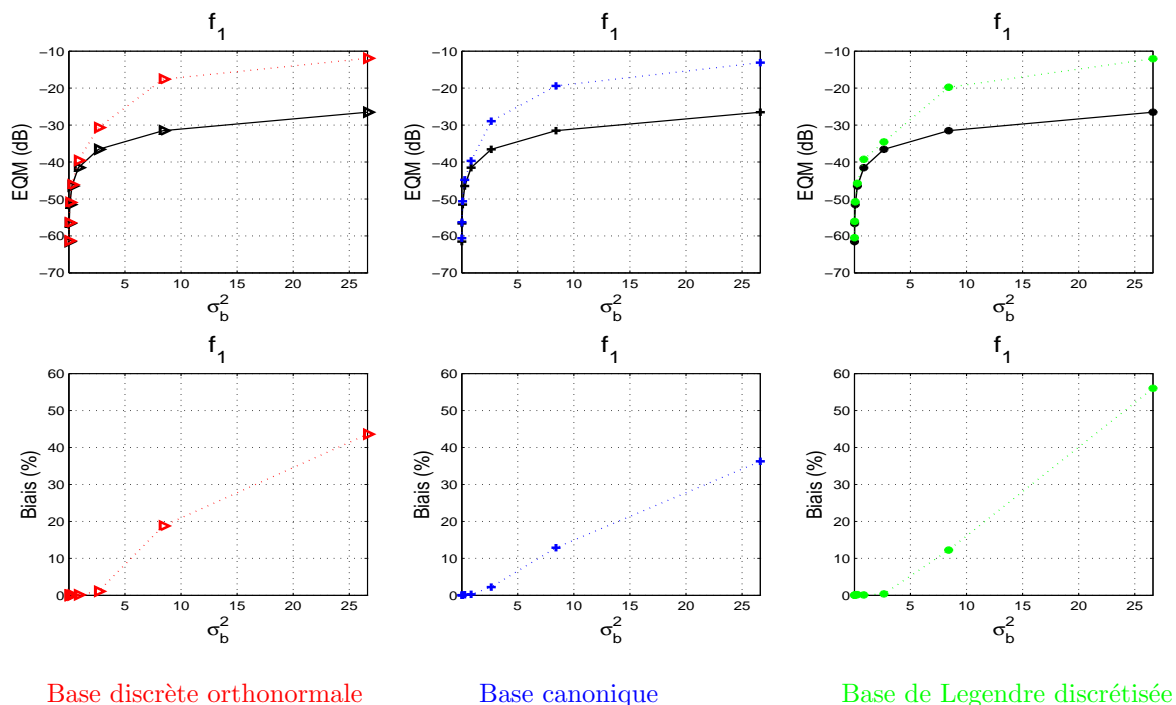


FIG. B.5 – EQM et biais (%) de f_1 du Signal 1 obtenus en utilisant la base discrète orthonormale (Δ), la base canonique (+) et la base de Legendre discrétisée (*) comparés avec BCR correspondants (—).

