

TRAITEMENT DU SIGNAL

STATISTIQUES
D'ORDRE
SUPÉRIEUR POUR
LE TRAITEMENT
DU SIGNAL

J.-L. Lacoume
P.-O. Amblard
P. Comon

Statistiques d'Ordre Supérieur pour le Traitement du Signal

This book was originally published by Masson, Paris, 1997.

Dunod, which was the new owner of the book stopped its publication and gave the copyright back to us in September 2003.

We have decided to make the book accessible on the internet. Anyone can download it, print it, copy it without altering its content, nor hide its origin.

J.L. Lacoume, P.O. Amblard and P. Comon

Remerciements

La réalisation d'un ouvrage sur les statistiques d'ordre supérieur nécessitait l'existence d'un environnement de qualité supérieure. Cet environnement, moteur de notre action, se développe sur plusieurs cercles.

- Il s'ouvre sur la communauté internationale de recherche en traitement du signal qui nous a constamment aiguillonnés.
- La brillante école française de traitement du signal et particulièrement l'action nationale du CNRS en traitement du signal et des images (GdR TdSI puis ISIS) nous a donné une impulsion forte.
- Nos collègues plus proches, les chercheurs du CEPHAG, du LTIRF, du département signal de l'ENST Paris, de l'équipe signal du Laboratoire de physique de l'ENS Lyon, de l'ETIS de l'ENSEA, les doctorants de plusieurs laboratoires, ont déblayé avec nous ce nouveau territoire.
- Nous devons beaucoup à *Jean-Marc Brossier, Valérie Carmillet, Pierre Marchand, Eric Moisan, Philippe Ravier* et *André Silvent* qui ont eu le courage, la patience et la gentillesse de relire tout ou partie de l'ouvrage et de l'enrichir par leurs critiques constructives.

Soyez ici tous remerciés, et ensemble, poursuivons la passionnante aventure de la connaissance et de l'action.

à Albert Cymermann, parti trop vite

Table des matières

Introduction	1
1 Introduction aux Statistiques d'Ordre Supérieur	5
1.1 Variables aléatoires réelles scalaires	5
1.1.1 Moments et cumulants	6
1.1.2 Exemples de variables aléatoires	8
1.2 Variables aléatoires multidimensionnelles	11
1.3 Variables aléatoires à valeurs complexes	14
1.3.1 Définitions	15
1.3.2 Standardisation	16
1.3.3 Circularité	16
1.4 Propriétés des moments et des cumulants	18
1.4.1 Multilinéarité	18
1.4.2 Invariance par translation	19
1.4.3 Variables aléatoires indépendantes	19
1.5 Statistiques d'ordre supérieur et densités de probabilité	20
1.5.1 Tendance vers la gaussianité	20
1.5.2 Développement d'une ddp autour d'une loi gaussienne	21
1.5.3 Critères de gaussianité et d'indépendance	23
1.6 Estimation des moments et cumulants	25
1.6.1 Estimation des moments	26
1.6.2 Estimation des cumulants	26
1.6.3 Estimation de l'asymétrie et du kurtosis	28
1.6.4 Estimation récursive	29
1.7 Conclusion	30
I Signaux monodimensionnels	31
2 Multicorrélations et multispectres	35
2.1 Multicorrélations	35
2.1.1 Définitions	36
2.1.2 Symétries	37

2.1.3	Signaux à mémoire limitée	38
2.1.4	Signaux à valeurs complexes	39
2.2	Multispectres	41
2.2.1	La représentation de Cramér	41
2.2.2	Définition des multispectres	43
2.2.3	Multispectres et représentation de Cramér	43
2.2.4	Moments ou cumulants?	44
2.2.5	Multispectres symétriques	45
2.2.6	Circularité des composantes spectrales	46
2.2.7	Symétries	46
2.2.8	Interprétation des multispectres	47
2.2.9	Signaux à valeurs complexes	47
2.3	Filtrage linéaire	49
2.3.1	Relations de filtrage en fréquence.	49
2.3.2	Relations de filtrage en temps	51
2.4	Echantillonnage et signaux à temps discret	52
2.4.1	L'échantillonnage	52
2.4.2	Signaux à bande limitée et échantillonnage	53
2.4.3	Sur la géométrie du support des multispectres	55
2.4.4	Multispectres des signaux à temps discret	56
2.4.5	Multispectres et composantes spectrales	58
2.4.6	Filtrage linéaire à temps discret	59
2.4.7	Relations de filtrage en fréquence	59
2.4.8	Relations de filtrage en temps	60
2.5	Ergodisme, signaux déterministes	61
2.5.1	Ergodisme	61
2.5.2	Définition étendue des multicorrélations et des multispectres	65
2.6	Estimation des multicorrélations et des multispectres	67
2.6.1	L'ergodisme revisité	67
2.6.2	Moments ou cumulants?	68
2.6.3	Estimation des multicorrélations	69
2.6.4	Estimation des multispectres	70
2.7	Illustration	78
2.8	Conclusion	84
2.9	Compléments	85
2.9.1	Note sur la <i>TFD</i>	85
2.9.2	Sur les classes de signaux	87
2.9.3	Le multipériodogramme construit sur les cumulants	88
2.9.4	Comportement asymptotique	89
2.9.5	Variance du bipériodogramme	89
2.9.6	Variance de la bicorrélation	91

3	Signaux et systèmes	93
3.1	Les systèmes échantillonnés	93
3.1.1	Présentation générale	93
3.1.2	Stabilité et causalité	95
3.2	Modélisation et filtrage	98
3.2.1	Modélisation systémique	98
3.2.2	Filtre blanchisseur. Les différents bruits blancs	99
3.2.3	Filtre inverse	100
3.2.4	Relation entre le filtre blanchisseur et le filtre inverse pour des signaux linéaires. Identification aveugle	100
3.2.5	Illustration	101
3.2.6	Conclusion	103
3.3	Identification aveugle (extralucide?)	104
3.4	Identification aveugle “en boucle”	104
3.4.1	L’algorithme du gradient stochastique	105
3.4.2	Sur les critères	108
3.4.3	Identification aveugle par le kurtosis	111
3.4.4	Les algorithmes de Godard et Sato	113
3.4.5	Illustration	113
3.5	Identification aveugle en fréquence	115
3.5.1	Principe	115
3.5.2	Mise en œuvre	116
3.5.3	Domaines d’application	120
3.6	Identification aveugle paramétrique	121
3.6.1	Introduction	121
3.6.2	Identification d’un filtre <i>MA</i>	124
3.6.3	Identification d’un filtre <i>AR</i>	130
3.6.4	Identification d’un filtre <i>ARMA</i>	131
3.6.5	Conclusion	133
3.7	Modélisation et estimation	133
3.7.1	Modélisation systémique	133
3.7.2	Signaux harmoniques	134
3.8	Conclusion	136
4	Signaux non stationnaires	137
4.1	Signaux cyclostationnaires	138
4.1.1	Cyclostationnarité d’ordre 2	138
4.1.2	Définitions	139
4.1.3	Influence de la cyclostationnarité sur les SOS	142
4.1.4	Éléments d’estimation	145
4.1.5	Application en Communication	148
4.2	Rappel sur le temps-fréquence	152
4.2.1	Représentations temps-fréquence	152

4.2.2	Classe de Cohen	154
4.2.3	Non-linéarité et distributions bilinéaires	155
4.3	Représentations temps-multifréquence	156
4.3.1	Cas des signaux déterministes	156
4.3.2	Extension aux signaux aléatoires	161
4.3.3	Quelques autres approches	163
5	Filtres de Volterra	165
5.1	Filtres de Volterra à temps continu	166
5.1.1	Définition en temps et interprétations	166
5.1.2	Critères physiques	167
5.1.3	Point de vue fréquentiel	169
5.1.4	Cas des signaux aléatoires : statistiques entrée-sortie	171
5.1.5	Modéliser et détecter une non-linéarité	173
5.2	Filtres de Volterra discrets	176
5.2.1	Définitions	176
5.2.2	Mise sous forme vectorielle	178
5.3	Identification en moyenne quadratique	182
5.3.1	Identification dans le domaine temporel	182
5.3.2	Identification dans le domaine fréquentiel	184
5.3.3	Identification orthogonale dans le domaine temporel	185
5.3.4	Implantation récursive	186
5.3.5	Identification récursive dans le domaine fréquentiel	189
5.3.6	Complexité numérique des algorithmes RLS	190
5.4	Application : soustraction de bruit	191
5.4.1	Filtres de Volterra et Soustraction de Bruit	192
5.4.2	Etude d'un cas réel	194
5.5	En guise de conclusion	198
II	Signaux multidimensionnels	203
6	Traitement d'antenne	207
6.1	Modélisation	207
6.1.1	Les signaux reçus	207
6.1.2	Les sources	208
6.1.3	La propagation	210
6.2	Formalisme et outils à l'ordre 2	213
6.2.1	Sous-espaces signal et bruit	213
6.2.2	La matrice spectrale	214
6.3	Formalisme et outils aux ordres supérieurs	215

6.3.1	La quadricovariance	216
6.3.2	Sous-espaces signal et bruit	217
6.3.3	Nombre de sources séparables	217
6.3.4	Les moments hybrides non linéaires	219
6.4	Formation de voies	221
6.4.1	Formation de voies à l'ordre 2	221
6.4.2	Formation de voies à l'ordre 4	223
6.5	Le goniomètre	224
6.5.1	Le goniomètre à l'ordre 2	224
6.5.2	Le goniomètre à l'ordre 4	225
6.5.3	Utilisation des moments hybrides non linéaires	227
6.6	Illustration	228
6.7	Conclusion	230
7	Séparation de sources	231
7.1	Modélisation	231
7.1.1	Modèle	231
7.1.2	Identifiabilité	233
7.1.3	Blanchiment spatial à l'ordre 2	235
7.2	Séparation de sources à l'ordre 2	236
7.2.1	Séparation de mélanges instantanés	236
7.2.2	Séparation de mélanges spectraux	237
7.2.3	Séparation de mélanges convolutifs	237
7.3	Séparation de mélanges simples	238
7.3.1	L'analyse en composantes indépendantes	238
7.3.2	Utilisation de l'indépendance statistique	238
7.3.3	Les systèmes de séparation	246
7.3.4	Les performances	254
7.3.5	Illustrations	256
7.4	Séparation de mélanges convolutifs	262
7.4.1	Le modèle convolutif	262
7.4.2	Séparation par les multi-outils	264
7.4.3	Séparation par les cumulants	267
7.5	Domaines d'application	268
	Bibliographie	269
	Glossaire	283
	Index	285

Contents

Introduction	1
1 Introduction to Higher-Order Statistics	5
1.1 Real scalar random variables	5
1.2 Multidimensional random variables	11
1.3 Complex valued random variables	14
1.4 Properties of moments and cumulants	18
1.5 Higher-Order Statistics and probability density functions	20
1.6 Estimation of moments and cumulants	25
1.7 Conclusion	30
I Monodimensional signals	31
2 Multicorrelations and multispectra	35
2.1 Multicorrelations	35
2.2 Multispectra	41
2.3 Linear filtering	49
2.4 Sampling and discrete time signals	52
2.5 Ergodism, deterministic signals	61
2.6 Estimation of multicorrelations and multispectra	67
2.7 Illustration	78
2.8 Conclusion	84
2.9 Complements	85
3 Signals and systems	93
3.1 Sampled systems	93
3.2 Modelisation and filtering	98
3.3 Blind identification	104
3.4 Recursive blind identification	104
3.5 Blind identification in the frequency domain	115
3.6 Parametric blind identification	121
3.7 Modelisation and estimation	133
3.8 Conclusion	136

4	Nonstationary signals	137
4.1	Cyclostationary signals	138
4.2	Elements of time-frequency analysis	152
4.3	Time-multifrequency representations	156
5	Volterra filters	165
5.1	Continuous time Volterra filters	166
5.2	Discrete Volterra filters	176
5.3	Mean square identification	182
5.4	Application : noise cancellation	191
5.5	Conclusion	198
II	Multidimensional signals	203
6	Array processing	207
6.1	Modelisation	207
6.2	Formalism and second-order tools	213
6.3	Formalism and higher-order tools	215
6.4	Beam forming	221
6.5	MUSIC	224
6.6	Illustration	228
6.7	Conclusion	230
7	Sources separation	231
7.1	Modelisation	231
7.2	Second-order sources separation	236
7.3	Separation of simple mixtures	238
7.4	Separation of convolutive mixtures	262
7.5	Domains of application	268
	Bibliography	269
	Glossary	283
	Index	285

Introduction

Le traitement du signal est associé à l'émergence d'une nouvelle "matière première", élément essentiel de l'environnement de l'homme du XX^e siècle : l'information. Cette nouvelle discipline se situe au point de convergence des mathématiques, qui lui fournissent un cadre conceptuel, de la physique, qui décrit les vecteurs de transmission de l'information, de l'électronique et de l'informatique, qui concourent aux développements des techniques matérielles et logicielles.

Le traitement du signal a fondé sa problématique sur la théorie des probabilités, sur l'analyse fonctionnelle et sur l'analyse numérique. Ces disciplines sont le cadre naturel de la modélisation en traitement du signal.

La théorie des probabilités a été particulièrement mise à contribution pour construire les modèles utilisés en traitement du signal. Cette branche très riche des mathématiques a été très développée depuis son origine dans la "théorie des jeux" pratiquée dès le XVII^e siècle par les précurseurs que furent Bernouilli, Fermat, Pascal et bien d'autres... Les concepts de base de la théorie des probabilités sont les notions de variable aléatoire et de fonction (signal) aléatoire. Ces grandeurs sont décrites complètement à travers des lois (ou densités) de probabilité. On se limite bien souvent à une description donnée par les moments.

La plupart des résultats méthodologiques et des techniques de traitement du signal sont fondés sur une description simplifiée des signaux faisant appel aux moments d'ordre 1 et 2. Les fonctions de corrélation et les densités spectrales de puissance (spectres) en sont les outils de base. Ces grandeurs jouent un rôle essentiel dans la description de signaux dans tous les domaines d'application et dans la définition d'opérateurs fondamentaux : filtre adapté, filtre de Wiener, filtre de Kalman... Le statut particulier des signaux gaussiens, qui sont très répandus et pour lesquels la description à l'ordre 2 est complète, a largement contribué à la diffusion des techniques de traitement limitées à l'ordre 2.

Le développement des moyens de traitement a permis dans un passé récent de prendre en compte une description plus subtile des signaux en utilisant des statistiques d'ordre supérieur à 2. Le cadre conceptuel est déjà très riche. Les probabilistes ont depuis longtemps développé une approche

générale des variables et des fonctions aléatoires. On peut trouver le fondement de ces modélisations dans [13, 29, 36, 37, 38, 60, 187]. Le fort courant de recherche sur ces nouvelles méthodologies qui se développe depuis une quinzaine d'années a suscité une série de colloques internationaux consacrés aux statistiques d'ordre supérieur [1, 2, 3, 4, 5], et de nombreuses revues scientifiques leur ont consacré des numéros spéciaux [6, 7, 8, 9, 10, 11, 12]. Ces études débouchent sur de nouveaux algorithmes de traitement dans de nombreux domaines d'applications. Citons l'astronomie où sont apparues les premières techniques d'ordre supérieur, la prospection sismique et les communications qui ont été un puissant moteur de développement des techniques de déconvolution aveugle, et plus généralement la géophysique, le contrôle non-destructif, les vibrations, la parole, le radar, le sonar. . .

