



HAL
open science

Blancheur et non-gaussianité pour la déconvolution aveugle de données brutiées : application aux signaux sismiques

Anthony Larue

► **To cite this version:**

Anthony Larue. Blancheur et non-gaussianité pour la déconvolution aveugle de données brutiées : application aux signaux sismiques. Traitement du signal et de l'image [eess.SP]. Institut National Polytechnique de Grenoble - INPG, 2006. Français. NNT : . tel-00097161

HAL Id: tel-00097161

<https://theses.hal.science/tel-00097161>

Submitted on 21 Sep 2006

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

THÈSE

pour obtenir le grade de

DOCTEUR DE L'INPG

Spécialité : « Signal Image Parole et Télécoms »

préparée au Laboratoire des Images et des Signaux de Grenoble

dans le cadre de l'École Doctorale « Électronique, Électrotechnique,
Automatique, Traitement du Signal »

présentée et soutenue publiquement

par

Anthony LARUE

le 13 septembre 2006

Titre :

**BLANCHEUR ET NON-GAUSSIANITÉ POUR LA
DÉCONVOLUTION AVEUGLE DE DONNÉES BRUITÉES :
APPLICATION AUX SIGNAUX SISMIQUES.**

Directeurs de thèse : Jérôme I. MARS et Christian JUTTEN

JURY

Madame	O. Macchi,	Présidente
Monsieur	J-M. Boucher,	Rapporteur
Monsieur	G. Demoment,	Rapporteur
Monsieur	D-T. Pham,	Examinateur
Monsieur	M. Van Der Baan,	Examinateur
Monsieur	J. I. Mars,	Directeur de thèse
Monsieur	C. Jutten,	Co-directeur de thèse

Remerciements

Je remercie tout d'abord mes deux directeurs de thèse qui ont suivi mon travail, pour leur soutien et leurs encouragements pendant ces trois années.

- Jérôme Mars, pour m'avoir expliqué les rudiments de la géophysique, pour ses perpétuels encouragements à la publication des résultats et son aide pour parfaire mon manuscrit.
- Christian Jutten, pour ses conseils d'expert et ses relectures très pointues de ma thèse et des différents articles.

Je leur suis très reconnaissant pour leur idée de réaliser un séjour de six mois à Leeds. Cette proposition est passée d'un exil peu attrayant à une expérience très enrichissante.

Je remercie également les membres de mon jury de thèse de m'avoir fait l'honneur d'évaluer mon travail et pour l'ensemble de leurs remarques pour l'amélioration de mon mémoire :

- Mme Odile Macchi qui a accepté de présider mon jury et pour son implication dans la séance de questions lors de ma soutenance,
- M. Jean-Marc Boucher pour avoir rapporté mon travail et pour ses questions percutantes,
- M. Guy Demoment pour son rapport d'expert du domaine de la déconvolution, et ses nombreuses remarques,
- M. Dinh Tuan Pham pour sa collaboration sur des aspects théoriques pendant ma thèse qui ont permis les développements du chapitre 5 et ses questions très pointues lors de ma soutenance,
- M. Mirko Van Der Baan pour l'intérêt qu'il a porté à mon travail et son accueil lors de mon séjour à Leeds.

Ensuite, je me dois de remercier un grand nombre de personnes pour ces trois années studieuses et animées passées au LIS. Tout d'abord, Jérôme pour son dynamisme et sa bonne humeur pour animer le groupe SIN, mais aussi pour nos longues discussions sportives et plus précisément sur l'ASSE. Cependant, je ne comprendrai jamais sa double passion verte et marseillaise !!

Je poursuivrai par les membres du bureau 138 pour les bons moments partagés, jugés bruyants par nos voisins. Chouchou pour ses passages furtifs dans le bureau entre la BU et la salle de karaté... Sebastian pour son pessimisme exacerbé... c'est sûr il ne nous a pas remonté le moral souvent !! Ma phrase préférée entendue au milieu de ma rédaction : "tu sais après (la rédaction) c'est encore pire".

Merci aussi à Barbara pour sa bonne humeur quasi permanente et l'année passée en collocation ponctuée par de nombreuses soirées. Pour une électricienne tu cuisines pas si mal que cela!! Cédric Gérot très certainement mon ami le plus improbable au départ, je suis heureux d'avoir participé à ton baptême du chaudron. Caroline pour les nombreuses pauses café partagées, les discussions matinales à notre arrivée et nos profondes analyses de l'humeur marsienne. Benoît pour nos débats de footex qui énervaient certains! Le chaudron c'est cool, ben ? Bertrand, le seul colocataire à la table basse frappée de la pomme! Julien pour son art de mélanger les épices. Ensuite, la liste est longue pour tous ceux avec qui j'ai partagé les pauses café, les repas au RU, les sandwiches de la cafet ou de longues discussions : Nicolas, Vincent, Pierre, Grégoire, Mathieu, Matthieu, Jeremy, Sébastien H., Isabelle, Éric, Maxime, François et de façon plus large tous les membres du laboratoire qui franchissent fréquemment la porte de la cafet. Enfin, Marion qui a le mérite de partager des soirées avec tous ces scientifiques. J'ai aussi une petite pensée pour les indispensables informaticiens du laboratoire Hervé et Jean-Marc.

Je remercie aussi tous ceux que j'ai croisés lors de mon séjour à Leeds. Tout d'abord, Raoul mon colocataire atypique pour sa patience à comprendre mon anglais très spécial... pour toutes les soirées à la maison et au pub etc... saches que la chartreuse ce n'est pas un alcool très fort. Ensuite, tous mes co-bureaux et collègues pour leur gentillesse et leurs efforts pour m'aider à améliorer mon anglais au bureau ou au pub et pour leur initiation au poker : James, Andy, James, Andy, Nicola, James, Ash, Lykke, Katya ... Merci aux francophones : Kristoff, Marielle, Nicolas pour les pauses repas en français... Un merci plus particulièrement à Christine pour les nombreuses virées touristiques organisées lors de mon séjour là bas. Le nord de l'Angleterre c'est pluvieux comme prévu mais c'est beau aussi.

Comme on oublie toujours des gens, ou que certains tiennent à être remerciés, cette phrase conçue sur le modèle des albums panini (les spécialistes comprendront) est pour eux. Je remercie _____ pour _____.

Je remercie enfin le lecteur car sans lui cette thèse perd beaucoup de sens, en espérant tout de même qu'il ne s'arrête pas sur cette page... beaucoup d'autres pages passionnantes suivent celle-là... courage.

J'ai d'énormes pensées envers ma famille pour leur soutien depuis le début de mes études. Mes parents pour m'avoir encouragé et toujours poussé vers le haut depuis le début. Mon frère pour les moments partagés depuis notre enfance et plus récemment pendant les week ends reposants à la campagne. J'ai une petite pensée aussi pour Odile mon institutrice à l'école primaire qui m'a appris à lire il y a longtemps et qui m'a fait l'amabilité de venir à ma soutenance.

Pour finir, je remercie amoureuxment Anouck, mon rayon de soleil depuis un an, pour sa relecture minutieuse de ma thèse et surtout pour tous les moments de bonheurs partagés.

Table des matières

Introduction	11
1 Imagerie sismique et déconvolution par les statistiques d'ordre supérieur	15
1.1 L'imagerie sismique	17
1.1.1 Acquisition sismique	17
1.1.2 Modèle convolutif pour les données	21
1.1.3 Approches "simples"	22
1.2 Statistiques d'ordre supérieur pour les variables et vecteurs aléatoires	23
1.2.1 Variables aléatoires scalaires	24
1.2.2 Vecteurs aléatoires réels	28
1.2.2.1 Définition générale	28
1.2.2.2 Propriétés des moments et cumulants	30
1.3 Les statistiques d'ordre supérieur pour l'analyse des signaux monodimensionnels	32
1.3.1 Multicorrélation et multispectre	33
1.3.2 Effet du filtrage sur les multicorrélations et les multispectres	38
1.4 État de l'art de la déconvolution	39
1.4.1 Filtrage de Wiener : ondelette connue	40
1.4.2 Déconvolution semi-aveugle : ondelette inconnue, phase connue	41
1.4.2.1 Égalisation spectrale	42
1.4.2.2 Méthodes dans le domaine temporel	42
1.4.3 Déconvolution aveugle : ondelette (module et phase) inconnue	46
1.4.3.1 Domaine fréquentiel	46
1.4.3.2 Domaine temporel	47
1.4.4 Résumé	50
2 Déconvolution aveugle dans le domaine temporel par le taux d'information mutuelle	53
2.1 Développements théoriques	55
2.2 Algorithme MAMV de déconvolution d'ondelette AR	57
2.2.1 Gradient du critère	57
2.2.2 Algorithme MAMV	58
2.2.3 Données simulées	60
2.3 Algorithme ARMV de déconvolution d'ondelette MA	62
2.3.1 Gradient du critère	63

2.3.2	Algorithme ARMV	64
2.3.3	Données simulées	65
2.4	Estimation de la fonction score	66
2.4.1	Estimateur de fonction score	67
2.4.1.1	Estimateur à noyaux	67
2.4.1.2	Estimation directe	68
2.4.2	Étude des performances des estimateurs	70
2.4.2.1	Fonctions scores théoriques	71
2.4.2.2	Comparaison des estimateurs	72
2.5	Conclusion	75
3	Le taux d'information mutuelle pour la déconvolution de données brui-	
	tées	77
3.1	Applications et performances de l'algorithme MAMV	79
3.1.1	Données réelles : explosions séismo-volcaniques	79
3.1.2	Performances de l'algorithme MAMV en présence de bruit gaussien .	82
3.2	Déconvolution par le taux d'information mutuelle dans le domaine fréquentiel	86
3.2.1	Critère de déconvolution	86
3.2.2	Gradient du critère et algorithme FBD	87
3.2.3	Réglages des hyperparamètres	90
3.2.4	Données simulées	91
3.3	Traitement de données réelles	93
3.3.1	Données réelles : explosions séismo-volcaniques	93
3.3.2	Données réelles : explosions sous-marines	96
3.4	Conclusions	98
4	Déconvolution par non-gaussianité	101
4.1	Déconvolution par non-gaussianité	104
4.1.1	Mesure de non-gaussianité : la néguentropie	104
4.1.2	Déconvolution par maximisation de la néguentropie : algorithme MA-	
	Nege	105
4.1.3	Résultats de simulations	107
4.2	MANege : un contexte général pour des méthodes existantes	109
4.2.1	Rappel sur les lois gaussiennes généralisées	110
4.2.2	Déconvolution par maximisation du kurtosis : algorithme MED . . .	112
4.2.3	Déconvolution par variation de norme : algorithme VARIMAX	115
4.2.4	Transformation exponentielle	118
4.2.5	Algorithme de Godfrey	120
4.3	Contraintes d'identifiabilité sur les critères	121
4.4	Concavité et estimation non biaisée	122
4.4.1	Contraintes théoriques	122
4.4.2	Cas de l'algorithme MANege	124
4.4.3	Cas des algorithmes MED et VARIMAX	125
4.5	Variance asymptotique des algorithmes	129
4.5.1	Calcul théorique	130
4.5.2	Borne minimale	131

4.5.3	Courbes de performances	132
4.6	Comparaison des algorithmes	135
4.6.1	Fonction score	135
4.6.2	Résultats de simulations	137
4.7	Résumé et conclusions	138
5	Robustesse en contexte bruité	143
5.1	Influence du bruit sur les critères	145
5.1.1	Algorithme MED : maximisation du kurtosis	146
5.1.2	Algorithme MANege : maximisation de la néguentropie	147
5.1.3	Algorithme FBD : minimisation du taux d'information mutuelle.	151
5.2	Expérimentations	152
5.2.1	Signaux simulés	152
5.2.2	Données réelles : explosions séismo-volcaniques	155
5.2.3	Données réelles : explosions sous-marines	158
5.2.4	Données réelles : profil sismique	159
5.3	Conclusions	163
	Conclusions et perspectives	165
	Annexes	171
A	Développement théorique relatif au chapitre 2	173
A.1	Preuve du lemme 2.1	173
A.2	Calcul de (2.15)	174
A.3	Preuve du lemme 2.2	175
B	Calcul des fonctions scores théoriques	177
B.1	Préliminaires	177
B.2	Cas de la réflectivité laplacienne	178
B.3	Cas de la réflectivité Bernoulli-gaussienne	179
B.4	Cas de réflectivité Bernoulli-laplacienne	180
B.5	Cas de la réflectivité uniformément distribuée	181
C	Règles de dérivation et gradient par rapport à une variable complexe	183
C.1	Définitions et relations générales	183
C.2	Définition du gradient	184
D	Performances asymptotiques de l'algorithme MANege	185
D.1	Matrice de covariance	185
D.2	Approximations et simplifications	187
	Bibliographie	189

Notations

$r(t)$	Signal de réflectivité
$w(t)$	Ondelette
$n(t)$	Bruit additif
$g(t)$	Filtre déconvolveur
$y(t)$	Sortie estimée
$X(\nu)$	Transformée de Fourier de $x(t)$
$C_{nX}(\cdot)$	Multicorrélation d'ordre n de X
$S_{nX}(\cdot)$	Multispectre d'ordre n de X
$I(Y)$	Taux d'information mutuelle de Y
$H(x)$	Entropie de x
$H^-(X)$	Néguentropie du processus X
\mathcal{K}_{4Y}	Kurtosis de Y
$p_Y(y)$	Densité de probabilité de Y
$\psi_Y(y)$	Fonction score de Y
\mathbf{a}	Vecteur a
\mathbf{A}	Matrice A

Introduction

Une des problématiques classiques en traitement du signal est l'inversion d'un modèle ou l'extraction de paramètres d'un modèle. Les domaines d'application sont très variés : télécommunications, imagerie (radar, sismique, rayons X, médical...), contrôle non destructif, astronomie, optique, électromagnétisme, automatique... En général, nous disposons d'un modèle direct construit à partir de considérations physiques. Le modèle physique peut être de forme et de complexité très variées. Néanmoins, il apparaît très souvent sous une forme assez générale comprenant un ou plusieurs signaux d'entrée, un système qui opère sur ces entrées pour nous fournir un ou plusieurs signaux de sortie appelés également observations. Le système est aussi appelé la relation entrée-sortie. Puis, nous cherchons à inverser ce modèle par des outils de traitement du signal, c'est-à-dire d'estimer les entrées et/ou le système. L'estimation peut également ne concerner que certains paramètres de ces deux parties du modèle. L'objectif du problème inverse est souvent d'augmenter la qualité des observations pour faciliter l'interprétation, ou d'extraire des grandeurs caractéristiques du modèle, ou même de permettre une décision.

L'inversion du modèle peut être placée dans le cadre plus général de l'estimation. La résolution d'un problème d'estimation commence par la spécification du modèle. De multiples modèles peuvent être mis en place (modèle déterministe ou comportant des composantes aléatoires et/ou des paramètres aléatoires). Une fois le modèle posé, nous souhaitons ensuite estimer ses parties encore indéterminées. Pour cela, les algorithmes d'estimation s'appuient sur un critère qui est construit en fonction des propriétés des composantes du modèle (signal, relation entrée-sortie, bruit...). Classiquement, la forme de la relation entrée-sortie est supposée connue. En revanche pour les signaux, nous choisissons souvent l'hypothèse de signaux aléatoires. Cette hypothèse permet de s'intéresser seulement aux signaux du point vue statistique et non déterministe, ce qui minimise les *a priori* faits sur les signaux.

Parmi les problèmes qui utilisent cette modélisation, nous pouvons citer les problèmes de séparation de sources ou d'analyses en composantes indépendantes. Pour ceux-ci la relation entrée-sortie est un mélange qui peut être de plusieurs types : linéaire instantané, linéaire convolutif, post non-linéaire... Nous notons aussi que le modèle de signal aléatoire peut varier, nous pouvons opter pour des modèles iid, markoviens, parcimonieux...

Si la relation entrée-sortie est un filtrage, nous avons un modèle direct convolutif. Nous considérerons qu'il n'y a qu'une entrée et qu'une sortie auquel cas le filtrage déforme le signal d'entrée et appauvrit son spectre, réduisant ainsi la résolution des observations. Nous appelons problème de déconvolution l'inversion d'un tel modèle direct convolutif. L'objectif de la déconvolution est d'augmenter la résolution des observations, nous souhaitons limiter les hypothèses sur le modèle direct ainsi nous optons pour la déconvolution dite aveugle,

i.e. utilisant seulement les données et des hypothèses statistiques sur le signal d'entrée.

Les signaux observés en imagerie sismique suivent ce modèle convolutif. Comme l'objectif de l'imagerie sismique est d'obtenir une cartographie du sous-sol, la déconvolution est un traitement pertinent pour augmenter la résolution de ces cartes. Les enjeux d'une imagerie de qualité sont importants dans plusieurs domaines d'application dont le génie civil, mais le domaine le plus porteur est celui de l'exploration pétrolière. L'amélioration de la résolution des données sismiques permet de diminuer les coûts des campagnes d'acquisition et facilite aussi l'interprétation. Remarquons que la technique d'investigation utilisée est également proche de celle développée dans de nombreuses applications d'imagerie (biomédical ou de contrôle non destructif). Du point de vue physique et expérimental, une source émet une onde dans le sol, et les changements de structure du sous-sol provoquent une réflexion/transmission de l'onde émise. La mesure de ces échos sur une antenne de capteurs et l'étude des variations d'amplitude et des variations de fréquence des ondes observées permet d'obtenir des informations sur la structure du sous-sol. Dans ce cadre, les signaux sismiques sont modélisés par la convolution de la séquence de réflectivité et de l'ondelette émise. La séquence de réflectivité porte l'information relative à la structure du sous-sol. L'ondelette représente la source (explosion, camion vibreur...). L'application de la déconvolution aux données sismiques a pour but de supprimer l'effet de l'ondelette et donc d'augmenter la résolution des données. Malheureusement, le modèle doit être complété avec un bruit additif sur les données pour être réaliste. Classiquement, une méthode de déconvolution sera performante si elle n'amplifie pas trop ce bruit additif. Ainsi, le compromis entre la déconvolution et l'amplification du bruit doit être un critère fort lors du développement de nouveaux algorithmes.

Pour la déconvolution, la communauté du traitement du signal a pour habitude d'utiliser l'hypothèse de blancheur du signal de réflectivité. Les premières méthodes (égalisation spectrale, Yule-Walker, Durbin) sont basées sur un modèle *a priori* gaussien des distributions, car elles utilisent seulement les statistiques d'ordre 2. Ces algorithmes sont simples, mais ne permettent pas l'estimation de la phase de l'ondelette. Cette estimation nécessite les statistiques d'ordre supérieur à 2. Pour des signaux avec des distributions symétriques, nous utilisons généralement les statistiques d'ordre 4 (tricorrélation, trispectre). Aucune justification n'a été jusqu'à maintenant apportée à ce choix de statistiques, hormis la théorie des statistiques d'ordre supérieur. Notre travail, qui s'inscrit dans ce cadre de la déconvolution aveugle, s'attache à résoudre le problème du choix des statistiques employées afin qu'elles soient adaptées au signal à déconvoluer. Nous proposons d'aborder ce problème avec des mesures statistiques générales, c'est-à-dire qui ne se limitent pas à des ordres supérieurs pré-déterminés. Tout d'abord, nous mesurerons l'indépendance statistique (au sens général) entre les échantillons à l'aide du taux d'information mutuelle. Une seconde approche sera fondée sur la non-gaussianité et l'utilisation de la néguentropie, qui mesure la déviation à la gaussienne avec la divergence de Kullback.

Ce mémoire s'organise de la façon suivante :

Le premier chapitre s'attache à présenter le principe de l'imagerie sismique. Ensuite,

nous présenterons les outils de statistiques d'ordre supérieur (cumulants, moments, multicorréaltions, multispectres...) nécessaires à notre approche du problème de déconvolution. Nous établirons un état de l'art des méthodes de déconvolution aveugle basées sur la blancheur du signal de réflectivité.

Le chapitre 2 débute par la présentation d'une mesure de blancheur utilisant toutes les statistiques d'ordre supérieur : le taux d'information mutuelle. Ensuite, nous décrirons deux algorithmes (MAMV et ARMV) de déconvolution aveugle utilisant le taux d'information mutuelle comme mesure de blancheur et un modèle paramétrique (soit MA, soit AR) temporel pour le filtre déconvolveur. Nous nous intéresserons aussi au problème d'estimation de la fonction score posé par ces algorithmes.

Le troisième chapitre met en évidence le problème du comportement des algorithmes en fonction du bruit. Nous proposons un nouvel algorithme FBD (Frequency Blind Deconvolution) de déconvolution aveugle en fréquence utilisant le taux d'information mutuelle et comportant un terme pour contrôler le rapport signal sur bruit de sortie.

Le quatrième chapitre marque une rupture par rapport aux développements précédents. Nous proposons de placer la blancheur comme hypothèse secondaire et de nous intéresser à la non-gaussianité de la réflectivité pour l'ajustement du filtre déconvolveur. Dans le cas des signaux sismiques, nous utilisons plus précisément la sur-gaussianité. Nous mesurons la déviation à la gaussienne à l'aide de la néguentropie et sa maximisation fera l'objet de l'algorithme MANeige. Nous aurons une longue discussion sur les propriétés des différents algorithmes de la littérature basés sur la sur-gaussianité.

Le dernier chapitre permet de caractériser à la fois en théorie mais aussi en pratique le comportement des différents algorithmes en fonction du bruit additif sur les données. Ce travail permettra principalement de déterminer laquelle des deux hypothèses de blancheur ou de sur-gaussianité est plus pertinente pour la déconvolution de données sismiques.

Des résultats de simulations et de traitements de données réelles seront présentés tout au long de ce mémoire. En particulier, le dernier point de ce manuscrit concernera la validation des méthodes existantes et surtout celles développées dans cette thèse sur un profil sismique réel.

Chapitre 1

Imagerie sismique et déconvolution par les statistiques d'ordre supérieur

Sommaire

1.1	L'imagerie sismique	17
1.1.1	Acquisition sismique	17
1.1.2	Modèle convolutif pour les données	21
1.1.3	Approches "simples"	22
1.2	Statistiques d'ordre supérieur pour les variables et vecteurs aléatoires	23
1.2.1	Variables aléatoires scalaires	24
1.2.2	Vecteurs aléatoires réels	28
1.2.2.1	Définition générale	28
1.2.2.2	Propriétés des moments et cumulants	30
1.3	Les statistiques d'ordre supérieur pour l'analyse des signaux monodimensionnels	32
1.3.1	Multicorrélation et multispectre	33
1.3.2	Effet du filtrage sur les multicorrélations et les multispectres	38
1.4	État de l'art de la déconvolution	39
1.4.1	Filtrage de Wiener : ondelette connue	40
1.4.2	Déconvolution semi-aveugle : ondelette inconnue, phase connue	41
1.4.2.1	Égalisation spectrale	42
1.4.2.2	Méthodes dans le domaine temporel	42
1.4.3	Déconvolution aveugle : ondelette (module et phase) inconnue	46
1.4.3.1	Domaine fréquentiel	46
1.4.3.2	Domaine temporel	47
1.4.4	Résumé	50

Avant de s'intéresser aux problèmes de déconvolution aveugle, il est important de présenter le principe des acquisitions sismiques pour parvenir au modèle des observations en fonction de l'ondelette émise et de la réflectivité. Nous insisterons sur les spécificités de ce domaine d'applications qui devront être prises en compte lors de la construction des algorithmes de déconvolution dans les chapitres suivants. Nous détaillons les outils statistiques nécessaires aux techniques de déconvolution, notamment les multicorrélations et les multispectres pour la mesure de blancheur d'un signal. Enfin, en utilisant ces outils statistiques, nous dresserons un état de l'art aussi complet que possible des techniques existantes.

1.1 L'imagerie sismique

L'objectif de l'imagerie sismique est d'obtenir une cartographie 2D ou 3D de la structure du sous-sol. Cette technique est similaire à toutes les techniques d'imagerie de type échographie. L'exploration sismique consiste à générer des ondes et à mesurer les échos de ces ondes après leur propagation dans le sol. Par exemple, la connaissance du temps de propagation et de la vitesse de propagation entre la source et les récepteurs va apporter des renseignements sur les propriétés (physiques, mécaniques, pétrologiques) des structures traversées. Évidemment, la motivation première de ce type d'imagerie est la localisation d'hydrocarbures pour les compagnies pétrolières. Mon travail de thèse s'inscrit d'ailleurs dans le projet BLInd Seismic Signals (BLISS) soutenu par l'Industry Technology Facilitator (ITF) qui est un consortium de compagnies pétrolières. Ce projet fédère l'Université de Leeds, le Laboratoire des Images et Signaux (LIS) et le Laboratoire de Modélisation de Calcul (LMC) de Grenoble. Dans cette partie, nous allons décrire le principe de l'acquisition sismique terrestre et marine pour introduire le modèle convolutif des données sismiques liant la réflectivité et l'ondelette émise.

1.1.1 Acquisition sismique

La prospection sismique consiste à produire des séismes artificiels (source) de très faible amplitude à des instants et des endroits prédéterminés. Les séismes provoqués sont enregistrés sur des dispositifs récepteurs. Suivant le type d'acquisition, il y a des spécificités en terme de dispositif de mesure et de type d'ondes mises en jeu. Nous proposons de séparer le cas d'acquisition terrestre (onshore) et d'acquisition marine (offshore).

Acquisition terrestre

Différents types de dispositifs sont utilisables dans le domaine de la prospection sismique terrestre. L'objectif de la campagne d'acquisition ainsi que la position de la cible géologique (profondeur, taille, résolution) dictent le type de sismique à mettre en oeuvre [Yil87, MGC99, Yil00a].

Les différents types de sources utilisés en prospection sismique terrestre sont :

- les chutes de masses qui sont de moins en moins utilisées à cause de leur faible maniabilité,

- les explosifs qui fournissent un signal bref temporellement mais demandant un temps de manipulation élevé. De plus, l'ondelette émise est mal contrôlée,
- les camions vibreurs qui permettent d'émettre un signal large bande assez bien maîtrisé dans le sous sol. Lors d'une campagne d'Imagerie à Haute Résolution (IHR) au puits des Goules en Auvergne réalisée par le LGIT de Grenoble [Bre05] à laquelle j'ai participé, la fréquence du *sweep* utilisé variait de quelques hertz (5 ou 7 Hz) à 80 Hz. Les figures 1.1 et 1.2 montrent des photos d'un camion vibreur et de la plaque vibrante du camion lors de cette campagne.



FIG. 1.1 – Camion vibreur pour l'acquisition sismique terrestre. FIG. 1.2 – Plaque vibrante du camion vibreur.

Les capteurs permettant de mesurer les vibrations résultantes des ondes sismiques sont des géophones. Classiquement, ils sont composés d'un aimant fixe autour duquel une bobine mobile peut coulisser. L'onde sismique provoque le mouvement de la bobine dont résulte un courant induit dans la bobine. Les géophones sont organisés en antenne uniforme. Pour une acquisition 2D ils sont placés sur une antenne rectiligne, et pour une acquisition 3D, nous les disposons sous forme d'un réseau plan. Pour collecter l'ensemble des données, les géophones sont connectés à des flûtes disposées le long de l'antenne, qui sont elles-mêmes connectées au camion enregistreur. La gamme de fréquence de résonance des géophones est comprise entre 2 et 35 Hz, classiquement la fréquence de résonance est de 10 Hz. Pour des besoins d'acquisition haute résolution, la gamme de la fréquence de résonance peut être augmentée jusqu'à 100Hz. Les derniers géophones MEMS (capteurs numériques) de la société Sercel permettent même d'avoir une réponse totalement plate entre 1Hz et 100Hz.

La sismique de surface est la technique la plus simple à mettre en oeuvre, avec une source et un réseau de capteurs à la surface. Nous obtenons une image de la structure du terrain à analyser en terme de temps de trajet. Lorsque l'onde sismique arrive sur une interface géologique, elle se réfléchit, se transmet à la couche sous-jacente ou se réfracte. Une interface géologique est généralement identifiée par un changement d'impédance acoustique.

Si l'angle d'incidence dépasse l'angle critique i_c , nous avons une acquisition appelée sismique de réfraction (figure 1.4). Dans ce cas, l'onde réfractée se propage le long de l'interface, et réémet tout le long de celle-ci une onde vers la surface avec le même angle critique. Lorsque l'angle d'incidence est inférieur à l'angle critique, il y a réflexion de l'onde (figure 1.3), nous avons alors une acquisition de sismique de réflexion. En sismique de surface, des ondes de surfaces, aussi appelées *ground-roll*, très énergétiques se propagent parallèlement à la surface. L'antenne de capteurs est beaucoup plus distante de la source en sismique réfraction (long offset) qu'en sismique réflexion (short offset). L'onde réfractée peut se propager sur plusieurs kilomètres le long de l'interface.

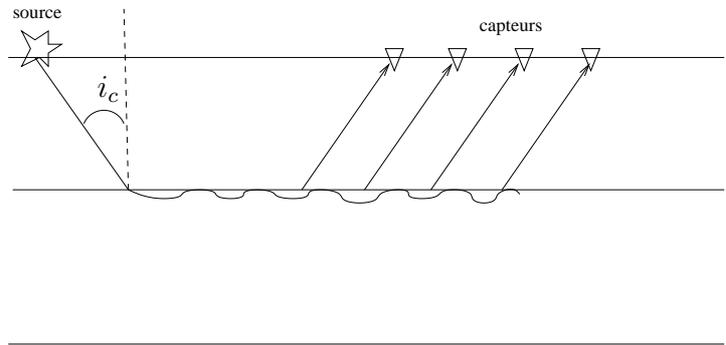
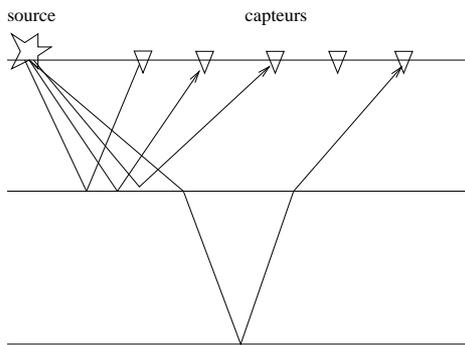


FIG. 1.3 – Campagne de sismique terrestre réflexion. FIG. 1.4 – Campagne de sismique terrestre réfraction.

Dans les zones jugées très intéressantes, les compagnies pétrolières ont recouru à la technique de sismique de puits. Généralement, l'antenne de réception est positionnée verticalement dans un puits creusé dans la terre. La source est toujours située en surface. Dans cette configuration, les ondes de surface n'altèrent pas la mesure mais des ondes de tube peuvent être générées à cause de l'existence d'une discontinuité entre la roche et le puits. Si nous plaçons la source à la verticale du puits, nous avons un Profil Sismique Vertical (PSV) sinon nous disposons d'une acquisition dite Profil Sismique Oblique (PSO). Nous avons parfois plusieurs puits afin d'explorer précisément la zone située entre les puits : nous parlons dans ce cas de sismique inter-puits. La sismique de puits est une technique beaucoup plus précise que la sismique de surface. La résolution est de l'ordre du mètre contre une dizaine de mètres pour les techniques sismiques réflexion et réfraction. Il est évident que la sismique de puits donnera des images de meilleure résolution (quelques mètres) que la sismique de surface. En revanche, le problème des puits réside dans le coût et le temps de forage qui sont très importants. Un des objectifs des compagnies pétrolières est de réduire la part de la sismique de puits en améliorant la précision de la sismique de surface notamment avec des nouvelles méthodes telles que la déconvolution.

Dans un premier temps en sismique, nous nous intéressons uniquement aux ondes de compression (ondes P) qui donnent un mouvement de particule parallèle à la direction de propagation de l'onde. Il existe aussi les ondes de cisaillement (ondes S) qui procurent une

vibration dans le plan perpendiculaire à la direction de propagation. Elles se divisent en deux catégories : les ondes SH (horizontale) et SV (verticale). Pour les ondes SH (respectivement SV) le mouvement de la particule se fait dans le plan perpendiculaire (respectivement parallèle) au plan de propagation. Nous parlons donc d'ondes polarisées. Les ondes de cisaillement (S) sont d'amplitudes plus faibles que les ondes de compression (P).

L'utilisation de géophones mesurant les mouvements suivant trois axes est de plus en plus répandue dans les systèmes d'acquisitions. Nous parlons de capteurs multicomposantes (vectoriels).

Lorsqu'une onde de type P ou SV atteint une interface géologique, elle donne en général naissance à quatre ondes, deux ondes transmises (P et SV) et deux ondes réfléchies (P et SV). Ce phénomène est appelé conversion. Ainsi, nous voyons que les enregistrements sur plusieurs composantes des géophones nous apportent une information plus riche (polarisation) avec un niveau de redondance permettant en premier lieu d'améliorer le rapport signal sur bruit.

Acquisition marine

De très nombreuses campagnes d'exploration sont réalisées en mer. En prospection sismique sous-marine, nous disposons des deux techniques de sismique réflexion et réfraction. La figure 1.5 illustre les principes de l'acquisition.

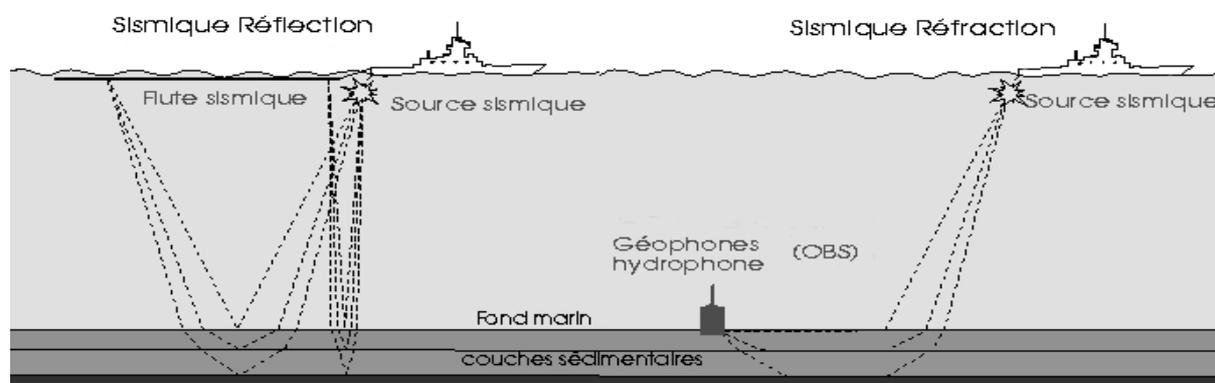


FIG. 1.5 – Acquisitions sismiques marines.

Ces deux techniques font intervenir les mêmes phénomènes physiques que pour la sismique terrestre. Pour la sismique réflexion marine, les capteurs utilisés sont des hydrophones qui mesurent des variations de pression par le biais d'un matériau piézoélectrique. Les hydrophones sont placés à intervalle régulier dans une flûte (câble ou *streamer*) tracté par le bateau à une faible profondeur (quelques mètres).

En sismique réfraction, le système de mesure est posé ou enfoui au fond de la mer. Les capteurs utilisés sont constitués généralement d'un géophone à 3 composantes et d'un hydrophone. Nous parlons dans ce cas d'un capteur 4 composantes (4C). Cet ensemble de mesures est appelé OBS (Ocean Bottom Sensor). Si les éléments sont reliés dans un câble

alors l'antenne d'OBS placée au fond de la mer est nommée OBC (Ocean Bottom Cable). Parfois, pour faire des acquisitions dans des lieux non accessibles par un bateau tractant un OBC (par exemple proche d'une plate-forme pétrolière), nous utilisons des OBS indépendants, ils sont appelés *Node*.

Pour la source, l'objectif est de créer une explosion sous-marine. La source la plus courante est un canon à air qui émet une bulle d'air sous pression dont l'explosion dans l'eau va générer une onde acoustique. L'utilisation d'un ensemble de canons à air permet d'augmenter la bande de l'ondelette envoyée (les fréquences les plus basses sont de l'ordre de 6 Hz).

1.1.2 Modèle convolutif pour les données

Pour la prospection sismique, il est intéressant de modéliser les données acquises en fonction de l'ondelette émise qui subit des modifications dues à la propagation dans le sol. En considérant les hypothèses suivantes :

- les réflecteurs dans le sous-sol sont horizontaux,
- la vitesse de propagation est constante dans les différentes couches (milieu homogène),
- l'onde émise est plane,
- les réflexions se font à incidence quasi-normale,
- l'ondelette ne change pas lors de son trajet dans le sous-sol, et elle est stationnaire,

alors une modélisation très simple consiste à écrire le signal reçu $d(t)$ comme une somme de N échos (réflexions) successifs de l'onde émise $w(t)$ dans le sol [Rob85, LY98] :

$$d(t) = \sum_{i=1}^N a_i w(t - t_i) + n(t), \quad (1.1)$$

où $n(t)$ est le bruit, a_i représente l'amplitude des coefficients de réflexion qui dépendent des impédances des différentes couches géologiques. La durée de trajet t_i nous renseigne sur la profondeur du réflecteur. Nous appelons séquence de réflectivité ou réflectivité le signal $r(t)$ comportant l'ensemble des réflecteurs. Il contient la signature des variations d'impédance du sous-sol :

$$r(t) = \sum_{i=1}^N a_i \delta(t - t_i). \quad (1.2)$$

Ainsi, nous obtenons le modèle convolutif (écrit pour les signaux échantillonnés) :

$$d(t) = w \star r(t) + n(t) = \sum_{\tau} w(\tau) r(t - \tau) + n(t), \quad (1.3)$$

\star est le produit de convolution, $w(t)$ représente l'ondelette émise, $r(t)$ la réflectivité, $d(t)$ l'observation et $n(t)$ est un bruit additif englobant l'ensemble des bruits physiques et de mesure. En fait, le modèle de réflectivité discrète proposé en (1.2) sous forme de somme

d'impulsions de Dirac est trop limitatif et est étendu classiquement à une réflectivité continue contenant des impulsions brèves et fortes [RT80, Yil87, Yil00b, RT01]. Le modèle de réflectivité continu obtenu est plus flexible. Ensuite, l'ondelette contient l'onde émise mais on peut y ajouter le couplage entre la source et le sous-sol et l'ensemble des effets de la propagation.

Le modèle utilisé par la suite est résumé par la figure 1.6 :

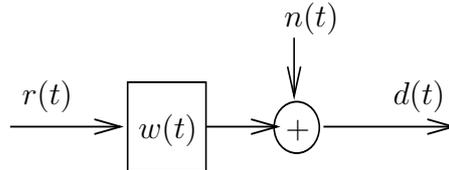


FIG. 1.6 – Modèle convolutif des données sismiques.

Dans le modèle physique, la réflectivité apparaît comme le filtre opérant sur l'ondelette. La convolution étant une opération commutative, nous préférons placer la réflectivité comme entrée du filtrage et l'ondelette comme filtre. De plus, ce changement est fait car les hypothèses statistiques de déconvolution aveugle seront faites sur la réflectivité $r(t)$ donc il est plus naturel de la placer comme excitation du système.

1.1.3 Approches "simples"

Des méthodes de déconvolution [Lav91, Lav93, IG93] utilisent une propriété très spécifique de la réflectivité pour estimer simultanément la réflectivité et l'ondelette. La réflectivité est modélisée suivant (1.2) comme une somme d'impulsions de Dirac pondérées et décalées. L'estimation complète de la réflectivité est équivalente à l'estimation des amplitudes des réflecteurs a_i et de leurs positions t_i . Dans un cadre statistique pour estimer a_i et t_i , la parcimonie temporelle de la réflectivité est décrite par un modèle Bernoulli-gaussien. Le signal de réflectivité est alors modélisé comme le produit de deux processus aléatoires :

$$r(t) = \gamma(t)q(t), \quad (1.4)$$

où $q(t)$ est un processus gaussien à moyenne nulle avec une variance σ^2 , et $\gamma(t)$ est une séquence de Bernoulli de paramètre λ , soit :

$$Pr[\gamma(t) = 0] = 1 - \lambda \quad \text{et} \quad Pr[\gamma(t) = 1] = \lambda. \quad (1.5)$$

Le processus de Bernoulli $\gamma(t)$ modélise la localisation temporelle t_1, \dots, t_N des réflecteurs alors que le processus gaussien $q(t)$ porte l'information sur l'amplitude des réflecteurs a_1, \dots, a_N . En général, les méthodes de déconvolution travaillent en deux étapes : (1) une étape de détection avec estimation des positions t_1, \dots, t_N des réflecteurs, (2) une étape d'estimation de l'amplitude des réflecteurs a_1, \dots, a_N qui donne lieu aussi à l'estimation de l'ondelette. Par exemple, Kormylo et Mendel [KM82] ont proposé une détection-estimation par le maximum de vraisemblance. La recherche de la position des réflecteurs n'est pas faite de façon exhaustive car elle impliquerait un temps de calcul trop important. Lavielle

[Lav93] a ajouté certains *a priori* pour déconvoluer un processus Bernoulli-gaussien dans un contexte bayésien. Ensuite l'extension au cas 2-D tient compte de la continuité spatiale avec un modèle markovien [IG93, Lav91]. L'approche bayésienne avec l'estimation par le Maximum a Posteriori (MAP) peut être résolue de façon itérative dans une boucle de filtrage de Kalman. Le cas de la représentation MA est traitée dans [GD89] et est affinée dans [IG90] avec une résolution sous-optimale plus rapide avec un algorithme de Viterbi. La généralisation aux modèles ARMA est donnée dans [GD87]. Une autre estimation sous optimale du problème MAP est possible à l'aide d'une méthode comportant une fenêtre glissante [Kaa98]. Le domaine d'application est élargi aux images astronomiques du télescope Hubble. Dans un contexte stochastique, le problème est résolu avec une méthode dédiée aux données incomplètes [Lav95, CGI96]. Cette approche est prolongée [DD97] par une modélisation dans l'espace des états pour l'estimation d'une ondelette AR et aussi pour une ondelette MA [RB99].

Ces algorithmes utilisent un modèle trop contraignant de la réflectivité. En effet dans des données réelles, les changements d'impédances n'apparaissent pas aussi clairement que le modèle Bernoulli-Gaussien le prétend et un niveau de bruit important gêne considérablement l'étape de détection. Cette prise de décision va se faire au détriment d'informations importantes cachées dans le bruit. Ainsi, une méthode de déconvolution aveugle plus générale peut être une bonne alternative. L'objectif de la déconvolution est de comprimer l'ondelette afin de mieux séparer deux échos successifs, soit augmenter la résolution des données. Une des hypothèses classiques est de supposer que la réflectivité est un processus dont les échantillons sont identiquement et indépendamment distribués (iid). Mais, avant de détailler ces méthodes de déconvolution, nous devons mettre en place les outils statistiques nécessaires. Nous débuterons par le cas des variables aléatoires réelles centrées. Ensuite, nous poursuivrons par le cas des vecteurs aléatoires réels pour parvenir à la définition des fonctions de multicorrélations et multispectres, qui seront les outils de base pour la mesure de blancheur d'un processus.

1.2 Statistiques d'ordre supérieur pour les variables et vecteurs aléatoires

Dans cette partie, nous allons définir les moments et cumulants d'ordre supérieur d'une variable aléatoire scalaire et vectorielle. Dans chaque cas, nous nous attacherons à montrer leur intérêt et nous intéresserons aux techniques d'estimation. Le cas scalaire nous permettra de caractériser un processus stochastique à un instant donné. Mais, en déconvolution aveugle, la blancheur représente les liens entre différents instants d'un signal. Ainsi nous serons amenés à nous intéresser aux outils statistiques d'ordre supérieur (moments, cumulants, fonction caractéristique) de vecteurs de variables aléatoires réels. A partir de ces outils, il en découle les fonctions de multicorrélations et les multispectres qui caractérisent la blancheur du signal aux ordres supérieurs dans les domaines fréquentiel et temporel. Nous nous limiterons par la suite à des variables, vecteurs ou processus aléatoires **centrés**. En effet, la moyenne est vue comme une grandeur déterministe qui peut être omise dans un cadre statistique. Les différents outils statistiques présentés dans cette partie ont été étudiés

en détail dans de nombreux ouvrages [BLF53, KS63, Bri81, Cul87, Com95, LAC97].

1.2.1 Variables aléatoires scalaires

Moments et première fonction caractéristique

Soit X une variable aléatoire réelle, que nous supposons **centrée** pour la suite afin d'éviter la distinction entre les moments centrés et non centrés. La fonction de répartition de la variable aléatoire X est notée $F_X(u)$. Elle est définie comme la probabilité d'apparition de la variable aléatoire X dans l'intervalle $] -\infty, u]$:

$$F_X(u) = \text{Proba}\{x \leq u\}. \quad (1.6)$$

Dans le cas d'une variable aléatoire continue, la densité de probabilité (ddp) notée $p_X(x)$ de la variable aléatoire X est :

$$p_X(x) = \frac{d}{dx} F_X(x).$$

Le moment d'ordre r de la variable aléatoire X est défini par :

$$\mu_{rX} = E[X^r] = \int_{-\infty}^{+\infty} x^r p_X(x) dx, \quad (1.7)$$

où $E[.]$ représente l'opérateur d'espérance mathématique. Le moment d'ordre 1 est la moyenne. Elle est nulle pour une variable aléatoire centrée. Le moment d'ordre 2 représente la variance de la variable aléatoire. On notera que pour une variable aléatoire dont la ddp est paire, tous les moments centrés d'ordre impair sont nuls.

La première fonction caractéristique est définie comme la Transformée de Fourier (TF) de la ddp :

$$\Phi_X(\nu) = E[e^{-j2\pi X\nu}] = \int_{-\infty}^{+\infty} p_X(x) e^{-j2\pi x\nu} dx. \quad (1.8)$$

La première fonction caractéristique est aussi appelée fonction génératrice des moments, car les moments d'ordre r s'expriment comme :

$$\mu_{rX} = (-j2\pi)^{-r} \left[\frac{d^r \Phi_X(\nu)}{d\nu^r} \right]_{\nu=0}. \quad (1.9)$$

En écrivant un développement de Taylor au voisinage de $\nu = 0$, les moments permettent d'approximer la première fonction caractéristique :

$$\Phi_X(\nu) = 1 + \sum_{r=1}^{+\infty} \frac{(-j2\pi\nu)^r \mu_{rX}}{r!}. \quad (1.10)$$

Avec ce développement, nous remarquons qu'en considérant tous les moments d'ordre inférieur à r nous faisons une approximation de la première fonction caractéristique. Ainsi, la première fonction caractéristique et les moments auront des propriétés similaires.

Cumulants et seconde fonction caractéristique

La seconde fonction caractéristique est définie comme le logarithme népérien complexe de la première fonction caractéristique [KS63] :

$$\Psi_X(\nu) = \ln[\Phi_X(\nu)]. \quad (1.11)$$

Cette fonction est dite génératrice des cumulants. Le cumulante κ_{rX} d'ordre r de X s'exprime en fonction de la dérivée de la seconde fonction caractéristique en zéro :

$$\kappa_{rX} = (-j2\pi)^{-r} \left[\frac{d^r \Psi_X(\nu)}{d\nu^r} \right]_{\nu=0}. \quad (1.12)$$

Les propriétés statistiques de la variable aléatoire X peuvent être décrites de façon équivalente par les moments et/ou les cumulants. En effet, les cumulants peuvent se déduire des moments de la façon suivante :

$$\kappa_{rX} = r! \sum_{m=0}^r \sum \frac{(-1)^{\rho-1} (\rho-1)!}{\pi_1! \dots \pi_m!} \left(\frac{\mu_{p_1 X}}{p_1!} \right)^{\pi_1} \dots \left(\frac{\mu_{p_m X}}{p_m!} \right)^{\pi_m}, \quad (1.13)$$

où la seconde somme est effectuée pour toutes les valeurs non négatives des π_i , p_i et ρ pour lesquelles $p_1\pi_1 + \dots + p_m\pi_m = r$ et $\pi_1 + \dots + \pi_m = \rho$.

Les premiers cumulants s'expriment simplement en fonction des moments centrés d'ordre inférieur ou égal par :

$$\begin{aligned} \kappa_{1X} &= \mu_{1X} = E[X], \\ \kappa_{2X} &= \mu_{2X} = E[X^2], \\ \kappa_{3X} &= \mu_{3X} = E[X^3], \\ \kappa_{4X} &= \mu_{4X} - 3(\mu_{2X})^2 = E[X^4] - 3E[X^2]^2. \end{aligned} \quad (1.14)$$

A l'aide de ces cumulants, nous définissons les cumulants standardisés d'ordre 3 et 4.

Asymétrie ou *skewness*

L'asymétrie ou *skewness* \mathcal{K}_{3X} d'une variable aléatoire X est le rapport entre le cumulante d'ordre 3 et le cumulante d'ordre 2 à la puissance $3/2$:

$$\mathcal{K}_{3X} = \frac{\kappa_{3X}}{(\kappa_{2X})^{3/2}} = \frac{E[X^3]}{E[X^2]^{3/2}}. \quad (1.15)$$

Le skewness mesure l'asymétrie de la ddp. Il est nul si la densité de probabilité présente un axe de symétrie [LAC97] autour de sa valeur moyenne (0 dans notre cas).

Kurtosis

Le kurtosis \mathcal{K}_{4X} d'une variable aléatoire X est le rapport entre le cumulants d'ordre 4 et le cumulants d'ordre 2 à la puissance 2 de la variable aléatoire X :

$$\mathcal{K}_{4X} = \frac{\kappa_{4X}}{(\kappa_{2X})^2} = \frac{E[X^4]}{E[X^2]^2} - 3. \quad (1.16)$$

Uniquement dans le cas de loi monomodale [MJ99], le kurtosis réalise une comparaison de la ddp de la variable aléatoire X par rapport à la loi gaussienne de même variance :

- Si la densité de probabilité de X tend vers 0 à l'infini plus rapidement que la loi gaussienne, la variable aléatoire est dite sous-gaussienne et le kurtosis est négatif.
- Si la densité de probabilité de X tend vers 0 à l'infini moins vite que la loi gaussienne, la variable aléatoire est dite sur-gaussienne et le kurtosis est positif.
- Si la densité de probabilité de X est gaussienne, le kurtosis est nul.

Le kurtosis caractérise l'aplatissement de la ddp par rapport à celui d'une distribution gaussienne. Nous venons de voir les définitions théoriques des moments et cumulants d'une variable aléatoire réelle, dans la suite, nous allons nous intéresser à l'estimation de ces grandeurs.

Estimation des moments et cumulants

Soit X une variable aléatoire réelle scalaire centrée et N réalisations indépendantes de X notées x_1, x_2, \dots, x_N .

Un estimateur du moment d'ordre r de X est :

$$\widehat{\mu_{rX}} = \frac{1}{N} \sum_{k=1}^N x_k^r. \quad (1.17)$$

Cet estimateur est non biaisé et sa variance est $\text{var}[\widehat{\mu_{rX}}] = \frac{1}{N}(\mu_{2rX} - (\mu_{rX})^2)$. Ainsi, cet estimateur $\widehat{\mu_{rX}}$ du moment d'ordre r de X est convergent car $\lim_{N \rightarrow \infty} \text{var}[\widehat{\mu_{rX}}] = 0$. L'estimateur $\widehat{\mu_{rX}}$ converge donc vers μ_{rX} avec une probabilité de 1 quand $N \rightarrow \infty$. De plus, $\widehat{\mu_{rX}}$ est asymptotiquement normal. Pour construire des estimateurs des cumulants d'ordre r de X , nous pouvons tout simplement utiliser (1.13) en remplaçant le moment μ_{rX} d'ordre r par son estimateur $\widehat{\mu_{rX}}$. Cependant, on observe généralement que ces estimateurs sont biaisés. Par exemple, pour le cumulants d'ordre 4, l'estimateur serait :

$$\widehat{\kappa_{4X}} = \widehat{\mu_{4X}} - 3(\widehat{\mu_{2X}})^2. \quad (1.18)$$

Avec les estimateurs des moments (1.17), l'estimateur $\widehat{\kappa_{4X}}$ du cumulants d'ordre 4 est :

$$\widehat{\kappa_{4X}} = \frac{1}{N} \sum_{k=1}^N x_k^4 - \frac{3}{N^2} \left(\sum_{k=1}^N x_k^2 \right)^2. \quad (1.19)$$

Pour calculer le biais, nous nous intéressons à $E[\widehat{\kappa_{4X}}]$. Pour simplifier, nous devons réécrire le deuxième terme de façon différente comme une somme de termes et non une somme au carré, nous avons :

$$\left(\sum_{k=1}^N x_k^2\right)^2 = \left(\sum_{k=1}^N x_k^2\right) \left(\sum_{j=1}^N x_j^2\right) = \sum_{k=1}^N x_k^4 + \sum_{k=1}^N x_k^2 \sum_{j=1, j \neq k}^N x_j^2. \quad (1.20)$$

En utilisant (1.20) dans l'estimateur du cumulante d'ordre 4 (1.19), nous obtenons :

$$E[\widehat{\kappa}_{4X}] = \left(N \frac{1}{N} - N \frac{3}{N^2}\right) \mu_{4X} - N(N-1) \frac{3}{N^2} (\mu_{2X})^2. \quad (1.21)$$

En reprenant la relation (1.13), nous pouvons écrire :

$$E[\widehat{\kappa}_{4X}] = \kappa_{4X} - \frac{3}{N} (\kappa_{4X} + 2(\mu_{2rX})^2). \quad (1.22)$$

L'estimateur $\widehat{\kappa}_{4X}$ basé sur la relation générale entre les moments et cumulants est biaisé. Cependant, ces estimateurs sont asymptotiquement non biaisés, puisque le biais tend vers 0 quand N tend vers l'infini.

Des estimateurs non biaisés des cumulants s'appuient sur les k -statistiques [KS63, chap. 12],[Cul87, chap. 4]. Les k -statistiques sont des fonctions symétriques qui ont été introduites par Fisher pour caractériser les distributions. Par exemple, en notant $s_q = \sum_{k=1}^N x_k^q$, les premières k -statistiques sont des estimateurs des cumulants :

$$\begin{aligned} \widehat{\kappa}_{1X} &= k_{1X} = \frac{1}{N} s_1, \\ \widehat{\kappa}_{2X} &= k_{2X} = \frac{1}{N(N-1)} (N s_2 - (s_1)^2), \\ \widehat{\kappa}_{3X} &= k_{3X} = \frac{1}{N(N-1)(N-2)} (N^2 s_3 - 3N s_2 s_1 + 2(s_1)^3), \\ \widehat{\kappa}_{4X} &= k_{4X} = \frac{1}{N(N-1)(N-2)(N-3)} ((N^3 + N^2) s_4 - 4(N^2 + N) s_3 s_1 + \\ &\quad - 3(N^2 - N)(s_2)^2 + 12N s_2 (s_1)^2 - 6(s_1)^4). \end{aligned} \quad (1.23)$$

Les k -statistiques k_{rX} sont des estimateurs non biaisés et convergents des cumulants [KS63, Cul87] *i.e.* $E[k_{rX}] = \kappa_{rX}$ et $\lim_{N \rightarrow \infty} \text{var}[k_{rX}] = 0$.

Pour les cumulants normalisés (skewness et kurtosis), il n'existe pas un estimateur non-biaisé indépendant de la distribution de la variable aléatoire [LAC97, p.28]. Le skewness et le kurtosis définis en (1.15) et (1.16) peuvent être estimés de la façon suivante :

$$\widehat{\mathcal{K}}_{3X} = \frac{k_{3X}}{(k_{2X})^{3/2}}, \quad (1.24)$$

$$\widehat{\mathcal{K}}_{4X} = \frac{k_{4X}}{(k_{2X})^2}. \quad (1.25)$$

Dans le cas d'une distribution gaussienne, ces estimateurs sont non biaisés et sont convergents car leurs variances tendent vers 0 quand N tend vers l'infini. Nous avons même plus précisément [KS63, chap.12] :

$$\text{var}[\widehat{\mathcal{K}}_{3X}] = \frac{6N(N-1)}{(N-2)(N+1)(N+3)} \sim \frac{6}{N}, \quad N \rightarrow \infty \quad (1.26)$$

$$\text{var}[\widehat{\mathcal{K}}_{4X}] = \frac{24N(N-1)^2}{(N-3)(N-2)(N+3)(N+5)} \sim \frac{24}{N}, \quad N \rightarrow \infty. \quad (1.27)$$

Nous avons détaillé les éléments statistiques permettant de décrire une variable aléatoire réelle scalaire, l'extension de ses notions au cas scalaire complexe est détaillé dans [AGL96a, AGL96b] et est repris notamment dans [LAC97, p.14] et [Vra03]. En déconvolution aveugle, nous serons amenés à mesurer les liens statistiques entre plusieurs variables. Ainsi nous proposons de nous intéresser dans le paragraphe suivant aux variables multidimensionnelles réelles centrées.

1.2.2 Vecteurs aléatoires réels

1.2.2.1 Définition générale

Soit $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$ un vecteur de n variables aléatoires réelles **centrées**, la fonction de répartition à n dimensions est définie par :

$$F_{\mathbf{X}}(\mathbf{u}) = \text{Proba}\{x_1 \leq u_1, x_2 \leq u_2, \dots, x_n \leq u_n\}, \quad (1.28)$$

où $\mathbf{u} = (u_1, \dots, u_n)^T$. La densité de probabilité (ddp) conjointe $p_{\mathbf{X}}(\mathbf{x})$ du vecteur de variables aléatoires \mathbf{X} est :

$$p_{\mathbf{X}}(\mathbf{x}) = \frac{\partial^n F_{\mathbf{X}}(\mathbf{x})}{\partial x_1 \partial x_2 \dots \partial x_n},$$

où $\mathbf{x} = (x_1, \dots, x_n)^T$ est le vecteur des valeurs prises par le vecteur aléatoire \mathbf{X} . Le moment croisé d'ordre r du vecteur aléatoire \mathbf{X} est défini par :

$$\mu_{r_1 \dots r_n \mathbf{X}} = E[X_1^{r_1} \dots X_n^{r_n}] = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} x_1^{r_1} \dots x_n^{r_n} p_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}, \quad (1.29)$$

avec $r_1 + \dots + r_n = r$. Comme pour les variables aléatoires scalaires, nous définissons les deux premières fonctions caractéristiques. La première fonction caractéristique $\Phi_{\mathbf{X}}(\boldsymbol{\nu})$, avec $\boldsymbol{\nu} = (\nu_1, \dots, \nu_n)^T$ le vecteur des fréquences, est définie comme la transformée de Fourier de dimension n de la ddp et la seconde est simplement le logarithme népérien de la première, soit :

$$\Phi_{\mathbf{X}}(\boldsymbol{\nu}) = E[e^{-j2\pi \mathbf{x}^T \boldsymbol{\nu}}] = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} p_{\mathbf{X}}(\mathbf{x}) e^{-j2\pi \mathbf{x}^T \boldsymbol{\nu}} dx_1 \dots dx_n, \quad (1.30)$$

$$\Psi_{\mathbf{X}}(\boldsymbol{\nu}) = \ln[\Phi_{\mathbf{X}}(\boldsymbol{\nu})]. \quad (1.31)$$

De façon identique aux relations (1.9) et (1.12) dans le cas de variable aléatoire scalaire, les moments et les cumulants s'expriment aussi avec les dérivées partielles des deux fonctions caractéristiques :

$$\mu_{r_1 \dots r_n \mathbf{X}} = (-j2\pi)^{-r} \left[\frac{\partial^m \Phi_{\mathbf{X}}(\boldsymbol{\nu})}{\partial \nu_1^{r_1} \dots \partial \nu_n^{r_n}} \right]_{\nu_1 = \dots = \nu_n = 0}, \quad (1.32)$$

$$\kappa_{r_1 \dots r_n \mathbf{X}} = (-j2\pi)^{-r} \left[\frac{\partial^r \Psi_{\mathbf{X}}(\boldsymbol{\nu})}{\partial \nu_1^{r_1} \dots \partial \nu_n^{r_n}} \right]_{\nu_1 = \dots = \nu_n = 0}. \quad (1.33)$$

Ainsi, les moments (respectivement les cumulants) permettent d'approximer la première (resp. la seconde) fonction caractéristique.

Cas particulier

Classiquement, nous utilisons les cumulants dans le cas particulier $r_1 = \dots = r_n = 1$. Comme $r = \sum_{i=1}^n r_i$ nous avons $r = n$ [Cul87, p.29],[Bri81, p.19],[LAC97, p.13]. Classiquement, pour simplifier les notations, nous noterons dans ce cas particulier le cumulant croisé d'ordre n par :

$$\kappa_{1 \dots 1 \mathbf{X}} = Cum[X_1, X_2, \dots, X_n]. \quad (1.34)$$

Les cumulants croisés d'un vecteur aléatoire n dimensionnel, $Cum[X_1, X_2, \dots, X_n]$ sont liés aux moments croisés d'ordre inférieur ou égal à n par la relation :

$$Cum[X_1, X_2, \dots, X_n] = \sum (-1)^{k-1} (k-1)! E \left[\prod_{i_1 \in n_1} X_{i_1} \right] E \left[\prod_{i_2 \in n_2} X_{i_2} \right] \dots E \left[\prod_{i_p \in n_p} X_{i_p} \right], \quad (1.35)$$

où la somme s'étend sur tous les ensembles $\{n_1, n_2, \dots, n_p : 1 \leq p \leq n\}$ formant une partition de $\{1, 2, \dots, n\}$ et k est le nombre d'éléments composant la partition. Par exemple, les premiers cumulants croisés des variables aléatoires centrées s'expriment par :

$$\begin{aligned} Cum[X_1, X_2] &= E[X_1 X_2], \\ Cum[X_1, X_2, X_3] &= E[X_1 X_2 X_3], \\ Cum[X_1, X_2, X_3, X_4] &= E[X_1 X_2, X_3, X_4] - E[X_1 X_2] E[X_3 X_4] \\ &\quad - E[X_1 X_3] E[X_2 X_4] - E[X_1 X_4] E[X_2 X_3]. \end{aligned} \quad (1.36)$$

Remarque : Les relations précédentes correspondent au prolongement des relations (1.14) dans le cas scalaire. Par exemple, le cumulant d'ordre 4 κ_{4X} de la variable scalaire X est identique au moment d'ordre 4 du vecteur $\mathbf{X} = (X, X, X, X)$ soit $\kappa_{4X} = Cum[X, X, X, X]$

Nous avons défini les moments et cumulants dans le cas d'un vecteur aléatoire. Dans la suite, nous allons présenter les principales propriétés des moments et cumulants.

1.2.2.2 Propriétés des moments et cumulants

L'ensemble des moments d'ordre inférieur à n et l'ensemble des cumulants d'ordre inférieur à n contiennent exactement les mêmes informations statistiques. Cependant, l'expression de celles-ci est différente et est plus au moins simple à obtenir. Le choix du formalisme des moments ou des cumulants se fera en fonction des propriétés recherchées dans le vecteur aléatoire.

Multilinéarité

Les cumulants et moments vérifient la propriété de multilinéarité. L'expression de cette propriété est donnée dans [Cul87] [LAC97, p.19]. En particulier, nous avons pour un réel $\lambda \in \mathbb{R}$ et $n + 1$ variables aléatoires réelles Y, X_1, X_2, \dots, X_n :

$$E[(\lambda X_1 + Y)X_2 \dots X_n] = \lambda E[X_1 X_2 \dots X_n] + E[Y X_2 \dots X_n], \quad (1.37)$$

$$Cum[(\lambda X_1 + Y), X_2 \dots, X_n] = \lambda Cum[X_1, X_2, \dots, X_n] + Cum[Y, X_2 \dots, X_n]. \quad (1.38)$$

Remarque : Les cumulants contrairement aux moments sont invariants par translation, mais en déconvolution aveugle nous nous limitons à des signaux centrés, par conséquent cette invariance n'est pas utile.

Indépendance statistique

Soient $\{X_1, \dots, X_n\} = \{X_1, \dots, X_i\} \cup \{X_{i+1}, \dots, X_n\}$ un ensemble de n variables aléatoires, tel que les deux sous-ensembles soient indépendants, alors :

$$E[X_1, X_2 \dots X_n] = E[X_1 X_2 \dots X_i] E[X_{i+1} \dots X_n], \quad (1.39)$$

$$Cum[X_1, X_2 \dots, X_n] = 0. \quad (1.40)$$

Ces relations ne font qu'approcher aux ordres supérieurs les relations vérifiées par les deux fonctions caractéristiques. En effet, en écrivant l'indépendance des deux ensembles de variables nous avons :

$$p_{X_1, \dots, X_n}(x_1, \dots, x_n) = p_{X_1, \dots, X_i}(x_1, \dots, x_i) p_{X_{i+1}, \dots, X_n}(x_{i+1}, \dots, x_n).$$

Après transformée de Fourier de l'équation précédente et le logarithme, les fonctions caractéristiques vérifient :

$$\Phi_{X_1, \dots, X_n}(\nu_1, \dots, \nu_n) = \Phi_{X_1, \dots, X_i}(\nu_1, \dots, \nu_i) \Phi_{X_{i+1}, \dots, X_n}(\nu_{i+1}, \dots, \nu_n), \quad (1.41)$$

$$\Psi_{X_1, \dots, X_n}(\nu_1, \dots, \nu_n) = \Psi_{X_1, \dots, X_i}(\nu_1, \dots, \nu_i) + \Psi_{X_{i+1}, \dots, X_n}(\nu_{i+1}, \dots, \nu_n). \quad (1.42)$$

Enfin, par dérivation de (1.41) (resp. (1.42)), nous trouvons les relations (1.39) (resp. (1.40)). La relation (1.40) avec les cumulants croisés permet de mettre plus facilement en évidence la notion d'indépendance statistique entre différentes variables aléatoires comparativement au formalisme des moments avec (1.39). Cette propriété des cumulants est très largement utilisée en traitement du signal, notamment dans les problèmes d'Analyse en Composantes Indépendantes (ACI) ou de séparation de sources [CS93, Com94, Car99].

Somme de variables aléatoires indépendantes

Soient $\mathbf{X} = (X_1, X_2, \dots, X_n)$ et $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)$ deux vecteurs de variables aléatoires, si les composantes de \mathbf{X} sont mutuellement indépendantes de celles de \mathbf{Y} , alors nous avons pour le cumulants d'ordre n , la relation suivante :

$$\text{Cum}[X_1 + Y_1, X_2 + Y_2, \dots, X_n + Y_n] = \text{Cum}[X_1, X_2, \dots, X_n] + \text{Cum}[Y_1, Y_2, \dots, Y_n]. \quad (1.43)$$

Cette relation peut être retrouvée en écrivant que $p_{\mathbf{X}+\mathbf{Y}} = p_{\mathbf{X}} \star p_{\mathbf{Y}}$, puis par transformée de Fourier nous obtenons pour la première fonction caractéristique $\Phi_{\mathbf{X}+\mathbf{Y}}(\boldsymbol{\nu}) = \Phi_{\mathbf{X}}(\boldsymbol{\nu})\Phi_{\mathbf{Y}}(\boldsymbol{\nu})$. Puis avec le logarithme népérien pour la seconde fonction caractéristique nous avons $\Psi_{\mathbf{X}+\mathbf{Y}}(\boldsymbol{\nu}) = \Psi_{\mathbf{X}}(\boldsymbol{\nu}) + \Psi_{\mathbf{Y}}(\boldsymbol{\nu})$. Enfin par dérivation, nous obtenons (1.43). Cette relation (1.43) approche à l'ordre n la relation générale vérifiée par la seconde fonction caractéristique. Cette propriété n'est pas vérifiée par les moments. Cette relation justifie le nom de "cumulants", puisque les cumulants d'une somme de variables indépendantes sont cumulatifs.

Nous venons de voir les principales propriétés des cumulants et moments. En pratique, se pose le problème de l'estimation de ces grandeurs, qui sera détaillé dans le paragraphe suivant.

Estimation des moments et cumulants multidimensionnels

Soit $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$ un vecteur aléatoire de n variables aléatoires réelles. On dispose de N réalisations indépendantes de chaque variable. Pour la variable aléatoire X_i , nous noterons les réalisations sous forme d'un vecteur $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{iN}) = (x_{ij})_{j=1}^N$.

Comme dans le cas scalaire (1.17), les moments croisés peuvent être estimés sans biais de la façon suivante :

$$\hat{\mu}_{r_1 \dots r_n \mathbf{X}} = \frac{1}{N} \sum_{j=1}^N x_{1j}^{r_1} x_{2j}^{r_2} \dots x_{nj}^{r_n}. \quad (1.44)$$

En disposant d'un estimateur des moments, nous pourrions ensuite estimer les cumulants en utilisant la relation (1.35). Mais nous obtiendrions comme dans le cas scalaire des estimateurs biaisés. Les estimateurs non biaisés des cumulants croisés sont données pour les données multidimensionnelles [KS63, chap. 13][Cul87, chap. 4] par les k -statistiques :

$$\begin{aligned}
\widehat{Cum}[X_p, X_q] &= \frac{1}{N(N-1)}(Ns_{pq} - s_p s_q), \\
\widehat{Cum}[X_p, X_q, X_r] &= \frac{1}{N(N-1)(N-2)}(N^2 s_{pqr} - N s_p s_{qr} - N s_q s_{rp} - N s_r s_{pq} + 2s_p s_q s_r), \\
\widehat{Cum}[X_p, X_q, X_r, X_t] &= \frac{1}{N(N-1)(N-2)(N-3)} \{ N^2(N+1)s_{pqrt} - N(N+1)s_p s_{qrt} \\
&\quad - N(N+1)s_q s_{rtp} - N(N+1)s_r s_{tpq} - N(N+1)s_t s_{pqr} \\
&\quad - N(N-1)s_{pq} s_{rt} - N(N-1)s_{pr} s_{qt} - N(N-1)s_{pt} s_{qr} \\
&\quad + 2N s_p s_q s_r s_t + 2N s_p s_r s_q s_t + 2N s_p s_q s_r s_t + 2N s_q s_r s_p s_t \\
&\quad + 2N s_q s_t s_{pr} + 2N s_r s_t s_{pq} - 6s_p s_q s_r s_t \},
\end{aligned}$$

où $s_p = \sum_{j=1}^N x_{pj}$, $s_{pq} = \sum_{j=1}^N x_{pj} x_{qj}$, $s_{pqr} = \sum_{j=1}^N x_{pj} x_{qj} x_{rj}$, $s_{pqrt} = \sum_{j=1}^N x_{pj} x_{qj} x_{rj} x_{tj}$ et les indices $p, q, r, t \in \{1, \dots, n\}$ ne sont pas nécessairement distincts. Dans le cas particulier de $p = q = r = t$, nous retrouvons le cas de la variable aléatoire scalaire donné par les équations (1.23).

Dans cette partie de rappel de statistiques, nous avons présenté les moments et les cumulants d'une variable aléatoire et d'un vecteur aléatoire. En traitement du signal, nous observons des signaux que nous souhaitons caractériser. En faisant l'hypothèse d'un processus iid, les échantillons $x(t)$ sont vus comme des réalisations d'un processus stochastique X , qui s'apparente à une variable aléatoire. Par conséquent, les outils développés dans le cas d'une variable aléatoire scalaire permettent de caractériser ce processus. Ensuite, si nous nous intéressons aux liaisons entre plusieurs processus stochastiques, nous serons amenés à utiliser les outils vectoriels. Par exemple, pour la séparation de sources, nous utilisons les cumulants pour décrire l'indépendance entre les sources estimées, soit à l'ordre 2 (Analyse en Composantes Principales) ou à l'ordre 3 ou au delà (fréquemment à l'ordre 4) [Car99, CS93, Com94] pour l'Analyse en Composantes Indépendantes. En déconvolution aveugle, nous souhaitons caractériser la dépendance temporelle entre les échantillons dans un processus stochastique, par exemple la blancheur qui est synonyme de l'indépendance statistique entre les échantillons successifs. Ainsi, chaque version décalée $x(t + \tau)$ d'une observation $x(t)$ peut être vue comme une variable aléatoire. Ensuite, nous pouvons quantifier la dépendance en utilisant les cumulants croisés entre les versions décalées du processus. Dans la partie suivante, nous allons présenter les outils nécessaires à l'analyse des signaux monodimensionnels stationnaires tels que les fonctions de multicorrélations et les multispectres.