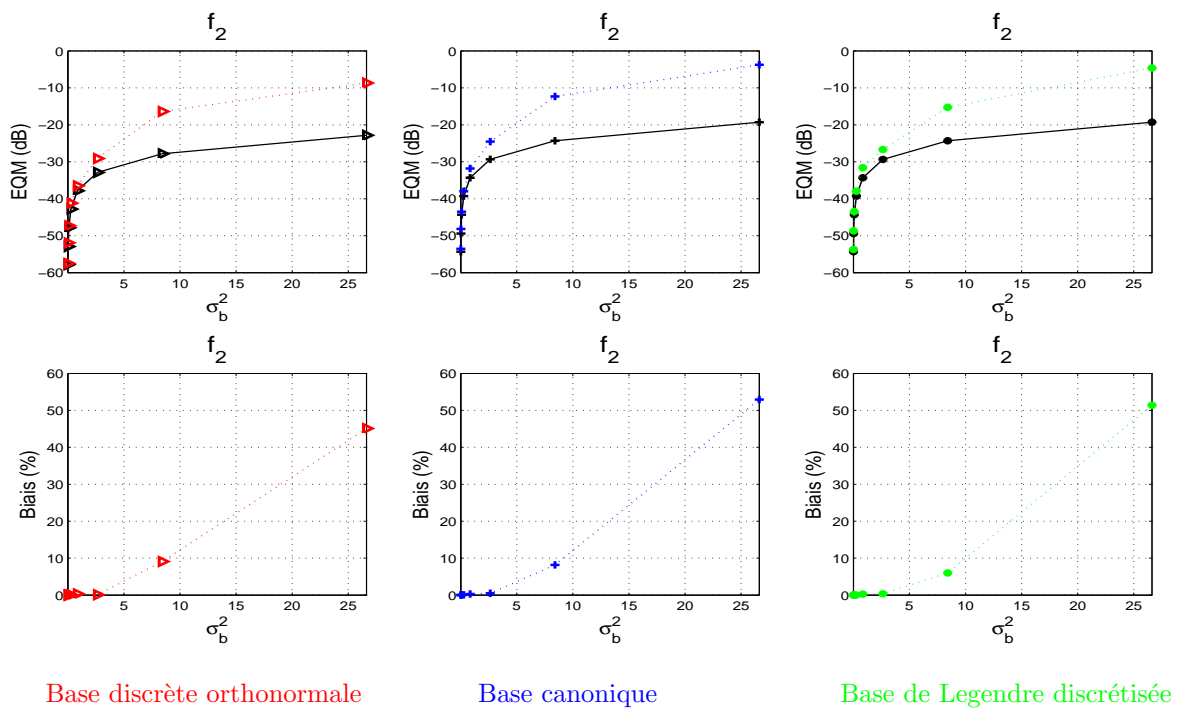


FIG. B.6 – EQM et biais (%) de f_2 du Signal 1 obtenus en utilisant la base discrète orthonormale (Δ), la base canonique (+) et la base de Legendre discrétisée (*) comparés avec BCR correspondants (—).

Bibliographie

- [Abu93] M.F. Aburdene. On the computation of discrete legendre polynomial coefficients. In *Multidimensional systems and signal processing*, pages 181–186, Kluwer Academic Publishers, Boston. Manufactured in The Netherlands, 1993.
- [Abu05] A. Abutaleb. Instantaneous frequency estimation when the amplitude is a stochastic process using stochastic calculus and bootstrapping. *Circuits Systems Signal Processing*, 24, No. 1 :35–52, 2005.
- [AD99] C. Andrieu et A. Doucet. Joint bayesian detection and estimation of noisy sinusoids via reversible jump mcmc. *IEEE Trans. Signal Processing*, 47 :2667–2676, 1999.
- [Ade97] R. M. Adelson. Frequency estimation from few measurements. *Digital Signal Processing*, 7 :47–54, 1997.
- [AF87] S. Anily et A. Federgruen. Simulated annealing methods with general acceptance probabilities. *Journal of Applied Probability*, 24 :657–667, 1987.
- [Aka83] H. Akaike. Information measure and model selection. *Bulltein INT. Statist. Inst.*, pages 50, 277–290, 1983.
- [Ang00] J. Angeby. Estimating signal parameters using the nonlinear instantaneous least squares approach. *IEEE Trans. on Signal Processing*, 48, No. 10, Oct. 2000.
- [AS64] M. Abramowitz et I.A. Stegen. Handbook of mathematical functions with formulas, graphs, and mathematical tables. Applied Mathematics Series 55, Chapter 22, Washington, D.C. : National Bureau of Standards, June 1964.
- [Bar93] S. Barbarossa. Detection and estimation of the instantaneous frequency of polynomial phase signals by multilinear time-frequency representations. In *Proc. IEEE SP Workshop Higher Order Stat.*, Lake Tahoe, CA, June 1993.
- [Bar95a] R. Baraniuk. Unitary equivalence : a new twist on signal processing. *IEEE Trans. on Signal Processing*, 43, No. 10 :1226–1234, Oct. 1995.
- [Bar95b] S. Barbarossa. Analysis of multicomponent lfm signals by a combined wigner-hough transform. *IEEE Trans. Signal Processing*, 43 :1511–1515, June 1995.
- [BB99] B. Barkat et B. Boashash. Instantaneous frequency estimation of polynomial FM signals using the peak of the PWVD : statistical performance in the presence of additive gaussian noise. *IEEE Trans. on Signal Processing*, 47, No. 9 :2480–2490, Sept. 1999.