Avant d'aborder l'étude des nouvelles descriptions de la réalité expérimentale utilisant les statistiques d'ordre supérieur, il est bon de s'interroger sur l'opportunité de ces développements.

La question est importante et la réponse ne peut être que conjecturale. A l'affirmation indiscutable du caractère plus général de cette approche, on voit objecter à juste titre la complexité accrue, et plus fondamentalement, l'applicabilité douteuse. Le mieux peut être l'ennemi du bien. Nous proposerons de relever ce défi en restant avertis des risques qu'il contient, mais qui sont somme toute associés à toute innovation. . .

Parmi les raisons invoquées à l'appui de l'utilisation des SOS, outre leur plus grand degré de généralité, nous pensons que l'argument le plus solide est la possibilité qu'elles offrent de résoudre des problèmes non accessibles à l'ordre 2 : nous citerons la séparation de sources sans modèle *a priori*, la déconvolution aveugle ou encore le filtrage non linéaire.

Laissons aux applications et à l'avenir le soin de trancher. Et consacrons nous à l'étude des SOS qui nous introduiront dans un monde de nouvelles techniques de traitement, et nous apporteront de grandes satisfactions intellectuelles à la mesure de leurs mystères. . .

Guide de lecture

Le but de cet ouvrage est de donner un panorama des innovations introduites au niveau des concepts, des méthodes et des applications par l'utilisation des statistiques d'ordre supérieur à deux. Nous avons eu le souci constant de montrer comment ces nouvelles méthodes se situent dans la ligne des techniques classiques d'ordre deux. Cette approche ne constitue pas une remise en cause des acquis essentiels du traitement du signal, mais un approfondissement de ses capacités de description et d'interprétation de notre environnement dans toutes ses composantes.

Nous débutons par une introduction aux statistiques d'ordre supérieur suivie par les deux parties constituant le corps de l'ouvrage. Dans les parties 1 et 2 nous présentons successivement les statistiques d'ordre supérieur appliquées aux signaux à une dimension, issus de la mesure réalisée sur un seul capteur, et aux signaux à plusieurs dimensions, obtenus en utilisant un réseau de plusieurs capteurs constituant une antenne.

Le chapitre 1 est une introduction aux statistiques d'ordre supérieur qui traite essentiellement des principaux outils utilisés dans les techniques d'ordre supérieur : les moments et les cumulants. Cette présentation sommaire a pour objectif de donner au lecteur les éléments mathématiques indispensables à la suite de l'exposé, accompagnés des méthodes élémentaires d'estimation des moments et des cumulants d'ordres supérieurs. Les lecteurs qui veulent pénétrer plus avant dans ce domaine de la théorie des probabilités pourront se reporter aux nombreux ouvrages qui présentent cette matière très riche.

La première partie est consacrée aux signaux à une dimension. Elle débute par la présentation des outils de base pour la description des propriétés d'ordre supérieur : les multicorrélations et les multispectres, qui font l'objet du chapitre 2. Nous insistons sur les particularités introduites par les signaux à valeurs complexes que l'on rencontre dans plusieurs domaines d'applications. Nous donnons également les techniques d'estimation des multicorrélations et des multispectres, et nous caractérisons ces mesures par leurs propriétés d'ordre 1 et 2 : moyenne et variance. Dans le chapitre 3, nous développons les méthodologies introduites par les statistiques d'ordre supérieur dans l'identification de systèmes linéaires, en insistant sur la possibilité, issue de l'utilisation des SOS, de réaliser une identification aveugle. Cette nouvelle technique, déjà mise en œuvre dans les systèmes de communication, en astronomie et en prospection sismique, peut être étendue à de nombreux autres domaines : contrôle non destructif, applications biomédicales... Le chapitre 4 donne la description des propriétés d'ordre supérieur des signaux cyclostationnaires et plus généralement non stationnaires. Ce domaine très riche s'inscrit dans la ligne des études sur les représentations temps-fréquence qui sont un des fleurons de la théorie du signal conduisant à de multiples applications. Dans ce thème nous ouvrons des pistes de recherches qui devraient conduire

à de nombreux développements. Le domaine très vaste et encore largement méconnu de l'identification de systèmes non linéaires est traité au chapitre 5, à travers l'exemple des filtres de Volterra. L'extension de la technique standard de minimisation de critères quadratiques donne une élégante solution formelle aux problèmes d'identification de filtres non linéaires. Nous montrons la faisabilité de ces méthodes qui requerront encore de nombreux efforts avant de déboucher sur des solutions pratiques bien maîtrisées.

La deuxième partie traite des signaux à plusieurs dimensions reçus sur un réseau de capteurs ou sur une antenne. Dans le chapitre 6 nous montrons, à partir des techniques classiques de traitement d'antenne, les possibilités apportées par les statistiques d'ordre supérieur. Le chapitre 7 sur la séparation de sources fait le point sur une méthode d'analyse des signaux dont le développement a été fortement impulsé par les statistiques d'ordre supérieur. Cette partie est illustrée par des simulations en séparation de sources simulées ou réelles (signaux de parole) et dans le vaste domaine de la réduction de bruit. Ces nouvelles techniques sont à l'origine de méthodologies débouchant sur de nombreux domaines d'applications : effet *cocktail-party*, téléphonie mains libres, radar, sonar, contrôle, diagnostic...

Sur la terminologie

Deux expressions très employées dans cet ouvrage nous ont posé des problèmes sémantiques et grammaticaux.

Nous avons convenu de considérer l'expression "ordre supérieur" comme un terme générique définissant l'ensemble des propriétés statistiques d'ordre supérieur à 2. Nous utilisons donc cette expression au singulier, sauf quand nous faisons référence explicitement aux éléments de cet ensemble.

La littérature concernant les statistiques d'ordre supérieur utilise indifféremment les mots *polyspectre* ou *multispectre* pour qualifier les représentations spectrales d'ordre supérieur des signaux. Nous avons choisi la dénomination *multispectre* qui allie deux termes ayant une racine latine, contrairement à la dénomination *polyspectre*, qui accouple une racine grecque à une racine latine!

CHAPITRE 1

Introduction aux Statistiques d'Ordre Supérieur

Les Statistiques d'Ordre Supérieur à 2 (SOS), autrement dit les moments et cumulants d'ordre supérieur à 2, sont utilisées essentiellement en complément aux statistiques d'ordre 2. Elles donnent une description plus complète des données et de leurs propriétés. Les SOS permettent la résolution de problèmes insolubles à l'ordre 2. Elles ont été également utilisées pour améliorer les solutions (conditionnement, identifiabilité...) déjà apportées par les techniques classiques. Les multicorrélations et les densités multispectrales que l'on définira au chapitre 2 sont issues des SOS de variables aléatoires multidimensionnelles.

Ce chapitre a pour but de donner les définitions et les propriétés nécessaires à l'introduction des SOS. Nous proposons également quelques éléments de la théorie de l'estimation des statistiques d'ordre supérieur.

Nous ne présentons dans ce chapitre que le matériel nécessaire pour la suite de l'ouvrage. Des traitements plus détaillés sont disponibles dans des ouvrages plus spécialisés que nous citons tout au long du texte.

1.1 Variables aléatoires réelles scalaires

Soit x une variable aléatoire à valeurs réelles (le cas de variables à valeurs complexes sera abordé dans la section 1.3). $F_x(u)$, fonction de répartition de x , est la probabilité d'apparition de la variable aléatoire dans le segment $] -\infty, u]$. Lorsque x admet une densité de probabilité $p_x(u)$, alors $dF_x(u) = p_x(u) du$. La densité de probabilité $p_x(u)$ est positive et a pour somme l'unité. Lorsque la fonction de répartition $F_x(u)$ est une fonction en escalier, elle n'admet pas de densité au sens des fonctions, mais au sens des distributions.

Les moments généralisés de x sont définis pour toute application réelle g

par

$$E[g(x)] = \int_{-\infty}^{+\infty} g(u) p_x(u) du.$$

On utilise souvent des fonctions polynômiales $g(u)$, conduisant aux moments "classiques" de différents ordres, tels que la moyenne ou le moment d'ordre 2. En utilisant des fonctions exponentielles, on associe aux variables aléatoires des fonctions caractéristiques.

La *première fonction caractéristique* de x est

$$\Phi_x(v) = E[e^{jvx}], \quad (1.1)$$

où j désigne la racine de -1 . Lorsque la variable aléatoire x admet une densité de probabilité $p_x(u)$, la première fonction caractéristique $\Phi_x(v)$ est sa transformée de Fourier

$$\Phi_x(v) = \int_{-\infty}^{+\infty} e^{jvu} p_x(u) du.$$

Dans ce cas, on retrouve la densité de probabilité à partir de la première fonction caractéristique par transformation de Fourier inverse

$$p_x(u) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-jvu} \Phi_x(v) dv.$$

La fonction caractéristique est continue en tout point et vaut 1 à l'origine. Elle est donc non nulle dans un voisinage de l'origine, sur lequel on pourra définir son logarithme népérien

$$\Psi_x(v) = \log(\Phi_x(v)).$$

Cette fonction est appelée *seconde fonction caractéristique*.

1.1.1 Moments et cumulants

Notons $\mu_{x(r)}$ les moments d'ordre r de x , lorsqu'ils existent,

$$\mu_{x(r)} = E[x^r],$$

et $\mu'_{x(r)}$ les moments centrés

$$\mu'_{x(r)} = E[(x - \mu_{x(1)})^r].$$

Les fonctions caractéristiques décrivent complètement la variable aléatoire à laquelle elles sont associées. En développant e^{jvx} dans l'expression (1.1) au

voisinage de l'origine et en identifiant avec le développement de Taylor de la première fonction caractéristique, on obtient les moments

$$\mu_{x(r)} = (-j)^r \left. \frac{d^r \Phi_x(v)}{dv^r} \right|_{v=0} = E[x^r].$$

Les dérivées de la seconde fonction caractéristique, prises à l'origine, définissent les *cumulants*

$$\kappa_{x(r)} = (-j)^r \left. \frac{d^r \Psi_x(v)}{dv^r} \right|_{v=0} = \text{Cum}[x, x, \dots, x],$$

qui sont les coefficients du développement en série de Taylor de la seconde fonction caractéristique.

Remarque sur les notations *Il est important de prendre garde aux notations. La notation consistant à écrire $\text{Cum}[x^3]$ au lieu de $\kappa_{x(3)}$ ou de $\text{Cum}[x, x, x]$ est dangereuse. En effet, x^3 est elle-même une variable aléatoire dont on peut calculer le cumulants d'ordre 3. L'écriture est donc nécessairement plus lourde que pour les moments.*

Les cumulants d'ordre r peuvent être calculés à partir des moments d'ordre inférieur ou égal à r [115]. Pour les ordres 1 à 4, ces relations sont

$$\begin{aligned} \kappa_{x(1)} &= \mu_{x(1)}, \\ \kappa_{x(2)} &= \mu'_{x(2)} = \mu_{x(2)} - \mu_{x(1)}^2, \\ \kappa_{x(3)} &= \mu'_{x(3)} = \mu_{x(3)} - 3\mu_{x(1)}\mu_{x(2)} + 2\mu_{x(1)}^3, \\ \kappa_{x(4)} &= \mu_{x(4)} - 4\mu_{x(3)}\mu_{x(1)} - 3\mu_{x(2)}^2 + 12\mu_{x(2)}\mu_{x(1)}^2 - 6\mu_{x(1)}^4 \end{aligned}$$

Dans le cas de variables aléatoires centrées ($\mu_{x(1)} = 0$), les expressions des cumulants se simplifient en

$$\begin{aligned} \kappa_{x(1)} &= 0, \\ \kappa_{x(2)} &= E[x^2], \\ \kappa_{x(3)} &= E[x^3], \\ \kappa_{x(4)} &= E[x^4] - 3E[x^2]^2. \end{aligned}$$

Lorsque la variable x est gaussienne, sa seconde fonction caractéristique est

$$\Psi_x(v) = j\mu_{x(1)}v - \frac{1}{2}\mu_{x(2)}v^2,$$

et ses cumulants d'ordre supérieur à 2 sont donc tous nuls. Cette propriété caractérise la loi gaussienne [113]. Les variables gaussiennes sont donc entièrement décrites par leurs propriétés au second ordre. Ceci explique pourquoi les chercheurs en traitement du signal se sont longtemps limités au second ordre. En “invokant” le théorème de la limite centrale, on peut penser que la plupart des signaux ont tendance à être gaussiens. Ce point de vue est limitatif, nous aurons l'occasion d'y revenir.

La variance, cumulant d'ordre 2 de x , $\text{Var}[x] = \kappa_{x(2)}$, caractérise la *puissance* de x . En centrant la variable aléatoire x et en la normalisant on obtient la variable standardisée

$$\tilde{x} = \frac{x - \mu_{x(1)}}{\sqrt{\kappa_{x(2)}}}.$$

On en déduit les cumulants standardisés qui sont les cumulants de la variable aléatoire standardisée. Le cumulant standardisé d'ordre 3, $\mathcal{K}_{x(3)} = \kappa_{\tilde{x}(3)}$, est dénommé *l'asymétrie*¹. L'asymétrie est nulle si la densité de probabilité de x possède un axe de symétrie. Le cumulant standardisé d'ordre 4 est le facteur d'aplatissement ou *kurtosis*². Le signe du kurtosis caractérise le comportement à l'infini des densités de probabilité unimodales. Les densités de probabilité unimodales possédant un kurtosis négatif sont dites sous-gaussiennes car elles tendent vers 0 à l'infini plus rapidement que la loi gaussienne. Les densités de probabilité unimodales à kurtosis positif, dites sur-gaussiennes, tendent moins vite vers 0 à l'infini que la loi gaussienne. Ces notions de sur- et sous-gaussianité perdent leur sens pour les densités de probabilité multimodales comme l'a montré [138].

Pour une variable aléatoire x centrée, l'asymétrie et le kurtosis s'écrivent

$$\begin{aligned} \mathcal{K}_{x(3)} &= \frac{E[x^3]}{E[x^2]^{3/2}}, \\ \mathcal{K}_{x(4)} &= \frac{E[x^4]}{E[x^2]^2} - 3. \end{aligned}$$

1.1.2 Exemples de variables aléatoires

Examinons quelques exemples de variables aléatoires.

Variable uniforme La variable aléatoire uniformément répartie dans le segment $[-a, a]$ a pour fonction caractéristique $\Phi(v) = \frac{\sin av}{av}$, pour moments d'ordre pair $\mu_{x(r)} = \frac{a^r}{r+1}$, et pour kurtosis $\mathcal{K}_{x(4)} = -\frac{6}{5}$.

¹Skewness en anglais.

² $\kappa\upsilon\rho\tau\omega\sigma\iota\varsigma$: action de courber, convexité.