1.3 Les statistiques d'ordre supérieur pour l'analyse des signaux monodimensionnels

Dans cette partie, nous allons introduire les principaux outils de statistiques d'ordre supérieur pour la caractérisation des signaux monodimensionnels réels centrés. Pour de plus amples détails, le lecteur pourra se reporter aux ouvrages suivants [BLF53, Doo53,

Sap78, Bri81, Pri81, BLP81, Pic94, LAC97]. Dans la suite, pour être en accord avec la partie précédente sur les statistiques d'ordre supérieur des variables aléatoires, nous devrions noter X_t le processus à l'instant t et $x(t)$ sa réalisation. Pour ne pas surcharger les notations, nous noterons de façon équivalente la réalisation $x(t)$ et le processus au même instant.

1.3.1 Multicorrélation et multispectre

Soit $x(t)$ un signal aléatoire centrés à valeurs réelles, la multicorrélation d'ordre r est basée sur le cumulants d'ordre r (défini dans le paragraphe 1.2.2) des valeurs du signal (variables aléatoires) aux instants t_0, t_1, \dots, t_{r-1} :

$$C_{rX}(\mathbf{t}) = \text{Cum}[x(t_0), x(t_1), \dots, x(t_{r-1})], \quad (1.45)$$

où $\mathbf{t} = (t_0, t_1, \dots, t_{r-1})^T$. D'après la définition, la multicorrélation d'ordre r est fonction de r instants contenus dans le vecteur \mathbf{t} . Si nous supposons le signal $x(t)$ stationnaire au sens strict, ses statistiques sont invariantes par changement de l'origine des temps. Ainsi, nous définissons les retards $\tau_i = t_i - t_0$, $i = 1, \dots, r - 1$. La multicorrélation d'ordre $r - 1$ est alors :

$$C_{rX}(\boldsymbol{\tau}) = \text{Cum}[x(t), x(t + \tau_1), \dots, x(t + \tau_{r-1})], \quad (1.46)$$

où $\boldsymbol{\tau} = (\tau_1, \dots, \tau_{r-1})^T$ est un vecteur de $r - 1$ retards. Ainsi, nous avons avec les multicorrélations une caractérisation dans le domaine temporel des statistiques du signal. Pour obtenir une analyse fréquentielle, si la multicorrélation d'ordre r admet une transformée de Fourier à r dimensions, nous définissons les multispectres d'ordre r [Ros62, Bri81] par :

$$S_{rX}(\boldsymbol{\nu}) = \text{TF}[C_{rX}(\boldsymbol{\tau})] = \int_{\mathbb{R}^{r-1}} C_{rX}(\boldsymbol{\tau}) \exp\{-2\pi j \boldsymbol{\nu}^T \boldsymbol{\tau}\} d\boldsymbol{\nu}, \quad (1.47)$$

où $\boldsymbol{\nu} = (\nu_1, \dots, \nu_{r-1})^T$ est le vecteur des fréquences et $\boldsymbol{\tau} = (\tau_1, \dots, \tau_{r-1})^T$ le vecteur des variables temporelles.

La définition précédente des multispectres est indirecte car elle s'appuie sur la multicorrélation définie au préalable. Une seconde définition équivalente mais directe utilise les représentations de Cramér $dX(\nu)$ du signal aléatoire [Cra46]. Le signal et sa représentation de Cramér sont liés par l'intégrale stochastique de Cramér [BLF53] :

$$x(t) = \int e^{j2\pi t\nu} dX(\nu). \quad (1.48)$$

Cette définition permet de prolonger la définition de la transformée de Fourier pour les signaux aléatoires. Dans le cas stationnaire, le multispectre d'ordre r peut être défini de façon équivalente par :

$$S_{rX}(\nu_0, \nu_1, \dots, \nu_{r-1}) \delta \left(\sum_{i=0}^{r-1} \nu_i \right) d\nu_0 d\nu_1 \dots d\nu_{r-1} = \text{Cum}[dX(\nu_0), dX(\nu_1), \dots, dX(\nu_{r-1})], \quad (1.49)$$

où $\delta(\cdot)$ est la distribution de Dirac. Grâce à l'hypothèse de stationnarité, la variable ν_0 peut être remplacée par $-\sum_{i=1}^{r-1} \nu_i$. Les définitions (1.47) et (1.49) sont totalement équivalentes.

Dans les deux définitions des multispectres (1.47) et (1.49), nous utilisons les cumulants ainsi nous définissons les multispectres de cumulants. Il existe aussi les multispectres de moments [BLF53], mais ils ne permettent pas de mesurer la blancheur d'un signal, ainsi nous avons fait le choix de ne pas les présenter ici.

Par permutation des retards ou des fréquences dans les arguments des définitions de la multicorrélation (1.46) et du multispectre (1.47) ou (1.49), il existe un ensemble de valeurs identiques de ces grandeurs. Dans la suite, nous allons présenter les spécificités des multicorrélations et des multispectres d'ordre 2, 3 et 4 en précisant les symétries de chaque fonction et leurs domaines de définition. Le problème de l'estimation sera aussi abordé.

Ordre 2

La multicorrélation d'ordre 2 est tout simplement la fonction d'autocorrélation $C_{2X}(\tau)$ du signal $x(t)$:

$$C_{2X}(\tau) = E[x(t)x(t + \tau)]. \quad (1.50)$$

La mesure d'autocorrélation $C_{2X}(\tau)$ mesure à l'ordre 2 la dépendance statistique entre les échantillons du processus $x(t)$ et $x(t + \tau)$ une version retardée du même processus. En pratique, sous hypothèse d'ergodicité nous remplaçons la moyenne statistique par une moyenne temporelle sur les T échantillons, ainsi nous estimons la fonction d'autocorrélation par :

$$\hat{C}_{2X}(\tau) = \frac{1}{T} \sum_{t=1}^T x(t)x(t + \tau). \quad (1.51)$$

Pour rappel, la fonction d'autocorrélation est paire : $C_{2X}(\tau) = C_{2X}(-\tau)$. Nous définissons en général la notion de signal à mémoire limitée. Un signal $x(t)$ est dit à mémoire limitée s'il existe une valeur du temps τ_M telle que si deux instants t_1 et t_2 sont séparés d'un retard supérieur à τ_M soit $|t_2 - t_1| > \tau_M$ alors $x(t_1)$ et $x(t_2)$ sont deux variables aléatoires indépendantes. En particulier, l'autocorrélation sera nulle si $\tau > \tau_M$. En définitive, le domaine d'estimation de l'autocorrélation peut se réduire à $\tau \in [0, \tau_M]$.

Le multispectre d'ordre 2 est tout simplement la densité spectrale de puissance $S_{2X}(\nu)$ définie comme la transformée de Fourier de la fonction d'autocorrélation. Nous avons :

$$S_{2X}(\nu) = TF[C_{2X}(\tau)] = E[|X(\nu)|^2]. \quad (1.52)$$

La dernière égalité découle de la définition (1.49) de Cramér. D'après (1.49), $S_{2X}(\nu) = Cum[X(-\nu), X(\nu)]$ et en reprenant l'expression du cumulants d'ordre 2 de (1.36), nous obtenons que $S_{2X}(\nu) = E[|X(\nu)|^2]$. Cette définition permet de montrer que nous pouvons estimer directement la densité spectrale de puissance par une moyenne statistique des spectres $X(\nu)$ de plusieurs réalisations du signal $x(t)$. Dans le cas classique, nous avons des signaux à bande limitée car ils devront être par la suite échantillonnés, *i.e.* $|X(\nu)| = 0$ si $|\nu| \geq B$ où B représente la largeur de bande de fréquence du signal. Ainsi, le multispectre d'ordre 2 est défini sur $\nu \in [-B, B]$ avec un axe de symétrie en $\nu = 0$. Son domaine d'estimation se réduira à $[0, B]$.

Ordre 3

A l'ordre 3, la fonction de multicorrélation est la bicorrélation $C_{3X}(\tau_1, \tau_2)$ définie par [Nan99] :

$$C_{3X}(\tau_1, \tau_2) = E[x(t)x(t + \tau_1)x(t + \tau_2)]. \quad (1.53)$$

Nous avons cinq symétries pour la bicorrélation :

$$\begin{aligned} C_{3X}(\tau_1, \tau_2) &= C_{3X}(\tau_2, \tau_1), & C_{3X}(\tau_1, \tau_2) &= C_{3X}(\tau_2 - \tau_1, -\tau_1), \\ C_{3X}(\tau_1, \tau_2) &= C_{3X}(-\tau_1, \tau_2 - \tau_1), & C_{3X}(\tau_1, \tau_2) &= C_{3X}(\tau_1 - \tau_2, -\tau_2), \\ C_{3X}(\tau_1, \tau_2) &= C_{3X}(-\tau_2, \tau_1 - \tau_2). \end{aligned}$$

Avec l'ensemble de ces symétries, le domaine d'estimation de la bicorrélation est représenté par la figure 1.7 :

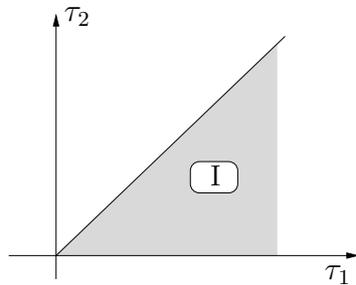


FIG. 1.7 – Zone de calcul de la bicorrélation.

L'estimation se fait dans le domaine I et nous retrouvons le reste du plan par les cinq symétries décrites précédemment.

De plus, si le signal $x(t)$ est à mémoire limitée, alors la bicorrélation $C_{3X}(\tau_1, \tau_2)$ est nulle dès que $|\tau_1| > \tau_M$ ou $|\tau_2| > \tau_M$ ou $|\tau_1 - \tau_2| > \tau_M$. Le support de la bicorrélation d'un signal à mémoire limitée est contenue dans l'hexagone de la figure 1.8 :

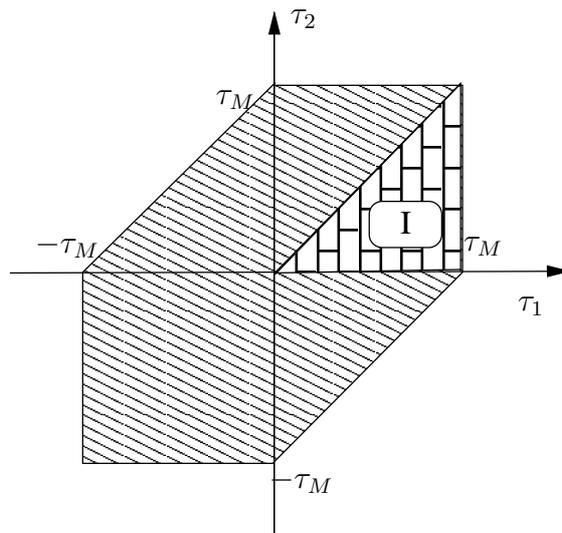


FIG. 1.8 – Zone de calcul de la bicorrélation pour un signal à mémoire limitée.

Avec l'ensemble des cinq symétries et la mémoire limitée, le domaine d'estimation I est réduit à un triangle.

Dans le domaine fréquentiel, le bispectre $S_{3X}(\nu_1, \nu_2)$ est défini comme la double transformée de Fourier de la bicorrélation :

$$S_{3X}(\nu_1, \nu_2) = TF[C_{3X}(\tau_1, \tau_2)]. \quad (1.54)$$

Avec la définition (1.49), par le biais de la représentation de Cramér, nous pouvons définir directement le bispectre à partir de la transformées de Fourier de X :

$$S_{3X}(\nu_1, \nu_2) = E[X(\nu_1)X(\nu_2)X^*(\nu_1 + \nu_2)]. \quad (1.55)$$

Pour un signal à bande limitée de largeur de bande B , le multispectre d'ordre 3 est défini sur un hexagone limité par $|\nu_1| < B$, $|\nu_2| < B$ et $|\nu_1 + \nu_2| < B$. Nous avons un domaine de la même forme que celui de la bicorrélation pour des signaux à durée limitée (Figure 1.8). D'après [LAC97, p.52], la présence d'un terme dépendant de la somme de deux fréquences implique que le bispectre du signal échantillonné coïncide avec celui du signal continu si et seulement si la fréquence d'échantillonnage F_E vérifie $F_E > 3B$. Les propriétés du multispectre d'ordre 3 sont étudiées précisément par Rossille [Ros95].

Ordre 4

La fonction de multicorrélation d'ordre 4 $C_{4X}(\tau_1, \tau_2, \tau_3)$, aussi appelée tricorrélation, est fonction de trois retards (τ_1, τ_2, τ_3) . Elle est définie comme le cumulante d'ordre 4 entre quatre versions décalées du signal $x(t)$:

$$\begin{aligned} C_{4X}(\tau_1, \tau_2, \tau_3) = & E[x(t)x(t + \tau_1)x(t + \tau_2)x(t + \tau_3)] \\ & - E[x(t)x(t + \tau_1)]E[x(t + \tau_2)x(t + \tau_3)] \\ & - E[x(t)x(t + \tau_2)]E[x(t + \tau_1)x(t + \tau_3)] \\ & - E[x(t)x(t + \tau_3)]E[x(t + \tau_1)x(t + \tau_2)]. \end{aligned} \quad (1.56)$$

Pour la fonction de tricorrélation, dans le cas d'un signal à mémoire limitée, le domaine de définition se réduit à un octaèdre qui est défini par un cube dont deux sommets sont coupés par des plans [Hue97, HLR99].

Pour la caractérisation fréquentielle, le multispectre d'ordre 4 ou trispectre $S_{4X}(\nu_1, \nu_2, \nu_3)$ est la transformée de Fourier à trois dimensions de la tricorrélation :

$$S_{4X}(\nu_1, \nu_2, \nu_3) = TF[C_{4X}(\tau_1, \tau_2, \tau_3)]. \quad (1.57)$$

Comme pour le multispectre d'ordre 3, avec la représentation de Cramér, nous pouvons estimer directement le trispectre par [Hue97, HLR99] :

$$\begin{aligned} S_{4X}(\nu_1, \nu_2, \nu_3) = & E[X(\nu_1)X(\nu_2)X(\nu_3)X^*(\nu_1 + \nu_2 + \nu_3)] \\ & - E[|X(\nu_1)|^2]E[|X(\nu_3)|^2]\delta(\nu_1 + \nu_2) \\ & - E[|X(\nu_1)|^2]E[|X(\nu_2)|^2]\delta(\nu_1 + \nu_3) \\ & - E[|X(\nu_1)|^2]E[|X(\nu_2)|^2]\delta(\nu_2 + \nu_3). \end{aligned} \quad (1.58)$$

Il existe un certain nombre de symétries pour ce trispectre, mais le plus intéressant est de voir que le trispectre du signal échantillonné et du signal à temps continu seront identiques uniquement si $F_E > 4B$ à cause du terme en $\nu_1 + \nu_2 + \nu_3$. De manière générale [LAC97, p.55], pour un signal $x(t)$ le multispectre d'ordre r du signal échantillonné coïncide avec celui du signal à temps continu si et seulement si :

$$F_E > rB, \quad (1.59)$$

où B est la largeur de bande du signal, *i.e.* $X(\nu) = 0$ si $|\nu| > B$. Nous venons de présenter les multispectres et les multicorrélations d'ordre 2, 3 et 4 qui sont les plus utilisés dans la littérature de la déconvolution. Dans le paragraphe suivant, nous donnons quelques éléments pour l'estimation de ces grandeurs.

Estimation

Pour l'estimation des multicorrélations, nous utilisons classiquement l'hypothèse d'ergodicité. Comme nous l'avons vu dans le cas de l'estimation des cumulants dans le paragraphe 1.2.2.2, nous remplaçons la moyenne statistique par une moyenne temporelle.

Pour les multispectres, il existe deux stratégies distinctes [LAC97, p.70]. La première est indirecte. Après estimation de la multicorrélation d'ordre r , nous en prenons la transformée de Fourier de dimension $r - 1$ pour obtenir le multispectre : cette technique est le multicorrélogramme. Pour améliorer la variance de l'estimateur, nous pouvons faire la moyenne sur plusieurs intervalle de temps : cet estimateur est aussi appelé le multipériodogramme moyenné [Bar48, BT58].

La seconde est une estimation directe : le multipériodogramme [RVN65, BR67, Bri81]. Elle est basée sur la représentation de Cramér des multispectres qui fournit des expressions du type de (1.55) ou (1.58). Nous estimons dans un premier temps le spectre $X(\nu)$ de $x(t)$, ensuite le multispectre d'ordre r est estimé par :

$$\widehat{S}_{rX}(\nu_1, \nu_2, \dots, \nu_{r-1}) = \widehat{Cum}[X(-\sum_{i=1}^{r-1} \nu_i), X(\nu_1), X(\nu_2), \dots, X(\nu_{r-1})]. \quad (1.60)$$

Cette estimation nécessite un estimateur du cumulants d'ordre r de variables complexes s'écrivant de la même façon que dans le cas réel (1.36). D'après [LAC97, p.72], cet estimateur n'est pas asymptotiquement efficace pour le bispectre. Pour obtenir un estimateur asymptotiquement efficace, nous devons mettre en oeuvre une opération de moyenne en fréquence : nous obtenons ainsi le multipériodogramme lissé.

Mesure de la blancheur

Après la mise en place des outils statistiques de multicorrélation et multispectre pour mesurer les liens statistiques existant entre plusieurs versions retardées du signal $x(t)$, nous devons écrire les relations induites par la blancheur du signal $x(t)$.

Un signal $x(t)$ est dit blanc ou iid si les échantillons de ce signal sont indépendants et identiquement distribués. Si le signal $x(t)$ est blanc alors la multicorrélation est nulle sauf si tous les retards sont nuls, et le multispectre est constant sur l'ensemble du domaine des fréquences :

$$C_{rX}(\tau_1, \dots, \tau_{r-1}) = \kappa_{rX} \delta(\tau_1, \dots, \tau_{r-1}), \quad (1.61)$$

$$S_{rX}(\nu_1, \dots, \nu_{r-1}) = \kappa_{rX}, \quad (1.62)$$

où κ_{rX} est le cumulante d'ordre r du signal $x(t)$ défini par (1.12) et $\delta(\cdot)$ est la distribution de Dirac. Ces relations prolongent celles à l'ordre 2 pour un signal blanc dont l'autocorrélation est égale à la distribution de Dirac et la densité spectrale de puissance (multispectre d'ordre 2) est constante.

Nous avons présenté les fonctions de multicorrélations pour caractériser la blancheur dans le domaine temporel avec les statistiques d'ordre supérieur et les multispectres pour le domaine fréquentiel. Nous devons nous intéresser à l'effet d'un filtrage sur ces grandeurs statistiques afin de mettre en place par la suite des méthodes de déconvolution aveugle.

1.3.2 Effet du filtrage sur les multicorrélations et les multispectres

Soient un signal $x(t)$ et un filtre linéaire homogène invariant dans le temps de réponse impulsionnelle $h(t)$, le signal $y(t) = h \star x(t)$ est le résultat le filtrage de $x(t)$ par $h(t)$. La figure suivante résume cette opération :

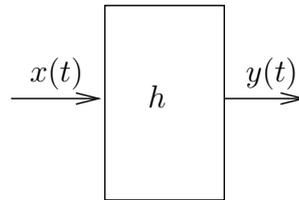


FIG. 1.9 – Filtrage.

Le filtrage nous fournit une relation de convolution dans le domaine temporel (écriture en signal échantillonné) et une multiplication dans le domaine fréquentiel :

$$y(t) = h \star x(t) = \sum_i h(i)x(t-i), \quad (1.63)$$

$$Y(\nu) = H(\nu)X(\nu), \quad (1.64)$$

Nous avons fait le choix d'écrire la relation directement dans le cas échantillonné pour permettre une relation directe avec les méthodes de déconvolution détaillées dans la partie suivante. Dans le domaine temporel (écriture pour un signal échantillonné), en utilisant la relation (1.63), les multicorrélations d'ordre r sont reliées par [Bri65, Bri81, NP93] :

$$C_{rY}(\boldsymbol{\tau}) = C_{rX}(\boldsymbol{\tau}) \star \left\{ \sum_t h(t) \prod_{i=1}^{r-1} h(t + \tau_i) \right\} = C_{rX}(\boldsymbol{\tau}) \star \left\{ \sum_t h(t)h(t + \tau_1), \dots, h(t + \tau_{r-1}) \right\}. \quad (1.65)$$

Dans le domaine fréquentiel, en utilisant la définition (1.49) avec la représentation $dY(\nu)$ de Cramér, où $dY(\nu) = H(\nu)dX(\nu)$, alors les multispectres sont exprimés par :

$$S_{rY}(\nu_1, \dots, \nu_{r-1}) = S_{rX}(\nu_1, \dots, \nu_{r-1}) H^* \left(\sum_{i=1}^{r-1} \nu_i \right) \prod_{i=1}^{r-1} H(\nu_i). \quad (1.66)$$

A priori, dans le domaine spectral les équations sont plus simples, surtout si nous nous limitons aux statistiques d'ordre 2, 3 ou 4. Ces relations donnent lieu à des méthodes de déconvolution : l'égalisation spectrale à l'ordre 2, et ensuite aux ordres 3 et 4 avec le bispectre et le trispectre. En déconvolution aveugle nous supposons qu'un processus $x(t)$ ou $y(t)$ (suivant l'approche utilisée) est blanc. Ainsi les relations (1.61) et (1.62) exprimant la blancheur du signal en temps ou en fréquence permettent de simplifier les relations dues au filtrage (1.65) et (1.66). Si nous utilisons les méthodes dans le domaine temporel, les filtres seront choisis à réponse impulsionnelle finie ce qui permettra de simplifier les relations et de stabiliser le problème inverse.

Remarque : Si nous nous intéressons aux statistiques d'ordre 2, en utilisant (1.66) avec $r = 2$, nous avons :

$$S_{2Y}(\nu) = |H(\nu)|^2 S_{2X}(\nu). \quad (1.67)$$

Ainsi, nous voyons que les statistiques d'ordre 2 ne prennent pas en compte l'information de phase du filtre. En revanche, par le biais des statistiques d'ordre 3 et 4, nous pouvons extraire l'information de phase du filtre.

1.4 État de l'art de la déconvolution

Dans ce paragraphe, nous nous proposons de faire un état de l'art des techniques de déconvolution linéaire basée sur la blancheur. Nous débuterons par le cas simple du filtrage de Wiener quand l'ondelette est supposée connue. Ensuite, si l'ondelette est inconnue, totalement ou en partie (phase ou module), nous entrons dans le domaine de la déconvolution semi-aveugle et aveugle. Dans ce contexte, nous ajoutons l'hypothèse de blancheur du signal de réflectivité $r(t)$. Nous mesurerons la blancheur du signal par les outils présentés dans le paragraphe précédent (multicorrélation, multispectre). Dans le cas semi-aveugle (module de l'ondelette inconnu et phase connue), les méthodes s'appuient sur les statistiques d'ordre 2 (Yule Walker, Durbin, égalisation spectrale). Enfin, en déconvolution aveugle (module et phase de l'ondelette inconnus), nous présenterons l'estimation du filtre par les statistiques d'ordre 3 (bicorrélation, bispectre) ou d'ordre 4 (tricorrélation, trispectre). Pour réaliser la déconvolution des données $d(t)$, nous estimerons un filtre inverse g qui permettra de calculer la sortie $y(t) = g \star d(t)$. Dans un cadre général le filtre g sera appelé le filtre déconvolveur. La figure 1.10 résume ces notations.

1.4.1 Filtrage de Wiener : ondelette connue

Si nous supposons l'ondelette $w(t)$ connue, en l'absence de bruit additif, le gain en fréquence $G(\nu)$ du filtre déconvolveur peut être choisi égal à $1/W(\nu)$. Mais avec la présence du bruit additif $n(t)$ sur les observations, la sortie de la déconvolution s'écrit comme $Y(\nu) = G(\nu)W(\nu)R(\nu) + G(\nu)N(\nu)$. Nous souhaitons que $Y(\nu) \approx R(\nu)$ mais le bruit additif vient "perturber" la sortie. Un bon compromis entre la déconvolution et l'amplification du bruit sera obtenu en minimisant l'erreur quadratique moyenne (EQM) entre la sortie $Y(\nu)$ et l'entrée $R(\nu)$ par rapport au gain en fréquence $G(\nu)$ du filtre déconvolveur :

$$G_{Wiener}(\nu) = \arg \min_{G(\nu)} \frac{E[|Y(\nu) - R(\nu)|^2]}{E[|R(\nu)|^2]}. \quad (1.68)$$

Nous ajoutons les hypothèses suivantes :

- la réflectivité $r(t)$ a une variance σ_R^2 ,
- la réflectivité $r(t)$ et le bruit $n(t)$ sont décorrélés,
- le bruit additif blanc a une variance σ_N^2 .

En utilisant le modèle de convolution-déconvolution, nous avons $Y(\nu) = G(\nu)[W(\nu)R(\nu) + N(\nu)]$, ainsi l'EQM s'écrit :

$$E[\{G(\nu)[W(\nu)R(\nu) + N(\nu)] - R(\nu)\} \{G(\nu)[W(\nu)R(\nu) + N(\nu)] - R(\nu)\}^*]. \quad (1.69)$$

Pour obtenir le minimum $G_{Wiener}(\nu)$, nous dérivons cette dernière équation par rapport à $G^*(\nu)$, ainsi $G_{Wiener}(\nu)$ vérifie :

$$E[\{G(\nu)[W(\nu)R(\nu) + N(\nu)] - R(\nu)\} \{W(\nu)^*R(\nu)^* + N(\nu)^*\}]. \quad (1.70)$$

Comme les grandeurs $G(\nu)$ et $W(\nu)$ sont déterministes, elle n'influe pas sur l'espérance mathématique. La décorrélation entre la réflectivité et le bruit implique que $E[R(\nu)N(\nu)] = 0$. De même, si nous conjuguons une des deux grandeurs. De plus, nous avons $E[|N(\nu)|^2] = \sigma_N^2$. Avec toutes ces remarques, nous parvenons à :

$$G_{Wiener}(\nu) = \frac{W^*(\nu)}{|W(\nu)|^2 + f_n^2}, \quad (1.71)$$

où f_n est appelé le facteur de bruit et vérifie $f_n = \sigma_N/\sigma_R$. La grandeur f_n peut être vue comme le rapport signal à bruit entre le bruit additif et le signal de réflectivité. Suivant, la prépondérance du terme $W(\nu)$ ou f_n , nous pouvons différencier deux bandes de fréquences. La première est la bande passante de l'ondelette (soit les fréquences pour lesquelles $|W(\nu)| \gg f_n$). Le filtre déconvolveur estimé tend vers le filtre déconvolveur théorique en absence de bruit, $1/W(\nu)$. Ainsi, dans cette bande de fréquence, nous avons une très bonne adéquation aux données. En revanche, dans la bande atténuée de l'ondelette (soit $|W(\nu)| \ll f_n$), le filtre de Wiener a pour gain $W^*(\nu)/f_n^2$. Dans cette bande de fréquence, l'adéquation aux données n'est pas du tout respectée au profit de l'atténuation du bruit. Le filtrage de Wiener est très intéressant par sa capacité à réaliser un compromis entre l'adéquation aux données (une bonne déconvolution) et une amplification modérée du bruit. Nous devons régler le facteur de bruit f_n . Pour cela, une estimation de la puissance du

bruit peut être faite sur une partie du signal comportant uniquement du bruit. Cependant, il nous manque l'information σ_R qui est liée à l'amplitude de l'ondelette. En effet, si nous multiplions l'ondelette par un coefficient α , nous divisons l'amplitude de la réflectivité par α . En pratique, nous prenons souvent f_n proportionnel à la valeur maximum du gain en fréquence de l'ondelette. Pour avoir une estimation de l'ondelette, plusieurs stratégies sont possibles, par exemple :

- mesure de l'ondelette avec un capteur de référence proche de la source,
- simulation de l'ondelette avec un modèle théorique de la source,
- estimation par recalage de l'arrivée principale de l'ondelette et moyenne spatiale.

Les bonnes performances du filtrage de Wiener dépendent de la qualité de l'estimation de l'ondelette et du facteur de bruit. L'avantage prépondérant reste ce compromis entre déconvolution et amplification du bruit.

1.4.2 Déconvolution semi-aveugle : ondelette inconnue, phase connue

Si l'ondelette est inconnue ou partiellement inconnue, nous devons faire intervenir des hypothèses supplémentaires portant généralement sur la réflectivité. Nous pouvons faire des hypothèses statistiques sur $r(t)$. Ensuite, il faut ajuster le filtre déconvolveur g afin que la sortie $y(t) = g \star d(t)$ tende à vérifier la même propriété. La figure 1.10 résume ce principe :

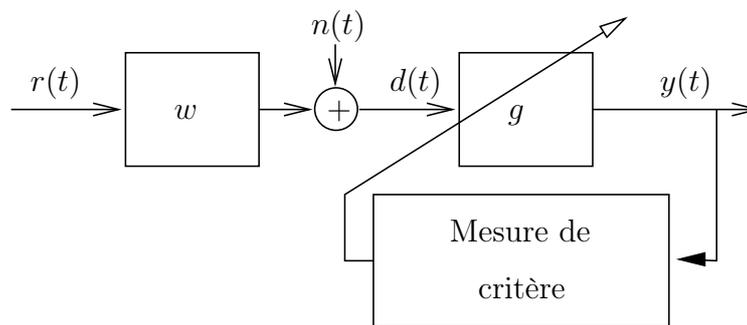


FIG. 1.10 – Le système de convolution-déconvolution.

L'hypothèse généralement utilisée est la blancheur du signal de réflectivité $r(t)$. Dans le cas sans bruit, si la réflectivité n'est pas gaussienne, le système est identifiable à une indétermination d'amplitude et un retard près, *i.e.* $y(t) = Ar(t - \tau)$, où A est le rapport d'amplitude et τ le retard. Nous nous appuyerons sur les outils tels que les multicorrélations et multispectres définis au paragraphe 1.3.1 pour mesurer la blancheur du signal de sortie. Si l'ondelette est inconnue, mais sa phase est en revanche connue, le plus simple et le plus naturel est d'utiliser les statistiques d'ordre 2. En effet, nous avons montré précédemment que les statistiques d'ordre 2 permettent l'estimation du module du gain du filtre $G(\nu)$ mais pas de la phase du filtre. L'ensemble des méthodes de déconvolution semi-aveugles et aveugles est reporté dans le tableau de la partie 1.4.4. Elles sont organisées par domaine de travail et par statistiques utilisées avec les avantages et inconvénients de chacune.

1.4.2.1 Égalisation spectrale

L'équation de convolution étant plus simple dans le domaine fréquentiel que dans le domaine temporel, il est aisé d'écrire en fréquence l'estimation du gain du filtre déconvolveur. Si nous souhaitons une sortie avec une densité spectrale de puissance $S_{2Y}(\nu)$ parfaitement constante, nous pouvons poser $|G(\nu)| = 1/\sqrt{S_{2D}(\nu)}$, où $S_{2D}(\nu)$ est la densité spectrale de puissance des données. Comme pour le filtrage de Wiener, nous ajoutons un facteur de bruit f_n pour limiter l'amplification du bruit :

$$|G(\nu)| = \frac{1}{\sqrt{S_{2D}(\nu) + f_n^2}}, \quad (1.72)$$

la densité spectrale de puissance de la sortie s'exprime par :

$$S_{2Y}(\nu) = \frac{S_{2D}(\nu)}{S_{2D}(\nu) + f_n^2}. \quad (1.73)$$

Nous remarquons que la densité spectrale de puissance de la sortie n'est constante que dans la bande de fréquence où la dsp des données est grande. Cette bande de fréquence correspond à la bande passante de l'ondelette. Nous avons, comme pour le filtrage de Wiener, un compromis entre la déconvolution (qualité du blanchiment) et l'amplification du bruit. L'implémentation en fréquence est simple, mais ne peut donner lieu à des algorithmes adaptatifs. De plus, nous nous heurtons au problème de l'estimation de la densité spectrale de puissance. L'implémentation dans le domaine temporel est à première vue plus complexe mais elle est possible en s'appuyant sur une paramétrisation simple du filtre.

1.4.2.2 Méthodes dans le domaine temporel

Avant d'estimer le filtre déconvolveur g , nous devons faire un bref rappel sur les différentes paramétrisations des filtres.

Paramétrisation MA, AR, ARMA

Le modèle le plus simple pour un filtre h est le modèle à réponse impulsionnelle finie (RIF) ne comportant qu'un nombre fini de coefficients non nuls. Nous avons pour un filtre causal :

$$H(z) = b_0 + b_1 z^{-1} + b_2 z^{-2} + \dots + b_P z^{-P}, \quad (1.74)$$

où P représente l'ordre du filtre. Ce type de filtre est aussi appelé MA pour Moving Average ou Moyenne Ajustée. L'estimation du filtre h se ramène à l'estimation du vecteur $\mathbf{B} = [b_0, \dots, b_P]$. Afin de modéliser des filtres résonants à réponse impulsionnelle infinie avec un nombre fini de coefficients, nous définissons les filtres autorégressifs (AR) :

$$H(z) = \frac{1}{a_0 + a_1 z^{-1} + a_2 z^{-2} + \dots + a_Q z^{-Q}}, \quad (1.75)$$

où Q est l'ordre du filtre. En posant $\mathbf{A} = [a_0, \dots, a_Q]$ le vecteur de paramètre décrit entièrement le filtre h . Le filtre inverse d'un filtre MA d'ordre P (resp. AR d'ordre Q) est un

filtre AR d'ordre P (resp. MA d'ordre Q) avec les mêmes coefficients.

De façon plus générale, nous pouvons définir le gain des filtres ARMA d'ordre (P, Q) par :

$$H(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2} + \dots + b_P z^{-P}}{a_0 + a_1 z^{-1} + a_2 z^{-2} + \dots + a_Q z^{-Q}}. \quad (1.76)$$

Suivant le type de filtre choisi (MA, AR, ARMA) pour le filtre déconvolueur, il existe différents algorithmes de déconvolution. Ils estiment les paramètres du filtre en rendant le signal de sortie $y(t)$ le plus blanc possible. Plusieurs façons d'aborder le problème de déconvolution existent dans le domaine temporel. Dans une première technique, nous écrivons d'abord les équations des multicorrélations ou multispectres entre la réflectivité $r(t)$ et les données $d(t)$ en faisant apparaître l'ondelette (ou les coefficients de sa paramétrisation). Puis, avec l'hypothèse de blancheur nous parvenons à des équations d'estimation de l'ondelette. Ensuite, nous filtrons les observations par le filtre déconvolueur à l'ondelette pour obtenir la sortie estimée. Une seconde technique consiste à considérer directement le filtre déconvolueur et la sortie. Nous écrivons les relations entre les multispectres de la sortie $y(t)$ et des données $d(t)$ et avec les hypothèses de blancheur sur la sortie $y(t)$ nous estimons le filtre déconvolueur. Les deux approches donnent exactement les mêmes équations d'estimation du filtre déconvolueur dans le cas sans bruit. Mais, la différence d'approche peut amener des confusions entre filtres estimés de type MA ou AR. En effet, si dans la première approche nous cherchons à estimer un filtre AR(Q), avec la seconde nous allons estimer un filtre MA($P=Q$). Dans la suite, nous privilégierons la seconde approche avec l'estimation du filtre déconvolueur par le biais de la mesure de blancheur sur la sortie. Nous faisons ce choix pour rendre la comparaison plus facile avec les algorithmes développés dans les chapitres suivants qui sont le coeur de notre travail et qui utilisent ce formalisme. Les méthodes présentées dans la suite se différencient au travers de deux aspects : (i) le modèle du filtre déconvolueur (MA, AR, ARMA) (ii) la mesure de blancheur de la sortie.

Algorithme de Yule-Walker

L'algorithme de Yule-Walker [LC77, WU83] consiste à prendre un modèle MA d'ordre P pour le filtre déconvolueur $g(t)$ défini par le vecteur de paramètres $\mathbf{B} = [b_0, b_1, \dots, b_P]^T$. Si le filtre est d'ordre P nous devons estimer ces $P + 1$ coefficients. En raison de l'indétermination d'échelle, nous pouvons normaliser le premier coefficient b_0 à 1 et nous avons donc P coefficients à déterminer $\mathbf{B} = [b_1, \dots, b_P]^T$. Nous savons que :

$$y(t) = d(t) + \sum_{i=1}^P b_i d(t-i). \quad (1.77)$$

En multipliant l'équation (1.77) par $d(t-n)$ pour $n > 0$ et en prenant l'espérance mathématique nous obtenons :

$$E[y(t)d(t-n)] = E[d(t)d(t-n)] + \sum_{i=1}^P b_i E[d(t-i)d(t-n)]. \quad (1.78)$$

Or $E[y(t)d(t-n)] = 0$ pour $n > 0$ car $d(t-n)$ peut s'écrire comme une combinaison linéaire du passé de $y(t-n)$, donc des échantillons $y(t-n)$, $y(t-n-1), \dots$. Or $y(t)$ est blanc à l'ordre deux donc sa fonction d'autocorrélation est $C_{2Y}(n) = \sigma_Y^2 \delta(n)$. Ainsi, l'équation (1.78) s'exprime par :

$$C_{2D}(n) + \sum_{i=1}^P b_i C_{2D}(|n-i|) = 0 \quad \forall n > 0. \quad (1.79)$$

En écrivant la relation précédente pour les valeurs de $n = 1, \dots, P$, nous obtenons le système d'équations suivant :

$$\begin{bmatrix} C_{2D}(0) & C_{2D}(1) & C_{2D}(2) & \dots & C_{2D}(P-1) \\ C_{2D}(1) & C_{2D}(0) & C_{2D}(1) & \dots & C_{2D}(P-2) \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ C_{2D}(P-1) & C_{2D}(P-2) & C_{2D}(P-3) & \dots & C_{2D}(0) \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ \vdots \\ b_P \end{bmatrix} = - \begin{bmatrix} C_{2D}(1) \\ C_{2D}(2) \\ \vdots \\ \vdots \\ C_{2D}(P) \end{bmatrix}. \quad (1.80)$$

L'algorithme de Yule-Walker estime les paramètres $\mathbf{B} = [b_1, \dots, b_P]^T$ en inversant le système (1.80). Nous pouvons tirer parti de la structure Toeplitz de la matrice pour réaliser une inversion rapide par récurrence sur l'ordre du filtre P [Pap81].

La méthode de Yule-Walker est très peu coûteuse en temps de calcul, mais le système d'équations est fréquemment mal conditionné dans le cas où l'ordre du filtre P est surestimé. Nous ajoutons une matrice λId à la matrice à inverser pour faciliter l'inversion. Id est la matrice identité et λ un réel positif petit. De plus, l'algorithme n'est pas très robuste au bruit additif sur les données. Enfin, il n'est pas sensible à la phase de l'ondelette car l'information de phase n'est pas conservée par la fonction d'autocorrélation. Pour la détermination de l'ordre, qui est le principal paramètre de l'algorithme, de nombreux critères ont été développés : AIC [Aka73], BIC [Sch78], MDL [Ris78]... La méthode de Yule-Walker estime le filtre à phase minimale et est aussi appelée déconvolution prédictive [Yil87, RT01]. En effet, l'équation (1.77) revient à prédire $y(t)$ en utilisant le passé des données $d(t)$.

Algorithme de Durbin

Pour l'algorithme de Durbin, l'ondelette répond à un modèle MA, donc le filtre déconvolveur est un modèle autorégressif. La méthode consiste à se ramener à l'algorithme de Yule-Walker. En effet, si $S_{2D}(\nu)$ est la densité spectrale de puissance de données résultant d'un filtre MA alors la densité spectrale de puissance "inverse" $\tilde{S}_{2D}(\nu) = 1/S_{2D}(\nu)$ correspond à un processus issu d'un filtrage autorégressif. Ainsi, l'algorithme de Durbin se décrit de la façon suivante :

1. Estimation de la fonction de corrélation $C_{2D}(\tau)$ des données ;
2. Estimation de la densité spectrale de puissance des données $S_{2D}(\nu) = TF[C_{2D}(\tau)]$;
3. Calcul $\tilde{S}_{2D}(\nu) = 1/S_{2D}(\nu)$;

4. Calcul de $\tilde{C}_{2D}(\tau) = TF^{-1}[\tilde{S}_{2D}(\nu)]$,
5. Application de l'algorithme de Yule Walker à $\tilde{C}_{2D}(\tau)$.

où $TF[\cdot]$ et $TF^{-1}[\cdot]$ sont respectivement les transformées de Fourier directe et inverse. Les coefficients du filtre AR estimés sont les mêmes que ceux du filtre MA recherché. L'inversion de la dsp ($1/S_{2D}(\nu)$) doit être faite avec précaution : en pratique nous effectuons $1/(S_{2D}(\nu) + f_n^2)$, où f_n est un facteur de bruit limitant les problèmes numériques.

Algorithme de Durbin-Yule-Walker

Pour estimer les filtres avec une paramétrisation ARMA, nous utilisons la méthode de Durbin-Yule-Walker. L'estimation des paramètres AR $[a_1, \dots, a_Q]$ et des paramètres MA $[b_1, \dots, b_P]$ est faite en deux étapes :

1. Estimation des paramètres MA ($\mathbf{B} = [b_1, \dots, b_P]$) avec un système similaire à celui de (1.80) de P équations construites en utilisant la relation (1.79) mais avec $n = Q+1, \dots, Q+P$. Nous noterons que des retards supérieurs à l'ordre Q de la partie AR du filtre déconvolveur permettent d'avoir une fonction d'autocorrélation indépendante de \mathbf{A} la partie AR.
2. Annulation de la partie MA avec le filtre déterminé dans l'étape précédente. Puis nous appliquons l'algorithme de Durbin sur les nouvelles données pour estimer les coefficients $\mathbf{A} = [a_1, \dots, a_Q]$ de la partie AR.

Cette méthode a les mêmes avantages et inconvénients que les méthodes de Yule-Walker et de Durbin. De plus, la détermination de la partie MA se fait avec une variance assez grande car l'estimation de la fonction d'autocorrélation pour des retards assez importants est difficile. Le système de Yule-Walker pour la partie MA étant très souvent mal-conditionné.

Détermination de la phase

Les méthodes précédentes permettent de déterminer le gain du filtre déconvolveur. Pour l'égalisation spectrale nous déterminons le module du filtre $|G(\nu)|$. Pour les méthodes de Yule-Walker et Durbin, nous estimons des filtres à phase minimale. La phase du filtre étant connue *a priori*, nous devons ensuite l'ajuster. La phase étant rarement connue de façon analytique, les hypothèses classiques sont de filtres à phase nulle ou minimale. Dans le cas d'un filtre à phase nulle, nous écrivons simplement $G(\nu) = |G(\nu)|$. Pour un filtre à phase minimale, la phase peut être déterminée à partir du module $|G(\nu)|$ par l'intermédiaire de la relation de Bode-Bayard :

$$\phi_G(\nu) = H[\log |G(\nu)|], \quad (1.81)$$

où $H[x]$ est la transformée de Hilbert de x .

Dans le cas de la sismique, il est difficile d'avancer des hypothèses sur la phase de l'ondelette. En effet, aucune justification physique ne permet de dire que l'ondelette est à phase minimale ou nulle. Par conséquent, la phase est souvent supposée inconnue. Nous sommes dans un contexte aveugle car nous connaissons uniquement les données.

1.4.3 Déconvolution aveugle : ondelette (module et phase) inconnue

Les statistiques d'ordre 2 étant insensibles à la phase de l'ondelette et sans hypothèse *a priori* sur la phase, nous devons utiliser les statistiques d'ordre supérieur à 2 pour la déterminer. Le module peut être déterminé avec les statistiques d'ordre 2 ou par celles d'ordre supérieur. Dans le domaine fréquentiel, nous estimons le module du filtre à l'aide des statistiques d'ordre 2, et nous chercherons uniquement la phase avec les statistiques d'ordre 3 ou 4 (bi ou trispectre). Dans le domaine temporel, les deux approches existent. Soit nous estimons le module à l'ordre 2 et cherchons la phase avec les statistiques d'ordre supérieur (kurtosis), soit nous cherchons le filtre déconvolveur directement, avec uniquement les statistiques d'ordre 3 ou 4 (bi ou tricorrélation).

1.4.3.1 Domaine fréquentiel

Dans ce paragraphe, nous présentons les méthodes développés précédemment dans le domaine fréquentiel pour l'estimation de la phase de l'ondelette basée sur le bispectre et trispectre définis au paragraphe 1.3.1.

Dans une première étape, nous déterminons le module du filtre déconvolveur $|G(\nu)|$ par l'intermédiaire d'une méthode décrite précédemment à l'ordre 2 (Égalisation spectrale, Yule-Walker, Durbin). Dans une seconde étape, la phase du filtre déconvolveur sera déterminée à l'aide du bispectre $S_{3D}(\nu_1, \nu_2)$ ou du trispectre $S_{4D}(\nu_1, \nu_2)$ des données.

Méthode du bispectre

Nous cherchons à lier le bispectre $S_{3Y}(\nu_1, \nu_2)$ de la sortie de la déconvolution $y(t)$ et celui des données $S_{3D}(\nu_1, \nu_2)$. D'après (1.66) avec $r = 3$, nous savons que :

$$S_{3Y}(\nu_1, \nu_2) = G(\nu_1)G(\nu_2)G^*(\nu_1 + \nu_2)S_{3D}(\nu_1, \nu_2). \quad (1.82)$$

La sortie de la déconvolution $y(t)$ doit être un processus blanc. Par conséquent son bispectre doit être constant et à phase nulle, *i.e* $S_{3Y}(\nu_1, \nu_2) = \kappa_{3Y}$, où κ_{3Y} est une constante réelle égale au cumulants d'ordre 3 de $y(t)$. Le bispectre des observations $S_{3D}(\nu_1, \nu_2)$ peut s'écrire de la façon suivante :

$$S_{3D}(\nu_1, \nu_2) = |S_{3D}(\nu_1, \nu_2)|e^{j\Psi_{3D}(\nu_1, \nu_2)}, \quad (1.83)$$

où $\Psi_{3D}(\nu_1, \nu_2)$ représente sa phase. En utilisant la notation $G(\nu) = |G(\nu)|e^{j\phi_G(\nu)}$ avec (1.82), nous obtenons pour les phases :

$$\phi_G(\nu_1) + \phi_G(\nu_2) - \phi_G(\nu_1 + \nu_2) + \Psi_{3D}(\nu_1, \nu_2) = 0. \quad (1.84)$$

Suivant (1.84), l'estimateur minimisant l'erreur quadratique moyenne de $\phi_G(\nu_1)$ dans (1.84) est [Lan90] :

$$\phi_G(\nu_1) = -\frac{1}{T} \sum_{\nu_2=0}^{T-1} \Psi_{3D}(\nu_1, \nu_2). \quad (1.85)$$

Cette relation est aussi liée à la symétrie hermitienne de la transformée de Fourier. En effet, si nous sommions $\phi_G(\nu_2)$ ou bien $\phi_G(\nu_1 + \nu_2)$ sur un ensemble de fréquences ν_2 décrivant toute la période T (T est le nombre d'échantillons de calcul de la TFD car nous travaillons sur des signaux échantillonnés), ces termes se compensent.

Comme toutes les statistiques d'ordre impair, le bispectre est nul ou quasi-nul pour des observations dont la densité de probabilité est symétrique, ce qui est le cas de la majorité des données réelles. Nous préférons donc utiliser les statistiques d'ordre 4 par le biais du trispectre $S_{4Y}(\nu_1, \nu_2, \nu_3)$.

Méthode du trispectre

En suivant la même démarche que précédemment et comme les trispectres sont liés d'après (1.66) par :

$$S_{4Y}(\nu_1, \nu_2, \nu_3) = G(\nu_1)G(\nu_2)G(\nu_3)G^*(\nu_1 + \nu_2 + \nu_3)S_{4D}(\nu_1, \nu_2, \nu_3). \quad (1.86)$$

L'estimateur de la phase du filtre déconvolveur au sens des moindres carrés devient :

$$\phi_G(\nu_1) = -\frac{1}{T^2} \sum_{\nu_3=0}^{T-1} \sum_{\nu_2=0}^{T-1} \Psi_{4D}(\nu_1, \nu_2, \nu_3), \quad (1.87)$$

où $\Psi_{4D}(\nu_1, \nu_2, \nu_3)$ est la phase du trispectre des observations. L'estimation de la phase avec le trispectre et le bispectre est indépendante de celle du module du filtre, ce qui est une limitation non négligeable. En effet, tout est fait comme si le module et la phase du filtre étaient indépendants, or cette hypothèse n'est pas réaliste. De plus, l'estimation de la phase du filtre est délicate pour plusieurs raisons. Tout d'abord, l'estimateur du trispectre possède une variance assez grande, ensuite les équations (1.84) ou (1.87) liant la phase du filtre à celle du bispectre ou du trispectre ne prennent pas en compte le bruit additif sur les données. Or, si l'ondelette est à bande limitée, le domaine de fréquence du trispectre ou du bispectre, où nous retrouvons l'information de l'ondelette sera lui aussi très limité. Pour une ondelette à bande limitée de largeur de bande B , (*i.e.* $|X(\nu)| = 0$ pour $|\nu| > B$) le domaine de définition du trispectre est approximativement la partie de l'espace définie par $\{(\nu_1, \nu_2, \nu_3) \in \mathbb{R}^3 \mid |\nu_1| < B, |\nu_2| < B, |\nu_3| < B, |\nu_1 + \nu_2 + \nu_3| < B\}$. Ainsi, en sommant dans (1.87) sur tout le domaine des fréquences nous pouvons penser que l'estimation de la phase du filtre déconvolveur sera de très mauvaise qualité dans le cas d'une ondelette à bande limitée.

1.4.3.2 Domaine temporel

Dans le domaine temporel, il existe deux approches distinctes. La première estime le filtre à phase minimale avec les statistiques d'ordre 2 et ensuite utilise l'information sur les pôles et les zéros pour déterminer la phase du filtre. Le critère souvent utilisé est le kurtosis. La seconde technique estime directement les coefficients du filtre dans le cas d'un filtre déconvolveur autorégressif à l'aide de la bicorrélation et tricorrélation.

Méthode indirecte à l'ordre 4 : le kurtosis

Pour estimer les filtres AR, MA, ARMA à phase non minimale, nous pouvons utiliser le cumulatif d'ordre 4 normalisé à savoir : le kurtosis de la sortie \mathcal{K}_{4Y} défini dans le paragraphe 1.2.1. On rappelle que le kurtosis mesure l'écart à la gaussienne. Or, d'après le théorème central limite, le filtrage d'une séquence iid tend à gaussianiser la distribution. Comme nous sommes dans le cas de réflectivité sur-gaussienne, la sortie du filtre de déconvolution doit avoir le kurtosis maximal. Dans le cas de signaux sous-gaussiens, le kurtosis sera négatif et devra être minimum. Dans les deux cas, il est équivalent à maximiser la valeur absolue du kurtosis.

Boumahdi [Bou95, Bou96] utilise la maximisation de la valeur absolue du kurtosis de sortie pour lever l'indétermination de phase dans le cas de filtre AR, MA ou ARMA. Il estime dans une première étape le filtre à phase minimale avec les techniques à l'ordre 2 présentées dans le paragraphe 1.4.2.2. Considérons un filtre ARMA(P,Q) dont les zéros sont $[\alpha_1, \dots, \alpha_P]$ et les pôles sont $[z_1, \dots, z_Q]$. Si nous modifions un ou plusieurs pôles ou zéros en prenant leurs symétriques par rapport au cercle unité, nous obtenons un filtre spectralement équivalent au filtre d'origine, c'est à dire le module du spectre du filtre ne change pas, seule la phase est modifiée. (Le symétrique d'un complexe z par rapport au cercle unité est $1/z^*$). Ainsi, après estimation du filtre à phase minimale, Boumahdi calcule la sortie pour tous les filtres ARMA d'ordre (P,Q) spectralement équivalents, soit 2^{P+Q} filtres, et sélectionne celui qui donne le kurtosis de sortie maximal en valeur absolue.

La blancheur à l'ordre 2 n'est pas modifiée par la seconde étape. Cette méthode est assez longue à mettre en oeuvre si P et Q sont élevés car le nombre de filtres spectralement équivalents est grand. Ensuite, l'estimation des pôles proches du cercle unité est très difficile avec les statistiques d'ordre 2 en présence de bruit. Dans ce cas, le kurtosis est utilisé pour estimer uniquement la phase du filtre. Une autre approche [Wig78, Don81, Whi88] consiste à maximiser la valeur absolue du kurtosis de sortie $|\mathcal{K}_{4Y}|$ pour estimer directement le filtre à phase non minimale de sortie. Nous noterons l'ambiguïté entre blancheur et sur-gaussianité. En effet, en maximisant $|\mathcal{K}_{4Y}|$ en fait, nous mesurons la sur-gaussianité et non la blancheur, et cette approche par la non-gaussianité fera l'objet des développements du chapitre 4.

Dans cette démarche, les statistiques d'ordre supérieur à 2 permettent uniquement de lever l'indétermination de la phase du filtre. Les démarches présentées par la suite estiment directement le filtre déconvolueur en lui octroyant un modèle AR(Q). Nous devons estimer le vecteur de paramètres $\mathbf{A} = [a_0, \dots, a_Q]$. Pour réaliser cette estimation, nous devons écrire les relations entre la bicorrélation ou la tricorrélation des données $d(t)$ et celles de la sortie $y(t)$ en nous appuyant sur l'équation (1.65). En fait, pour simplifier les écritures, $d(t)$ sera vu comme le résultat du filtrage de $y(t)$ par un filtre MA dont les coefficients seront ceux du filtre AR recherché, *i.e.* $\mathbf{B} = \mathbf{A}$. De plus, l'hypothèse de blancheur de $y(t)$ est incorporée.

Méthode à l'ordre 3 directe : bicorrélation

Les premiers développements sont l'oeuvre de Giannakis *et al.* [Gia87, GM89]. Ils uti-

lisent la bicorrélation pour l'estimation d'un filtre direct MA à phase non minimale, soit un filtre déconvolveur autorégressif. Avec l'équation (1.65) à l'ordre 3, nous avons pour la bicorrélation :

$$\kappa_{3X} \sum_{i=0}^Q a_i a_{i+\tau_1} a_{i+\tau_2} = C_{3D}(\tau_1, \tau_2) \quad [\tau_1, \tau_2] \in [-Q, Q]^2. \quad (1.88)$$

Par convention, $a_i = 0$ si $i < 0$ ou $i > Q$, et $a_0 = 1$. La première idée, très simple, est de sélectionner des valeurs particulières de τ_1 et τ_2 dans la relation précédente pour que la somme se réduise à un terme unique. Par exemple, si nous choisissons $\tau_1 = 0$ et $\tau_2 = Q$, alors nous avons :

$$\kappa_{3X} a_0^2 a_Q = C_{3D}(0, Q). \quad (1.89)$$

Ensuite, nous obtenons avec $\tau_1 = Q$ la relation suivante :

$$\kappa_{3X} a_{\tau_2} a_Q a_0 = C_{3D}(Q, \tau_2), \quad (1.90)$$

alors les coefficients du filtre déconvolveur vérifient :

$$a_\tau = \frac{C_{3D}(Q, \tau)}{C_{3D}(0, Q)}. \quad (1.91)$$

Dans (1.91), l'estimation est très sensible au biais des estimateurs de la bicorrélation. De plus, les valeurs de la bicorrélation pour des retards importants sont très faibles ce qui rend d'autant plus difficile l'estimation avec ce type de relation. Pour remédier à cela, un second algorithme se base simultanément sur les statistiques d'ordre 2 et 3. Avec l'équation (1.65), à l'ordre 2 et 3 et en se limitant au cas où les deux retards pour la bicorrélation sont égaux, nous avons :

$$\kappa_{2X} \sum_{i=\max(0, -\tau)}^{\min(Q, Q-\tau)} a_i a_{i+\tau} = C_{2D}(\tau), \quad \tau = -Q, \dots, Q, \quad (1.92)$$

$$\kappa_{3X} \sum_{i=\max(0, -\tau)}^{\min(Q, Q-\tau)} a_i a_{i+\tau}^2 = C_{3D}(\tau, \tau), \quad \tau = -Q, \dots, Q. \quad (1.93)$$

Ensuite, en réunissant les équations (1.92) et (1.93) précédentes en une seule, on estime les coefficients $\mathbf{A} = [a_1, \dots, a_P]$ du filtre déconvolveur par récurrence. Une extension pour l'estimation des filtres ARMA non causaux est proposée par Giannakis et Swami [GS90]. Elle s'inspire grandement de la méthode de Yule-Walker-Durbin à l'ordre 2, avec une estimation en deux temps pour la partie AR et MA en faisant intervenir des retards importants pour supprimer l'effet de la partie AR. Enfin, la version multidimensionnelle MIMO (Multi-Input Multi-Output) est détaillée dans [GIM89]. Le mauvais conditionnement du problème, dû à l'annulation de la bicorrélation pour les signaux à distributions de probabilité symétrique, amène naturellement aux statistiques d'ordre 4.

Méthode à l'ordre 4 directe : tricorrélation

Pour la tricorrélation, nous avons d'après l'équation (1.65) :

$$\kappa_{4Y} \sum_{i=0}^Q a_i a_{i+\tau_1} a_{i+\tau_2} a_{i+\tau_3} = C_{4D}(\tau_1, \tau_2, \tau_3). \quad (1.94)$$

Giannakis a proposé une méthode très proche de celle de la bicorrélation décrite précédemment, en sélectionnant des valeurs particulières des retards τ_1, τ_2, τ_3 [Gia87]. L'estimation est faite par :

$$a_\tau = \frac{C_{4D}(Q, Q, \tau)}{C_{3D}(Q, Q, 0)}, \quad \tau = 1, \dots, Q. \quad (1.95)$$

Cette méthode est limitée par l'instabilité numérique due à la variance élevée d'estimation de la tricorrélation. Comon [Com92] propose de réduire l'instabilité en utilisant un plus grand nombre de valeurs de la tricorrélation de la sortie. Cela définit un système d'équation surdéterminé où il y a plus d'équations que d'inconnues. Comon s'intéresse à la famille d'équations suivante :

$$a_{\tau_3} C_{4D}(\tau_1, \tau_2, Q) - a_{\tau_1} C_{4D}(\tau_3, \tau_2, Q) = 0, \quad 0 \leq \tau_3 < \tau_1 \leq Q \text{ et } 0 \leq \tau_2 \leq Q, \quad (1.96)$$

$$a_{\tau_1}^2 C_{4D}(\tau_2, \tau_2, Q) - a_{\tau_2}^2 C_{4D}(\tau_1, \tau_1, Q) = 0, \quad 0 \leq \tau_2 < \tau_1 \leq Q. \quad (1.97)$$

Les relations précédentes donnent respectivement $Q(Q+1)^2/2$ et $Q(Q+1)/2$ équations. Les relations (1.97) sont difficilement utilisables en raison de l'indétermination de signe sur les coefficients a_τ . Avec (1.96), les coefficients sont estimés au sens des moindres carrés du système linéaire surdéterminé. Une extension MIMO est proposée dans [Com92].

1.4.4 Résumé

Le tableau suivant résume l'ensemble des principales méthodes de déconvolution semi-aveugles et aveugles, en les classant suivant les statistiques utilisées et le domaine de travail. Les principaux avantages (mentionnés par "+") et inconvénients (mentionnés par "-") de chacune des méthodes sont donnés.

Toutes les méthodes présentées dans ce chapitre sont indépendantes de la distribution du signal de réflectivité. Le seul cas écarté est la distribution gaussienne qui est la condition pour l'identifiabilité de l'ondelette. En effet, nous sélectionnons les statistiques d'ordre supérieur dans l'ordre croissant pour lever l'indétermination de la phase de l'ondelette. Nous nous intéresserons jamais à la distribution du signal. Nous ne pouvons pas assurer que nous sélectionnons les meilleures statistiques pour décrire la blancheur du signal. Nous rappelons que l'utilisation des statistiques d'ordre 2 est équivalente au choix d'une distribution *a priori* gaussienne pour les signaux. Dans le chapitre suivant, nous présenterons une nouvelle mesure de blancheur : le taux d'information mutuelle. Cette mesure utilise l'ensemble des statistiques d'ordre supérieur par le biais de l'entropie.

TAB. 1.1 – Résumé des méthodes de déconvolution utilisant la blancheur.

		Domaine de travail	
		Temps	Fréquence
Ordre 2	Yule Walker et Durbin		Egalisation spectrale
	+ : simple + : nombre fini de coefficients		+ : simple + : régularisation vis-à-vis du bruit additif
	- : détermination de l'ordre - : sensibilité au bruit		- : estimation de la DSP - : réglage du facteur de bruit
Ordre 3	Bicorrélation [Gia87, GIM89, GS90]		Bispectre
	+ : simplicité avec nombre fini de coefficients + : utilise ordre 2 et 3		+ : simple + : choix de la stratégie de l'estimation du module
	- : détermination de l'ordre - : sensibilité de la phase au bruit - : n'utilise que certains retards		- : sensibilité de la phase au bruit - : domaine de définition du bispectre limité - : estimation séparée du module et de la phase
Ordre 4	Kurtosis [Bou95, Bou96]		Trispectre
	+ : simplicité		+ : simple
	- : temps calcul		+ : choix de la stratégie de l'estimation du module
	- : détermination phase et module séparée		- : estimation du trispectre difficile
	- : pôles et zéros proches du cercle unité difficiles à déterminer		- : sensibilité au bruit
	Tricorrélation [Com92]		
+ : nombre fini de coefficients + : équations non linéaires mais inversibles			
- : n'utilise pas toutes les valeurs de la tricorrélation - : bruit additif non pris en compte			

Chapitre 2

Déconvolution aveugle dans le domaine temporel par le taux d'information mutuelle

Sommaire

2.1	Développements théoriques	55
2.2	Algorithme MAMV de déconvolution d'ondelette AR	57
2.2.1	Gradient du critère	57
2.2.2	Algorithme MAMV	58
2.2.3	Données simulées	60
2.3	Algorithme ARMV de déconvolution d'ondelette MA	62
2.3.1	Gradient du critère	63
2.3.2	Algorithme ARMV	64
2.3.3	Données simulées	65
2.4	Estimation de la fonction score	66
2.4.1	Estimateur de fonction score	67
2.4.1.1	Estimateur à noyaux	67
2.4.1.2	Estimation directe	68
2.4.2	Étude des performances des estimateurs	70
2.4.2.1	Fonctions scores théoriques	71
2.4.2.2	Comparaison des estimateurs	72
2.5	Conclusion	75

Après avoir présenté le principe de l'imagerie sismique, le modèle des données et les outils statistiques et un état de l'art des méthodes de déconvolution semi-aveugles et aveugles, nous proposons dans ce chapitre une mesure de blancheur utilisant l'ensemble des statistiques d'ordre supérieur : le taux d'information mutuelle (TIM). Dans un premier temps, nous présenterons les fondements théoriques du taux d'information mutuelle. Nous nous attacherons à simplifier son expression pour la rendre par la suite utilisable dans un algorithme de déconvolution aveugle. Nous détaillerons un algorithme de déconvolution [TSiCJ01] basé sur la minimisation du taux d'information mutuelle pour estimer le filtre déconvolveur g de déconvolution des données dans le domaine temporel. Cet algorithme utilise un modèle MA pour identifier le filtre déconvolveur. Nous proposons une extension de cet algorithme avec un modèle AR du filtre déconvolveur. Pour chacun de ces deux algorithmes, nous présenterons quelques résultats de déconvolution sur des données synthétiques pour mettre en évidence leurs aptitudes à estimer des filtres à phase quelconque grâce à l'utilisation de l'ensemble des statistiques d'ordre supérieur. Dans la suite, nous insisterons sur le problème de l'estimation de la fonction score qui apparaît dans ces deux algorithmes. Cette estimation est équivalente à la sélection des statistiques les plus adaptées pour la déconvolution des données. Nous présenterons deux stratégies d'estimation par une méthode à noyaux et une méthode directe avec un modèle polynomial. Nous étudierons les performances de ces estimateurs. Le cas des distributions typiques des signaux sismiques sera bien entendu analysé avec précision. Après comparaison de ces estimateurs, nous sélectionnerons celui offrant le meilleur compromis entre robustesse et précision pour nos applications avec un temps de calcul acceptable.

2.1 Développements théoriques

Dans le domaine de la séparation de sources [Pha01, Pha02], on mesure classiquement l'indépendance d'un vecteur de variables aléatoires $\mathbf{z} = (z_1, \dots, z_n)$ de dimension n par l'information mutuelle définie par :

$$I(\mathbf{z}) = \sum_{i=1}^n H(z_i) - H(z_1, z_2, \dots, z_n), \quad (2.1)$$

où $H(z_i)$ est l'entropie de Shannon marginale de z_i , *i.e.* $H(z_i) = - \int_{\mathbb{R}} p_{z_i}(u) \log p_{z_i}(u) du$. De même, $H(z_1, z_2, \dots, z_n)$ est l'entropie de Shannon conjointe, *i.e.* $H(\mathbf{z}) = - \int_{\mathbb{R}^n} p_{\mathbf{z}}(\mathbf{u}) \log p_{\mathbf{z}}(\mathbf{u}) d\mathbf{u}$.

L'information mutuelle $I(\mathbf{z})$ possède la propriété intéressante d'être toujours positive. De plus, $I(\mathbf{z})$ s'annule si et seulement si les composantes du vecteur \mathbf{z} sont mutuellement indépendantes. Ainsi, l'information mutuelle est une mesure intéressante d'indépendance de variables aléatoires. Dans le cas de signaux markoviens, nous pouvons aussi nous intéresser à la définition de l'information mutuelle avec l'entropie conditionnelle du point vue temporel [HJP03, LJH04].

Dans le contexte de la déconvolution aveugle, nous considérons des processus stochastiques, *i.e.* des séquences de longueurs infinies, pour lesquelles la notion d'entropie n'est pas adaptée, par exemple le terme conjoint deviendrait infini dans le cas d'une séquence de longueur infinie. En revanche, nous pouvons utiliser la notion de taux d'entropie d'un processus stochastique $Z = \{Z_t\}$ définie par [CT91] selon la formule suivante :

$$\mathcal{H}(Z) = \lim_{T \rightarrow \infty} \frac{1}{T} H(Z_1, \dots, Z_T). \quad (2.2)$$

Le taux d'entropie a été aussi utilisé par Papoulis [Pap81]. Suivant [CT91], cette limite existe pour un processus stationnaire. Le taux d'information mutuelle (TIM) du processus stochastique $Z = \{Z_t\}$ est défini par :

$$I(Z) = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^{t=T} H(Z_t) - \mathcal{H}(Z), \quad (2.3)$$

où $H(Z_t)$ représente l'entropie marginale du processus Z_t et $\mathcal{H}(Z)$ le taux d'entropie du processus Z défini par (2.2). Pour clarifier les notations, Z représente le processus avec tous les échantillons $z(1), \dots, z(T)$ pris conjointement, alors que Z_t correspond au processus d'un seul échantillon $z(t)$. L'information mutuelle mesure l'indépendance entre des variables aléatoires. De même le taux d'information mutuelle $I(Z)$ mesure l'indépendance entre les échantillons. $I(Z)$ est toujours positif et s'annule si et seulement si Z est asymptotiquement un processus iid [CT91]. Ainsi, le TIM apparaît comme une mesure de blancheur intéressante pour la déconvolution aveugle. Cependant, l'expression du taux d'information mutuelle nécessite des simplifications pour parvenir au développement d'un algorithme de déconvolution aveugle. La première simplification proposée vient du caractère stationnaire du processus, la seconde s'attachera à simplifier le terme du taux d'entropie.

Sous hypothèse de stationnarité, l'entropie marginale du processus Z_t est indépendante du temps. Tous les termes de la somme dans l'équation (2.3) sont égaux : $H(Z_1) = \dots = H(Z_T)$, ainsi nous avons :

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^{t=T} H(Z_t) = H(Z_\tau), \quad \forall \tau. \quad (2.4)$$

En ajoutant l'hypothèse de stationnarité, le TIM (2.3) s'exprime simplement par :

$$I(Z) = H(Z_\tau) - \mathcal{H}(Z). \quad (2.5)$$

Dans (2.5), τ est arbitrairement choisi mais il n'y a aucune raison de privilégier une valeur particulière de τ . De plus, une estimation convenable de l'entropie marginale $H(Z_\tau)$ nécessite plusieurs réalisations du processus Z_τ , donc, en pratique, pour estimer cette entropie nous utiliserons tous les échantillons $z(1), \dots, z(T)$ comme T réalisations du processus Z_τ . Par la suite, nous noterons de façon équivalente l'entropie du processus Z_τ et celle d'une réalisation $z(\tau)$, soit $H(Z_\tau) = H(z(\tau)), \forall \tau$. Dans le contexte de la déconvolution aveugle, l'obtention d'une fonction coût plus simple que la définition du TIM en (2.5) est possible [TSiCJ01, Pha01, ZC04] en simplifiant le terme du taux d'entropie de la sortie Y à l'aide du lemme suivant :

Lemme 2.1. *Soit X un processus stochastique stationnaire et h un filtre linéaire invariant dans le temps, le taux d'entropie de la sortie $Y = h \star X$ vérifie (démonstration en annexe A.1) :*

$$\mathcal{H}(Y) = \mathcal{H}(h \star X) = \mathcal{H}(X) + \frac{1}{2\pi} \int_0^{2\pi} \log \left| \sum_{t=-\infty}^{+\infty} h(t) e^{-jt\theta} \right| d\theta. \quad (2.6)$$

Avec ce lemme appliqué à $Y = g \star D$, où D est le processus des observations, et en incorporant la simplification due à la stationnarité de (2.4), le TIM de la sortie déconvoluée (2.5) s'exprime par :

$$I(Y) = H(y(\tau)) - \mathcal{H}(D) - \frac{1}{2\pi} \int_0^{2\pi} \log \left| \sum_{t=-\infty}^{+\infty} g(t) e^{-jt\theta} \right| d\theta. \quad (2.7)$$

Comme le taux d'entropie des données $\mathcal{H}(D)$ est indépendant du filtre déconvolveur g , nous pouvons mesurer la blancheur de la sortie de la déconvolution Y de façon équivalente au taux d'information mutuelle (2.7) par le biais de :

$$\tilde{I}(Y) = H(y(\tau)) - \frac{1}{2\pi} \int_0^{2\pi} \log \left| \sum_{t=-\infty}^{+\infty} g(t) e^{-jt\theta} \right| d\theta. \quad (2.8)$$

$\tilde{I}(Y)$ aura la même propriété que le taux d'information mutuelle (2.7) d'être minimale quand le processus Y sera iid. Ce critère mesure la blancheur avec l'ensemble des statistiques d'ordre supérieur grâce à l'entropie. Pour des systèmes de Wiener composés d'une cascade d'un filtre linéaire homogène w et d'une fonction non linéaire inversible, Taleb *et al.* [TSiCJ01] estiment le filtre déconvolveur g dans le domaine temporel, en minimisant (2.8) par rapport à la réponse impulsionnelle $g(t)$. En choisissant un nombre fini de coefficients pour la réponse impulsionnelle, l'algorithme fait l'hypothèse d'un filtre déconvolveur MA. La méthode se posera donc en concurrente de la méthode de Yule-Walker. Une méthode assez similaire mais restreinte au système linéaire a été proposée avec l'utilisation de l'entropie de Renyi [EHP⁺04]. Nous pouvons montrer que l'algorithme qui minimise $\tilde{I}(Y)$ par rapport à la réponse impulsionnelle du filtre $g(t)$ est équivalent à une méthode de maximum de vraisemblance (MV). En effet, pour le maximum de vraisemblance nous supposons connue la distribution du processus Y recherché alors que, ici, elle sera estimée à chaque itération. En référence au modèle MA du filtre déconvolveur et à la méthode du maximum de vraisemblance, nous nommerons cet algorithme MAMV, soit :

$$g^{MAMV}(t) = \arg \min_{g(t)} \tilde{I}(Y). \quad (2.9)$$

Pour trouver ce minimum, Taleb *et al.* mettent en place une descente de gradient par rapport aux coefficients du filtre déconvolveur $g(t)$. Dans la section suivante nous développons le calcul de ce gradient. Puis, nous détaillons la structure complète de l'algorithme.