- [BCSG04] A.O. Boudraa, J.C. Cexus, F. Salzenstein et L. Guillon. IF estimation using empirical mode decomposition and nonlinear teager energy operator. In *ICASSP*, pages 45–48, Montreal, Canada, 2004.
- [Ben04] M. Benedir. Robust polynomial wigner-ville distribution for the analysis of polynomial phase signals in α -stable noise. In *ICASSP*, pages 17–21, July 2004.
- [Bjo90] Bjornstad. Predictive likelihood : a review. *Staist. Science*, pages 5, 242–265, 1990.
- [BLP81] A. Blanc-Lapierre et B. Picinbono. Fonctions aléatoires. Ed. Masson, Paris, 1981.
- [BMS97] S. Barbarossa, R. Mamei et A. Scaglione. Adaptive detection of polynomial-phase signals embedded in noise using high order ambiguity functions. *Proc. 31st Asilomar Conf. Signals, Syst. Comput., Pacific Grove, CA*, Nov. 1997.
- [BPS96] S. Barbarossa, A. Porchia et A. Scaglione. Multiplicative multilag higher order ambiguity function. In *Proc. ICASSP*, Atlanta, May 7-11 1996.
- [Bro70] C.G. Broyden. The convergence of a class of double-rank minimization algorithms. Part I : General considerations. *Journal of the Institute of Mathematics and Its Applications*, 6 :76–90, 1970.
- [BS94] J.M. Bernardo et A.F.M. Smith. Bayesian theory. In *Editor Wiley*, New York, 1994.
- [BS04] B. Barkat et L. Stankovic. Analysis of polynomial FM signals corrupted by heavy-tailed noise. *Signal Processing*, 84 :69–75, 2004.
- [BSET97] CK.L. Bell, Y. Steinberg, Y. Ephraim et H.L.V Trees. Extended ziv-zakai lower bound for vector parameter estimation. *IEEE Trans. on information theory*, 43 :624–636, 1997.
- [BSG98] S. Barbarossa, A. Scaglione et G.B. Giannakis. Product high-order ambiguity function for multicomponent polynomial-phase signal modeling. *IEEE Trans. on Signal Processing*, 49 :691–708, March 1998.
- [CFL04] P. Ciblat, P. Forster et P. Larzabal. Harmonic retrieval in non-circular complex-valued multiplicative noise : Barankin bound. *EUSIPCO, Vienne, Autriche*, 2004.
- [CG95] S. Chib et E. GreenBerg. Understanding the Metropolis Hastings algorithm. *The American Statisticien*, pages 49 :327–335, 1995.
- [Chi96] S. Chib. Markov chain Monte Carlo simulation methods in econometrics. *Econometric theory*, pages 12 :409–431, 1996.
- [CL99] V. Chen et H. Ling. Joint Time-Frequency analysis for radar signal and image processing. In *IEEE Signal Processing Magazine*, pages 81–93, March 1999.
- [CLV99] L. Cohen, P. Loughlin et D. Vakman. On an ambiguity in the definition of the amplitude and phase of a signal. *Signal Processing*, 79 :301–307, 1999.

- [Coh66] L. Cohen. Generalized phase-space distributions functions. *Journal of Math. and Phys.*, 7, No. 5 :781–786, 1966.
- [Coh89] L. Cohen. Time-frequency distributions-A review. *Proceedings of the IEEE*, 77 :941–982, July 1989.
- [Coh93] L. Cohen. Time-frequency analysis. In *Pretince Hall*, New Jersey, 1993.
- [CPD03] H. Cottureau, J.M. Piasco et M. Davy. Tow approaches for the estimation of time-varying amplitude multi-chirp signals. In *ICASSP*, pages IV.657–IV.660, 2003.
- [CS95] C.K. Catchpole et P.J.B. Slater. Bird song : Biological themes and variations. Cambridge university press, 1995.
- [Dav75] P.J. Davis. Interpolation and approximation. Dover, 1975.
- [DCK00] R.M. Dizaji, N. Ross Chapman et R. Lynn Kirilin. Time-frequency matched field processing for time varying sources. In *IEEE Sensor Array and Multi-channel Signal Processing Workshop 2000*, pages 117–120, March 2000.
- [DDI05] C. Dubois, M. Davy et J. Idier. Tracking of time-frequency components using particle filtering. In *ICASSP*, pages IV.9–IV.12, Philadelphia, US, 2005.
- [Des05] M. Desbazaille. Représentations temps-fréquence appliquées à des signaux de bilan de masse de glaciers. *Rapport de stage de DEA au laboratoire des images et des signaux, sous la direction de N. Martin*, Février-Juin 2005.
- [DIS03] M. Dakovic, V.N. Ivanovic et L. Stankovic. On the S-method based instantaneous frequency estimation. In *ICASSP*, Hong Kong, 2003.
- [DMD05] M. Davy, P. Del Morel et A. Doucet. Méthodes Monte Carlo séquentielles pour l’analyse spectrale bayésienne. *GRETSI*, 2005.
- [DPST03] J. Dréo, A. Petrowsk, P. Siarry et E. Taillard. Métaheuristiques pour l’optimisation difficile. Eyrolles, 61,Bld Saint-Germain, Paris, July 2003.
- [DS99] I. Djurovic et L. Stankovic. Time-frequency representation based on the reasigned S-method. *Signal Processing*, 77 :115–120, 1999.
- [Ele04] C.P.H. Elemans. How do birds sing? sound analysis - mechanical modelling - muscular control. Ph. D thesis, Experimental Zoology Group, Wageningen University, The Netherlands, Nov. 2004.
- [FF95] B. Friedlander et J.M. Francos. Estimation of amplitude and phase parameters of multicomponent signals. *IEEE Trans. on Signal Processing*, 43 :917–926, Apr. 1995.
- [Fla93] P. Flandrin. Temps-fréquence. Editions Hermes, 14 rue Lantiez, Paris, 1993.
- [FP99] A. Francos et M. Porat. Analysis and synthesis of multicomponent signals using positive Time-Frequency distributions. *IEEE Trans. on Signal Processing*, 47 :493–504, Feb. 1999.
- [FP03] P. Flandrin et P.Gonçalves. Sur la décomposition modale empirique. In *GRETSI*, 2003.