Variable gaussienne Pour une variable gaussienne centrée, de loi $p_x(u) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{u^2}{2\sigma^2})$, la fonction caractéristique s'écrit $\phi(v) = \exp(-\sigma^2 v^2/2)$. Les moments d'ordre impair sont nuls et les moments d'ordre pair valent

$$\mu_{x(2r)} = \sigma^{2r} \frac{(2r)!}{r! 2^r}.$$

La seconde fonction caractéristique d'une variable gaussienne est un polynôme d'ordre 2, ce qui implique la nullité de ses cumulants d'ordre supérieur ou égal à 3.

Variables exponentielles Une variable aléatoire positive, distribuée selon la loi exponentielle

$$\begin{aligned} p_x(u) &= \frac{1}{\sigma} \exp(-\frac{u}{\sigma}) \text{ si } u \geq 0 \\ p_x(u) &= 0 \text{ si } u < 0 \end{aligned}$$

a pour fonction caractéristique $\phi_x(v) = \frac{1}{1-j\sigma v}$. Les moments vérifient la relation de récurrence $\mu_{x(r)} = r\sigma\mu_{x(r-1)}$ pour $r \geq 2$. La moyenne est $\mu_{x(1)} = \sigma$. Les premiers moments centrés sont : $\mu'_{x(2)} = \sigma^2$, $\mu'_{x(3)} = 2\sigma^3$, $\mu'_{x(4)} = 9\sigma^4$. Ses cumulants sont $\kappa_{x(r)} = (r-1)!\sigma^r$. L'asymétrie et le kurtosis valent donc $\kappa_{x(3)} = 2$ et $\mathcal{K}_{x(4)} = 6$.

Une variable aléatoire distribuée selon une loi de Laplace bilatère $p_x(u) = \frac{1}{2a} \exp(-\frac{|u|}{a})$, a pour fonction caractéristique $\phi(v) = \frac{1}{1+a^2 v^2}$, et admet pour coefficient d'aplatissement (kurtosis) $\mathcal{K}_{x(4)} = 3$.

Variable de Gauss généralisée Une variable aléatoire x est dite gaussienne généralisée si sa densité de probabilité s'écrit

$$p_x(u) = B \exp(-A|u|^g),$$

où g est un nombre réel positif. Les coefficients A et B sont introduits afin de normaliser la somme de $p_x(\cdot)$ à 1, et la variance à 1. Ces coefficients sont

$$A = \frac{\Gamma(3/g)}{\Gamma(1/g)}, \quad B = g \frac{\Gamma(3/g)^{1/2}}{\Gamma(1/g)^{3/2}}.$$

où $\Gamma(\cdot)$ est la fonction gamma, $\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt$.

Les moments d'ordre impair sont nuls et les moments d'ordre pair sont, $\forall g \geq 1$,

$$\mu_{x(r)} = \frac{\Gamma(1/g)^{r-1}}{\Gamma(3/g)^r} \Gamma\left(\frac{2r+1}{g}\right).$$

On retrouve bien $\mu_{x(2)} = 1$. Si $g = 2$, on est dans le cas gaussien, et on retrouve les résultats donnés plus haut en utilisant les relations $\Gamma(1/2) = \sqrt{\pi}$ et $\Gamma(n+1) = n\Gamma(n)$. En faisant tendre g vers l'infini, on retrouve le cas uniforme évoqué plus haut. Si $g = 1$ on obtient une variable de Laplace bilatère.

Variable du Chi-deux Une variable aléatoire χ^2 suit une loi du Chi-deux à ν degrés de liberté si elle est la somme des carrés de ν variables gaussiennes indépendantes u_i , chacune de variance unité. Le paramètre de non-centralité de χ^2 est la somme des carrés des moyennes des u_i .

Le cumulants d'ordre r d'une variable aléatoire x suivant une loi du Chi-deux à ν degrés de liberté et de paramètre de non-centralité λ est

$$\kappa_{x(r)} = 2^{r-1}(r-1)!(\nu + r\lambda).$$

En particulier dans le cas central, $\kappa_{x(4)} = 48\nu$, et le kurtosis vaut $12/\nu$. Pour $\nu = 2$ et $\lambda = 0$, on retrouve les cumulants de la variable exponentielle.

Signal sinusoïdal Supposons que l'on reçoive un signal sinusoïdal pur de puissance 1, et que l'on considère ses valeurs comme des réalisations successives d'une variable aléatoire (ce qui est bien évidemment incorrect, mais possible). Alors cette variable aurait pour densité de probabilité

$$p_x(u) = \frac{1}{\pi\sqrt{2}} \frac{1}{\sqrt{1+u^2/2}}.$$

Nous aurons en outre $\mu_{x(2)} = 1$ (qui correspond à la puissance), $\mu_{x(2r+1)} = 0$, et $\mu_{x(4)} = \frac{3}{2}$. Le kurtosis vaut donc $-\frac{3}{2}$.

Variable de Poisson Une variable aléatoire de Poisson de paramètre λ est un variable discrète prenant les valeurs $k = 0, 1, \dots$ avec la probabilité

$$p_k = \frac{\lambda^k e^{-\lambda}}{k!}.$$

Une variable de Poisson a la particularité d'avoir tous ses cumulants égaux. En effet, la première fonction caractéristique est

$$\begin{aligned} E[e^{jkv}] &= \sum_{k=0}^{\infty} e^{jkv} \frac{\lambda^k e^{-\lambda}}{k!} \\ &= e^{-\lambda} \exp\{\lambda e^{jv}\}. \end{aligned}$$

Par suite, la seconde fonction caractéristique vaut $-\lambda + \lambda e^{jv}$. La définition des cumulants montre alors qu'ils sont tous égaux à λ . Ce fait est bien connu aux ordres 1 et 2 puisque la moyenne et la variance sont égales.

Variable binaire Considérons une variable aléatoire prenant deux valeurs z_1 et z_2 avec les probabilités P_1 et P_2 respectivement. Si cette variable est standardisée, alors

$$z_1 = \frac{1}{a}, \quad z_2 = -a, \quad P_1 = \frac{a^2}{1+a^2}, \quad P_2 = \frac{1}{1+a^2},$$

où a est un réel positif fixé. Le moment d'ordre r vaut

$$\mu_{x^{(r)}} = \frac{1}{1+a^2} \left[\frac{1}{a^{r-2}} + (-1)^r a^r \right].$$

On vérifie notamment que $\mu_{x^{(1)}} = 0$ et $\mu_{x^{(2)}} = 1, \forall a \in \mathbb{R}^+$. Le kurtosis d'une variable binaire est donc de la forme

$$\kappa_{x^{(4)}} = \frac{a}{1+a^2} \left[\frac{1}{a^3} + a^3 \right] - 3.$$

1.2 Variables aléatoires multidimensionnelles

On supposera dorénavant que les variables aléatoires sont centrées, sauf mention explicite.

On peut représenter les variables aléatoires à plusieurs dimensions par le vecteur colonne

$$\mathbf{x}^T = (x_1, x_2, \dots, x_N).$$

De la même façon que dans le cas scalaire, on définit la fonction caractéristique conjointe de N variables aléatoires x_n par la relation

$$\Phi_{\mathbf{x}}(\mathbf{v}) \stackrel{\text{def}}{=} E \left[e^{j \sum_n v_n x_n} \right] = E \left[e^{j \mathbf{v}^T \mathbf{x}} \right].$$

où $\mathbf{v}^T = (v_1, v_2, \dots, v_N)$. Si les composantes x_n du vecteur aléatoire \mathbf{x} admettent une densité conjointe $p_{\mathbf{x}}(\mathbf{u})$, alors la fonction caractéristique de \mathbf{x} est donnée par la transformée de Fourier de cette densité

$$\Phi_{\mathbf{x}}(\mathbf{v}) = \int_{\mathbb{R}^N} e^{j \mathbf{v}^T \mathbf{u}} p_{\mathbf{x}}(\mathbf{u}) \, d\mathbf{u}.$$

Comme précédemment, la seconde fonction caractéristique est $\Psi_{\mathbf{x}}(\mathbf{v}) = \log \Phi_{\mathbf{x}}(\mathbf{v})$. Les fonctions caractéristiques peuvent servir à générer les moments et les cumulants.

Les cumulants d'ordre 2 sont des grandeurs à deux indices, qui peuvent être rangées dans une matrice (la matrice de *covariance*)

$$\kappa_{\mathbf{x}^{(2)}}_{ij} = \text{Cum} [x_i, x_j]$$

Lorsqu'on manipule des données à plusieurs dimensions, il peut devenir inutile de préciser l'ordre du moment ou du cumulants considéré. Ainsi, la matrice de covariance s'écrit simplement $\kappa_{\mathbf{x}ij} = Cum[x_i, x_j]$. Nous disposons donc de deux notations, par exemple $\kappa_{\mathbf{x}(2)} = \kappa_{\mathbf{x}ij}$, qui seront utilisées selon le contexte de travail. Ainsi, lorsque l'on manipule des vecteurs aléatoires, l'ordre de la statistique utilisée correspond au nombre des indices. Par exemple,

$$\begin{aligned}\kappa_{\mathbf{x}ijk} &= Cum[x_i, x_j, x_k] = \kappa_{\mathbf{x}(3)}, \\ \kappa_{\mathbf{x}iii} &= Cum[x_i, x_i, x_i] = \kappa_{x_i(3)}\end{aligned}$$

sont des cumulants d'ordre 3, et

$$\begin{aligned}\kappa_{\mathbf{x}hijk} &= Cum[x_h, x_i, x_j, x_k] = \kappa_{\mathbf{x}(4)}, \\ \kappa_{\mathbf{x}iiii} &= Cum[x_i, x_i, x_i, x_i] = \kappa_{x_i(4)}.\end{aligned}$$

sont d'ordre 4.

En développant l'exponentielle $e^{\mathbf{v}^T \mathbf{x}}$ en série autour de $\mathbf{v} = \mathbf{0}$, les coefficients des termes $v_i v_j \dots v_k$ de degré r , sont $j^r \mu_{\mathbf{x}ij\dots k}/r!$ [115, 148], donnant pour les moments

$$\mu_{\mathbf{x}i_1 i_2 \dots i_r} = (-j)^r \left. \frac{\partial^r \Phi_x(\mathbf{v})}{\partial v_{i_1} \partial v_{i_2} \dots \partial v_{i_r}} \right|_{\mathbf{v}=\mathbf{0}} = E[x_{i_1} \dots x_{i_r}]. \quad (1.2)$$

Pour obtenir les cumulants, on procède de même en remplaçant la première fonction caractéristique par la seconde. Les cumulants sont donc définis par

$$\kappa_{\mathbf{x}i_1 i_2 \dots i_r} = (-j)^r \left. \frac{\partial^r \Psi_x(\mathbf{v})}{\partial v_{i_1} \partial v_{i_2} \dots \partial v_{i_r}} \right|_{\mathbf{v}=\mathbf{0}} = Cum[x_{i_1}, \dots, x_{i_r}]. \quad (1.3)$$

Comme dans le cas scalaire, il est possible d'établir des relations entre les moments et les cumulants en développant la fonction logarithme en série entière. On obtient pour les cumulants d'ordre 2

$$\kappa_{\mathbf{x}ij} = \mu_{\mathbf{x}ij} - \mu_{\mathbf{x}i} \mu_{\mathbf{x}j}.$$

Pour écrire l'ensemble de ces relations de façon plus compacte, il est utile d'introduire des conventions d'écriture. On conviendra d'écrire une somme de k termes se déduisant les uns des autres par permutation d'indices par une *notation crochet* [148].

Quelques bons exemples valent mieux qu'un long discours :

$$\begin{aligned}[3] \delta_{ij} \delta_{kl} &= \delta_{ij} \delta_{kl} + \delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk}, \\ [3] a_{ij} b_k c_{ijk} &= a_{ij} b_k c_{ijk} + a_{ik} b_j c_{ijk} + a_{jk} b_i c_{ijk}.\end{aligned}$$

La présence du crochet entraîne une sommation implicite. Les termes à r indices sont ici des tenseurs d'ordre r complètement symétriques. Le nombre de monômes distincts que l'on peut obtenir par permutation est égal à l'entier figurant entre crochets. On écrira

$$[3]x_i\delta_{jk}, [6]x_ix_j\delta_{kl}, [10]x_ix_jx_k\delta_{lm}, [35]A_{ijk}B_{abcd}C_{ijkabcd}.$$

Dans le cas non centré, les cumulants d'ordre 3 et 4 sont alors donnés en fonction des moments par les expressions compactes

$$\kappa_{\mathbf{x}ijk} = \mu_{\mathbf{x}ijk} - [3]\mu_{\mathbf{x}i}\mu_{\mathbf{x}jk} + 2\mu_{\mathbf{x}i}\mu_{\mathbf{x}j}\mu_{\mathbf{x}k}, \quad (1.4)$$

$$\begin{aligned} \kappa_{\mathbf{x}ijkl} &= \mu_{\mathbf{x}ijkl} - [4]\mu_{\mathbf{x}i}\mu_{\mathbf{x}jkl} - [3]\mu_{\mathbf{x}ij}\mu_{\mathbf{x}kl} \\ &+ 2[6]\mu_{\mathbf{x}i}\mu_{\mathbf{x}j}\mu_{\mathbf{x}kl} - 6\mu_{\mathbf{x}i}\mu_{\mathbf{x}j}\mu_{\mathbf{x}k}\mu_{\mathbf{x}l}. \end{aligned} \quad (1.5)$$

Dans le cas centré, ces expressions se simplifient en

$$\begin{aligned} \kappa_{\mathbf{x}ij} &= \mu_{\mathbf{x}ij}, \\ \kappa_{\mathbf{x}ijk} &= \mu_{\mathbf{x}ijk}, \\ \kappa_{\mathbf{x}ijkl} &= \mu_{\mathbf{x}ijkl} - [3]\mu_{\mathbf{x}ij}\mu_{\mathbf{x}kl}. \end{aligned}$$

Il est intéressant de comparer ces expressions avec celles du cas scalaire : pour les retrouver, il suffit de remplacer $n[m]$ par nm . Dans le cas centré, aux ordres 5 et 6

$$\begin{aligned} \kappa_{\mathbf{x}ijklm} &= \mu_{\mathbf{x}ijklm} - [10]\mu_{\mathbf{x}ij}\mu_{\mathbf{x}klm}, \\ \kappa_{\mathbf{x}ijklmn} &= \mu_{\mathbf{x}ijklmn} - [15]\mu_{\mathbf{x}ij}\mu_{\mathbf{x}klmn} \\ &- [10]\mu_{\mathbf{x}ijk}\mu_{\mathbf{x}lmn} + 2[15]\mu_{\mathbf{x}ij}\mu_{\mathbf{x}kl}\mu_{\mathbf{x}mn}. \end{aligned}$$

A partir de la définition de la seconde fonction caractéristique, il est possible d'écrire les relations générales liant moments et cumulants de manière compacte [36, 188]. Les cumulants sont liés aux moments par la formule dite de Leonov et Shirayev (donnée ici à l'ordre r)

$$Cum[x_1, \dots, x_r] = \sum (-1)^{k-1} (k-1)! E \left[\prod_{i \in v_1} x_i \right] \cdot E \left[\prod_{j \in v_2} x_j \right] \cdots E \left[\prod_{k \in v_p} x_k \right],$$

où la sommation s'étend sur tous les ensembles $\{v_1, v_2, \dots, v_p : 1 \leq p \leq r\}$ formant une partition de $\{1, 2, \dots, r\}$. Dans cette formule, k est le nombre d'éléments composant la partition.