2.2 Algorithme MAMV de déconvolution d'ondelette AR

2.2.1 Gradient du critère

Nous allons calculer le gradient du critère (mesure de blancheur) défini par (2.8) par rapport à la réponse impulsionnelle du filtre déconvolveur $g(t)$. Dans le calcul, nous laissons

très souvent des sommes infinies, mais en fait nous cherchons un nombre fini de coefficients pour la réponse impulsionnelle du filtre déconvolveur, soit un vecteur $\mathbf{g} = [g(0), \dots, g(P)]$. Pour la dérivation du terme entropique de (2.8), nous utilisons la définition de l'entropie, soit :

$$\frac{\partial H(y(\tau))}{\partial g(t)} = -\frac{\partial}{\partial g(t)} E[\log p_Y(y(\tau))]. \quad (2.10)$$

Après permutation des opérateurs de dérivation et d'espérance mathématique et en écrivant les dérivées de fonctions composées, l'expression précédente s'exprime comme :

$$\frac{\partial H(y(\tau))}{\partial g(t)} = -E \left[\frac{d \log p_Y(y(\tau))}{dy(\tau)} \frac{\partial (g \star d)(\tau)}{\partial g(t)} \right]. \quad (2.11)$$

Pour progresser dans le calcul, nous définissons la fonction score du processus Y par :

$$\psi_Y(u) = -\frac{d}{du} \log p_Y(u). \quad (2.12)$$

Avec l'équation de convolution $y(t) = g \star d(t)$, nous pouvons écrire que :

$$\frac{\partial (g \star d)(\tau)}{\partial g(t)} = \frac{\partial \left\{ \sum_{u=-\infty}^{+\infty} g(u) d(\tau - u) \right\}}{\partial g(t)} = d(\tau - t). \quad (2.13)$$

À l'aide de la définition de la fonction score et de la relation (2.13), la dérivation du terme entropique (2.11) s'exprime simplement par :

$$\frac{\partial H(y(\tau))}{\partial g(t)} = E[\psi_Y(y(\tau)) d(\tau - t)]. \quad (2.14)$$

Pour la dérivation du second terme de (2.8) nous avons (démonstration en annexe A.2) :

$$\frac{\partial}{\partial g(t)} \left\{ \frac{1}{2\pi} \int_0^{2\pi} \log \left| \sum_{t=-\infty}^{+\infty} g(t) e^{-jt\theta} \right| d\theta \right\} = \bar{g}(-t), \quad (2.15)$$

où $\bar{g}(t)$ est le filtre déconvolveur de $g(t)$ (soit $\bar{g} \star g(t) = \delta(t)$ où $\delta(t)$ est le filtre identité). Avec (2.14) et (2.15), le gradient du critère s'exprime par :

$$\frac{\partial \tilde{I}(Y)}{\partial g(t)} = E[\psi_Y(y(\tau)) d(\tau - t)] - \bar{g}(-t). \quad (2.16)$$

Après le calcul du gradient du critère, nous allons expliquer par la suite la structure de l'algorithme MAMV proposé par Taleb *et al.* [TSiCJ01]

2.2.2 Algorithme MAMV

Afin de minimiser le taux d'information mutuelle de la sortie par rapport aux coefficients de la réponse impulsionnelle du filtre déconvolveur, nous pourrions appliquer une descente de gradient classique avec une mise à jour du filtre déconvolveur de la façon suivante :

$$g(t) \leftarrow g(t) - \mu \frac{\partial \tilde{I}(Y)}{\partial g(t)} = g(t) - \mu E[\psi_Y(y(\tau))d(\tau - t)] + \mu \bar{g}(-t), \quad (2.17)$$

où μ est un réel positif, appelé le pas du gradient. Implicitement dans ce cas, nous nous intéresserons à des variations du filtre déconvolveur g de la forme :

$$g \rightarrow g + \epsilon, \quad (2.18)$$

où ϵ est un "petit" filtre. Dans l'expression (2.17), nous devons estimer le filtre \bar{g} qui est le filtre déconvolveur de g . En fait, en s'inspirant des problèmes de séparation de sources, Taleb *et al.* proposent de s'intéresser à des petites variations du filtre déconvolveur $g(t)$ proportionnelles au filtre $g(t)$ soit :

$$g \rightarrow g + \epsilon \star g. \quad (2.19)$$

Ainsi la mise à jour du filtre déconvolveur peut être envisagée de la façon suivante :

$$g(t) \leftarrow g(t) - \mu \frac{\partial \tilde{I}(Y)}{\partial g(t)} \star g^- \star g(t), \quad (2.20)$$

avec $g^-(t)$ le filtre dont la réponse impulsionnelle est celle de $g(t)$ retournée, soit $g^-(t) = g(-t)$. Cette méthode de gradient s'inspire [SiC00, SiCTJ00, TSiCJ01] des développements faits dans le cadre des problèmes de séparation de sources. Lors de la minimisation d'un critère $J(B)$, nous effectuons la mise à jour suivante pour la matrice séparante B :

$$B \leftarrow B - \mu \frac{\partial J(B)}{\partial B} B^T B. \quad (2.21)$$

De cette manière et d'après [Car98], nous obtenons un algorithme équivariant vis-à-vis du filtre direct w . Ses performances de convergence ne dépendent pas de la forme du filtre direct w mais uniquement de l'initialisation. Après développement de (2.20), comme $g^-(t)$ est le filtre déconvolveur de $\bar{g}(-t)$, *i.e.* $g^-(t) \star \bar{g}(-t) = \delta(t)$ et $x(\tau - t) \star g^-(t) = y(\tau - t)$, la descente de gradient est faite par :

$$g(t) \leftarrow g(t) + \mu \{-E[\psi_Y(y(\tau))y(\tau - t)] + \delta\} \star g(t), \quad (2.22)$$

où

$$\gamma_{y, \psi_Y}(t) = E[\psi_Y(y(\tau))y(\tau - t)] \quad (2.23)$$

représente la corrélation entre la sortie déconvoluée $y(\tau)$ et la fonction score $\psi_Y(y(\tau))$. $\delta(\cdot)$ est le filtre identité, *i.e.* $x \star \delta(t) = x(t)$. Nous noterons, que si $\gamma_{y, \psi_Y}(t) = \beta \delta(t)$ (avec β une constante), alors nous sommes à un point stationnaire pour $g(t)$.

Dans le cas gaussien, $\psi_Y(y(\tau))$ est une fonction linéaire donc $\gamma_{y, \psi_Y}(t)$ est proportionnelle à $C_{2Y}(t)$ la fonction d'autocorrélation de Y . Ainsi, l'objectif de cet algorithme revient à trouver une sortie blanche à l'ordre 2. Mais dans le cas général, $\psi_Y(y(\tau))$ admet un développement plus complexe. Par conséquent $\gamma_{y, \psi_Y}(t)$ sera la corrélation entre la sortie $y(\tau)$ et une fonction non linéaire $\psi_Y(y(\tau))$ de $y(\tau)$. L'algorithme fait donc effectivement intervenir l'ensemble des statistiques d'ordre supérieur. La sélection de ces statistiques d'ordre supérieur est faite en fonction de la distribution de la sortie $y(\tau)$ à chaque itération, par

l'intermédiaire de l'estimation de la fonction score. En définitive, l'algorithme est basé sur l'ensemble des statistiques d'ordre supérieur et ses performances devraient *a priori* assez peu dépendre de la distribution du signal d'entrée. Nous noterons que la relation (2.22) ne nécessite pas l'estimation de $\bar{g}(t)$ contrairement à (2.17).

L'algorithme MAMV s'organise de la façon suivante :

1. Initialisation de $g(t)$ et $y(t)$,
2. Estimation de la fonction score ψ_Y ,
3. Estimation de $\gamma_{y,\psi_Y}(t) = E[\psi_Y(y(\tau))y(\tau - t)]$,
4. Mise à jour du filtre déconvolveur par (2.22),
5. Calcul du signal déconvolué $y(t)$,
6. Normalisation de $y(t)$ et $g(t)$.

Nous itérons les étapes 2 à 6 jusqu'à convergence (μ est le pas de gradient qui est un réel positif, sa valeur est ajustée pour assurer la convergence). La normalisation permet de lever l'indétermination d'échelle sur le système de déconvolution $y(t)$ et $g(t)$. En général, nous nous fixons comme contrainte que le signal de sortie est de puissance unitaire. L'estimation de la fonction score est l'étape la plus coûteuse en terme de temps de calcul et celle-ci conditionnera les performances de l'algorithme. Ce problème sera détaillé dans la section 2.4. Après la présentation de l'algorithme MAMV, nous allons analyser des résultats sur des données synthétiques avec des ondelettes simples (à phase minimale et nulle).

2.2.3 Données simulées

Nous souhaitons montrer à l'aide de cette simulation que l'algorithme MAMV est capable d'estimer des filtres à phase quelconque. En effet, il utilise implicitement les statistiques d'ordre supérieur à 2 si les signaux ne sont pas gaussiens par le biais de l'entropie ou de la fonction score suivant si nous analysons la fonction coût de l'algorithme ou son gradient. Nous choisissons dans les figures 2.1 et 2.2 :

- une réflectivité Bernoulli-laplacienne de 400 échantillons avec 20 réflecteurs,
- une ondelette AR de 6 coefficients $[1, 0.4, 0.5, 0.45, 0.4, 0.1]$,
- Pour la figure 2.1 une ondelette à phase minimale ,
- Pour la figure 2.2 une ondelette à phase nulle,
- un bruit additif gaussien blanc avec un RSB de 18dB,

puis, nous déconvoluons les observations avec l'algorithme MAMV avec un filtre à 30 coefficients.

Dans les figures 2.1 et 2.2, nous traçons en (a) la réflectivité, en (b) les observations après convolution de la réflectivité par l'ondelette à phase nulle ou minimale et l'ajout du bruit gaussien. Pour analyser la qualité de la déconvolution, nous avons en (c) la réflectivité estimée et nous comparons en (d) (resp. (e)) la phase (resp. module) du filtre estimé par l'algorithme MAMV et la valeur théorique.

Avec les figures 2.1 et 2.2 (a) et (c), nous remarquons que l'estimation de la réflectivité est très satisfaisante quelle que soit la phase de l'ondelette. Cette remarque est corroborée

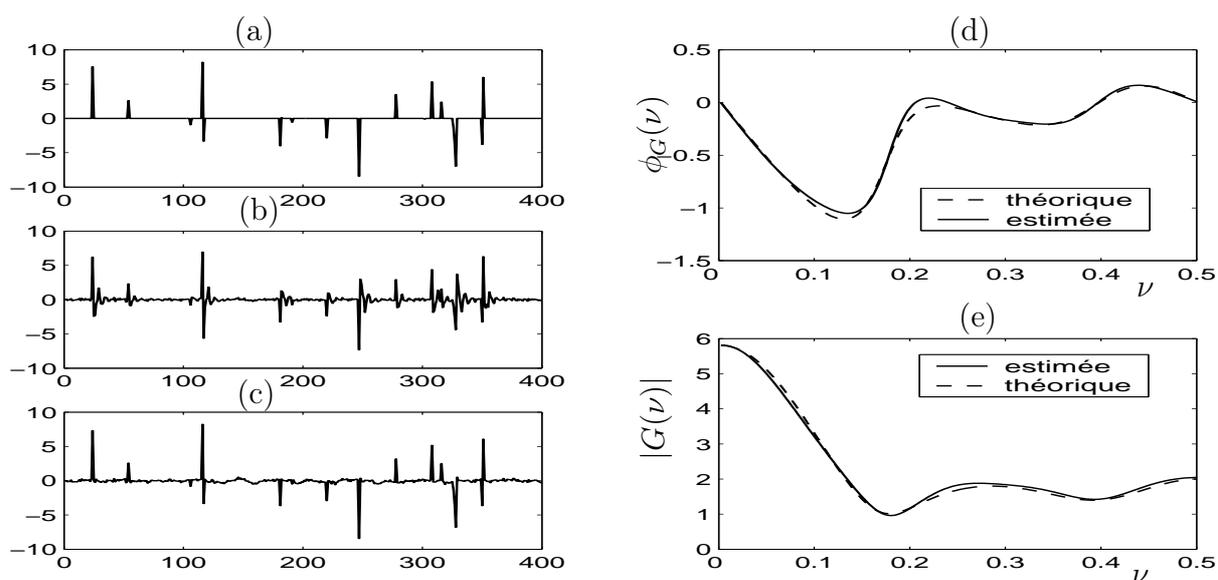


FIG. 2.1 – Déconvolution de données sismiques synthétiques avec une ondelette **AR à phase minimale**. (a) réflectivité Bernoulli-laplacienne (b) observation avec ondelette AR à phase minimale, et RSB= 18dB, (c) réflectivité estimée par MAMV, (d) phase théorique (rad) en pointillé et phase estimée du filtre déconvolveur (trait plein) (e) idem que (d) pour module du filtre.

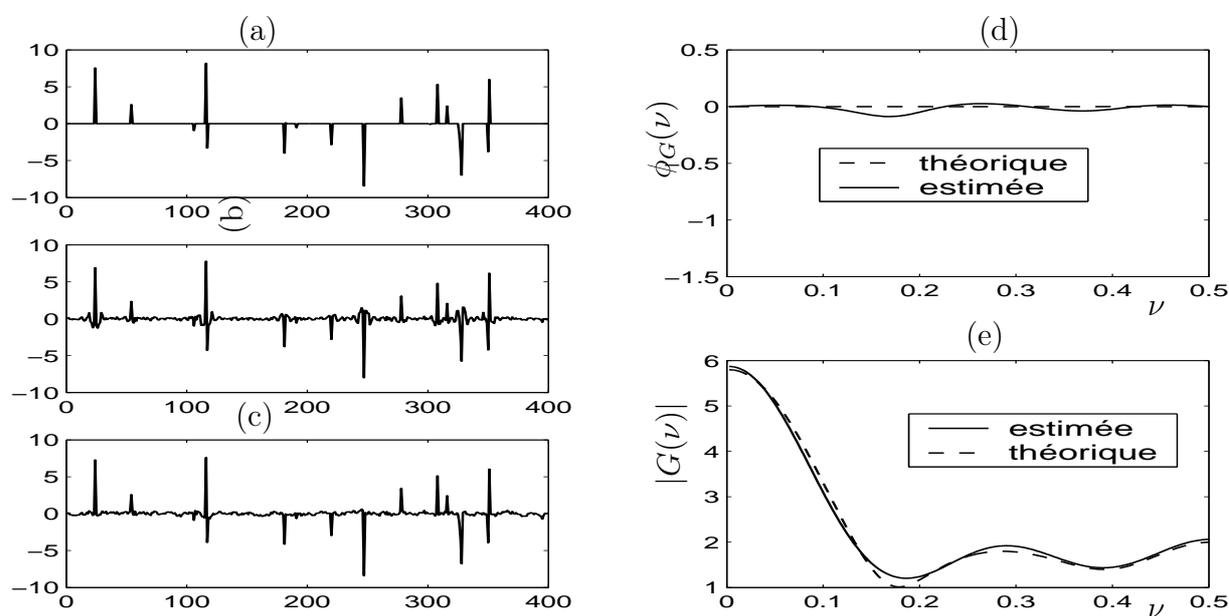


FIG. 2.2 – Déconvolution de données sismiques synthétiques avec une ondelette **AR à phase nulle**. (a) réflectivité Bernoulli-laplacienne (b) observation avec ondelette AR à nulle, et RSB= 18dB, (c) réflectivité estimée par MAMV, (d) phase théorique (rad) en pointillé et phase estimée du filtre déconvolveur (trait plein) (e) idem que (d) pour module du filtre.

par une très bonne estimation du module en (e). Enfin, la phase estimée en (d) est proche de la phase théorique. Pour le cas à phase minimale, l'estimation est quasi parfaite. En ce qui concerne le filtre à phase nulle, l'erreur est plus grande (figure 2.2(d)). La plus grande difficulté à estimer le filtre à phase nulle vient du fait que seules les statistiques d'ordre supérieur à 2 tendent à estimer le filtre à phase nulle, celles à l'ordre 2 sont insensibles à la phase de l'ondelette. Ainsi, dans le cas de la figure 2.1, pour l'estimation du filtre à phase minimale, nous pouvons parvenir à des résultats identiques avec l'algorithme de Yule-Walker basé sur les statistiques d'ordre 2. En revanche, si le filtre est à phase non minimale, l'intervention des statistiques d'ordre supérieur est cruciale et comme la fonction score comporte peu d'échantillons dans sa partie non linéaire, l'estimation est donc rendue plus difficile.

L'estimation d'un filtre déconvolveur MA, *i.e.* une ondelette de type AR, est possible par minimisation du taux d'information mutuelle de la sortie. Une extension naturelle est d'envisager d'autres modèles paramétriques temporels du filtre déconvolveur. Suivant les données, il est important de pouvoir changer de modèle paramétrique afin de choisir le mieux adapté. Ce choix est généralement guidé par le nombre et la facilité d'identification de ces paramètres. Nous essayons de réduire au maximum le nombre de paramètres pour faciliter l'identification et la stabilité des algorithmes. Ainsi, nous proposons dans la section suivante l'estimation de filtres inverses répondant à un modèle AR. Cette approche qui se nommera ARMV, permettra l'inversion d'une ondelette MA.

2.3 Algorithme ARMV de déconvolution d'ondelette MA

Dans cette section, nous proposons de minimiser le taux d'information mutuelle de la sortie $y(t)$ par rapport à un modèle paramétrique autorégressif (AR) du filtre déconvolveur. Dans un premier temps, nous pourrions nous baser sur la relation (1.75) avec un vecteur de paramètres $\mathbf{A} = [a_0, \dots, a_P]$ définissant le filtre autorégressif $g(t)$. En réalité, il ne faut pas se limiter à un filtre causal (le vecteur de paramètres doit pouvoir comporter des coefficients pour des retards négatifs). Afin de simplifier les écritures du gradient, nous pouvons nous intéresser au filtre MA $b(t)$ de coefficients \mathbf{B} , avec les mêmes coefficients que le filtre $g(t)$, soit $\mathbf{B} = \mathbf{A}$. Ainsi, nous pouvons écrire soit $y(t) = g \star d(t)$ ou alors $b \star y(t) = d(t)$. Les deux expressions sont identiques, ainsi les problèmes de la détermination du filtre $b(t)$ ou du filtre $g(t)$ sont équivalents. La détermination du filtre $b(t)$ donnera des équations plus simples. Afin d'estimer ce filtre, nous souhaitons minimiser le taux d'information mutuelle de sortie soit :

$$g^{ARMV}(t) = \arg \min_{g(t) \text{ AR}} \tilde{I}(Y). \quad (2.24)$$

Pour obtenir un algorithme utilisable, nous devons modifier l'expression (2.8) de $\tilde{I}(Y)$ pour l'écrire en fonction du filtre MA $b(t)$. Dans (2.8), nous devons modifier le second terme faisant apparaître la réponse impulsionnelle $g(t)$. Comme les filtres $g(t)$ et $b(t)$ sont les inverses (soit $g \star b(t) = \delta(t)$), nous avons en utilisant l'expression des réponses en fréquence des filtres $g(t)$ et $b(t)$:

$$\frac{1}{2\pi} \int_0^{2\pi} \log \left| \sum_{t=-\infty}^{+\infty} g(t)e^{-jt\theta} \right| d\theta = -\frac{1}{2\pi} \int_0^{2\pi} \log \left| \sum_{t=-\infty}^{+\infty} b(t)e^{-jt\theta} \right| d\theta. \quad (2.25)$$

Ainsi, le critère de déconvolution peut s'écrire comme :

$$\tilde{I}(Y) = H(y(\tau)) + \frac{1}{2\pi} \int_0^{2\pi} \log \left| \sum_{t=-\infty}^{+\infty} b(t)e^{-jt\theta} \right| d\theta. \quad (2.26)$$

Nous allons maintenant minimiser ce critère (2.26) par rapport aux coefficients $b(t)$. Ainsi dans la sous-section suivante, nous allons calculer le gradient par rapport aux coefficients $b(t)$. Nous avons un nombre fini de coefficients pour $b(t)$ contrairement au filtre $g(t)$.

2.3.1 Gradient du critère

Nous allons calculer le gradient du critère défini par (2.26) par rapport à la réponse impulsionnelle du filtre $b(t)$. Le calcul du gradient du terme intégral est identique à celui de l'algorithme MAMV. Il est détaillé en annexe A.2 et le résultat est donné en (2.15). De la même façon, nous obtenons :

$$\frac{\partial}{\partial b(t)} \left\{ \frac{1}{2\pi} \int_0^{2\pi} \log \left| \sum_{t=-\infty}^{+\infty} b(t)e^{-jt\theta} \right| d\theta \right\} = \bar{b}(-t), \quad (2.27)$$

où $\bar{b}(t)$ est le filtre déconvolveur de $b(t)$, soit en fait $g(t)$.

Pour la dérivation du terme entropique, comme pour l'algorithme MAMV dans la relation (2.10), nous utilisons la définition de l'entropie, soit :

$$\frac{\partial H(y(\tau))}{\partial b(t)} = -\frac{\partial}{\partial b(t)} E[\log p_Y(y(\tau))] = E \left[-\frac{\partial}{\partial b(t)} \log p_Y(y(\tau)) \right]. \quad (2.28)$$

Ensuite, comme pour (2.11) de l'algorithme MAMV, nous faisons apparaître la fonction score de Y définie par (2.12). Nous obtenons :

$$\frac{\partial H(y(\tau))}{\partial b(t)} = E \left[\psi_Y(y(\tau)) \frac{\partial y(\tau)}{\partial b(t)} \right]. \quad (2.29)$$

Or, nous avons posé $b \star y(\tau) = d(\tau)$, soit :

$$d(\tau) = (b \star y)(\tau) = \sum_{i=-\infty}^{i=+\infty} b(i)y(\tau - i). \quad (2.30)$$

En dérivant l'expression précédente par rapport à $b(t)$ et puisque $d(\tau)$ est indépendant de $b(t)$, on a :

$$\begin{aligned} 0 &= \frac{\partial d(\tau)}{\partial b(t)} = \frac{\partial (b \star y)(\tau)}{\partial b(t)} = \frac{\partial \left\{ \sum_{i=-\infty}^{+\infty} b(i)y(\tau - i) \right\}}{\partial b(t)}, \\ &= y(\tau - t) - \sum_{i=-\infty}^{+\infty} b(i) \frac{\partial y(\tau - i)}{\partial b(t)}. \end{aligned} \quad (2.31)$$

Nous définissons le processus :

$$\theta_t(\lambda) = \frac{\partial y(\lambda)}{\partial b(t)}. \quad (2.32)$$

Avec la définition (2.32), nous pouvons écrire la relation (2.31) sous la forme suivante :

$$\sum_{i=-\infty}^{+\infty} b(i)\theta_t(\tau - i) = -y(\tau - t). \quad (2.33)$$

Ainsi, $\theta_t(\tau)$ est un processus autorégressif de processus d'innovation $-y(\tau - t)$ et de filtre $b(\tau)$. Avec (2.29) et (2.27), nous avons :

$$\frac{\partial I(Y)}{\partial b(t)} = E[\psi_Y(y(\tau))\theta_t(\tau)] + \bar{b}(-t). \quad (2.34)$$

Après le calcul du gradient du critère du taux d'information mutuelle par rapport aux paramètres du filtre déconvolveur, nous allons maintenant détailler l'algorithme ARMV.

2.3.2 Algorithme ARMV

Contrairement au modèle de filtre déconvolveur MA, nous ne pouvons pas mettre en place de façon simple un algorithme de gradient relatif. En effet, si nous effectuons la convolution de $\frac{\partial I(Y)}{\partial b(t)}$ de la relation (2.34) avec le filtre $b(t)$ et $b(-t)$, nous constatons que le dernier terme contenant $\bar{b}(-t)$ se simplifie mais aucune simplification n'apparaît sur le premier terme. Ainsi, pour cet algorithme ARMV, nous nous limiterons à une méthode de gradient classique. Nous devons estimer le filtre déconvolveur $\bar{b}(-t)$. La mise à jour se fera par :

$$b(t) \leftarrow b(t) - \mu (E[\psi_Y(y(\tau))\theta_t(\tau)] + \bar{b}(-t)). \quad (2.35)$$

L'algorithme ARMV de minimisation du critère (2.26) s'organise de la façon suivante :

1. Initialisation du filtre $b(t)$ et de la sortie $y(t)$,
2. Estimation de la fonction score ψ_Y ,
3. Estimation du processus $\theta_t(\cdot)$ avec la relation (2.33),
4. Mise à jour du filtre déconvolveur avec la relation (2.35),
5. Calcul du signal déconvolué $y(t)$,
6. Normalisation de $y(t)$ et $b(t)$.

Nous itérons les étapes 2 à 6 jusqu'à la convergence (μ est le pas du gradient (réel positif)). La normalisation permet de lever l'indétermination d'amplitude sur le problème de déconvolution. Comme pour la méthode d'estimation d'un filtre déconvolveur MA, nous pouvons faire le parallèle avec la méthode du maximum de vraisemblance. Il suffit de remplacer la densité de probabilité du signal déconvolué, supposée connue dans une méthode du maximum de vraisemblance, par l'estimée à chaque itération de la densité de probabilité du signal déconvolué. Après la mise en place de l'algorithme ARMV pour la déconvolution d'ondelette de type MA, nous proposons une simulation pour montrer l'efficacité de cet algorithme par rapport à l'algorithme de Durbin.

2.3.3 Données simulées

Nous présentons à la figure 2.3 des résultats de déconvolution de données synthétiques par l'algorithme ARMV. Le modèle de l'ondelette étant un modèle MA, soit un filtre déconvolveur autorégressif, nous comparons les résultats avec l'algorithme de Durbin présenté dans le paragraphe 1.4.2.2 se basant sur les statistiques d'ordre 2 dans le domaine temporel. Nous utilisons une réflectivité Bernoulli-laplacienne de 20 réflecteurs pour 400 échantillons. L'ondelette comporte 6 coefficients $[1, 0.4, 0.5, 0.45, 0.4, 0.1]$. L'inversion se fait par les algorithmes de Durbin et ARMV avec des filtres ayant une longueur de 10 coefficients.

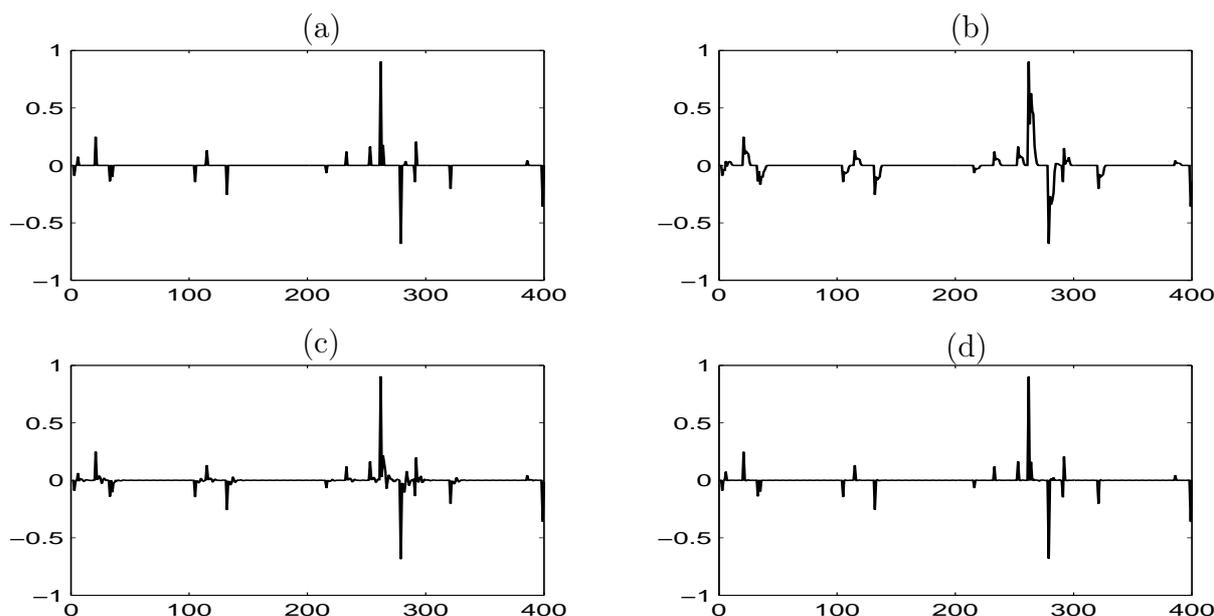


FIG. 2.3 – Exemple de déconvolution par l'algorithme ARMV : (a) réflectivité Bernoulli-laplacienne simulée de 20 réflecteurs pour 400 échantillons, (b) données obtenues par convolution de la réflectivité par une ondelette MA de coefficients $[1, 0.4, 0.5, 0.45, 0.4, 0.1]$, (c) déconvolution par la méthode de Durbin avec un filtre à 10 coefficients, (d) idem avec l'algorithme ARMV.

Nous remarquons que l'ondelette avec une longueur de 6 coefficients augmente la largeur de chacun des réflecteurs. Après déconvolution, nous avons une bonne estimation de la réflectivité (figure 2.3(a)) par l'algorithme de Durbin (figure 2.3(c)) et par l'algorithme ARMV (figure 2.3(d)). L'estimation par l'algorithme ARMV est meilleure. En effet, pour l'algorithme de Durbin, nous notons quelques résidus après les réflecteurs. Nous pouvons voir principalement cet effet dans la partie entre les échantillons 250 et 300 où nous avons un ensemble de plusieurs réflecteurs très proches. L'imperfection de l'estimation de la méthode de Durbin provient de l'étape d'inversion de la densité spectrale de puissance qui doit être faite avec précaution (voir section 1.4.2.2). En conclusion, l'algorithme ARMV est intéressant car il permet d'étendre l'algorithme MAMV à l'estimation de filtre déconvolveur AR.

Ainsi, nous pourrions tenter de le caractériser en comparaison des algorithmes à l'ordre 2, notamment à propos de sa robustesse au bruit. En fait, l'algorithme souffre d'un problème de stabilité indéniable. En effet, l'aspect itératif de l'algorithme n'assure pas l'estimation

d'un filtre déconvolveur AR stable à chaque itération. Par conséquent, sans amendement à l'algorithme, nous nous heurtons à un grave problème de stabilité. Ainsi, il n'apparaît pas nécessaire de caractériser plus finement cet algorithme. L'extension à un modèle ARMA est possible. Dans ce cas, le filtre déconvolveur $g(t)$ est décomposé en deux parties : $b(t)$ représentant la partie AR du filtre et $h(t)$ représentant la partie MA. En partant de $y(t) = g \star d(t)$, nous écrivons que $b \star y(t) = h \star d(t)$. Ensuite, nous écrivons le taux d'information mutuelle de la sortie en fonction de ces deux filtres. Enfin nous calculons le gradient par rapport aux filtres $b(t)$ et $h(t)$ qui sont à réponse impulsionnelle finie. Nous ne pousserons pas les développements plus loin, car nous avons les mêmes problèmes de stabilité pour la partie AR exprimée par le filtre $b(t)$ que pour la méthode ARMV.

Avant de poursuivre, la caractérisation des algorithmes ARMV et MAMV, nous devons revenir sur le point crucial des deux algorithmes : l'estimation de la fonction score. Elle permet la sélection des statistiques d'ordre supérieur pour la déconvolution. Dans la section suivante, nous proposons deux estimateurs de fonctions scores : (i) une méthode à noyaux, (ii) un estimateur direct avec un modèle polynomial. Ensuite, nous comparons les performances de ces estimateurs dans le cas de signaux avec des distributions typiques de réflectivité sismique mélangée à un bruit blanc gaussien. L'objectif de cette comparaison sera de sélectionner, suivant des critères de robustesse et de précision, le meilleur estimateur de fonction score ayant tout de même un temps de calcul acceptable.

2.4 Estimation de la fonction score

La fonction score est le point clé qui apparaît dans les problèmes de séparation de sources [PGJ92, Car98] et en déconvolution aveugle [TSiCJ01], comme dans les algorithmes MAMV et ARMV. En pratique, certains auteurs approximent la fonction score par un modèle non linéaire tel une tangente hyperbolique ou cubique [PGJ92, ZC04]. Ces approximations sont liées à un *a priori* sur la distribution du processus. En effet, suivant un modèle non linéaire de la fonction score, nous pouvons remonter à la ddp du processus par simple intégration. Par exemple, une approximation de la fonction score avec un polynôme de degré trois, est équivalente à prendre une ddp *a priori* gaussienne généralisée avec un coefficient de gaussianité de 4 pour le processus (*i.e.* sous-gaussien). Si on applique cette approximation cubique à un signal sur-gaussien, cela peut induire de mauvaises performances pour l'algorithme de déconvolution. Nous pouvons aussi utiliser un modèle paramétrique de la distribution du processus comme une distribution lambda ou béta [EKK00], ou une distribution gaussienne généralisée [CCA00, ER99]. Ces techniques estiment dans un premier temps les paramètres caractéristiques de la distribution avec la variance et le kurtosis. Dans un second temps, on applique la formule théorique de la fonction score en y incorporant les paramètres estimés de la loi. Enfin, une autre paramétrisation s'inspire de la modélisation paramétrique AR d'une densité spectrale de puissance [BV00]. Pour les algorithmes de déconvolution aveugle, nous avons besoin d'un estimateur de fonction score avec un temps de calcul assez restreint et, si possible, qui s'adapte à des distributions de signaux très variés. Dans la suite, nous présenterons deux estimateurs de la littérature et nous ferons une étude de leurs performances dans le cas de distributions typiques des signaux sismiques.

2.4.1 Estimateur de fonction score

2.4.1.1 Estimateur à noyaux

Avec la définition (2.12) de la fonction score d'une variable aléatoire, nous pouvons estimer la fonction score par $\hat{\psi}_X = -\frac{\hat{p}'_X}{\hat{p}_X}$ où \hat{p}_X est la ddp estimée du processus X . Ainsi, nous sommes ramenés au problème plus classique d'estimation d'une ddp.

À l'aide de T échantillons $x(1), \dots, x(T)$ d'une réalisation du processus X , nous pouvons estimer la ddp p_X par la méthode des noyaux [Sil82] :

$$\hat{p}_X(x) = \frac{1}{Th} \sum_{n=1}^T K\left(\frac{x - x(n)}{h}\right), \quad (2.36)$$

où K est le noyau qui doit être une ddp, c'est à dire $\int_{\mathbb{R}} K(x)dx = 1$, et h est le paramètre de largeur de bande ajusté en fonction du nombre d'échantillons. Dans notre cas, comme nous souhaitons dériver cet estimateur, le noyau doit être de classe \mathcal{C}^1 . Classiquement, nous choisissons le noyau gaussien $\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$.

Notre algorithme nécessitant l'estimation de la ddp \hat{p}_X pour toutes les valeurs $x(1), \dots, x(T)$, le calcul de l'estimée en T points demande T^2 estimations du noyau. Pham [Pha03, Pha04] a proposé un algorithme rapide pour l'estimation des ddp, des fonctions scores, de l'information mutuelle et de l'entropie pour des applications de séparation de sources. L'algorithme est basé sur un estimateur à noyaux. Le gain en temps de calcul est obtenu en utilisant un noyau à support borné. Le noyau est défini par la fonction spline cardinal d'ordre 3 qui est la fonction spline la plus simple de classe \mathcal{C}^1 :

$$K(u) = \begin{cases} 3/4 - u^2 & |u| \leq 1/2, \\ (3/2 - |u|)^2/2 & 1/2 \leq |u| \leq 3/2, \\ 0 & \text{sinon.} \end{cases} \quad (2.37)$$

La fonction spline d'ordre 3 est la triple convolution de la fonction indicatrice qui est nulle partout sauf sur l'intervalle $[0; 1]$ où elle vaut 1. Nous tirons partie de la largeur finie du noyau en estimant la ddp sur une grille régulière de T points espacés de h . Une valeur empirique de h est donnée dans [Pha04]. En d'autres termes, si nous supposons que X est un processus centré, avec le changement de variable $X' = X/h$, (2.36) nous obtenons :

$$\hat{p}_X(ih) = \frac{1}{Th} \sum_{n=1}^T K(i - x'(n)). \quad (2.38)$$

La somme se réduit en fait à trois termes, car $K(i - x'(n))$ est non nul si et seulement si $i = \langle x'(n) \rangle$ ou $i = \langle x'(n) \rangle \pm 1$, où $\langle x \rangle$ est la valeur entière de x . Ainsi, nous pouvons calculer rapidement $\hat{p}_X(ih)$ avec la démarche suivante : nous initialisons $\hat{p}_X(ih)$ à 0, puis pour $n = 1, \dots, T$, nous mettons à jour de la façon suivante :

$$\hat{p}_X(ih + \langle x'(n) \rangle h) = \hat{p}_X(ih + \langle x'(n) \rangle h) + \frac{1}{Th} \sum_{n=1}^T K(i + \langle x'(n) \rangle - x'(n)), \quad i = -1, 0, 1. \quad (2.39)$$

L'estimation de la ddp nécessite $3T$ estimations du noyau au lieu de T^2 pour un estimateur à noyau classique. Ensuite, nous choisissons d'estimer $\log p_X$ par $\widehat{\log p_X} = \log \hat{p}_X$ sur la grille jh . Puis, l'estimateur de la fonction score $\hat{\psi}_X = -\frac{d}{dx} \widehat{\log p_X}$ au point de grille $(j - 1/2)h$ est obtenu par différenciation de $-\widehat{\log p_X}$ calculée sur la grille jh , soit $-\frac{1}{h} [\log \hat{p}_X(jh) - \log \hat{p}_X((j-1)h)]$. Ensuite, nous interpolons cette estimation pour obtenir l'estimateur de fonction score, soit :

$$\hat{\psi}_X((j+v)h) = \frac{1}{h} \left\{ \left(\frac{1}{2} - v \right) \log \hat{p}_X((j-1)h) + 2v \log \hat{p}_X(jh) - \left(\frac{1}{2} + v \right) \log \hat{p}_X((j+1)h) \right\}, \quad (2.40)$$

avec $v \in [-\frac{1}{2}, \frac{1}{2}]$. L'estimateur est principalement appliqué en séparation de sources dans cette version simple [Tal99, TJ99, Pha02]. Il existe une version multidimensionnelle pour l'estimation de densité de probabilité ou de fonction score conditionnelles qui est notamment utile pour des modèles markoviens en séparation de sources [HJP03, Lar03, LJH04].

Les estimateurs à noyaux permettent d'avoir une assez bonne estimation mais le temps de calcul est relativement long. Pham a proposé une technique pour réduire le temps de calcul, cependant il reste encore conséquent. Afin de simplifier l'estimation, de nombreuses techniques tentent d'approximer directement la fonction score par des modèles paramétriques. Nous allons présenter un estimateur direct de la fonction score avec un modèle polynomial.

2.4.1.2 Estimation directe

L'indépendance peut être approximée par les statistiques d'ordre supérieur avec les cumulants [Tug93, Cad96] ou le kurtosis [Bou96, LH98]. En approximant, la ddp $p_X(x)$ d'un processus X de puissance unitaire avec un développement d'Edgeworth [Com94], ou de Gram-Charlier [TJ99, BZ02], nous obtenons une approximation polynomiale de la fonction score faisant apparaître les cumulants ou moments d'ordre supérieur. Cependant, nous n'avons aucune assurance que l'estimation est correcte car nous tronquons les développements sans justifier que les termes tronqués sont négligeables par rapport à ceux conservés. Dans [TJ99], il est montré que pour une distribution uniforme l'approximation est très mauvaise.

Dans ce paragraphe, nous détaillons une estimation directe de ψ_X suivant un modèle paramétrique $\hat{\psi}_X(x, \mathbf{p})$ où \mathbf{p} est le vecteur de paramètres, qui sera estimé par minimisation de l'erreur quadratique [PGJ92, TJ99] :

$$\mathcal{E}(\mathbf{p}) = \frac{1}{2} E[(\psi_X(x) - \hat{\psi}_X(x, \mathbf{p}))^2]. \quad (2.41)$$

Le gradient de $\mathcal{E}(\mathbf{p})$ par rapport au vecteur paramètre \mathbf{p} est :

$$\frac{\partial \mathcal{E}(\mathbf{p})}{\partial \mathbf{p}} = E \left[\left\{ \psi_X(x) - \hat{\psi}_X(x, \mathbf{p}) \right\} \frac{\partial \hat{\psi}_X}{\partial \mathbf{p}}(x, \mathbf{p}) \right]. \quad (2.42)$$

Pour rendre l'estimation du vecteur de paramètres possible, nous pouvons utiliser le lemme suivant (démonstration en annexe A.3) :

Lemme 2.2. *Soit X une variable aléatoire et $\psi_X(x)$ sa fonction score. Si f est une fonction continue et dérivable sur \mathbb{R} (sauf en un nombre fini de points x_i), et satisfaisant la relation :*

$$\lim_{|x| \rightarrow +\infty} p_X(x)f(x) = 0, \quad (2.43)$$

alors nous avons :

$$E[f(x)\psi_X(x)] = E[f'(x)]. \quad (2.44)$$

A l'aide du lemme avec $f = \frac{\partial \hat{\psi}_X}{\partial \mathbf{p}}$, le vecteur de paramètres \mathbf{p} vérifiant $\frac{\partial \mathcal{E}(\mathbf{p})}{\partial \mathbf{p}} = 0$, satisfait :

$$E \left[\hat{\psi}_X(x, \mathbf{p}) \frac{\partial \hat{\psi}_X}{\partial \mathbf{p}}(x, \mathbf{p}) \right] = E \left[\frac{\partial^2 \hat{\psi}_X}{\partial \mathbf{p} \partial x}(x, \mathbf{p}) \right]. \quad (2.45)$$

Les paramètres estimés ne dépendent pas directement de la fonction score ψ_X . En fait la fonction score est présente dans l'opérateur de moyenne statistique $E[\cdot]$ par l'intermédiaire de la ddp de X . Si nous choisissons un modèle paramétrique $\hat{\psi}_X(x, \mathbf{p})$ linéaire par rapport au vecteur paramètre \mathbf{p} , nous pouvons trouver une estimation explicite de ce vecteur \mathbf{p} . Le modèle le plus simple répondant à cette condition est un modèle polynomial :

$$\hat{\psi}_X(x, \mathbf{p}) = \sum_{i=0}^N p_i x^i = \mathbf{p}^T \mathbf{K}(x), \quad (2.46)$$

en notant $\mathbf{K}(x) = [1, x, x^2, \dots, x^N]^T$ et $\mathbf{p} = [p_0, p_1, \dots, p_N]^T$. Avec le modèle polynomial (2.46), l'équation d'estimation des paramètres (2.45) s'exprime par le système suivant :

$$\begin{bmatrix} 1 & \mu_1 & \mu_2 & \dots & \mu_N \\ \mu_1 & \mu_2 & \mu_3 & \dots & \mu_{N+1} \\ \mu_2 & \mu_3 & \mu_4 & \dots & \mu_{N+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mu_N & \mu_{N+1} & \mu_{N+2} & \dots & \mu_{2N} \end{bmatrix} \begin{bmatrix} p_0 \\ p_1 \\ p_2 \\ \vdots \\ p_N \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 2\mu_1 \\ \vdots \\ N\mu_{N-1} \end{bmatrix}, \quad (2.47)$$

où $\mu_i = E[x^i]$ est le moment d'ordre i de X .

Nous pouvons remarquer dans (2.47) que la meilleure approximation (au sens des moindres carrés) de la fonction score avec un polynôme de degré N , nécessite l'estimation des moments d'ordre inférieur à $2N$. En revanche, un développement de Gram-Charlier et d'Edgeworth à l'ordre N ne demande que les N premiers moments, soit utilise moins d'information sur le signal. Cette remarque explique en partie l'imprécision des développements de Gram-Charlier et Edgeworth notée dans [TJ99]. Pour nos données sismiques, nous pouvons supposer que les distributions de probabilité sont paires, donc les fonctions scores sont impaires. Ainsi, pour limiter le nombre de paramètres nous pouvons choisir un modèle polynomial ne comportant que des puissances impaires tel que :

$$\hat{\psi}_X(x, \mathbf{p}) = p_1x + p_3x^3 + \dots + p_Nx^N, \quad (2.48)$$

où $N = 2Q + 1$ est un entier impair. Ce modèle est équivalent à supprimer les colonnes impaires de la matrice de (2.47). Avec la parité des ddp, tous les moments impairs sont nuls donc $\mu_{2j+1} = E[x^{2j+1}] = 0$ ainsi nous pouvons aussi enlever toutes les lignes impaires car elles représentent des équations triviales. A l'aide de ces deux remarques, on obtient un système de taille deux fois plus petite défini uniquement avec les moments d'ordre pair et qui présente donc un meilleur conditionnement :

$$\begin{bmatrix} \mu_2 & \mu_4 & \mu_6 & \dots & \mu_{N+1} \\ \mu_4 & \mu_6 & \mu_8 & \dots & \mu_{N+3} \\ \mu_6 & \mu_8 & \mu_{10} & \dots & \mu_{N+5} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mu_{N+1} & \mu_{N+3} & \mu_{N+5} & \dots & \mu_{2N} \end{bmatrix} \begin{bmatrix} p_1 \\ p_3 \\ p_5 \\ \vdots \\ p_N \end{bmatrix} = \begin{bmatrix} 1 \\ 3\mu_2 \\ 5\mu_4 \\ \vdots \\ N\mu_{N-1} \end{bmatrix}. \quad (2.49)$$

Après la présentation des estimateurs à noyaux et d'un estimateur direct avec un modèle polynomial, il est opportun de comparer leurs performances pour différentes distributions et en fonction du rapport signal à bruit pour effectuer un choix raisonné.

2.4.2 Étude des performances des estimateurs

Les performances des estimateurs vont dépendre très certainement de la distribution du signal. Intuitivement, la méthode à noyaux est plus générale et aura une capacité d'adaptation plus grande en fonction de la distribution du signal comparativement à l'estimateur direct avec un polynôme. Un autre problème est le choix de l'ordre du polynôme. L'augmentation de l'ordre doit normalement donner une amélioration de l'estimation. Nous proposons dans cette partie de comparer plusieurs estimateurs : (i) l'estimateur à noyaux, l'estimation directe avec un polynôme (ii) d'ordre 7 et (iii) d'ordre 11. Nous ferons une étude sur des signaux de réflectivité ayant une distribution de type :

- Bernoulli-laplacienne,
- Bernoulli-gaussienne,
- laplacienne,
- uniforme.

Nous proposons de faire cette étude dans le cadre bruité ainsi, le processus $X = R + N$ où R est le signal de réflectivité distribué suivant une des quatre distributions précédemment décrites et N un bruit additif gaussien blanc centré indépendant de la réflectivité. Le choix de ce signal X est fait suivant le modèle de déconvolution dans un contexte bruité. En effet, la sortie estimée résultera de la superposition de la réflectivité recherchée et d'un bruit additif. Les performances des estimateurs seront tracées en fonction du rapport signal sur bruit. Avant la comparaison des estimateurs, nous devons déterminer les fonctions scores théoriques de ces processus dans les quatre cas.

2.4.2.1 Fonctions scores théoriques

Pour le calcul de la fonction score théorique de $X = R + N$, nous devons tout d'abord calculer la densité de probabilité $p_X(x)$ de X . Nous supposons la réflectivité et le bruit indépendants, ainsi, nous savons que $p_X(x) = p_R \star p_N(x)$. Les détails des calculs dans les quatre cas sont donnés en annexe B. La figure 2.4 représente la fonction score pour les quatre signaux décrits précédemment pour trois valeurs de rapport signal à bruit (0, 3 et 10 dB).

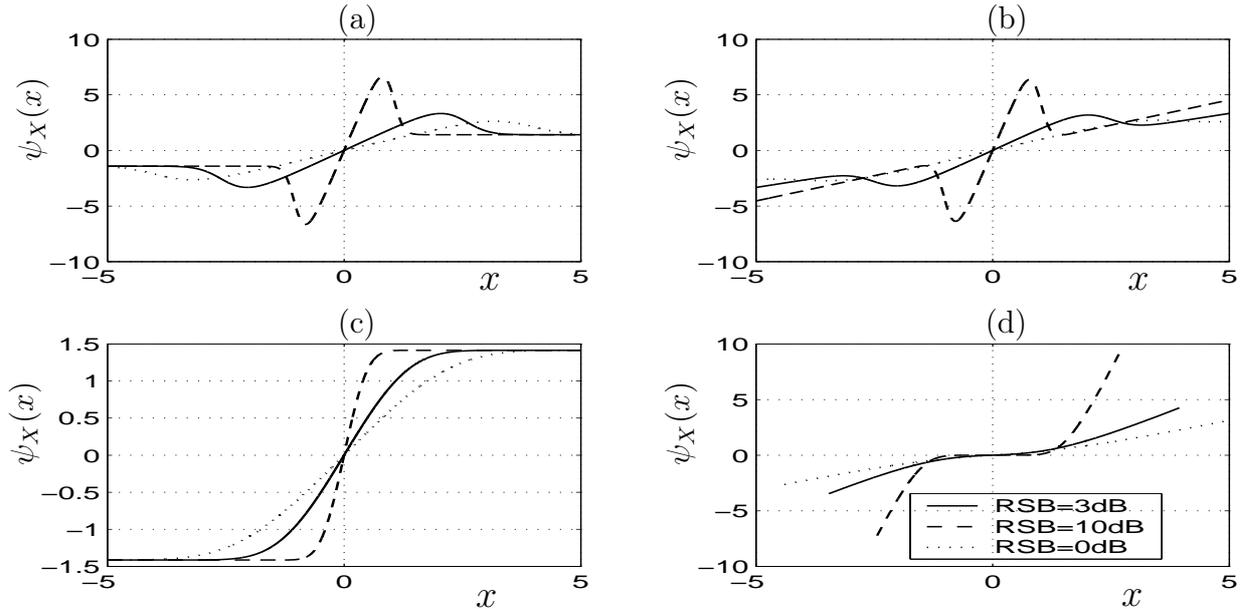


FIG. 2.4 – Fonctions scores théoriques pour une réflectivité distribuée selon une loi : (a) Bernoulli-laplacienne, (b) Bernoulli-gaussienne, (c) laplacienne, (d) Uniforme dans $[-\sqrt{3}; \sqrt{3}]$ avec un bruit additif blanc gaussien : RSB=0dB (...); RSB=3dB (-); RSB=10dB(- -).

Pour les signaux Bernoulli-laplacien et Bernoulli-gaussien, le rapport signal sur bruit est défini directement comme $10 \log_{10}(\sigma^2/\sigma_N^2)$, avec σ la variance de la loi laplacienne ou gaussienne représentant les réflecteurs et σ_N la variance du bruit. La puissance de la réflectivité est en fait $\lambda\sigma^2$. λ est le paramètre de la loi Bernoulli défini en (1.5). Le paramètre des lois de Bernoulli est pris égal à $1/20$. Ainsi, le rapport des puissances est $\lambda\sigma^2/\sigma_N^2$ donc le RSB défini précédemment n'est pas le rapport des puissances. En fait, ce rapport signal à bruit peut être vu comme un rapport signal sur bruit local autour d'un réflecteur. Pour toutes les distributions, la fonction score théorique est linéaire pour les faibles valeurs de $|x|$. Cette partie représente les échantillons liés au bruit. La droite est d'équation x/σ_N^2 : c'est l'équation de la fonction score du bruit gaussien. Ainsi, comme la puissance du signal de réflectivité est prise constante, la pente diminue lorsque le RSB diminue. Pour les grandes valeurs de $|x|$, la fonction score théorique tend vers la fonction score de la réflectivité.

Pour la loi laplacienne (figure 2.4(c)), c'est une constante de valeur σ . Pour le signal Bernoulli-gaussien (figure 2.4(b)), c'est une droite de pente $1/(\sigma^2 + \sigma_N^2)$. Pour la loi uniformément distribuée entre $[-\sqrt{3}, \sqrt{3}]$ (figure 2.4(d)), la fonction score croît très rapidement

à l'image de la fonction score théorique de la loi uniforme qui présente des singularités verticales en $\pm\sqrt{3}$. En définitive, si le RSB est très faible, la fonction score de X tend vers une droite qui représente le bruit, et pour les grandes valeurs du RSB, la fonction score comporte des singularités importantes qui rendront l'estimation difficile.

2.4.2.2 Comparaison des estimateurs

Avant de passer à une comparaison très générale, la figure 2.5 montre un exemple d'estimation dans le cas d'un signal Bernoulli-laplacien avec un RSB de 10dB et un paramètre $\lambda = 1/20$ pour la loi de Bernoulli. Nous comparons (i) l'estimateur à noyaux, (ii) une estimation directe avec un polynôme d'ordre 7 et (iii) avec un polynôme d'ordre 11.

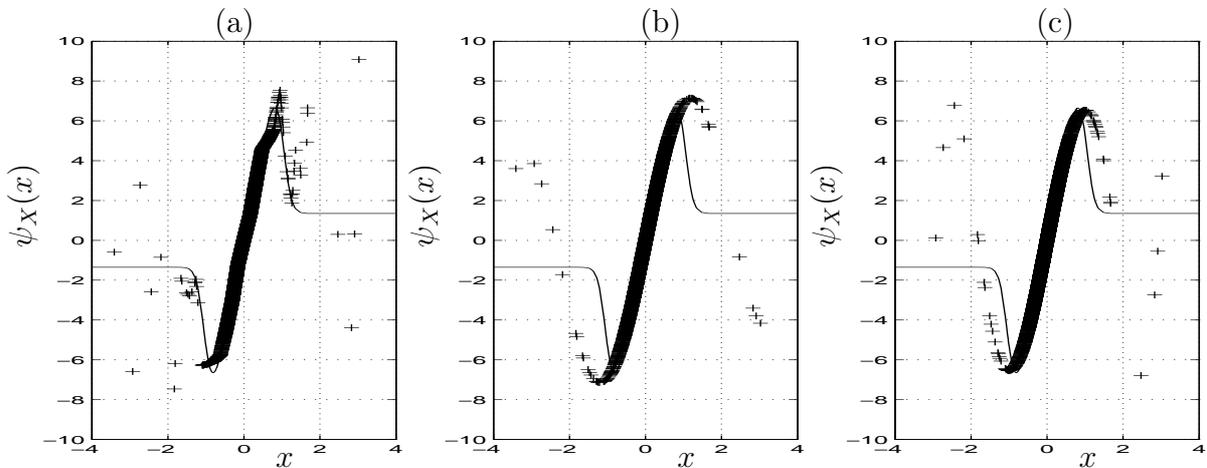


FIG. 2.5 – Exemple d'estimation de la fonction score Bernoulli-laplacienne avec un RSB=10dB et $\lambda = 1/20$. Fonction score théorique en trait continu. (a) estimateur à noyaux, (b) polynôme de degré 7, (c) degré 11.

Sur les allures des estimées, nous remarquons que les trois estimateurs ont des bonnes performances pour les valeurs de x proches de zéro. La partie linéaire correspondant au bruit est bien estimée, en revanche, pour les grandes valeurs, les estimées sont beaucoup plus éloignées des valeurs théoriques. Cette erreur d'estimation est due au faible nombre d'échantillons disponibles dans cet intervalle pour l'estimation.

Nous pouvons comparer les performances des estimateurs grâce à l'erreur quadratique moyenne (EQM) normalisée de la fonction score $\hat{\psi}_X(x)$ définie par :

$$\mathcal{E}(\hat{\psi}_X) = \frac{E[\{\hat{\psi}_X(x) - \psi_X(x)\}^2]}{E[\psi_X(x)^2]}. \quad (2.50)$$

Pour la figure 2.5, le tableau 2.1 résume les performances en terme d'EQM. Nous calculons l'EQM sur l'ensemble des échantillons, mais aussi l'EQM sur l'ensemble défini par les 20% des échantillons de plus forte amplitude en valeur absolue, soit les échantillons correspondant approximativement aux réflecteurs. L'EQM partielle permet d'avoir une indication de la précision sur la partie "signal" du processus bruité.

TAB. 2.1 – Comparaison des EQM en dB des estimations de la figure 2.5.

	Noyaux	ordre 7	ordre 11
EQM totale	-17.8	-14.8	-15.1
EQM partielle	-11.8	-8.01	-7.96

Ces valeurs montrent que, dans l'ensemble, la méthode à noyaux est plus performante que l'approximation polynomiale. Le gain est de l'ordre de 3dB pour l'EQM partielle. Ensuite, pour les deux approximations polynomiales d'ordre 7 et 11, les résultats sont similaires. Afin de tirer de plus amples conclusions pour le choix de l'estimateur, nous devons faire une caractérisation plus large des estimateurs en fonction du rapport signal à bruit et des distributions du signal de réflectivité. Par la suite, nous proposons pour chaque valeur de RSB, de calculer l'erreur quadratique moyenne pour 1000 configurations aléatoires. Le bruit et la réflectivité sont tirés aléatoirement. Le paramètre des lois de Bernoulli est toujours pris égal à 1/20.

Nous comparons sur la figure 2.6 trois estimateurs, (i) l'estimateur à noyaux de Pham, l'estimateur direct avec un modèle polynomial (ii) de degré 7 et (iii) de degré 11.

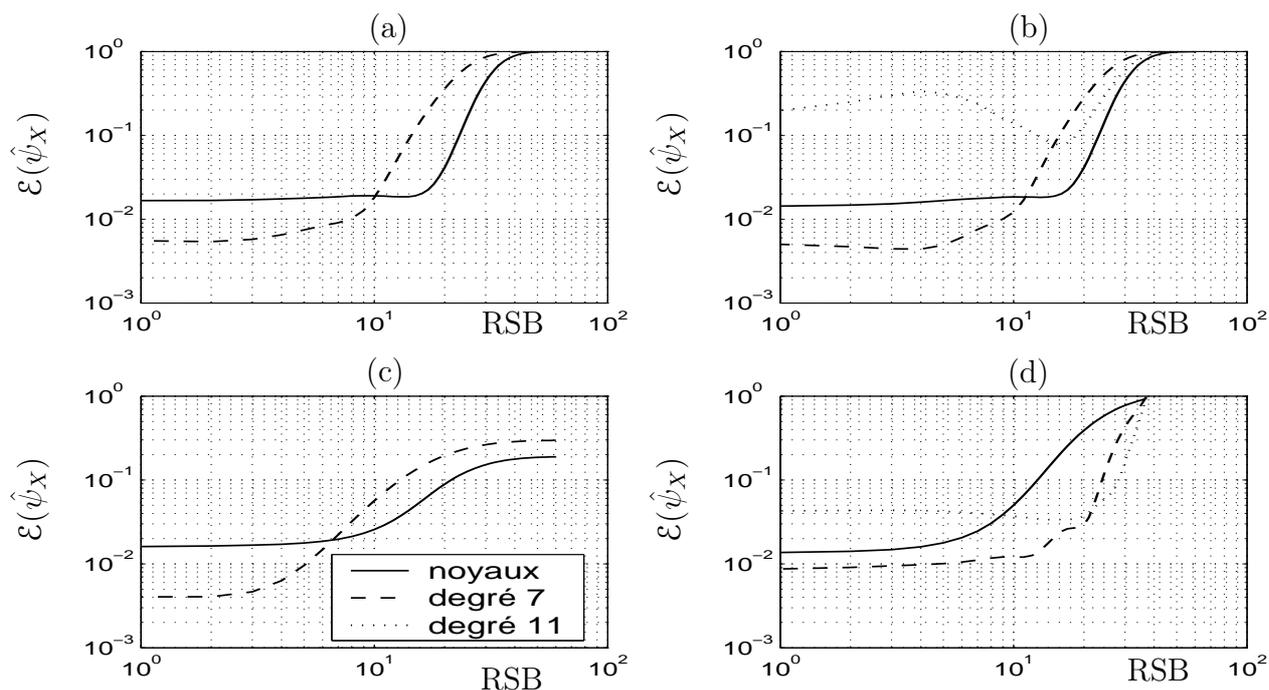


FIG. 2.6 – Erreur quadratique moyenne des estimateurs de la fonction score en fonction du RSB. Estimateur à noyaux (—), estimation directe avec un polynôme de degré 7 (- -) et de degré 11 (-.). Loi des signaux de réflectivité : (a) Bernoulli-laplacien, (b) Bernoulli-gaussien, (c) laplacien, (d) uniforme sur $[-\sqrt{3}; \sqrt{3}]$ avec un bruit additif gaussien.

Pour les quatre signaux d'entrée différents, les estimateurs à noyaux et l'approximation de degré 7 donnent des courbes de performances ayant des allures assez similaires. Pour les

plus petits RSB, *i.e.* où le bruit est dominant, la fonction score est proche de celle d'un processus gaussien (*i.e.* $\psi_X(x) = x/\sigma_N^2$, où σ_N est l'écart type du bruit gaussien). Ainsi, pour ces RSB, un simple modèle polynomial permet d'avoir une estimation convenable. L'estimateur à noyaux a des performances légèrement inférieures à celles de l'approximation polynomiale de degré 7, mais elles sont toutefois très acceptables. Le polynôme de degré 7 donne de meilleures performances que celui de degré 11 car il a un nombre plus faible de coefficients à estimer ce qui limite certainement la variance d'estimation.

Pour les plus grands RSB (lorsque le signal est prédominant par rapport au bruit), l'erreur quadratique moyenne d'estimation de la fonction score augmente pour toutes les distributions et tous les estimateurs. En fait, pour les grands RSB, les fonctions scores sont proches des fonctions scores des réflectivités sans bruit. Comme nous le remarquons sur la figure 2.4(a)(b)(d), pour les signaux Bernoulli-laplacien ou Bernoulli-gaussien ou uniforme, les fonctions ψ_X ont une allure complexe avec des variations très fortes et qui tendent même à comporter des singularités. D'autre part, à cause de la parcimonie, nous avons un grand nombre d'échantillons quasi nuls qui se situent dans la partie linéaire de la fonction score. En revanche, un nombre faible d'échantillons de valeurs importantes permettent d'estimer la partie complexe de la fonction score. Ainsi, tous les estimateurs ont des performances assez faibles.

De même pour les signaux Bernoulli-laplacien et Bernoulli-gaussien, nous avons théoriquement une discontinuité en zéro avec la dérivée d'un Dirac. Pour le processus laplacien, la fonction score est un échelon de Heaviside (figure 2.4(c)), qui est difficile à estimer. Ces remarques expliquent les problèmes d'estimation rencontrés par les 3 estimateurs. Cependant, l'estimateur à noyaux s'accommode mieux de ces difficultés que l'approximation polynomiale de degré 7 pour les signaux, Bernoulli-gaussien, laplacien et Bernoulli-Laplacien. Pour le signal uniforme, c'est le contraire, car la convexité de la fonction score théorique convient mieux au modèle polynomial. Enfin, nous pourrions penser qu'augmenter le degré du polynôme permettrait d'augmenter les performances de l'estimateur. En effet, en passant d'un degré 7 à 11 nous ajoutons deux paramètres. Certes, nous gagnons en flexibilité sur le modèle paramétrique de l'estimateur. Mais, ce gain de flexibilité se fait au détriment d'une estimation difficile des paramètres. Nous nous heurtons à des problèmes de conditionnement de la matrice du système (2.47). L'inversion est même tout simplement impossible dans de nombreux cas. Pour les distributions Bernoulli-laplacienne (en (a)) et laplacienne (en (c)), nous n'avons pas représenté les performances de l'estimateur avec le polynôme d'ordre 11, car cette estimation n'a pas de sens avec une erreur quadratique supérieure à 0dB. Nous pouvons voir sur la figure 2.6(b)(d), pour les signaux Bernoulli-gaussien et uniforme que le polynôme de degré 11 donne des résultats moins bons que le polynôme de degré 7.

En conclusion, nous utiliserons dans toutes nos expériences, l'estimateur à noyaux avec un noyau à base de fonction spline cubique. Cette estimation nécessite l'ajustement d'un seul paramètre, la largeur de bande du noyau h mais un réglage automatique est possible [Pha04] et donne des résultats corrects pour toutes les distributions étudiées en sismique. Le choix de l'estimateur polynomial poserait le problème du réglage du degré du polynôme. Nous avons vu par exemple que le passage d'un degré 7 à 11 faisait chuter drastiquement

les performances. Le choix de la méthode à noyaux l'emporte en particulier en terme de robustesse. En effet, une très mauvaise estimation de la fonction score lors d'une itération peut totalement déstabiliser l'algorithme de déconvolution.

2.5 Conclusion

Dans ce chapitre, nous avons présenté une mesure de blancheur d'un processus : le taux d'information mutuelle. Cette mesure s'apparente à la mesure d'indépendance de sources par l'information mutuelle, utilisée très largement en séparation de sources. Le taux d'information mutuelle permet de mesurer l'indépendance au sens large grâce à l'entropie. Puis, nous avons détaillé le premier algorithme de déconvolution basé sur le taux d'information mutuelle, qui minimise cette grandeur en fonction d'un nombre fini de coefficients de la réponse impulsionnelle. Cet algorithme est nommé MAMV car il s'apparente à une méthode de maximum de vraisemblance pour l'estimation d'un filtre MA. Cette estimation du filtre déconvolveur se fait par une descente de gradient. Nous avons proposé une extension avec l'algorithme ARMV pour l'estimation des filtres inverses AR. Cette extension est faisable, mais ne peut être retenue par la suite car l'algorithme est très instable. L'extension au cas ARMA n'est pas envisagée pour les mêmes raisons. L'étape importante de ces algorithmes est l'estimation de la fonction score car elle permet de sélectionner les statistiques utilisées pour la mesure de blancheur en fonction de la ddp du signal de sortie. Nous avons détaillé ce problème pour sélectionner un estimateur à noyaux suivant des critères de robustesse et de précision afin d'obtenir de bonnes performances pour les algorithmes et sans risquer des problèmes d'instabilité dus à une mauvaise estimation de la fonction score. Dans la suite, nous souhaitons caractériser le comportement de l'algorithme MAMV en présence de bruit additif gaussien, et notamment en comparaison à celles des algorithmes à l'ordre 2.

Chapitre 3

Le taux d'information mutuelle pour la déconvolution de données bruitées

Sommaire

3.1 Applications et performances de l'algorithme MAMV	79
3.1.1 Données réelles : explosions séismo-volcaniques	79
3.1.2 Performances de l'algorithme MAMV en présence de bruit gaussien	82
3.2 Déconvolution par le taux d'information mutuelle dans le do-	
 maine fréquentiel	86
3.2.1 Critère de déconvolution	86
3.2.2 Gradient du critère et algorithme FBD	87
3.2.3 Réglages des hyperparamètres	90
3.2.4 Données simulées	91
3.3 Traitement de données réelles	93
3.3.1 Données réelles : explosions séismo-volcaniques	93
3.3.2 Données réelles : explosions sous-marines	96
3.4 Conclusions	98

Dans le chapitre précédent, on a proposé deux algorithmes de déconvolution basés sur des modèles paramétriques temporels du filtre déconvolueur MA et AR. Seul l'algorithme MAMV utilisant le modèle MA est intéressant en terme de stabilité de la minimisation par la technique de gradient. Cet algorithme est capable d'estimer des filtres à phase quelconque par l'intermédiaire des statistiques d'ordre supérieur. En déconvolution, les performances des algorithmes se détériorent en général très vite en présence de bruit. Nous remarquons qu'en théorie le modèle n'est pas inversible. Ainsi, pour de nombreuses applications, le filtrage de Wiener est souvent utilisé en raison de sa robustesse vis-à-vis du bruit. Par conséquent, nous devons nous attacher à caractériser au mieux les performances de l'algorithme MAMV en présence de bruit additif sur les données. Nous réaliserons une comparaison à la méthode de Yule-Walker (statistiques d'ordre 2). Nous débuterons ce chapitre par la présentation des résultats de l'algorithme MAMV sur des données réelles séismo-volcaniques en les comparant avec l'égalisation spectrale et la méthode de Yule-Walker. L'objectif de cette comparaison est d'évaluer l'intérêt des statistiques d'ordre supérieur pour la déconvolution de données réelles. Ensuite, une étude des performances en terme d'erreur quadratique moyenne en fonction du RSB pour des configurations simulées sera menée. Puis, nous proposerons un algorithme de déconvolution aveugle en fréquence (FBD) mesurant la blancheur du signal de sortie et comportant aussi un terme de régularisation pour limiter l'amplification du bruit additif. Nous finirons par des tests de ce nouvel algorithme sur des données synthétiques et réelles d'explosions sous marines et séismo-volcaniques. L'objectif sera de mettre en avant cette recherche de compromis entre qualité de la déconvolution et amplification du bruit inspirée du filtrage de Wiener.

3.1 Applications et performances de l'algorithme MAMV

Dans un premier temps, nous comparons l'algorithme MAMV à la méthode de Yule-Walker (statistiques d'ordre 2 dans le domaine temporel) et à l'égalisation spectrale (ordre 2 en fréquence) pour la déconvolution de données réelles séismo-volcaniques présentant des effets résonnants. Enfin, nous ferons une étude des performances de l'algorithme MAMV en présence de bruit additif gaussien en comparaison avec la méthode de Yule-Walker.

3.1.1 Données réelles : explosions séismo-volcaniques

L'application suivante envisage le cas d'un séismogramme du volcan Purace de Colombie. Lors d'une explosion à l'intérieur du volcan (séismes naturels), la cheminée du volcan agit comme une cavité résonnante. Nous enregistrons avec un géophone positionné à la surface du volcan à la fréquence d'échantillonnage de 100Hz un évènement à longue période [LGM02] avec une importante résonance (figure 3.1(a)). En utilisant le modèle de convolution de la figure 1.10, l'observation sera vue comme la convolution entre le signal de l'explosion (t) et le filtre $w(t)$ résonnant relatif à la cheminée. Ainsi, le signal $y(t)$ obtenu par déconvolution contient l'information relative à l'excitation du volcan. Le filtre résonnant direct w donne des informations sur la géométrie du volcan et la structure interne du volcan. Le processus d'excitation (explosion) est supposé blanc car il est bref dans le domaine temporel et que, généralement, le spectre d'une explosion est quasi blanc. L'application de la déconvolution à ces données a pour but de séparer au mieux les effets résonnants et ceux

provenant de l'excitation du volcan. En fonction des fréquences présentes dans le signal, il est possible de caractériser la structure du volcan. Pour l'instant, les sismologues ne parviennent pas à tirer d'informations de l'excitation car son estimation est délicate. De plus, l'effet résonnant du filtre direct w est modélisé classiquement par un filtre autorégressif, donc nous pouvons appliquer préférentiellement des algorithmes utilisant des filtres inverses MA pour ces données. Pour la figure 3.1, nous appliquons l'égalisation spectrale (b), et deux techniques dédiées aux filtres inverses MA : la méthode de Yule-Walker (c) utilisant les statistiques d'ordre 2 dans le domaine temporel et l'algorithme MAMV (d) détaillé précédemment. Nous choisissons pour ces deux méthodes des filtres avec 30 coefficients afin d'être certains d'avoir un nombre suffisant de coefficients pour permettre l'extraction de toutes les fréquences de résonance. Après vérification par des simulations, nous constatons qu'un choix de 10 à 20 coefficients donne des résultats similaires. Pour la comparaison, les signaux déconvolués sont normalisés à une amplitude maximale de 10.