- [Fri93] B. Friedlander. Parametric signal analysis using the polynomial phase transform. In *Proc. IEEE Workshop Higher-order Statistics*, pages 151–159, Lake Tahoe, CA, 7-9 June 1993.
- [Gab46] D. Gabor. Theory of communications. *Journal of Institute of Electrical Engineers IEE*, 93, No. 3 :429–457, Feb. 1946.
- [GBM⁺96] R. Gribonval, E. Bacry, S. Mallat, Ph. Depalle et X. Rodet. Analysis of sound signals with high resolution matching pursuit. In *Proc. of Third Int. Symposium on Time-Frequency and Time-Scale Analysis*, pages 125–128, June 1996.
- [GDL94] Z. Guo, L.G. Durand et H.C. Lee. Comparison of time-frequency distribution techniques for analysis of simulated doppler ultrasound signals of the femoral artery. *IEEE Trans. on Biomedical Engineering*, 41, No. 4, Apr. 1994.
- [Gey92] C.J. Geyer. Practical Markov chain Monte Carlo. *Statistical Science* 7(4), pages 473–511, 1992.
- [GF99] S. Golden et B. Friedlander. Maximum likelihood estimation, analysis and applications of exponential polynomials signals. *IEEE Trans. on Signal Processing*, 47 :1493–1501, June 1999.
- [GFR94] W.L. Goffe, G.D. Ferrier et J. Rogers. Global optimization and statistical functions with simulated annealing. *Journal of Econometrics*, 60(1-2) :65–99, 1994.
- [Gir99] J.M. Girault. Apport des techniques du traitement du signal à l’analyse et detection de signaux emboliques. Thèse de Doctorat de l’Université de Tours Signaux et Instrumentations-EA2102, Laboratoire d’Ultrasons, Dec. 1999.
- [GM05] T.J. Gardner et M.O. Magnasco. Instantaneous frequency decomposition : An application to spectrally sparse sounds with fast frequency modulations. *Journal Acoustic Society Am.*, 117 :2896, 2005.
- [GR92] A. Gelman et D.B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, pages 7(4) :457–511, 1992.
- [Gre68] C.H. Greenewalt. Bird song : Acoustics and physiology. Smithsonian Institution press, Washington, 1968.
- [GRS96] W.R. Gilks, S. Richardson et D.G. Spiegelhalter. Markov chain Monte Carlo in practise. *Chapman and Hall*, 1996.
- [GTC05] M. Gay, E. Thibert et C. Chauveau. Bilan de masse du glacier de sarnnes, 55ième et 56ième saisons 2002-2003-2004. *Campagne de mesure 2003-2004*, pages 1–7, 9 et 10 Mars 2005.
- [Her00] L. Herault. Rescaled simulated annealing- accelerating convergence of simulated annealing by rescaling the states energies. *Journal of Heuristics*, 6 :215–252, 2000.
- [HG89] B. Hajek et G.Sasaki. Simulated annealing- to cool or not. *Systems and control letters*, 12 :443–447, 1989.

- [HPSBB97] F. Hlawatsch, A. Papandreou-Suppappola et G. Faye Boudreaux-Bartels. The hyperbolic class of quadratic time-frequency representations - part II : subclasses, intersection with the affine and power classes, regularity and unitary. *IEEE Trans. on Signal Processing*, 45 :303–315, Feb. 1997.
- [HRA86] M.D. Huang, F. Romea et Sangiovanni-Vincentalli A. An efficient general cooling schedule for simulated annealing. In *IEEE International Conference on Computer Aided design*, pages 381–384, Santa Clara, March 1986.
- [HSL⁺98] N.E Huang, Z. Shen, S.R Long, M.L.Wu Shih, Q. Zheng, N.C. Yen, C.C Tung et H.H Liu. The empirical mode decomposition and Hilbert spectrum for non-linear and non-stationary times series analysis. In *Proc. Roy. Soc. London*, pages 903–955, 1998.
- [ICQ04] C. Ioana, C. Cornu et A. Quinquis. Polynomial phase signal processing via warped higher-order ambiguity function. In *Proc. of EUSIPCO*, pages 1159–1162, Vienna, Austria, Sept. 2004.
- [Ing93] L. Ingber. Simulated annealing : practice versus theory. *Mathematical and Computer Modelling*, 18 :29–57, Nov 1993.
- [Ioa03] C. Ioana. Contribution à la caractérisation des structures temps-fréquence non-linéaires. Thèse de doctorat de l’université de Bretagne occidentale, UFR sciences et techniques, ENSIETA-Bretagne, 2003.
- [IQ02] C. Ioana et A. Quinquis. Underwater transitory signal characterization using time-frequency warping operator. In *6th European Conference on Underwater acoustic - ECUA*, Gdansk, 2002.
- [IQ03] C. Ioana et A. Quinquis. High-order warped based ambiguity function. In *IEEE-EURASIP workshop on Non-linear Signal and Image Processing*, Grado, Italy, 2003.
- [IZ01] M.Z. Ikram et G.T. Zhou. Estimation of multicomponent polynomial phase signals of mixed orders. *Signal Processing*, 81 :2293–2308, Nov. 2001.
- [Jef61] H. Jeffreys. Theory of probability. *Oxford University Press, 3rd edition*, pages 115 :89–99, 1961.
- [JJDS02] J. Fine J. J Dreesbeke et GI Saporta. Méthodes Bayésiennes en statistiques. Editions TECHNIP, Paris, 2002.
- [JMVL05a] M. Jabloun, N. Martin, M. Vieira et F. Leonard. Maximum likelihood parameter estimation of short-time multicomponent signals with nonlinear AM/FM modulation. In *IEEE workshop on Statistical Signal Processing*, Bordeaux, France, July 2005.
- [JMVL05b] M. Jabloun, N. Martin, M. Vieira et F. Leonard. Multicomponent signal : Local analysis and estimation. In *EUSIPCO*, Antalya, Turkey, Sept 2005.
- [JMVL07a] M. Jabloun, N. Martin, M. Vieira et F. Leonard. Estimation of the instantaneous amplitude and frequency of non-stationary short-time signals. *Accepted with minor changes in Signal Processing*, 2007.