Illustrons le fonctionnement de cette formule aux premiers ordres. A l'ordre 2, les partitions possibles sont (1,2) et (1)(2). Donc $Cum[x_1, x_2] = (-1)^0 0! E[x_1 x_2] + (-1)^{2-1} (2-1)! E[x_1] E[x_2]$ donnant la formule connue de

la covariance. A l'ordre 3, les partitions possibles sont $(1, 2, 3)$, $(1)(2, 3)$, et $(1)(2)(3)$. Remarquons qu'il y a trois partitions du type $(1)(2, 3)$: ce nombre de partitions correspond au nombre apparaissant dans la notation crochet définie précédemment. Le cumulants d'ordre 3 s'écrit alors

$$\begin{aligned} Cum[x_1, x_2, x_3] &= (-1)^0 0! E[x_1 x_2 x_3] + (-1)^{2-1} (2-1)! [3] E[x_1] E[x_2 x_3] \\ &+ (-1)^{3-1} (3-1)! E[x_1] E[x_2] E[x_3] \end{aligned}$$

qui redonnent bien la formule (1.4).

A l'ordre 4, les partitions sont

- $(1, 2, 3, 4)$ au nombre de 1, $k - 1 = 0$,
- $(1)(2, 3, 4)$ au nombre de 4, $k - 1 = 1$,
- $(1)(2)(3, 4)$ au nombre de 6, $k - 1 = 2$
- $(1, 2)(3, 4)$ au nombre de 3, $k - 1 = 1$,
- $(1)(2)(3)(4)$ au nombre de 1, $k - 1 = 3$.

On retrouve alors la relation (1.5) en appliquant la formule de Leonov et Shirayev. Cette formule peut être inversée pour donner les moments en fonction des cumulants. La formule inverse est plus simple que la formule directe, puisqu'elle ne fait apparaître que des sommes de partitions, soit

$$E[x_1, \dots, x_r] = \sum Cum[x_i, i \in v_1] \cdot Cum[x_j, j \in v_2] \cdots Cum[x_k, k \in v_p],$$

où $Cum[x_k, k \in v_p]$ représente le cumulants des variables x_k , k variant dans l'ensemble v_p d'une partition. Encore une fois, la somme s'étend sur toutes les partitions possibles de $\{1, 2, \dots, r\}$. A l'ordre 4, on a alors immédiatement

$$\begin{aligned} \mu_{\mathbf{x}ijkl} &= \kappa_{\mathbf{x}ijkl} + [4]\kappa_{\mathbf{x}i} \kappa_{\mathbf{x}jkl} + [3]\kappa_{\mathbf{x}ij} \kappa_{\mathbf{x}kl} \\ &+ [6]\kappa_{\mathbf{x}i} \kappa_{\mathbf{x}j} \kappa_{\mathbf{x}kl} + \kappa_{\mathbf{x}i} \kappa_{\mathbf{x}j} \kappa_{\mathbf{x}k} \kappa_{\mathbf{x}l}. \end{aligned}$$

Les cumulants de variables multidimensionnelles apparaîtrons naturellement dans le chapitre 2 pour la définition des multicorrélations et des multispectres de signaux aléatoires. Les variables aléatoires multidimensionnelles sont à la base des modélisations en traitement d'antenne (chapitre 6) et en séparation de sources (chapitre 7).

1.3 Variables aléatoires à valeurs complexes

Des domaines aussi différents que les communications ou le traitement d'antenne utilisent des variables aléatoires à valeurs complexes. L'étude de ces variables s'est développée considérablement ces dernières années.

1.3.1 Définitions

Une variable aléatoire à valeurs complexes, comme l'a souligné Fortet [77], peut être représentée par une variable aléatoire réelle de dimension 2. Ainsi, une variable aléatoire complexe z admet une densité si et seulement si ses parties réelle et imaginaire admettent une densité conjointe.

La fonction caractéristique de la variable vectorielle complexe \mathbf{z} , avec $\mathbf{z} = \mathbf{x} + j\mathbf{y}$, \mathbf{x} à valeurs dans \mathbb{R}^N , \mathbf{y} à valeurs dans \mathbb{R}^N , est

$$\Phi_{\mathbf{z}}(\mathbf{u}) \stackrel{\text{def}}{=} E \left[e^{j[\mathbf{x}^T \mathbf{v} + \mathbf{y}^T \mathbf{w}]} \right] = E \left[e^{j \operatorname{Re} [\mathbf{z}^t \mathbf{u}]} \right], \quad (1.6)$$

avec $\mathbf{u} = \mathbf{v} + j\mathbf{w}$.

Une autre description des variables aléatoires complexes, prenant directement en compte leur caractère complexe, a été proposée par [15]. Dans cette approche on associe au vecteur aléatoire complexe \mathbf{z} son conjugué \mathbf{z}^* . La densité de probabilité est fonction de \mathbf{z} et \mathbf{z}^* . La fonction caractéristique est donnée par

$$\Phi_{\mathbf{z}, \mathbf{z}^*}(\mathbf{u}, \mathbf{u}^*) \stackrel{\text{def}}{=} E \left[e^{j \frac{\mathbf{z}^t \mathbf{u} + \mathbf{u}^t \mathbf{z}}{2}} \right],$$

qui est identique à (1.6).

Cette approche montre que la description statistique d'une variable aléatoire complexe fait intervenir des moments et des cumulants croisés entre la variable et sa conjuguée. Les notations doivent alors prendre en compte la présence de termes conjugués et de termes non conjugués. Les indices ou les ordres correspondant aux termes conjugués apparaissent en exposant, alors que les indices ou ordres des termes non conjugués sont placés en indices. Par exemple, considérons une variable z . Ses moments sont écrits

$$\mu_z \binom{q}{p} = E [z^p z^{*q}],$$

et ses cumulants

$$\kappa_z \binom{q}{p} = \operatorname{Cum} \left[\underbrace{z, \dots, z}_p, \underbrace{z^*, \dots, z^*}_q \right].$$

Soit maintenant le vecteur aléatoire complexe \mathbf{z} de composantes z_i . Ses moments sont

$$\mu_{\mathbf{z}} \binom{j_1 \dots j_q}{i_1 \dots i_p} = E [z_{i_1} \dots z_{i_p} z_{j_1}^* \dots z_{j_q}^*],$$

et ses cumulants

$$\kappa_{\mathbf{z}} \binom{j_1 \dots j_q}{i_1 \dots i_p} = \operatorname{Cum} [z_{i_1}, \dots, z_{i_p}, z_{j_1}^*, \dots, z_{j_q}^*].$$

Ces statistiques peuvent être obtenues à partir des fonctions caractéristiques, de la même manière que dans le cas réel.

1.3.2 Standardisation

Nous avons introduit dans la section 1.1.1 l'opération de *standardisation* pour les variables aléatoires scalaires. Cette opération peut aussi être définie dans le cas multivariable, pour des vecteurs à valeurs réelles ou complexes. Soit \mathbf{x} un vecteur aléatoire de moyenne $\mu_{\mathbf{x}(1)}$ et de matrice de covariance C , strictement définie positive ($C = \kappa_{\mathbf{x}ij}$ pour des valeurs réelles et $C = \kappa_{\mathbf{x}i}^j$ pour des valeurs complexes). La standardisation est la transformation affine qui associe à \mathbf{x} un vecteur aléatoire centré dont la matrice de covariance est l'identité I . Le vecteur aléatoire standardisé est

$$\tilde{\mathbf{x}} = W(\mathbf{x} - \mu_{\mathbf{x}(1)}).$$

La matrice W est solution de la relation

$$WCW^\dagger = I.$$

La matrice W , qui diagonalise la covariance de \mathbf{x} , n'est pas unique. Par convention on choisit

$$W = \Lambda^{-1/2}U^\dagger,$$

Λ étant la matrice diagonale des valeurs propres de C et U la matrice unitaire dont les colonnes sont les vecteurs propres normalisés de C .

Le vecteur $\tilde{\mathbf{x}}$ est défini à une matrice multiplicative près de la forme ΔP , où Δ est diagonale et constituée d'éléments de module 1, et P est une matrice de permutation. On range souvent les valeurs propres dans l'ordre de leurs valeurs décroissantes.

1.3.3 Circularité

Les variables aléatoires complexes sont la plupart du temps obtenues par transformation de Fourier (TF) de variables réelles, ce qui leur confère une structure très particulière. Les variables aléatoires complexes obtenues de cette façon ne sont donc pas de simples variables aléatoires à deux composantes réelles, mais des contraintes lient ces deux composantes. C'est pourquoi il est pertinent d'introduire les variables aléatoires dites *circulaires*.

Définition de la circularité Nous dirons qu'un vecteur aléatoire complexe \mathbf{z} , de dimension N , est circulaire (ou circulaire au sens fort), si et seulement si, $\forall \theta$, \mathbf{z} et $\mathbf{z}e^{j\theta}$ ont les mêmes propriétés statistiques.

Les propriétés suivantes sont des conditions nécessaires et suffisantes de circularité :

- Si \mathbf{z} est une variable aléatoire circulaire, sa fonction caractéristique vérifie la relation

$$\Phi_{\mathbf{z}}(e^{j\theta} \mathbf{u}) = \Phi_{\mathbf{z}}(\mathbf{u}), \quad \forall \theta.$$

De manière équivalente dans ce cas circulaire, la fonction caractéristique, fonction des variables complexes \mathbf{u} et \mathbf{u}^* , ne dépend que du produit tensoriel $\mathbf{u}\mathbf{u}^\dagger$, c'est-à-dire uniquement de termes de la forme $u_i u_j^*$.

- La densité de probabilité de \mathbf{z} , si elle existe, vérifie la même propriété d'invariance par multiplication par $e^{j\theta}$. La densité de probabilité définie à partir de \mathbf{z} et \mathbf{z}^* , n'est également fonction que du produit tensoriel $\mathbf{z}\mathbf{z}^\dagger$ soit de termes du type $z_i z_j^*$.

Dans le plan complexe, la multiplication par le facteur $e^{j\theta}$ correspond à une rotation d'angle θ . Ainsi la circularité d'une variable complexe est définie par l'invariance par rotation de sa densité de probabilité.

- Soit \mathbf{z} un vecteur aléatoire complexe, dont les moments existent à tous les ordres. \mathbf{z} est circulaire si et seulement si tous ses moments de la forme

$$\mu_{\mathbf{z}}^{(q)} = E \left[\prod_{\Sigma a_i = p} z_i^{a_i} \prod_{\Sigma b_j = q} z_j^{*b_j} \right]$$

sont nuls dès que $p \neq q$.

En effet, si z est circulaire, les moments de \mathbf{z} et de $\mathbf{z} e^{j\alpha}$ sont égaux, puisque ces deux variables aléatoires ont même loi. En particulier, l'égalité $\mu_{\mathbf{z}}^{(q)} = \mu_{\mathbf{z} \exp(j\alpha)}^{(q)}$ entraîne que

$$\mu_{\mathbf{z}}^{(q)} = \mu_{\mathbf{z}}^{(q)} e^{j\alpha(p-q)},$$

ce qui prouve la propriété.

Les propriétés précédentes montrent qu'une variable aléatoire scalaire complexe circulaire vérifie $E[z] = 0$, $E[z^2] = 0$, $E[z^2 z^*] = 0 \dots$ Autrement dit, pour une variable complexe circulaire, les moments ou cumulants comportant un nombre différent de termes conjugués que de termes non conjugués sont nuls. Dans la suite, nous aurons besoin de la définition restrictive suivante.

Circularité à l'ordre r On dira qu'un vecteur aléatoire complexe \mathbf{z} est circulaire à l'ordre r s'il vérifie

$$E \left[\prod_{\Sigma a_i = p} z_i^{a_i} \prod_{\Sigma b_j = q} z_j^{*b_j} \right] = 0$$

pour tout couple (p, q) d'entiers positifs tel que $p + q \leq r$ et $p \neq q$.

Une définition alternative repose encore sur l'invariance par rotation de la densité de probabilité. Nous avons vu que la variable z est circulaire au sens strict si z et $ze^{j\theta}$ ont même densité de probabilité pour tout θ . Ainsi, la densité de probabilité d'une variable circulaire au sens strict est invariante par rotation.

Alors, *une variable est circulaire à l'ordre r si sa densité de probabilité est invariante par rotation d'angle $2\pi/(r+1)$, ou de façon équivalente si z et $ze^{2\pi j/(r+1)}$ ont mêmes propriétés statistiques.*

Cas Gaussien Dans le cas gaussien, la circularité à l'ordre 2 entraîne la circularité à tous les ordres, et est caractérisée par deux propriétés liant les parties réelle et imaginaire des moments d'ordre 2. Posons $\mathbf{z} = \mathbf{x} + j\mathbf{y}$. Si \mathbf{z} est circulaire, $E[\mathbf{z}\mathbf{z}^T] = 0$ implique que $E[\mathbf{x}\mathbf{x}^T - \mathbf{y}\mathbf{y}^T] = 0$ et que $E[\mathbf{x}\mathbf{y}^T + \mathbf{y}\mathbf{x}^T] = 0$. Autrement dit, \mathbf{x} et \mathbf{y} ont la même matrice de covariance, et leur covariance croisée est antisymétrique. C'est ainsi qu'ont été définies les variables gaussiennes complexes circulaires [94, 202].

Cas des N -PSK Il est usuel en communication d'utiliser des variables aléatoires discrètes complexes, prenant comme valeurs les racines N -ième de l'unité de façon équiprobable. Les signaux en découlant sont appelés des N -PSK (Phase Shift Keying). Soit z une telle variable. Elle prend ses valeurs dans $\{\exp(2\pi jk/N), k = 0, \dots, N-1\}$, équiprobablement avec la probabilité $1/N$. Cette variable est alors circulaire à l'ordre $N-1$. En effet, z et $z \exp(2\pi j/N)$ ont même probabilité (notons que pour N pair, les moments d'ordre impair sont nuls, et la circularité est donc d'ordre $N-2$).

La propriété de circularité des N -PSK sera utilisée au chapitre 4 dans une application de reconnaissance de modulation.

Différentes définitions de la circularité ont été passées en revue, et analysées plus en profondeur dans [171].

1.4 Propriétés des moments et des cumulants

Ce paragraphe concerne l'étude des moments et des cumulants. Nous donnons les propriétés élémentaires de ces statistiques utiles pour la suite de l'ouvrage. Il existe un grand nombre d'autres propriétés que les lecteurs pourront trouver dans [148] par exemple.