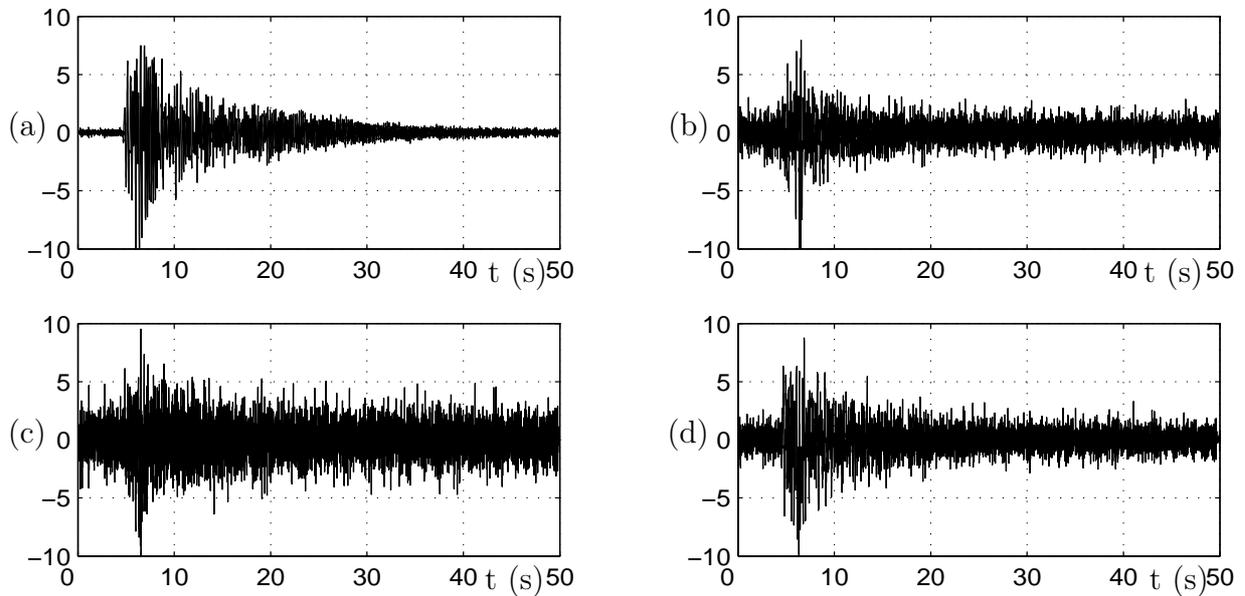


FIG. 3.1 – Déconvolution du séismogramme du volcan Purace : (a) observation, déconvolutions par (b) égalisation spectrale (c) Yule-Walker, (d) MAMV.

Sur la figure 3.1, les trois méthodes estiment un signal d'excitation situé entre les instants 5 s et 10 s. Dans la suite du signal, nous pouvons dire que nous avons uniquement du bruit. L'égalisation spectrale est la méthode qui amplifie le moins le bruit, suivie de la méthode MAMV et enfin de l'algorithme de Yule-Walker. Sur les signaux déconvolués, nous remarquons que nous avons réduit les effets de résonance des observations. Avec l'algorithme de Yule-Walker et MAMV, nous notons, qu'il reste encore une enveloppe exponentielle décroissante sur les données : tous les effets résonnants n'ont donc pas été supprimés. Afin d'avoir une comparaison plus précise, nous pouvons caractériser le niveau de bruit sur chacun des quatre signaux en estimant l'écart-type sur la partie comprises entre les instants 10 s et 50 s. Ensuite, pour mesurer la blancheur, nous estimons la mesure de blancheur $\tilde{I}(Y)$

(2.8) issue du taux d'information mutuelle. Le tableau 3.1 résume ces mesures.

TAB. 3.1 – Comparaison de la déconvolution du séismogramme du volcan Purace.

	Données	Egalisation spectrale	Yule-Walker	MAMV
Écart type du bruit	0.67	0.82	1.45	0.9
Mesure de blancheur $\tilde{I}(Y)$	1.46	-0.52	-0.21	-0.22

Le tableau 3.1 montre effectivement que l'égalisation spectrale donne le niveau de bruit le plus faible. Ce résultat était prévisible, car l'égalisation spectrale est la seule des trois méthodes prenant en compte le bruit additif. Il en est de même pour la blancheur mesurée par le taux d'information mutuelle. En fait les résultats de l'estimation de l'algorithme MAMV et de la méthode de Yule-Walker sont quasi identiques en raison du bruit additif important. L'égalisation spectrale estime une sortie plus blanche en terme d'information mutuelle en comparaison à l'algorithme MAMV qui minimise le taux d'information mutuelle. Cette remarque peut paraître paradoxale. Une explication est liée au modèle du filtre déconvolveur utilisé. En effet, pour l'égalisation spectrale, nous travaillons en fréquence, en revanche, pour l'algorithme MAMV, nous possédons seulement un nombre fini de coefficients temporels à estimer (30 ici). Ainsi, nous pouvons avancer comme justification le nombre de paramètres qui est plus faible pour l'algorithme MAMV donc il fournit une sortie moins blanche que pour l'égalisation spectrale.

De plus, en sortie nous mesurons le taux d'information mutuelle de $y(t) = g \star w \star r(t) + g \star n(t)$ et non pas seulement de la partie provenant de la réflectivité $g \star w \star r(t)$. Nous ajustons le filtre déconvolveur g pour augmenter la blancheur du terme $g \star w \star r(t)$ dans le même temps nous diminuons la blancheur du terme $g \star n(t)$. Ainsi, nous maximisons la blancheur de la somme $g \star w \star r(t) + g \star n(t)$.

L'égalisation spectrale ne blanchit pas totalement $g \star w \star r(t)$ mais elle a l'avantage de limiter grandement la puissance du terme $g \star n(t)$. Enfin, pour terminer la comparaison, nous traçons sur la figure 3.2 le filtre direct estimé par les trois algorithmes afin d'identifier les fréquences de résonance extraites du signal observé. Nous effectuons un zoom pour les fréquences réduites entre 0 et 0.15, afin d'analyser la partie intéressante du spectre.

Sur la figure 3.2(a), le spectre du filtre direct estimé par la méthode de l'égalisation spectrale comporte beaucoup de fréquences de résonance. Le spectre étant très peu lisse, il est difficile d'identifier les fréquences de résonance. Par contre, pour les méthodes de Yule-Walker (b) et l'algorithme MAMV (c), nous estimons respectivement 3 et 4 fréquences de résonance malgré 30 coefficients pour le filtre MA inverse. Théoriquement, les 30 coefficients permettent d'estimer 15 fréquences de résonance. Ainsi, ces tracés de spectre de filtre déconvolveur corroborent la remarque faite sur les allures temporelles des signaux : l'égalisation spectrale élimine mieux les effets résonnants que les méthodes de Yule-Walker et MAMV. Nous noterons que les pics sont plus étroits pour l'algorithme MAMV que pour l'algorithme de Yule-Walker. Ainsi, l'algorithme MAMV estime plus facilement des

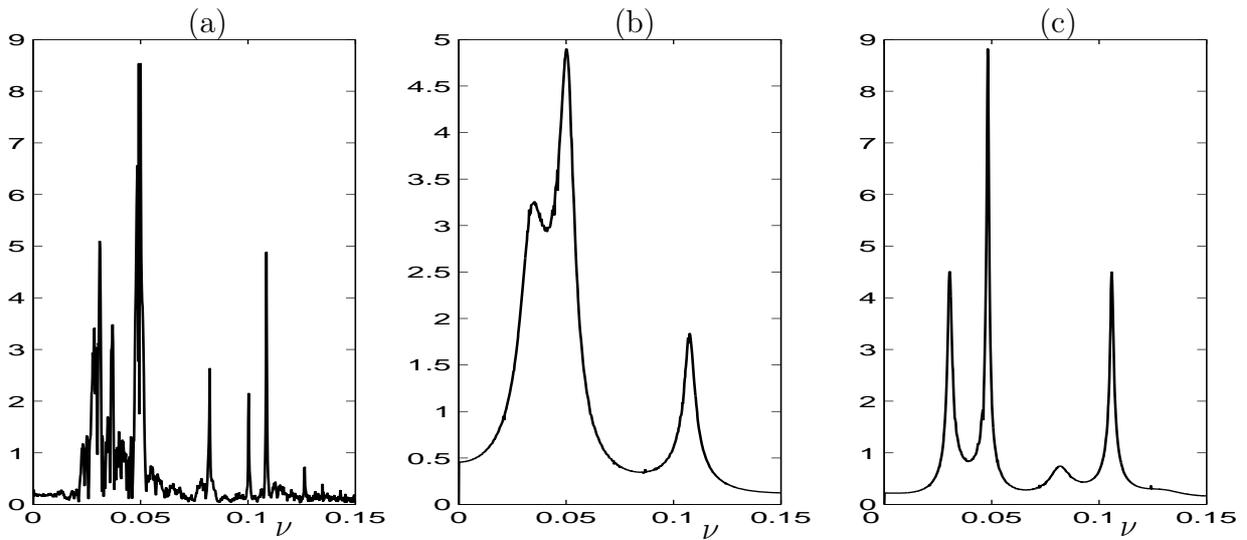


FIG. 3.2 – Spectre du filtre direct estimé (échelle linéaire) lors de la déconvolution du séismogramme du volcan Purace (figure 3.1) par : (a) égalisation spectrale (b) Yule-Walker, (c) MAMV.

zéros proches du cercle unité que l’algorithme de Yule-Walker. En conclusion sur ces données séismo-volcaniques, l’égalisation spectrale donne des résultats supérieurs à l’algorithme MAMV en terme d’extraction des effets résonnants et d’amplification du bruit. Cependant, l’algorithme MAMV ne donne pas de résultats meilleurs que l’algorithme de Yule-Walker pour une complexité plus importante. En fait, nous notons seulement l’extraction d’une fréquence de résonance de plus, et, une meilleure résolution fréquentielle. Nous proposons d’approfondir la comparaison de la méthode MAMV et de Yule-Walker en contexte bruité avec des signaux simulés.

3.1.2 Performances de l’algorithme MAMV en présence de bruit gaussien

Pour expliquer ces faibles améliorations de performance avec l’utilisation de l’ensemble des statistiques d’ordre supérieur (taux d’information mutuelle) par rapport à un simple blanchiment à l’ordre 2 (Yule-Walker), nous devons analyser l’influence d’un bruit additif gaussien sur les données. Nous proposons d’étudier cet effet dans un cadre totalement contrôlé avec des signaux simulés et un RSB choisi. Nous rappelons que les deux méthodes MAMV et Yule-Walker estiment un filtre déconvolveur MA (soit une ondelette ou filtre direct autorégressif). Le cadre de l’expérience est le suivant :

- La réflectivité suit une loi Bernoulli-gaussienne de paramètre $1/20$ et comporte 400 échantillons,
- L’ondelette est autorégressive avec les coefficients $[1, 0.4, 0.5, 0.45, 0.4, 0.1]$,
- Le bruit additif est blanc gaussien centré et indépendant de la réflectivité.

Nous effectuerons une étude pour un rapport signal à bruit (RSB) variant de $1dB$ à $50dB$. Pour chaque valeur de RSB nous effectuons 1000 configurations aléatoires, c’est-à-

dire que le bruit et la réflectivité sont choisis de façon aléatoire. Ensuite, nous résoudrons le problème de déconvolution avec l'algorithme MAMV et la méthode de Yule-Walker. Pour comparer les performances de ces deux algorithmes nous nous intéresserons à l'erreur quadratique moyenne (EQM) de sortie définie par :

$$\mathcal{E}(y) = \frac{E[(y(t) - r(t))^2]}{E[r(t)^2]}. \quad (3.1)$$

La figure 3.3 représente l'erreur quadratique moyenne de sortie en fonction du RSB des données.

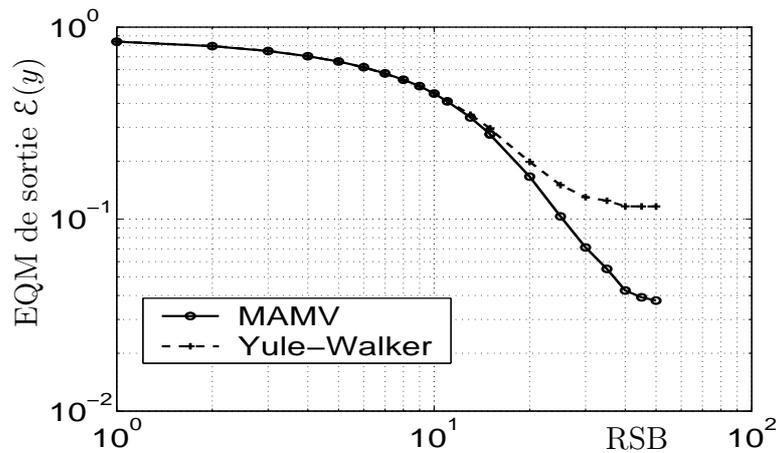


FIG. 3.3 – Comparaison de l'erreur quadratique (en dB) de sortie entre la réflectivité estimée et la réflectivité théorique en fonction du RSB. Algorithme MAMV (trait continu) et la méthode de Yule-Walker (trait discontinu).

Nous remarquons que les performances de l'algorithme MAMV en contexte peu bruité sont en moyenne meilleures que celle de l'algorithme de Yule-Walker. Nous notons une différence de l'ordre de 6dB pour les forts RSB. En revanche pour les faibles RSB (en contexte très bruité) les performances des deux algorithmes sont similaires car les deux courbes se superposent sur la figure 3.3. Pour poursuivre la comparaison, il est intéressant de considérer les coefficients du filtre déconvolveur estimés par les deux méthodes. En effet, même si nous estimons parfaitement le filtre déconvolveur (soit $g = w^{-1}$) alors la sortie estimée est $y(t) = r(t) + g \star n(t)$. Ainsi l'erreur quadratique d'estimation de la réflectivité n'est pas nulle en raison de la présence du bruit additif sur les données. Cependant, il est intéressant de savoir si les méthodes présentées sont capables ou non d'estimer correctement les coefficients du filtre déconvolveur même en présence de bruit. En effet, si le bruit n'implique pas de biais sur l'estimation du filtre déconvolveur alors, nous pourrions envisager un post-traitement pour supprimer ou limiter l'effet du bruit. Dans la figure 3.4, nous représentons en fonction du RSB le biais normalisé $b(\hat{a}_i)$ de l'estimation \hat{a}_i du coefficient a_i défini par :

$$b(\hat{a}_i) = \frac{E[\hat{a}_i - a_i]}{a_i} \quad \text{avec } i = 1 \dots 5, \quad (3.2)$$

avec $[a_1, a_2, a_3, a_4, a_5] = [0.4, 0.5, 0.45, 0.4, 0.1]$ (le coefficient a_0 n'est pas considéré car nous normalisons le filtre par rapport à son premier coefficient).

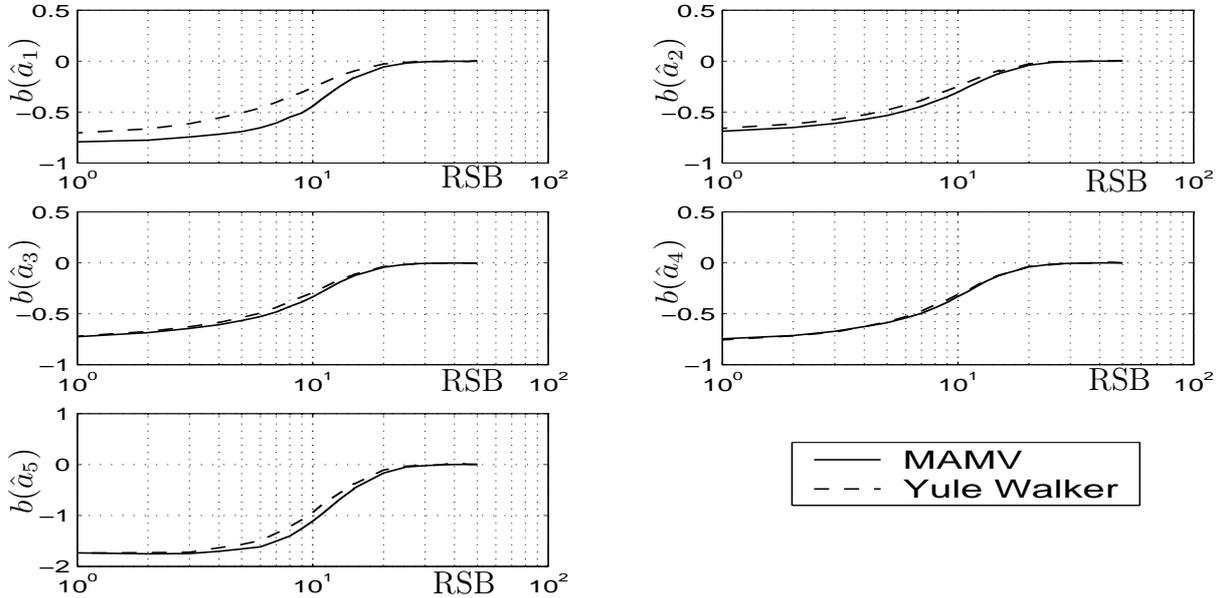


FIG. 3.4 – Biais normalisé $b(\hat{a}_i)$ (échelle linéaire) de l'estimateur \hat{a}_i pour la méthode MAMV (trait continu) et Yule-Walker (trait discontinu) en fonction du RSB des données.

Cette figure 3.4 montre que le biais est quasi identique pour la méthode MAMV et Yule-Walker, quel que soit le RSB. Pour les forts RSB, le biais est nul comme nous pouvons le prévoir car les méthodes sont appliquées sur des données respectant scrupuleusement le modèle. En revanche, pour les faibles RSB, le biais est très important pour les deux méthodes. Nous avons fait le choix de coefficients tous positifs, ainsi comme le biais est négatif, tous les coefficients sont sous-estimés. En effet, la variance du terme $g \star n(t)$ est $\sum_{i=0}^P a_i^2 \sigma_N^2$. Comme nous mesurons la blancheur du terme $g \star w \star r(t) + g \star n(t)$, l'algorithme a tendance à limiter la puissance de la partie bruit sur le signal déconvolué. Ainsi, les coefficients seront sous-estimés. Pour des RSB très faibles, le filtre tendra vers le filtre identité car l'observation est constituée presque uniquement du bruit (qui est déjà blanc). Ainsi, les coefficients \hat{a}_i tendent vers 0 pour $i \neq 0$: le biais normalisé $b(\hat{a}_i)$ tend vers -1 comme le montrent les tracés de la figure 3.4. Le cas du dernier coefficient est légèrement différent car sa valeur est faible et son importance sur l'allure du signal de sortie sera négligeable.

Pour finir notre analyse, la figure 3.5 trace l'erreur quadratique $\mathcal{E}(\hat{a}_i)$ de l'estimateur \hat{a}_i en fonction du RSB.

Avec la figure 3.5, nous remarquons pour les faibles RSB que les deux estimateurs sont équivalents pour les cinq coefficients du filtre. Nous pouvons noter un comportement un peu différent pour le dernier coefficient. Sa valeur étant faible (0.1), l'écart d'erreur quadratique sur ce coefficient aura des effets assez limités sur le signal déconvolué. Pour l'algorithme de Yule-Walker, l'EQM est quasi constante en fonction du RSB pour tous les coefficients.

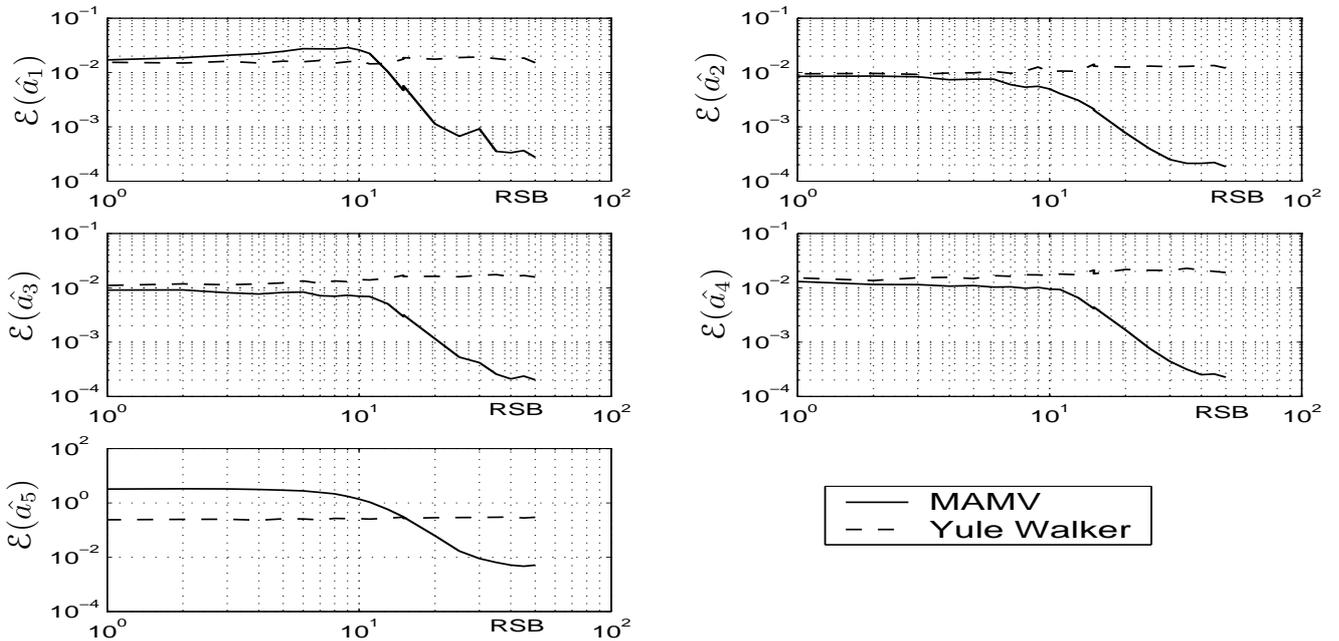


FIG. 3.5 – Comparaison de l'EQM $\mathcal{E}(\hat{a}_i)$ de l'estimateur \hat{a}_i du coefficient du filtre en fonction du RSB pour la méthode MAMV et de Yule-Walker.

En revanche, pour l'algorithme MAMV les performances s'améliorent grandement quand le RSB augmente. Dans la figure 3.4, le biais pour les deux algorithmes est identique ainsi, si l'EQM est plus faible pour l'algorithme MAMV, cela implique que l'estimateur MAMV a une variance beaucoup plus faible. D'après la figure 3.5, pour des RSB supérieurs à 10 dB nous avons un gain significatif de performance en terme de variance d'estimation par rapport à la méthode de Yule-Walker. En revanche en dessous de 10dB, les performances sont quasi identiques. En conclusion, en contexte bruité avec un bruit gaussien, les performances de l'algorithme MAMV, en terme d'EQM de sortie et d'EQM sur l'estimation des coefficients du filtre, tendent vers celles de la méthode à l'ordre 2 de Yule-Walker. Cette conclusion est en accord avec les tracés des fonctions scores théoriques de la figure 2.4 (chapitre 2) qui tendent vers une droite pour des faibles RSB. Nous avons remarqué à l'aide de l'équation (2.22) que si la fonction score est linéaire, ainsi nous sélectionnons pour la mesure de blancheur les statistiques d'ordre 2, donc l'algorithme est équivalent à un blanchiment à l'ordre 2.

Dans cette section, l'application de l'algorithme MAMV à des données réelles sismovolcaniques, et la comparaison avec d'autres méthodes, fait apparaître des performances proches de l'algorithme de Yule-Walker basé sur les statistiques d'ordre 2. Une étude plus précise en contexte bruité a démontré que la présence d'un bruit additif gaussien important tend à rendre la méthode tout simplement équivalente à la méthode à l'ordre 2. Dans la section suivante, nous proposons de prendre en compte le bruit additif sur les données dans la construction d'un algorithme se basant sur le taux d'information mutuelle de la sortie. Comme pour le filtrage de Wiener, pour limiter l'amplification du bruit, il faut en général ne

pas avoir un gain trop grand du filtre déconvolveur dans la bande de fréquence atténuée par l'ondelette. L'information mutuelle mesure la blancheur dans le domaine temporel, ainsi, dans un premier temps, il est naturel de proposer des algorithmes travaillant dans le domaine temporel. Dans la section suivante, nous proposons un algorithme de déconvolution dans le domaine fréquentiel car il est plus facile de contrôler l'amplification du bruit en fréquence dans le cas d'ondelette à bande passante limitée et de bruit additif blanc. De plus, l'approche fréquentielle permet de s'affranchir du choix du modèle paramétrique du filtre.

3.2 Déconvolution par le taux d'information mutuelle dans le domaine fréquentiel

Dans un premier temps, nous allons détailler la construction de la fonction coût de l'algorithme utilisant le taux d'information mutuelle en y associant le problème de l'amplification du bruit additif. Cet algorithme est fondé sur une formulation qui comporte évidemment le taux d'information mutuelle de la sortie ainsi que deux termes de régularisation. Le premier assurera la continuité de la réponse en fréquence du filtre déconvolveur. Le second limitera l'amplification du bruit en pénalisant les grandes valeurs du spectre du filtre déconvolveur. Cet algorithme sera nommé FBD (Frequency Blind Deconvolution).

3.2.1 Critère de déconvolution

Dans l'équation de la mesure de blancheur $\tilde{I}(Y)$ (2.8), nous pouvons exprimer l'intégrale en utilisant le gain en fréquence du filtre déconvolveur g . Ainsi, nous utilisons la transformée de Fourier discrète $\mathbf{G} = [G(0), \dots, G(T-1)]$ du filtre déconvolveur :

$$G(\nu) = \sum_{t=0}^{T-1} g(t) \exp\left(-\frac{j2\pi t\nu}{T}\right), \quad \forall \nu = 0, \dots, T-1. \quad (3.3)$$

En approximant l'intégrale par une méthode des rectangles, la mesure de blancheur (2.8) peut être estimée par :

$$\tilde{I}(Y) = H(y(\tau)) - \frac{1}{T} \sum_{\nu=0}^{T-1} \log |G(\nu)|. \quad (3.4)$$

La première idée est de minimiser directement la relation (3.4) par rapport à la réponse en fréquence du filtre déconvolveur $\mathbf{G} = [G(0), \dots, G(T-1)]$. En fait, si nous adoptons cette démarche, nous devons estimer T paramètres complexes (en réalité $T/2$ car nous avons la symétrie hermitienne pour un filtre réel). Cette estimation de $T/2$ nombres complexes pour calculer en sortie T échantillons réels du signal $y(t)$ nous place dans un schéma d'un problème inverse mal conditionné. Par exemple, dans le cas d'une réflectivité Bernoulli-gaussienne ou Bernoulli-laplacienne, sans bruit additif sur les données, nous obtenons la solution triviale comportant un seul réflecteur. Ce réflecteur est à la position du réflecteur principal de la réflectivité simulée. Ce signal comportant un seul réflecteur est bien effectivement le signal le plus blanc ressemblant à la réflectivité simulée. Par conséquent, nous devons ajouter des contraintes sur le filtre déconvolveur pour parvenir à une solution

réaliste et acceptable. Ces contraintes sont ajoutées sous forme d'*a priori* sur la réponse en fréquence du filtre déconvolveur. Nous ajoutons tout d'abord une contrainte de continuité, *i.e.* qui contrôle la différence $|G(\nu) - G(\nu + 1)|$. Enfin, pour limiter l'amplification du bruit sur la sortie, comme dans le filtrage de Wiener, nous devons limiter les très grandes valeurs de $|G(\nu)|$ par un second terme de régularisation.

Suite à ces remarques, nous proposons le critère de déconvolution aveugle en fréquence définie par :

$$J(\mathbf{G}) = H(y(\tau)) - \frac{1}{T} \sum_{\nu=0}^{T-1} \log |G(\nu)| + \lambda_1 \sum_{\nu=0}^{T-1} |G(\nu) - G(\nu + 1)|^2 + \lambda_2 \sum_{\nu=0}^{T-1} |G(\nu)|^p, \quad (3.5)$$

où λ_1 , λ_2 et p sont trois hyperparamètres.

La contrainte de continuité est totalement définie en utilisant la périodicité de la transformée de Fourier, soit $G(T) = G(0)$. Le premier terme de régularisation pondéré par l'hyperparamètre λ_1 contraint le filtre déconvolveur à avoir une réponse en fréquence continue. Plus précisément, il équivaut à un choix de distribution gaussienne pour $|G(\nu) - G(\nu + 1)|$, avec un maximum de la densité pour $|G(\nu) - G(\nu + 1)| = 0$. En pratique, ce terme améliore la stabilité de l'algorithme et ses performances. En effet, la contrainte de continuité influe comme une réduction du nombre de paramètres à estimer.

Le dernier terme, pondéré par λ_2 , pénalise (avec la norme \mathbb{L}^p) les grandes valeurs du gain en fréquence du filtre déconvolveur. Ce terme est équivalent au facteur de bruit classique du filtrage de Wiener : il permet un compromis entre une déconvolution de qualité (blancheur du signal) et l'amplification du bruit. Dans le cas d'une ondelette à bande passante limitée et d'un bruit additif blanc (densité spectrale de puissance constante), nous ne devons pas amplifier trop largement la bande atténuée de l'ondelette pour reconstruire cette bande de fréquence de la réflectivité car dans le même temps nous amplifions le bruit. Nous pouvons interpréter ce critère comme un critère de Maximum *a Posteriori* (MAP). Il est équivalent de prendre une distribution *a priori* gaussienne pour la loi $p(|G(\nu)| \mid |G(\nu - 1)|)$ de $|G(\nu)|$ conditionnellement à $|G(\nu - 1)|$ et une distribution gaussienne généralisée avec un coefficient p pour la loi *a priori* marginale de $|G(\nu)|$. Dans la suite, nous souhaitons minimiser ce critère (3.5) par rapport au vecteur complexe \mathbf{G} suivant une méthode de gradient.

Ainsi, dans la partie suivante, nous détaillons le calcul du gradient et enfin l'algorithme FBD de déconvolution aveugle dans le domaine fréquentiel.

3.2.2 Gradient du critère et algorithme FBD

Nous devons calculer le gradient d'une fonction réelle $J(\mathbf{G})$ par rapport à un vecteur complexe $\mathbf{G} = [G(0), \dots, G(T - 1)]$ pour trouver le minimum de la fonction coût $J(\mathbf{G})$. L'annexe C présente les éléments mathématiques nécessaires à la définition du gradient d'une fonction réelle d'une variable complexe [Bra83, VDB94]. Elle définit la dérivée par rapport à une variable complexe ainsi que les règles de base de dérivation. Suivant cette annexe C, nous devons calculer la dérivée de la fonction coût $J(\mathbf{G})$ par rapport à $G^*(\theta)$, $\theta = 0, \dots, T - 1$ (voir annexe C). En posant $|G(\nu)|^2 = G(\nu)G^*(\nu)$, le second terme du critère (3.5) se réécrit avec les variables $G(\nu)$ et $G^*(\nu)$ comme :

$$\frac{\partial}{\partial G^*(\theta)} \frac{1}{T} \sum_{\nu=0}^{T-1} \log |G(\nu)| = \frac{1}{2T} \frac{\partial}{\partial G^*(\theta)} \sum_{\nu=0}^{T-1} \log [G(\nu)G^*(\nu)]. \quad (3.6)$$

Comme la dérivation se passe (voir annexe C) comme si $G(\theta)$ et $G^*(\theta)$ étaient indépendants, nous avons :

$$\frac{\partial}{\partial G^*(\theta)} \frac{1}{T} \sum_{\nu=0}^{T-1} \log |G(\nu)| = \frac{1}{2T} \frac{1}{G^*(\theta)}. \quad (3.7)$$

De même, avec la propriété de périodicité de la réponse en fréquence, *i.e.* $G(-1) = G(T-1)$ et $G(T) = G(0)$, la dérivation des termes de régularisation donne :

$$\frac{\partial}{\partial G^*(\theta)} \sum_{\nu=0}^{T-1} |G(\nu) - G(\nu+1)|^2 = 2G(\theta) - G(\theta+1) - G(\theta-1), \quad (3.8)$$

$$\text{et} \quad \frac{\partial}{\partial G^*(\theta)} \sum_{\nu=0}^{T-1} |G(\nu)|^p = \frac{p}{2} \frac{|G(\theta)|^p}{G^*(\theta)}. \quad (3.9)$$

Pour le terme entropique, nous utilisons la définition de l'entropie ($H(u) = -E[\log p_U(u)]$) de la même manière que pour le calcul de (2.10) dans le cas de l'algorithme MAMV, nous avons ainsi :

$$\frac{\partial}{\partial G^*(\theta)} H(y(\tau)) = -\frac{\partial}{\partial G^*(\theta)} E[\log p_Y(y(\tau))]. \quad (3.10)$$

Après permutation des opérateurs d'espérance mathématique et de dérivation dans (3.10), nous devons calculer $\frac{\partial}{\partial G^*(\theta)} \log p_Y(y(\tau))$. Pour cela, nous utilisons la règle de dérivation en chaîne donnée par (C.6) pour obtenir :

$$\frac{\partial}{\partial G^*(\theta)} \log p_Y(y(\tau)) = \frac{\partial}{\partial y(\tau)} \log p_Y(y(\tau)) \frac{\partial}{\partial G^*(\theta)} y(\tau) + \frac{\partial}{\partial y(\tau)} \log p_Y(y(\tau)) \frac{\partial}{\partial G^*(\theta)} y^*(\tau). \quad (3.11)$$

Comme $y(\tau)$ est à valeurs réelles, nous avons :

$$\frac{\partial}{\partial G^*(\theta)} y(\tau) = 0, \quad (3.12)$$

$$\text{et} \quad \frac{\partial}{\partial G^*(\theta)} y^*(\tau) = \left(\frac{\partial}{\partial G(\theta)} y(\tau) \right)^*. \quad (3.13)$$

De plus, la relation de convolution $y(t) = (g \star d)(t)$ permet d'exprimer $\frac{\partial}{\partial G(\theta)} y(\tau)$ comme :

$$\frac{\partial}{\partial G(\theta)} y(\tau) = \frac{\partial}{\partial G(\theta)} \frac{1}{T} \sum_{\nu=0}^{T-1} G(\nu) D(\nu) \exp\left(\frac{j2\pi\nu\tau}{T}\right) \quad (3.14)$$

$$= \frac{1}{T} D(\theta) \exp\left(\frac{j2\pi\theta\tau}{T}\right), \quad (3.15)$$

où $D(\nu)$ est le spectre des observations $d(t)$.

Avec la définition (2.12) de la fonction score du chapitre précédent, et les relations (3.13) et (3.15), l'équation (3.11) s'exprime par :

$$\frac{\partial}{\partial G^*(\theta)} \log p_Y(y(\tau)) = -\frac{1}{2T} \psi_Y(y(\tau)) D^*(\theta) \exp\left(-\frac{j2\pi\theta\tau}{T}\right). \quad (3.16)$$

Finalement, pour estimer la dérivation du terme entropique (3.10), nous devons réaliser une moyenne statistique, qui sera équivalente à une moyenne temporelle sous hypothèse d'ergodicité. Nous estimons (3.10) par :

$$\frac{\partial \widehat{H}(y(\tau))}{\partial G^*(\theta)} = \frac{1}{2T^2} D^*(\theta) \sum_{\tau=1}^T \psi_Y(y(\tau)) \exp\left(-\frac{j2\pi\theta\tau}{T}\right). \quad (3.17)$$

Ainsi, nous devons estimer la fonction score aux T valeurs des échantillons $y(1), \dots, y(T)$. En notant $\Psi_Y(\theta)$ la transformée de Fourier de la fonction score $\psi_Y(y(\tau))$, (3.17) devient :

$$\frac{\partial \widehat{H}(y(\tau))}{\partial G^*(\theta)} = \frac{1}{2T^2} \Psi_Y(\theta) D^*(\theta). \quad (3.18)$$

Finalement, avec (3.7), (3.8), (3.9) et (3.18), le gradient du critère (3.5) par rapport $G(\nu)$ peut être estimé par¹ :

$$\widehat{\nabla J(\mathbf{G})} = \frac{1}{2T^2} \Psi_Y(\nu) D^*(\nu) - \frac{1}{2T} \frac{1}{G^*(\nu)} + \lambda_1 (2G(\nu) - G(\nu + 1) - G(\nu - 1)) + \lambda_2 \frac{p}{2} \frac{|G(\nu)|^p}{G^*(\nu)}. \quad (3.19)$$

Après le calcul du gradient, nous pouvons maintenant construire l'algorithme de déconvolution aveugle en fréquence basé sur la minimisation du taux d'information mutuelle de sortie auquel nous avons ajouté une régularisation vis-à-vis du bruit. La minimisation est faite avec une descente de gradient. La structure de l'algorithme FBD (Frequency domain Blind Deconvolution) est proche de celle de l'algorithme MAMV détaillée à la section 2.2.

L'algorithme FBD est le suivant :

1. Initialisation du filtre déconvolveur $G(\nu)$ et de la sortie $y(t)$;
2. Estimation de la fonction score ψ_Y ;
3. Estimation du gradient (3.19) ;
4. Mise à jour du filtre déconvolveur $G(\nu) \leftarrow G(\nu) - \mu \widehat{\nabla J(\mathbf{G})}$;
5. Calcul de la sortie de déconvolution $y(t)$;
6. Étape de normalisation.

¹Nous notons de nouveau la fréquence ν , alors que les dérivées précédentes étaient calculées par rapport à $G^*(\theta)$ pour éviter la confusion avec la variable ν utilisée dans la définition du critère. Mais, ν est une notation plus commune pour représenter une fréquence

Nous itérons les étapes 2 à 6 jusqu'à convergence. L'étape de normalisation est nécessaire pour lever l'indétermination d'échelle sur la sortie. En général, nous forçons la sortie à avoir une puissance unitaire. μ est le pas du gradient.

De la même façon que pour l'algorithme MAMV de la section 2.2 et l'algorithme ARMV de la section 2.3, l'estimation de la fonction score est équivalente à la sélection quasi-optimale des statistiques d'ordre supérieur pour la déconvolution de la sortie. Cette sélection est faite en fonction de la distribution de la sortie. D'après le chapitre 2, nous choisissons l'estimateur à noyaux de Pham. Après la présentation de l'algorithme FBD, nous allons maintenant aborder succinctement le problème du réglage des hyperparamètres λ_1 , λ_2 et p .

3.2.3 Réglages des hyperparamètres

Pour l'hyperparamètre λ_2 associé au troisième terme de l'équation (3.5), nous pouvons proposer une technique de réglage automatique. Ce terme est équivalent à une distribution *a priori* gaussienne généralisée pour $|G(\nu)|$. Cette distribution est définie par $p_G(g) = B \exp(-\lambda_2 |g|^p)$, où B est un terme de normalisation et nous avons² :

$$E[|G(\nu)|^2] = \frac{\Gamma(3/p)}{\Gamma(1/p)\lambda_2^{2/p}},$$

où Γ est la fonction gamma définie par $\Gamma(x) = \int_0^{+\infty} t^{x-1} \exp(-t) dt$. Par ailleurs, nous pouvons estimer $E[|G(\nu)|^2]$ avec la formule des interférences à l'aide des densités spectrales de puissance des observations et de la sortie, définies respectivement par $S_{2D}(\nu)$ et $S_{2Y}(\nu)$:

$$S_{2Y}(\nu) = |G(\nu)|^2 S_{2D}(\nu).$$

Comme la sortie $y(t)$ est blanche et de puissance unitaire, nous avons $S_{2Y}(\nu) = 1$. Par conséquent, nous avons la relation suivante pour λ_2 :

$$\lambda_2 = \left(\frac{\Gamma(3/p)}{\Gamma(1/p) E \left[\frac{1}{S_{2D}(\nu)} \right]} \right)^{p/2}. \quad (3.20)$$

Cette estimation de λ_2 est correcte dans le cas où la densité spectrale $S_{2D}(\nu)$ ne s'annule pour aucune fréquence. Nous pourrions envisager des alternatives estimant $E[|G(\nu)|^2]$ en utilisant la valeur de $|G(\nu)|$ estimée par une méthode du type égalisation spectrale nécessitant aussi le réglage d'un paramètre de bruit.

En général, nous ajustons λ_1 et λ_2 séparément, avec des essais successifs en observant la sortie de la déconvolution et le spectre du filtre déconvolveur (Nous débutons avec la valeur de λ_2 donnée par (3.20)) :

- si $G(\nu)$ n'est pas assez continue, λ_1 est augmenté,
- si la sortie $y(t)$ est trop bruitée, λ_2 est augmenté.

²Pour plus de détails sur les lois gaussiennes généralisées voir le paragraphe 4.2.1

Cette stratégie est assez efficace et donne une méthode simple et assez robuste pour régler λ_1 et λ_2 . L'ajustement n'a pas besoin d'être très fin. En effet, une variation $\lambda_i^*/2 < \lambda_i < 2\lambda_i^*$ autour de la valeur empirique λ_i^* n'a pas de grande influence sur les performances de la déconvolution.

Enfin, le réglage du coefficient de sur-gaussianité p n'est pas un point déterminant, si nous réglons correctement λ_2 . En effet, le couple (p, λ_2) définit complètement la distribution *a priori* pour $|G(\nu)|$. Si nous cherchons un filtre déconvolveur avec une réponse en fréquence comportant principalement des pics de résonance, nous devons prendre un *a priori* sur-gaussien pour la réponse en fréquence, donc p devra être inférieur à 2. Au contraire, si la réponse est assez constante, avec des valeurs resserrées autour de la moyenne nous devons prendre un *a priori* sous-gaussien (p supérieur à 2). Néanmoins, si le coefficient de sur-gaussianité p est mal réglé, un bon réglage de λ_2 (qui est équivalent à spécifier la variance de la distribution) peut compenser ce mauvais réglage.

Après la description de l'algorithme et le problème pratique du réglage des hyperparamètres, nous allons illustrer notre propos avec des résultats de simulations.

3.2.4 Données simulées

Dans cette première expérience, nous souhaitons montrer la possibilité d'estimation de filtres à phase quelconque, grâce à l'utilisation de l'ensemble des statistiques d'ordre supérieur. Les données simulées des figures 3.7 et 3.8 sont construites avec :

- une réflectivité Bernoulli-laplacienne de 400 échantillons avec 50 réflecteurs,
- une ondelette ARMA (10,10) dont les pôles et les zéros sont pris aléatoirement 2 à 2 conjugués (figure 3.6),
- une ondelette à phase minimale (figure 3.7),
- une ondelette à phase nulle (figure 3.8).

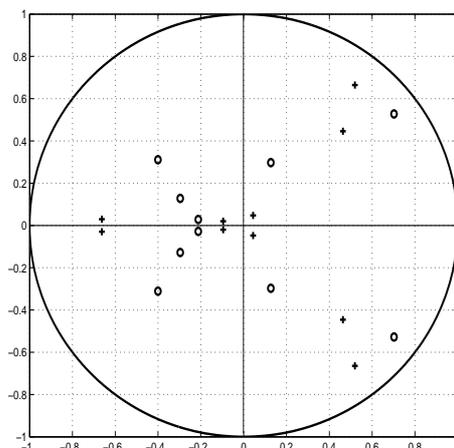


FIG. 3.6 – Position des pôles (+) et des zéros (o) de l'ondelette ARMA (10,10) utilisée pour les simulations des figures 3.7 et 3.8.

Nous déconvoluons les observations avec l'algorithme FBD. Dans les figures 3.7 et 3.8, nous traçons en (a) la réflectivité, en (b) les observations après convolution de la réflectivité par l'ondelette à phase nulle ou minimale. Nous noterons que l'ondelette est à réponse impulsionnelle assez longue en raison de la présence de résonance. Pour analyser la qualité de la déconvolution, nous traçons en (c) la réflectivité estimée et nous comparons en (d) (resp. (e)) la phase (resp. module) du filtre estimé par l'algorithme FBD avec sa valeur théorique.

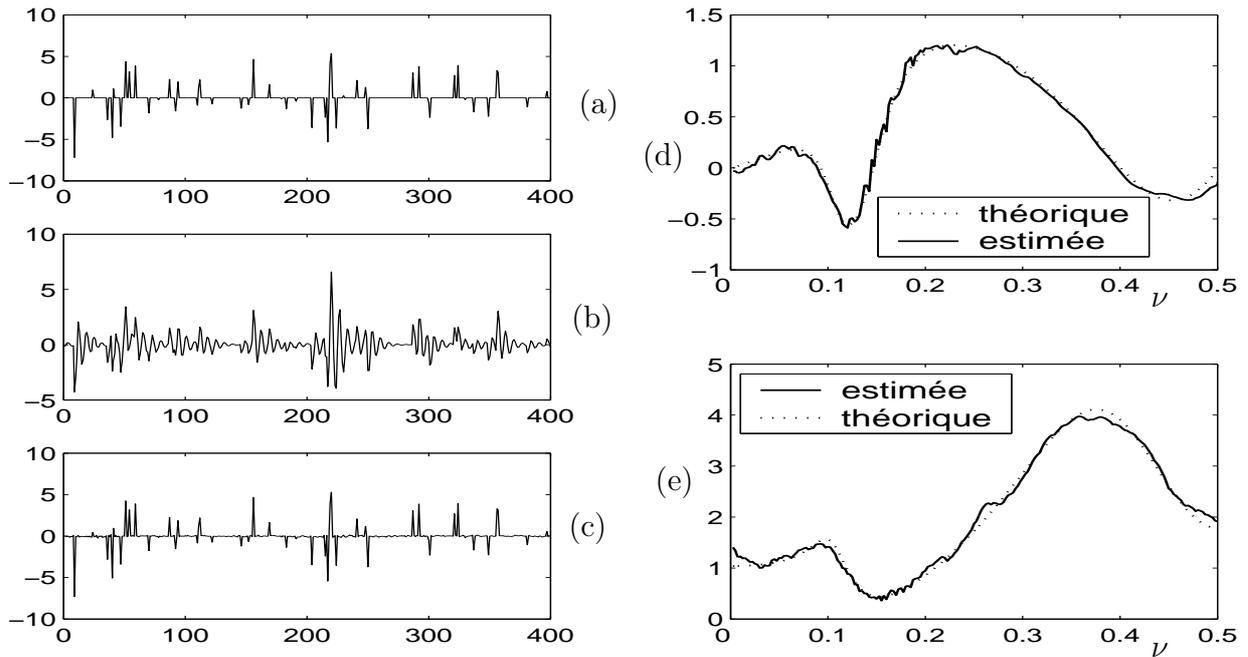


FIG. 3.7 – Déconvolution de données sismiques synthétiques avec une ondelette ARMA à phase minimale ($\lambda_1 = 0.1$, $\lambda_2 = 0$, $p = 2$) : (a) réflectivité Bernoulli-laplacienne, (b) observation avec **ondelette ARMA à phase minimale**, (c) réflectivité estimée par FBD, (d) phase théorique (pointillé) et phase estimée du filtre déconvolveur (trait plein) (e) idem que (d) pour le module du filtre.

En étudiant les figures 3.7 et 3.8 (a) et (c), nous remarquons que l'estimation de la réflectivité est très satisfaisante que la phase de l'ondelette soit nulle (figure 3.8) ou minimale (figure 3.7). En effet, en (e), nous avons une bonne estimation du module du filtre déconvolveur. La légère erreur d'estimation s'explique par le grand nombre de paramètres ($T/2$ complexes) à estimer par l'algorithme. La contrainte de continuité permet d'avoir une réponse assez lisse. Nous ne devons pas trop augmenter l'hyperparamètre λ_1 qui tend à lisser la réponse en fréquence. En effet, dans ce cas, nous serons dans l'impossibilité d'estimer des parties comportant des variations rapides, comme aux alentours de la fréquence réduite 0.12. Ainsi, le réglage du paramètre λ_1 doit prendre en compte ce compromis.

Enfin, la phase estimée en (d) est proche de la phase théorique. L'erreur d'estimation est légèrement supérieure dans le cas d'un filtre à phase nulle. La plus grande difficulté rencontrée lorsque nous estimons le filtre à phase nulle est due au fait que seules les statistiques d'ordre supérieur à 2 tendent à estimer le filtre à phase nulle (celles d'ordre 2 sont

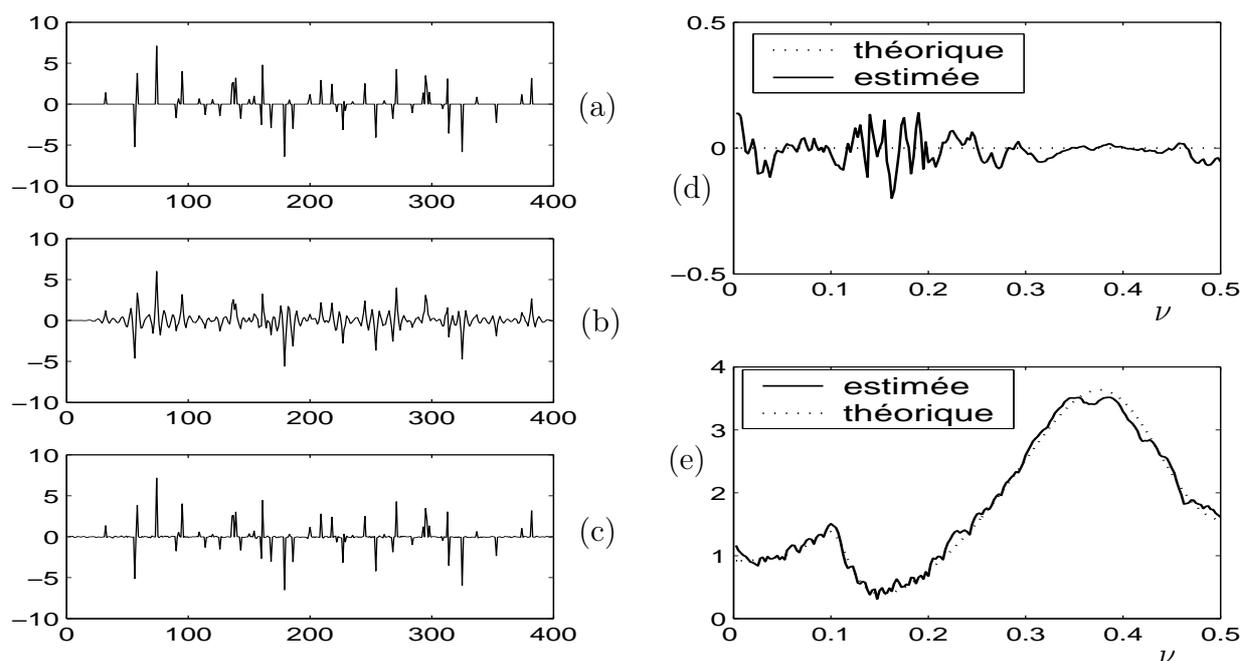


FIG. 3.8 – Idem figure 3.7 avec ondelette ARMA à phase nulle.

insensibles à la phase de l'ondelette).

Nous avons montré que l'algorithme FBD est capable d'estimer des filtres à phase quelconque par l'intermédiaire des statistiques d'ordre supérieur. De plus, la simulation précédente montre que l'approche fréquentielle permet de s'adapter à un filtre quelconque sans avoir à gérer le problème du choix du modèle paramétrique. Mais pour justifier pleinement l'approche fréquentielle, nous devons traiter des signaux bruités. Nous proposons dans la suite de comparer l'algorithme FBD aux techniques à l'ordre 2 (Yule-Walker, Durbin et égalisation spectrale) pour des données réelles d'explosions sous-marines et séismo-volcaniques. Pour ces dernières données, nous utiliserons aussi l'algorithme MAMV, pour montrer l'apport du terme de régularisation vis-à-vis du bruit. De plus, les deux jeux de données réelles feront appel à des filtres directs de type très différents. Ainsi, nous pourrions juger de la capacité d'adaptation de l'algorithme FBD en fonction du type de filtre direct.

3.3 Traitement de données réelles

3.3.1 Données réelles : explosions séismo-volcaniques

Nous souhaitons voir l'apport de notre algorithme sur les données séismo-volcaniques déjà traitées dans la sous-section 3.1.1. Nous rappelons que l'objectif pour ces données est de séparer les effets d'excitation et des effets résonnants caractéristiques de la géométrie et de la structure du volcan. Lors des premiers tests, l'apport des statistiques d'ordre supérieur avec l'algorithme MAMV n'était pas probant par rapport aux méthodes à l'ordre

2. Mais, nous avons vu l'utilité de la régularisation vis-à-vis du bruit pour l'égalisation spectrale. Comme dans la section 3.1.1 (figure 3.1 et 3.2), nous tracerons la sortie $y(t)$ de la déconvolution (colonne de gauche de la figure 3.9) et le spectre du filtre direct estimé avec un zoom pour les fréquences réduites entre 0 et 0.15 (colonne de droite de la figure 3.9). Nous comparons quatre algorithmes : l'égalisation spectrale, la méthode de Yule-Walker, l'algorithme MAMV du chapitre 2 et l'algorithme FBD présenté dans ce chapitre.

Pour l'algorithme FBD, nous avons choisi un paramètre $p = 1.5$ car les effets résonnants du filtre direct impliquent que le filtre déconvolveur aura une distribution *a priori* pour $|G(\nu)|$ sur-gaussienne. Le phénomène d'excitation est localisé entre les instants 5 s et 10 s. En analysant les allures temporelles estimées par les cinq algorithmes, nous pouvons conclure que l'algorithme FBD donne la meilleure séparation des effets de résonance et d'excitation avec une très faible amplification du bruit. Le tableau 3.2 évalue les performances des cinq algorithmes avec l'écart-type du bruit sur les 35 dernières secondes et mesure la blancheur de la sortie avec la grandeur $\tilde{I}(Y)$ (2.8) provenant du taux d'information mutuelle.

TAB. 3.2 – Comparaison des performances des méthodes de déconvolution pour le séismogramme du volcan Purace de la figure 3.9.

	Données	Yule-Walker	MAMV	Egalisation spectrale	FBD
Écart-type du bruit	0.67	1.45	0.9	0.82	0.43
Mesure de blancheur $\tilde{I}(Y)$	1.46	-0.21	-0.22	-0.52	-0.38

Le tableau 3.2 confirme que l'algorithme FBD donne le meilleur compromis entre la qualité de la déconvolution et le niveau de bruit. En effet, la sortie estimée par l'algorithme FBD a l'écart-type de la partie bruit la plus faible, et est la plus blanche après celle fournie par l'égalisation spectrale. Ainsi, en comparaison à l'algorithme MAMV qui s'appuie sur la même mesure de blancheur que l'algorithme FBD, nous notons la pertinence et l'apport de la régularisation vis-à-vis du bruit pour obtenir une meilleure estimation du signal d'excitation. Pour finir l'analyse des résultats, nous pouvons nous intéresser au filtre direct estimé et aux fréquences de résonance présentes dans la partie de fréquence réduite entre 0 et 0.15. Les cinq algorithmes donnent les mêmes fréquences de résonance prépondérantes avec une résolution plus ou moins fine. L'algorithme de Yule-Walker donne la résolution la plus mauvaise avec seulement trois fréquences de résonance. Sur l'allure temporelle de la déconvolution par l'algorithme de Yule-Walker, nous remarquons l'enveloppe exponentielle décroissante qui est synonyme de la présence résiduelle d'effets résonnants. Ensuite, l'algorithme MAMV estime des pôles plus proches du cercle unité car les pics sont plus étroits, mais nous extrayons seulement quatre fréquences de résonance. Pour l'égalisation spectrale, nous avons beaucoup de fréquences de résonance mais avec une mauvaise résolution. Pour l'algorithme FBD, nous avons des pics très étroits et l'estimation de nombreuses fréquences de résonance. Nous notons des résonances aux fréquences 0.05 et 0.1 : par conséquent nous pouvons dire que c'est le fondamental et le premier mode d'une vibration. Si nous observons avec plus de précision, nous détectons une petite résonance à 0.0415 et deux résonances plus

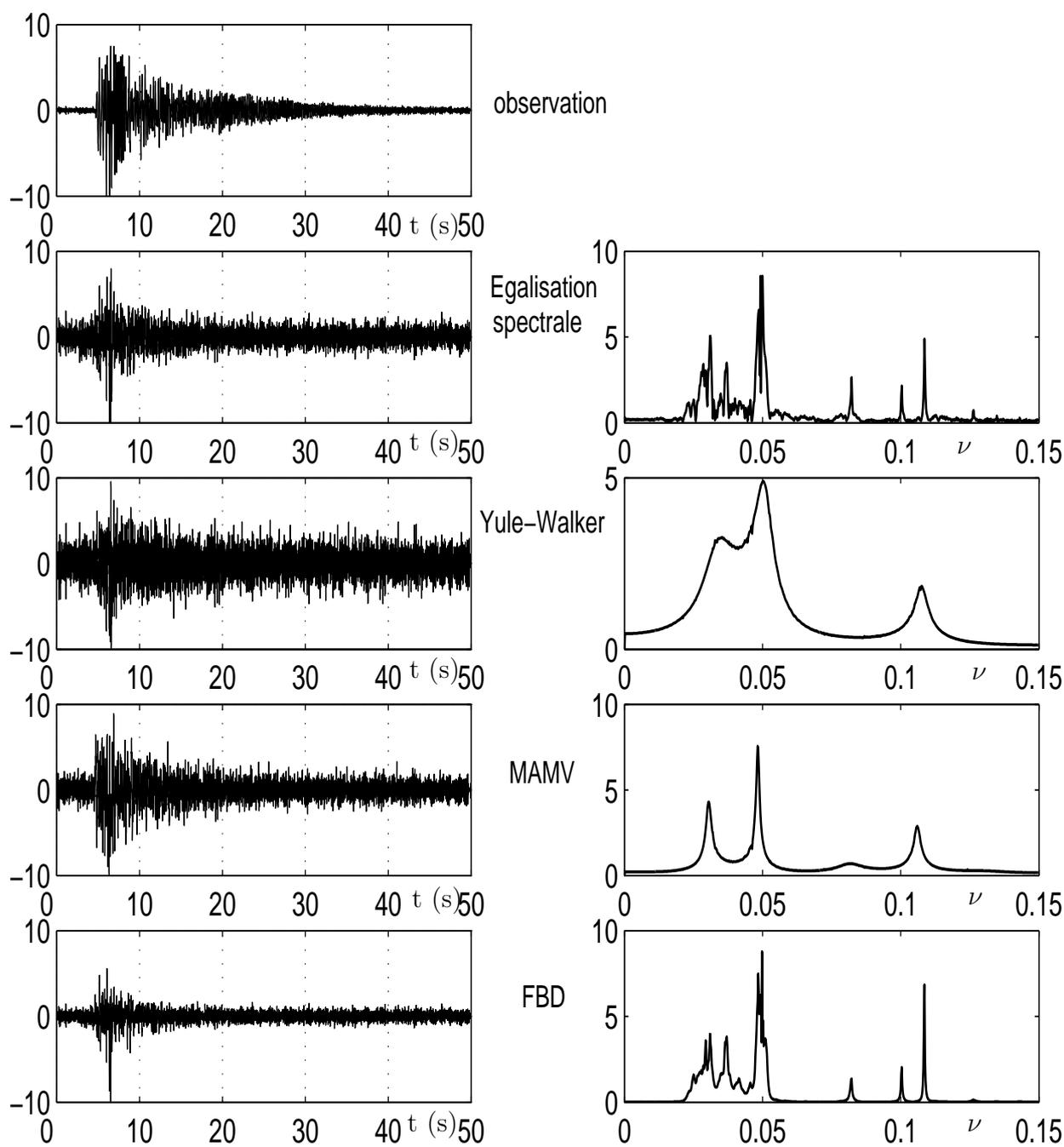


FIG. 3.9 – Déconvolution du séismogramme du volcan Purace. Sortie estimée (colonne de gauche) et réponse en fréquence (zoom pour les fréquences réduites entre 0 et 0.15) du filtre direct estimée (colonne de droite) par : l'égalisation spectrale, la méthode de Yule-Walker avec un filtre d'ordre 30, l'algorithme MAMV avec un filtre de 30 coefficients, l'algorithme FBD avec $\lambda_1 = 10$, $\lambda_2 = 5$ et $p = 1.5$.

claires à 0.082 et 0.126, que nous interprétons comme le fondamental et les deux premiers modes d'une autre vibration.

En conclusion, l'algorithme FBD permet d'avoir une bonne résolution en fréquence, avec une réponse en fréquence assez lisse grâce au premier terme de régularisation. La prise en compte du bruit avec le second terme de régularisation apporte une forte amélioration par rapport à l'algorithme MAMV qui utilise exactement la même mesure de blancheur. Pour ces données, le filtre direct était du type autorégressif, nous proposons de traiter d'autres données ayant un filtre direct plutôt du type MA.

3.3.2 Données réelles : explosions sous-marines

La figure 3.11(a) montre l'enregistrement réel d'une explosion sous-marine dans une piscine. Comme décrit par la figure 3.10, la source S est une source explosive et le capteur P mesure les variations de pression. Le signal est composé d'une onde directe ①, d'une réflexion au fond de la piscine ② avec un coefficient de réflexion proche de -1 et d'une réflexion à la surface ③ avec un coefficient de réflexion proche de 1.

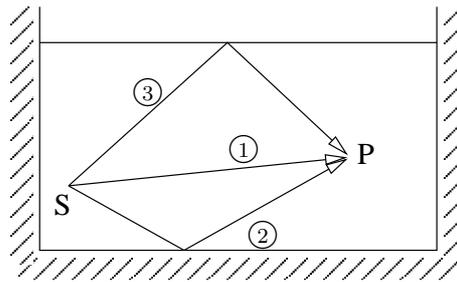


FIG. 3.10 – Expérience d'explosion sous-marine en piscine : (S) source explosive, (P) capteur de pression.

Dans cette application, nous pouvons utiliser un modèle de convolution entre la source explosive et le signal de réflectivité. Le signal de réflectivité comportera les informations relatives aux temps de propagation et aux coefficients de réflexion. Le filtre direct w représente l'onde émise par le signal d'explosion caractéristique de l'effet "bulle"³, avec une croissance exponentielle de la pression suivie d'une décroissance exponentielle. Nous pouvons modéliser cette onde par un modèle MA [Bou96] non causal. Dans la figure 3.11, nous comparons les résultats de la déconvolution des observations (a) suivant trois méthodes : égalisation spectrale (b), méthode de Durbin avec un filtre non causal d'une longueur de 40 échantillons (c), algorithme FBD (d).

Tout d'abord, à propos du réglage des paramètres de l'algorithme FBD, la valeur de λ_2 donnée par l'équation (3.20) est pertinente car nous avons une ondelette très large bande, et donc la densité spectrale des données ne s'annule jamais. Ensuite, nous prenons $p = 3$ car le gain du filtre déconvolveur n'a pas de bande de fréquence où il s'annule, ainsi nous prenons

³Une bulle d'air est créée, son diamètre augmente jusqu'à une pression de rupture et son diamètre redécroit jusqu'à disparition de la bulle

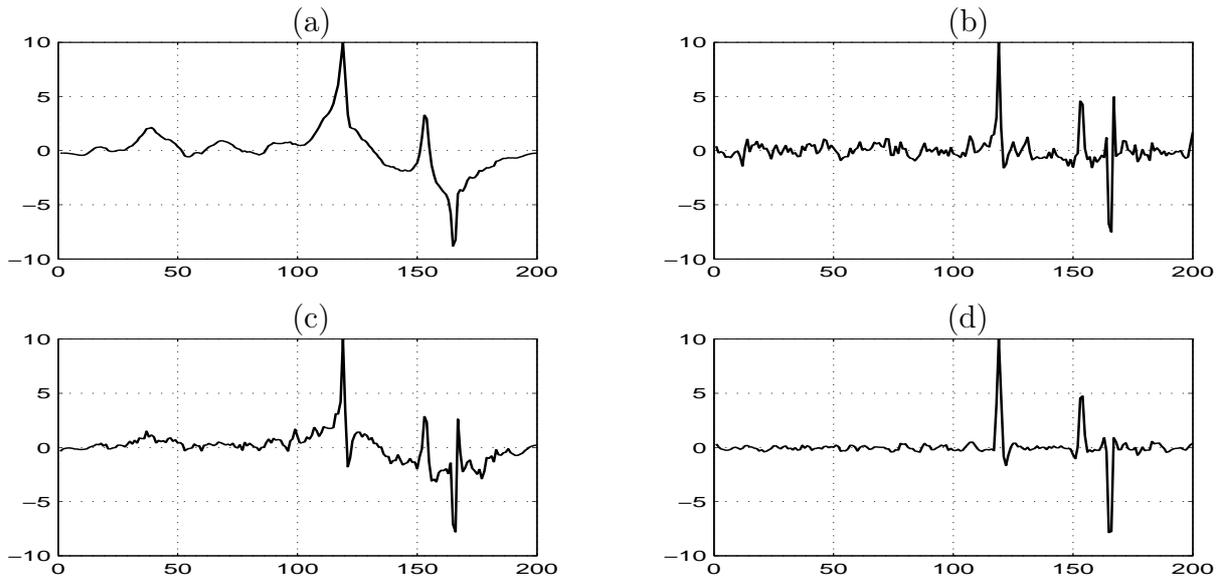


FIG. 3.11 – Déconvolution de données d’explosions sous-marines : (a) Observation, (b) égalisation spectrale, (c) méthode de Durbin, (d) algorithme FBD avec $\mu = 0.01$, $\lambda_1 = 0.1$, $\lambda_2 = 0.1$ (valeur donnée par 3.20) et $p = 3$.

un *a priori* sous-gaussien pour la distribution de $|G(\nu)|$. Nous remarquons que les trois méthodes donnent une bonne estimation de la position temporelle des trois réflecteurs. Pour mesurer plus précisément les performances et avoir une analyse plus objective, le tableau 3.3 mesure trois critères. Comme les algorithmes sont basés sur la blancheur des signaux, nous proposons de mesurer la blancheur de la sortie par $\tilde{I}(Y)$ (2.8). Ensuite, nous mesurons la largeur à 10% du premier réflecteur *i.e.* le nombre d’échantillons entre les deux échantillons ayant une valeur égale à 10% de la valeur maximale du réflecteur. Finalement, le dernier critère mesure l’écart-type du bruit sur les 100 premiers échantillons ne comportant pas de réflecteurs. Afin de faciliter la comparaison, nous avons normalisé tous les signaux pour obtenir un premier réflecteur d’amplitude 10.

TAB. 3.3 – Comparaison de la déconvolution des explosions sous-marines de la figure 3.11.

	Observation	Durbin	Egalisation spectrale	FBD
Écart-type du bruit	0.63	0.39	0.5	0.19
Mesure de blancheur $\tilde{I}(Y)$	2.1	0.32	-0.5	-0.31
Largeur du premier réflecteur	21	14	5	3

Clairement, l’algorithme FBD (figure 3.11(d)) donne la réflectivité avec les pics les mieux identifiés et un niveau de bruit très faible. Concernant la mesure de blancheur $\tilde{I}(Y)$, l’égalisation spectrale donne de meilleures performances que l’algorithme FBD. En fait, $\tilde{I}(Y)$ mesure la blancheur de $y = \hat{r} = r + n'$, où r est le signal de réflectivité désiré. Si l’entrée $r(t)$ n’est pas parfaitement blanche et/ou si les observations $d(t)$ sont bruitées, la meilleure

estimation de la sortie (en terme de blancheur) $y = g \star [w \star r + n]$ est la plus iid possible. Mais même avec une estimation parfaite du filtre déconvolveur $g = w^{-1}$, nous obtenons la sortie $y = r + g \star n$, *i.e.* une version bruitée de la réflectivité théorique. La régularisation faite par rapport au bruit dans l'algorithme FBD tend à limiter le terme $g \star n$ pour permettre une bonne estimation de $r(t)$. Mais dans le même temps, cette régularisation se fait au détriment de la blancheur de la sortie estimée $y(t)$. En fait, le critère minimisé par l'algorithme FBD n'est pas $\tilde{I}(Y)$ mais la fonction coût régularisée $J(\mathbf{G})$ de (3.5), et donc les minima des deux critères $\tilde{I}(Y)$ et $J(\mathbf{G})$ sont différents. Cette remarque confirme que le critère de blancheur pris sans régularisation ne donne pas la meilleure estimation de la réflectivité, sauf si les données ne sont pas bruitées. En d'autres termes, seuls les algorithmes prenant en compte le bruit additif peuvent donner des algorithmes efficaces pour les données réelles. De plus, dans la figure 3.11(b) et (c), nous observons l'apparition de pics secondaires juste après les réflecteurs pour les méthodes de Durbin et de l'égalisation spectrale, qui n'apparaissent pas avec l'algorithme FBD.

En conclusion, l'algorithme FBD donne le meilleur compromis entre la qualité de la déconvolution et l'amplification du bruit. Avec ces données, nous avons pu mettre en évidence l'apport de la régularisation vis-à-vis de bruit. Cet algorithme peut estimer aisément la phase nulle ou quasi nulle de l'ondelette par l'intermédiaire de l'ensemble des statistiques d'ordre supérieur.

3.4 Conclusions

Dans ce chapitre, nous avons vu l'importance de prendre en compte le bruit additif dans la construction du critère de déconvolution. Même l'utilisation de l'ensemble des statistiques d'ordre supérieur ne permet pas de s'affranchir du problème du bruit. En effet, notre étude détaillée de l'algorithme MAMV en présence de bruit gaussien montre que les performances de cet algorithme tendent vers celles de la méthode de Yule-Walker (statistiques d'ordre 2). Pour prendre en compte le bruit additif, nous proposons une approche fréquentielle inspirée par le compromis réalisable entre la déconvolution et l'amplification du bruit dans le domaine fréquentiel avec le filtrage de Wiener ou l'égalisation spectrale. La fonction coût comporte trois parties : (i) le taux d'information mutuelle de la sortie (ii) une contrainte de continuité sur le spectre du filtre déconvolveur (iii) une régularisation vis-à-vis du bruit additif. Des méthodes ont été décrites pour le réglage des paramètres. Des simulations permettent de mettre en évidence la capacité de l'algorithme FBD à estimer des filtres à phase quelconque. Ensuite, des tests sur des données réelles d'explosions sous-marines et séismo-volcaniques montrent la capacité de notre algorithme à s'adapter à des données très différentes grâce à son approche fréquentielle et aussi à limiter l'amplification du bruit. En effet, l'algorithme FBD a de bonnes performances aussi bien sur des filtres directs AR que MA. Cet algorithme a fait l'objet d'un article [LMJ06] et de deux communications dans des conférences [LJ05, LMJ04].

Nous voyons lors de la mise en place des algorithmes de déconvolution aveugle que nous sommes gênés par la double hypothèse de blancheur sur la réflectivité et le bruit. Ainsi, le filtre déconvolveur tend à blanchir le terme $g \star w \star r(t)$ mais dans le même temps il

colore le terme $g \star n(t)$. En définitive, pour augmenter la robustesse au bruit, nous devons tenter de relâcher la contrainte de blancheur sur la réflectivité et nous tourner vers d'autres hypothèses comme la sur-gaussianté. De plus, la blancheur n'est pas totalement vérifiée par les réflectivités réelles, qui ont un spectre dit bleu, avec des hautes fréquences plus importantes que les basses fréquences.

Chapitre 4

Déconvolution par non-gaussianité

Sommaire

4.1	Déconvolution par non-gaussianité	104
4.1.1	Mesure de non-gaussianité : la néguentropie	104
4.1.2	Déconvolution par maximisation de la néguentropie : algorithme MANege	105
4.1.3	Résultats de simulations	107
4.2	MANege : un contexte général pour des méthodes existantes .	109
4.2.1	Rappel sur les lois gaussiennes généralisées	110
4.2.2	Déconvolution par maximisation du kurtosis : algorithme MED .	112
4.2.3	Déconvolution par variation de norme : algorithme VARIMAX .	115
4.2.4	Transformation exponentielle	118
4.2.5	Algorithme de Godfrey	120
4.3	Contraintes d'identifiabilité sur les critères	121
4.4	Concavité et estimation non biaisée	122
4.4.1	Contraintes théoriques	122
4.4.2	Cas de l'algorithme MANege	124
4.4.3	Cas des algorithmes MED et VARIMAX	125
4.5	Variance asymptotique des algorithmes	129
4.5.1	Calcul théorique	130
4.5.2	Borne minimale	131
4.5.3	Courbes de performances	132
4.6	Comparaison des algorithmes	135
4.6.1	Fonction score	135
4.6.2	Résultats de simulations	137
4.7	Résumé et conclusions	138

Dans les deux chapitres précédents, nous avons utilisé l'hypothèse de blancheur pour concevoir un algorithme de déconvolution aveugle le plus robuste possible par rapport au bruit additif et qui permette l'estimation de la phase du filtre. Nous avons noté le problème lié à la prépondérance accordée aux hypothèses de blancheur du signal de réflectivité et de blancheur du bruit additif. Dans ce chapitre, nous proposons de nous intéresser à la déconvolution aveugle basée sur l'hypothèse de non-gaussianité de la réflectivité. Nous savons d'après le théorème limite centrale, que la convolution d'un processus blanc non gaussien par un filtre augmente la gaussianité du signal. La figure 4.1 illustre cette remarque.