- [JMV07b] M. Jabloun, N. Martin, M. Vieira et F. Leonard. A new flexible approach to estimate highly nonstationary signals. *Accepted in IEEE Trans. on Signal Processing*, July 2007.
- [JVLM04] M. Jabloun, M. Vieira, F. Leonard et N. Martin. Local orthonormal decomposition for both instantaneous amplitude and frequency of highly nonstationary discrete signals. In *IMA : The Institute of Mathematics and Its Applications*, pages 107–110, Cirencester, United Kingdom, Dec 2004.
- [JVLM05] M. Jabloun, M. Vieira, F. Leonard et N. Martin. A AM/FM single component signal reconstruction using a nonsequential time segmentation and polynomial modeling. In *NSIP : IEEE workshop on Nonlinear Signal and Image Processing*, Sapporo, Japan, May 2005.
- [Kaa98] W. Kaakour. Modélisation et identification des signaux à phase polynomiale. Thèse de doctorat en Automatique et Informatique Appliquée, Ecole Centrale de Nantes, 1998.
- [Kat95] V. Katkovnik. Local polynomial periodogram for time-varying frequency estimation. *South African Statist. J.*, 29 :169–198, 1995.
- [Kat97] V. Katkovnik. Nonparametric estimation of instantaneous frequency. *IEEE Trans. on Signal Processing*, 43, No. 1 :183–189, Jan. 1997.
- [Kat00] V. Katkovnik. Minimax lower bounds for nonparametric estimation of the instantaneous frequency and time-varying amplitude of a harmonic signal. *Signal Processing*, 80 :577–595, 2000.
- [Kay88] S.M. Kay. Modern spectral estimation : Theory and application. Editor A.V. Oppenheim, Prentice Hall, Englewood Cliffs, New Jersey, 1988.
- [KGV83] S. Kirkpatrick, C.D. Gelatt et Jr.M.P. Vecchi. Optimization by simulated annealing. *Science*, 220-4598 :670–679, May 1983.
- [KS98] V. Katkovnik et L. Stankovic. Periodogram with varying and data-driven window length. *Signal Processing*, 67 :345–358, 1998.
- [LA87] P.J.M. Van Laarhoven et E.H.L. Aarts. Simulated annealing : Theory and applications. Kluwer Academic Publishers, 1987.
- [LAC97] J.L. Lacoume, P.O. Amblard et P. Comon. Statistiques d'ordre supérieur pour le traitement du signal. Ed. Masson, Paris, 1997.
- [Leo00] F. Leonard. Spectrogramme de phase et spectrogramme de fréquence. *Traitement du signal*, 17, No 4 :269–286, 2000.
- [LL05] K. Lu et X. Liu. Enhanced visibility of maneuvering targets for high-frequency over-the-horizon radar. *IEEE Trans. on Antennas and Propagation*, 53 :404–411, Jan. 2005.
- [Loc00] M. Locatelli. Simulated annealing algorithm for continuous global optimization : convergence conditions. *Journal of Optimization theory and applications*, 104(1) :121–154, 2000.