1.4.1 Multilinéarité

Les moments et les cumulants satisfont la propriété de *multilinéarité*. Soient deux vecteurs aléatoires \mathbf{x} et \mathbf{y} liés par la relation linéaire $\mathbf{y} = \mathbf{A}\mathbf{x}$, où

A est une matrice quelconque. Alors les moments et cumulants de \mathbf{y} sont des fonctions linéaires de chacune des composantes A_{ij} . Par exemple, on a

$$\text{Cum}[y_i, y_j, y_k^*] = \sum_{a,b,c} A_{ia} A_{jb} A_{kc}^* \text{Cum}[x_a, x_b, x_c],$$

Pour démontrer ces relations, il suffit de remarquer que $\Phi_{A\mathbf{x}}(\mathbf{u}) = \Phi_{\mathbf{x}}(A^\dagger \mathbf{u})$ et d'utiliser (1.3, page 12).

C'est grâce à la multilinéarité que les moments et cumulants méritent la dénomination de *tenseurs*³. Cette propriété se réduit dans le cas scalaire à une simple relation d'homogénéité

$$\kappa_{\lambda x(r)} = \lambda^r \kappa_{x(r)}.$$

En particulier, quelles que soient les variables aléatoires x, y, z_1, \dots, z_p , réelles ou complexes, scalaires ou multidimensionnelles, indépendantes ou non, nous avons

$$\begin{aligned} \text{Cum}[x + y, z_1, \dots, z_p] &= \text{Cum}[x, z_1, \dots, z_p] + \text{Cum}[y, z_1, \dots, z_p], \\ \text{Cum}[\lambda x, z_1, \dots, z_p] &= \lambda \text{Cum}[x, z_1, \dots, z_p]. \end{aligned}$$

1.4.2 Invariance par translation

Les cumulants sont invariants par translation déterministe : si $\mathbf{y} = \mathbf{x} + \mathbf{t}$, où \mathbf{t} est déterministe, alors les cumulants d'ordre supérieur ou égal à 2 de \mathbf{y} sont identiques à ceux de \mathbf{x} . En effet, la translation provoque un déphasage de la première fonction caractéristique, qui implique que les secondes fonctions caractéristiques de \mathbf{y} et \mathbf{x} sont liées par $\Psi_{\mathbf{y}}(\mathbf{v}) = j\text{Re}[\mathbf{t}^\dagger \mathbf{v}] + \Psi_{\mathbf{x}}(\mathbf{v})$. Cette relation montre que la translation n'affecte que le cumulants d'ordre 1, c'est-à-dire la moyenne.

La propriété d'invariance par translation déterministe n'est évidemment pas vérifiée par les moments.

1.4.3 Variables aléatoires indépendantes

L'indépendance statistique de variables aléatoires implique la factorisation des moments (via la factorisation des premières fonctions caractéristiques).

Soient \mathbf{x} et \mathbf{y} deux vecteurs aléatoires indépendants, à valeurs réelles ou complexes, de dimensions respectives n et p , et soit $\mathbf{z}^T = (\mathbf{x}^T \mathbf{y}^T)$ (concaténation des vecteurs \mathbf{x} et \mathbf{y}). Alors, d'après la remarque précédente, la première fonction caractéristique de \mathbf{z} est le produit des premières fonctions

³La multilinéarité implique, en effet, que dans un changement de base les cumulants se transforment comme des tenseurs.

caractéristiques de \mathbf{x} et \mathbf{y} . La seconde fonction caractéristique de \mathbf{z} s'écrit alors $\Psi_{\mathbf{z}}(\mathbf{u}, \mathbf{v}) = \Psi_{\mathbf{x}}(\mathbf{u}) + \Psi_{\mathbf{y}}(\mathbf{v})$. Les cumulants croisés entre \mathbf{x} et \mathbf{y} sont donc nuls. En effet, d'une façon générale, on a

$$\begin{aligned} \kappa_{\mathbf{z}} \, i_1, \dots, i_n, j_1, \dots, j_p &= \text{Cum} [x_{i_1}, \dots, x_{i_n}, y_{j_1}, \dots, y_{j_p}] \\ &= (-j)^r \frac{\partial^r \Psi_{\mathbf{z}}(\mathbf{u}, \mathbf{v})}{\partial u_{i_1} \dots \partial u_{i_n} \partial v_{j_1} \dots \partial v_{j_p}} \Big|_{\mathbf{v}=\mathbf{0}}. \end{aligned}$$

où $r = n + p$. Comme $\Psi_{\mathbf{z}}(\mathbf{u}, \mathbf{v}) = \Psi_{\mathbf{x}}(\mathbf{u}) + \Psi_{\mathbf{y}}(\mathbf{v})$, le cumulants précédent est nul dès que l'un des i_k et l'un des j_k sont non nuls simultanément. Autrement dit, les cumulants croisés de \mathbf{x} et \mathbf{y} sont nuls.

Lorsque, l'on examine le cumulants d'un groupe de variables aléatoires, si ce groupe peut être scindé en deux sous-groupes de variables indépendantes, le cumulants est nul. En d'autres termes, si x et y sont indépendantes, alors quelle que soit z

$$\text{Cum} [x, y, z] = 0.$$

Cette propriété est à la base des techniques de séparation de sources qui seront présentées au chapitre 7.

Cette propriété fondamentale des cumulants implique la suivante, qui constitue la propriété d'additivité. Le cumulants de la somme de vecteurs aléatoires indépendants est la somme des cumulants : $\kappa_{\mathbf{x}+\mathbf{y}}(r) = \kappa_{\mathbf{x}}(r) + \kappa_{\mathbf{y}}(r)$ et ce quel que soit l'ordre r . Cette propriété, bien connue à l'ordre 2 (variance), est évidemment fautive pour les moments.

1.5 Statistiques d'ordre supérieur et densités de probabilité

Dans ce paragraphe, nous examinons les liens entre statistiques d'ordre supérieur et densités de probabilité. Commençons par le théorème de la limite centrale.

1.5.1 Tendances vers la gaussianité

Le théorème de la limite centrale a une grande importance car il permet d'approcher la loi de certains estimateurs par la loi gaussienne, et d'accéder à l'ordre de grandeur de ses cumulants successifs.

Considérons N variables aléatoires scalaires indépendantes, $x(n)$ pour $1 \leq n \leq N$, chacune de cumulants d'ordre r bornés, notés $\kappa_{(r)}(n)$. Posons

$$\bar{\kappa}_{(r)} = \frac{1}{N} \sum_{n=1}^N \kappa_{(r)}(n) \text{ et } y = \frac{1}{\sqrt{N}} \sum_{n=1}^N (x(n) - \bar{\kappa}_{(1)}).$$

La variable aléatoire y est la somme normalisée de N variables aléatoires indépendantes. Lorsque $N \rightarrow \infty$ la variable aléatoire y tend en loi⁴ vers une variable aléatoire gaussienne.

Les cumulants de y , notés $\lambda_{(r)}$, sont donnés par

$$\begin{aligned}\lambda_{(1)} &= 0, \\ \lambda_{(2)} &= \bar{\kappa}_{(2)}, \\ \lambda_{(r)} &= \frac{1}{N^{r/2-1}} \bar{\kappa}_{(r)}, \quad \forall r \geq 2.\end{aligned}$$

Ces relations nous indiquent comment les cumulants d'ordre $r > 2$ de y tendent vers 0 lorsque N tend vers l'infini. Celles-ci résultent de la propriété d'additivité des cumulants de sommes de variables aléatoires indépendantes (voir page 20). Les cumulants de la variable y s'écrivent

$$\kappa_{y(r)} = N^{-r/2} \sum_{n=1}^N \kappa_{x^{(n)}(r)},$$

ce qui prouve que $\lambda_{(r)} = N^{1-r/2} \bar{\kappa}_{(r)}$ par définition de $\bar{\kappa}_{(r)}$.

1.5.2 Développement d'une ddp autour d'une loi gaussienne

Lorsqu'une variable aléatoire est non gaussienne, il est intéressant de décrire les écarts au caractère gaussien. Cette idée intuitive trouve une justification mathématique dans le développement de la densité au voisinage de la loi gaussienne. Il existe plusieurs types de développement, notamment ceux de Gram-Charlier et de Edgeworth. Ces deux développements ne diffèrent que par la façon d'ordonner les termes. En faisant référence au théorème de la limite centrale, le développement de Edgeworth rassemble les termes qui correspondent à la même puissance (négative) de N . Pour cette raison nous avons retenu ce développement. Notons que pour des lois symétriques, et jusqu'à l'ordre 4, les deux développements sont identiques.

Développement en série d'une densité Voyons tout d'abord quelques généralités sur les développements en série de densités de probabilité. Soit une variable aléatoire scalaire x de seconde fonction caractéristique $\Psi_x(v)$, supposée être voisine d'une fonction $\Psi_o(v)$. Par définition, $\Psi_x(v)$ génère les cumulants de x dans son développement en série entière

$$\Psi_x(u) = \kappa_{1J}v + \frac{1}{2!}\kappa_{2}(Jv)^2 + \frac{1}{3!}\kappa_{3}(Jv)^3 + \dots,$$

⁴Cela signifie que la densité de probabilité de y converge uniformément vers une loi de Gauss lorsque N tend vers l'infini.

où κ_r désigne le cumulants d'ordre r $\kappa_{x(r)}$. Appelons λ_r le cumulants d'ordre r dans le développement en série de $\Psi_o(v)$, et $\eta_r = \kappa_r - \lambda_r$. Alors la différence des fonctions caractéristiques s'écrit

$$\Psi_x(v) - \Psi_o(v) = \sum_{r=1}^{\infty} \frac{1}{r!} \eta_r (jv)^r.$$

En introduisant les coefficients β_k définis par

$$\exp\left[\sum_{r=1}^{\infty} \frac{1}{r!} \eta_r (jv)^r\right] = \sum_{j=0}^{\infty} \frac{1}{k!} \beta_k (jv)^k,$$

il est possible de développer $p_x(u)$ autour de $p_o(u)$

$$p_x(u) = p_o(u) \sum_{k=0}^{\infty} \frac{1}{k!} \beta_k h_k(u), \quad (1.7)$$

où les fonctions $h_k(u)$ sont définies par

$$h_k(u) = \frac{(-1)^k}{p_o(u)} \frac{d^k p_o}{du^k}(u).$$

Le développement (1.7) ne revêt une forme simple que pour certaines densités $p_o(u)$ particulières, notamment celles pour lesquelles les fonctions $h_k(u)$ sont des polynômes.

Développement en série d'Edgeworth Le développement en série d'Edgeworth permet d'approcher une densité lorsque $p_o(u)$ est gaussienne. Pour simplifier les expressions, et sans restreindre la généralité, on se place dans le cas gaussien standardisé. Dans ce cas, les fonctions $h_k(u)$ sont les polynômes de Hermite définis par la relation de récurrence

$$\begin{aligned} h_0(u) &= 1, \\ h_1(u) &= u, \\ h_{k+1}(u) &= u h_k(u) - \frac{d}{du} h_k(u). \end{aligned}$$

Par exemple, $h_2(u) = u^2 - 1$ et $h_3(u) = u^3 - 3u, \dots$

Comme nous l'avons indiqué en introduction le développement d'Edgeworth rassemble les termes, en faisant référence au théorème de la limite centrale, selon les puissances de N . Les coefficients des premiers termes du développement sont listés dans le tableau (1.1). Ainsi le développement en

Ordre	
$N^{-1/2}$	κ_3
N^{-1}	$\kappa_4 \quad \kappa_3^2$
$N^{-3/2}$	$\kappa_5 \quad \kappa_3\kappa_4 \quad \kappa_3^3$
N^{-2}	$\kappa_6 \quad \kappa_3\kappa_5 \quad \kappa_3^2\kappa_4 \quad \kappa_4^2 \quad \kappa_3^4$
$N^{-5/2}$	$\kappa_7 \quad \kappa_3\kappa_6 \quad \kappa_3^2\kappa_5 \quad \kappa_4^2\kappa_3 \quad \kappa_3^5 \quad \kappa_4\kappa_5 \quad \kappa_3^3\kappa_4$

Tableau 1.1: Liste des termes intervenant dans le développement de Edgeworth. Ces termes sont rangés en fonction de leur vitesse de décroissance lorsqu'ils concernent la somme de N variables aléatoires indépendantes.

série de Edgeworth de la densité $p_x(u)$ autour de $p_o(u)$ s'écrit à l'ordre $1/N$

$$\begin{aligned}
 p_x(u)/p_o(u) &= 1 \\
 &+ \frac{1}{3!} \kappa_3 h_3(u) \\
 &+ \frac{1}{4!} \kappa_4 h_4(u) + \frac{10}{6!} \kappa_3^2 h_6(u).
 \end{aligned}$$

Le développement en série de Edgeworth se généralise au cas multidimensionnel [148].

1.5.3 Critères de gaussianité et d'indépendance

La déconvolution aveugle, présentée au chapitre 3, fait appel à des mesures de la distance d'une densité de probabilité à la loi gaussienne. La séparation de sources (chapitre 7), quant à elle, est fondée sur la notion d'indépendance statistique. L'entropie, la divergence de Kullback et l'information mutuelle permettent d'introduire des critères de gaussianité et d'indépendance.

Entropie et gaussianité L'observation d'un vecteur aléatoire \mathbf{x} de densité de probabilité $p_{\mathbf{x}}(\mathbf{u})$ apporte une quantité d'information quantifiée par l'entropie

$$S(p_{\mathbf{x}}) = - \int p_{\mathbf{x}}(\mathbf{u}) \log p_{\mathbf{x}}(\mathbf{u}) d\mathbf{u}.$$

Considérons l'ensemble des vecteurs aléatoires dont la moyenne et la matrice de covariance sont fixées. Dans cet ensemble, le vecteur aléatoire ayant une entropie maximale et apportant donc une information maximale, est le vecteur aléatoire gaussien \mathbf{x}_g .

On démontre cette propriété par la méthode des multiplicateurs de Lagrange. Appliquée à ce problème de maximisation sous contraintes, elle conduit à chercher l'extremum de $\int p_{\mathbf{x}}(\mathbf{u}) [\log p_{\mathbf{x}}(\mathbf{u}) +$

$Q(\mathbf{u})]d\mathbf{u}$, $Q(\mathbf{u})$ étant un polynôme en \mathbf{u} de degré 2. En différenciant par rapport à $p_{\mathbf{x}}$ on obtient $\int \delta p_{\mathbf{x}}(\mathbf{u})[1 + \log p_{\mathbf{x}}(\mathbf{u}) + Q(\mathbf{u})]d\mathbf{u} = 0$. $\log p_{\mathbf{x}}(\mathbf{u})$ est donc un polynôme en \mathbf{u} de degré 2 ce qui démontre la propriété annoncée.

Pour chiffrer le déficit d'entropie, par rapport au vecteur aléatoire gaussien, d'un vecteur aléatoire \mathbf{x} appartenant à l'ensemble on introduit la *néguentropie*

$$J(p_{\mathbf{x}}) = S(p_{\mathbf{x}_g}) - S(p_{\mathbf{x}}). \quad (1.8)$$

Cette grandeur peut être utilisée comme une mesure de distance à la gaussianité. Cette notion de distance a été développée par [65], nous y reviendrons au chapitre 3.