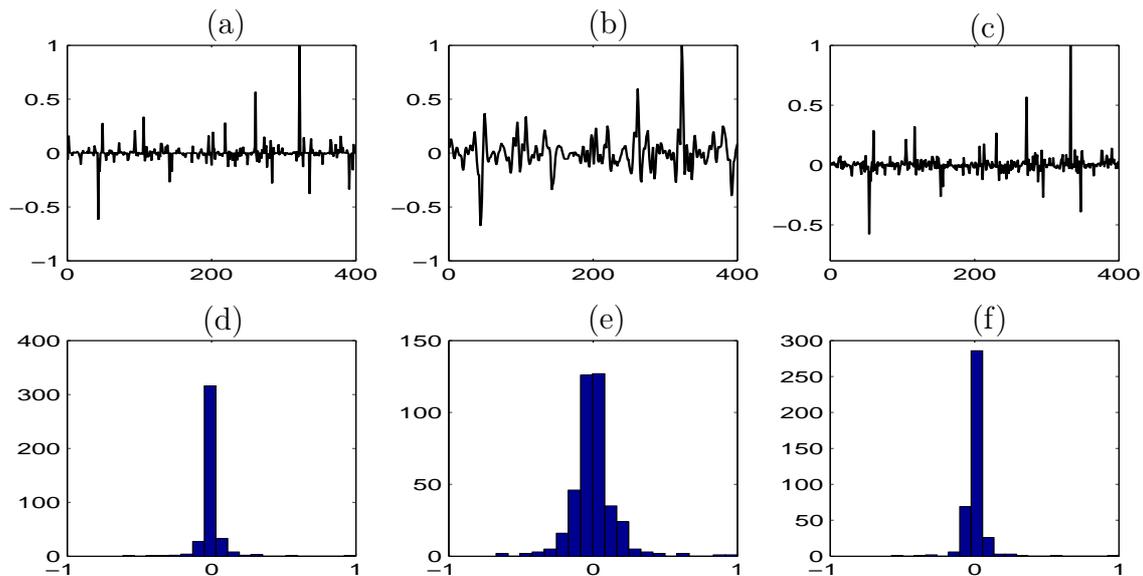


FIG. 4.1 – Illustration de l'hypothèse de sur-gaussianité : (a) séquence de réflectivité simulée, et son histogramme (d) ; (b) observation simulée après convolution de la réflectivité (a) et d'une ondelette, et l'histogramme de l'observation (e) ; (c) réflectivité estimée et son histogramme (f).

Sur cette figure 4.1(a), nous avons une réflectivité simulée avec un modèle sur-gaussien, la majorité de ces valeurs sont faibles en amplitude et seules certaines valeurs sont significatives. Ces dernières représentant les réflecteurs. L'histogramme de la réflectivité en figure 4.1(d) montre effectivement une distribution sur-gaussienne avec un grand nombre d'échantillons de valeurs quasi-nulles. Après convolution de la réflectivité par une ondelette, les observations (figure 4.1(b)) admettent une distribution (figure 4.1(e)) plus gaussienne que la réflectivité. Ainsi, l'idée de base des algorithmes proposés dans ce chapitre est d'ajuster le filtre déconvolveur $g(t)$ suivant l'hypothèse de sur-gaussianité de la sortie à la place de la blancheur dans les chapitres 2 et 3. En fait, la sur-gaussianité est une alternative à la blancheur pour la sélection des statistiques d'ordre supérieur utilisées par l'algorithme.

Dans une première section, nous présenterons la néguentopie, qui est une mesure de déviation à la gaussienne. Ensuite, nous détaillerons son utilisation dans un algorithme (MANege) de déconvolution aveugle dans le domaine temporel. Nous poursuivrons par des résultats sur des données synthétiques. Dans la seconde section, nous ferons le parallèle entre les méthodes existantes et l'algorithme proposé précédemment. En effet, de nombreuses

méthodes fondées sur la minimisation de l'entropie du signal de sortie ou des critères de parcimonie ont été proposées [Cla77, Wig78, Gra79, OU79, Wig85, SVC94]. En réalité, nous verrons que toutes ces méthodes sont des cas particuliers de l'algorithme MANeGe maximisant la négentropie avec pour chaque méthode un choix différent de distribution *a priori* pour la sortie. Les sections suivantes concerneront les performances des algorithmes avec une étude comparative. Tout d'abord, nous nous intéresserons aux contraintes suffisantes à l'identifiabilité du système, puis la section 4.4 permettra de justifier la concavité des critères nécessaires à la stabilité des algorithmes. Puis, nous étudierons les conditions pour avoir une estimation non biaisée. Dans la section 4.5, nous nous intéresserons aux performances asymptotiques théoriques avec la matrice de covariance de l'estimée des différents algorithmes en fonction de la distribution de la réflectivité. Nous poursuivrons dans la section 4.6 par la présentation de résultats de simulations pour illustrer les résultats théoriques précédemment présentés. Enfin, la dernière section fera la synthèse, sous forme d'un tableau récapitulatif, des méthodes avec les avantages et les inconvénients de chacune.

4.1 Déconvolution par non-gaussianité

Dans cette section, nous allons présenter une mesure de déviation à la gaussienne : la négentropie. Ensuite, on proposera l'algorithme MANeGe de déconvolution aveugle s'appuyant sur la maximisation de la négentropie. Nous finirons la section avec des résultats de cet algorithme sur des signaux simulés.

4.1.1 Mesure de non-gaussianité : la négentropie

Pour caractériser la non-gaussianité de la sortie de la déconvolution, nous devons utiliser des outils statistiques. La première mesure est bien évidemment le kurtosis, comme nous l'avons mentionné dans la sous-section 1.2.1. Mais, pour avoir une caractérisation plus large, nous pouvons utiliser la divergence de Kullback [CT91, LAC97] qui mesure l'écart entre deux densités de probabilité $p_X(x)$ et $p_Z(z)$ par :

$$K(p_X, p_Z) = \int p_X(u) \log \frac{p_X(u)}{p_Z(u)} du. \quad (4.1)$$

La divergence de Kullback est positive si $p_X \neq p_Z$ et est nulle si et seulement si $p_X = p_Z$. Cette mesure porte le nom de divergence et non distance, car elle ne vérifie pas les propriétés mathématiques d'une distance (par exemple la symétrie). Nous pouvons lier cette mesure à l'entropie utilisée dans les chapitres précédents, notamment pour la définition du taux d'information mutuelle. L'entropie d'un processus X étant définie par :

$$H(X) = -E[\log p_X(x)] = - \int p_X(x) \log p_X(x) dx, \quad (4.2)$$

la divergence de Kullback entre p_X et p_Z s'écrit donc :

$$K(p_X, p_Z) = -H(X) - \int p_X(u) \log p_Z(u) du. \quad (4.3)$$

Nous souhaitons mesurer la déviation à la gaussienne du processus Y . Ainsi, nous définissons le processus Y_G comme le processus gaussien ayant la même variance σ et la même valeur moyenne (ici elle est supposée nulle) que Y . La néguentropie est alors définie comme la divergence de Kullback entre Y et Y_G :

$$H^-(Y) = K(p_Y, p_{Y_G}) = -H(Y) - \int p_Y(u) \log p_{Y_G}(u) du. \quad (4.4)$$

Nous pouvons développer le dernier terme de la néguentropie en utilisant l'expression de la loi de Y_G :

$$\int p_Y(u) \log p_{Y_G}(u) du = E \left[\log \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{Y^2}{2\sigma^2}} \right] = -\frac{1}{2} \log(2\pi e\sigma^2), \quad (4.5)$$

où σ représente la variance de Y . Finalement la néguentropie de Y s'exprime par :

$$H^-(Y) = \frac{1}{2} \log(2\pi e\sigma^2) - H(Y). \quad (4.6)$$

La néguentropie est toujours positive, et est nulle si et seulement si le processus Y est gaussien. Ainsi, la néguentropie est une mesure de divergence à la gaussienne [CT91]. Nous proposons de l'utiliser pour résoudre le problème de déconvolution aveugle. En effet, pour estimer une sortie $y(t)$ qui soit similaire à la réflectivité $r(t)$, nous pouvons ajuster le filtre déconvolveur $g(t)$ pour rendre le signal de sortie $y(t)$ le moins gaussien possible. Nous proposons dans la suite l'algorithme MANege (Moving Average Negentropy) estimant le filtre déconvolveur en maximisant la néguentropie, soit :

$$g^{MANege}(t) = \arg \max_{g(t)} H^-(Y). \quad (4.7)$$

Comme dans le chapitre 2, pour parvenir au développement d'un algorithme, dans l'expression (4.6) de la néguentropie, nous supposons que l'entropie du processus Y est identique à l'entropie d'une réalisation $y(\tau)$, soit $H(Y) = H(y(\tau))$. Ainsi, la néguentropie s'écrit :

$$H^-(Y) = \frac{1}{2} \log(2\pi e\sigma^2) - H(y(\tau)). \quad (4.8)$$

4.1.2 Déconvolution par maximisation de la néguentropie : algorithme MANege

Nous souhaitons maximiser la néguentropie par rapport au filtre déconvolveur $g(t)$. Nous devons calculer le gradient de la néguentropie $H^-(Y)$ de Y par rapport à la réponse impulsionnelle du filtre déconvolveur $g(t)$, celle-ci comportera un nombre fini de coefficients (soit un modèle MA pour le filtre déconvolveur comme pour la méthode MAMV du chapitre 2). Pour la dérivation du premier terme de (4.8), nous avons :

$$\frac{\partial}{\partial g(t)} \frac{1}{2} \log(2\pi e\sigma^2) = \frac{1}{2\sigma^2} \frac{\partial \sigma^2}{\partial g(t)}. \quad (4.9)$$

En écrivant que $\sigma^2 = E[y(\tau)^2]$, nous avons :

$$\frac{\partial}{\partial g(t)} \frac{1}{2} \log(2\pi e\sigma^2) = \frac{1}{\sigma^2} E \left[y(\tau) \frac{\partial y(\tau)}{\partial g(t)} \right]. \quad (4.10)$$

Or, nous avons $\frac{\partial y(\tau)}{\partial g(t)} = d(\tau - t)$, ainsi, nous obtenons :

$$\frac{\partial}{\partial g(t)} \frac{1}{2} \log(2\pi e\sigma^2) = \frac{1}{\sigma^2} E [y(\tau)d(\tau - t)]. \quad (4.11)$$

Pour le calcul du gradient de l'entropie $H(y(\tau))$ présent dans l'équation (4.8), nous répétons la démarche mise en place pour l'algorithme MAMV (section 2.2) qui nous avait permis de parvenir à l'expression (2.14). Finalement, en utilisant (4.11) et (2.14), la dérivation de la néguentropie par rapport au coefficient de la réponse impulsionnelle $g(t)$ est :

$$\frac{\partial}{\partial g(t)} H^-(Y) = \frac{1}{\sigma^2} E [y(\tau)d(\tau - t)] - E[\psi_Y(y(\tau))d(\tau - t)], \quad (4.12)$$

où ψ_Y est la fonction score de Y . À partir de (4.12), nous pourrions utiliser une méthode de gradient classique pour la maximisation de la néguentropie. Mais, comme pour l'algorithme MAMV, nous mettons en place une méthode de gradient relatif (voir section 2.2.2) pour avoir un algorithme équivariant, *i.e.* dont les performances de convergence ne dépendent pas de la forme de l'ondelette. La mise à jour du filtre déconvolveur $g(t)$ se fera par :

$$g(t) \leftarrow g(t) + \mu \frac{\partial H^-(Y)}{\partial g(t)} \star g^- \star g(t), \quad (4.13)$$

où μ est le pas de gradient (réel positif) et $g^-(t) = g(-t)$ est le filtre $g(t)$ retourné. Ainsi, nous avons :

$$g(t) \leftarrow g(t) + \mu \left\{ -E[\psi_Y(y(\tau))y(\tau - t)] + \frac{1}{\sigma^2} E [y(\tau)y(\tau - t)] \right\} \star g(t). \quad (4.14)$$

Il est intéressant de regrouper les deux espérances pour obtenir la mise à jour sous la forme suivante :

$$g(t) \leftarrow g(t) + \mu E \left[\left(-\psi_Y(y(\tau)) + \frac{y(\tau)}{\sigma^2} \right) y(\tau - t) \right] \star g(t). \quad (4.15)$$

L'expression précédente fait apparaître dans la parenthèse la différence entre la fonction score $\psi_Y(y)$ du processus Y et celle du processus gaussien de même variance qui est y/σ^2 . Ainsi, dans la relation précédente, nous calculons la corrélation entre $y(\tau)$ et une fonction $\phi_Y(y(\tau))$ non linéaire de $y(\tau)$ que nous définissons par :

$$\phi_Y(y(\tau)) = \psi_Y(y(\tau)) - \frac{y(\tau)}{\sigma^2}. \quad (4.16)$$

Les points stationnaires du critère de la néguentropie vérifient $E[\phi_Y(y(\tau))y(\tau - t)] = \delta(t)$, où $\delta(t)$ est le filtre identité. Cette égalité est équivalente à une décorrélacion de $y(\tau)$ et $\phi_Y(y(\tau))$.

Dans un premier temps, en comparaison à l'algorithme MAMV, la seule différence des algorithmes se situe dans le choix de la fonction non linéaire ϕ_Y qui pour l'algorithme

MAMV est la fonction score ψ_Y . En fait, le terme soustrait $\frac{y(\tau)}{\sigma^2}$ à la fonction score va réduire la contribution des statistiques d'ordre 2 dans l'algorithme de négentropie par rapport à l'algorithme MAMV. Dans la section 3.1, nous avons montré que la partie linéaire de la fonction score rendait les performances de l'algorithme MAMV identiques à celles des méthodes à l'ordre 2. Ainsi, avec cette maximisation de la négentropie, nous pouvons espérer de meilleures performances en terme de robustesse au bruit gaussien en comparaison à l'algorithme MAMV.

L'algorithme MANege (Moving Average Negentropy) maximisant la négentropie dans le domaine temporel comporte les étapes suivantes :

1. Initialisation de $g(t)$ et $y(t)$,
2. Estimation de la fonction score ψ_Y ,
3. Mise à jour du filtre déconvolveur par (4.15),
4. Calcul du signal déconvolué $y(t)$,
5. Normalisation de $y(t)$ et $g(t)$.

Nous itérons les étapes 2 à 5 jusqu'à convergence. L'étape de normalisation permet de lever l'indétermination d'échelle sur $y(t)$ et $g(t)$. En général, la puissance du signal de sortie est fixée unitaire.

Comme pour les algorithmes BFD et MAMV, plusieurs stratégies sont possibles pour le choix de l'estimateur de la fonction score (voir la section 2.4). Nous pouvons choisir dans un cadre général un estimateur à noyaux. Dans la sous-section suivante, comme pour les algorithmes MAMV et FBD nous mettrons en évidence la possibilité d'estimer des filtres à phase non-minimale avec l'algorithme MANege, grâce à l'utilisation des statistiques d'ordre supérieur à 2.

4.1.3 Résultats de simulations

Pour illustrer l'algorithme MANege, nous utilisons un cadre de simulation assez similaire aux figures 2.1 et 2.2 de l'algorithme MAMV ou aux figures 3.7 et 3.8 de l'algorithme FBD. Nous choisissons :

- une réflectivité Bernoulli-laplacienne de 400 échantillons avec 20 réflecteurs,
- une ondelette AR de coefficients $[1, 0.4, 0.5, 0.45, 0.4, 0.1]$,
- une ondelette à phase minimale pour la figure 2.1,
- une ondelette à phase nulle pour la figure 2.2,
- un bruit additif gaussien blanc avec un RSB de 18dB,

puis, nous déconvoluons les observations grâce à l'algorithme MANege avec un filtre déconvolveur de 30 coefficients.

Dans les figures 4.2 et 4.3, nous traçons en (a) la réflectivité et en (b) les observations qui résultent de la convolution entre la réflectivité et une ondelette à phase nulle ou minimale et l'ajout du bruit gaussien. Pour analyser la qualité de la déconvolution, nous avons en (c) la réflectivité estimée et nous comparons en (d) (resp. (e)) la phase (resp. le module) du filtre estimé par l'algorithme MANege à sa valeur théorique.

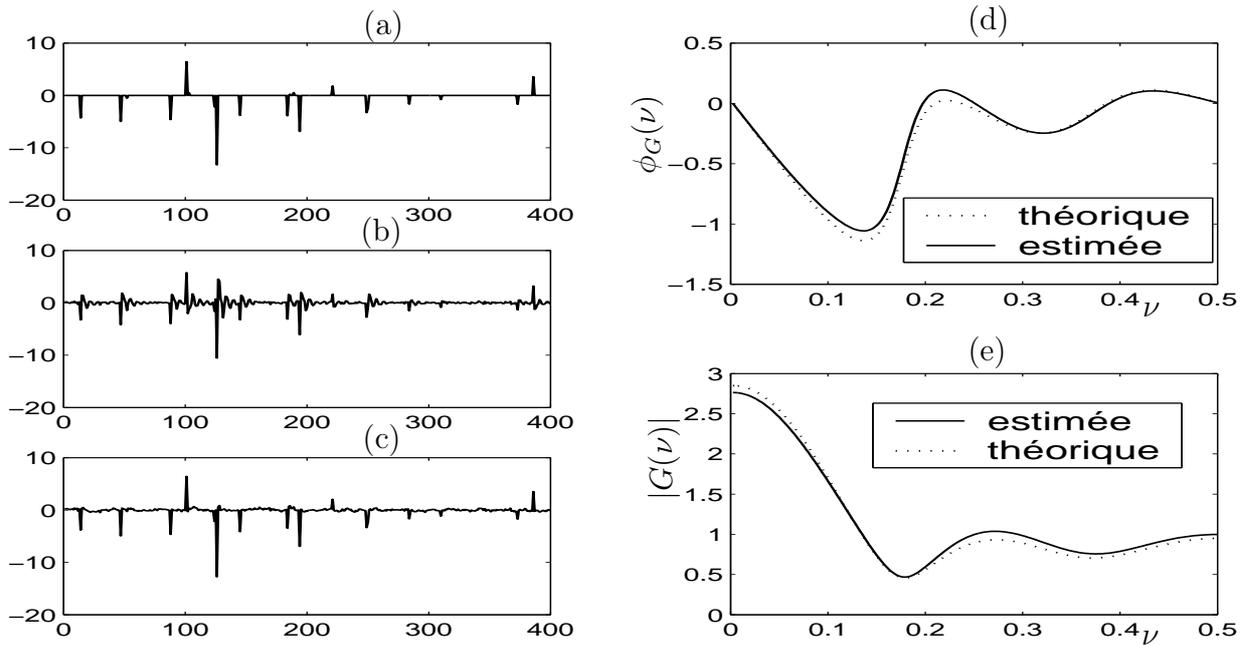


FIG. 4.2 – Déconvolution de données sismiques synthétiques avec une ondelette AR à phase minimale : (a) réflectivité Bernoulli-laplacienne (b) observation avec ondelette AR à **phase minimale**, et RSB= 18dB, (c) réflectivité estimée par l’algorithme MANeige, (d) phase théorique (en pointillé) et phase (rad)estimée du filtre déconvolue (trait plein) (e) idem que (d) pour le module du filtre.

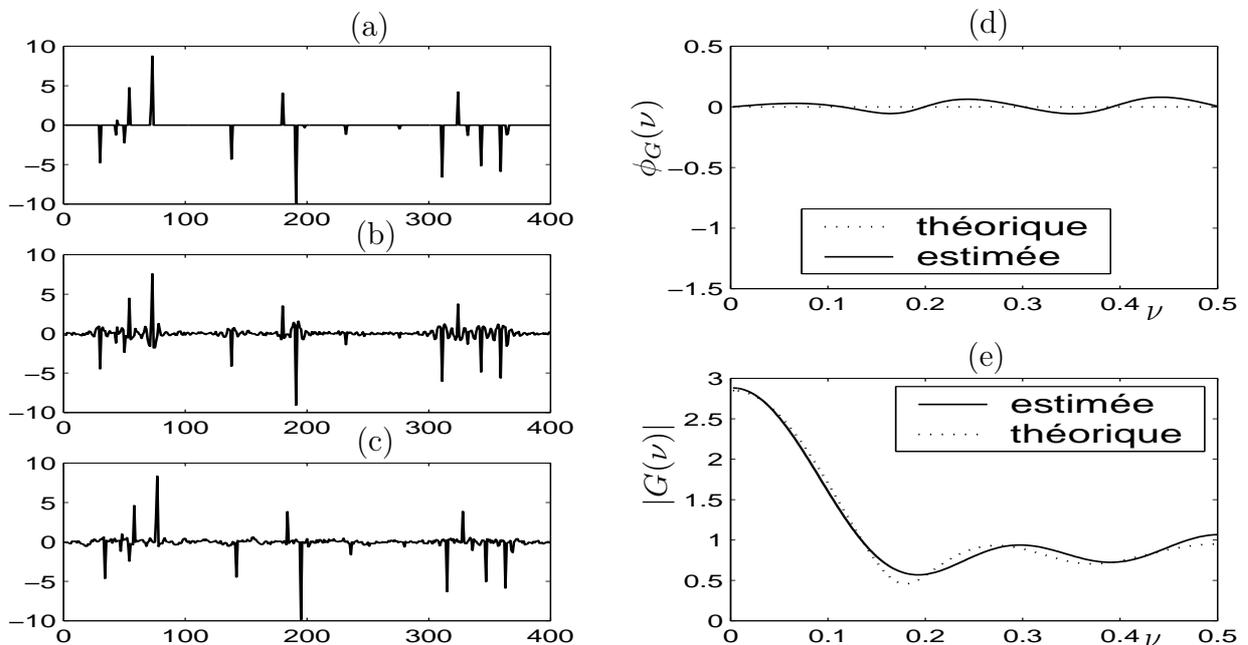


FIG. 4.3 – Idem figure 4.2 avec ondelette AR à phase nulle.

En analysant sur les figures 4.2 (ondelette AR à phase minimale) et 4.3 (ondelette AR à phase nulle) la réflectivité simulée en (a) et la réflectivité estimée en (c), nous notons une très bonne estimation. Les légères variations sur la réflectivité pour les parties entre les réflecteurs sont liées à la présence du bruit sur les données. Ensuite, avec les tracés de la phase en (d) et du module en (e) du filtre estimé et du filtre théorique, nous remarquons une très bonne estimation du filtre, même si nous avons très largement surestimé le nombre de coefficients du filtre déconvolveur (30 au lieu de 7). Cette expérience montre la possibilité d'estimer des filtres à phase quelconque par maximisation de la néguentropie. Cette estimation est rendue possible par l'utilisation de l'ensemble des statistiques d'ordre supérieur par le biais de la fonction non linéaire ϕ_Y de l'équation (4.16) qui intervient dans la corrélation pour le calcul du gradient (4.15).

Dans la section suivante, nous proposons de montrer que de nombreuses méthodes existantes sont des cas particuliers de cet algorithme MANege.

4.2 MANege : un contexte général pour des méthodes existantes

Un grand nombre de méthodes a été proposé en avançant des arguments de sur-gaussianité ou de parcimonie. Nous proposons de détailler ces différentes méthodes et de montrer qu'elles peuvent être vues comme des cas particuliers de l'algorithme MANege, avec un choix particulier de la fonction score ψ_Y . Le choix de la fonction score est lié en fait à un choix de distribution *a priori* pour la sortie $y(t)$. Ce travail a déjà été abordé, mais de façon légèrement différente par Godfrey [God78], par Donoho [Don81], et enfin par Walden [Wal85].

En effet, Godfrey et Walden proposaient comme contexte le rapport de vraisemblance. Cette méthode est classique pour le test d'hypothèse binaire. Dans ce cadre, nous souhaitons savoir si les échantillons $r(1), \dots, r(T)$ du signal $r(t)$ sont tirés de façon iid selon une hypothèse H_1 ou H_0 . Sous l'hypothèse H_0 , nous supposons que la distribution est $p(r|H_0)$ de variance σ . De même sous l'hypothèse H_1 , la distribution est $p(r|H_1)$ avec la même variance. La décision pour l'hypothèse vérifiée est prise grâce à un test du rapport de vraisemblance qui s'écrit dans le cas d'échantillons iid :

$$\frac{\prod_{t=1}^T p(r(t)|H_1)}{\prod_{t=1}^T p(r(t)|H_0)} \underset{H_0}{\overset{H_1}{\geq}} \eta, \quad (4.17)$$

où η est un seuil qui peut être déterminé selon différentes stratégies. Ensuite, en prenant le log de l'expression précédente, nous obtenons une différence de log-vraisemblance :

$$\sum_{t=1}^T \log(p(r(t)|H_1)) - \sum_{t=1}^T \log(p(r(t)|H_0)) \underset{H_0}{\overset{H_1}{\geq}} \log \eta. \quad (4.18)$$

Pour faire un parallèle avec la néguentropie, nous pouvons remarquer que les termes de l'équation (4.18) s'écrivent comme des estimations d'entropie sous les hypothèses H_1 et

H_0 :

$$\frac{1}{T} \sum_{t=1}^T \log(p(r(t)|H_1)) = \widehat{E}[\log p(r|H_1)] = -\widehat{H}(R|H_1), \quad (4.19)$$

de même pour H_0 . Ainsi, le test d'hypothèse (4.18) peut s'écrire comme une différence d'entropies :

$$-\widehat{H}(R|H_1) + \widehat{H}(R|H_0) \underset{H_0}{\overset{H_1}{\geq}} \frac{1}{T} \log \eta. \quad (4.20)$$

Pour lier cette relation à l'algorithme MANege, il nous suffit de prendre pour H_0 l'hypothèse d'une distribution gaussienne et pour H_1 une autre distribution de même variance et caractéristique de la réflectivité recherchée. Maximiser la distance entre la distribution gaussienne et celle du processus R est équivalent à faire en sorte que l'hypothèse H_1 soit prépondérante par rapport à l'hypothèse H_0 . Ainsi, les deux approches : maximisation de la néguentropie et avec le rapport de vraisemblance sont équivalentes. Mais notre présentation de la néguentropie est mieux adaptée au cadre de la déconvolution. Si nous choisissons une hypothèse H_1 opposée à la vraie distribution de la réflectivité, par exemple, sous-gaussienne pour H_1 alors que la réflectivité est sur-gaussienne. Alors, nous noterons qu'il faut que l'hypothèse H_0 soit plus probable que H_1 . En effet, l'hypothèse sous-gaussienne est moins probable que l'hypothèse gaussienne. Par conséquent, il faut minimiser à la place de maximiser l'entropie sous H_1 . Ce problème apparaîtra de nouveau pour l'algorithme MED qui sera détaillé dans la sous-section 4.2.2.

4.2.1 Rappel sur les lois gaussiennes généralisées

Avant de présenter les différentes méthodes développées sur l'hypothèse de sur-gaussianité, nous devons définir les lois gaussiennes généralisées (GG). Une variable aléatoire X est distribuée suivant une loi de probabilité gaussienne généralisée centrée de paramètre (α, β) si :

$$p_X(x) = \frac{\alpha}{2\beta\Gamma(\frac{1}{\alpha})} e^{-|\frac{x}{\beta}|^\alpha}, \quad (4.21)$$

où $\Gamma(u)$ est la fonction Gamma définie par :

$$\Gamma(u) = \int_0^{+\infty} t^{u-1} e^{-t} dt. \quad (4.22)$$

α est appelé l'exposant de la loi GG ou le coefficient de gaussianité. Nous notons les cas particuliers : $\alpha = 1$ correspond à une distribution laplacienne et $\alpha = 2$ à une loi gaussienne. Si $\alpha < 2$, nous avons une loi sur-gaussienne et si $\alpha > 2$ une loi sous-gaussienne. Le paramètre β est un facteur d'échelle, comme la variance dans le cas gaussien. Les moments d'ordre r (r pair car les moments d'ordre impair sont nuls) d'une distribution gaussienne généralisée sont :

$$\mu_{rX} = \int_{-\infty}^{+\infty} x^r p_X(x) dx = \beta^r \frac{\Gamma(\frac{r+1}{\alpha})}{\Gamma(\frac{1}{\alpha})}. \quad (4.23)$$

Pour $r = 2$, nous trouvons l'expression de la variance :

$$\sigma^2 = \beta^2 \frac{\Gamma(\frac{3}{\alpha})}{\Gamma(\frac{1}{\alpha})}. \quad (4.24)$$

De manière générale, pour tout réel $r > -1$ nous pouvons montrer que¹ :

$$\mu_{r|X|} = E[|X|^r] = \beta^r \frac{\Gamma(\frac{r+1}{\alpha})}{\Gamma(\frac{1}{\alpha})}. \quad (4.25)$$

Nous souhaitons écrire la néguentropie définie par la relation (4.8) pour un processus X distribué selon une loi gaussienne généralisée de paramètre (α, β) . Tout d'abord, nous savons que son entropie s'écrit à l'aide de l'expression de la ddp (4.21) :

$$H(X) = -E[\log p_X(x)] = -\log \frac{\alpha}{2\beta\Gamma(\frac{1}{\alpha})} + E\left[\left|\frac{X}{\beta}\right|^\alpha\right]. \quad (4.26)$$

Avec la définition (4.8), la néguentropie du processus GG X s'écrit :

$$H^-(X) = \frac{1}{2} \log(2\pi e\sigma^2) - H(X) = \frac{1}{2} \log(2\pi e\sigma^2) + \log \frac{\alpha}{2\beta\Gamma(\frac{1}{\alpha})} - E\left[\left|\frac{X}{\beta}\right|^\alpha\right]. \quad (4.27)$$

Pour finir les simplifications, nous utilisons la relation entre la variance et le paramètre β donnée en (4.24). Ainsi, la néguentropie s'exprime par :

$$H^-(X) = \frac{1}{2} \log\left(\frac{\pi e \alpha^2 \Gamma(\frac{3}{\alpha})}{2\Gamma(\frac{1}{\alpha})^3}\right) - \left(\frac{\Gamma(\frac{3}{\alpha})}{\Gamma(\frac{1}{\alpha})}\right)^{\alpha/2} E\left[\frac{|X|^\alpha}{(\sigma^2)^{\alpha/2}}\right]. \quad (4.28)$$

Avec l'expression précédente de la néguentropie pour un processus suivant une ddp gaussienne généralisée de coefficient (α, β) , la maximisation de la néguentropie de la sortie avec une loi *a priori* GG de paramètre α est équivalente à la minimisation du terme suivant :

$$V_\alpha(X) = E\left[\frac{|X|^\alpha}{(\sigma^2)^{\alpha/2}}\right] = \frac{E[|X|^\alpha]}{E[X^2]^{\alpha/2}}. \quad (4.29)$$

Dans la suite, nous proposons de détailler les différentes méthodes proposées se basant sur des hypothèses de sur-gaussianité (MED [Wig78], VARIMAX [Cla77, Gra79], avec une transformation exponentielle [OU79] et l'algorithme de Godfrey [God78]). Nous tenterons de lier chacune de ces méthodes à l'algorithme MANege. Souvent les algorithmes sont présentés comme des algorithmes SIMO (Single Input Multiple Output). En effet, l'ondelette est supposée constante et nous cherchons le filtre déconvolveur en tirant parti de l'information de toutes les traces. Nous disposons de N traces $d_i(t)$ pour $i = 1 \dots N$. La réflectivité

¹La définition de moment est généralisée à r non entier pour les besoins futurs. La contrainte $r > -1$ est liée à l'existence de l'espérance mathématique. Par abus de langage, pour $\mu_{r|X|}$ nous parlerons de moment d'ordre r de $|X|$

est estimée avec le même filtre g sur toutes les traces par $y_i(t) = g \star d_i(t)$. Dans tous les algorithmes, les auteurs réalisent simplement la moyenne du critère sur toutes les traces. Par souci de clarté, nous présenterons le cas de système de déconvolution SISO. Dans la section 4.7, le tableau 4.2 résume l'ensemble des méthodes présentées dans la suite de cette section en listant les avantages et inconvénients de chacune.

4.2.2 Déconvolution par maximisation du kurtosis : algorithme MED

Wiggins [Wig78] propose d'ajuster le filtre déconvolveur pour maximiser \mathcal{K}_{4Y} le kurtosis de la sortie $y(\tau)$. Le kurtosis est défini par (1.16) au paragraphe 1.2.1 comme le rapport du cumulants d'ordre 4 et du cumulants d'ordre 2 au carré. Il est possible de l'estimer par :

$$\widehat{\mathcal{K}}_{4Y} = \frac{\frac{1}{T} \sum_{\tau=1}^T y(\tau)^4}{\left(\frac{1}{T} \sum_{\tau=1}^T y(\tau)^2\right)^2} - 3. \quad (4.30)$$

Le kurtosis est nul pour une distribution gaussienne, négatif pour une distribution sous-gaussienne et positif pour une distribution sur-gaussienne. En sismique, la réflectivité étant généralement sur-gaussienne, Wiggins propose de maximiser le kurtosis pour rendre la sortie $y(\tau)$ la plus sur-gaussienne possible. Ensuite, deux voies légèrement différentes sont envisageables pour parvenir à un algorithme de déconvolution permettant l'estimation du filtre déconvolveur $g(t)$ par maximisation du kurtosis. La première consiste à dériver l'estimée du kurtosis $\widehat{\mathcal{K}}_{4Y}$ de l'équation (4.30) par rapport aux P coefficients de la réponse impulsionnelle du filtre déconvolveur $g(t)$. La seconde consiste à dériver l'expression théorique du kurtosis \mathcal{K}_{4Y} qui est donnée en (1.16), puis d'estimer le gradient théorique. Dans un souci de cohérence avec l'algorithme MANege présenté précédemment, nous préférons poursuivre suivant le second point de vue. En définitive, les deux approches sont identiques pour les critères simples comme celui du kurtosis. Nous noterons que Wiggins suit la première démarche.

Nous calculons le gradient du kurtosis \mathcal{K}_{4Y} de (1.16), par rapport aux coefficients de la réponse impulsionnelle du filtre déconvolveur $g(t)$. Nous avons pour $t = 0, \dots, P$:

$$\frac{\partial}{\partial g(t)} \mathcal{K}_{4Y} = \frac{\partial}{\partial g(t)} \frac{E[y(\tau)^4]}{E[y(\tau)^2]^2} = \frac{1}{E[y(\tau)^2]^2} E \left[4y(\tau)^3 \frac{\partial y(\tau)}{\partial g(t)} \right] - 2 \frac{E[y(\tau)^4]}{E[y(\tau)^2]^3} E \left[2y(\tau) \frac{\partial y(\tau)}{\partial g(t)} \right]. \quad (4.31)$$

Puis, en utilisant $\frac{\partial y(\tau)}{\partial g(t)} = d(\tau - t)$ et en reprenant les notations du chapitre 1 : $\mu_{2Y} = E[y(\tau)^2]$ et $\mu_{4Y} = E[y(\tau)^4]$, nous obtenons :

$$\frac{\partial}{\partial g(t)} \mathcal{K}_{4Y} = \frac{\partial}{\partial g(t)} \frac{E[y(\tau)^4]}{E[y(\tau)^2]^2} = \frac{4}{\mu_{2Y}^2} E \left[y(\tau)^3 d(\tau - t) \right] - \frac{4\mu_{4Y}}{\mu_{2Y}^3} E \left[y(\tau) d(\tau - t) \right]. \quad (4.32)$$

Dans l'équation précédente, nous voyons apparaître des fonctions de corrélation entre la sortie $y(\tau)$ et les données $d(\tau)$ mais aussi entre $y^3(\tau)$ et $d(\tau)$. Pour maximiser le kurtosis

de sortie plutôt que d'utiliser une méthode de gradient comme pour l'algorithme MANege, Wiggins propose une approche différente en cherchant la solution de :

$$\frac{\partial \mathcal{K}_{4Y}}{\partial g(t)} = 0. \quad (4.33)$$

Avec l'équation (4.32), l'annulation du gradient conduit à l'équation d'estimation des coefficients du filtre déconvolueur g suivante :

$$\sum_{t'=0}^P g(t') E [d(\tau - t') d(\tau - t)] = \frac{\mu_{2Y}}{\mu_{4Y}} E [y(\tau)^3 d(\tau - t)], \quad t = 0, \dots, P. \quad (4.34)$$

En réorganisant sous forme matricielle les équations qui proviennent de (4.34) pour les différentes valeurs de t , nous obtenons pour les coefficients $g(t)$ placés dans un vecteur $\mathbf{g} = [g(0), \dots, g(P)]^T$ l'équation d'estimation suivante :

$$\mathbf{R}_{DD} \mathbf{g} = \frac{\mu_{2Y}}{\mu_{4Y}} \mathbf{r}_{Y^3D}, \quad (4.35)$$

où \mathbf{r}_{Y^3D} est le vecteur de corrélation de $y(\tau)^3$ et des données $d(\tau)$ pour les P premiers retards. \mathbf{R}_{DD} est la matrice d'autocorrélation des données de taille $(P + 1) \times (P + 1)$. La matrice \mathbf{R}_{DD} a une structure de type Toeplitz comme dans le cas de l'algorithme de Yule-Walker (de déconvolution prédictive présentée dans le paragraphe 1.4.2.2 avec le système d'équations (1.80)). Wiggins [Wig78] propose l'algorithme MED (Minimum Entropy Deconvolution) suivant :

1. Estimation de \mathbf{R}_{DD} ,
2. Initialisation de la réponse impulsionnelle du filtre déconvolueur $\mathbf{g} = [g(0), \dots, g(P)]^T$,
3. Estimation des moments μ_{4Y} et μ_{2Y} ,
4. Estimation de la corrélation \mathbf{r}_{Y^3D} ,
5. Inversion de l'équation (4.35), pour calculer, $\mathbf{g} = \frac{\mu_{2Y}}{\mu_{4Y}} \mathbf{R}_{DD}^{-1} \mathbf{r}_{Y^3D}$,
6. Estimation de la sortie $y(t)$,
7. Normalisation de $g(t)$ et $y(t)$.

Puis nous itérons les étapes 3 à 7 jusqu'à la convergence. L'inversion de la matrice \mathbf{R}_{DD} peut être faite de façon rapide par l'algorithme de Levinson car elle est de type Toeplitz. L'initialisation peut être quelconque en excluant toutefois le filtre comportant uniquement des coefficients nuls, car sinon l'algorithme n'évolue pas. Cet algorithme est nommé MED, bien qu'il ne minimise pas directement l'entropie de la sortie : la dénomination est donc mal choisie.

Pour réaliser l'analogie entre cet algorithme MED maximisant le kurtosis et l'algorithme MANege maximisant la néguentropie, nous pouvons nous intéresser au critère ou au gradient du critère.

L'expression (4.29) de la néguentropie d'un processus aléatoire suivant une ddp gaussienne généralisée permet l'analogie entre la néguentropie et le kurtosis. Avec cette relation, on démontre que la minimisation (et non la maximisation comme dans l'algorithme MED)

du kurtosis est équivalente à la maximisation de la néguentropie dans l'algorithme MANeGe avec une distribution *a priori* pour la réflectivité gaussienne généralisée avec un coefficient $\alpha = 4$. Ce problème de maximisation ou minimisation peut s'expliquer en partie par la propriété du kurtosis. Si nous avons affaire à des signaux sous-gaussiens comme le préconise l'*a priori* avec $\alpha = 4$, nous minimiserions le kurtosis et le parallèle entre les deux algorithmes serait évident. En fait, pour éloigner la distribution de la gaussienne, nous avons une alternative soit sur-gaussianiser, soit sous-gaussianiser la distribution de la sortie. Ainsi, avec un *a priori* sous-gaussien, l'algorithme de la néguentropie (MANeGe) écarte la sortie de la distribution gaussienne en la sous-gaussianisant. Comme nous avons à traiter des signaux sur-gaussiens avec un *a priori* sous-gaussien, nous nous heurtons à un problème de cohérence. Ainsi, la minimisation du kurtosis qui serait équivalente à l'algorithme MANeGe se transforme en une maximisation en raison du mauvais choix de l'*a priori*. Nous retrouvons la remarque que nous avons faite à la fin de la présentation sur le test d'hypothèse où le choix d'un mauvais *a priori* (sous-gaussien au lieu de sur-gaussien) entraînait une minimisation au lieu d'une maximisation de l'entropie sous H_1 .

A propos du gradient du critère, nous pouvons aussi écrire l'équivalence entre le gradient du kurtosis en (4.32) et le gradient de la néguentropie en (4.12). En effet, la fonction score d'une variable aléatoire gaussienne généralisée de coefficient de gaussianité 4 est :

$$\psi_X(x) = \frac{4x^3}{\beta^4}. \quad (4.36)$$

En reportant dans l'expression du gradient de la néguentropie (4.12), nous obtenons :

$$\frac{\partial}{\partial g(t)} H^-(Y) = \frac{1}{\sigma^2} E[y(\tau)d(\tau - t)] - \frac{4}{\beta^4} E[y(\tau)^3 d(\tau - t)]. \quad (4.37)$$

Enfin pour identifier la dernière relation avec le gradient du kurtosis (4.32), nous utilisons le lien entre β et le moment d'ordre 4 donné en (4.23) pour $r = 4$, soit :

$$\mu_{4Y} = \beta^4 \frac{\Gamma(\frac{5}{4})}{\Gamma(\frac{1}{4})}. \quad (4.38)$$

Or $\frac{\Gamma(\frac{5}{4})}{\Gamma(\frac{1}{4})} = \frac{1}{4}$, ainsi, nous pouvons finir l'identification entre les deux gradients. L'identification est réalisée à un signe près en raison de la maximisation au lieu de la minimisation du kurtosis comme l'indique l'algorithme MANeGe avec un *a priori* GG d'exposant 4.

L'approche de MED semble assez paradoxale, car nous choisissons une distribution *a priori* sous-gaussienne pour la réflectivité, alors que nous savons que la réflectivité a une distribution sur-gaussienne. Cependant, la méthode a de bonnes performances pour les distributions très sur-gaussiennes, car le kurtosis caractérise assez bien la distance à la gaussianité. D'autre part, cette méthode est intéressante car la maximisation est faite de façon simple et la convergence est très rapide. En pratique, nous notons une convergence quasi certaine mais nous reviendrons sur les aspects de convergence et de biais dans les sections suivantes. En revanche, dans le cas de signaux sous-gaussiens, il faut minimiser le kurtosis afin d'éloigner la distribution du signal de sortie $y(t)$ de la gaussienne. Dans ce cas,

l'algorithme MED ne convient pas, car, il converge vers un maximum du kurtosis et non vers un minimum. En effet, l'algorithme MED annule le gradient, mais rien ne permet de justifier le type d'extremum (minimum ou maximum) vers lequel nous convergeons. Ainsi, dans cette configuration, il faudra utiliser une autre méthode d'optimisation comme celle du gradient relatif mise en place pour l'algorithme MANeige. Cependant, dans le cadre des applications aux signaux sismiques, ce problème ne se pose pas car nous avons généralement des distributions sur-gaussiennes.

En pratique, surtout pour des données réelles sismiques, la communauté géophysicienne a coutume de dire que l'algorithme MED a tendance à privilégier le réflecteur principal et à atténuer les réflecteurs secondaires [Wig78, Wig85, BB97]. Cet effet est très visible principalement si nous surestimons la longueur du filtre, car nous laissons alors beaucoup de degrés de liberté à l'algorithme pour diminuer la gaussianité des observations. Nous retrouverons cet aspect dans le chapitre 5 lors de l'application à un profil sismique réel. Pour palier à cet inconvénient, plusieurs modifications de l'algorithme ont été proposées. L'algorithme MED est capable d'estimer la phase de l'ondelette mais d'après les géophysiciens brise la structure des données. Quant à la déconvolution prédictive, elle a l'avantage d'être très stable, de bien conserver la structure des données mais ne permet la détermination de la phase. Par conséquent, nous pouvons penser à associer les deux méthodes. Dans un premier temps, Wiggins propose d'estimer le filtre à phase minimale avec les statistiques d'ordre 2 et d'utiliser le kurtosis pour lever l'indétermination entre minimum et maximum de phase [Wig85]. Cette méthode s'apparente au travail sur les pôles et zéros des filtres ARMA proposé par Boumahdi [Bou96, Bou95] et expliqué brièvement dans la sous-section 1.4.3.2. L'idée suivante est de construire un critère comme une somme pondérée du kurtosis et d'un critère de déconvolution prédictive. Cette idée a fait l'objet d'un algorithme appelé CMED (Constrained Minimum Entropy Deconvolution) proposé par Boadu et Brown [BB97]. Un paramètre permet de régler l'importance relative des deux termes, il est ajusté de façon manuelle. Ensuite, une estimation dans le domaine fréquentiel du filtre déconvolveur est faite par Sacchi *et al.* [SVC94] de la même façon que [Gra79]. Une extension de l'algorithme MED à la déconvolution d'images est décrite dans [WB98].

4.2.3 Déconvolution par variation de norme : algorithme VARI-MAX

Après Wiggins, certains auteurs comme Gray et Claerbout [Cla77, Gra79] proposent de déconvoluer les signaux sismiques à l'aide de critère de variation de norme.

Nous pouvons définir la norme α d'un processus aléatoire par :

$$\|Y\|_{\alpha} = (E [|y(\tau)|^{\alpha}])^{1/\alpha}. \quad (4.39)$$

Dans le cas particulier $\alpha = 2$, nous avons la norme 2, classiquement utilisée. Le kurtosis est tout simplement le rapport entre la norme 4 et la norme 2 élevé à la puissance 4. Avec l'inégalité de Jensen, nous pouvons montrer que :

$$\|Y\|_{\alpha_1} < \|Y\|_2 < \|Y\|_{\alpha_2}, \quad \text{si } \alpha_1 < 2 < \alpha_2. \quad (4.40)$$

L'écart entre les normes sera d'autant plus grand que la loi sera sur-gaussienne. Nous avons coutume de dire que la sensibilité aux grands événements de la norme α augmente en même temps que α . Choisissons la norme 2 comme référence, avec le choix de $\alpha = 4$, cette remarque permet de justifier que nous maximisons le kurtosis pour les lois sur-gaussiennes. Ensuite, pour généraliser cette idée mais avec $\alpha < 2$, Gray et Claerbout proposent de déconvoluer les signaux sismiques en minimisant le critère de variation de norme suivant :

$$V_\alpha(Y) = \frac{E[|y(\tau)|^\alpha]}{E[y(\tau)^2]^{\alpha/2}}. \quad (4.41)$$

Le critère $V_\alpha(Y)$ est équivalent au rapport des normes α et 2 de Y élevé à la puissance α . Nous minimisons car $\alpha < 2$. Gray et Claerbout minimisent cette grandeur avec un algorithme identique à l'algorithme MED de Wiggins. Ils calculent le gradient du critère par rapport aux coefficients de la réponse impulsionnelle du filtre déconvolveur pour $t = 0, \dots, P$:

$$\frac{\partial V_\alpha(Y)}{\partial g(t)} = \frac{\partial}{\partial g(t)} \frac{E[|y(\tau)|^\alpha]}{E[y(\tau)^2]^{\alpha/2}} = \frac{\alpha}{\mu_{2Y}^{\alpha/2}} E[\text{sign}(y(\tau))|y(\tau)|^{\alpha-1}d(\tau-t)] - \frac{\alpha\mu_{\alpha|Y|}}{\mu_{2Y}^{\alpha/2+1}} E[y(\tau)d(\tau-t)], \quad (4.42)$$

où $\mu_{\alpha|Y|}$ est le moment d'ordre α de $|Y|$ et $\text{sign}(z)$ est la fonction signe valant 1 si $z > 0$ et -1 sinon. Après annulation du gradient (4.42) du critère de variation de norme, nous pouvons déduire l'équation d'estimation des P coefficients du filtre déconvolveur g :

$$\sum_{t'=0}^P g(t') E[d(\tau-t')d(\tau-t)] = \frac{\mu_{2Y}}{\mu_{\alpha|Y|}} E[\text{sign}(y(\tau))|y(\tau)|^{\alpha-1}d(\tau-t)] \quad t = 0, \dots, P. \quad (4.43)$$

Pour trouver la réponse impulsionnelle du filtre déconvolveur, ils utilisent la même formulation matricielle que Wiggins dans l'algorithme MED. En effet, en réorganisant sous forme matricielle le système de d'équations qui provient de (4.43) pour les différentes valeurs de $t = 0, \dots, P$, nous obtenons :

$$\mathbf{R}_{DD}\mathbf{g} = \frac{\mu_{2Y}}{\mu_{\alpha|Y|}} \mathbf{r}_{\text{sign}(Y)|Y|^{\alpha-1}D}. \quad (4.44)$$

La seule différence avec l'algorithme MED est le second membre de (4.42) qui fait apparaître un terme de corrélation entre les données $d(\tau)$ et une transformation non linéaire de la sortie $y(\tau)$. L'algorithme a la même structure que l'algorithme MED décrit dans la sous-section précédente avec une recherche itérative de la solution, qui comprend une inversion du système (4.44) à chaque itération. Les premiers développements de Gray et Claerbout [Gra79] ont été faits avec $\alpha = 1$, puis ont été étendus à toutes valeurs $1 < \alpha < 2$. L'extension peut se faire pour toute valeur de α . Si $\alpha > 2$ nous devons maximiser $V_\alpha(Y)$ comme pour l'algorithme MED. L'algorithme revient à ajuster le filtre déconvolveur pour annuler le gradient donc l'aspect minimisation ou maximisation n'apparaît plus. Par la suite, nous nommerons cet algorithme VARIMAX.

D'après le calcul de la néguentropie d'un processus GG de coefficient de gaussianité α en (4.29), il est évident que la minimisation de la variation de norme avec $\alpha < 2$ pour des signaux sur-gaussiens est équivalente à la maximisation de la néguentropie de la sortie avec une distribution *a priori* gaussienne généralisée avec un coefficient de gaussianité α . Ce parallèle avait déjà été avancé par Gray [Gra79], mais il est valable uniquement pour $\alpha < 2$ (resp. $\alpha > 2$) appliquée à une sortie sur-gaussienne (resp. sous-gaussienne). Le kurtosis pour l'algorithme MED est un cas particulier avec $\alpha = 4$ et la minimisation est remplacée par une maximisation. Le parallèle peut être aussi fait avec le gradient du critère de la même façon que pour l'algorithme MED. La fonction score du processus GG de coefficient α est :

$$\psi_X(x) = \frac{\alpha \text{sign}(x) |x|^{\alpha-1}}{\beta^\alpha}. \quad (4.45)$$

En reportant dans l'expression du gradient de la néguentropie (4.12), nous avons :

$$\frac{\partial}{\partial g(t)} H^-(Y) = \frac{1}{\sigma^2} E[y(\tau) d(\tau - t)] - \frac{\alpha}{\beta^\alpha} E[\text{sign}(x) x^{\alpha-1} d(\tau - t)]. \quad (4.46)$$

Enfin, pour identifier la dernière relation avec le gradient du critère VARIMAX (4.42), nous utilisons le lien entre β et le moment d'ordre α de $|Y|$ qui est donné en (4.25). Mais, il faut ajouter la relation de récurrence vérifiée par la fonction Gamma :

$$\Gamma(x + 1) = x\Gamma(x). \quad (4.47)$$

Ainsi, en appliquant cette relation pour $x = \alpha$ nous obtenons :

$$\mu_{\alpha|Y} = \frac{\beta^\alpha}{\alpha}. \quad (4.48)$$

Avec la relation (4.48), le gradient de la néguentropie avec un *a priori* GG de paramètre α de l'équation (4.46) est équivalent au gradient (4.42) du critère variation maximum de norme.

Pour des signaux sismiques, Walden [WH86] montre en s'appuyant sur des séquences de réflectivité réelles que leurs distributions peuvent être modélisées par des distributions GG de coefficient de gaussianité compris entre 0.6 et 1.5. Il estime le coefficient de gaussianité par maximum de vraisemblance. Par conséquent, nous pourrions utiliser un coefficient α compris entre 0.6 et 1.5.

Tout d'abord, pour un problème évident de stabilité, il faudra *a priori* que $\alpha > 1$, car dans le gradient (4.42), nous avons des termes en $y(\tau)^{\alpha-1}$. Ces termes peuvent devenir très grands pour des valeurs faibles de $y(\tau)$. Elles sont *a priori* nombreuses en raison de la sur-gaussianité de la réflectivité recherchée. Ce problème est aussi remarqué pour l'algorithme MANeige. Ensuite, en pratique pour $1 < \alpha < 2$, Gray a mis en évidence des problèmes de convergence de l'algorithme de minimisation de variation de norme. En effet, la résolution itérative de l'équation annulant le gradient ne permet pas la convergence de l'algorithme vers le minimum, et, au contraire il tend à maximiser le critère. Pour palier à ce problème, Gray [Gra79] modifie la structure de l'algorithme d'optimisation pour assurer la minimisation au lieu de la maximisation. Ce problème ne se retrouve pas dans l'algorithme MANeige qui

possède en plus la propriété d'équivariance grâce à l'utilisation du gradient relatif. Claerbout [Cla77] propose d'utiliser $\alpha = 1.5$ comme valeur réalisant le compromis d'un *a priori* sur-gaussien et permettant une optimisation rapide et certaine. Dans la section suivante, nous reviendrons sur ce choix en ce qui concerne les performances. En effet, si α tend vers 1 par valeur supérieure, alors la minimisation devient de plus en plus difficile car la convexité du critère est faible. C'est un problème classique rencontré pour les critères de norme \mathbb{L}^1 . Pour cet algorithme avec $\alpha = 1.5$, Claerbout parle de déconvolution parcimonieuse. La parcimonie correspond en fait à un simple *a priori* GG d'exposant $\alpha = 1.5$. Cette définition de la parcimonie s'intéresse à la ddp des signaux, plus récemment des visions plus générales sont apparues [BZJM06].

4.2.4 Transformation exponentielle

Ooe et Ulrych [OU79] voulant tirer parti de la robustesse au bruit de l'algorithme MED de Wiggins maximisant le kurtosis, mais notant la tendance à supprimer les plus petits réflecteurs, ont suggéré de ne pas mesurer la sur-gaussianité directement sur la sortie estimée $y(\tau)$ mais sur une nouvelle sortie $z(\tau)$ qui est une fonction non linéaire de $y(\tau)$. Tout d'abord, nous pouvons remarquer que l'algorithme MED est équivalent à appliquer l'algorithme VARIMAX avec $\alpha = 1$ sur une sortie $z(\tau) = y(\tau)^2$. Ooe et Ulrych proposent une non-linéarité du type :

$$z = 1 - \exp \left\{ -\frac{1}{2} \left(\frac{Cy}{y_{max}} \right)^2 \right\}, \quad (4.49)$$

où y_{max} est le maximum de $|y|$ et C un paramètre à régler. Cette transformation a pour effet d'amplifier de façon plus importante les valeurs intermédiaires de $y(\tau)$ au détriment des petites valeurs de $y(\tau)$. Ainsi, les valeurs intermédiaires ont une plus grande influence dans le critère de sur-gaussianité. Pour illustrer cette remarque, la figure 4.4 trace la transformation exponentielle pour les valeurs de $C = 3$ et $C = 2.38$ mais également la droite $z = |y|$ et la courbe $z = y^2$. Comme les algorithmes de déconvolution sont insensibles à un facteur d'échelle sur les signaux, les axes sont normalisés à leur valeur maximale pour faciliter la comparaison

Ooe et Ulrych recommandent de choisir $\sqrt{2} < C < 3$ pour avoir un compromis acceptable entre la suppression du bruit et une bonne sensibilité de l'algorithme aux petits événements. Le critère de choix de C est lié à la position du point d'intersection de la non linéarité et de la courbe $z = |y|$. Ce point doit jouer le rôle de séparateur des échantillons appartenant au bruit et ceux représentant les réflecteurs. L'objectif est de limiter l'importance des échantillons du bruit dans le critère. Ainsi, les échantillons situés avant le point d'intersection doivent appartenir au bruit. Ceux situés au delà du point doivent être les réflecteurs. Par exemple, avec la courbe de la figure 4.4, nous notons que les valeurs au centre du graphique augmentent leur contribution au sein du critère. Par conséquent, la valeur du paramètre C ne doit pas être trop grande au risque d'augmenter l'importance des échantillons provenant des plus faibles valeurs d'échantillons, c'est-à-dire du bruit.

Après la transformation exponentielle, Ooe et Ulrych estiment le filtre déconvolveur $g(t)$ en minimisant un critère de sur-gaussianité sur $z(\tau)$. Ils minimisent le critère de variation de norme de Gray et Claerbout avec $\alpha = 1$ (distribution *a priori* pour $z(\tau)$ laplacienne). La

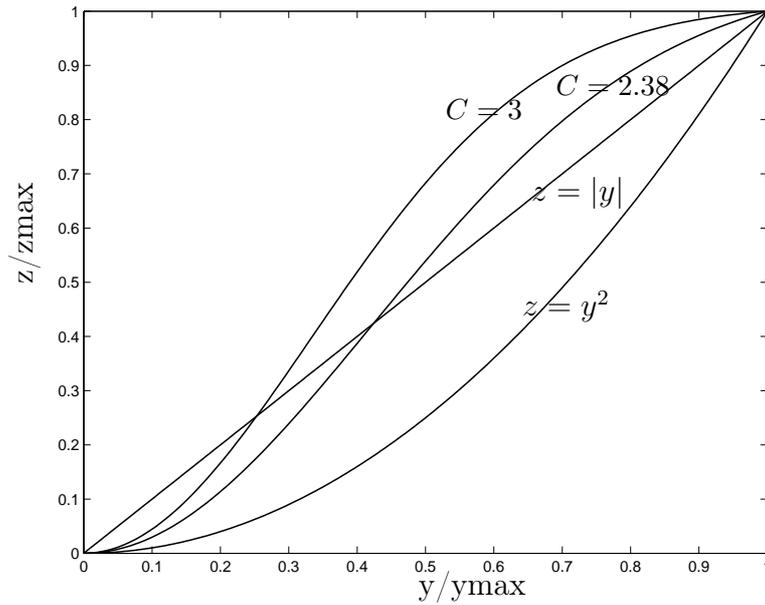


FIG. 4.4 – Transformée exponentielle de Ooe et Ulrych pour $C=3$, $C=2.38$ et $z = y^2$ et $z = |y|$.

courbe $z = |y|$ correspond à l'utilisation de l'algorithme VARIMAX avec $\alpha = 1$ directement sur la sortie estimée $y(\tau)$. De même, la courbe d'équation $z = y^2$ correspond à l'algorithme MED. Cette approche rend aussi l'algorithme plus stable que l'algorithme VARIMAX avec $\alpha = 1$ appliqué directement sur la sortie.

L'idée de cette méthode est assez simple. Cependant, le parallèle entre cette approche et l'algorithme MANege avec une distribution *a priori* fixée pour la sortie $y(\tau)$ n'est pas évident à réaliser. Tout d'abord, si nous voulons conserver la symétrie des distributions, la non-linéarité (4.49) doit être une fonction impaire de y . En effet, si les ddp ne sont pas symétriques, il sera difficile de justifier que le maximum de variation de norme mesure la sur-gaussianité du signal estimé. En fait, nous pouvons définir la non-linéarité par :

$$z = \text{sign}(y) \left(1 - \exp \left\{ -\frac{1}{2} \left(\frac{Cy}{y_{\max}} \right)^2 \right\} \right). \quad (4.50)$$

Ce changement de définition ne change en rien l'algorithme car en réalité le critère n'est pas sensible au signe de $z(\tau)$, mais ses fondements théoriques requièrent une ddp symétrique. Ensuite, Ooe et Ulrych utilisent l'algorithme VARIMAX avec $\alpha = 1$. Ainsi, comme nous l'avons montré dans le paragraphe précédent, cet algorithme revient à choisir une distribution *a priori* laplacienne pour $z(\tau)$. Or d'après la définition (4.50), z est bornée entre -1 et 1 , donc l'*a priori* est mal choisi. Pour poursuivre il faudrait prendre une loi *a priori* laplacienne dont nous tronquons la queue de distribution pour $z(\tau)$, et ensuite appliquer le changement de variable aléatoire. En définitive, l'idée de la transformation exponentielle est simple, mais il est impossible de la lier à l'algorithme MANege avec un *a priori* simple pour la distribution de la réflectivité estimée.

4.2.5 Algorithme de Godfrey

Dans un cadre plus formel, Godfrey [God78] propose de maximiser le critère d'entropie relative suivant :

$$J(g) = \frac{1}{T} \left(\sum_{\tau=1}^T \log f(y(\tau)) - \sum_{\tau=1}^T \log h(y(\tau)) \right). \quad (4.51)$$

Cette expression est identique aux rapports de vraisemblance dans le cas du test d'hypothèse que nous avons mentionné au début de la section 4.2. Godfrey utilise pour h une loi gaussienne de variance σ et pour f une loi de Cauchy définie sous la forme suivante :

$$f(y) = \frac{\Gamma(m)}{\Gamma(m - \frac{1}{2})\sqrt{(2m-3)\pi\sigma^2}} \frac{1}{\left(1 + \frac{y^2}{(2m-3)\sigma^2}\right)^m}. \quad (4.52)$$

La loi de Cauchy a un paramètre de forme $m > 1.5$ et un paramètre d'échelle σ . Lorsque m augmente, la loi de Cauchy tend vers la loi gaussienne. Le critère d'entropie relative $J(g)$ est en fait l'estimation de la négentropie de la sortie $y(\tau)$ définie en (4.8) avec une distribution *a priori* de type Cauchy pour $y(\tau)$. Nous noterons que la loi de Cauchy est sur-gaussienne pour toutes les valeurs de m car la ddp tend moins vite vers 0 à l'infini que la gaussienne.

La fonction score théorique pour une loi de Cauchy est définie par :

$$\psi_Y(y) = \frac{2my}{(2m-3)\sigma^2} \frac{1}{1 + \frac{y^2}{(2m-3)\sigma^2}}. \quad (4.53)$$

Pour réaliser la maximisation du critère $J(g)$, Godfrey propose la même méthode itérative que l'algorithme MED ou que l'algorithme VARIMAX. Cependant, Godfrey remarque des problèmes de convergence en raison de la minimisation du critère au lieu de la maximisation.

Pour rendre adaptatif son algorithme au cours des itérations, Godfrey [God78] propose de faire évoluer le paramètre m et la variance à chaque itération. Le paramètre m est estimé à l'aide du rapport des moments. Nous pouvons montrer pour une loi de Cauchy que :

$$\frac{\mu_{1|Y}^2}{\mu_{2Y}} = \frac{2}{\pi} \frac{\Gamma^2(m-1)}{\Gamma^2(m-\frac{1}{2})} \left(m - \frac{3}{2}\right), \quad (4.54)$$

où $\mu_{1|Y} = E[|y|]$. A chaque itération, nous devons estimer $\mu_{1|Y}$ et μ_{2Y} et résoudre de façon implicite l'équation (4.54) pour estimer m . La variance est estimée simplement par le moment d'ordre 2. Dans son article, Godfrey propose aussi d'estimer la densité de probabilité de la sortie estimée à chaque itération avec une approximation constante par morceau. Cette approche est équivalente à un estimateur à noyaux avec un noyau de type uniforme. Cependant, ce noyau n'est pas dérivable donc la dérivation de la ddp pour parvenir à la fonction score pose des problèmes de singularité. Le principe de l'algorithme de Godfrey est proche des développements que nous avons proposés avec l'algorithme MANeGe. Nous

noterons que Godfrey ne fait pas apparaître le concept de fonction score.

Nous pouvons envisager de nombreux choix de la fonction score du processus de sortie, ce qui correspond à un choix de loi *a priori* pour la réflectivité estimée. Walden [Wal85] utilise aussi une fonction score définie par la fonction tanh. Ce choix est aussi classiquement fait en séparation de sources pour des signaux avec des sources sur-gaussiennes [ER99, CCA00, ZCA04].

Pour conclure, dans cette section nous avons présenté les principales méthodes de déconvolution développées dans le cadre des données sismiques utilisant la sur-gaussianité de la réflectivité pour l'estimation du filtre déconvolveur $g(t)$: toutes sont des cas particuliers de l'algorithme MANeige. Dans la suite, nous proposons de comparer les différents algorithmes d'un point de vue théorique. Dans la section suivante, nous nous intéresserons plus particulièrement à l'identifiabilité du système dans le cadre non bruité par les différents critères.

4.3 Contraintes d'identifiabilité sur les critères

Dans ce paragraphe, nous souhaitons justifier que les algorithmes précédents permettent l'inversion du modèle. En effet, nous n'avons pas encore montré clairement l'identifiabilité du système par le biais des algorithmes présentés dans le cas où la séquence de réflectivité est iid et non-gaussienne. Cette section reprend en partie les arguments avancés par Donoho [Don81].

Pour l'algorithme MANeige, l'inversibilité est justifiée par le théorème centrale limite. Pour tous les algorithmes, nous pouvons écrire que l'estimée $\hat{g}(t)$ du filtre déconvolveur $g(t)$ est faite par maximisation d'un critère $J(Y)$ portant sur la sortie Y :

$$\hat{g}(t) = \max_{g(t)} J(Y). \quad (4.55)$$

Par exemple, pour l'algorithme MED, c'est le kurtosis. Pour l'algorithme VARIMAX avec $\alpha < 2$, nous prendrons l'opposé (ou l'inverse) du critère pour parler de maximisation au lieu de minimisation. Nous devons nous intéresser aux contraintes sur $J(Y)$ pour que son optimum coïncide avec le filtre déconvolveur théorique. Donoho [Don81] montre que seulement deux conditions sont suffisantes pour assurer cette consistance du critère. La première contrainte est l'invariance par changement d'échelle du critère $J(Y)$. Elle est liée à l'indétermination d'échelle sur le filtre déconvolveur et la sortie estimée. Ensuite, le critère doit vérifier la propriété suivante :

Propriété :

Soient a_i des coefficients non nuls, X une variable aléatoire, et X_i des réalisations de cette variable aléatoire, alors le critère $J(X)$ doit vérifier pour une combinaison linéaire $\sum_i a_i X_i$ comportant au minimum deux termes l'inégalité suivante :

$$J\left(\sum_i a_i X_i\right) < J(X). \quad (4.56)$$

Nous pouvons analyser cette propriété comme une décroissance stricte du critère pour une combinaison linéaire de variables aléatoires. Nous imposons à la somme de contenir au moins deux termes car sinon nous avons l'égalité des deux termes $J(X)$ et $J(a_1 X_1)$ par le biais de l'indétermination d'échelle. Pour l'algorithme MANege, l'inégalité est vérifiée car la négentropie d'une somme de variables aléatoires est inférieure strictement à la négentropie de chacune des variables. Pour l'algorithme VARIMAX, il n'est pas simple de justifier que $V_\alpha(\sum_i a_i X_i) < V_\alpha(X)$ pour $\alpha > 2$, et, $V_\alpha(\sum_i a_i X_i) > V_\alpha(X)$ pour $\alpha < 2$. Donoho justifie à l'aide de la notion de distance développée par Zolotarev [Zol76] que pour $\alpha > 2$ (*a priori* sous-gaussien) la condition est vérifiée. Ainsi, l'algorithme VARIMAX pour $\alpha > 2$ permet l'identifiabilité du système. Pour $\alpha < 2$, aucune preuve n'est avancée pour démontrer l'identifiabilité, mais aucun contre-exemple ne montre que le système peut ne pas être identifiable.

En définitive, les critères précédemment utilisées dans les algorithmes MANege, MED et VARIMAX pour $\alpha > 2$ permettent l'identifiabilité du filtre déconvolveur $g(t)$. En revanche, nous ne pouvons rien conclure pour l'algorithme VARIMAX avec $\alpha < 2$. Pour l'algorithme de Godfrey, il semble difficile de prouver l'identifiabilité. Dans la suite, nous noterons par $g^0(t)$ le filtre déconvolveur théorique. Dans la section suivante pour comparer ces différents estimateurs, nous allons détailler les conditions sous lesquelles le critère sera concave et l'estimation se fera sans biais.

4.4 Concavité et estimation non biaisée

En général, pour la construction d'un algorithme performant, nous souhaitons avoir un critère avec une bonne courbure et permettant une estimation non biaisée en moyenne. Si nous minimisons (resp. maximisons) un critère, nous souhaitons qu'il soit convexe (resp. concave). La convexité/concavité permet l'optimisation par des techniques simples de gradient. Nous débuterons en présentant les conditions théoriques suffisantes pour assurer la concavité et l'estimation non biaisée, puis nous rechercherons si ces conditions sont vérifiées ou non pour les algorithmes MANege, MED et plus généralement VARIMAX.

4.4.1 Contraintes théoriques

Tous les algorithmes précédents sont équivalents soit à la maximisation de la négentropie, soit à l'annulation du gradient de la négentropie (4.12) qui sera estimé sous hypothèse d'ergodicité par :

$$\frac{\partial \widehat{H^-(Y)}}{\partial g(t)} = -\frac{1}{T} \sum_{\tau} \phi_Y(y(\tau)) d(\tau - t), \quad (4.57)$$

où $\phi_Y(y(\tau))$ est définie par (4.16). Pour que la méthode soit sans biais, le critère doit être minimum pour $g(t) = g^0(t)$, c'est-à-dire que le gradient doit s'annuler en $g^0(t)$:

$$0 = \left. \frac{\partial \widehat{H}^-(Y)}{\partial g(t)} \right|_{g=g^0} = -\frac{1}{T} \sum_{\tau} \phi_Y(y(\tau)) d(\tau - t), \quad \forall t. \quad (4.58)$$

D'après Donoho [Don81], il existe trois conditions au point $g = g^0$ de régularité suffisantes pour assurer l'égalité (4.58) et la concavité du critère :

$$(R1) \quad E[\phi_Y(y)] = 0, \quad (4.59)$$

$$(R2) \quad E[\phi'_Y(y)] > 0, \quad (4.60)$$

$$(R3) \quad E[y\phi_Y(y)] = 0. \quad (4.61)$$

Les conditions (R1) et (R2) sont liées aux problèmes des M -estimateurs traités par Huber [Hub77]. Nous avons choisi de ne pas faire le parallèle des critères précédents avec les M -estimateurs, mais le lecteur pourra se reporter aux deux articles [Don81, Wal85] pour de plus amples explications. La condition (R2) est une condition de concavité (hessien négatif) pour assurer que l'optimum trouvé soit un maximum et non un minimum.

Il faut relier la condition (R3) à la convergence de la moyenne de l'estimée du gradient (4.58) à zéro. En effet, nous souhaitons que :

$$E \left[\frac{1}{T} \sum_{\tau} \phi_Y(y(\tau)) d(\tau - t) \right] = 0, \quad \forall t. \quad (4.62)$$

Afin de supprimer les données de cette expression, nous pouvons convoluer la dernière expression par $g(t)$, nous avons donc :

$$E \left[\frac{1}{T} \sum_{\tau} \phi_Y(y(\tau)) y(\tau - t) \right] = 0, \quad \forall t. \quad (4.63)$$

Après permutation de la somme et de l'espérance et en utilisant l'indépendance entre $\phi_Y(y(\tau))$ et $y(\tau - t)$ pour $t \neq 0$ car le processus de sortie $y(\tau)$ est iid et que la sortie est à valeur moyenne nulle, nous obtenons la condition (R3).

La condition (R1) est utile uniquement si nous avons des signaux non centrés, car sinon la simplification précédente n'est pas possible. (R1) permet de supprimer l'effet des valeurs moyennes des signaux dans l'estimation du gradient. La condition (R1) est vérifiée pour toute fonction $\phi_Y(y)$ impaire. Or d'après la définition de la fonction par $\phi_Y(y(\tau)) = \psi_Y(y(\tau)) - \frac{y(\tau)}{\sigma^2}$, cette propriété est vérifiée pour tous les algorithmes car les fonctions scores $\psi_Y(y)$ choisies ou estimées sont impaires avec l'hypothèse de distribution symétrique.