- [Loc01] M. Locatelli. Convergence and first hitting time of simulated annealing algorithm for continuous global optimization. *Mathematical Methods of Operations research, Springer-Verlag*, 54 :171–199, 2001.
- [Mar77] S. Martin. Analyse et reconstitution de la série des bilans annuels du glacier de sarnnes, sa relation avec les fluctuations du niveau de trois glaciers du Massif du Mont Blanc (Bossons, Argentière, Mer de Glace. *Z. Gletscherkd. Glazialgeol*, 13(1-2) :127–153, 1977.
- [Mar07] N. Martin. Advanced signal processing and condition monitoring session - Keynote adress. *ICWMA*, 2007.
- [MC97] P.R.S. Mendonça et L.P. Calôba. New simulated annealing algorithms. In *IEEE Internatinal Synposium on Circuits and Systems*, pages 1668–1671, Hong Kong, 1997.
- [Mic91] L. Miclo. *Evolution de l'énergie libre. Application à la convergence des algorithmes du recuit simulé*. Université de Paris-6, 1991.
- [Mit86] Mitra. D. mitra and f. romeo and a. sangiovanni-vicentelli. *Journal of Applied Probability*, 18 :747–771, 1986.
- [MKQ93] P. Maragos, J.F. Kaiser et T.F. Quatieri. Energy separation in signal modulations with application to speech analysis. *IEEE Transactions on Signal Processing*, 41, No. 10 :3024–3051, Oct. 1993.
- [ML05] G.B. Mindlin et R. Laje. *The physics on birdsong*. Springer, Heidelberg, Germany, 2005.
- [MMB93] D. Maiwald, J.W. Dalle Molle, et J.F. Böhme. Model identification and validation of nonstationary seismic signal. In *IEEE Signal Processing Workshop on High-Order Statistics*, pages 319–322, June 1993.
- [MMM06] K. Miettinen, M.M. Makela et H. Maaranen. Efficient hybrid methods for global continuous optimization based on simulated annealing. *Computers and operations research, elseiver*, 33 :1102–1116, 2006.
- [Mor69] N. Morrison. *Introduction to sequential smoothing and prediction*. McGraw-Hill Book Company, New York, Second Edition, Vol 1 1969.
- [MRR⁺53] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller et E. Teller. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21 :1087–1092, June 1953.
- [MT93] S.P. Meyn et R.L. Tweedie. *Markov chains and stochastic stability*. Springer-Verlag, London, 1993.
- [MZ02] M.R. Morelande et A.M. Zoubir. Model selection of random amplitude polynomial phase signals. *IEEE Trans. on Signal Processing*, 50, No. 3 :578–589, March 2002.
- [NGH⁺01] L. Nolle, A. Goodyear, A.A. Hopgood, P.D. Picton et N.S. Braithwaite. On step width adaptation in simulated annealing for continuous parameter optimisation. In *B. Reusch (Ed.) : Fuzzy Days 2001, Springer-Verlag*, pages 589–598, Berlin Heidelberg, 2001.

- [O'S04] P. O'Shea. A fast algorithm for estimating the parameters of a quadratic FM signal. *IEEE Trans. on Signal Processing*, 52, No. 2 :385–392, Feb 2004.
- [Pic96] B. Picinbono. Second-order complex random vectors and normal distributions. *IEEE Trans. on Signal Processing*, 44, No. 10 :2637–2640, Oct. 1996.
- [Pin62] M. Pincus. A closed form solution for certain programming problems. *Oper. Res.*, pages 690–694, 1962.
- [PKBB93] A. Papandreou, S.M. Kay et G.F. Boudreaux-Bartels. The hyperbolic class of quadratic time-frequency representation - part-I : Constant-Q warping, the hyperbolic paradigm properties and members. *IEEE Trans. on Signal Processing*, 41 :3425–3444, Dec. 1993.
- [PMT96] L.R. Padoverse, N. Martin et J.M. Terriez. Temps-fréquence pour l'identification des caractéristiques dynamiques d'un pylône de téléphérique. *Traitement du Signal*, 13 :211–223, 1996.
- [Por93] B. Porat. Digital processing of random signals. In *Pretince Hall*, New Jersey, 1993.
- [PP91] S. Peleg et B. Porat. Estimation and classification of polynomial-phase signals. *IEEE Trans. on Informations Theory*, 37 :422–430, March 1991.
- [PPF93] S. Peleg, B. Porat et B. Friedlander. The achievable accuracy in estimating the instantaneous phase and frequency of a constant amplitude. *IEEE Trans. on Signal Processing*, 71 :2216–2224, June 1993.
- [QDMH92] T.F. Quatieri, R.B. Dunn, R.J. McAulay et T.E. Hanna. Underwater signal enhancement using a sine-wave representation. In *IEEE Oceans '92 Proceedings*, pages 449–454, Oct. 1992.
- [QM91] T.F. Quatieri et R.J. McAulay. Peak-to-RMS reduction of speech based on a sinusoidal model. *IEEE Transactions on Signal Processing*, 39, No. 2 :273–288, Feb 1991.
- [RB74] D. Rife et R. Boorstyn. Single-tone parameter estimation from discrete-time observations. *IEEE Trans. on information theory*, IT-20, 1974.
- [RFL05] A. Renaux, P. Forster et P. Larzabal. A new deraivation of the bayesian bounds for parameter estimation. *IEEE workwhop on Statistical Signal Processing : SSP, Bourdeaux, France.*, 2005.
- [RGG95] G.O. Roberts, A. Gelman et W. Gilks. Weak convergence and optimal scaling of random walk Metropolis algorithms. *Technical report, Staistical Laboratory, University of Cambridge*, pages 16–94, 1995.
- [Rob96] C. P. Robert. Méthodes de Monte Carlo par chaînes de markov. Economica, Paris, 1996.
- [Rob06] C. P. Robert. Le choix Bayésien : Principes et pratiques. Springer-Verlag France, Paris, 2006.