La divergence de Kullback La néguentropie peut se réécrire sous une forme qui fait apparaître une autre mesure utilisée classiquement en statistique pour caractériser la distance entre deux densités de probabilité : la divergence⁵ de Kullback.

La divergence de Kullback mesure la distance entre deux densités de probabilité, $p_{\mathbf{v}}(\mathbf{u})$ et $p_{\mathbf{w}}(\mathbf{u})$, par

$$k(p_{\mathbf{v}}, p_{\mathbf{w}}) = \int p_{\mathbf{v}}(\mathbf{u}) \log \frac{p_{\mathbf{v}}(\mathbf{u})}{p_{\mathbf{w}}(\mathbf{u})} d\mathbf{u}.$$

La divergence de Kullback est négative si $p_{\mathbf{v}} \neq p_{\mathbf{w}}$, elle est nulle si $p_{\mathbf{v}} = p_{\mathbf{w}}$.

Cette propriété se démontre à partir de la relation $\log w \leq w - 1$, l'égalité ayant lieu pour $w = 1$

$$k(p_{\mathbf{v}}, p_{\mathbf{w}}) = \int p_{\mathbf{v}}(\mathbf{u}) \log \frac{p_{\mathbf{w}}(\mathbf{u})}{p_{\mathbf{v}}(\mathbf{u})} d\mathbf{u} \leq \int p_{\mathbf{v}}(\mathbf{u}) \left(\frac{p_{\mathbf{w}}(\mathbf{u})}{p_{\mathbf{v}}(\mathbf{u})} - 1 \right) d\mathbf{u} = 0.$$

Dans la relation (1.8), on a

$$S(p_{\mathbf{x}_g}) = - \int p_{\mathbf{x}_g}(\mathbf{u}) \log p_{\mathbf{x}_g}(\mathbf{u}) d\mathbf{u} = - \int p_{\mathbf{x}}(\mathbf{u}) \log p_{\mathbf{x}_g}(\mathbf{u}) d\mathbf{u},$$

par suite de l'égalité des moments d'ordre 1 et 2 des vecteurs aléatoires \mathbf{x} et \mathbf{x}_g . La néguentropie de x est donc

$$J(p_{\mathbf{x}}) = \int p_{\mathbf{x}}(\mathbf{u}) \log \frac{p_{\mathbf{x}}(\mathbf{u})}{p_{\mathbf{x}_g}(\mathbf{u})} d\mathbf{u} = k(p_{\mathbf{x}}, p_{\mathbf{x}_g}),$$

et est donc égale à la divergence de Kullback entre la densité de \mathbf{x} et la densité du vecteur aléatoire gaussien ayant les mêmes moments d'ordre 1 et 2.

On peut également appliquer la divergence de Kullback à la mesure de l'indépendance statistique.

⁵Le terme de divergence est utilisé car cette mesure n'est pas une distance au sens mathématique du terme.

Critères d'indépendance Soit un vecteur aléatoire \mathbf{x} de densité de probabilité $p_{\mathbf{x}}(\mathbf{u})$. Si ce vecteur est formé de variables indépendantes sa densité de probabilité est le produit des densités de probabilité marginales de chacune de ses composantes $\prod_i p_{x_i}(u_i)$. La divergence de Kullback entre $p_{\mathbf{x}}(\mathbf{u})$ et $\prod_i p_{x_i}(u_i)$ donne une mesure de l'indépendance statistique des composantes de \mathbf{x} que l'on appelle *l'information mutuelle*

$$I(p_{\mathbf{x}}) = k(p_{\mathbf{x}}, \prod_i p_{x_i}(u_i)) = \int p_{\mathbf{x}}(\mathbf{u}) \log \frac{p_{\mathbf{x}}(\mathbf{u})}{\prod_i p_{x_i}(u_i)} d\mathbf{u}.$$

Associons au vecteur aléatoire \mathbf{x} le vecteur aléatoire gaussien \mathbf{x}_g ayant les mêmes moments d'ordre 1 et 2, on peut alors écrire

$$\begin{aligned} J(p_{\mathbf{x}}) - \sum_i J(p_{x_i}) &= \int p_{\mathbf{x}}(\mathbf{u}) \log \frac{p_{\mathbf{x}}(\mathbf{u})}{p_{\mathbf{x}_g}(\mathbf{u})} d\mathbf{u} - \sum_i \int p_{x_i}(u_i) \log \frac{p_{x_i}(u_i)}{p_{x_{g_i}}(u_i)} du_i \\ &= \int p_{\mathbf{x}}(\mathbf{u}) \left[\log \frac{p_{\mathbf{x}}(\mathbf{u})}{p_{\mathbf{x}_g}(\mathbf{u})} - \log \frac{\prod_i p_{x_i}(u_i)}{\prod_i p_{x_{g_i}}(u_i)} \right] d\mathbf{u} \end{aligned}$$

car $\int p_{\mathbf{x}}(\mathbf{u}) \log \prod_i p_{x_i}(u_i) d\mathbf{u} = \sum_i \int p_{x_i}(u_i) \log p_{x_i}(u_i) du_i$. En manipulant l'expression entre crochets de la dernière expression, on obtient

$$J(p_{\mathbf{x}}) - \sum_i J(p_{x_i}) = I(p_{\mathbf{x}}) - I(p_{\mathbf{x}_g}),$$

donnant finalement

$$I(p_{\mathbf{x}}) = I(p_{\mathbf{x}_g}) + J(p_{\mathbf{x}}) - \sum_i J(p_{x_i}).$$

En appelant C la matrice de covariance du vecteur aléatoire \mathbf{x}_g , un calcul direct donne

$$I(p_{\mathbf{x}_g}) = \frac{1}{2} \log \frac{\prod_i C_{ii}}{\text{Det}[C]},$$

$\text{Det}[C]$ étant le déterminant de la matrice C .

Le terme $J(p_{\mathbf{x}}) - \sum_i J(p_{x_i})$, formé de négentropies, chiffre l'écart à la gaussianité. Ce terme ne dépend donc que des statistiques d'ordre supérieur à 2.

Ces résultats nous indiquent que l'information mutuelle du vecteur aléatoire \mathbf{x} est la somme de termes associés aux statistiques d'ordre 2 et de termes qui font apparaître les propriétés statistiques d'ordre supérieur à 2. Cette interprétation sera très utile en séparation de sources (chapitre 7).

1.6 Estimation des moments et cumulants

L'utilisation des statistiques d'ordre supérieur passe par leur estimation. Ce paragraphe présente quelques éléments sur l'estimation, essentiellement dans le cas scalaire. Commençons par une remarque.

Remarque sur le centrage Les développements théoriques sont généralement très simplifiés si les grandeurs aléatoires en jeu sont centrées. Or, pratiquement, ceci n'est pas toujours le cas. Les traitements commencent alors en général par une étape de centrage des données à traiter. Lorsque le nombre de données est grand, on peut supposer que le centrage est effectué de manière parfaite.

Nous supposerons dans la suite que les variables aléatoires sont centrées.

1.6.1 Estimation des moments

Soit x une variable aléatoire scalaire centrée, et $x_n, 1 \leq n \leq N$, N réalisations de x . L'estimateur le plus simple des moments d'ordre r de x est alors⁶

$$\widehat{\mu}_{(r)} = \frac{1}{N} \sum_{n=1}^N x_n^r$$

Cet estimateur est évidemment non biaisé puisque $E[\widehat{\mu}_{(r)}] = \mu_{(r)}$. De plus, si les x_n sont des réalisations indépendantes de x , alors la variance de l'estimateur est donné par $\text{Var}[\widehat{\mu}_{(r)}] = \text{Var}[x^r]/N$. Une conclusion à tirer de ce résultat est que la variance de l'estimation d'une grandeur d'ordre r est en O_{2r}/N , où O_{2r} est un terme d'ordre $2r$.

Pour finir, notons que l'estimateur des moments est un estimateur consistant puisque

$$\lim_{N \rightarrow +\infty} \text{Var}[\widehat{\mu}_{(r)}] = 0$$

Ayant à disposition des estimateurs des moments, nous pouvons construire des estimateurs de cumulants, en utilisant la formule de Leonov et Shirayev.

1.6.2 Estimation des cumulants

Une présentation détaillée de la théorie de l'estimation des cumulants peut être trouvée dans [115, 118, 148]. Nous nous contentons de développer quelques idées en nous appuyant sur l'estimation du cumulants d'ordre 4.

La première idée pour estimer un cumulants est de tirer partie de la formule de Leonov et Shirayev, qui relie les cumulants aux moments. Un estimateur de cumulants est obtenu en remplaçant dans cette formule les moments par leurs estimateurs.

Le cumulants d'ordre 4 s'écrit en fonction des moments selon

$$\kappa_{(4)} = \mu_{(4)} - 3\mu_{(2)}^2.$$

Un estimateur de cumulants d'ordre 4 est alors

$$\widehat{\kappa}_{(4)} = \widehat{\mu}_{(4)} - 3\widehat{\mu}_{(2)}^2.$$

⁶Dans la suite, l'indice x relatif au nom de la variable est supprimé.

L'étude de cet estimateur commence par l'évaluation de son biais.

Etude du biais Réécrivons l'estimateur sous la forme

$$\widehat{\kappa}_{(4)} = f(N) \sum_{i=1}^N x_i^4 + g(N) \sum_{i,j=1}^N x_i^2 x_j^2, \quad (1.9)$$

où $f(N) = 1/N$ et $g(N) = -3/N^2$. La moyenne de $\widehat{\kappa}_{(4)}$ est alors

$$E[\widehat{\kappa}_{(4)}] = (Nf(N) + Ng(N))\mu_{(4)} + N(N-1)g(N)\mu_{(2)}^2 \quad (1.10)$$

soit encore

$$E[\widehat{\kappa}_{(4)}] = \kappa_{(4)} - 3(\kappa_{(4)} + 2\mu_{(2)}^2)/N$$

L'estimateur est donc biaisé. Toutefois, le biais tend vers 0 lorsque le nombre d'échantillons tend vers l'infini. Dans les situations pratiques où le nombre de données est grand, $\widehat{\kappa}_{(4)}$ est un estimateur approximativement non biaisé. Toutefois, si l'échantillon est de taille courte, le biais peut devenir gênant. Pour ces situations, il existe des estimateurs non biaisés des cumulants : les k -statistiques.

k -statistiques L'écriture de $\widehat{\kappa}_{(4)}$ sous la forme (1.9) montre que les fonctions $f(N)$ et $g(N)$ sont libres : on peut les choisir de sorte à annuler le biais.

En reprenant l'équation (1.10), on s'aperçoit que le biais est nul si

$$Nf(N) + Ng(N) = 1 \text{ et } N(N-1)g(N) = -3,$$

conditions satisfaites par

$$\begin{cases} f(N) = \frac{N+2}{N(N-1)} \\ g(N) = \frac{-3}{N(N-1)} \end{cases}$$

L'estimateur ainsi défini, appelé k -statistique, est noté $\widehat{k}_{(4)}$ et s'écrit explicitement

$$\widehat{k}_{(4)} = \frac{N+2}{N(N-1)} \sum_{i=1}^N x_i^4 - \frac{3}{N(N-1)} \sum_{i,j=1}^N x_i^2 x_j^2.$$

Les k -statistiques sont donc les estimateurs non biaisés des cumulants. Elles se définissent d'une façon très générale, pour des variables scalaires ou multidimensionnelles. P. Mc Cullagh en fait un exposé très détaillé dans [148]. Examinons maintenant la variance d'estimation d'un cumulant d'ordre 4.

Variance d'estimation Les calculs de variance sont en général fastidieux. Des formules exactes existent (voir [115, 118, 148]), mais les résultats approchés suffisent pour caractériser l'estimation. On montre notamment qu'à l'ordre $1/N$, la variance de la k -statistique est

$$\begin{aligned} \text{Var} \left[\widehat{k_{(4)}} \right] &= \frac{1}{N} \left(\kappa_{(8)} + 16\kappa_{(6)}\kappa_{(2)} + 48\kappa_{(5)}\kappa_{(3)} + 34\kappa_{(4)}^2 \right. \\ &\quad \left. + 72\kappa_{(4)}\kappa_{(2)}^2 + 144\kappa_{(3)}^2\kappa_{(2)} + 24\kappa_{(2)}^4 \right), \end{aligned}$$

ce résultat étant identique pour $\widehat{\kappa_{(4)}}$, estimateur approximativement non biaisé du cumulants.

Les estimateurs des cumulants sont donc consistants puisque la variance d'estimation tend vers 0 lorsque le nombre de données tend vers l'infini.

Ayant caractérisé les deux premiers ordres statistiques des estimateurs, il est intéressant d'examiner les ordres supérieurs, afin de mieux caractériser la distribution des erreurs commises.

Tendance vers la gaussianité Les cumulants d'ordre supérieur à 2 des estimateurs peuvent être calculés, mais les calculs deviennent extrêmement pénibles (nous renvoyons encore à [115, 118, 148] pour plus de précisions). Toutefois, quelques arguments simples peuvent être avancés.

Tout d'abord, le cumulants d'ordre q de l'estimateur du cumulants d'ordre r a des termes dominants en $1/N^{q-1}$. Ceci montre que, asymptotiquement, les cumulants d'ordre supérieur ou égal à 3 s'annulent : l'estimateur tend à devenir une variable aléatoire gaussienne (notons que l'écart type de cette variable décroît en \sqrt{N}). Ce résultat est intéressant, et est souvent à la base de méthodes d'inférence statistique.

Toutefois, un examen un peu plus approfondi révèle que le cumulants d'ordre q de l'estimateur du cumulants d'ordre r se comporte comme O_{rq}/N^{q-1} , où O_{rq} est un terme statistique d'ordre rq . En général, ce terme peut être très grand, de sorte que la convergence vers la loi gaussienne peut devenir très lente. L'utilisation de résultats asymptotiques pour des échantillons de taille courte peut donc conduire à des erreurs importantes lorsque l'on manipule des statistiques d'ordre supérieur.

1.6.3 Estimation de l'asymétrie et du kurtosis

En ce qui concerne les cumulants standardisés, il n'existe pas d'estimateur non biaisé qui soit indépendant de la distribution de la variable aléatoire étudiée x . L'asymétrie et le kurtosis, qui sont les grandeurs standardisées qui retiendront essentiellement notre intérêt, seront estimés par les grandeurs biaisées suivantes

$$\widehat{\mathcal{K}}_{x(3)} = \widehat{k_{(3)}} / \widehat{k_{(2)}}^{3/2},$$

$$\widehat{\mathcal{K}}_{x(4)} = \widehat{\kappa}_{(4)} / \widehat{\kappa}_{(2)}^2.$$

Plus généralement, on conviendra de noter

$$\widehat{\mathcal{K}}_{x(r)} = \widehat{\kappa}_{(r)} / \widehat{\kappa}_{(2)}^{r/2}. \quad (1.11)$$

En raison de leur complexité, les propriétés statistiques des estimateurs standardisés $\widehat{\mathcal{K}}_{x(r)}$ définis en (1.11) ont été étudiées seulement de façon approchée pour de grandes valeurs de N [118, 148]. On notera en particulier qu'ils sont biaisés au premier ordre, et qu'ils sont corrélés (leur biais dépend d'ailleurs des cumulants d'ordre plus élevé) [115].