Dans les sous-sections suivantes, nous allons vérifier si les algorithmes MANeige, MED et VARIMAX répondent aux contraintes (R2) et (R3).

4.4.2 Cas de l'algorithme MANege

Pour l'algorithme MANege, nous pouvons vérifier les conditions (R2) et (R3). Pour la condition (R2), nous devons montrer que :

$$E[\phi'_Y(y)] = E[\psi'_Y(y)] - E\left[\frac{1}{\sigma^2}\right] > 0. \quad (4.64)$$

L'inégalité précédente s'écrit :

$$E[\psi'_Y(y)] E[y^2] > 1. \quad (4.65)$$

En utilisant le lemme 2 de la sous-section 2.4.1.2 concernant l'estimation directe des fonctions scores, nous avons :

$$E[\psi'_Y(y)] = E[\psi_Y(y)^2]. \quad (4.66)$$

D'après, l'inégalité de Cauchy-Schwarz, nous savons que :

$$E[y^2] E[\psi_Y(y)^2] > (E[|y\psi_Y(y)|])^2. \quad (4.67)$$

Nous avons l'inégalité stricte car le cas d'égalité est caractérisé par une dépendance linéaire entre $\psi_Y(y)$ et y , soit une fonction score d'une loi gaussienne. Or, nous supposons que nous n'avons pas à traiter des signaux gaussiens sinon le système n'est pas identifiable. Ensuite, nous devons montrer que le terme de droite de l'équation (4.67) est égal à 1. Nous ferons l'hypothèse supplémentaire que la fonction score est une fonction impaire, *i.e.* que la distribution de Y est symétrique, et que Y est à valeur moyenne nulle. Ainsi, le second membre de (4.67) s'écrit :

$$E[|y\psi_Y(y)|] = 2 \int_0^{+\infty} y\psi_Y(y)p_Y(y)dy = -2 \int_0^{+\infty} yp'_Y(y)dy = 2 \int_0^{+\infty} p_Y(y)dy = 1. \quad (4.68)$$

En définitive avec la dernière relation et (4.67), nous avons démontré que la propriété (R2) était vérifiée dans le cas de distribution symétrique et à valeur moyenne nulle. Nous avons aussi supposé que nous connaissions la fonction score du processus Y , ou que nous l'estimions parfaitement.

Pour la condition (R3), nous devons montrer que :

$$E[y\psi_Y(y)] - E\left[\frac{y^2}{\sigma^2}\right] = 0. \quad (4.69)$$

Le second terme vaut de façon évidente 1. En utilisant les restrictions identiques et la même méthode de calcul que pour (4.68), le premier vaut également 1. Ainsi, la condition (R3) est vérifiée pour l'algorithme MANege. Nous venons de montrer que les trois conditions suffisantes pour la concavité et l'estimation non biaisée de l'algorithme MANege sont vérifiées. En conclusion, l'algorithme MANege estimera sans biais le filtre déconvolveur dans le cas d'un système sans bruit et le critère sera concave, ce qui permet une optimisation par une descente de gradient. Pour l'algorithme MANege, nous avons tout de même supposé

que l'estimation de la fonction score était parfaite. Si l'estimation est imparfaite, nous ne pouvons rien conclure.

4.4.3 Cas des algorithmes MED et VARIMAX

Dans la suite, nous nous intéressons aux cas des algorithmes MED et VARIMAX. Nous rappelons que, pour tous ces algorithmes, la condition (R1) est vérifiée car les fonctions scores choisies sont impaires. Dans un premier temps, nous pouvons vérifier les conditions (R2) et (R3) pour l'algorithme MED et VARIMAX qui utilisent un *a priori* gaussien généralisé avec un coefficient α . Nous débuterons par la condition de concavité (R2).

Condition de concavité (R2)

Avec l'expression de la fonction score théorique d'une loi gaussienne généralisée (4.45), pour la condition (R2), nous devons montrer que :

$$E \left[-\frac{1}{\sigma^2} + \frac{\alpha(\alpha-1)|y|^{\alpha-2}}{\beta^\alpha} \right] > 0. \quad (4.70)$$

Il faut noter que β est lié à la variance σ par la relation (4.24). L'inégalité précédente est équivalente à :

$$\alpha(\alpha-1) \left(\frac{\Gamma(\frac{3}{\alpha})}{\Gamma(\frac{1}{\alpha})} \right)^{\alpha/2} E \left[\left| \frac{y}{\sigma} \right|^{\alpha-2} \right] > 1. \quad (4.71)$$

Nous remarquons que pour $\alpha < 1$, l'inégalité ne sera pas vérifiée : la courbure du critère ne permettra par une optimisation par une descente de gradient Ceci avait aussi été remarqué par Gray et Claerbout [Gra79, Cla77]. Ensuite, nous notons simplement que pour $\alpha = 4$ (soit l'algorithme MED), le terme d'espérance est simplement unitaire, et le terme restant vaut 1.37. Ainsi, l'algorithme MED vérifie la condition (R2), quelle que soit la distribution du signal de réflectivité.

Plus généralement pour l'algorithme VARIMAX, nous ne pouvons pas calculer simplement le terme comportant l'espérance mathématique. En effet, nous avons besoin de la ddp de Y pour faire ce calcul. Nous pouvons cependant nous intéresser à quelques cas particuliers. Par exemple, si la sortie Y suit une loi de probabilité GG de paramètres (γ, η) alors que nous avons choisi une distribution *a priori* GG de paramètre (α, β) . Dans ce cas particulier, le terme espérance de (4.71) s'exprime avec le moment d'ordre $\alpha - 2$ de $|Y|$ qui suit une loi GG de paramètres (γ, η) , dont l'expression est donnée en (4.25). Ainsi, la condition (R2) est équivalente à :

$$f(\alpha, \gamma) = \alpha(\alpha-1) \left(\frac{\Gamma(\frac{3}{\alpha})}{\Gamma(\frac{1}{\alpha})} \right)^{\alpha/2} \left(\frac{\Gamma(\frac{1}{\gamma})}{\Gamma(\frac{3}{\gamma})} \right)^{\alpha/2-1} \frac{\Gamma(\frac{\alpha-1}{\gamma})}{\Gamma(\frac{1}{\gamma})} > 1. \quad (4.72)$$

Dans un premier temps, nous pouvons étudier le cas $\gamma = \alpha$, c'est-à-dire que nous avons choisi le bon exposant de la loi GG. Il n'est pas évident de voir que la condition (R2) est vérifiée dans l'équation (4.72). Nous traçons la fonction $f(\alpha, \alpha)$ sur la figure 4.5(a) pour

$\alpha \in [1; 10]$. Nous notons que la fonction est toujours supérieure à 1, la condition (R2) est donc vérifiée. En fait, nous aurions pu conclure directement en nous reportant à l'étude de l'algorithme MANege pour laquelle nous avons supposé parfaitement connue la distribution de sortie. En effet, notre cas d'étude est identique à l'algorithme MANege avec une loi connue, qui est ici une loi GG d'exposant α .

Ensuite, nous proposons de nous pencher sur le cas plus général $\alpha \neq \gamma$. L'étude de $f(\alpha, \gamma)$ n'est pas simple. La figure 4.5(b) représente en fonction de α et γ , le test de validité de (R2) défini par $f(\alpha, \gamma) > 1$. Les zones blanches correspondent aux valeurs de α et γ où la contrainte (R2) est vérifiée et les noires les valeurs pour lesquelles la contrainte n'est pas vérifiée.

Comme nous l'avons dit dans la sous section précédente, pour $\alpha = 4$ soit l'algorithme MED, la contrainte (R2) est vérifiée pour toutes les valeurs de γ . Ensuite, nous voyons que la valeur $\alpha = 2$ joue le rôle de séparateur naturel entre deux zones de la figure pour les lois sous-gaussiennes ($\alpha > 2$) et sur-gaussiennes ($\alpha < 2$). Si nous choisissons un *a priori* sur-gaussien pour une loi GG sur-gaussienne alors (R2) est automatiquement vérifiée. Il en est de même pour les *a priori* sous-gaussiens pour les signaux sous-gaussiens. Ensuite, si nous choisissons un *a priori* sur-gaussien (resp. sous-gaussien) pour un signal sous-gaussien (resp. sur-gaussien), les comportements sont différents. Par exemple, nous remarquons qu'en prenant un coefficient de gaussianité supérieur à 4, la condition (R2) est vérifiée même pour les signaux sous-gaussiens. Néanmoins, nous ne pouvons rien conclure pour d'autres distributions, car nous n'avons démontré cette propriété que pour $\alpha = 4$. En revanche, le choix d'un *a priori* sur-gaussien pour un signal sous-gaussien est très pénalisant pour la condition (R2).

Suite à l'étude réalisée avec les signaux GG, nous pourrions conclure que le choix d'un *a priori* sous-gaussien ($\alpha > 2$) permet d'assurer la condition (R2) pour tous les signaux (sous et sur-gaussiens). De plus, nous pourrions penser que les *a priori* sur-gaussiens fonctionnent pour tous les signaux sur-gaussiens. Pour contredire ces conclusions rapides, nous proposons de nous intéresser au cas d'un processus Y obtenu par :

$$y(t) = \text{sign}(z(t))|z(t)|^{2/\gamma}, \quad (4.73)$$

où Z est un processus gaussien iid et γ un paramètre. Si $\gamma = 2$ nous avons un signal gaussien, si $\gamma > 2$, nous avons un signal sous-gaussien, et respectivement sur-gaussien pour $\gamma < 2$. Nous utilisons parfois cette technique pour simuler simplement des réflectivités sur-gaussiennes. En calculant les moments de Y avec la fonction Γ , nous obtenons :

$$\mu_{\lambda|Y|} = E[|y|^\lambda] = 2^{\frac{\lambda}{\gamma}} \frac{\Gamma(\frac{\lambda}{\gamma} + \frac{1}{2})}{\sqrt{\pi}}. \quad (4.74)$$

En utilisant la dernière relation, la condition (R2) décrite par (4.71) fournit l'inégalité suivante :

$$f(\alpha, \gamma) = \alpha(\alpha - 1) \left(\frac{\Gamma(\frac{3}{\alpha})}{\Gamma(\frac{1}{\alpha})} \right)^{\alpha/2} (\sqrt{\pi})^{\alpha/2-2} \frac{\Gamma(\frac{\alpha-2}{\gamma} + \frac{1}{2})}{\left(\Gamma(\frac{2}{\gamma} + \frac{1}{2}) \right)^{\alpha/2-1}} > 1. \quad (4.75)$$

Dans la relation (4.75), le calcul fait apparaître le moment d'ordre $\frac{2(\alpha-2)}{\gamma}$ de la loi gaussienne de Z , et qui existe pour des raisons d'intégrabilité si et seulement si $\frac{2(\alpha-2)}{\gamma} > -1$. Ainsi, certaines valeurs de α et γ sont incompatibles. La figure 4.5(c) représente le test (4.75) avec en blanc, les zones de validité, en noir, les zones de non-respect et en grisé, la partie où la condition n'existe pas pour des raisons de non-intégrabilité. Sur la figure 4.5(c), nous notons que pour l'algorithme MED ($\alpha = 4$), la condition (R2) est toujours vérifiée. Pour les signaux sur-gaussiens ($\gamma < 2$), si nous souhaitons que la condition de concavité (R2) soit vérifiée et que les moments existent, le choix d'un *a priori* GG sur-gaussien n'est pas simple. Ce contre-exemple permet de montrer que, pour les *a priori* sous-gaussiens avec $\alpha > 4$, il existe des signaux pour lesquels la condition (R2) n'est pas vérifiée (zone en bas à droite de la figure 4.5(c)).

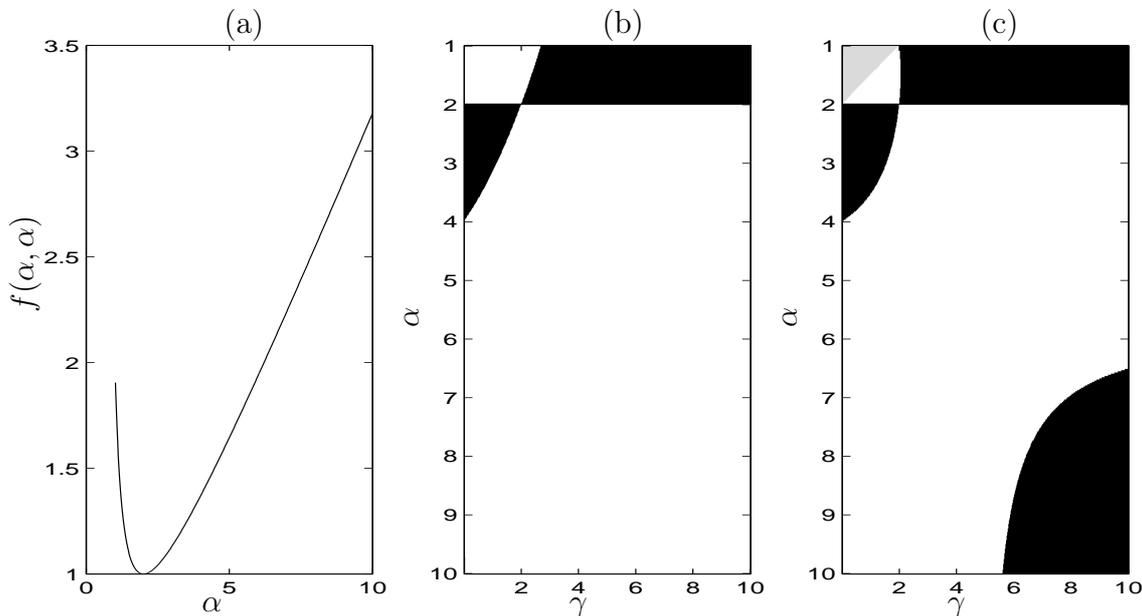


FIG. 4.5 – Étude de la contrainte (R2) : (a) $f(\alpha, \alpha)$ pour l'algorithme VARIMAX, (b) test de validité de (R2) pour l'algorithme VARIMAX avec un *a priori* GG de paramètre α pour Y suivant une loi GG de paramètre γ . En blanc (resp. noir) condition (R2) vérifiée (resp. non vérifiée). (c) idem (b) mais pour un processus suivant une loi décrite en (4.73) de paramètre γ . En gris, la zone de non-existence des moments.

En conclusion, la condition (R2) est automatiquement vérifiée dans 2 cas seulement :

- Pour l'algorithme MED quelle que soit la ddp du signal de sortie,
- Pour l'algorithme VARIMAX, si l'*a priori* α est exact.

Il est difficile de tirer de plus amples conclusions générales. Néanmoins, pour les cas non justifiés théoriquement, nous pouvons penser que le choix du paramètre α est plus critique pour les *a priori* sur-gaussiens que pour les *a priori* sous-gaussiens. En effet, nous devons choisir $\alpha > 1$ pour assurer la stabilité de l'algorithme VARIMAX. Cependant, nous sommes parfois amenés à estimer des moments qui n'existent pas théoriquement (figure 4.5(c)). Cette estimation risque de déstabiliser l'algorithme VARIMAX. Ce compromis stabilité-adéquation de l'*a priori* a été observé en pratique par Claerbout [Cla77] qui pro-

posait de prendre $\alpha = 1.5$.

Précédemment, nous avons proposé une longue discussion sur la condition (R2) car elle permet d'assurer la concavité du critère qui est nécessaire à la stabilité de l'algorithme. Nous allons nous intéresser dorénavant à la condition (R3) d'estimation non biaisée avec les algorithmes VARIMAX et MED.

Condition d'estimation non biaisée (R3)

La condition (R3) est suffisante pour une estimation sans biais du filtre déconvolveur, elle s'exprime pour l'algorithme VARIMAX par :

$$E[y\phi_Y(y)] = 0 = E\left[-\frac{y^2}{\sigma^2} + \frac{\alpha|y|^\alpha}{\beta^\alpha}\right]. \quad (4.76)$$

Si la sortie admet une distribution GG de coefficient α alors la condition (R3) est vérifiée. Pour parvenir à cette conclusion, nous pouvons nous reporter à l'expression du moment d'ordre α de $|Y|$ en (4.25), où nous utilisons le résultat de l'algorithme MANeige pour lequel la vraie distribution de la sortie est connue. Mais, dans le cas général, nous ne pouvons rien conclure. Nous ne proposons pas de faire une étude pour différentes distributions de Y . La condition (R3) montre que nous aurons un biais d'estimation du filtre déconvolveur. Cependant, si l'*a priori* est choisi correctement, le biais sera faible. Nous pouvons même réécrire la condition (R3) de l'équation (4.76) par :

$$E[|y|^\alpha] = \frac{\beta^\alpha}{\alpha}. \quad (4.77)$$

Si nous définissons Y_α la variable aléatoire GG de coefficient α de même variance et de même valeur moyenne que Y , alors, l'équation (4.25) est équivalente à dire que les moments d'ordre α de $|Y_\alpha|$ et $|Y|$ sont égaux. Ainsi, la condition (R3) est vérifiée si la sortie a le même moment d'ordre α que la loi GG de coefficient α . Suivant la même stratégie que Godfrey pour l'ajustement du paramètre de la loi de Cauchy (sous-section 4.2.5) grâce à l'estimation en fonction des moments (4.54), nous pourrions envisager d'ajuster α à chaque itération et le critère d'ajustement serait la vérification de la contrainte (R3) soit :

$$E\left[\left|\frac{y}{\sigma}\right|^\alpha\right] = \mu_{\alpha|\frac{|Y|}{\sigma}} = \frac{1}{\alpha} \left(\frac{\Gamma(\frac{1}{\alpha})}{\Gamma(\frac{3}{\alpha})}\right)^{\alpha/2}. \quad (4.78)$$

Nous noterons que pour les *a priori* de Cauchy de l'algorithme de Godfrey, nous ne pouvons pas exprimer facilement la condition (R3). L'ajustement proposé par Godfrey par le rapport des moments n'assure pas l'estimation sans biais. Cependant, il tend par son ajustement à choisir le meilleur *a priori* possible. Ainsi nous pouvons penser que le biais sera réduit. Dans un contexte général, Donoho propose, pour rendre la condition (R3) satisfaite de modifier la fonction ϕ_Y comme suit :

$$\phi_Y^{new}(y) = \phi_Y(y) - \frac{E[y\phi_Y(y)]}{E[y^2]}y. \quad (4.79)$$

Cette opération est équivalente à une orthogonalisation de la fonction $\phi_Y(y)$ par rapport à y . Par ailleurs, nous pouvons voir cette modification comme un ajout d'un terme linéaire sur la fonction $\phi_Y(y)$. Avec l'équation du gradient de l'algorithme de la néguentropie, nous avons montré que, pour le minimum nous avons :

$$E[\phi_Y(y(\tau))y(\tau - t)] = \delta(t). \quad (4.80)$$

Avec la nouvelle fonction ϕ_Y^{new} , nous serons au minimum lorsque :

$$E[\phi_Y(y(\tau))y(\tau - t)] - \frac{E[y\phi_Y(y)]}{E[y^2]}E[y(\tau)y(\tau - t)] = \delta(t). \quad (4.81)$$

Cette expression peut être vue comme le gradient d'un critère optimisant une somme pondérée de la néguentropie et d'un terme de blanchiment à l'ordre 2. Cette idée a fait l'objet de l'algorithme CMED (Constrained Minimum Entropy Deconvolution) [BB97]. En effet, Boadu et Brown construisent un critère utilisant la maximisation du kurtosis et la décorrélation à l'ordre 2 utilisée dans la déconvolution prédictive. Si la condition (R3) est quasi vérifiée (biais faible) alors la modification de la fonction $\phi_Y(y)$ n'apporte qu'une modification mineure à l'algorithme et à ses performances.

Enfin, nous avons fait le choix de ne pas traiter le cas de l'algorithme de Godfrey, qui prend en compte une loi *a priori* de Cauchy, car nous parviendrons très certainement aux mêmes résultats, *i.e.* que le biais est nul si la sortie suit effectivement une loi de Cauchy de même paramètre que l'*a priori*.

Nous avons montré dans les sous-sections précédentes, que le biais d'estimation sera nul si nous connaissons ou estimons parfaitement la fonction score du signal de sortie. L'algorithme MANeige n'assure aucunement un biais nul si l'estimation de la fonction score n'est pas parfaite. Cependant, il a l'avantage de pouvoir s'adapter et le biais sera d'autant plus faible que l'estimation de la fonction score sera bonne. Ensuite, pour la concavité du critère qui permet l'optimisation par une méthode de gradient, l'algorithme MED fournit un critère concave quelle que soit la distribution de la réflectivité. En revanche, pour l'algorithme VARIMAX en général, nous ne pouvons rien assurer. Cependant, il ressort que les *a priori* sous-gaussiens donnent en général une meilleure garantie de concavité. De plus, les *a priori* sur-gaussiens peuvent faire appel à l'estimation de moments qui n'existent pas théoriquement, ce qui sera très préjudiciable à la stabilité de l'algorithme. Ainsi, la stabilité des algorithmes et d'une estimation non biaisée ne sont pas assurées. Nous proposons dans la section suivante de nous intéresser aux performances des algorithmes avec le calcul de la variance asymptotique des estimateurs.

4.5 Variance asymptotique des algorithmes

Ces calculs ont été principalement développés par Donoho [Don81], et, Walden [Wal85] en a poursuivi l'analyse. Nous calculerons la variance asymptotique de l'estimée proposée dans les différents algorithmes, *i.e.* quand le nombre d'échantillons T tend vers l'infini.

4.5.1 Calcul théorique

L'étude de la variance asymptotique de l'estimée permet d'affiner la caractérisation de l'algorithme en complément du biais d'estimation. En fait, nous nous restreignons aux performances asymptotiques car ce sont les seules qui sont calculables et exploitables sans de trop grandes difficultés. Les coefficients de retard nul $g(0)$ et $g^0(0)$ sont normalisés à 1. En général, la variance asymptotique est intéressante lorsque nous supposons que le vecteur estimé $\hat{\mathbf{g}} = [\hat{g}(1), \dots, \hat{g}(P)]^T$ à partir de T échantillons converge vers la vraie valeur du vecteur $\mathbf{g}^0 = [g^0(1), \dots, g^0(P)]^T$ à la vitesse de \sqrt{T} , où T est le nombre d'échantillons, *i.e.* :

$$\sqrt{T}(\hat{\mathbf{g}} - \mathbf{g}^0) \rightarrow \mathcal{N}(0, \Sigma), \quad T \rightarrow \infty. \quad (4.82)$$

Asymptotiquement, la déviation $\sqrt{T}(\hat{\mathbf{g}} - \mathbf{g}^0)$ converge vers une loi gaussienne multivariée de moyenne nulle et de matrice de covariance Σ quand T tend vers l'infini. Nous noterons que cette approche est valide dans le cas d'une estimation sans biais. Or, d'après le paragraphe précédent, nous ne pouvons pas affirmer que le biais est toujours nul car les trois conditions données par Donoho ne sont pas valides dans tous les cas. Néanmoins, nous proposons de poursuivre l'étude en supposant que le biais est quasi nul. Cette hypothèse est réaliste si l'estimation de la fonction score est correcte pour l'algorithme MANeige ou dans le cas des algorithmes VARIMAX ou de Godfrey, si l'*a priori* choisi est proche de la vraie distribution. Sinon, nous supposons que nous avons eu recours à l'orthogonalisation proposée par Donoho avec la nouvelle fonction ϕ_Y^{new} de l'équation (4.79) qui assure une estimation sans biais.

En utilisant les développements proposés dans le cadre des M -estimateurs [Hub77], Donoho [Don81] montre que la matrice de covariance peut s'approcher sous la forme suivante (démonstration en annexe D) :

$$\Sigma = \frac{V(\phi_Y, Y)}{\sigma^2} \mathbf{R}^{-1}, \quad (4.83)$$

où

$$V(\phi_Y, Y) = \frac{E[\phi_Y^2(Y)]}{(E[\phi_Y'(Y)])^2}, \quad (4.84)$$

et la matrice \mathbf{R} de taille $P \times P$ est de terme générique :

$$\mathbf{R}_{ij} = \frac{\sum_v w(i+v)w(j+v)}{\sum_t w(t)^2}. \quad (4.85)$$

La matrice \mathbf{R} est en fait la matrice de covariance des données $d(t)$. En effet, avec l'hypothèse que la réflectivité est iid, nous pouvons écrire que :

$$E[d(\tau+i)d(\tau+j)] = E \left[\sum_{t_1} w(\tau+i-t_1)r(t_1) \sum_{t_2} w(\tau+j-t_2)r(t_2) \right], \quad (4.86)$$

$$= \sum_{t_1} \sum_{t_2} w(\tau+i-t_1)w(\tau+j-t_2)E[r(t_1)r(t_2)], \quad (4.87)$$

$$= E[r(t)^2] \sum_{t_1} w(\tau+i-t_1)w(\tau+j-t_1), \quad (4.88)$$

$$= E[r(t)^2] \sum_t w(t+i)w(t+j). \quad (4.89)$$

Ainsi, \mathbf{R} est la matrice de covariance des données :

$$\mathbf{R}_{ij} = \frac{E[d(\tau+i)d(\tau+j)]}{E[d(\tau)^2]} = \frac{\sum_t w(t+i)w(t+j)}{\sum_t w(t)^2}. \quad (4.90)$$

L'expression (4.83) de la matrice de covariance se divise en deux termes. Le premier :

$$A(\phi_Y, y) = \frac{V(\phi_Y, Y)}{\sigma^2} \quad (4.91)$$

dépend uniquement de la réflectivité à estimer et du critère utilisé et le second \mathbf{R}^{-1} dépend uniquement du filtre direct (l'ondelette). Pour comparer les différents algorithmes nous proposons de nous intéresser seulement au terme $A(\phi_Y, Y)$ qui est appelé fonction de mérite de l'algorithme. Nous noterons que la technique de gradient relatif dans l'algorithme MANeige fera disparaître le terme \mathbf{R}^{-1} représentant la dépendance des performances au filtre direct grâce à la propriété d'équivariance.

4.5.2 Borne minimale

La fonction de mérite est minorée par la borne de Cramèr-Rao définie par :

$$A(\phi_Y, Y) > B(Y) = [I(Y)Var(Y) - 1]^{-1}, \quad (4.92)$$

où

$$I(Y) = E \left[-\frac{d^2}{dy^2} \log p_Y(y) \right] = E [\psi_Y'(y)] \quad (4.93)$$

est l'information de Fisher. En fait, dans cette définition, il apparaît la distribution et la fonction score de la sortie théorique Y^0 . Pour différencier ces grandeurs de celles de la sortie estimée Y , nous utiliserons des notations du type $\psi_Y^0(y)$. Avec cette notation et avec la définition de l'information de Fisher, la borne $B(Y)$ peut se récrire :

$$B(Y) = (E[\psi_Y^0'(y)]\sigma^2 - 1)^{-1}. \quad (4.94)$$

Dans cette équation, nous pouvons faire intervenir la fonction $\phi_Y^0(y)$ qui est la fonction ϕ_Y définie en (4.16) pour Y^0 la sortie théorique afin d'obtenir :

$$B(Y) = \frac{1}{\sigma^2 E[\phi_Y^0'(y)]}. \quad (4.95)$$

Cette borne permet d'obtenir une borne inférieure pour les performances des estimateurs. Nous pouvons montrer qu'elle est atteinte dans le cas où la fonction score est parfaitement estimée ou connue, soit $Y = Y^0$. Nous devons utiliser l'expression de $A(\phi_Y, y) = \frac{V(\phi_Y, Y)}{\sigma^2}$ avec $V(\phi_Y, Y)$ donnée en (4.84) et celle de la fonction $\phi_Y(y)$ en (4.16). Si la fonction score est parfaitement estimée, nous avons :

$$E[\phi_Y^2(Y)] = E \left[\psi_Y^2(y) - 2\frac{y\psi_Y(y)}{\sigma^2} + \frac{y^2}{\sigma^4} \right], \quad (4.96)$$

et en reprenant les relations (4.66) et (4.69) nous pouvons écrire :

$$E[\phi_Y^2(Y)] = E[\phi_Y'(Y)]. \quad (4.97)$$

Ainsi, nous avons montré que la fonction de mérite $A(\phi_Y, y)$ de l'algorithme MANege est égale à la borne $B(Y)$ dans le cas où la distribution est parfaitement connue ou estimée. Dans le cas particulier d'une distribution gaussienne, nous pouvons voir que la fonction $\phi_Y(y)$ est nulle (ou quasi) car la fonction score $\psi_Y(y) \approx \frac{y}{\sigma^2}$ donc la borne $B(Y)$ tend vers l'infini. Il est normal que l'estimation dans ce cas soit très difficile car nous utilisons un critère de non-gaussianité sur des signaux gaussiens.

4.5.3 Courbes de performances

Dans la figure 4.6 (issue de [Wal85]), nous représentons les performances par le terme $A(\phi_Y, y)$ pour différents algorithmes en fonction de γ , l'exposant de la loi GG suivie par la réflectivité. La figure 4.6(a) montre les performances des algorithmes de la famille VARIMAX :

- 1 : algorithme MED,
- 2 : VARIMAX avec $\alpha = 1.34$,
- 3 : VARIMAX avec $\alpha = 1.2$,
- 4 : VARIMAX avec $\alpha = 1$.

La borne minimale $B(Y)$ est également tracée. Pour la figure 4.6(b), nous traçons les performances de l'algorithme de Godfrey avec un *a priori* de Cauchy pour trois valeurs de m : $m=6$, $m=2.1$, $m=1.6$.

Nous remarquons sur la figure 4.6(a) que l'ensemble des courbes des fonctions de mérite se situe évidemment au-dessus de la borne $B(Y)$. Les algorithmes tendent vers les mêmes performances lorsque que le coefficient de gaussianité tend vers 2, soit une réflectivité gaussienne. Dans ce cas, toutes les méthodes sont équivalentes à la déconvolution prédictive car les statistiques d'ordre 2 décrivent totalement le signal. Enfin, l'algorithme MED est celui qui a les performances les moins bonnes, comme nous pouvions le prédire, car il utilise un *a priori* mal adapté aux signaux sur-gaussiens ($\gamma < 2$) auxquels on s'intéresse. En fait, à γ fixé, l'algorithme qui a les meilleures performances est celui qui choisit le coefficient de gaussianité α le plus proche de γ . Il est impossible de différencier les courbes pour les valeurs proches de 1. Ainsi, nous ne pouvons détecter d'éventuels croisements entre les courbes lors du passage de γ par une valeur de α choisie par un des algorithmes. En revanche, pour les faibles valeurs de γ , nous remarquons que les performances décroissent lorsque que le coefficient de gaussianité de l'*a priori* augmente. De plus, elles s'écartent de plus en plus de la borne $B(Y)$.

D'après Walden [Wal85], pour les séquences de réflectivité en imagerie sismique, le coefficient de gaussianité est compris entre 0.6 et 1.5. Suivant la figure 4.6(a), si nous omettons l'algorithme MED, le choix de l'exposant de gaussianité de l'algorithme est peu déterminant pour les performances. Néanmoins, nous devons choisir un *a priori* sur-gaussien avec α tendant vers 1. Mais ce choix va à l'opposé des contraintes de consistance de l'algorithme et du biais d'estimation décrits précédemment.

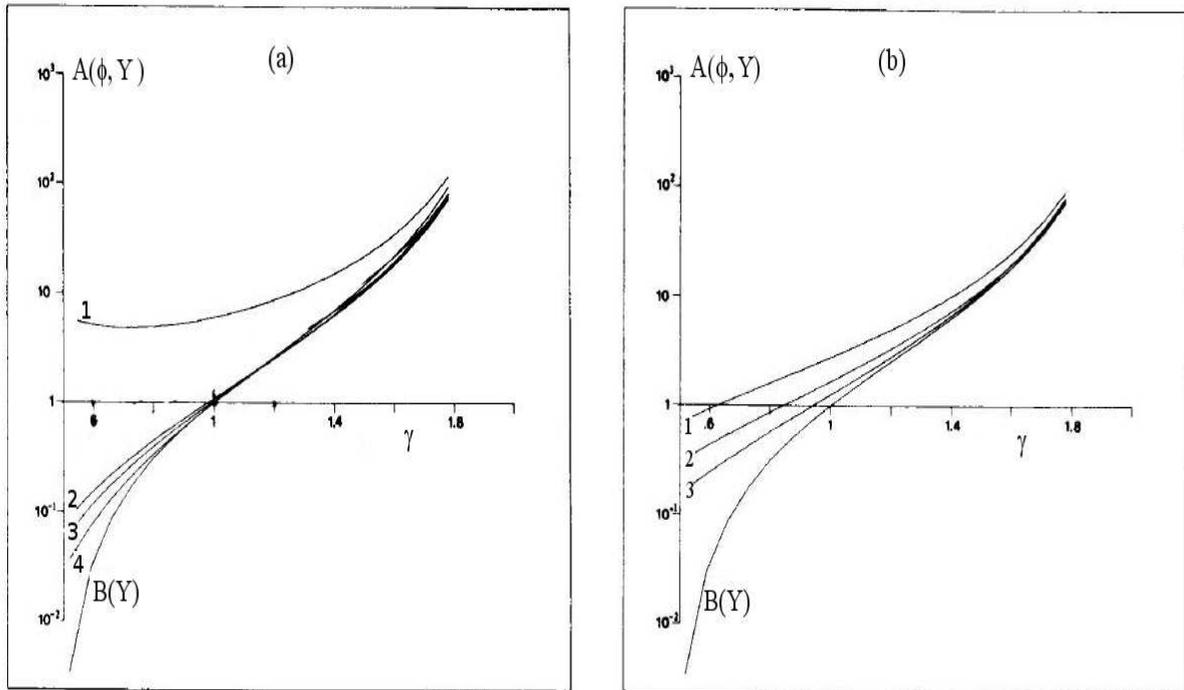


FIG. 4.6 – Performances des algorithmes représentées par la fonction de mérite $A(\phi_Y, y)$ en fonction de l'exposant γ de la loi GG suivie par la réflectivité. (a) comparaison des algorithmes de la famille VARIMAX : 1 : algorithme MED soit $\alpha = 4$, 2 : $\alpha = 1.34$, 3 : $\alpha = 1.2$, 4 : VARIMAX avec $\alpha = 1$. (b) algorithme de Godfrey avec un *a priori* de Cauchy : 1 : $m=6$, 2 : $m=2.1$, 3 : $m=1.6$. $B(Y)$ est la borne inférieure de toutes les performances (d'après Walden [Wal85]).

Dans la section 4.4, nous avons montré que le choix de $\alpha = 4$ permet de s'assurer de la concavité du critère, ce qui est synonyme d'une convergence certaine de l'algorithme. En revanche, pour les *a priori* sur-gaussiens, nous ne pouvons rien assurer. En définitive, le choix de α doit se faire suivant un compromis entre une estimation précise et une convergence certaine.

Pour les *a priori* de Cauchy dans l'algorithme de Godfrey, la figure 4.6(b) montre que les performances sont les meilleures pour $m = 1.6$ qui correspond à l'*a priori* le plus sur-gaussien des 3. Nous noterons que les performances sont cependant inférieures à celles de l'algorithme VARIMAX avec un *a priori* sur-gaussien (Figure 4.6(a)).

À la figure 4.7, nous représentons les mêmes courbes qu'à la figure 4.6 mais pour une distribution de réflectivité de Cauchy de paramètre m . Ainsi, nous traçons les performances de la fonction de mérite $A(\phi_Y, Y)$ en fonction du paramètre m de la loi de Cauchy suivie par la réflectivité. La figure 4.7(a) montre les performances pour les algorithmes utilisant un *a priori* GG, et la figure 4.7(b) pour l'algorithme de Godfrey utilisant l'*a priori* de Cauchy pour différentes valeurs du paramètre de la loi. La borne $B(Y)$ est aussi représentée.

D'après la figure 4.7, nous remarquons que la borne $B(Y)$ est supérieure pour les signaux très sur-gaussiens suivant une distribution de Cauchy (m tendant vers 1.5) en comparaison à ceux suivant une loi GG (α tendant vers 1) (Figure 4.6). Sur la figure 4.7(b), pour les *a priori* de Cauchy, nous notons que c'est l'*a priori* avec le coefficient le plus proche de la bonne valeur, qui donne les meilleures performances. En effet, pour m grand, c'est la courbe 3 avec $m = 6$ qui donne les meilleures performances. Puis, lorsqu'on s'approche des valeurs de 1.6 et 2.1, nous notons des croisements des courbes de performance, car le coefficient de l'*a priori* le plus proche de la vraie valeur de m change. Le croisement le plus net se situe autour de l'abscisse $m = 3.3$ entre les courbes 2 et 3 caractéristiques de $m = 2.1$ et $m = 6$. Enfin, pour les valeurs de m faibles, l'*a priori* de $m = 1.6$ (très sur-gaussien) donne les meilleures performances et approche de très près la borne $B(Y)$.

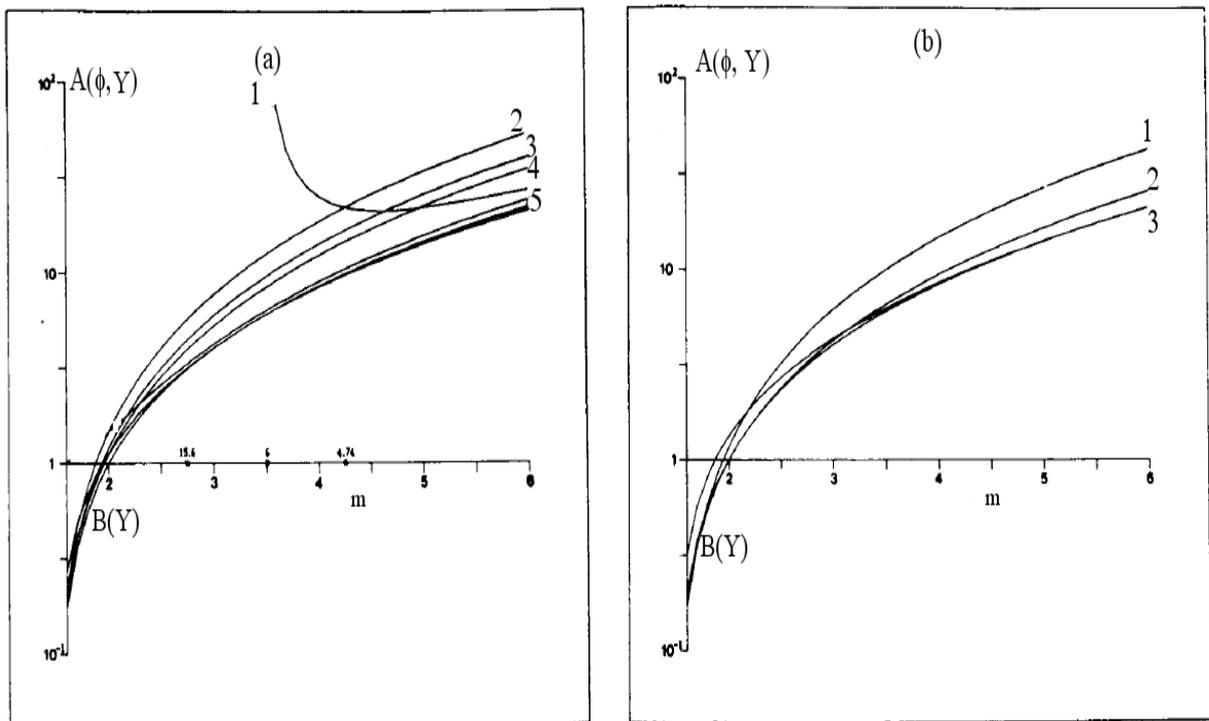


FIG. 4.7 – Performances des algorithmes représentées par la fonction mérite $A(\phi_Y, y)$ en fonction de l'exposant γ de la loi GG suivie par la réflectivité. (a) comparaison des algorithmes de la famille VARIMAX : 1 : algorithme MED soit $\alpha = 4$, 2 : $\alpha = 1.34$, 3 : $\alpha = 1.2$, 4 : VARIMAX avec $\alpha = 1$. (b) algorithme de Godfrey avec un *a priori* de Cauchy : 1 : $m=1.6$, 2 : $m=2.1$, 3 : $m=6$. $B(Y)$ est la borne inférieure de toutes les performances (D'après Walden [Wal85]).

Si nous réalisons le traitement avec les algorithmes utilisant un *a priori* GG pour une réflectivité suivant en réalité une loi de Cauchy de paramètre m (figure 4.7(a)), nous notons que l'algorithme MED donne les plus faibles performances lorsque $m < 4$. Nous pouvons justifier cette propriété par le fait que l'*a priori* est mal choisi car il est sous-gaussien alors

que les données suivent une loi sur-gaussienne. Ainsi, pour les grandes valeurs de m , à m fixé, la fonction de mérite diminue lorsque le coefficient de gaussianité décroît. Pourtant le signal de sortie reste sur-gaussien, mais le choix d'un *a priori* sous-gaussien comme pour l'algorithme MED est meilleur qu'un *a priori* sur-gaussien. Pour les faibles valeurs de m , le classement, en terme de performance, s'inverse et les fonctions de mérite des algorithmes tendent vers la borne $B(Y)$.

A propos des figures 4.6 et 4.7 concernant les tracés des fonctions de mérite de l'algorithme VARIMAX (pour différentes valeurs de α) et de l'algorithme de Godfrey (pour différentes valeurs de m) pour des distributions *a priori* GG (figure 4.6) et de Cauchy (figure 4.7), nous pouvons conclure que les meilleures fonctions de mérite des algorithmes approchent assez bien la borne $B(Y)$. Ensuite, le choix de la distribution *a priori* la plus proche possible de la distribution réelle permet d'obtenir les meilleures performances. Ainsi, pour des signaux sur-gaussiens, l'algorithme MED donne des résultats de moins bonne qualité que les autres algorithmes de type VARIMAX utilisant un *a priori* sur-gaussien. De même pour l'algorithme de Godfrey, le choix d'une loi de Cauchy avec $m = 1.6$ permet d'obtenir les meilleures performances. Nous rappelons que ces calculs sont réalisés en supposant que l'estimation se fait sans biais et que la convergence est assurée. Or nous avons vu dans la sous-section précédente que ces hypothèses sont soumises aux conditions (R1), (R2) et (R3).

Dans la section suivante, nous proposons de nous intéresser à une caractérisation plus expérimentale des performances des différents algorithmes pour plusieurs types de signaux de réflectivité.

4.6 Comparaison des algorithmes

Pour réaliser la comparaison des algorithmes et pour dégager les avantages et inconvénients de chacun, nous débuterons en traçant les différentes fonctions scores utilisées selon les *a priori* choisis. Nous analyserons leurs formes et nous conclurons sur leurs aptitudes à décrire les signaux sismiques. Enfin, nous réaliserons des comparaisons chiffrées sur des signaux simulés.

4.6.1 Fonction score

Pour faire une synthèse de l'ensemble des méthodes, nous proposons de tracer (figure 4.8), les fonctions score *a priori* utilisées par les différents algorithmes : MED, VARIMAX et l'algorithme de Godfrey. Les fonctions score sont normalisées pour que leurs maxima valent 1. De plus, les fonctions étant impaires, nous ne représentons que la partie positive.

Sur la figure 4.8(a), nous traçons les fonctions scores de lois gaussiennes généralisées qui sont implicitement utilisées par l'algorithme VARIMAX, pour les valeurs $\alpha = 1.1$, $\alpha = 1.6$, $\alpha = 2$, $\alpha = 4$. La valeur $\alpha = 1.1$ correspond à un *a priori* presque laplacien qui est la valeur minimale de α pour des raisons de stabilité déjà mentionnées. La valeur $\alpha = 2$ correspond à un *a priori* gaussien, la fonction score est une droite, ce cas n'est jamais utilisé mais en fait il permet de séparer les *a priori* sous-gaussiens ($\alpha > 2$) et sur-gaussiens ($\alpha < 2$). Pour

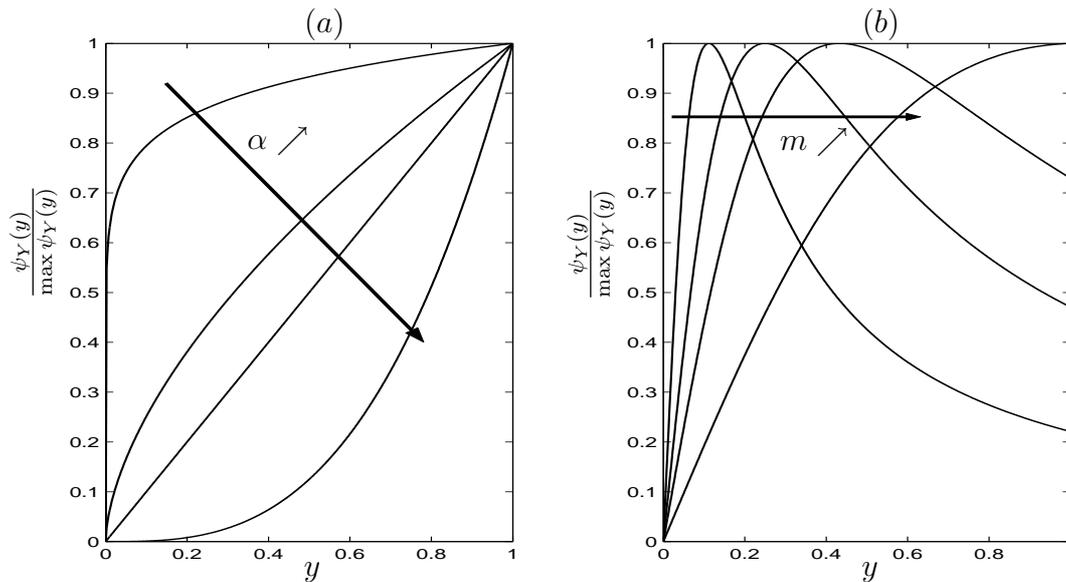


FIG. 4.8 – Fonctions score *a priori* utilisées par les algorithmes : (a) Algorithme VARIMAX : loi GG de paramètre $\alpha = 1.1, \alpha = 1.6, \alpha = 2, \alpha = 4$ (algorithme MED), (b) algorithme de Godfrey : loi de Cauchy de paramètre $m = 1.6, m = 2, m = 3, m = 10$.

$\alpha = 4$, nous trouvons le cas particulier de l'algorithme MED.

L'allure de la fonction score permet de juger de l'importance relative de chaque échantillon dans l'estimation du gradient en fonction de sa valeur. Pour l'algorithme MED, l'*a priori* avec $\alpha = 4$ privilégie les grandes valeurs des échantillons au détriment des plus faibles. Au contraire, pour les *a priori* sur-gaussiens, les échantillons ont une importance quasi comparable, en particulier dans le cas laplacien ($\alpha = 1$ non représenté ici) où la fonction score est constante. Si nous utilisons des distributions avec un coefficient de gaussianité inférieur à 1, nous obtenons une singularité en zéro qui pose des problèmes de stabilité comme nous l'avons déjà évoqué. Dans ce cas, la fonction score est plus grande pour les échantillons de faibles valeurs ou du moins passe par un maximum. Nous pouvons rapprocher cela de l'allure des fonctions score théoriques tracées à la figure 2.4 pour des signaux Bernoulli-laplacien en (a) et Bernoulli-gaussien en (b) mélangés à du bruit gaussien. Nous avons une fonction score théorique passant par un maximum avant de décroître pour les grandes valeurs.

En choisissant une distribution *a priori* de Cauchy, nous obtenons une fonction score avec une allure plus proche de celle tracée pour des signaux sismiques sur la figure 2.4(a) et (b). Sur la figure 4.8(b), nous avons tracé la fonction score pour une loi de Cauchy de paramètre $m = 1.6, m = 2, m = 3, m = 10$. Pour rappel, le paramètre de la loi de Cauchy doit vérifier $m > 1.5$. Nous remarquons que pour $m = 10$, la fonction score tend vers une droite. En effet pour m tendant vers l'infini, la loi de Cauchy tend vers une loi gaussienne. Pour m décroissant de 3 à 1.6, l'abscisse du maximum de la fonction score se rapproche de zéro, et la valeur pour y grand diminue. Cependant, nous noterons une assez grande sensibilité au paramètre m , car le passage de $m = 1.6$ à $m = 2$ donne un grand changement de l'allure de la fonction score.

4.6.2 Résultats de simulations

Dans cette partie, nous proposons d'illustrer les résultats théoriques précédemment développés à l'aide des mesures de performances moyennes dans le cadre de simulations. Il est difficile de mesurer l'influence de tous les paramètres car nous pouvons agir sur :

- l'ondelette (modèle, longueur, type),
- la distribution de la réflectivité (sa forme, le nombre de réflecteurs pour les lois discrètes),
- la longueur du filtre de déconvolution.

Pour simplifier, nous utiliserons une ondelette autorégressive définie par le vecteur de paramètres : $[1, 0.4, 0.5, 0.45, 0.4, 0.1]$. Ainsi, le modèle MA du filtre déconvolveur utilisé par tous les algorithmes basés sur la sur-gaussianité sera parfaitement adapté. Le tableau 4.1 mesure l'erreur quadratique moyenne sur 200 réalisations en dB de l'estimation de la sortie pour 6 algorithmes de déconvolution par la sur-gaussianité :

- l'algorithme MANege avec une estimation de la fonction score par une méthode à noyaux,
- l'algorithme VARIMAX avec $\alpha = 4$ donc équivalent à l'algorithme MED,
- l'algorithme VARIMAX avec $\alpha = 1.6$,
- l'algorithme VARIMAX avec $\alpha = 1.1$,
- l'algorithme de Godfrey avec $m = 1.6$,
- l'algorithme de Godfrey avec m adaptatif.

Nous utiliserons des données synthétiques de 400 échantillons répondant à des modèles Bernoulli-laplacien (BL), Bernoulli-gaussien (BG), laplacien. Nous ferons varier la longueur P du filtre de déconvolution, elle prendra les valeurs 9, 20, ou 30. Les différents cas seront numérotés pour faciliter l'analyse des résultats.

Cas	Réflectivité	P	MANege Noyaux	Godfrey		Gray		
				$m = 1.6$	adaptatif	$\alpha = 1.1$	$\alpha = 1.6$	MED : $\alpha = 4$
①	BG $\lambda = 1/20$	9	-25.1	-25.8	-23.5	-24.6	-19.7	-17.7
②	BG $\lambda = 1/20$	20	-22.8	-24.3	-20.1	-22.1	-16.2	-12.1
③	Laplacienne	9	-16.7	-16.8	-16.3	-17.2	-16.6	-6.0
④	Laplacienne	20	-12.8	-11.8	-12.6	-13.2	-12.8	-1.9
⑤	Laplacienne	30	-10.8	-8.6	-10.7	-10.8	-10.8	-0.7
⑥	BL $\lambda = 1/20$	9	-24.1	-24.9	-23.2	-23.5	-19.9	-19.8
⑦	BL $\lambda = 1/20$	20	-21.3	-22.8	-19.65	-19.9	-15.7	-12.5
⑧	BL $\lambda = 1/5$	9	-24.1	-24.9	-23.2	-23.5	-19.9	-19.8

TAB. 4.1 – Erreur quadratique (en dB) moyenne sur 200 réalisations de l'estimation de la réflectivité par les algorithmes utilisant la sur-gaussianité en fonction de la longueur du filtre et de la distribution de réflectivité.

Dans un premier temps, nous avons choisi des réflectivités sur-gaussiennes pour se placer dans un cadre assez proche des applications sismiques. Ce choix de distributions sur-gaussiennes implique des résultats faibles pour l'algorithme MED. Il est le seul à faire le choix d'un *a priori* sous-gaussien. Les performances sont acceptables pour les distributions

BG et BL (cas ①, ②, ⑥, ⑦ et ⑧), en revanche, l'EQM est très grande pour les distributions laplaciennes (cas ③, ④ et ⑤), avec une valeur tendant même vers 0 dB.

Pour la sensibilité à la surestimation du nombre de coefficients du filtre déconvolveur (7 théoriquement), nous remarquons que l'algorithme MED est le plus sensible. Pour l'algorithme MED, le passage de 9 à 20 coefficients entraîne une baisse de performance de $5.6dB$ (cas ① à ②) au lieu de $3.5 dB$ au maximum pour les autres méthodes pour les mêmes cas. Pour les lois laplaciennes, le passage de 9 à 20 coefficients (cas ③ et ④) induit la plus grande hausse l'EQM pour l'algorithme MED, qui affichait déjà dans le cas ③ à 9 coefficients les plus mauvaises performances. L'augmentation de 20 à 30 coefficients (cas ④ et ⑤) perturbe tous les algorithmes. Les résultats de l'algorithme MED ne peuvent pas être analysés car les performances sont déjà trop faibles dans le cas ④. Dans tous les cas, nous pouvons dire que tous les algorithmes excepté MED donnent des résultats très proches quelle que soit la distribution de la réflectivité et le nombre de coefficients. Nous pourrions écarter de ce groupe l'algorithme VARIMAX avec $\alpha = 1.6$ qui, dans les cas ②, ⑥, ⑦ et ⑧, a des performances légèrement inférieures de quelques dB aux autres algorithmes. Pour les distributions moins sur-gaussiennes (laplacienne dans les cas ③, ④ et ⑤ et BL avec $\lambda = 1/5$ pour le cas ⑧) nous observons un nivellement des performances des algorithmes par rapport aux autres cas de signaux très sur-gaussiens.

En conclusion de ces expériences, le choix d'une bonne loi *a priori* permet de s'assurer de bonnes performances, même si nous surestimons la longueur du filtre. En revanche, un *a priori* mal choisi avec un trop grand nombre de coefficients tend à faire augmenter largement l'EQM de sortie. Ces résultats permettent de tirer des conclusions en accord avec les caractérisations théoriques précédentes, notamment celles liées au calcul de la covariance asymptotique. Néanmoins, pour des données non bruitées répondant parfaitement au modèle, il est difficile de mesurer la stabilité des algorithmes. En effet, dans l'expérience précédente, nous avons fixé un grand nombre d'itérations (200) pour les différents algorithmes, mais rien ne mesure dans cette expérience la vitesse de convergence.

4.7 Résumé et conclusions

Le tableau 4.2 résume les différents algorithmes présentés dans la section 4.2, en spécifiant le principe de base de l'algorithme, et les principaux avantages et inconvénients des algorithmes. Nous rappelons aussi les *a priori* sur les distributions de la sortie estimée en lien avec l'algorithme MANege.

TAB. 4.2 – Résumé des méthodes de déconvolution utilisant la sur-gaussianité.

Méthode	Principe	<i>a priori</i>	Avantages	Inconvénients
MANege	Maximisation de la néguentropie Gradient relatif	Aucun	Adaptatif non paramétrique /ddp Aucun <i>a priori</i> sur les ddp Equivariant Identifiabilité et concavité R_2 Estimation non biaisée R_3	Estimation de la fonction score
MED	Maximise le kurtosis Annule le gradient	Gaussien généralisé de coeff. 4	Simplicité Convergence rapide Concavité (R_2) assurée \forall ddp Identifiabilité du système prouvée	<i>A priori</i> non adapté Atténue les petits réflecteurs
VARIMAX	Minimise la variation de norme Annule le gradient	Gaussien généralisé d'exposant α	Simplicité Meilleure préservation des petits réflecteurs que MED	Optimisation non certaine Si $\alpha < 1$ instable Concavité non assurée Identifiabilité non prouvée
Transformation Exponentielle	Transformation exponentielle de la sortie VARIMAX avec $\alpha = 1$???	Simplicité Meilleure préservation des petits réflecteurs que MED	Aucune justification statistique Réglage d'un paramètre de non-linéarité
Godfrey	Maximisation de la néguentropie Annule le gradient (Cf MED) Adapte le paramètre m avec le rapport des moments	Cauchy de paramètre m	Adaptatif paramétrique Fonction score sans singularité	Estimation de m sensible Optimisation défailante

En définitive, dans ce chapitre, nous avons proposé de baser les algorithmes de déconvolution sur l'hypothèse de sur-gaussianité de la réflectivité au lieu de la blancheur (chapitres 2 et 3). Nous noterons cependant que l'hypothèse de blancheur est aussi nécessaire dans ce cas pour justifier l'identifiabilité du système et elle est mise en évidence dans le gradient de l'algorithme MANege. La sur-gaussianité constitue une hypothèse discriminante entre le bruit et la réflectivité. Elle permet une approche alternative pour la sélection des statistiques utilisées dans la déconvolution aveugle. En réalité, dans le gradient de la négentropie, nous avons clairement une mesure de blancheur. Mais, les statistiques choisies comportent peu de statistiques d'ordre 2. Cette approche est intéressante car le bruit a une influence largement moindre pour la sélection des statistiques dans l'algorithme MAMV avec la mesure d'indépendance au sens large avec le taux d'information mutuelle.

Nous avons proposé l'algorithme MANege maximisant la négentropie par rapport à la réponse impulsionnelle du filtre déconvolveur $g(t)$. Par la suite, nous avons montré que l'algorithme MANege permettait de créer un cadre généralisant pour de nombreux algorithmes proposés dans la littérature. Nous avons débuté par l'algorithme le plus connu (MED) qui maximise le kurtosis de sortie. Cet algorithme MED est identique à l'algorithme MANege avec une distribution *a priori* de sortie gaussienne généralisée de coefficient 4. Cependant, nous noterons que le parallèle a révélé le passage d'une minimisation à une maximisation lors du choix d'un *a priori* sous-gaussien au lieu de sur-gaussien. Puis, nous avons traité de la famille des algorithmes VARIMAX qui choisissent un *a priori* GG avec un exposant quelconque. L'algorithme de Godfrey maximisant l'entropie relative est similaire à l'algorithme MANeg. Godfrey fait le choix d'une distribution *a priori* de Cauchy et propose un algorithme semi-adaptatif paramétrique. Il met à jour en fonction de la distribution de la sortie le paramètre de la loi de Cauchy.

Ensuite, en nous appuyant sur les travaux de Donoho [Don81], nous avons détaillé les conditions assurant l'identifiabilité du système, la concavité et l'estimation sans biais du filtre déconvolveur. Nous rappelons que seuls l'algorithme MANege et les algorithmes VARIMAX utilisant un *a priori* sous-gaussien (comme MED) permettent d'assurer l'identifiabilité du système. Pour les autres, aucune démonstration ne permet de conclure positivement ou négativement sur l'identifiabilité du système. Pour la concavité et l'estimation non biaisée, les algorithmes doivent remplir un ensemble de trois conditions. Nous avons montré que seule la connaissance exacte de la distribution du signal de sortie permet de remplir ces trois conditions. Cependant, l'algorithme MED vérifie la condition de concavité quelque soit la distribution du signal, ce qui le rend intéressant. D'autre part, si le coefficient α est inférieur à 1, les algorithmes de type VARIMAX conduisent à des critères non concaves. Pour $1 < \alpha < 2$, nous pouvons trouver des configurations ne vérifiant pas la condition de concavité. Ainsi, pour faire le choix d'un *a priori* sur-gaussien, nous devons prendre $\alpha < 2$ mais ne pas trop s'approcher de 1 pour éviter les problèmes de stabilité de l'algorithme. Dans la partie traitant des performances théoriques, nous avons aussi calculé la variance asymptotique des différents algorithmes en fonction de la distribution du signal. Nous avons observé que les performances asymptotiques s'améliorent si la distribution *a priori* approchait la vraie distribution du signal. Ainsi, l'algorithme MED aura des performances inférieures à l'algorithme VARIMAX avec $\alpha < 2$ pour la déconvolution

de données sur-gaussiennes, mais il compense par une meilleure concavité, ce qui, en pratique, rend la convergence de l'algorithme plus certaine. Enfin, des mesures de performance en terme d'erreur quadratique moyenne dans un cadre de simulations apportent les mêmes conclusions que les calculs théoriques, en montrant l'importance du choix d'un bon *a priori*.

Dans ce chapitre, nous avons caractérisé les différents algorithmes de déconvolution basés sur la sur-gaussianité en fonction de la distribution de la réflectivité. La première idée pour ces développements était de relâcher la contrainte de blancheur de la réflectivité et la sur-gaussianité s'est posée comme une hypothèse candidate intéressante, notamment par rapport à la présence du bruit additif (supposé généralement) gaussien. Dans le chapitre suivant, nous proposons de faire une étude de l'influence du bruit sur l'estimation du filtre déconvolveur. En effet, il apparaît que la sortie estimée sera satisfaisante seulement si le signal est quasi blanc ou très sur-gaussien, et, si le bruit additif présent sur les données n'est pas amplifié. Une méthode de déconvolution aveugle pouvant conduire à de bonnes performances pour des données réelles doit réaliser ce compromis entre une bonne déconvolution et une amplification raisonnable du bruit.

Chapitre 5

Robustesse en contexte bruité

Sommaire

5.1	Influence du bruit sur les critères	145
5.1.1	Algorithme MED : maximisation du kurtosis	146
5.1.2	Algorithme MANege : maximisation de la néguentropie	147
5.1.3	Algorithme FBD : minimisation du taux d'information mutuelle.	151
5.2	Expérimentations	152
5.2.1	Signaux simulés	152
5.2.2	Données réelles : explosions séismo-volcaniques	155
5.2.3	Données réelles : explosions sous-marines	158
5.2.4	Données réelles : profil sismique	159
5.3	Conclusions	163

Dès le chapitre 3, la présence du bruit additif sur les données nous avait intéressée. Nous avons constaté qu'en contexte bruité, les performances de l'algorithme MAMV étaient équivalentes aux algorithmes utilisant les statistiques d'ordre 2. Ainsi, dans le chapitre 3, nous avons proposé l'algorithme FBD qui intégrait une régularisation prenant en compte le bruit additif. Cette régularisation tendait à augmenter le RSB de sortie toutefois la sélection des statistiques d'ordre supérieur utilisées n'était pas beaucoup modifiée par rapport à l'algorithme MAMV. Dans le chapitre 4, compte tenu de l'ambiguïté des hypothèses de blancheur du bruit et de la réflectivité qui limite certainement les performances de la déconvolution, nous nous sommes intéressés à l'hypothèse de sur-gaussianité pour la sélection des statistiques intervenant dans la déconvolution.

Dans ce chapitre, nous allons comparer la robustesse vis-à-vis du bruit des différents algorithmes. En effet, la sortie Y s'écrit comme $g \star w \star R + g \star N$. Ainsi, la sortie estimée se divise en deux termes. Le premier $Y_R = g \star w \star R$ est relatif à la réflectivité à estimer et le second $Y_N = g \star N$ est le bruit sur la sortie. En l'absence de bruit, sous hypothèse d'identifiabilité, nous avons $g = w^{-1}$ soit $Y = R$. Mais, la présence du bruit dégrade l'estimation du filtre déconvolveur et donne aussi un bruit sur la sortie $g \star N$. Pour obtenir une méthode performante, nous devons réaliser le compromis entre l'estimation idéale de la réflectivité et un bruit additif de sortie de variance peu élevée. Par exemple, ce compromis est bien traité dans le filtrage de Wiener.

Nous commencerons par écrire des développements théoriques des critères des algorithmes MED, MANeGe et FBD afin d'étudier le comportement des critères en présence de bruit. Nous nous attacherons plus particulièrement à mettre en évidence la tendance des critères à effectuer une régularisation pour limiter l'amplification du bruit en sortie. Nous nous placerons dans le cadre réaliste d'une ondelette passe-bande et d'un bruit additif blanc. Nous étudierons le gain du filtre déconvolveur dans la bande atténuée de l'ondelette contenant uniquement du bruit appelé l'espace nul¹. Ensuite, une expérience avec des signaux simulés permettra d'illustrer ces développements théoriques. Dans la dernière section, nous présenterons les résultats de la déconvolution par ces trois algorithmes sur des données réelles d'explosions séismo-volcaniques et sous-marines. Nous finirons par la déconvolution d'un profil sismique réel. Ce travail a été initié lors de mon séjour à Leeds puis a été réalisé en collaboration avec Dinh-Tuan Pham pour la partie théorique et a fait l'objet de deux communications [LVDBMJ05, LP06].

5.1 Influence du bruit sur les critères

Dans cette section, nous effectuons le développement des algorithmes MED, MANeGe et FBD des développements de leurs critères respectifs en présence de bruit. En effet, nous souhaitons isoler les effets de la partie bruit Y_N sur le critère. Nous analyserons si un compromis apparaît au sein du critère entre la déconvolution et l'amplification du bruit. En outre, nous débuterons par l'algorithme MED, puis, nous poursuivrons par les algorithmes MANeGe et FBD.

¹"null space" en anglais.

5.1.1 Algorithme MED : maximisation du kurtosis

Pour l'algorithme MED maximisant le kurtosis du signal de sortie, le critère est souvent dit insensible au bruit additif gaussien car il fait appel aux statistiques d'ordre 4. Cet argument est faux car la variance de Y au dénominateur dépend de la variance de Y_N .

Nous pouvons écrire le kurtosis de la sortie estimée Y en fonction du kurtosis de la partie "signal" $Y_R = g \star w \star R$. Nous revenons à la définition du kurtosis (1.16) dans le chapitre 1 comme le rapport du cumulants d'ordre 4 κ_{4Y} et le cumulants (ou moment) d'ordre 2 au carré. L'indépendance entre le bruit N et la réflectivité R entraîne pour le cumulants d'ordre 4 de la sortie Y :

$$\kappa_{4Y} = \kappa_{4Y_R} + \kappa_{4Y_N}. \quad (5.1)$$

Le bruit N est supposé gaussien, ce qui implique que $Y_N = g \star N$ est gaussien alors κ_{4Y_N} son cumulants d'ordre 4 est nul. Finalement, le kurtosis \mathcal{K}_{4Y} de la sortie Y s'exprime en fonction du kurtosis \mathcal{K}_{4Y_R} de la partie signal par :

$$\mathcal{K}_{4Y} = \mathcal{K}_{4Y_R} \left[\frac{\text{var}(Y_R)}{\text{var}(Y)} \right]^2, \quad (5.2)$$

où $\text{var}(\cdot)$ représente la variance. Pour poursuivre, nous pouvons développer le rapport des variances à l'aide des gains du filtre déconvolveur $G(\nu)$ et de l'ondelette $W(\nu)$. La réflectivité R étant supposée de variance unitaire, la variance de $Y_R = g \star w \star R$ est :

$$\text{var}(Y_R) = \int_0^1 |G(\nu)W(\nu)|^2 d\nu, \quad (5.3)$$

où ν est la fréquence réduite. Pour la variance de Y nous avons :

$$\text{var}(Y) = \int_0^1 |G(\nu)|^2 [|W(\nu)|^2 + \sigma_N^2] d\nu, \quad (5.4)$$

avec σ_N^2 la variance du bruit N . En reprenant (5.3) et (5.4), dans l'expression (5.2) du kurtosis de sortie, nous obtenons :

$$\mathcal{K}_{4Y} = \mathcal{K}_{4Y_R} \left\{ \int_0^1 \Delta(\nu) \frac{|W(\nu)|^2}{|W(\nu)|^2 + \sigma_N^2} d\nu \right\}^2, \quad (5.5)$$

où

$$\Delta(\nu) = \frac{|G(\nu)|^2 [|W(\nu)|^2 + \sigma_N^2]}{\int_0^1 |G(\lambda)|^2 [|W(\lambda)|^2 + \sigma_n^2] d\lambda}. \quad (5.6)$$

L'expression (5.5) montre que le kurtosis de la sortie Y peut s'écrire comme le produit de deux facteurs. Le premier est indépendant du bruit N , il s'exprime comme le kurtosis de la partie signal $Y_R = g \star w \star R$, qui représente la qualité de l'estimation de la réflectivité (soit l'adéquation aux données). Le second facteur concentre toute l'influence du bruit sur le critère de l'algorithme MED. La maximisation du kurtosis de la sortie Y est équivalente à maximiser le produit de ces deux facteurs, ce qui entraîne un compromis entre les deux. Le premier facteur tend à estimer le filtre déconvolveur $g = w^{-1}$, alors que le

second joue le rôle de régularisation vis-à-vis du bruit. En effet, $\Delta(\nu)$ peut être interprété comme un poids de barycentre car $\int_0^1 \Delta(\nu) d\nu = 1$. Ainsi, maximiser le second facteur de (5.5) a pour effet de concentrer tous les poids $\Delta(\nu)$ autour des fréquences ν pour lesquelles $|W(\nu)|^2 / [|W(\nu)|^2 + \sigma_N^2]$ est maximum, *i.e.* $|W(\nu)|$ est maximum. Ce facteur a pour effet de faire tendre $|G(\nu)|$ à zéro pour toutes les valeurs ν sauf celles pour lesquelles $|W(\nu)|$ est maximum. En définitive, le gain du filtre déconvolveur tend vers 0 dans l'espace nul. L'algorithme MED comporte une régularisation naturelle vis-à-vis du bruit additif gaussien.

Pour les algorithmes de la famille VARIMAX, nous ne sommes pas encore parvenus à écrire une relation simple similaire à (5.5). Pour les algorithmes utilisant un coefficient α entier pair, nous pouvons obtenir la même écriture que (5.5) en nous appuyant sur les propriétés des cumulants d'ordre supérieur. Mais ce résultat n'est pas très intéressant car nous n'employons jamais ces valeurs de α . Il serait plus approprié d'obtenir une caractérisation pour $1 < \alpha < 2$. Pour $\alpha = 1$, soit un *a priori* laplacien, l'écriture de $E[|Y_R + Y_N|]$ est possible. En effet $|Y_R + Y_N| = -Y_R + Y_N + 2 \max(Y_R, -Y_N)$ et la loi du maximum de Y_R et Y_N s'écrit avec les fonctions de répartition de Y_R et Y_N . Cependant, il est difficile d'analyser la formule finale.

5.1.2 Algorithme MANeige : maximisation de la négentropie

Puisque $Y = Y_R + Y_N$, et que les variables aléatoires de cette somme sont indépendantes et que la seconde est gaussienne, la négentropie vérifie :

$$H^-(Y) < H^-(Y_R). \quad (5.7)$$

De plus, la négentropie de la sortie estimée $H^-(Y)$ sera d'autant plus grande que la variance du bruit Y_N sera faible. Ainsi, maximiser la négentropie de la sortie Y limite l'augmentation de la variance du bruit sur la sortie Y_N .

Pour l'algorithme MANeige maximisant la négentropie de la sortie, nous pourrions nous limiter à l'inégalité (5.7) pour justifier que nous avons une limitation de la variance du bruit sur le signal estimé de sortie. En fait, nous pouvons écrire des équations plus précises. Ces développements s'inspirent de l'écriture de l'entropie d'une variable aléatoire polluée par une autre dans le cadre de la séparation de sources étudiées par Pham [Pha05]. Dans le cadre de la séparation de sources, Pham écrit le développement au second ordre de l'entropie d'une variable aléatoire Y , qui est perturbée par une combinaison linéaire des composantes d'un vecteur \mathbf{Z} supposées indépendantes de Y :

$$H(Y + \mathbf{a}^T \mathbf{Z}) = H(Y) + \frac{1}{2} E[\psi'_Y(Y)] \text{var}(\mathbf{a}^T \mathbf{Z}) + o(\|\mathbf{a}\|^2), \quad (5.8)$$

où \mathbf{a} représente le vecteur des coefficients affectés à chaque variable aléatoire du vecteur \mathbf{Z} et $\|\cdot\|$ est la norme euclidienne. ψ_Y est la fonction score. Ce développement est valable si $\|\mathbf{a}\|^2$ est négligeable par rapport à $E[Y - E[Y]]^2 / E[\|\mathbf{Z} - E[\mathbf{Z}]\|^2]$, *i.e.* si la variation $\mathbf{a}^T \mathbf{Z}$ est de variance très faible par rapport au processus Y . Ainsi, la fonction $o(\|\mathbf{a}\|^2)$ signifie que ce terme est négligeable par rapport à $E[Y - E[Y]]^2 / E[\|\mathbf{Z} - E[\mathbf{Z}]\|^2]$. En définitive, nous pouvons remplacer la fonction $o(\|\mathbf{a}\|^2)$ par une fonction $o\left(\frac{\text{var}(\mathbf{a}^T \mathbf{Z})}{\text{var}(Y)}\right)$.