- [Rod97] X. Rodet. Musical sound signal Analysis/Synthesis : Sinusoidal+residual and elementary waveform models. In *IEEE Time-Frequency ant Time-Scale Workshop 97*, August 1997.
- [RT96] G. O. Roberts et R. L. Tweedie. Geometric convergence and central limit theorems for multidimensional hastings and metropolis algorithms. *Biometrika*, Vol. 83, pages 95–110, 1996.
- [SB99] A. Scaglione et S. Barbarossa. Statistical analysis of the Product High-order Ambiguity Function. *IEEE Trans. on Information Theory*, 45, No. 1, Jan. 1999.
- [SD89] P. Siarry et G. Dreyfus. *La méthode du recuit simulé : théorie et applications*. 10 rue Vauquelin, 75005 Paris, 1989.
- [SDT05] L. Stankovic, M. Dakovic et T. Thayaparan. Algorithm for signal decomposition by using the S-method. In *EUSIPCO*, Antalya, Turkey, 2005.
- [SFD98] M.L. Dent S.M. Farabaugh et R.J. Dooling. Hearing and vocalizations of wild-caught australian budgerigars (*melopsittacus undulatus*). *Journal of Comparative Psychology*, 112, No. I :74–81, 1998.
- [Sia94] P. Siarry. *La méthode du recuit simulé en électronique. Comparaison avec d'autres méthodes d'optimisation. Application dans d'autres domaines*. Université de Paris-sud (Orsay), 1994.
- [SIP96] L. Stankovic, V. Ivanovic et Z. Petrovic. Unified appraoch to noise analysis in the wigner distribution and spectrogram. *Ann. Télécommun.*, 51, No. 11-12 :585–594, 1996.
- [SK02] S. Saha et S.M. Kay. Maximum likelihood parameter estimation of superimposed chirps using Monte Carlo importance sampling. *IEEE Trans. on Signal Processing*, 50, No.2 :224–230, Feb. 2002.
- [Sén02] S. Sénécal. Méthodes de simulation monte-carlo par chaînes de markov pour l'estimation de modèles. Application en séparation de sources et en égalisation. Thèse de doctorat, Institut National Polytechnique de Grenoble, 2002.
- [Sta94] L. Stankovic. A method of time-frequency analysis. *IEEE Trans. on Signal Processing*, 42, No. 1 :225–229, Jan. 1994.
- [Sta98] L. Stankovic. On the realization of the polynomial wigner-ville distribution for multicomponent signals. *IEEE Signal Processing Letters*, 5, 1998.
- [Tas89] P. Tassi. *Méthodes statistiques*. 49, rue Héricat, 75015, 1989.
- [Tay58] A.E. Taylor. Functional analysis. Wiley Editor, 1958.
- [Tie94] L. Tierney. Markov chains for exploring posterior distributions. *Annals of Statics*, pages 22 :1701–1762, 1994.
- [Tre68] H.L.V. Trees. Detection estimation and modulation theory, part I. Massachusetts institute of technology, Jhon Wiley and Sons, 1968.

- [Tre76] S.A. Treutter. Introduction to discrete-time signal processing. Jhon Wiley and sons, New York, 1976.
- [TVF97] C. Theys, M. Vieira et A. Ferrari. Bayesian estimation of the parameters of a polynomial phase signal using MCMC methods. *ICASSP*, pages 3553–3556, 1997.
- [TVG99] C. Theys, M. Vieira et G. Alenquin. A reversible jump sampler for polynomial phase signals. *ICASSP*, pages 3553–3556, 1999.
- [Vak96] D. Vakman. On the analytic signal, the Teager-Kaiser energy algorithm and other methods for defining amplitude and frequency. *IEEE Trans. on Signal Processing*, 44, No. 4 :791–797, 1996.
- [VB00] F. Vincent et O. Besson. Estimating time-varying DOA and doppler shift in radar array processing. *IEE Proc. Radar, Sonar Navigations*, 147, No. 6, Dec 2000.
- [Vie99] M. Vieira. Estimation Bayésienne par des méthodes MCMC : Application à la surveillance des moteurs asynchrones. Université de Nice- Sophia-Antipolis, Thèse de doctorat, 1999.
- [Vil48] J. Ville. Théorie et applications de la notion du signal analytique. *Câbles et transmissions, 2ème A.*, No. 1 :61–74, 1948.
- [Vil58] J. Ville. Theory and applications of the notion of complex signal. *translated by I. Seline in RAND Tech. Rpt, Santa Monica, CA.*, T. 92, 1958.
- [Vin01] C. Vincent. Influence of climate change over the 20th century on four french glacier mass balances. *Journal of Geophysical Research*, 107, no D19 :104–108, 2001.
- [VL84] D. Vanderbilt et SG. Louise. A Monte Carlo simulated annealing approach to optimization over continuous variables. *Journal of Computational Physics*, 56 :259–271, 1984.
- [VLJM04] M. Vieira, F. Leonard, M. Jabloun et N. Martin. Short time single polynomial phase using Legendre functions. In *Proc. of EUSIPCO*, pages 793–796, Vienna, Austria, Sept 2004.
- [Vra03] V. Vrabie. Statistiques d’ordre supérieur : Applications en géophysique et en électrotechnique. Thèse de Doctorat à l’INPG, Grenoble, 2003.
- [VVR95] M. Vallo, C. Vincent et L. Reynaud. Sensibilité des bilans de masse glaciaires aux fluctuations climatiques. *La Houille Blanche*, 5/6 :104–108, 1995.
- [ZA97] A.M. Zoubir et I. Ameer. Bootstrap analysis of polynomial amplitude and phase signals. In *IEEE TENCON - Speech and Image Technologies for Computing and Telecommunications*, pages 843–846, 1997.
- [ZGS96] G. Zhou, G.B. Giannakis et A. Swami. On polynomial phase signals with time-varying amplitudes. *IEEE Trans. on Signal Processing*, 44, No. 4 :848–861, Apr. 1996.