Toutefois, il existe des résultats exacts dans le cas gaussien. On montre en effet [115, 140, 148]

$$\begin{aligned} E \left[\widehat{\mathcal{K}}_{x(3)} \right] &= 0, \\ E \left[\widehat{\mathcal{K}}_{x(4)} \right] &= 0, \\ \text{Var} \left[\widehat{\mathcal{K}}_{x(3)} \right] &= \frac{6N(N-1)}{(N-2)(N+1)(N+3)} \approx \frac{6}{N}, \\ \text{Var} \left[\widehat{\mathcal{K}}_{x(4)} \right] &= \frac{24N(N-1)^2}{(N-3)(N-2)(N+3)(N+5)} \approx \frac{24}{N}. \end{aligned}$$

Nous arrêtons ici ce petit aperçu sur la théorie de l'estimation des cumulants. Encore une fois, nous invitons les lecteurs courageux à consulter [115, 118, 148].

1.6.4 Estimation récursive

Il est possible d'estimer les cumulants sans passer explicitement par l'intermédiaire des moments. Donnons-en un exemple qui parlera de lui-même. Pour le cumulante d'ordre 4 d'une variable aléatoire centrée, nous pouvons en effet adopter le schéma d'estimation

$$\begin{aligned} \widehat{\kappa}_{(4)} &= \frac{1}{N} \sum_{n=1}^N [x(n)^4 - 3x(n)^2 \widehat{\kappa}_{(2)}], \\ \widehat{\kappa}_{(2)} &= \frac{1}{M} \sum_{m=1}^M x(m)^2. \end{aligned}$$

Ce type d'estimateur a été proposé récemment par P.O. Amblard et J.M. Brossier [14], et présente un intérêt dans une implantation récursive (en ligne). Une version récursive est

$$\begin{cases} \widehat{\kappa}_{(2)_t} &= (1 - \gamma_t) \widehat{\kappa}_{(2)_{t-1}} + \gamma_t x(t)^2 \\ \widehat{\kappa}_{(4)_t} &= (1 - \alpha_t) \widehat{\kappa}_{(4)_{t-1}} + \alpha_t (x(t)^4 - 3x(t)^2 \widehat{\kappa}_{(2)_{t-1}}). \end{cases}$$

Cet algorithme est asymptotiquement non biaisé dans le cas d'un processus i.i.d. (échantillons indépendants et identiquement distribués). Ceci provient du fait que $x(t)^4 - 3x(t)^2\widehat{\kappa}_{(2)_{t-1}}$ est un estimateur sans biais du cumulants d'ordre 4, dans le cas i.i.d. De plus, l'algorithme est très robuste vis-à-vis de γ_t . En effet, il est montré dans [14] que dans un contexte adaptatif pour lequel on choisit $\gamma_t = \gamma$ et $\alpha_t = \alpha$ pour s'adapter aux non-stationnarités, prendre $\gamma = 1$ ne perturbe pas trop la variance asymptotique de l'estimation. Ce choix permet d'estimer le cumulants directement en fonction des données par une seule récursion qui s'écrit

$$\widehat{\kappa}_{(4)_t} = (1 - \alpha)\widehat{\kappa}_{(4)_{t-1}} + \alpha(x(t)^4 - 3x(t)^2x(t-1)^2).$$

1.7 Conclusion

Nous avons axé cette présentation rapide des statistiques d'ordre supérieur sur les propriétés utilisées dans la suite de l'ouvrage. Ce sujet très vaste est abordé dans plusieurs ouvrages. Les lecteurs soucieux d'approfondir leurs connaissances pourront se reporter à [29, 36, 115, 148, 170], notamment pour les problèmes d'estimation dans le cas multivariable.

Nous avons néanmoins insisté sur certains aspects (comme les variables complexes), qui nous semblent avoir bénéficié de peu d'égard dans ces ouvrages.

Partie I

Signaux monodimensionnels

Dans les nombreux domaines du traitement du signal, on dispose de capteurs délivrant des informations que l'on associe dans un signal à une ou plusieurs dimensions. L'étude de ces signaux est en général menée en deux étapes. Dans la première, traitée dans la partie 1 de cet ouvrage, on s'intéresse à chacun des signaux considérés indépendamment les uns des autres. La seconde étape, développée dans la partie 2, concerne l'étude des relations entre les différents signaux, en les considérant comme un signal à plusieurs dimensions.

L'étude du signal à une dimension est donc la première tâche qui attend le traicteur de signaux. Ce signal peut être représenté dans de nombreux domaines, ceci apportant une riche moisson d'informations.

Il existe de nombreuses situations dans lesquelles les signaux à une dimension jouent un rôle essentiel. Dans les techniques de radar ou de sonar, le signal dans sa représentation temporelle donne la distance de la cible, alors que dans sa représentation fréquentielle il conduit à une mesure de la vitesse par effet Doppler. Les informations contenues dans un électrocardiogramme ou dans un électroencéphalogramme permettent de déterminer l'état d'un patient. L'observation du contenu spectral d'un signal vibratoire ou acoustique d'une machine renseigne sur son régime de fonctionnement, et est utilisé pour le diagnostic. Le signal de parole véhicule l'acquis de l'hominisation. . .

Le traitement du signal a développé de nombreuses méthodes de caractérisation et de représentation des signaux stationnaires à une dimension. Nous montrons, dans le chapitre 2 présentant les multicorrélations et les multispectres, comment ces méthodes de description et de caractérisation peuvent être enrichies par la prise en compte de propriétés d'ordre supérieur à 2. Pour les signaux non stationnaires, le chapitre 4 étend aux ordres supérieurs les outils performants de description constitués par les représentations temps-fréquences. Une attention spéciale est portée aux signaux cyclostationnaires qui se rencontrent, en particulier, dans les télécommunications ou dans le contrôle et le diagnostic des machines. Cet ensemble de moyens nouveaux constitue ce que nous conviendrons d'appeler les multi-outils.

Les statistiques d'ordre supérieur ont permis le développement de techniques originales, non accessibles aux approches classiques. Le chapitre 3 donne un panorama des informations supplémentaires, issues de la prise en compte des statistiques d'ordre supérieur, dans l'étude et l'identification des systèmes linéaires. Il s'agit là du problème à une dimension type. En utilisant les statistiques d'ordre supérieur et seulement la sortie d'un système linéaire, on peut identifier complètement ce système. La réalisation d'une déconvolution aveugle (ou autodidacte, ou myope. . .) est alors possible.

Les systèmes non linéaires, décrits au chapitre 5, sont aussi un domaine privilégié pour les statistiques d'ordre supérieur. La corrélation (ordre 2) ou le spectre sont insensibles aux non-linéarités. Ces non-linéarités se manifestent, en particulier, par l'apparition de couplages entre des fréquences différentes. Ces couplages ne sont pas apparents sur le spectre d'un signal stationnaire

dont les composantes spectrales sont décorréélées. Ils apparaissent seulement, nous le verrons au chapitre 5, sur les statistiques d'ordre supérieur à 2.

Un important travail de découverte nous attend dans l'étude des signaux à une dimension. Après avoir introduit les multi-outils voyons leurs applications principales.

CHAPITRE 2

Multicorrélations et multispectres

Les propriétés d'ordre supérieur des signaux aléatoires stationnaires peuvent être décrites, comme le sont les propriétés d'ordre 2, dans le domaine du temps ou dans le domaine des fréquences. L'étude dans le domaine du temps conduit aux multicorrélations, l'étude dans le domaine des fréquences aux multispectres.

Les différentes situations expérimentales produisent divers types de signaux [147]. Les signaux aléatoires sont décrits par leurs propriétés statistiques : c'est à partir des grandeurs moyennes associées à ces signaux, moments et cumulants (chapitre 1), que l'on introduira de manière naturelle les multicorrélations et les multispectres. On rencontre également des signaux déterministes : nous montrerons, à partir des propriétés d'ergodisme, comment on peut étendre les définitions des multicorrélations et des multispectres à ces signaux.

Le développement des moyens numériques de traitement conduit à utiliser de plus en plus des signaux à temps discret, obtenus par échantillonnage des signaux à temps continu. Nous considérerons ces deux types de signaux : à temps discret et à temps continu. Cela nous permettra de décrire les signaux naturels, issus du capteur, qui sont à temps continu et les signaux utilisés en pratique par les systèmes modernes de traitement qui sont dans la quasi-totalité des cas à temps discret.

2.1 Multicorrélations

Les multicorrélations sont définies à partir des cumulants (chapitre 1) d'ordre supérieur à 2 comme la corrélation est définie à partir du cumulants d'ordre 2 [36, 87, 121, 150, 161, 162]. Nous allons traiter d'abord le cas des signaux à temps continu. Nous présenterons ensuite le cas des signaux à temps discret.

2.1.1 Définitions

Soit un signal aléatoire $x(t)$ à valeurs réelles, la multicorrélation d'ordre p de $x(t)$ est le cumulants d'ordre p des valeurs (variables aléatoires) du signal $x(t)$ aux instants t_0, t_1, \dots, t_{p-1}

$$C_{x^{(p)}}(\mathbf{t}) = \text{Cum}[x(t_0), x(t_1), \dots, x(t_{p-1})], \quad (2.1)$$

où $\mathbf{t} = (t_0, t_1, \dots, t_{p-1})$. Dans le cas général, la multicorrélation d'ordre p est fonction d'un temps absolu t_0 et des $p - 1$ temps t_i . Lorsque le signal $x(t)$ est stationnaire au sens strict, ses statistiques sont invariantes par changement de l'origine des temps. Un des temps peut être pris comme origine (nous choisissons $t = t_0$ ici). La multicorrélation n'est alors fonction que des $p - 1$ écarts de temps $\tau_i = t_i - t$. La multicorrélation d'ordre p est alors

$$C_{x^{(p)}}(\boldsymbol{\tau}) = \text{Cum}[x(t), x(t + \tau_1), \dots, x(t + \tau_{p-1})].$$

où $\boldsymbol{\tau} = (\tau_1, \dots, \tau_{p-1})$. On retrouve la corrélation classique (cumulant d'ordre 2)

$$C_{x^{(2)}}(\tau_1) = \text{Cum}[x(t), x(t + \tau_1)],$$

la bicorrélation (cumulant d'ordre 3)

$$C_{x^{(3)}}(\tau_1, \tau_2) = \text{Cum}[x(t), x(t + \tau_1), x(t + \tau_2)],$$

la tricorrélation (cumulant d'ordre 4)

$$C_{x^{(4)}}(\tau_1, \tau_2, \tau_3) = \text{Cum}[x(t), x(t + \tau_1), x(t + \tau_2), x(t + \tau_3)] \quad (2.2)$$

et ainsi de suite ...

Les multicorrélations, définies à partir des cumulants, peuvent s'exprimer en fonction des moments en utilisant les relations entre les cumulants et les moments données au chapitre 1. Lorsque les signaux sont centrés, ce qui est souvent le cas, la corrélation et la bicorrélation s'expriment directement en fonction des moments

$$\begin{aligned} C_{x^{(2)}}(\tau_1) &= E[x(t)x(t + \tau_1)], \\ C_{x^{(3)}}(\tau_1, \tau_2) &= E[x(t)x(t + \tau_1)x(t + \tau_2)]. \end{aligned}$$

La tricorrélation s'exprime en fonction du moment du quatrième ordre et de la corrélation

$$\begin{aligned} C_{x^{(4)}}(\tau_1, \tau_2, \tau_3) &= E[x(t)x(t + \tau_1)x(t + \tau_2)x(t + \tau_3)] \\ &- C_{x^{(2)}}(\tau_1)C_{x^{(2)}}(\tau_2 - \tau_3) \\ &- C_{x^{(2)}}(\tau_2)C_{x^{(2)}}(\tau_3 - \tau_1) \\ &- C_{x^{(2)}}(\tau_3)C_{x^{(2)}}(\tau_1 - \tau_2). \end{aligned}$$

Il est naturel de se demander, comme l'ont fait certains auteurs, si la définition des multicorrélations en fonction des moments ne serait pas préférable à la définition en fonction des cumulants. Nous reviendrons sur ce point en 2.1.3 et nous donnerons des arguments en faveur de la définition en fonction des cumulants. Le terme multicorrélation désignera la multicorrélation définie à partir du cumulants. Dans certaines situations, il pourra être utile de considérer les multicorrélations définies en fonction des moments : nous les introduirons en les dénommant multicorrélations des moments.

2.1.2 Symétries

Plaçons-nous dans le cas des signaux stationnaires. Dans cette situation, les différents instants auxquels on prélève le signal pour calculer le cumulants donnant la multicorrélation jouent un rôle symétrique. On peut donc choisir arbitrairement l'instant de référence¹. De même, l'ordre dans lequel on fait apparaître les différents retards comme arguments de la multicorrélation est arbitraire. Il existe donc un ensemble de valeurs identiques de la multicorrélation. Ceci entraîne, pour la multicorrélation, des symétries dans l'espace des retards.

On peut évaluer simplement le nombre de valeurs égales de la multicorrélation en fonction de l'ordre. Considérons une multicorrélation d'ordre p , mettant en jeu p instants. Il y a p choix possibles de l'instant de référence et, pour chaque choix de l'instant de référence, il y a $(p-1)!$ ordres possibles des $p-1$ autres instants. On voit ainsi qu'une multicorrélation d'ordre p aura, dans l'espace des retards, $p!$ valeurs égales. Pour la corrélation ($p=2$), il y a deux valeurs égales, ce qui se traduit par la symétrie bien connue

$$C_{x(2)}(\tau_1) = C_{x(2)}(-\tau_1).$$

Il serait fastidieux de détailler les symétries des multicorrélations dans le cas général. Pour illustrer notre propos, traitons le cas de la bicorrélation pour laquelle il y a $3! = 6$ valeurs égales. En prenant l'origine en t , on obtient

$$C_{x(3)}(\tau_1, \tau_2) = C_{x(3)}(\tau_2, \tau_1),$$

en prenant l'origine en $t + \tau_1$, on a

$$C_{x(3)}(\tau_1, \tau_2) = C_{x(3)}(-\tau_1, \tau_2 - \tau_1),$$

$$C_{x(3)}(\tau_1, \tau_2) = C_{x(3)}(\tau_2 - \tau_1, -\tau_1),$$

et en prenant l'origine en $t + \tau_2$, on obtient

$$C_{x(3)}(\tau_1, \tau_2) = C_{x(3)}(-\tau_2, \tau_1 - \tau_2),$$

$$C_{x(3)}(\tau_1, \tau_2) = C_{x(3)}(\tau_1 - \tau_2, -\tau_2).$$

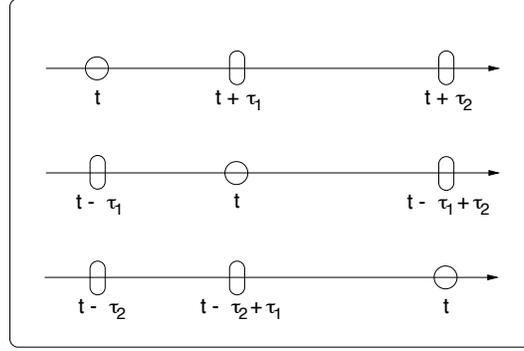


Figure 2.1: *Les différentes combinaisons de trois instants donnant la même bicorrélation. Ces combinaisons induisent les symétries de la bicorrélation.*

La figure (2.1) montre comment on obtient ces différentes bicorrélations à partir des valeurs du signal aux mêmes instants en modifiant la date de l'instant de référence (indiquée par un cercle sur la figure (2.1)) et l'ordre des autres instants.