Nous pouvons voir $Y_N = g \star N$ comme un mélange des processus issus du décalage de N qui sont indépendants de Y . Ainsi, en étendant le développement (5.8), nous avons au second ordre l'approximation suivante :

$$H(Y_R + Y_N) = H(Y_R) + \frac{1}{2} E [\psi'_{Y_R}(Y)] \text{var}(Y_N) + o\left(\frac{\text{var}(Y_N)}{\text{var}(Y_R)}\right). \quad (5.9)$$

Ensuite, nous souhaitons écrire $H^-(Y_R + Y_N)$ la néguentropie de $Y_R + Y_N$. D'après la définition de la néguentropie (4.8) et compte tenu de l'indépendance de Y_R et Y_N , nous avons :

$$H^-(Y_R + Y_N) = \frac{1}{2} \log(2\pi e (\text{var}(Y_R) + \text{var}(Y_N))) - H(Y_R + Y_N) \quad (5.10)$$

Avec l'équation précédente et l'approximation (5.9) de l'entropie, nous obtenons pour la néguentropie :

$$H^-(Y_R + Y_N) = H^-(Y_R) + \frac{1}{2} \log\left(1 + \frac{\text{var}(Y_N)}{\text{var}(Y_R)}\right) - \frac{1}{2} E [\psi'_{Y_R}(Y)] \text{var}(Y_N) + o\left(\frac{\text{var}(Y_N)}{\text{var}(Y_R)}\right). \quad (5.11)$$

Ensuite, on utilise le développement² $\log(1+x) \approx x + o(x)$ pour obtenir le développement de la néguentropie au second ordre suivant :

$$H^-(Y_R + Y_N) = H^-(Y_R) + \frac{1}{2} \left(-E [\psi'_Y(Y)] + \frac{1}{\text{var}(Y_R)} \right) \text{var}(Y_N) + o\left(\frac{\text{var}(Y_N)}{\text{var}(Y_R)}\right). \quad (5.12)$$

Dans l'équation (5.12), nous pouvons faire intervenir la fonction non-linéaire $\phi_{Y_R}(Y) = \psi_{Y_R}(Y) - \frac{Y}{\text{var}(Y_R)}$ (définie en (4.16)) pour écrire :

$$H^-(Y_R + Y_N) = H^-(Y_R) - \frac{1}{2} E [\text{var}(Y_R) \phi'_{Y_R}(Y)] \frac{\text{var}(Y_N)}{\text{var}(Y_R)} + o\left(\frac{\text{var}(Y_N)}{\text{var}(Y_R)}\right). \quad (5.13)$$

Dans le terme central du membre de droite, nous faisons apparaître la variance de Y_R pour mettre en évidence le rapport des variances de Y_N et Y_R . L'analyse sera facilitée car le terme $E [\text{var}(Y_R) \phi'_{Y_R}(Y)]$ est indépendant de la variance de Y_R contrairement au terme $E [\phi'_{Y_R}(Y)]$. A la fin de cette sous-section (page 150), nous effectuons une remarque pour donner un autre moyen pour parvenir à ce développement.

D'après le développement (5.13), la maximisation de la néguentropie de la sortie tend à maximiser la somme de la néguentropie de la partie signal Y_R et du second terme. Nous notons que le premier terme $H^-(Y_R)$ représente le critère dans le cas non bruité et le second terme concentre toute la dépendance du critère à la présence du bruit. D'après la condition (R2) de la sous-section 4.4 et plus particulièrement l'équation (4.60), nous savons que $E [\phi'_{Y_R}(Y)] > 0$ (nous pouvons nous reporter à l'équation (4.97)). Ainsi la maximisation de la néguentropie de la sortie Y tendra conjointement à maximiser la néguentropie de

²pour rappel le log provient de la définition de l'entropie et représente le logarithme népérien.

la partie signal $Y_R = g \star w \star R$ et à minimiser le rapport $\frac{\text{var}(Y_N)}{\text{var}(Y_R)}$. La minimisation de ce rapport de variances a tendance à donner le meilleur rapport signal sur bruit en sortie, soit la variance de la partie bruit $Y_N = g \star N$ la plus faible possible. Pour affiner notre analyse, en nous inspirant de l'équation (5.5) écrite pour le kurtosis, nous pouvons écrire le rapport des variances avec les gains en fréquence de l'ondelette $W(\nu)$ et du filtre déconvolveur $G(\nu)$ sous la forme :

$$\frac{\text{var}(Y_N)}{\text{var}(Y_R)} = \left\{ \frac{1}{\sigma_N^2} \int_0^1 \Delta(\nu) |W(\nu)|^2 d\nu \right\}^{-1}, \quad (5.14)$$

avec

$$\Delta(\nu) = \frac{|G(\nu)|^2}{\int_0^1 |G(\lambda)|^2 d\lambda}. \quad (5.15)$$

Comme pour le cas du kurtosis, nous notons que $\Delta(\nu)$ peut être vu comme un poids de barycentre car $\int_0^1 \Delta(\nu) d\nu = 1$. La maximisation de la négentropie induit la minimisation du rapport défini par (5.14), donc la maximisation l'intégrale. Cela est équivalent à concentrer les poids $\Delta(\nu)$ pour les grandes valeurs de $|W(\nu)|$. Ainsi, le gain du filtre déconvolveur estimée par l'algorithme MANege tend vers 0 dans l'espace nul.

Pour juger de l'importance du bruit dans le terme de négentropie de (5.13), nous pouvons nous intéresser au cas où Y suit une loi gaussienne généralisée de paramètre α . Notons $h(\alpha)$ la fonction :

$$h(\alpha) = -\frac{\frac{1}{2} E [\text{var}(Y_R) \phi'_{Y_R}(Y_R)]}{H^-(Y_R)}. \quad (5.16)$$

Cette fonction $h(\alpha)$ mesure l'importance du bruit dans le critère. À partir de (5.13) nous pouvons déduire :

$$\frac{H^-(Y_R + Y_N) - H^-(Y_R)}{H^-(Y_R)} = h(\alpha) \frac{\text{var}(Y_N)}{\text{var}(Y_R)} + o\left(\frac{\text{var}(Y_N)}{\text{var}(Y_R)}\right). \quad (5.17)$$

Le rapport des variances peut être vu comme le RSB. Pour trouver l'expression de la fonction $h(\alpha)$, il suffit de reprendre l'expression de la négentropie d'un processus GG de paramètre α en (4.28) et l'expression de la fonction non linéaire ϕ_Y . Puis, en se servant de l'expression des moments de la loi GG en (4.25), nous obtenons :

$$h(\alpha) = \frac{1 - \frac{\alpha(\alpha-1)\Gamma(\frac{3}{\alpha})\Gamma(1-\frac{1}{\alpha})}{\Gamma(\frac{1}{\alpha})^2}}{\log\left(\frac{\pi e \alpha^2 \Gamma(\frac{3}{\alpha})}{2\Gamma(\frac{1}{\alpha})^3}\right) - \frac{2\Gamma(1+\frac{1}{\alpha})}{\Gamma(\frac{1}{\alpha})}}. \quad (5.18)$$

La figure 5.1 représente la fonction $h(\alpha)$ pour les valeurs de α supérieures à 1 qui assurent l'existence du développement de Taylor.

Nous remarquons que la fonction $h(\alpha)$ est minimale pour $\alpha = 2$ soit un processus Y gaussien. Le biais sur la négentropie induit par la présence du bruit augmente lorsque le

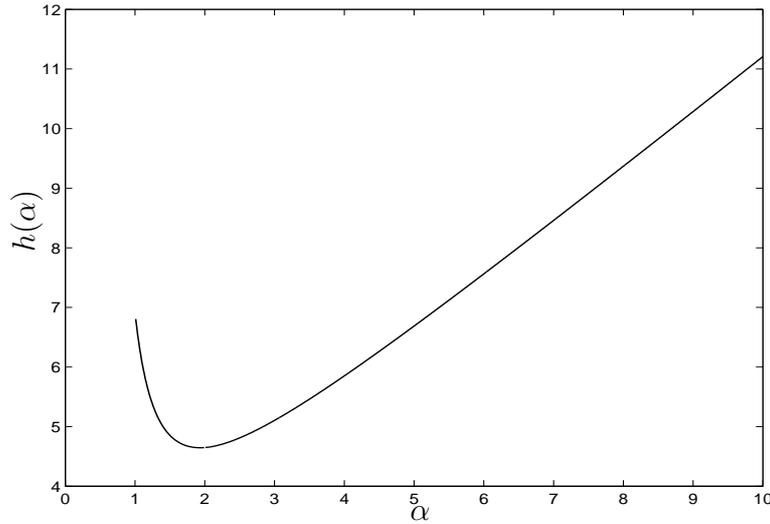


FIG. 5.1 – Importance $h(\alpha)$ du terme de second ordre relatif au bruit dans le développement de la négentropie d'une loi gaussienne généralisée de coefficient α .

coefficient de gaussianité s'écarte de 2. En terme de robustesse vis-à-vis du bruit, il est difficile de conclure plus précisément. En effet, le biais induit sur la mesure de la négentropie n'est pas obligatoirement préjudiciable pour réaliser un bon compromis entre la déconvolution et l'amplification du bruit. Si le critère est indépendant de la présence du bruit, l'algorithme risque d'amplifier grandement le bruit. Nous pouvons dire que plus le signal de sortie recherché est non-gaussien, on s'attend à ce que le critère prendra plus en compte la présence du bruit gaussien et donc limitera son amplification.

Remarque :

Nous aurions pu obtenir l'approximation (5.13) directement avec un développement de Taylor de la négentropie $H^-(Y)$. Pour une fonction $f(x)$, nous avons :

$$f(x_0 + h) = f(x_0) + f'(x_0)\|h\| + \frac{1}{2}f''(x_0)\|h\|^2 + o(\|h\|^2). \quad (5.19)$$

Nous prendrons pour la fonction f la négentropie, $h = Y_N = g \star N$ et $x_0 = Y_R = g \star w \star R$. En nous inspirant du calcul du gradient de la négentropie par rapport au filtre déconvolveur $g(t)$ dans le chapitre 4 et l'équation (4.12), le gradient de la négentropie $H^-(Y)$ par rapport à Y est :

$$\frac{\partial}{\partial Y} H^-(Y) = E \left[\frac{Y}{\text{var}(Y)} \right] - E[\psi_Y(Y)] = -E[\phi_Y(Y)]. \quad (5.20)$$

Sous l'hypothèse de signaux centrés suivant une distribution symétrique la dérivée première est nulle. Ensuite, si nous calculons la dérivée seconde, nous obtenons simplement :

$$\frac{\partial^2 H^-(Y)}{\partial Y^2} = \frac{1}{\text{var}(Y)} - E[\psi'_Y(Y)] = -E[\phi'_Y(Y)]. \quad (5.21)$$

Enfin, la norme utilisée pour les processus aléatoires est simplement la norme 2 :

$$\|X\| = \sqrt{E[X^2]}. \quad (5.22)$$

Ainsi, nous retrouvons l'expression (5.13) avec le développement de Taylor de la négentropie. Nous noterons que ce développement existe si la négentropie est au moins de classe \mathcal{C}^2 . Par exemple pour les processus GG, il faut que $\alpha > 1$.

5.1.3 Algorithme FBD : minimisation du taux d'information mutuelle.

L'algorithme FBD, décrit au chapitre 3, a été construit afin d'augmenter la robustesse au bruit additif des algorithmes MAMV, et, plus généralement des méthodes basées sur la blancheur. Dans le critère (3.5) de l'algorithme FBD, nous omettrons la partie qui contraint le filtre déconvolveur à avoir une réponse en fréquence continue. Ce terme est justifié pour assurer la stabilité de l'algorithme FBD mais n'a aucun effet sur la valeur du RSB de sortie. En revenant à une écriture avec les fréquences continues et en utilisant la notation de fréquence réduite pour ν , nous pouvons considérer de façon équivalente à $J(\mathbf{G})$ de (3.5) la fonction coût suivante :

$$J_1(\mathbf{G}) = H(Y) - \int_0^1 \log |G(\nu)| d\nu + \lambda_2 \int_0^1 |G(\nu)|^p d\nu. \quad (5.23)$$

Or, la première intégrale du membre de droite de l'expression précédente peut s'écrire :

$$\int_0^1 \log |G(\nu)| d\nu = \frac{1}{2} \int_0^1 \log |G(\nu)|^2 S_{2D}(\nu) d\nu - \frac{1}{2} \int_0^1 \log S_{2D} d\nu, \quad (5.24)$$

où $S_{2D}(\nu)$ est la densité spectrale de puissance des données D . Puis, avec la relation des interférences liant les densités spectrales de la sortie Y et des données D nous avons :

$$\int_0^1 \log |G(\nu)| d\nu = \frac{1}{2} \int_0^1 \log S_{2Y}(\nu) d\nu - \frac{1}{2} \int_0^1 \log S_{2D}(\nu) d\nu. \quad (5.25)$$

Ensuite, nous pouvons faire intervenir dans l'expression (5.23) la négentropie de la sortie $H^-(Y) = \frac{1}{2} \log(2\pi e \sigma_Y^2) - H(Y)$, définie aussi en (4.8), où σ_Y^2 est la variance de Y . La fonction coût $J_1(\mathbf{G})$ de (5.23) est égale à une constante près à :

$$J_2(\mathbf{G}) = -H^-(Y) + \frac{1}{2} \int_0^1 \log \frac{\sigma_Y^2}{S_{2Y}(\nu)} d\nu + \lambda_2 \int_0^1 |G(\nu)|^p d\nu. \quad (5.26)$$

L'algorithme FBD minimise à une constante près la fonction coût $J_2(\mathbf{G})$. Comme nous l'avions expliqué dans le chapitre 3 le dernier terme permet de limiter les grandes valeurs de $|G(\nu)|^2$. D'autre part, le second terme sera minimal si la densité spectrale de puissance est constante car nous rappelons que $\int_0^1 S_{2D}(\nu) d\nu = \sigma_Y^2$. Ainsi, le second terme est une mesure de constance de la densité spectrale de puissance de la sortie donc tend à blanchir à l'ordre 2 la sortie estimée. Comme les deux premiers termes expriment la blancheur par le taux d'information mutuelle, nous retrouvons que le critère de négentropie et le blanchiment à l'ordre 2 sont équivalents à la déconvolution par le taux d'information mutuelle. Une autre

façon de formuler cette remarque est que la néguentropie réalise un blanchiment en prenant en compte des statistiques d'ordre strictement supérieur à deux. La minimisation de la fonction coût $J_2(\mathbf{G})$ tend à maximiser la néguentropie et à réaliser un blanchiment à l'ordre 2, en excluant les grandes valeurs pour le gain du filtre déconvolveur. Ces trois termes sont donc optimisées conjointement. Nous avons vu dans le paragraphe précédent que le terme de la néguentropie fait tendre le gain du filtre déconvolveur vers 0 dans l'espace nul. Le dernier terme fait de même. Pour poursuivre, nous proposons donc de nous intéresser au terme central de blanchiment à l'ordre 2.

La densité spectrale de sortie s'exprime par :

$$S_{2Y}(\nu) = |G(\nu)|^2[|W(\nu)|^2 + \sigma_N^2]. \quad (5.27)$$

Ainsi, le terme de blanchiment à l'ordre 2 tend à estimer le filtre de déconvolution suivant :

$$|G(\nu)| = \frac{\text{Constant}}{[|W(\nu)|^2 + \sigma_N^2]^{1/2}} = |G_{\text{Wiener}}(\nu)| \left[1 + \frac{\sigma_N^2}{|W(\nu)|^2} \right]^{1/2}. \quad (5.28)$$

Ce gain est égal au produit du gain du filtre de Wiener avec un terme qui peut devenir très grand pour certaines fréquences ν pour lesquelles $\sigma_n^2/|W(\nu)|^2$ est grand, soit $|W(\nu)|^2$ est petit. En fait, le gain du filtre blanchisseur à l'ordre 2 est toujours borné, ainsi, le bruit n'est pas amplifié de façon trop forte. Mais, le gain du filtre est maximum dans l'espace nul, il tend vers une constante. L'amplification du bruit est supérieure à celle obtenue avec le filtrage de Wiener, car dans l'espace nul le gain tend vers une constante au lieu de zéro pour le filtrage de Wiener. De même, l'algorithme FBD est moins robuste que les algorithmes MED et MANege, qui estiment un gain du filtre déconvolveur tendant vers zéro dans l'espace nul.

5.2 Expérimentations

Pour illustrer les développements théoriques de la section précédente, nous allons dans un premier temps comparer les algorithmes MED, MANege et FBD pour la déconvolution de données synthétiques bruitées. Enfin, nous présenterons la déconvolution de données réelles. Nous traiterons le cas de données d'explosions séismo-volcaniques et sous-marines et nous finirons par la déconvolution d'un profil sismique réel par plusieurs algorithmes. Lors de ces tests, nous chercherons à confronter les deux hypothèses de blancheur et sur-gaussianité. Au sein de la seconde classe d'algorithmes, nous comparerons les algorithmes MED et MANege.

5.2.1 Signaux simulés

Nous souhaitons comparer avec une série de réflectivité simulée la robustesse vis-à-vis du bruit des algorithmes MED, MANege et FBD. Nous utiliserons le filtrage de Wiener comme référence. La figure 5.2 représente les signaux de la simulation. La réflectivité (figure 5.2(a)) est simulée suivant un modèle sur-gaussien. Elle est blanche comme l'indique sa

densité spectrale de puissance (figure 5.2(d)) et a une distribution sur-gaussienne d'après son histogramme (figure 5.2(e)). Pour mettre en évidence la régularisation par rapport au bruit, nous utilisons une ondelette à bande passante limitée. De plus, nous souhaitons avoir une ondelette à phase quelconque (ni minimale, ni nulle, ni maximale). La figure 5.2(b) représente l'ondelette choisie qui est un mélange d'ondelettes de Ricker avec des phases différentes. Après convolution entre la réflectivité et l'ondelette, puis l'ajout d'un bruit additif blanc gaussien avec un RSB de 8 dB, nous obtenons les observations de la figure 5.2(c). Il est intéressant d'analyser l'allure de la dsp des données sur la figure 5.2(f). Nous notons que l'ondelette a une bande passante comprise entre 0 et 0.25 en fréquence réduite. La bande de fréquences entre 0.25 et 0.5 est l'espace nul car elle comporte uniquement le bruit additif blanc.

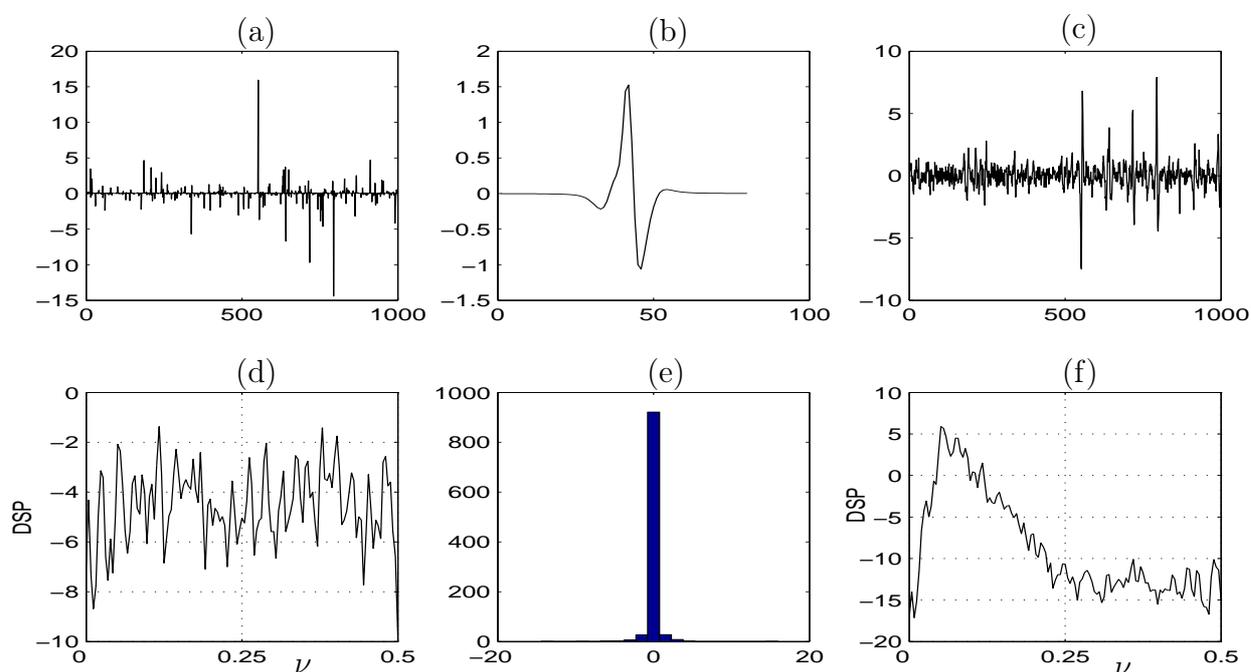


FIG. 5.2 – Données simulées pour la comparaison des algorithmes MED, MANeGe et FBD : (a) réflectivité sur-gaussienne, (b) ondelette, (c) observation avec un RSB=8dB, (d) Densité spectrale de puissance (dsp) de la réflectivité (en dB) (e) histogramme de la réflectivité (f) dsp des observations.

Nous proposons de déconvoluer les données simulées par 4 méthodes :

- Le filtrage de Wiener en fréquence avec des conditions optimales : ondelette théorique et rapport signal sur bruit connus,
- L'algorithme MED avec un filtre à 20 coefficients,
- L'algorithme MANeGe avec un filtre à 20 coefficients,
- L'algorithme FBD avec $\lambda_1 = 10$ et $\lambda_2 = 0.1$.

La figure 5.3 représente les signaux de sortie estimés par les quatre algorithmes. Nous trouvons sur la première ligne les sorties estimées par le filtrage de Wiener (a), l'algorithme MED (b), l'algorithme MANeGe (c) et l'algorithme FBD (d). Sur la seconde ligne, nous

représentons les dsp respectives des signaux estimés.

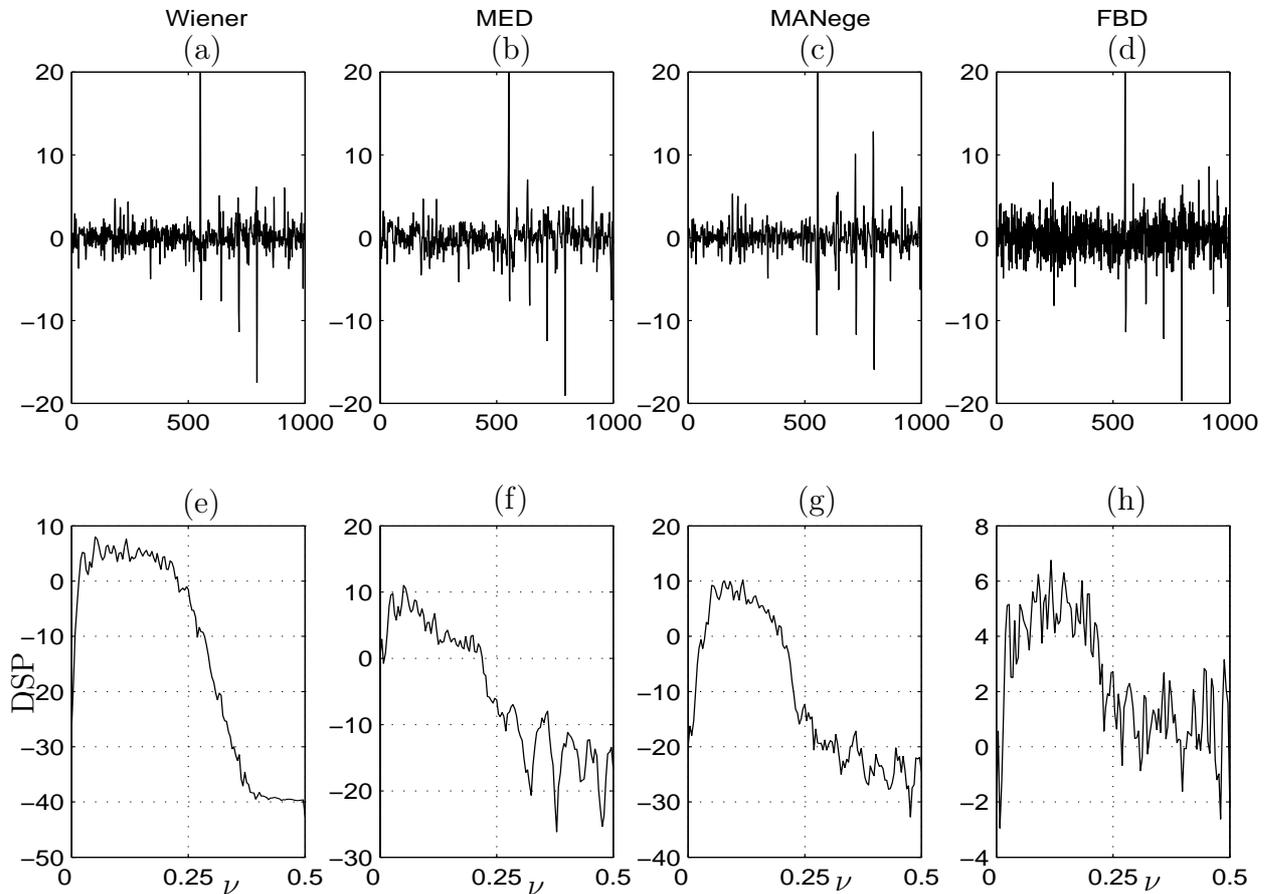


FIG. 5.3 – Comparaison de la déconvolution des données de la figure 5.2, sorties estimées et dsp en dB. (a) et (e) : filtrage de Wiener, (b) et (f) : algorithme MED, (c) et (g) algorithme MANege, (d) et (h) algorithme FBD.

Les sorties estimées par les différents algorithmes de la figure 5.3 doivent être comparées à la réflectivité théorique de la figure 5.2(a). Les signaux estimés sont normalisés pour avoir une valeur maximale de 20. Pour juger de la qualité des méthodes de déconvolution, nous devons tout d'abord vérifier si les grands réflecteurs sont bien estimés. Ensuite, sur les allures temporelles, nous devons localiser des réflecteurs d'amplitude moyenne ou faible qui sont mis en évidence par l'algorithme. Par exemple, si plusieurs réflecteurs se séparent après déconvolution alors qu'ils étaient indiscernables sur les données. Enfin, le niveau de bruit dans l'allure temporelle ne doit pas être trop grand au risque de masquer des réflecteurs. Ces informations se retrouvent aussi dans la densité spectrale des signaux estimés. L'augmentation de la résolution des données est réalisée en enrichissant le spectre des données. Ainsi, la déconvolution doit augmenter la bande passante des signaux, voire même blanchir le spectre. Mais, nous sommes en présence de bruit avec une ondelette à bande limitée. Nous devons donc mesurer la différence de la valeur de la dsp dans la bande passante de l'on-

delette et dans l'espace nul. Cet écart sera caractéristique du rapport signal à bruit de sortie.

En analysant les allures temporelles des sorties estimées par les quatre algorithmes de la figure 5.3, nous notons que le filtrage de Wiener permet le meilleur compromis entre déconvolution et amplification du bruit. Ce résultat est logique car la méthode est optimale au sens de l'erreur quadratique moyenne. Par ailleurs, les estimées des algorithmes MED et MANeGe sont très proches si nous analysons seulement les allures temporelles. Pour affiner l'analyse, nous pouvons utiliser les dsp des sorties. Nous notons que la différence entre la bande passante de l'ondette et l'espace nul est de 20dB pour l'algorithme MED contre au moins 25dB pour l'algorithme MANeGe. Ainsi, nous avons une plus faible amplification du bruit dans l'espace nul (fréquences entre 0.25 et 0.5) pour l'algorithme MANeGe. Cependant, la bande passante est légèrement moins large avec l'algorithme MANeGe qu'avec l'algorithme MED. En effet, la transition (aux alentours de 0.25) se fait plus vite pour l'algorithme MANeGe que pour l'algorithme MED. De cette observation, nous pouvons dire que l'algorithme MED aura une meilleure résolution sur ce jeu de données. Nous pouvons ajouter à cette remarque que pour l'algorithme MANeGe, nous notons que, pour les deux réflecteurs importants situés entre le 750 et 800^{ème} échantillons, nous avons un pic secondaire juste à côté du réflecteur. Ainsi, il est difficile de différencier les algorithmes MED et MANeGe, mais nous retiendrons leur aptitude à limiter l'amplification du bruit sans aucune régularisation particulière. Enfin, pour l'algorithme FBD, nous avons un bruit très important sur l'allure temporelle qui noie un grand nombre de réflecteurs d'amplitude moyenne (par exemple le groupe de réflecteurs situé aux alentours de l'échantillon 250). Avec la dsp, nous voyons que, conformément à notre argumentation dans la partie précédente, la méthode tend à blanchir le spectre à l'ordre 2, et nous avons un différentiel de seulement 3dB entre les deux bandes de fréquence. Pour finir l'analyse et pour illustrer les développements théoriques de la section précédente, nous traçons sur la figure 5.4, le gain en fréquence des filtres inverses estimés par les quatre algorithmes.

La figure 5.4 permet de totalement corroborer nos développements théoriques de la section précédente. En effet, les algorithmes MED et MANeGe estiment un filtre avec un gain qui tend vers 0 dans l'espace nul. Nous noterons que l'algorithme MANeGe a tendance à limiter plus l'amplification du bruit que l'algorithme MED. Néanmoins, leurs performances sont moindres mais davantage éloignées de celles du filtrage de Wiener. En revanche, l'algorithme FBD estime un filtre dont le gain tend vers une constante assez grande pour ces mêmes fréquences.

Nous ne présenterons pas les illustrations, mais dans la communication [LVDBMJ05], nous avons montré dans le cadre d'une simulation que la robustesse de l'algorithme MED était mise à mal si la distribution de la réflectivité est bimodale. En effet, dans ce cas, le kurtosis ne permet pas de caractériser la sur-gaussianité de la distribution [MJ99].

5.2.2 Données réelles : explosions séismo-volcaniques

Nous proposons de poursuivre avec le traitement de données réelles séismo-volcaniques du volcan Purace. Ces données ont déjà été utilisées dans les sous-sections 3.1.1 et 3.3.1.

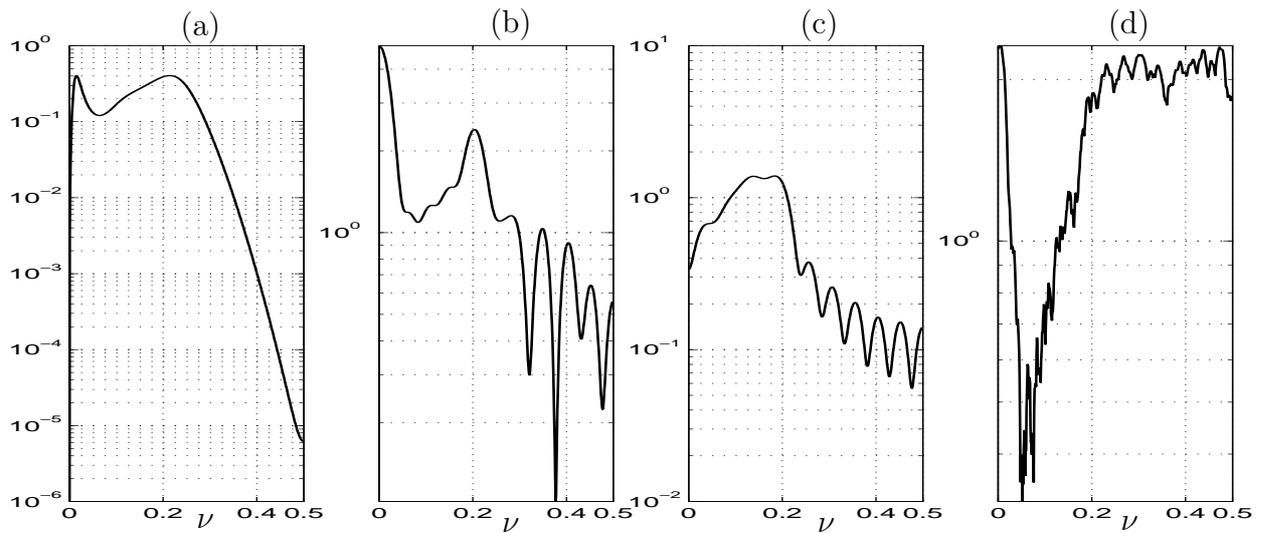


FIG. 5.4 – Gain en fréquence (en dB) du filtre déconvolué estimé pour la déconvolution des données de la figure 5.2 par : (a) filtrage de Wiener, (b) algorithme MED, (c) algorithme MANege, (d) algorithme FBD.

La partie de gauche de la figure 5.5 représente la sortie estimée lors de la déconvolution des données représentées en (a) par les algorithmes FBD, MANege et MED. La colonne de droite représente le spectre du filtre direct estimé (en échelle linéaire), ce qui permet d’analyser les fréquences de résonance extraites par la déconvolution. Nous effectuons un zoom pour les fréquences réduites entre 0 et 0.15.

L’analyse des allures temporelles (colonne de gauche de la figure 5.5) montre que l’algorithme FBD permet une meilleure estimation de l’excitation que les deux algorithmes (MED et MANege) basés sur la sur-gaussianité. L’algorithme MED ne supprime pas tous les effets résonnants car nous remarquons la présence d’une enveloppe exponentielle décroissante sur le signal déconvolué (Figure 5.5(d)). Pour l’algorithme MANege, la sortie comporte plusieurs parties, avec une sorte de répétition de phénomènes d’excitation. Nous notons cependant le très faible niveau de bruit dans la partie finale du signal déconvolué. Avec les spectres des filtres directs estimés (colonne de droite), l’algorithme FBD apparaît comme l’algorithme estimant le plus grand nombre de fréquences de résonance et avec une très bonne résolution. L’algorithme MED (Figure 5.5(e)) estime seulement trois fréquences de résonance. L’algorithme MANege (Figure 5.5(g)) estime quatre fréquences de résonance. Pour la fréquence située à environ 0.045, la valeur est en accord avec l’algorithme FBD et MED. En revanche, pour la fréquence à 0.12, nous n’avons pas la même valeur que pour l’algorithme FBD. Or, d’après l’étude menée dans le chapitre 3 et plus précisément la figure 3.9, les fréquences estimées par l’algorithme FBD sont en accord avec celles de l’égalisation spectrale. Or, nous pouvons prendre l’égalisation spectrale comme référence car en se basant sur le spectre des données elle estime obligatoirement les bonnes fréquences de résonance. En conclusion, les algorithmes MANege et MED (hypothèse de sur-gaussianité) donnent des résultats moins bons que l’algorithme FBD (hypothèse de blancheur). Pour ces données séismo-volcaniques,

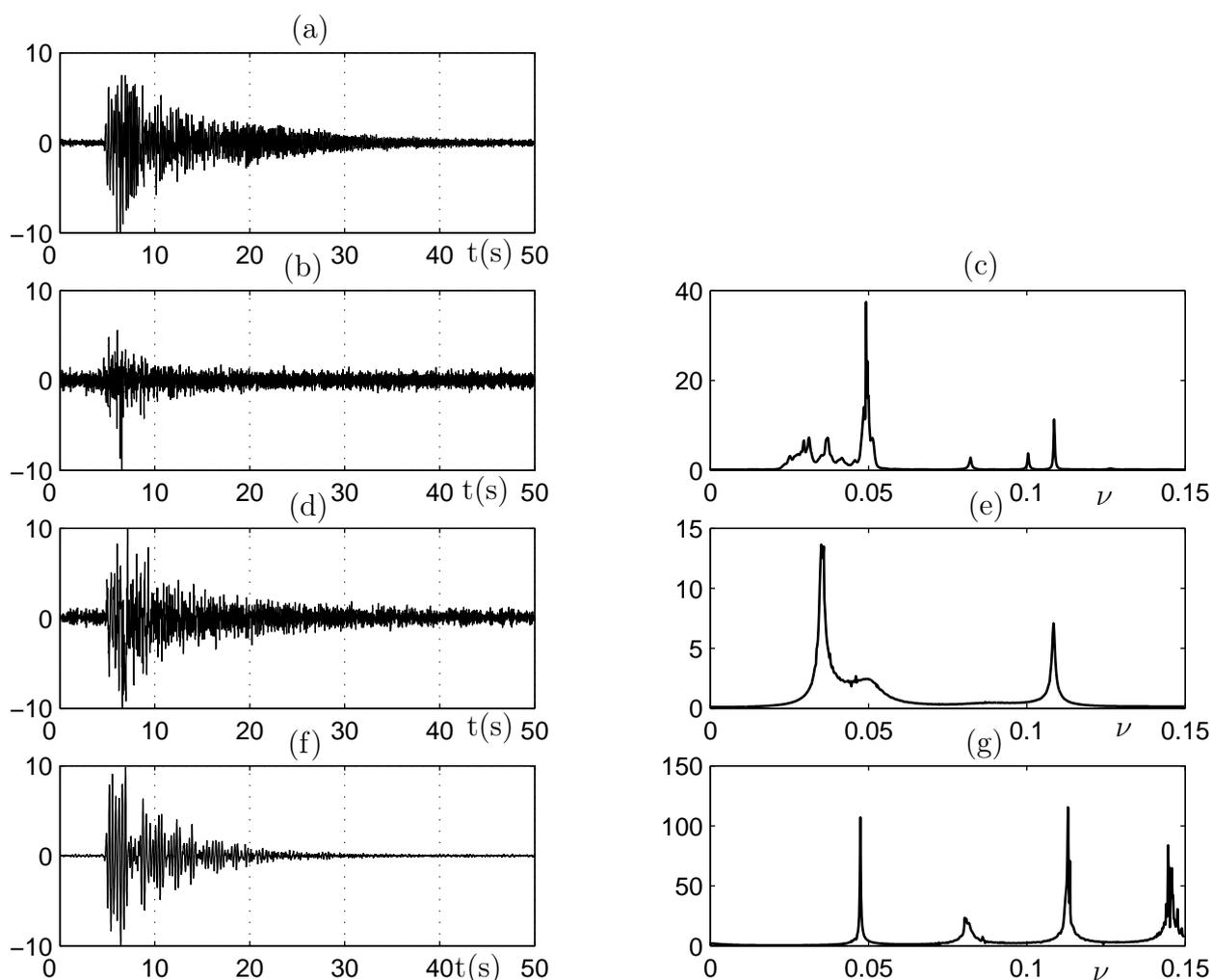


FIG. 5.5 – Comparaison des hypothèses de blancheur et de sur-gaussianité pour la déconvolution de données séismo-volcaniques : (a) observations. Sortie estimée (resp. gain du filtre direct) par : (b) algorithme FBD (resp. (c)), (d) algorithme MED (resp. (e)), (f) algorithme MANe (resp. (g)).

nous ne pouvons pas conclure que l'hypothèse de blancheur est meilleure que l'hypothèse de sur-gaussianité, car l'algorithme MAMV (blancheur) donne aussi des résultats très bruités (voir figure 3.9 et l'analyse correspondante). En fait, la qualité de la déconvolution de l'algorithme FBD dépend en grande partie de son approche fréquentielle. En effet, la présence de résonances importantes rend difficile l'estimation du filtre déconvolveur dans le domaine temporel. Le modèle autorégressif permet de telles résonances avec des pôles proches du cercle unité, mais l'estimation de ces pôles est réputée difficile. De plus, la sur-gaussianité n'apparaît pas de façon très claire sur le signal d'excitation contrairement à des signaux de réflectivité.

5.2.3 Données réelles : explosions sous-marines

Nous proposons de reprendre les données d'explosions sous-marines de la figure 3.11. La figure 5.6 représente les données en (a) et la déconvolution de ces dernières par l'algorithme FBD (b), l'algorithme MED (c) et l'algorithme MANeGe en (d). Les signaux sont normalisés en amplitude pour faciliter la comparaison. Nous rappelons que l'objectif concernant ces données est de trouver la position des trois réflecteurs de façon précise.

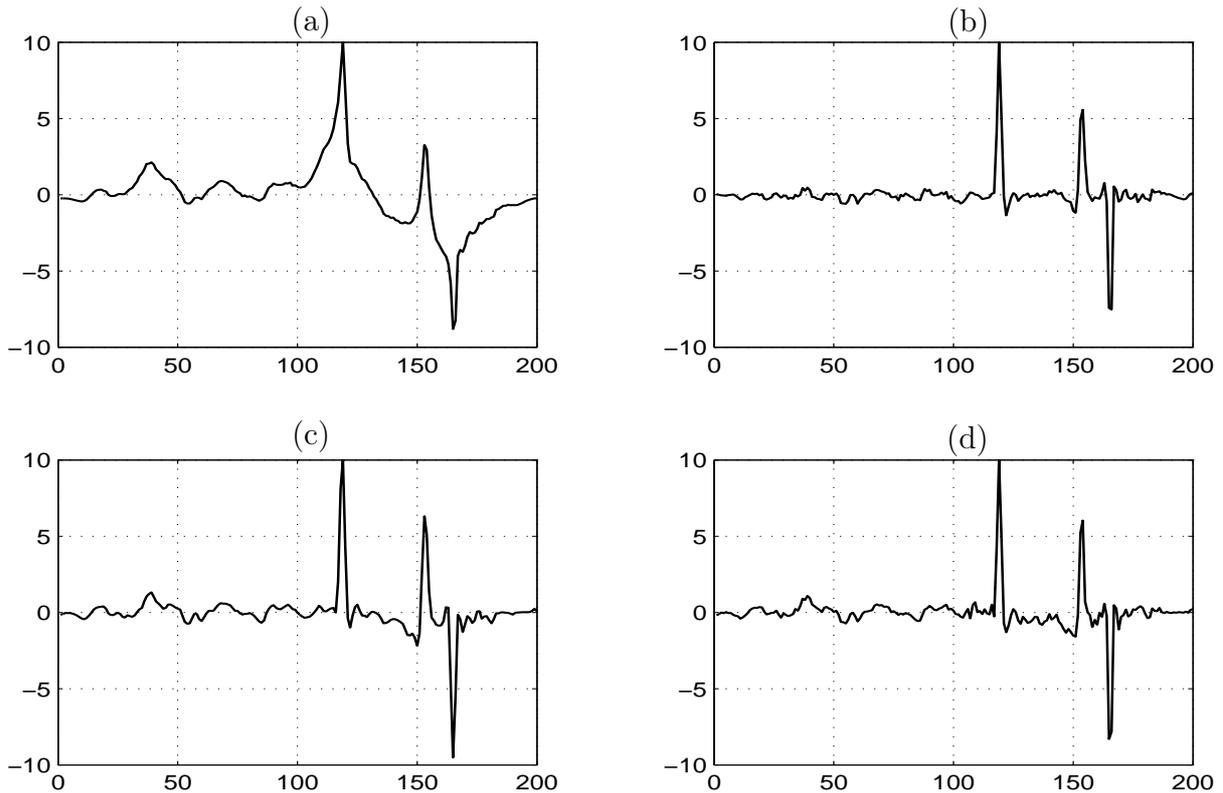


FIG. 5.6 – Comparaison des hypothèses de blancheur et de sur-gaussienité pour la déconvolution de données d'explosions sous-marines : (a) observations. Sortie estimée par : (b) algorithme FBD, (c) algorithme MED, (d) algorithme MANeGe.

La figure 5.6 montre que les résultats des trois algorithmes (MED, MANeGe et FBD) sont très semblables. Les pics sont de largeur identique et d'amplitude comparable. Nous rappelons que dans le chapitre 3, la comparaison avec les autres méthodes de déconvolution aveugle basées sur la blancheur avait montré que l'algorithme FBD donnait la meilleure sortie estimée en terme de largeur de réflecteurs et d'amplification du bruit. En conclusion, cette expérience montre que l'hypothèse de sur-gaussienité permet d'avoir sur ces données des résultats identiques au meilleur algorithme basé sur la blancheur, cela avec une complexité moindre. Ces résultats sont justifiés notamment grâce à une bonne robustesse au bruit additif des méthodes s'appuyant sur la sur-gaussienité. Ces données sont un bon exemple car elles ont une distribution très sur-gaussienne. Nous avons tenté de déconvoluer ces données avec l'algorithme MAMV, mais en vain, car la sélection des statistiques d'ordre

2 empêche d'estimer le filtre déconvolveur à phase nulle. En revanche, l'algorithme MANege qui utilise le même modèle de filtre déconvolveur donne un résultat intéressant.

5.2.4 Données réelles : profil sismique

Nous proposons pour terminer ce chapitre de présenter la déconvolution d'un profil sismique fourni par la compagnie Shell. La figure 5.7 représente le profil de départ. Ce profil comporte 103 traces (capteurs) et nous avons sélectionné une durée comportant 350 échantillons. La période d'échantillonnage est de 4 ms. Le pas inter-capteurs n'est pas connu. Nous disposons de données comportant 103 capteurs alors que tous les algorithmes de déconvolution utilisent un modèle monodimensionnel (SISO). En supposant l'ondelette constante sur toute l'antenne de capteurs, nous déduisons que le filtre déconvolveur sera le même pour toutes les traces. Ainsi, le filtre déconvolveur est estimé en optimisant la moyenne des critères SISO de chacune des traces. Cette technique est utilisée par tous les algorithmes développés dans le cadre de l'imagerie sismique [Cla77, God78, Wig78, Gra79].

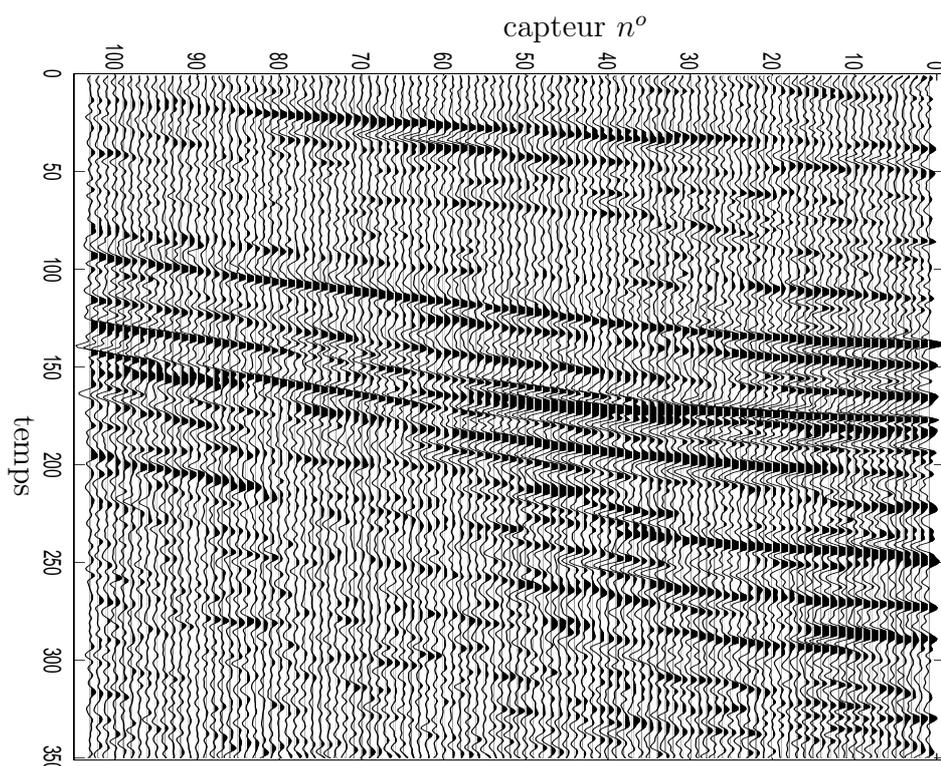


FIG. 5.7 – Profil sismique fourni par Shell.

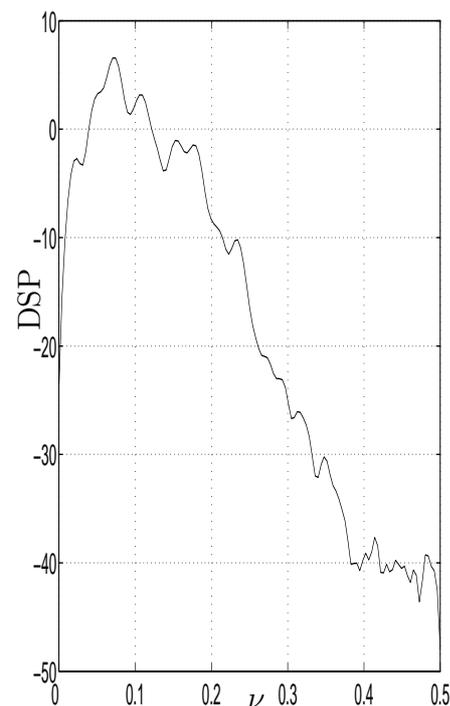


FIG. 5.8 – DSP de la trace n°1.

Dans un souci de clarté, nous allons expliquer plus précisément le mode de représentation. Pour mettre en évidence les grandes amplitudes positives sur les traces qui nous donneront les réflecteurs de signe positif, nous colorions en noir les arches positives suivant

un paramètre de saturation. Dans la suite, nous réglerons ce paramètre de saturation pour obtenir la meilleure représentation. En effet, le paramètre ne doit pas être trop petit sinon nous n'avons aucun réflecteur visible. Mais, il ne doit pas être trop grand sinon tout le profil est noir et aucun réflecteur n'est mis en évidence. Chaque trace, qui correspond à un capteur, est représentée verticalement. Avec cette représentation, le temps est sur l'axe vertical car il peut être assimilé à la profondeur. Ainsi, les réflecteurs apparaîtront horizontaux. Les données ont subi des pré-traitements. Nous ne les connaissons pas tous, mais un filtrage passe-bas grossier du bruit a été fait. Un contrôle automatique de gain a aussi été effectué pour supprimer les effets d'atténuation le long de l'antenne et également en fonction de la profondeur. D'autre part, une pré-déconvolution à l'aide d'un filtrage adapté a très certainement été réalisée car les données ont déjà une bonne résolution. La figure 5.8 représente la densité spectrale de puissance (en échelle logarithmique) de la première trace du profil des données 5.7. Nous notons que l'allure du spectre est passe-bas comme les données simulées dans la figure 5.2. La bande passante se termine pour la fréquence réduite 0.3 et l'espace nul s'étend entre les fréquences 0.3 et 0.5. Nous proposons de déconvoluer le profil réel de la figure 5.7 par trois algorithmes. Nous déconvoluons avec l'égalisation spectrale (Figure 5.9 et 5.10), l'algorithme MED (Figure 5.11 et 5.12) et l'algorithme MANege (Figure 5.13 et 5.14). Pour ces deux derniers algorithmes nous utilisons des filtres de longueur 16 coefficients. Pour analyser les résultats, nous tracerons côte-à-côte le profil estimé et la dsp de la première trace estimée pour juger de l'augmentation de la bande passante et de l'amplification du bruit. Nous n'avons pas représenté les résultats de l'algorithme FBD, car ils sont vraiment de très mauvaise qualité. Nous expliquons en grande partie ces mauvaises performances par l'amplification du bruit additif.

Pour mener à bien l'analyse, nous partirons des observations les plus faciles pour finir par les détails des résultats. Tout d'abord, l'algorithme MED (Figure 5.11) casse la structure des données. Il privilégie le réflecteur de plus forte énergie et le compresse, mais, dans le même temps, il réduit grandement l'amplitude des réflecteurs secondaires. Cette remarque avait déjà été faite par Wiggins [Wig78], et cette tendance est très connue dans le domaine de la déconvolution en géophysique. Sur le profil déconvolué (Figure 5.11), quelques réflecteurs d'énergie comparables à celle du plus grand sont encore visibles, mais le profil déconvolué n'est pas satisfaisant. Nous notons l'apparition de quelques événements haute fréquence notamment dans la partie supérieure droite. Cette amplification des hautes fréquences est normale car nous essayons d'enrichir le spectre pour augmenter la résolution. L'augmentation de la bande passante est visible en comparant la dsp de la première trace après déconvolution par l'algorithme MED (Figure 5.12) et la dsp des données (Figure 5.8). La bande passante est augmentée pour les fréquences proches de 0.3 tout en conservant une faible dsp dans l'espace nul afin d'avoir un bon RSB de sortie. Ce résultat est en accord avec la partie théorique du début de chapitre mais aussi avec les expériences sur les signaux

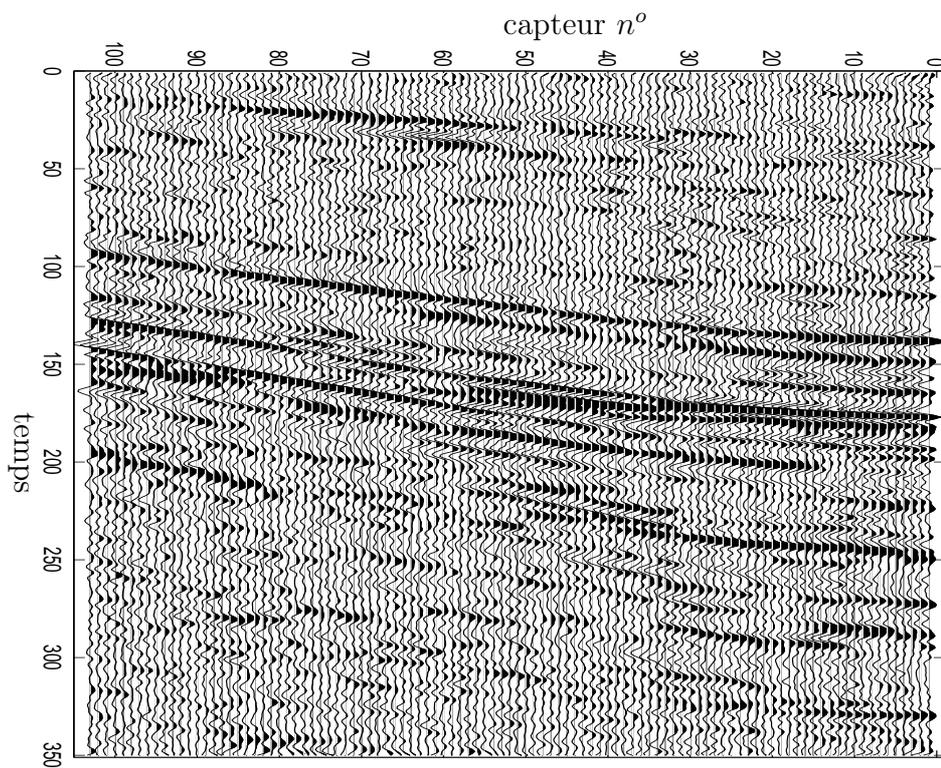


FIG. 5.9 – Déconvolution par l'égalisation spectrale.

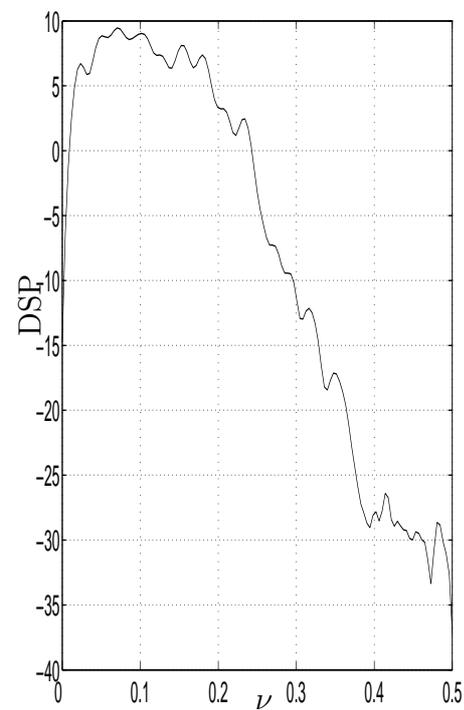
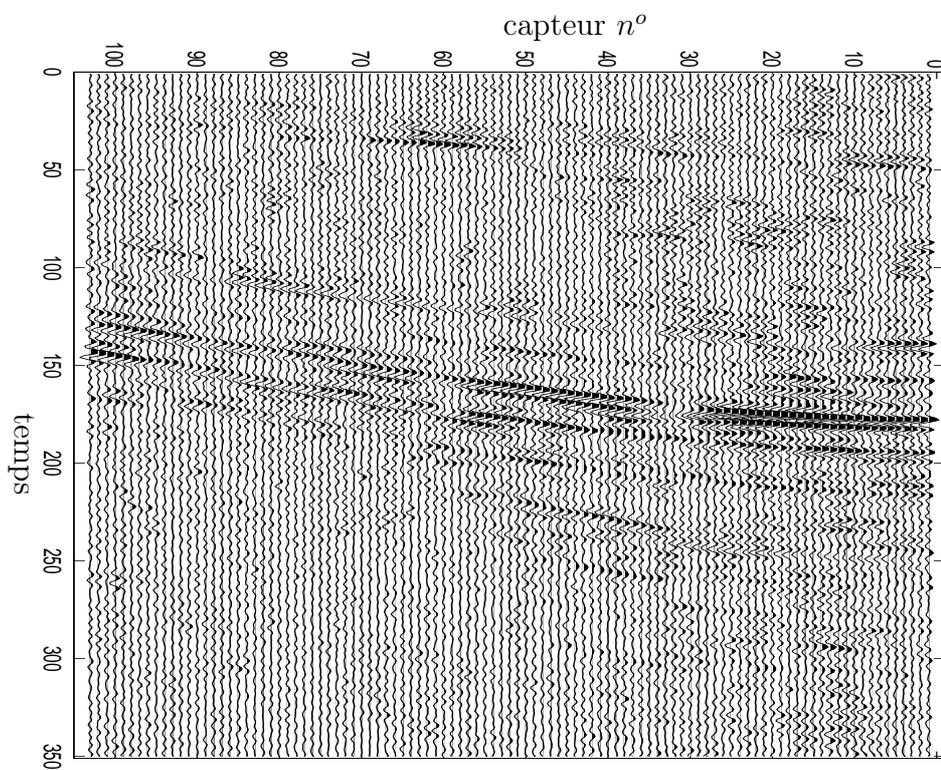
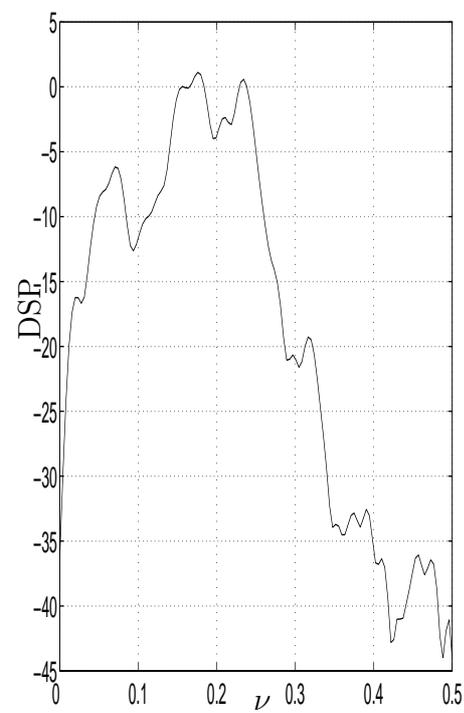
FIG. 5.10 – DSP de la trace $n^{\circ}1$.

FIG. 5.11 – Déconvolution par l'algorithme MED.

FIG. 5.12 – DSP de la trace $n^{\circ}1$.

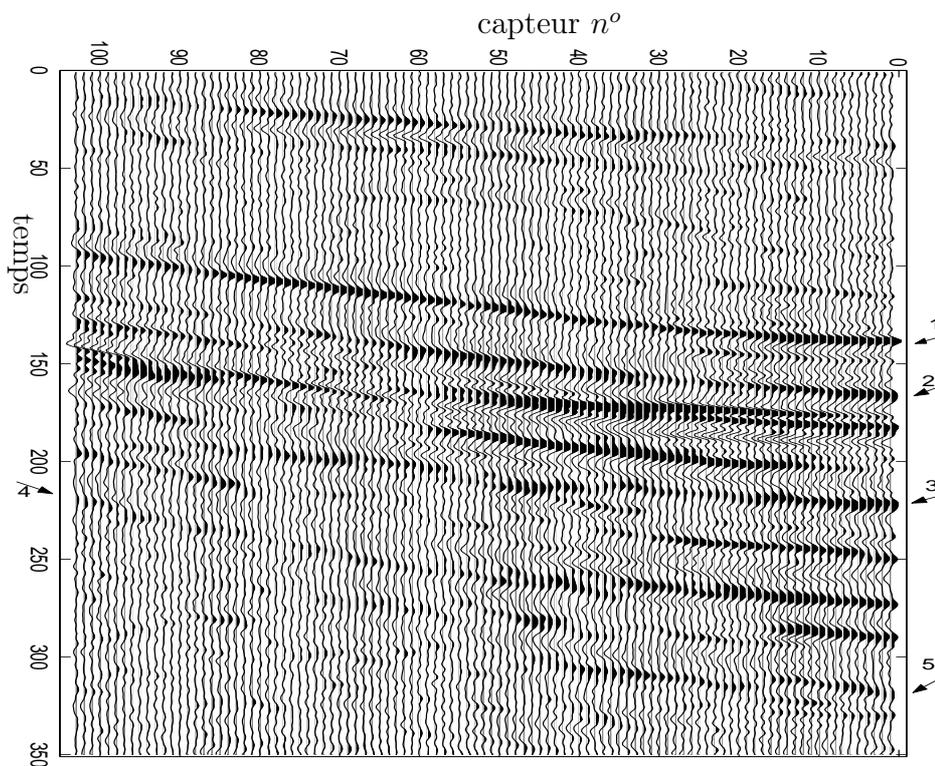
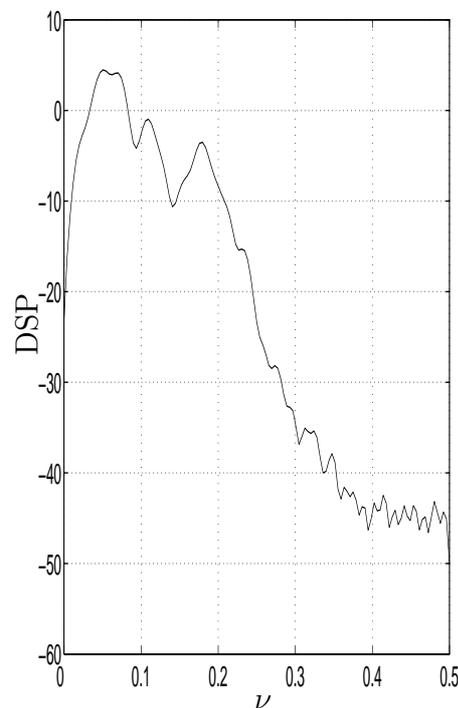


FIG. 5.13 – Déconvolution par l'algorithme MANeGe.

FIG. 5.14 – DSP de la trace $n^{\circ}1$.

simulés.

Nous notons que la structure des réflecteurs est bien conservée pour l'algorithme de l'égalisation spectrale (Figure 5.9 et 5.10) et l'algorithme MANeGe (Figure 5.13 et 5.14). Nous ne perdons pas d'informations contrairement à l'algorithme MED. Pour les deux algorithmes, le tracé du profil ne montre pas d'amplification trop grande de bruit haute fréquence. L'égalisation spectrale (Figure 5.9) obtient un bruit tout de même légèrement supérieur à l'algorithme MANeGe (Figure 5.13). Cette observation est confirmée par le tracé de la dsp de la première trace. La différence entre la bande passante et l'espace nul est de 40dB pour l'algorithme MANeGe (Figure 5.14) et seulement de 35dB pour l'égalisation spectrale (Figure 5.10). Nous remarquons aussi que l'égalisation spectrale élargit plus la bande passante que l'algorithme MANeGe. En effet, la dsp de la sortie a une valeur plus grande aux fréquences voisines de 0.3 pour l'égalisation spectrale comparativement à l'algorithme MANeGe. Ce gain de bande passante se fait obligatoirement au détriment de l'amplification du bruit. Pour l'égalisation spectrale, il est difficile de localiser une zone où un réflecteur apparaît très clairement grâce à la déconvolution.

Pour l'algorithme MANeGe (Figure 5.13), nous avons mis en évidence au moins cinq réflecteurs. Nous les désignons par des flèches sur la figure 5.13. Le réflecteur 1 est très bien mis en évidence, avec une très bonne continuité latérale, le rebond dû à l'ondelette visible sur les données initiales a été supprimé par l'algorithme MANeGe. Pour la flèche 2, nous avons une bonne conservation du réflecteur. En fait, nous pouvons mentionner que l'algo-

rithme ne le supprime pas alors qu'il est très proche de deux réflecteurs très énergétiques. Les flèches 3, 4 et 5 montrent des réflecteurs mis en évidence grâce à la déconvolution avec l'algorithme MANege. Ce résultat est très intéressant car ces réflecteurs se situent à des instants grands donc ils représentent des interfaces géologiques profondes. Cet aspect est important pour les géophysiciens, car l'atténuation des ondes avec la profondeur limite les performances de l'imagerie sismique.

5.3 Conclusions

Ce dernier chapitre de ce manuscrit est très important pour le côté applicatif de notre travail. En effet, il permet de comprendre le comportement des algorithmes FBD, MED et MANege pour des données bruitées. Nous avons présenté une étude théorique des critères de ces trois algorithmes pour voir si une régularisation naturelle était faite par les algorithmes pour limiter l'amplification du bruit de sortie. Nous avons montré que le gain dans l'espace nul tendait vers 0 pour les algorithmes MED et MANege. En revanche, l'algorithme FBD comporte un terme blanchisseur à l'ordre 2 qui tend à estimer un gain constant dans l'espace nul pour le filtre déconvolveur. En définitive, les algorithmes MED et MANege permettent un compromis naturel entre l'estimation de la réflectivité et l'amplification du bruit. Les résultats avec des données simulées illustrent ces résultats théoriques.

Ensuite, nous avons présenté des résultats avec la présentation de trois types de données réelles. Les premières, relatives à des explosions séismo-volcaniques, ne montrent pas de résultats en accord avec les développements théoriques. Nous avançons comme première explication que le filtre direct résonnant est très difficile à estimer avec le modèle temporel utilisé par les algorithmes MED et MANege. De plus, le caractère sur-gaussien du signal d'excitation n'est pas très marqué. Par conséquent, les algorithmes basés sur la sur-gaussianité ne sont pas dans la configuration la plus favorable. Pour les données d'explosions sous-marines, les deux hypothèses (blancheur ou sur-gaussianité) ont des performances identiques. Néanmoins, l'hypothèse de blancheur (algorithme FBD) nécessite un terme de régularisation pour le bruit additif. Enfin, nous avons terminé ce chapitre par le traitement d'un profil sismique. Pour ce dernier, nous opposons les hypothèses de blancheur et sur-gaussianité, mais nous comparons aussi les algorithmes (MED et MANege) basés sur la sur-gaussianité. L'algorithme FBD ne donne absolument aucun résultat pour ce jeu de données. Malgré une faible amplification du bruit, l'algorithme MED n'est pas intéressant car il met en évidence seulement le principal réflecteur et fait disparaître les réflecteurs de plus faibles énergies. Ensuite, l'égalisation spectrale conserve la structure des données et n'amplifie pas le bruit. Mais cette méthode ne met pas en évidence de nouveaux réflecteurs. Finalement, l'algorithme MANege a les meilleures performances, en travaillant en aveugle, en n'amplifiant pas le bruit et en augmentant la résolution des données avec la mise en évidence de nouveaux réflecteurs.

Conclusions et perspectives

Conclusions

Le travail de recherche présenté dans ce mémoire a permis d'adapter certaines méthodes de déconvolution aveugle aux données bruitées, et plus particulièrement aux signaux sismiques. Nous avons privilégié l'approche aveugle afin d'éviter de poser des *a priori* sur la réflectivité et/ou l'ondelette. Notre objectif final était de proposer des méthodes aveugles robustes au bruit additif dans un cadre assez général, mais qui prennent en compte les spécificités des signaux sismiques : ondelette à bande passante limitée, bruit additif, réflectivité blanche et/ou sur-gaussienne.

La communauté du traitement du signal a pour habitude d'utiliser l'hypothèse de blancheur du signal d'entrée (la réflectivité en sismique). Dans les trois premiers chapitres, nous nous sommes intéressés à cette classe d'algorithmes. Les premières méthodes étaient basées sur un modèle *a priori* gaussien des distributions. En effet, les méthodes (égalisation spectrale, Yule-Walker, Durbin) utilisent seulement les statistiques d'ordre 2, elles ne permettent donc pas l'estimation de la phase de l'ondelette. L'indétermination de phase est levée avec les statistiques d'ordre supérieur à 2. Nous avons coutume d'utiliser les statistiques d'ordre 4 (tricorrélation, trispectre) indépendamment des signaux traités. Nous proposons deux voies pour effectuer ce choix des statistiques : la mesure de blancheur par le taux d'information mutuelle, et la mesure de la non-gaussianité par la néguentropie.

Dans un premier temps, nous avons proposé d'utiliser le taux d'information mutuelle comme mesure de blancheur utilisant toutes les statistiques d'ordre supérieur. Nous avons détaillé l'algorithme MAMV qui minimise le taux d'information mutuelle de la sortie par rapport à la réponse impulsionnelle du filtre déconvolveur. Implicitement, cet algorithme utilise un modèle MA. Le parallèle avec le maximum de vraisemblance et l'estimation de la fonction score à chaque itération montrent l'adaptabilité de l'algorithme lorsqu'on choisit les statistiques d'ordre supérieur en fonction de la distribution du signal. Nous avons proposé une extension au modèle autorégressif du filtre déconvolveur avec l'algorithme ARMV. Malheureusement, ce prolongement ne donne pas satisfaction en raison d'un problème de stabilité. Des expérimentations avec des signaux simulés bruités et des signaux réels montrent que les performances de l'algorithme MAMV sont similaires aux méthodes à l'ordre 2 en raison de la présence du bruit additif gaussien sur les données. Le bruit gaussien domine le choix des statistiques donc l'algorithme utilise en fait préférentiellement les statistiques d'ordre 2. Ainsi, l'algorithme MAMV permet l'estimation de la phase de l'ondelette mais n'apporte pas d'amélioration très significative pour la robustesse vis-à-vis du bruit additif comparativement aux méthodes à l'ordre 2.

Pour continuer dans la même voie, nous avons proposé une approche fréquentielle avec l'algorithme FBD (Frequency Blind Deconvolution). Il minimise une fonction coût comportant une mesure de blancheur avec le taux d'information mutuelle de la sortie et une régularisation pour limiter l'amplification du bruit. Le RSB de sortie est amélioré en limitant les grandes valeurs de la réponse en fréquence du filtre déconvolveur. Nous avons démontré (chapitre 3) l'apport important de l'approche fréquentielle de l'algorithme FBD qui permet à la fois de s'adapter à plusieurs types de données et d'avoir un meilleur RSB de sortie comparativement aux méthodes existantes (MAMV incluse). Malgré ces résultats encourageants, nous notons que le bruit est toujours influent dans le choix des statistiques.

Le chapitre 4 s'inspire plus de la littérature de déconvolution dans le domaine géophysique et nous avons proposé d'ajuster le filtre déconvolveur suivant un critère de non-gaussianité de la réflectivité estimée. En fait, nous utilisons plus précisément la sur-gaussianité de la réflectivité. Nous avons appuyé nos réflexions sur la néguentropie, qui mesure la déviation à la gaussienne à l'aide de toutes les statistiques d'ordre supérieur. Du point de vue algorithmique, nous avons proposé l'algorithme MANeige qui maximise la néguentropie. Il se base aussi sur la blancheur, mais les équations de base utilisent la néguentropie, ainsi, il tend à limiter l'emploi des statistiques d'ordre 2 au profit des statistiques d'ordre supérieur à deux. De nombreuses méthodes proposées, principalement par la communauté travaillant sur des applications sismiques, sont équivalentes à notre algorithme MANeige avec un choix *a priori* pour la distribution de la réflectivité. Notre algorithme MANeige est le seul à réunir toutes les propriétés d'identifiabilité du système, de stabilité, d'estimation non biaisée et donne la covariance d'estimation la plus faible.