Résumé

Ce travail de recherche est consacré à l'élaboration et le développement d'une nouvelle méthode d'estimation et de reconstruction de signaux fortement non-stationnaires, modulés non-linéairement à la fois en amplitude et en fréquence. L'estimation de tels signaux dans un contexte très bruité est un problème délicat et les méthodes existantes de la littérature présentent plusieurs inconvénients dans ce cas.

Nous avons montré comment une approche locale permet une meilleure adaptabilité du modèle à la nature des variations locales des amplitudes et des fréquences instantanées. Les résultats de l'estimation sont par conséquent améliorés. L'originalité de la méthode proposée tient à l'application de modèles paramétriques bien adaptés sur des segments temporels de courtes durées extraits du signal étudié. Nous avons proposé une stratégie de segmentation puis une stratégie de fusion des segments estimés permettant la reconstruction du signal dans la totalité de sa durée. L'approche proposée permet de s'affranchir d'un modèle global du signal requérant un ordre d'approximation élevé.

La validation de l'efficacité de l'estimation a été effectuée au préalable sur un segment temporel court. Le modèle considéré localement consiste en une approximation polynomiale de la fréquence et de l'amplitude exprimée dans une base polynomiale discrète et orthonormale que nous avons calculée. Cette base permet de réduire le couplage entre les paramètres du modèle. Nous proposons et comparons deux techniques différentes pour estimer ces derniers. La première est fondée sur la maximisation de la fonction de vraisemblance en utilisant la technique d'optimisation stochastique le recuit simulé. Tandis que la deuxième se base sur une approche Bayésienne employant les méthodes MCMC simulées par l'algorithme de Metropolis-Hastings.

Nous montrons, sur des simulations et également sur des signaux réels, que l'approche proposée fournit de bons résultats d'estimation par comparaison à celles de la HAF.

Mots-clés : Signaux non stationnaires, estimation paramétrique, modélisation polynomiale, représentations temps-fréquence, optimisation numérique, recuit simulé, méthodes MCMC, bornes de Cramer-Rao.

Abstract

This work concentrates on the estimation and reconstruction of highly non-stationary signals having both non-linear amplitude and frequency modulations. We propose a new method for estimating the above mentioned signals with high additive noises. Most of the previously published work in this domain is plagued by shortcomings and at times even fails to address the problem of estimation.

We have shown that by using a local approach, which provides flexibility to fit to the local variations of the instantaneous amplitude and frequency, the estimation results are improved significantly. The novelty of the presented work consists in the use of parametric models well-adapted and defined on short time segments. We have also proposed a new technique for obtaining short time segments from the entire signal. Finally, we propose a merging strategy to reconstruct the entire signal along with its modulations. The proposed estimation approach is great interest as it does not require a higher order model to estimate the entire signal.

Initially, we validate the estimation efficiency of the local model considering a short time segment. This local model uses polynomial approximations of both local instantaneous amplitude and frequency, decomposed using a discrete orthonormal base which we derived. We also affirm that this base reduces the coupling of model parameters. Hence, the estimation accuracy is enhanced. Furthermore, we have compared two techniques to estimate the model parameters. The first approach is based upon the maximization of the likelihood function by means of an optimization stochastic technique named Simulated Annealing. The second approach uses the MCMC methods with Metropolis-Hastings algorithm.

We demonstrate using simulations and real applications that the proposed approach improves both performance and robustness of signal estimation as compared to the existing Higher Ambiguity function based technique.

Keywords : Non-stationary signals, parametric estimation, polynomial modeling, Time-frequency representation, optimization, simulated annealing, MCMC methods.