Le dernier chapitre caractérise la sensibilité des algorithmes basés sur la blancheur ou la sur-gaussianité au bruit additif blanc gaussien sur les données. Des développements théoriques démontrent que notre algorithme MANeige et aussi l'algorithme MED, tous deux basés sur la sur-gaussianité, sont beaucoup plus robustes au bruit que notre algorithme FBD, qui est, lui, le meilleur algorithme de déconvolution basé sur la blancheur. Le gain du filtre déconvolveur pour les algorithmes MED et MANeige tend vers zéro dans l'espace nul, alors que pour l'algorithme FBD il converge vers une constante. Les algorithmes MED et MANeige estiment des filtres dont le gain est proche de celui du filtrage de Wiener, tout en se plaçant dans un cadre aveugle. Les algorithmes sont comparés sur des données simulées et réelles. En particulier, pour le profil sismique réel fourni par la compagnie Shell, l'algorithme FBD est totalement inopérant. L'algorithme MED augmente seulement la résolution du réflecteur le plus énergétique et supprime les autres. L'algorithme MANeige est extrêmement performant, il fait notamment apparaître des réflecteurs à une grande profondeur. En définitive, par les signaux sismiques, nous pouvons conclure que l'hypothèse de sur-gaussianité permet d'avoir des résultats avec un bon rapport signal sur bruit de sortie et une bonne compression de l'ondelette. L'algorithme MANeige se positionne donc comme le plus performant des algorithmes.

Perspectives

En ce qui concerne les perspectives à court terme, il faudrait réaliser une comparaison plus importante entre les différents algorithmes se basant sur la sur-gaussianité. L'objectif serait de connaître *a priori* permettant le meilleur compromis entre la stabilité de l'algorithme et une bonne estimation de la sortie.

Ensuite, l'orthogonalisation de la fonction ϕ_Y proposée par Donoho pour assurer l'annulation du biais d'estimation n'a pas été analysée plus particulièrement. Tout d'abord, cette orthogonalisation permettra de généraliser l'algorithme CMED de Boadu et Brown [BB97]. En effet, l'algorithme CMED proposait d'utiliser une fonction coût définie comme une somme pondérée entre les critères de déconvolution prédictive (statistiques d'ordre 2) et de maximisation du kurtosis. Le changement de fonction ϕ_Y permet d'utiliser n'importe quel critère de sur-gaussianité et effectue un ajustement automatique de l'hyperparamètre réglant le compromis entre les deux termes.

Un autre axe de perfectionnement des méthodes de déconvolution aveugle basées sur la sur-gaussianité serait l'estimation de la fonction non-linéaire ϕ_Y en fonction de la distribution de la sortie Y . Pour l'instant, nous estimons la fonction non linéaire ϕ_Y par $\hat{\phi}_Y = \hat{\psi}_Y(y) - \frac{y}{\hat{\sigma}_Y^2}$, soit à partir de l'estimation de la fonction score ψ_Y . En fait, la fonction score comporte une grande partie linéaire proche de $\frac{y}{\hat{\sigma}_Y^2}$, donc l'estimation de ϕ_Y n'est pas précise. Nous pourrions étudier s'il est possible d'estimer directement ϕ_Y , par exemple par minimisation de l'erreur quadratique moyenne à un modèle paramétrique. Nous pouvons voir la fonction ϕ_Y comme un développement de la fonction score de Y autour de celle de la gaussienne. Pour les densités de probabilité, Gram-Charlier propose un développement d'une distribution par rapport à une distribution gaussienne. Cependant, en partant de ce dernier pour arriver à celui de la fonction score, le développement n'est pas précis. Il serait intéressant de développer directement la fonction score et non la densité de probabilité.

Un point reste à approfondir dans le chapitre 5 concernant l'étude de la robustesse au bruit des différents critères. En effet, pour l'algorithme MED, le bruit est supposé blanc et gaussien afin que le développement soit valable. Pour l'algorithme MANeGe, le développement nécessite que le bruit soit blanc, mais l'hypothèse de gaussianité n'est pas nécessaire. Ainsi, une étude théorique et pratique en présence de bruit distribué suivant une loi non gaussienne peut être intéressante. Enfin, le développement de la néguentropie est éventuellement prolongeable à un bruit non blanc en s'inspirant du travail de Pham [Pha05].

Une autre perspective de ces travaux est beaucoup plus vaste. Nous avons des méthodes de déconvolution SISO alors que nous disposons d'une antenne de capteurs, de plus ceux-ci sont parfois vectoriels (multi-composantes). Ainsi, nous avons une redondance d'informations que nous utilisons de façon très partielle en effectuant la moyenne de critères SISO. Des premiers travaux ont été menés pour prendre en compte la dimension spatiale.

Dans le cadre d'un stage de Master [Ngu06], nous avons proposé un algorithme prenant en compte la dimension spatiale pour réaliser un sur-échantillonnage temporel. Nous sommes partis du constat que l'algorithme MED était assez performant et très robuste au bruit, mais avait tendance à privilégier le réflecteur principal. Ainsi, nous proposons de sélectionner une partie des observations comportant un seul réflecteur et d'estimer le filtre

déconvolveur sur cette partie. Puis, le filtre estimé est appliqué sur l'ensemble du profil. Nous nous heurtons au problème du faible nombre d'échantillons de la partie sélectionnée, qui ne permet pas d'estimer correctement les statistiques nécessaires à l'algorithme. Si nous travaillons sur une antenne assez courte, nous pouvons supposer que les capteurs voisins mesurent le même signal mais avec des instants d'échantillonnage différents. Nous réorganisons les traces en une seule. La réorganisation se fait par un réarrangement suivant une relation d'ordre des échantillons en prenant soin de découper le signal en plusieurs lobes. Les premiers résultats obtenus sont très prometteurs. Théoriquement, nous pouvons montrer que l'algorithme de déconvolution fait apparaître des fonctions d'intercorrélations entre les traces, ce qui est très intéressant si les bruits sur chaque trace sont décorrélés.

Pour le filtrage [PM05] et l'estimation de paramètres [Mir05] des ondes (polarisation, déphasage, rapport d'amplitude), des outils dédiés aux signaux multidimensionnels ont été développés. La redondance d'information entre capteurs et/ou composantes permet d'obtenir des meilleures performances notamment de réduire l'influence du bruit sur les critères. Cette voie n'a pas été explorée pour la déconvolution. Des outils comme les quaternions et les matrices polynomiales pourraient éventuellement permettre de modéliser le problème de déconvolution de données multidimensionnelles.

Dans toutes les méthodes vues précédemment, nous cherchons à estimer la réflectivité et pas forcément l'ondelette. Une autre approche consiste à estimer l'ondelette et ensuite d'utiliser le filtrage de Wiener. En pratique, le filtrage de Wiener est très souvent utilisé pour des données réelles en raison de ses très bonnes performances. Avec un post-doctorant, nous avons proposé une estimation de l'ondelette basée sur la transformée homomorphique des données. Nous supposons que l'ondelette est invariante sur l'antenne ou sous-antenne considérée. La transformée homomorphique permet de passer du modèle multiplicatif dans le domaine fréquentiel à un modèle additif. Nous utilisons l'ensemble des traces en estimant l'ondelette avec la DVS³ (Décomposition en valeurs singulières) de la matrice constituée des transformées homomorphiques de chacune des traces. L'estimation du module de l'ondelette est de bonne qualité, en revanche le repliement de phase empêche l'estimation de la phase de l'ondelette. Pour lever cette ambiguïté, nous pourrions peut-être utiliser la transformée homomorphique pour les statistiques d'ordre supérieur. Ainsi, nous prendrons en compte le lien statistique entre les différents canaux de fréquence.

Nous nous sommes intéressés dans ce manuscrit au caractère sur-gaussien de la réflectivité et cette approche est la plus performante vis-à-vis du bruit additif. Comme le suggérait Claerbout [Cla77], la sur-gaussianité peut être vue comme une caractérisation de la parcimonie du signal. Dans ce cas, nous définissons la parcimonie à l'aide de la distribution du signal. Claerbout proposait simplement de prendre une loi *a priori* laplacienne. Il faudrait avoir une définition plus large de la parcimonie. Les récents développements réalisés dans le domaine de la séparation de sources [BZJM06, AABZJ06] pourraient constituer un point de départ.

Enfin, nous remarquons que la résolution est toujours limitée par l'impossibilité de retrouver les hautes fréquences de la séquence de réflectivité car l'ondelette a une bande

³SVD : Singular Value Decomposition en anglais

passante étroite. Un traitement avec uniquement un filtre déconvolveur linéaire ne permettra pas de retrouver certaines fréquences. Nous pourrions envisager d'utiliser les basses fréquences pour nous aider à retrouver les hautes fréquences. Seuls des traitements non linéaires peuvent permettre de telles estimations. Pour terminer, suite aux contacts avec les compagnies pétrolières, nous pouvons mentionner leur intérêt pour le développement ou l'adaptation des méthodes de déconvolution dans le domaine $\tau - p$ plus connu sous le nom de transformée de Radon. Ce domaine est utilisé pour une majorité des traitements appliqués aux données. Au contraire, la déconvolution est réalisée dans le domaine d'origine. Avec cette adaptation, la déconvolution pourrait être insérée facilement à différents niveaux de la chaîne de traitements.

Annexes

Annexe A

Développement théorique relatif au chapitre 2

A.1 Preuve du lemme 2.1

Nous souhaitons prouver que :

$$\mathcal{H}(Y) = \mathcal{H}(g \star D) = \mathcal{H}(D) + \frac{1}{2\pi} \int_0^{2\pi} \log \left| \sum_{t=-\infty}^{+\infty} g(t) \exp(-jt\theta) \right| d\theta. \quad (\text{A.1})$$

Dans un premier temps, nous devons montrer que pour deux processus stochastiques X et Y , nous avons :

$$I(X, h \star Y) = I(X, Y). \quad (\text{A.2})$$

En fait, l'information mutuelle entre deux processus est définie par :

$$I(X, h \star Y) = \mathcal{H}(X) - \mathcal{H}(X/h \star Y), \quad (\text{A.3})$$

et si nous supposons que h est un filtre inversible, connaître Y est équivalent à connaître $h \star Y$, ainsi l'incertitude sur X avec la connaissance de $h \star Y$ est égale à l'incertitude de X avec la connaissance de Y :

$$\mathcal{H}(X/h \star Y) = \mathcal{H}(X/Y). \quad (\text{A.4})$$

Avec (A.4), la relation (A.2) est immédiate. Maintenant, considérons Z un processus stochastique stationnaire indépendant de X , nous avons :

$$I(X + Z, Z) = \mathcal{H}(X + Z) + \mathcal{H}(Z) - \mathcal{H}(X + Z, Z). \quad (\text{A.5})$$

L'indépendance de X et Z implique que :

$$\mathcal{H}(X + Z, Z) = \mathcal{H}(Z) + \mathcal{H}(X). \quad (\text{A.6})$$

En combinant (A.5) et (A.6), nous obtenons :

$$I(X + Z, Z) = \mathcal{H}(X + Z) - \mathcal{H}(X). \quad (\text{A.7})$$

De même, pour $h \star Z$ et $h \star X$ qui sont des processus indépendants, nous avons :

$$I(h \star X + h \star Z, h \star Z) = \mathcal{H}(h \star X + h \star Z) - \mathcal{H}(h \star X). \quad (\text{A.8})$$

En utilisant à deux reprises (A.2), nous savons que :

$$I(h \star X + h \star Z, h \star Z) = I(X + Z, Z), \quad (\text{A.9})$$

ainsi avec (A.7) et (A.8), nous avons l'égalité suivante pour tout Z :

$$\mathcal{H}(h \star X + h \star Z) - \mathcal{H}(X + Z) = \mathcal{H}(h \star X) - \mathcal{H}(X). \quad (\text{A.10})$$

Cette expression montre que la différence entre le taux d'entropie à l'entrée et celui de la sortie est indépendant de la distribution de l'entrée donc :

$$\mathcal{H}(h \star X) - \mathcal{H}(X) = \mathcal{H}(h \star Z) - \mathcal{H}(Z). \quad (\text{A.11})$$

En particulier, si nous choisissons pour Z un processus G blanc gaussien stationnaire, nous avons donc :

$$\mathcal{H}(h \star X) - \mathcal{H}(X) = \mathcal{H}(h \star G) - \mathcal{H}(G). \quad (\text{A.12})$$

Or d'après le théorème de Szegö-Kolmogorov-Krein [Whi54], le deuxième membre est :

$$\mathcal{H}(h \star G) - \mathcal{H}(G) = \frac{1}{4\pi} \int_0^{2\pi} \log S(\theta) d\theta, \quad (\text{A.13})$$

avec $S(\theta)$ le spectre de $h \star G$. Mais, comme G est blanc nous avons :

$$S(\theta) = \left| \sum_{t=-\infty}^{t=+\infty} h(t) e^{-jt\theta} \right|. \quad (\text{A.14})$$

Finalement, avec (A.12) et (A.14) nous obtenons que :

$$\mathcal{H}(h \star X) - \mathcal{H}(X) = \frac{1}{2\pi} \int_0^{2\pi} \log \left| \sum_{t=-\infty}^{t=+\infty} h(t) e^{-jt\theta} \right|^2 d\theta. \quad (\text{A.15})$$

A.2 Calcul de (2.15)

Nous cherchons à calculer la dérivation suivante :

$$\frac{\partial}{\partial g(t)} \left\{ \frac{1}{2\pi} \int_0^{2\pi} \log \left| \sum_{t'=-\infty}^{t'=+\infty} g(t') e^{-jt'\theta} \right| d\theta \right\}. \quad (\text{A.16})$$

Si nous notons :

$$G(\theta) = \sum_{t=-\infty}^{t=+\infty} g(t) e^{-jt\theta}, \quad (\text{A.17})$$

alors (A.16) s'écrit de la façon suivante :

$$\frac{\partial}{\partial g(t)} \left\{ \frac{1}{4\pi} \int_0^{2\pi} \log G(\theta) G(-\theta) d\theta \right\}. \quad (\text{A.18})$$

Ainsi, nous avons :

$$\begin{aligned} \frac{\partial}{\partial g(t)} \left\{ \frac{1}{2\pi} \int_0^{2\pi} \log \left| \sum_{t'=-\infty}^{+\infty} g(t') e^{-jt'\theta} \right| d\theta \right\} &= \frac{1}{4\pi} \int_0^{2\pi} \frac{e^{jt\theta}}{G(-\theta)} d\theta + \frac{1}{4\pi} \int_0^{2\pi} \frac{e^{-jt\theta}}{G(\theta)} d\theta \\ &= \Re \left[\frac{1}{2\pi} \int_0^{2\pi} \frac{e^{-jt\theta}}{G(\theta)} d\theta \right], \end{aligned} \quad (\text{A.19})$$

où $\Re[z]$ est la partie réelle du complexe z . Comme le filtre $g(t)$ est réel, la partie imaginaire de l'intégrale est nulle, ainsi nous parvenons à l'expression (2.15) :

$$\frac{\partial}{\partial g(t)} \left\{ \frac{1}{2\pi} \int_0^{2\pi} \log \left| \sum_{t=-\infty}^{+\infty} g(t) e^{-jt\theta} \right| d\theta \right\} = \bar{g}(-t). \quad (\text{A.20})$$

A.3 Preuve du lemme 2.2

Si nous supposons f dérivable sur $\mathbb{R} \setminus \{x_i, i = 1 \dots N\}$. Le premier terme de (2.44) s'exprime par :

$$\begin{aligned} E[f(x)\psi_X(x)] &= \sum_{i=0}^N \int_{x_i^+}^{x_{i+1}^-} p_X(x) f(x) \psi_X(x) dx, \\ &= - \sum_{i=0}^N \int_{x_i^+}^{x_{i+1}^-} p'_X(x) f(x) dx. \end{aligned} \quad (\text{A.21})$$

Par convention, $x_0 = -\infty$ et $x_{N+1} = +\infty$. Et nous utilisons la notation suivante pour les intégrales :

$$\int_{a^+}^{b^-} g(x) dx = \lim_{\alpha \rightarrow a^+} \lim_{\beta \rightarrow b^-} \int_{\alpha}^{\beta} g(x) dx. \quad (\text{A.22})$$

En intégrant par parties chaque intégrale de (A.21), nous avons :

$$E[f(x)\psi_X(x)] = - \sum_{i=0}^N [p_X(x) f(x)]_{x_i^+}^{x_{i+1}^-} + \sum_{i=0}^N \int_{x_i^+}^{x_{i+1}^-} p_X(x) f'(x) dx. \quad (\text{A.23})$$

La continuité de la fonction f et la condition (2.43) assurent que la première somme s'annule. La seconde somme peut se regrouper en une seule intégrale, ainsi nous obtenons :

$$E[f(x)\psi_X(x)] = \int_{-\infty}^{+\infty} p_X(x) f'(x) dx = E[f'(x)]. \quad (\text{A.24})$$

Annexe B

Calcul des fonctions scores théoriques

B.1 Préliminaires

Dans cette annexe, nous détaillons le calcul de la fonction score d'un processus $X = R + N$ dans le cas de quatre distributions différentes du processus R . N est un bruit additif gaussien blanc supposé indépendant de R . Dans un premier temps, nous calculons la densité de probabilité du processus X par la relation :

$$p_X(x) = p_R \star p_N(x) = \int_{\mathbb{R}} p_R(r)p_N(x - r)dr, \quad (\text{B.1})$$

puis la fonction score sera calculée simplement par $\psi_X(x) = -\frac{d \log p_X(x)}{dx} = -\frac{p'_X(x)}{p_X(x)}$. La variance du bruit sera notée σ_N .

Dans la suite, nous aurons besoin des fonctions erf et erfc définies par :

$$\text{erf}(y) = \frac{2}{\sqrt{\pi}} \int_0^y e^{-u^2} du, \quad (\text{B.2})$$

$$\text{erfc}(y) = \frac{2}{\sqrt{\pi}} \int_y^{+\infty} e^{-u^2} du. \quad (\text{B.3})$$

Pour les limites de ces fonctions, nous avons :

$$\lim_{y \rightarrow +\infty} \text{erf}(y) = 1, \quad (\text{B.4})$$

$$\lim_{y \rightarrow +\infty} \text{erfc}(y) = 0, \quad (\text{B.5})$$

$$\lim_{y \rightarrow -\infty} \text{erf}(y) = -1, \quad (\text{B.6})$$

$$\lim_{y \rightarrow -\infty} \text{erfc}(y) = 2, \quad (\text{B.7})$$

$$\text{erf}(0) = 0, \quad (\text{B.8})$$

$$\text{erfc}(0) = 1. \quad (\text{B.9})$$

$$(\text{B.10})$$

Les dérivées de ces fonctions sont :

$$\operatorname{erf}'(y) = \frac{2}{\sqrt{\pi}} e^{-y^2}, \quad (\text{B.11})$$

$$\operatorname{erf}'(y) = -\frac{2}{\sqrt{\pi}} e^{-y^2}. \quad (\text{B.12})$$

B.2 Cas de la réflectivité laplacienne

La réflectivité suit une distribution de probabilité laplacienne de variance σ soit :

$$p_R(r) = \frac{1}{\sqrt{2}\sigma} \exp\left(-\frac{|r|\sqrt{2}}{\sigma}\right). \quad (\text{B.13})$$

Dans ce cas, la densité de probabilité du processus X résultat de la superposition de la réflectivité laplacienne et du bruit gaussien s'écrit comme :

$$p_X(x) = \int_{\mathbb{R}} p_R(r) p_N(x-r) dr = \int_{-\infty}^{+\infty} \frac{1}{2\sigma\sigma_N\sqrt{\pi}} \exp\left(-\frac{|r|\sqrt{2}}{\sigma}\right) \exp\left(-\frac{(x-r)^2}{2\sigma_N^2}\right) dr. \quad (\text{B.14})$$

Pour supprimer la valeur absolue, nous coupons l'intégrale précédente en deux parties pour les r positifs et négatifs, pour celle des r négatifs nous nous ramenons à une intégrale entre 0 et $+\infty$ en remplaçant r par $-r$. Soit, nous avons :

$$p_X(x) = \frac{1}{2\sigma\sigma_N\sqrt{\pi}} \left[\int_0^{+\infty} \exp\left(-\left\{\frac{r\sqrt{2}}{\sigma} + \frac{(x-r)^2}{2\sigma_N^2}\right\}\right) dr + \int_0^{+\infty} \exp\left(-\left\{\frac{r\sqrt{2}}{\sigma} + \frac{(x+r)^2}{2\sigma_N^2}\right\}\right) dr \right]. \quad (\text{B.15})$$

Pour exprimer les intégrales précédentes à l'aide de la fonction erf définie en (B.2), nous devons modifier les arguments des exponentielles en les écrivant comme le début du développement d'un terme au carré :

$$\frac{r\sqrt{2}}{\sigma} + \frac{(x-r)^2}{2\sigma_N^2} = \frac{1}{2\sigma_N^2} \left(r - x + \frac{\sqrt{2}\sigma_N^2}{\sigma} \right)^2 + \frac{\sqrt{2}}{\sigma} \left(x - \frac{\sigma_N^2}{\sqrt{2}\sigma} \right), \quad (\text{B.16})$$

$$\frac{r\sqrt{2}}{\sigma} + \frac{(x+r)^2}{2\sigma_N^2} = \frac{1}{2\sigma_N^2} \left(r + x + \frac{\sqrt{2}\sigma_N^2}{\sigma} \right)^2 - \frac{\sqrt{2}}{\sigma} \left(x + \frac{\sigma_N^2}{\sqrt{2}\sigma} \right). \quad (\text{B.17})$$

Si nous utilisons les développements (B.16) et (B.17) pour calculer la ddp de (B.14) nous avons besoin de l'expression d'intégrales du type suivant :

$$\frac{1}{\sqrt{2\pi}\sigma_N} \int_0^{+\infty} \exp\left(-\frac{(r+m)^2}{2\sigma_N^2}\right) dr = \frac{1}{2} \operatorname{erfc}\left(\frac{m}{\sqrt{2}\sigma_N}\right). \quad (\text{B.18})$$

Ainsi, en utilisant la relation précédente, et en développant les arguments des exponentielles de la relation (B.15) par (B.16) et (B.17), nous obtenons pour la densité de probabilité :

$$p_X(x) = \frac{1}{2\sqrt{2}\sigma} \left\{ \exp \left[-\frac{\sqrt{2}}{\sigma} \left(x - \frac{\sigma_N^2}{\sqrt{2}\sigma} \right) \right] \operatorname{erfc} \left[-\frac{x}{\sqrt{2}\sigma_N} + \frac{\sigma_N}{\sigma} \right] + \exp \left[\frac{\sqrt{2}}{\sigma} \left(x + \frac{\sigma_N^2}{\sqrt{2}\sigma} \right) \right] \operatorname{erfc} \left[\frac{x}{\sqrt{2}\sigma_N} + \frac{\sigma_N}{\sigma} \right] \right\}. \quad (\text{B.19})$$

Ensuite pour le calcul de la fonction score, il suffit de dériver la relation précédente. Pour simplifier l'écriture, nous noterons $a_1(x)$ le premier terme à l'intérieur de l'accolade et $a_2(x)$ le second. Après diverses simplifications, l'expression de la fonction score théorique est :

$$\psi_X(x) = \frac{\sqrt{2}}{\sigma} \frac{a_2(x) - a_1(x)}{a_1(x) + a_2(x)}. \quad (\text{B.20})$$

Après simplification du terme $\exp\left(\frac{\sigma_N^2}{\sigma^2}\right)$ dans les coefficients $a_1(x)$ et $a_2(x)$ nous obtenons l'expression de la fonction score :

$$\psi_X(x) = \frac{\sqrt{2}}{\sigma} \frac{\exp \left[-\frac{\sqrt{2}}{\sigma} x \right] \operatorname{erfc} \left[-\frac{x}{\sqrt{2}\sigma_N} + \frac{\sigma_N}{\sigma} \right] - \exp \left[\frac{\sqrt{2}}{\sigma} x \right] \operatorname{erfc} \left[\frac{x}{\sqrt{2}\sigma_N} + \frac{\sigma_N}{\sigma} \right]}{\exp \left[-\frac{\sqrt{2}}{\sigma} x \right] \operatorname{erfc} \left[-\frac{x}{\sqrt{2}\sigma_N} + \frac{\sigma_N}{\sigma} \right] + \exp \left[\frac{\sqrt{2}}{\sigma} x \right] \operatorname{erfc} \left[\frac{x}{\sqrt{2}\sigma_N} + \frac{\sigma_N}{\sigma} \right]}. \quad (\text{B.21})$$

Pour x tendant vers l'infini, c'est soit le terme a_1 ou a_2 qui est prépondérant suivant si nous étudions en $+\infty$ ou $-\infty$. En utilisant les limites de la fonction erfc données en (B.5) (B.7), nous avons :

$$\psi_X(x) \sim \operatorname{sign}(x) \frac{\sqrt{2}}{\sigma} \quad \text{quand } x \rightarrow \infty. \quad (\text{B.22})$$

Ainsi, la fonction score tend la fonction score de la distribution laplacienne, ce qui est normal puisque les grandes valeurs correspondent au signal.

B.3 Cas de la réflectivité Bernoulli-gaussienne

Le signal de réflectivité est décrit par de nombreux auteurs comme un signal Bernoulli-gaussien. Nous avons défini plus précisément l'écriture d'un tel processus dans le paragraphe 1.1.3. Le processus résulte du produit d'un processus de Bernoulli de paramètre λ et d'un processus gaussien de variance σ . Ainsi, la densité de probabilité de la réflectivité peut s'écrire de la façon suivante :

$$p_R(r) = (1 - \lambda)\delta(r) + \frac{\lambda}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{r^2}{2\sigma^2} \right). \quad (\text{B.23})$$

Ensuite, après ajout du bruit gaussien de variance σ_N , comme la somme de deux variables aléatoires gaussiennes indépendantes est une gaussienne donc la variance est la somme des variances, la densité de probabilité du processus X est :

$$p_X(x) = \frac{(1-\lambda)}{\sqrt{2\pi\sigma_N^2}} \exp\left(-\frac{x^2}{2\sigma_N^2}\right) + \frac{\lambda}{\sqrt{2\pi(\sigma^2 + \sigma_N^2)}} \exp\left(-\frac{x^2}{2(\sigma^2 + \sigma_N^2)}\right). \quad (\text{B.24})$$

Après simple dérivation, de la densité de probabilité, la fonction score est :

$$\psi_X(x) = \frac{\frac{(1-\lambda)x}{\sigma_N^3} \exp\left(-\frac{x^2}{2\sigma_N^2}\right) + \frac{\lambda x}{(\sigma^2 + \sigma_N^2)^{3/2}} \exp\left(-\frac{x^2}{2(\sigma^2 + \sigma_N^2)}\right)}{\frac{1-\lambda}{\sigma_N} \exp\left(-\frac{x^2}{2\sigma_N^2}\right) + \frac{\lambda}{\sqrt{\sigma^2 + \sigma_N^2}} \exp\left(-\frac{x^2}{2(\sigma^2 + \sigma_N^2)}\right)}. \quad (\text{B.25})$$

Pour que les réflecteurs soient visibles dans le bruit il faut que $\sigma \gg \sigma_N$. En utilisant cette remarque, nous pouvons dire que pour les x proches de zéros, la fonction score $\psi_X(x)$ se comporte comme x/σ_N^2 , soit comme la fonction score du bruit. En revanche, pour les grandes valeurs de x , la fonction score $\psi_X(x)$ est équivalente à $x/(\sigma^2 + \sigma_N^2)$ soit approximativement x/σ^2 . Cette droite représente la fonction score théorique du processus gaussien modélisant l'amplitude des réflecteurs.

B.4 Cas de réflectivité Bernoulli-laplacienne

Quelques approches choisissent de modéliser la réflectivité par un processus Bernoulli-laplacienne. Le signal de réflectivité est le produit d'un processus de Bernoulli et d'un processus laplacien. De la même façon que pour le signal Bernoulli-gaussien, la réflectivité a la distribution suivante :

$$p_R(r) = (1-\lambda)\delta(r) + \frac{\lambda}{\sqrt{2}\sigma} \exp\left(-\frac{|r|\sqrt{2}}{\sigma}\right). \quad (\text{B.26})$$

Ensuite, pour calculer la distribution du processus $X = R + N$, nous allons devoir calculer la convolution entre les densités de probabilités $p_R(r)$ de R et $p_N(n)$ de N . Il apparaît la convolution entre une loi laplacienne et une gaussienne. Ce calcul est fait dans le paragraphe B.2, nous pourrions réutiliser l'expression (B.19). Ainsi, la densité de probabilité de la somme d'une réflectivité Bernoulli-laplacienne et d'un bruit gaussien est :

$$p_X(x) = \frac{(1-\lambda)}{\sqrt{2\pi\sigma_N^2}} \exp\left(-\frac{x^2}{2\sigma_N^2}\right) + \frac{\lambda}{2\sqrt{2}\sigma} \left\{ \exp\left[-\frac{\sqrt{2}}{\sigma}\left(x - \frac{\sigma_N^2}{\sqrt{2}\sigma}\right)\right] \operatorname{erfc}\left[-\frac{x}{\sqrt{2}\sigma_N} + \frac{\sigma_N}{\sigma}\right] + \exp\left[\frac{\sqrt{2}}{\sigma}\left(x + \frac{\sigma_N^2}{\sqrt{2}\sigma}\right)\right] \operatorname{erfc}\left[\frac{x}{\sqrt{2}\sigma_N} + \frac{\sigma_N}{\sigma}\right] \right\}. \quad (\text{B.27})$$

Ensuite, nous devons dériver la relation précédente pour obtenir la fonction score. Pour la dérivation de la partie entre accolade comprenant des exponentielles et fonction erfc, le

calcul a déjà été fait pour le cas de la fonction score de la réflectivité laplacienne. Ainsi, la fonction score a l'expression suivante :

$$\psi_X(x) = \frac{\frac{(1-\lambda)x}{\sqrt{2\pi\sigma_N^2}} \exp\left(-\frac{x^2}{2\sigma_N^2}\right) + \frac{\lambda \exp\left[\frac{\sigma_N^2}{\sigma^2}\right]}{2\sqrt{2}\sigma^2} \left\{ \exp\left[-\frac{\sqrt{2}}{\sigma}x\right] \operatorname{erfc}\left[-\frac{x}{\sqrt{2}\sigma_N} + \frac{\sigma_N}{\sigma}\right] - \exp\left[\frac{\sqrt{2}}{\sigma}x\right] \operatorname{erfc}\left[\frac{x}{\sqrt{2}\sigma_N} + \frac{\sigma_N}{\sigma}\right] \right\}}{\frac{(1-\lambda)}{\sqrt{2\pi\sigma_N^2}} \exp\left(-\frac{x^2}{2\sigma_N^2}\right) + \frac{\lambda \exp\left[\frac{\sigma_N^2}{\sigma^2}\right]}{4\sigma} \left\{ \exp\left[-\frac{\sqrt{2}}{\sigma}x\right] \operatorname{erfc}\left[-\frac{x}{\sqrt{2}\sigma_N} + \frac{\sigma_N}{\sigma}\right] + \exp\left[\frac{\sqrt{2}}{\sigma}x\right] \operatorname{erfc}\left[\frac{x}{\sqrt{2}\sigma_N} + \frac{\sigma_N}{\sigma}\right] \right\}}. \quad (\text{B.28})$$

B.5 Cas de la réflectivité uniformément distribuée

Nous supposons que la réflectivité est uniformément distribué entre $[-a; a]$ où a est un réel positif. Si nous souhaitons que la réflectivité soit de puissance unitaire, nous prenons $a = \sqrt{3}$. En reprenant (B.1) dans notre cas particulier nous avons :

$$p_X(x) = \int_{\mathbb{R}} p_R(r)p_N(x-r)dr = \int_{-a}^a \frac{1}{2a} \frac{1}{\sqrt{2\pi\sigma_N^2}} e^{-\frac{(x-r)^2}{2\sigma_N^2}} dr. \quad (\text{B.29})$$

Nous effectuons dans l'intégrale précédente le changement de variable $u = \frac{r-x}{\sqrt{2}\sigma_N}$, nous obtenons :

$$p_X(x) = \frac{1}{2\sqrt{\pi}a} \int_{-\frac{a+x}{\sqrt{2}\sigma_N}}^{\frac{a-x}{\sqrt{2}\sigma_N}} e^{-u^2} du. \quad (\text{B.30})$$

En s'appuyant sur la définition (B.2) de la fonction erf, la densité de probabilité $p_X(x)$ s'écrit :

$$p_X(x) = \frac{1}{4a} \left[\operatorname{erf}\left(\frac{a-x}{\sqrt{2}\sigma_N}\right) - \operatorname{erf}\left(-\frac{a+x}{\sigma_N}\right) \right]. \quad (\text{B.31})$$

Après dérivation, la fonction score s'exprime de la façon suivante :

$$\psi_X(x) = \sqrt{\frac{2}{\pi\sigma_N^2}} \frac{\exp\left(\frac{a-x}{\sqrt{2}\sigma_N}\right) - \exp\left(-\frac{a+x}{\sigma_N}\right)}{\operatorname{erf}\left(\frac{a-x}{\sqrt{2}\sigma_N}\right) - \operatorname{erf}\left(-\frac{a+x}{\sigma_N}\right)}. \quad (\text{B.32})$$

Annexe C

Règles de dérivation et gradient par rapport à une variable complexe

C.1 Définitions et relations générales

Soit z une fonction complexe d'une variable complexe x , nous définissons la dérivée de z par rapport à x par [Bra83, VDB94] :

$$\frac{\partial z}{\partial x} = \frac{1}{2} \left(\frac{\partial z}{\partial x_R} - j \frac{\partial z}{\partial x_I} \right), \quad (\text{C.1})$$

où x_R et x_I sont les parties réelles et imaginaires de x . Cette définition est écrite comme si les parties réelle et imaginaire étaient indépendantes, *i.e.* que la dérivée d'un terme dépendant seulement de la partie réelle x_R en fonction de la partie imaginaire x_I est égal à zéro. Parfois, dans la définition (C.1) nous omettons le coefficient $1/2$, sa présence permet de conserver les identités classiques de la dérivation réelle par exemple $\frac{\partial z^n}{\partial z} = n z^{n-1}$. En revanche, si z est une fonction réelle d'une variable réelle, vu comme une fonction complexe d'une variable complexe ($x_R = x$), nous avons :

$$\left(\frac{\partial z}{\partial x} \right)_{(z,x) \in \mathbb{C}^2} = \frac{1}{2} \left(\frac{\partial z}{\partial x} \right)_{(z,x) \in \mathbb{R}^2}. \quad (\text{C.2})$$

Avec la définition (C.1), nous avons :

$$\frac{\partial z}{\partial x^*} = \frac{1}{2} \left(\frac{\partial z}{\partial x_R} + j \frac{\partial z}{\partial x_I} \right). \quad (\text{C.3})$$

Nous avons les identités remarquables suivantes :

$$\left(\frac{\partial z^*}{\partial x} \right)^* = \frac{\partial z}{\partial x}, \quad (\text{C.4})$$

$$\frac{\partial yz}{\partial x} = y \frac{\partial z}{\partial x} + z \frac{\partial y}{\partial x}. \quad (\text{C.5})$$

La règle de dérivation en chaîne est :

$$\frac{\partial z(x)}{\partial w} = \frac{\partial z}{\partial x} \frac{\partial x}{\partial w} + \left(\frac{\partial z^*}{\partial x} \right)^* \frac{\partial x^*}{\partial w}. \quad (\text{C.6})$$

C.2 Définition du gradient

Soit f une fonction réelle d'un vecteur complexe $\mathbf{x} = [x_1, \dots, x_n]^T$, nous définissons le gradient complexe par :

$$\nabla f(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}^*} = \begin{bmatrix} \frac{\partial f}{\partial x_1^*} \\ \vdots \\ \frac{\partial f}{\partial x_n^*} \end{bmatrix}. \quad (\text{C.7})$$

Il est normal à la surface définie par $f(\mathbf{x}) = \text{constant}$ et est nul pour tous les extremum de f . Il peut être utilisé pour la recherche de minimum par une technique de gradient.

Annexe D

Performances asymptotiques de l'algorithme MANeige

D.1 Matrice de covariance

Nous souhaitons calculer la matrice de covariance asymptotique du vecteur estimé $\widehat{\mathbf{g}}_T = [\hat{g}(1), \dots, \hat{g}(P)]^T$ avec T échantillons¹. La dénomination asymptotique signifie que le nombre d'échantillons T tend vers l'infini. Nous noterons $\mathbf{g}^0 = [g^0(1), \dots, g^0(P)]^T$ la vraie valeur à estimer. Le coefficient pour le retard nul est omis car, nous normalisons le filtre pour avoir $\hat{g}(0) = g^0(0) = 1$. La matrice de covariance est définie comme :

$$\Sigma = \lim_{T \rightarrow \infty} E[(\widehat{\mathbf{g}}_T - \mathbf{g}^0)(\widehat{\mathbf{g}}_T - \mathbf{g}^0)^T]. \quad (\text{D.1})$$

Pour l'écriture des vecteurs, nous avons fait le choix de réponse impulsionnelle causale, mais les équations restent valables pour une réponse non causale. Dans le cadre des M -estimateurs, la matrice de covariance asymptotique peut être calculée par [Hub77] :

$$\Sigma = \lim_{T \rightarrow \infty} T H_T^{-1} E[\nabla_T \nabla_T] H_T^{-1}, \quad (\text{D.2})$$

où ∇_T est le vecteur gradient estimé du critère avec T échantillons, soit dans notre cas :

$$(\nabla_T)_i = \frac{1}{T} \sum_t \phi_Y(y(t)) d(t-i), \quad (\text{D.3})$$

et H_T est la matrice du Hessien du critère. Dans notre cas :

$$(H_T)_{ij} = \frac{1}{T} E \left[\sum_t \phi'_Y(y(t)) d(t-j) d(t-i) \right]. \quad (\text{D.4})$$

Mais, l'expression (D.2) est valable uniquement dans le cas où le filtre direct vérifie $\sum_u w(u)^2 = 1$. Ainsi, si cette condition n'est pas vérifiée nous normalisons le filtre direct soit $w(u)$ est remplacé par $\frac{w(u)}{\sqrt{\sum_u w(u)^2}}$ et donc dans le même temps le filtre déconvolveur subit

¹Le lecteur devra se méfier de la confusion entre T le nombre d'échantillons et l'opérateur de transposition T

la transformation inverse de $g(t)$ en $g(t)\sqrt{\sum_u w(u)^2}$. Par conséquent dans le cas général la matrice de corrélation asymptotique s'écrit :

$$\Sigma = \sum_t w(t)^2 \lim_{T \rightarrow \infty} TH_T^{-1} E[\nabla_T \nabla_T] H_T^{-1}. \quad (\text{D.5})$$

Dans un premier temps, nous nous intéressons au calcul du terme central de l'équation (D.5). Nous avons :

$$E(\nabla_T \nabla_T^T)_{ij} = E \left[\frac{1}{T} \sum_t \phi_Y(y(t)) d(t-i) \frac{1}{T} \sum_s \phi_Y(y(s)) d(s-i) \right]. \quad (\text{D.6})$$

Ensuite pour poursuivre, nous supposons que la convergence se fait sans biais ainsi, la réflectivité estimée $y(t)$ est identique à la réflectivité théorique $r(t)$, par conséquent nous pouvons lier les observations $d(t)$ et la réflectivité estimée $y(t)$ par le filtre direct soit l'ondelette. Nous avons donc :

$$E(\nabla_T \nabla_T^T)_{ij} = E \left[\frac{1}{T^2} E \sum_t \phi_Y(y(t)) \sum_{t'} w(t') y(t-i-t') \frac{1}{T} \sum_s \phi_Y(y(s)) \sum_{s'} w(s') y(s-i-s') \right]. \quad (\text{D.7})$$

Nous pouvons récrire l'équation précédente de façon plus condensé par :

$$E(\nabla_T \nabla_T^T)_{ij} = \frac{1}{T^2} \sum_{tst's'} w(t') w(s') E[\phi_Y(y(t)) \phi_Y(y(s)) y(t-i-t') y(s-j-s')]. \quad (\text{D.8})$$

Dans cette somme sur quatre indices, l'espérance mathématique de beaucoup de termes est nulle. Nous utiliserons l'hypothèse que la sortie Y est iid ainsi, nous pourrions séparer les espérances des termes ne se situant pas au même instant. De plus, d'après les contraintes fixées par Donoho, nous savons que $E[\phi_Y(y(t))] = 0$ et $E[y(t)\phi_Y(y(t))] = 0$. Il ne faut pas oublier aussi que la moyenne de $y(t)$ est nulle : $E[y(t)] = 0$. Par conséquent, il reste dans la somme de (D.8) seulement les termes pour $t = s$ et $t-i-t' = s-j-s'$, soit $t = s$ et $i+t' = j+s'$. Dans la suite, nous posons $u = i+t' = j+s'$. Ainsi, (D.8) s'écrit simplement :

$$E(\nabla_T \nabla_T^T)_{ij} = \frac{1}{T^2} \sum_{tu} w(u-i) w(u-j) E[\phi_Y(y(t))^2 y(t-u)^2]. \quad (\text{D.9})$$

Ensuite, dans la somme tous les termes s'écriront sous la forme identique $E[\phi_Y(y)^2] E[y^2]$ sauf pour $u = 0$ car $\phi_Y(y(t))^2$ et $y(t)^2$ ne sont pas indépendants. Nous obtenons :

$$E(\nabla_T \nabla_T^T)_{ij} = \frac{1}{T^2} \left\{ \sum_t w(-i) w(-j) E[\phi_Y(y)^2 y^2] + \sum_{t,u \neq 0} w(u-i) w(u-j) E[\phi_Y(y)^2] E[y^2] \right\}. \quad (\text{D.10})$$

Comme les termes ne dépendent pas de t , nous sommes par rapport à t T termes égaux. De plus, pour avoir une somme sur u quelconque, nous ajoutons et retirons le terme pour $u = 0$, et nous effectuons le changement de variable $v = u - i - j$, pour obtenir :

$$E(\nabla_T \nabla_T^T)_{ij} = \frac{1}{T} E[\phi_Y(y)^2] E[y^2] \sum_t w(t)^2 \{\mathbf{R}_{ij} + e_i e_j U(\phi_Y, Y)\}, \quad (\text{D.11})$$

avec $\mathbf{R}_{ij} = \frac{\sum_v w(i+v)w(j+v)}{\sum_t w(t)^2}$ et $e_i = \frac{w(-i)}{\sqrt{\sum_t w(t)^2}}$, et $U(\phi_Y, Y) = \frac{E[\phi_Y(y)^2 y^2]}{E[\phi_Y(y)^2]E[y^2]} - 1$. La matrice \mathbf{R}_1 est définie par son terme générique :

$$(\mathbf{R}_1)_{ij} = \mathbf{R}_{ij} + e_i e_j U(\phi_Y, Y). \quad (\text{D.12})$$

Ensuite, nous nous intéressons au terme du hessien (D.4) qui a la même expression que le terme $E(\nabla_T \nabla_T^T)_{ij}$ dans (D.6) en remplaçant, $\phi_Y(y)^2$ par $\phi'_Y(y)$. Néanmoins, nous devons tout récrire pour faire une démonstration correcte car ces deux fonctions ne vérifient pas les mêmes propriétés. Nous avons, après utilisation de l'équation de convolution $d = w \star y$:

$$(H_T)_{ij} = \frac{1}{T} \sum_{tt's'} E[w(s')w(t')\phi'_Y(y(t))y(t-j-s')y(t-i-t')]. \quad (\text{D.13})$$

Puis, nous utilisons l'hypothèse de blancheur sur Y pour écrire :

$$(H_T)_{ij} = \frac{1}{T} \left\{ \sum_t E[w(-j)w(-i)\phi'_Y(y(t))y(t)y(t)] + \sum_{t,u \neq 0} w(u-j)w(u-i)E[\phi'_Y(y(t))]E[y(u)^2] \right\}. \quad (\text{D.14})$$

Par les mêmes techniques que permettant d'obtenir (D.11), nous avons :

$$(H_T)_{ij} = E[\phi'_Y(y)] E[y^2] \sum_t w(t)^2 \{\mathbf{R}_{ij} + e_i e_j W(\phi_Y, Y)\}, \quad (\text{D.15})$$

avec \mathbf{R}_{ij} et e_i défini précédemment et $W(\phi_Y, Y) = \frac{E[\phi'_Y(y)y^2]}{E[\phi'_Y(y)]E[y^2]} - 1$. La matrice \mathbf{R}_0 est définie par son terme générique suivant :

$$(\mathbf{R}_0)_{ij} = \mathbf{R}_{ij} + e_i e_j W(\phi_Y, Y). \quad (\text{D.16})$$

Finalement, nous obtenons pour la matrice de covariance asymptotique l'expression suivante :

$$\Sigma = \frac{E[\phi_Y^2(Y)]}{(E[\phi'_Y(Y)])^2 \sigma^2} \mathbf{R}_0^{-1} \mathbf{R}_1 \mathbf{R}_0^{-1}. \quad (\text{D.17})$$

D.2 Approximations et simplifications

Donoho [Don81] propose de simplifier les expressions provenant du gradient (D.11) et celle du hessien (D.15) en négligeant le second terme de l'accolade dépendant de $U(\phi_Y, Y)$ ou $W(\phi_Y, Y)$. En effet, $trace(\mathbf{R}) = P$ la longueur du filtre, alors que $trace(ee^T) = 1$. Pour

justifier, l'approximation, il faut montrer de plus que $U(\phi_Y, Y)$ et $W(\phi_Y, Y)$ ne sont pas très grands. Nous pouvons utiliser les inégalités de Hölder définie par :

$$|E[AB]| \leq E[|A|^{1/p}]^p E[|B|^{1/q}]^q \quad \text{avec } p + q = 1. \quad (\text{D.18})$$

Les inégalités de Hölder prolonge l'inégalité de Schwarz, en effet si $p = q = \frac{1}{2}$, nous obtenons l'inégalité de Schwarz. Ensuite, si nous prenons $B = 1$ et $p \leq 1$ et $q = 1 - p$, nous avons l'inégalité de Jensen :

$$|E[A]| \leq E[|A|^{1/p}]^p. \quad (\text{D.19})$$

Ainsi, pour $U(\phi_Y, Y)$, en utilisant l'inégalité de Hölder (D.18) pour $A = \phi_Y(y)^2$ et $B = y^2$ avec $p = q = \frac{1}{2}$ nous avons :

$$U(\phi_Y, Y) = \frac{E[\phi_Y(y)^2 y^2]}{E[\phi_Y(y)^2] E[y^2]} - 1 \leq \frac{E[|\phi_Y(y)|]^2 E[|y|^2]}{E[\phi_Y(y)^2] E[y^2]} - 1. \quad (\text{D.20})$$

Ensuite, nous utilisons l'inégalité de Jensen (D.19) pour $p = \frac{1}{2}$ et $A = |\phi_Y(y)|$ puis $A = |y|$ pour montrer que :

$$E[|\phi_Y(y)|] \leq E[|\phi_Y(y)|^2]^{1/2}, \quad (\text{D.21})$$

$$E[|y|] \leq E[|y|^2]^{1/2}. \quad (\text{D.22})$$

Les deux inégalités ci-dessous sont justifiables aussi par l'inégalité de Schwarz. En utilisant les inégalités (D.21) et (D.22), pour majorer (D.20), nous pouvons proposer l'encadrement suivant :

$$-1 \leq U(\phi_Y, Y) \leq 0. \quad (\text{D.23})$$

Pour la fonction $W(\phi_Y, Y)$, nous utilisons la même démarche, cependant, il faut tenir compte que la fonction $\phi'_Y(y)$ peut être négative. Ainsi, la valeur absolue de l'inégalité (D.18) doit être conservée dans les inégalités. En effet, pour $W(\phi_Y, Y)$, nous avons omis la valeur absolue pour obtenir (D.20). Suivant, cette remarque, nous obtenons que :

$$|W(\phi_Y, Y)| \leq 1 + \frac{|E[\phi'_Y(y)y^2]|}{E[\phi'_Y(y)] E[y^2]}. \quad (\text{D.24})$$

Nous notons d'après la contrainte (R2) posé par Donoho, nous avons $E[\phi'_Y(y)] > 0$. Et nous avons aussi besoin de l'inégalité : $E[\phi'_Y(y)] < E[|\phi'_Y(y)|]$, pour parvenir à la majoration suivante :

$$|W(\phi_Y, Y)| \leq 2. \quad (\text{D.25})$$

En définitive, pour le terme provenant du gradient (D.11), nous pouvons négliger le terme second terme $e_i e_j U(\phi_Y, Y)$ par rapport à \mathbf{R}_{ij} ainsi la matrice \mathbf{R}_1 est approximée par la matrice \mathbf{R} . De même pour le hessien (D.15), le terme $e_i e_j W(\phi_Y, Y)$ est négligeable par rapport à \mathbf{R}_{ij} ainsi la matrice \mathbf{R}_0 est approximée par la matrice \mathbf{R} . A l'aide de ces deux

approximations, nous obtenons finalement pour la matrice de covariance asymptotique (D.5) s'écrit :

$$\Sigma = \frac{V(\phi_Y, Y)}{\sigma^2} \mathbf{R}^{-1}, \quad (\text{D.26})$$

où σ^2 est la variance de Y et

$$V(\phi_Y, Y) = \frac{E[\phi_Y^2(Y)]}{(E[\phi_Y'(Y)])^2}. \quad (\text{D.27})$$

Bibliographie

- [AABZJ06] A. Ali-Amini, M. Babaie-Zadeh et C. Jutten. A new approach for sparse decomposition and sparse source separation. In *EUSIPCO*, Florence, September 2006.
- [AGL96a] P-O. Amblard, M. Gaeta et J-L. Lacoume. Statistics for complex variables and signals -part I : Variables. *Signal Processing*, 53 :1–13, 1996.
- [AGL96b] P-O. Amblard, M. Gaeta et J-L. Lacoume. Statistics for complex variables and signals -part II : Signals. *Signal Processing*, 53 :15–25, 1996.
- [Aka73] H. Akaike. Information theory as an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory*, pages 267–281, Budapest, 1973.
- [Bar48] M.S. Bartlett. Smoothing periodograms from time series with continuous spectra. *Nature*, 161 :686–687, May 1948.
- [BB97] F.K. Boadu et R.J. Brown. Constrained minimum entropy deconvolution. *Canadian Journal of Exploration Geophysics*, 33 :32–45, December 1997.
- [BLF53] A. Blanc-Lapierre et R. Fortet. *Théorie des fonctions aléatoires*. Masson edition, 1953.
- [BLP81] A. Blanc-Lapierre et B. Picinbono. *Fonctions aléatoires*. Paris, Masson edition, 1981.
- [Bou95] M. Boumahdi. *Déconvolution aveugle utilisant les statistiques d'ordre supérieur : application à des données sismiques*. PhD thesis, INPG, Grenoble, Juin 1995.
- [Bou96] M. Boumahdi. Blind identification using the kurtosis with applications to field data. *Signal Processing*, 48(3) :205–216, 1996.
- [BR67] D.R. Brillinger et M. Rosenblatt. Asymptotic theory of k-th order spectra. In Harris Ed., editor, *Spectral Analysis of Time Series*, pages 153–188. Wiley, 1967.
- [Bra83] D.H. Brandwood. A complex gradient operator and its application in adaptive array theory. *IEE Proceedings*, 130(1) :11–16, February 1983.
- [Bre05] F. Brenguier. *Imagerie sismique de haute résolution pour l'évaluation des risques naturels*. PhD thesis, Université Joseph Fourier, Grenoble, 2005.
- [Bri65] D.R. Brillinger. An introduction to polyspectra. In *Ann. Math. Stat.*, volume 36, pages 1351–1374, 1965.
- [Bri81] D.R. Brillinger. *Times Series, Data analysis and Theory*. Holden-Day, 1981.

- [BT58] R.B. Blackman et J.W. Tuckey. *The measure of power spectra from the view of communication engineering*. Dover Publication Inc., New York, 1958.
- [BV00] J-F. Bercher et C. Vignat. Estimating the entropy of a signal with applications. *IEEE Trans. on Signal Processing*, 48(6) :1687–1694, June 2000.
- [BZ02] M. Babaie-Zadeh. *On blind source separation in convolutive and nonlinear mixtures*. PhD thesis, Institut National Polytechnique de Grenoble, 2002.
- [BZJM06] M. Babaie-Zadeh, C. Jutten et A. Mansour. Sparse ICA via cluster-wise PCA. *Neurocomputing*, To appear, 2006.
- [Cad96] J.A. Cadzow. Blind deconvolution via cumulant extrema. *IEEE Signal Processing Magazine*, pages 24–42, May 1996.
- [Car98] J-F. Cardoso. Blind signal separation : statistical principles. *Proceedings of the IEEE*, 86(10) :2009–2025, October 1998.
- [Car99] J-F Cardoso. High-order contrasts for independent component analysis. *Neural Computation*, 11(1) :157–192, 1999.
- [CCA00] S. Choi, A. Cichocki et S. Amari. Flexible independent component analysis. *Journal of VLSI Signal Processing*, 20 :25–38, 2000.
- [CGI96] F. Champagnat, Y. Goussard et J. Idier. Unsupervised deconvolution of sparse spike trains using stochastic approximation. *IEEE Trans. on Geoscience and Remote Sensing*, 44(12) :2988–2998, December 1996.
- [Cla77] J.F. Claerbout. Parsimonious deconvolution. *Stanford Exploration Project*, (13) :1–9, 1977.
- [Com92] P. Comon. MA identification using fourth order cumulants. *Signal Processing*, 26, 1992.
- [Com94] P. Comon. Independent component analysis, a new concept ? *Signal Processing*, 36(3) :287–314, April 1994.
- [Com95] P. Comon. Quelques développements récents en Traitement du Signal. Habilitation à Diriger les Recherches, Université de Nice, UFR Sciences, 18 sept 1995.
- [Cra46] H. Cramér. *Mathematical methods of statistics*. Princeton Univ. Press, Princeton, 1946.
- [CS93] J-F. Cardoso et A. Souloumiac. Blind beamforming for non gaussian signals. *IEE-Proceedings*, 140(6) :362–370, December 1993.
- [CT91] T.M. Cover et J.A. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications, 1991.
- [Cul87] P.Mc. Cullagh. *Tensor Methods in Statistics*. Chapman and Hall Ltd, 1987.
- [DD97] A. Doucet et P. Duvaut. Bayesian estimation of state-space models applied to deconvolution of Bernoulli-Gaussian processes. *Signal Processing*, 57 :147–161, 1997.
- [Don81] D.L. Donoho. *On minimum entropy deconvolution*. Applied Time Series Analysis II, New York, 1981.

- [Doo53] J.L. Doob. *Stochastic processes*. Wiley, New York, 1953.
- [EHP⁺04] D. Erdogmus, K.E. Hild, J. Principe, M Lazaro et I. Santamaria. Adaptive blind deconvolution of linear channels using Renyi's entropy with Parzen window estimation. *IEEE Trans. on Signal Processing*, 52(6) :1489–1498, June 2004.
- [EKK00] J. Eriksson, J. Karvanen et V. Koivunen. Source distribution adaptive maximum likelihood estimation of ICA model. In *Proc. of the Int. Conf. ICA'00*, pages 227–232, Helsinki, 2000.
- [ER99] R. Everson et S. Roberts. Independent component analysis : A flexible nonlinearity and decorrelating manifold approach. *Neural computation*, 11 :1957–1983, 1999.
- [GD87] Y. Goussard et G. Demoment. Détection-estimation récursive rapide de séquences Bernoulli-gaussiennes. *Traitement du Signal*, 4(5) :377–388, 1987.
- [GD89] Y. Goussard et G. Demoment. Recursive deconvolution of Bernoulli-Gaussian processes using a MA representation. *IEEE Trans. on Geoscience and Remote Sensing*, 27(4) :384–394, July 1989.
- [Gia87] G.B. Giannakis. Cumulants : A powerful tool in signal processing. *Proceedings of the IEEE*, 75(9) :1333–1334, September 1987.
- [GIM89] G.B. Giannakis, Y. Inouye et J.M. Mendel. Cumulant based identification of multichannel Moving-Average models. *IEEE Trans. on Automatic Control*, 34(7) :783–787, July 1989.
- [GM89] G.B. Giannakis et J.M. Mendel. Identification of nonminimum phase systems using higher order statistics. *IEEE Trans. on Acoustics Speech and Signal Processing*, 37(3) :360–377, March 1989.
- [God78] R. Godfrey. An information theory approach to deconvolution. *Stanford Exploration Project*, 15 :157–181, 1978.
- [Gra79] W. Gray. *Variable norm deconvolution*. PhD thesis, Stanford University, 1979.
- [GS90] G.B. Giannakis et A. Swami. On estimating noncausal nonminimum phase ARMA of non-Gaussian process. *IEEE Trans. on Acoustics Speech and Signal Processing*, 38(3) :478–494, March 1990.
- [HJP03] S. Hosseini, C. Jutten et D-T. Pham. Markovian source separation. *IEEE Trans. on Signal Processing*, 51(12) :3009–3019, 2003.
- [HLR99] C. Huet et J. Le Roux. Linear system blind identification based on fourth order spectral analysis. *Signal Processing*, 77(2) :209–228, 1999.
- [Hub77] P.J. Huber. *Robust statistical procedures*. Society Independent of Applied Mathematics, 1977.
- [Hue97] C. Huet. *Factorisation des spectres d'ordre supérieur de signaux complexes, application en identification aveugle de systèmes et reconstruction d'images*. PhD thesis, Université de Nice, 1997.

- [IG90] J. Idier et Y. Goussard. Stack algorithm for recursive deconvolution of Bernoulli-Gaussian processes. *IEEE Trans. on Geoscience and Remote Sensing*, 28(5) :975–978, September 1990.
- [IG93] J. Idier et Y. Goussard. Multichannel seismic deconvolution. *IEEE Trans. on Geoscience and Remote Sensing*, 31(5) :961–980, October 1993.
- [Kaa98] K.F. Kaaresen. Evaluation and applications of the iterated maximization method for sparse deconvolution. *IEEE Trans. on Signal Processing*, 46(3) :609–624, March 1998.
- [KM82] J.J. Kormylo et J.M. Mendel. Maximum likelihood detection and estimation of Bernoulli-Gaussian processes. *IEEE Trans. on Information Theory*, 28(3) :482–488, May 1982.
- [KS63] M.G. Kendall et A. Stuart. *The Advanced Theory of Statistics, Volume 1 - Distribution Theory, second edition*. Charles Griffin and Company Ltd, London, 1963.
- [LAC97] J-L. Lacoume, P-O. Amblard et P. Comon. *Statistiques d'ordre supérieur pour le traitement du signal*. Masson edition, 1997.
- [Lan90] A. Lannes. Remarkable algebraic structure of phase closure imaging and their algorithmic implications in apertures synthesis. *Journal. Opt. Soc. Amer.*, 7 :500–512, 1990.
- [Lar03] A. Larue. Séparation de sources markoviennes. Master's thesis, Université Paris Sud, Orsay, Juin 2003.
- [Lav91] M. Lavielle. 2-D Bayesian deconvolution. *Geophysics*, 56(12) :2008–2018, 1991.
- [Lav93] M. Lavielle. Bayesian deconvolution of Bernoulli-Gaussian processes. *Signal Processing*, 33 :67–79, 1993.
- [Lav95] M. Lavielle. A stochastic algorithm for parametric and non parametric estimation in the case of incomplete data. *Signal Processing*, 42 :3–17, 1995.
- [LC77] L.R. Lines et R.W. Clayton. A new approach to vibroseis deconvolution. *Geophysical Prospecting*, 25 :417–433, 1977.
- [LGM02] P. Lesage, F. Glangeaud et J. Mars. Applications of autoregressive models and time-frequency analysis to the study of a volcanic tremor and long-period events. *Journal of Volcanology and Geothermal Research*, 114(3) :391–417, 2002.
- [LH98] S-F. Lei et R.P. Hamernik. Construction of a joint peak-interval histogram using higher-order cumulant-based inverse filtering. In *International Conference on Acoustics Speech and Signal Processing*, volume 6, pages 3617–3620, Seattle, May 1998.
- [LJ05] J. Larue, A. and Mars et C. Jutten. Déconvolution aveugle dans le domaine fréquentiel fondée sur le taux d'information mutuelle. In *GRETSI*, pages 1101–1104, Louvain-la-Neuve, Septembre 2005.
- [LJH04] A. Larue, C. Jutten et S. Hosseini. Markovian source separation in post non-linear mixtures. In *Proc. of the Int. Conf. ICA'04*, pages 702–709, Grenade, September 2004.

- [LMJ04] A. Larue, J. Mars et C. Jutten. Blind frequency deconvolution : A new approach using mutual information rate. In *74th Annual Meeting of SEG*, volume 23, pages 1941–1945, Denver, October 2004.
- [LMJ06] A. Larue, J. Mars et C. Jutten. Frequency-domain blind deconvolution based on mutual information rate. *IEEE Trans. on Signal Processing*, 54(5) :1771–1781, mai 2006.
- [LP06] A. Larue et D-T. Pham. Comparison of supergaussianity and whiteness assumptions for blind deconvolution in noisy context. In *EUSIPCO*, Florence, September 2006.
- [LVDBMJ05] A. Larue, M. Van Der Baan, J. Mars et C. Jutten. Sparsity or whiteness : what criterion to use for blind deconvolution of seismic data ? In *74th Annual Meeting of SEG*, volume 24, pages 1642–1646, Denver, November 2005.
- [LY98] H. Luo et L. Yanda. The application of blind channel identification techniques to prestack seismic deconvolution. *Proceedings of the IEEE*, 86(10) :2082–2089, October 1998.
- [MGC99] J-L. Mari, F. Glangeaud et F. Coppens. *Signal processing for geologists and geophysicists*. Technip edition, 1999.
- [Mir05] S. Miron. *Méthodes multilinéaires et hypercomplexes en traitement d'antenne multicomposante haute résolution*. PhD thesis, Institut National Polytechnique de Grenoble, Grenoble, Octobre 2005.
- [MJ99] A. Mansour et C. Jutten. What should we say about the kurtosis? *IEEE Signal Processing Letters*, 6(12) :321–322, December 1999.
- [Nan99] A.K. Nandi. *Blind estimation using Higher-order statistics*. Kluwer Academic publishers, 1999.
- [Ngu06] E. Nguyen. Blancheur et sur-gaussianité pour la déconvolution de données multidimensionnelles. Master's thesis, Institut National Polytechnique de Grenoble, Grenoble, Juin 2006.
- [NP93] C.L. Nikias et A.P. Petropulu. *Higher-order spectra analysis, Signal processing framework*. Prentice-Hall, Englewood Cliffs edition, 1993.
- [OU79] M. Ooe et T.J. Ulrych. Minimum entropy deconvolution with an exponential transformation. *Geophysical Prospecting*, 27 :458–473, 1979.
- [Pap81] A. Papoulis. Maximum entropy and spectral estimation : a review. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 29(6) :1176–1186, December 1981.
- [PGJ92] D-T. Pham, P. Garat et C. Jutten. Separation of a mixture of independent sources through a maximum likelihood approach. In *EUSIPCO*, volume 2, pages 771–774, Brussels, September 1992.
- [Pha01] D-T. Pham. Contrast functions for blind separation and deconvolution sources. In *Proc. of the Int. Conf. ICA'01*, pages 37–42, San-Diego, December 2001.
- [Pha02] D-T. Pham. Mutual information approach to blind separation of stationary sources. *IEEE Trans. on Information Theory*, 48(7) :pp.1935–1946, July 2002.

- [Pha03] D-T. Pham. Fast algorithm for estimating mutual information, entropies and scores functions. In *Proc. of the Int. Conf. ICA '03*, pages 17–22, Nara Japan, April 2003.
- [Pha04] D-T. Pham. Fast algorithm for mutual information based independent component analysis. *IEEE Trans. on Signal Processing*, 52(10) :2690–2700, October 2004.
- [Pha05] D-T. Pham. Entropy of a variable slightly contaminated with another. *IEEE Signal Processing Letters*, 12(7) :536–539, July 2005.
- [Pic94] B. Picinbono. *Signaux aléatoires*. Dunod edition, 1994.
- [PM05] C. Paulus et J. Mars. Wideband spectral matrix filtering for multicomponent sensors array. *Signal Processing*, 85(9) :1723–1743, September 2005.
- [Pri81] M.B. Priestley. *Spectral Analysis and times series*. Academic Press, London, 1981.
- [RB99] O. Rosec et J-M. Boucher. Bayesian estimation of non-minimum phase wavelets applied to marine reflection seismic data. In *International Conference on Acoustics Speech and Signal Processing*, volume 5, pages 2797–2800, Phoenix, March 1999.
- [Ris78] J. Rissanen. Modelling by the shortest data description. *Automatica*, 14 :465–471, 1978.
- [Rob85] E.A. Robinson. Seismic time invariant convolutional model. *Geophysics*, 50(12) :2742–2751, December 1985.
- [Ros62] M. Rosenblatt. *Random Processes*. Oxford press, 1962.
- [Ros95] D. Rossille. *Reconstruction de signaux à partir du bispectre, application à l'astronomie : effet de l'échantillonnage et de la stationnarité sur les spectres d'ordre supérieur*. PhD thesis, Université de Nice, Juin 1995.
- [RT80] E.A. Robinson et S. Treitel. *Geophysical Signal Analysis*. Prentice Hall, 1980.
- [RT01] E.A. Robinson et S. Treitel. *Geophysical signal analysis*. Society of Exploration Geophysicists, 2001.
- [RVN65] M. Rosenblatt et J.W. Van Ness. Estimation of the bispectrum. In *Ann. Math. Stat.*, volume 36, pages 139–148, 1965.
- [Sap78] G. Saporta. *Théories et méthodes de la statistique*. Paris, Technip edition, 1978.
- [Sch78] G. Schwarz. Estimating the dimension of a model. *Annals of statistics*, 6 :461–464, 1978.
- [SiC00] J. Solé i Casals. *Contribució a la separació de fonts i a les seves aplicacions*. PhD thesis, Universitat politècnica de Catalunya, June 2000.
- [SiCTJ00] J. Solé i Casals, A. Taleb et C. Jutten. Parametric approach to blind deconvolution of nonlinear channels. In *ESANN*, pages 21–26, Bruges Belgium, April 2000.
- [Sil82] B.W. Silverman. *Density estimation for statistics and data analysis*. Chapman and Hall, London, 1982.

- [SVC94] M.D. Sacchi, R.D. Velis et A.H. Cominguez. Minimum entropy deconvolution with frequency-domain constraints. *Geophysics*, 59(6) :938–945, June 1994.
- [Tal99] A. Taleb. *Séparation de sources dans les mélanges non-linéaires*. PhD thesis, Institut National Polytechnique de Grenoble, Grenoble, Septembre 1999.
- [TJ99] A. Taleb et C. Jutten. Source separation in post nonlinear mixtures. *IEEE Trans. on Signal Processing*, 47(10) :2807–2820, October 1999.
- [TSiCJ01] A. Taleb, J. Solé i Casals et C. Jutten. Quasi-nonparametric blind inversion of Wiener systems. *IEEE Trans. on Signal Processing*, 49(5) :917–924, May 2001.
- [Tug93] J. K. Tugnait. Estimation of linear parametric models using inverse filter criteria and higher order statistics. *IEEE Trans. on Signal Processing*, 41(11) :3196–3199, November 1993.
- [VDB94] A. Van Den Bos. Complex gradient operator and hessian. *IEE Proceedings on Vision, Image and Signal Processing*, 141 :380–382, 1994.
- [Vra03] V. Vrabie. *Statistiques d'ordre supérieur : Applications en géophysique et électrotechnique*. PhD thesis, INPG, Grenoble, Octobre 2003.
- [Wal85] A.T. Walden. Non-Gaussian reflectivity, entropy, and deconvolution. *Geophysics*, 50(12) :2862–2888, December 1985.
- [WB98] H-S. Wu et J. Barba. Minimum entropy restoration of star field images. *IEEE Trans. on Systems Man, and Cybernetics, Part B : Cybernetics*, 28(2) :227–231, April 1998.
- [WH86] A.T. Walden et J.W.J. Hosken. The nature of the non-gaussianity of primary reflection coefficients and its significance for deconvolution. *Geophysical Prospecting*, 34 :1038–1066, 1986.
- [Whi54] P. Whittle. *Some recent contributions to the theory of stationary processes*. Almqvist and Wiksell, Stockholm, Sweden, H. Wold edition, 1954.
- [Whi88] R.E. White. Maximum kurtosis phase correction. *Geophysics*, 95 :371–389, 1988.
- [Wig78] R.A. Wiggins. Minimum entropy deconvolution. *Geoexploration*, 16 :21–35, 1978.
- [Wig85] R.A. Wiggins. Entropy-guided deconvolution. *Geophysics*, 50 :2720–2726, 1985.
- [WU83] C. Walker et T.J. Ulrych. Autoregressive recovery of the acoustic impedance. *Geophysics*, 48(10) :1338–1350, October 1983.
- [Yil87] O. Yilmaz. *Seismic data processing*, volume 2 of *Investigations in Geophysics*. Society of Exploration Geophysicists, 1987.
- [Yil00a] O. Yilmaz. *Seismic data analysis*, volume I. Society of Exploration Geophysicists, 2000.
- [Yil00b] O. Yilmaz. *Seismic data analysis*, volume II. Society of Exploration Geophysicists, 2000.

- [ZC04] L. Zhang et A. Cichocki. Multichannel blind deconvolution of nonminimum-phase systems using filter decomposition. *IEEE Trans. on Signal Processing*, 52(5) :1430–1441, May 2004.
- [ZCA04] L. Zhang, A. Cichocki et S. Amari. Self-adaptive blind source separation based on activation function adaptation. *IEEE Trans. on Neural Networks*, 15(2) :233–244, 2004.
- [Zol76] V.M. Zolotarev. Approximation of distribution of sums of independent random variables with values in infinite dimensional spaces. *Theory Prob. Applied*, 21 :721–737, 1976.

Résumé

Nous nous intéressons à la déconvolution aveugle de signaux bruités et plus précisément de signaux d'imagerie sismique. Pour réaliser l'inversion par une approche linéaire, nous souhaitons effectuer une sélection des statistiques d'ordre supérieur adaptées à la distribution du signal à déconvoluer. Pour cela, nous nous appuyons sur l'hypothèse de blancheur ou de non-gaussianité.

Nous proposons une approche avec le taux d'information mutuelle comme mesure de blancheur et une autre basée sur la non-gaussianité du signal de sortie mesurée par la négentropie. Après le développement d'algorithmes dans le domaine temporel et fréquentiel, nous caractérisons l'influence sur les critères du bruit additif présent sur les données.

Nous démontrons que l'hypothèse de non-gaussianité est plus robuste à la présence d'un bruit additif blanc et gaussien sur les données. Cette approche permet pour des données synthétiques et réelles un très bon compromis entre la qualité de la déconvolution et l'amplification du bruit.

Mots-clés : Déconvolution, statistiques d'ordre supérieur, blancheur, non gaussianité, négentropie, taux d'information mutuelle, données bruitées, imagerie sismique

Abstract

This thesis deals with the blind deconvolution of noisy data. We consider the case of seismic data. The inversion of the model need to select higher order statistics according to the distribution of the signals. To solve that, we use the assumptions of whiteness or of nongaussianity.

We propose blind déconvolution algorithm in time domain and frequency domain. We measure whiteness by mutual information rate and nongaussianity with the negentropy. Afterwards, we study the sensitivity of the different algorithm with respect to a white Gaussian additive on the data. Theoretically and in practice on real and synthetic data, non-gaussianity appears as the method which provides the better trade off between déconvolution quality and noise amplification.

Keywords : Deconvolution, higher order statistics, whiteness, non-gaussianity, negentropy, mutual information rate, noidy data, seismic imagery

Laboratoire des Images et des Signaux
ENSIEG, Domaine Universitaire, BP 46,
38402 St-Martin-d'Hères Cedex, France