Dans le plan τ_1, τ_2 les relations suivantes se traduisent par différentes symétries. Il suffit alors de calculer la bicorrélation dans le secteur I (figure (2.2)). Cette remarque sera très utile dans l'estimation des multicorrélations.

2.1.3 Signaux à mémoire limitée

Pour préciser la structure des multicorrélations introduisons les signaux à mémoire limitée. Cette situation donnera un premier élément de réponse à la question du choix entre les moments et les cumulants dans la définition des multicorrélations.

Un signal $x(t)$ sera 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 par un retard supérieur à τ_M

$$|t_1 - t_2| \geq \tau_M,$$

alors $x(t_1)$ et $x(t_2)$, valeurs de ce signal aux instants t_1 et t_2 , sont deux variables aléatoires indépendantes.

La multicorrélation d'un signal à mémoire limitée est nulle lorsque deux, au moins, des instants de référence sont séparés par un retard supérieur à τ_M . Ce résultat est issu d'une propriété des cumulants (chapitre 1) : lorsque deux, au moins, des variables aléatoires intervenant dans un cumulant sont indépendantes le cumulant est nul. Cette propriété est assez naturelle : il paraît logique que les multicorrélations d'un signal à mémoire limitée aient

¹Nous verrons au chapitre 4 que la situation est différente dans le cas de signaux non stationnaires.

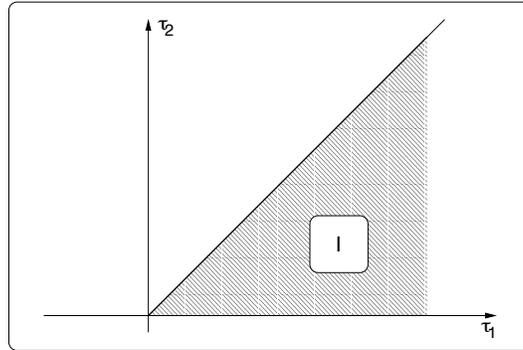


Figure 2.2: Zone de calcul de la bicorrélation. La bicorrélation dans tout le plan se déduit de la bicorrélation dans la zone I par différentes symétries.

un support limité dans l'espace des retards. Cette propriété est fautive pour les multicorrélations des moments, et ce résultat constitue un premier argument en faveur de la définition des multicorrélations à partir des cumulants et non des moments.

A titre d'exemple appliquons cette propriété à la corrélation et à la bicorrélation. La corrélation $C_{x(2)}(\tau_1)$ d'un signal à mémoire limitée est nulle si

$$|\tau_1| \geq \tau_M,$$

et le support de la corrélation est contenu dans le segment $[-\tau_M, +\tau_M]$.

La bicorrélation $C_{x(3)}(\tau_1, \tau_2)$ d'un signal à mémoire limitée est nulle dès que l'une des inégalités suivantes est vérifiée :

$$\begin{aligned} |\tau_1| &\geq \tau_M \\ |\tau_2| &\geq \tau_M \\ |\tau_1 - \tau_2| &\geq \tau_M. \end{aligned}$$

Le support de la bicorrélation d'un signal à mémoire limitée est contenu dans l'hexagone représenté sur la figure (2.3).

2.1.4 Signaux à valeurs complexes

Il est intéressant de considérer le cas des signaux à valeurs complexes que l'on rencontre assez fréquemment en traitement du signal, par exemple dans le signal analytique, après une transformation de Fourier ou en traitement d'antenne...

Comme pour les signaux à valeurs réelles, la multicorrélation d'ordre $p+q$ est un cumulants d'ordre $p+q$ mettant en jeu les valeurs du signal $x(t)$ à $p+q$ instants. Lorsque le signal $x(t)$ est à valeurs complexes, on peut définir $p+q$

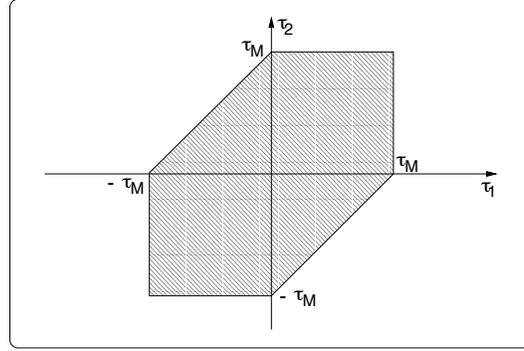


Figure 2.3: L'hexagone de la bicorrélation de signaux à mémoire limitée.

multicorrélations différentes selon le nombre de termes complexes conjugués [16, 109, 157]

$$C_{x^{(p)}}^{(q)}(\mathbf{t}) = \text{Cum}[x(t_0), x(t_1), \dots, x(t_{p-1}), x^*(t_p), \dots, x^*(t_{p+q-1})].$$

La stationnarité rend cette multicorrélation invariante par changement de l'origine des temps : la multicorrélation ne dépend que des $p + q - 1$ écarts de temps. Nous définissons alors les multicorrélations d'ordre $p + q$ du signal stationnaire $x(t)$ par

$$C_{x^{(p)}}^{(q)}(\boldsymbol{\tau}) = \text{Cum}[x(t), x(t + \tau_1), \dots, x(t + \tau_{p-1}), x^*(t - \tau_p), \dots, x^*(t - \tau_{p+q-1})].$$

où $\boldsymbol{\tau} = (\tau_1, \dots, \tau_{p+q-1})$. Dans la notation proposée, p est le nombre de termes complexes non conjugués, q le nombre de termes complexes conjugués, et enfin, les termes conjugués sont placés après les termes non conjugués.

Cette notation, qui peut paraître *a priori* complexe, a plusieurs avantages. Son intérêt essentiel est de mettre en évidence l'existence de $p + q$ multicorrélations *a priori* différentes pour les signaux à valeurs complexes². Nous montrerons sur des exemples l'existence de ces $p + q$ multicorrélations.

Cette notation englobe le cas des signaux à valeurs réelles pour lesquels les $p + q$ multicorrélations sont égales. C'est sans doute la raison qui a conduit à ne pas introduire de distinction entre les $p + q$ multicorrélations possibles. Notons enfin, nous y reviendrons, qu'il existe des classes de signaux pour lesquelles certaines multicorrélations sont nulles.

Indiquons enfin qu'il faut ajouter le retard dans les termes non conjugués et le retrancher dans les termes conjugués pour conserver la relation

²Il existe en fait des relations de conjugaisons entre certaines de ces multicorrélations. Pour cette raison, nous choisissons de ne jamais conjuguer le terme $x(t)$ correspondant à l'instant de référence.

de transformation de Fourier (théorème de Wiener-Kintchine [199]) entre les multicorrélations et les multispectres [16].

Les multicorrélations donnent une description des signaux dans le domaine des retards. Comme dans l'étude à l'ordre 2 on peut se placer dans le domaine des fréquences : on obtient alors les multispectres.

2.2 Multispectres

Les multispectres sont définis dans le cas de signaux aléatoires stationnaires. Nous verrons au chapitre 4 comment on peut traiter le cas non stationnaire par extension des représentations temps-fréquence. Les multispectres sont obtenus par transformation de Fourier multidimensionnelle des multicorrélations. Nous allons, comme pour les multicorrélations, envisager les cas des signaux à temps continu et des signaux à temps discret. Le passage du temps continu au temps discret nous conduira à revenir sur le théorème d'échantillonnage. Nous présenterons le cas des signaux à valeurs réelles avant d'étendre les définitions aux signaux à valeurs complexes. Pour donner le sens physique des multispectres, nous ferons appel à la représentation fréquentielle des signaux aléatoires stationnaires ou représentation de Cramér [29, 60]. Rappelons rapidement, sans entrer dans une analyse théorique complète, les fondements de cette représentation.

2.2.1 La représentation de Cramér

Les signaux déterministes $x(t)$ utiles pour le praticien ont, en général, une transformée de Fourier $X(\nu)$ qui peut être une fonction ou une distribution. $X(\nu)$ est dénommée la représentation fréquentielle du signal $x(t)$. On sait que la représentation temporelle $x(t)$ ou la représentation fréquentielle $X(\nu)$ contiennent toutes les informations sur le signal. Enfin le module et la phase du nombre complexe $X(\nu)$ sont le module et la phase de la composante du signal $x(t)$ à la fréquence ν . Cette décomposition du signal sur la base des fréquences pures est très utile, particulièrement dans tous les problèmes de filtrage linéaire. La transformation de Fourier inverse

$$x(t) = \int_{-\infty}^{+\infty} X(\nu) e^{2\pi j\nu t} d\nu,$$

permet d'interpréter le signal $x(t)$ comme une somme d'exponentielles éternelles (fréquences pures) d'amplitude complexe $X(\nu)$.

La définition de la représentation fréquentielle des signaux aléatoires est plus délicate. En effet la transformation de Fourier, donnant la représentation fréquentielle des signaux aléatoires par

$$X(\nu) = \int_{-\infty}^{+\infty} x(t) e^{-2\pi j\nu t} dt,$$

n'est pas convergente, en moyenne quadratique, lorsque $x(t)$ est un signal aléatoire stationnaire.

On trouvera dans Cramér [60] et Blanc-Lapierre et Fortet [29] la présentation des résultats suivants qui étendent aux signaux aléatoires la notion de représentation fréquentielle. Tout signal aléatoire stationnaire $x(t)$, appartenant à la classe $f^{(2)}$ (voir le paragraphe 2.9.2 et [29]), peut s'écrire selon l'expression intégrale

$$x(t) = \int e^{2\pi\nu t} dX(\nu).$$

La difficulté vient de la nature mathématique de $dX(\nu)$ qui est une mesure (au sens mathématique) aléatoire. $dX(\nu)$ est appelé *incrément spectral* à la fréquence ν . Les propriétés de $dX(\nu)$ sont données en général à l'ordre 2 (nous allons généraliser plus loin). On a

$$E[dX(\nu)] = 0,$$

et

$$E[dX(\nu_0)dX^*(\nu_1)] = S_{x(2)}(\nu_0)\delta(\nu_1 - \nu_0)d\nu_0d\nu_1, \quad (2.3)$$

où $S_{x(2)}(\nu_0)$ est la densité spectrale de puissance du signal³ $x(t)$. Cette relation nous montre que

- deux composantes de fréquences différentes sont décorrélées,
- la puissance du signal $x(t)$ est

$$P_x = \int_{-\infty}^{+\infty} S_{x(2)}(\nu_0)d\nu_0 = \int_{-\infty}^{+\infty} E[|dX(\nu_0)|^2];$$

$E[|dX(\nu_0)|^2]$ est donc la puissance du signal $x(t)$ dans la bande $d\nu$.

On peut donner une image intuitive plus simple de $dX(\nu)$ [29]. Considérons une batterie de filtres très sélectifs disjoints, recouvrant l'axe des fréquences. Appliquons le signal aléatoire $x(t)$ à l'entrée de ces filtres. Soit $F_\nu[x(t)]$ la sortie du filtre centré sur la fréquence ν . Lorsque la bande passante des filtres tend vers 0 la sortie du filtre centré sur la fréquence ν tend vers $F_\nu(\nu)e^{2\pi\nu t}dX(\nu)$, $F_\nu(\nu)$ étant le gain complexe du filtre F_ν . L'équation (2.3), associée aux relations de filtrage, nous indique que

- les sorties de deux filtres disjoints sont décorrélées,

³En général on suppose que le signal étudié possède une densité spectrale, on pourrait généraliser en admettant que $S_{x(2)}(\nu_0)$ peut être une distribution...

- la puissance de la sortie du filtre F_{ν_0} est

$$E[|F_{\nu_0}(x(t))|^2] \simeq S_{x(2)}(\nu_0) \int |F_{\nu_0}(\nu)|^2 d\nu,$$

soit, en normalisant le gain complexe du filtre,

$$E[|F_{\nu_0}(x(t))|^2] \simeq S_{x(2)}(\nu_0) \Delta\nu.$$

La puissance de la sortie d'un filtre est donc le spectre du signal multiplié par la bande passante du filtre.

2.2.2 Définition des multispectres

Le multispectre d'ordre p existe si la multicorrélation d'ordre p admet une transformée de Fourier à p dimensions au sens des fonctions ou au sens des distributions.

Le multispectre d'ordre p est⁴ [36, 176] :

$$S_{x(p)}(\boldsymbol{\nu}) = \int C_{x(p)}(\boldsymbol{\tau}) \exp\{-2\pi j \boldsymbol{\nu}^T \boldsymbol{\tau}\} d\boldsymbol{\tau}.$$

où $\boldsymbol{\nu} = (\nu_1, \nu_2, \dots, \nu_{p-1})$ et $\boldsymbol{\tau} = (\tau_1, \tau_2, \dots, \tau_{p-1})$. On peut revenir à la multicorrélation à partir du multispectre par transformation de Fourier inverse

$$C_{x(p)}(\boldsymbol{\tau}) = \int S_{x(p)}(\boldsymbol{\nu}) \exp\{2\pi j \boldsymbol{\nu}^T \boldsymbol{\tau}\} d\boldsymbol{\nu}. \quad (2.4)$$

2.2.3 Multispectres et représentation de Cramér

Le signal stationnaire $x(t)$ ayant une représentation de Cramér $dX(\nu)$

$$x(t) = \int e^{2\pi j \nu t} dX(\nu),$$

la multicorrélation d'ordre p de ce signal est

$$\begin{aligned} C_{x(p)}(\boldsymbol{\tau}) &= \text{Cum}[x(t), x(t + \tau_1), \dots, x(t + \tau_{p-1})] \\ &= \int \text{Cum}[dX(\nu_0), dX(\nu_1), \dots, dX(\nu_{p-1})] \\ &\times e^{2\pi j (\sum_{i=0}^{p-1} \nu_i t + \sum_{i=1}^{p-1} \nu_i \tau_i)}. \end{aligned} \quad (2.5)$$

⁴Les intégrales sont prises de $-\infty$ à $+\infty$, sauf mention explicite. De plus les intégrales multiples sont représentées sous forme d'un seul signe somme – s'il n'y a pas ambiguïté – et les éléments différentiels notés sous forme vectorielle.

Le signal étant stationnaire, sa multicorrélation ne doit pas dépendre de t . Cette condition de stationnarité impose

$$\nu_0 + \sum_{i=1}^{p-1} \nu_i = 0. \quad (2.6)$$

Cette relation définit un hyperplan dénommé *multiplicité stationnaire* [169]. En identifiant les intégrants dans les relations (2.4) et (2.5) on obtient

$$\text{Cum}[dX(\nu_0), dX(\nu_1), \dots, dX(\nu_{p-1})] = S_{x(p)}(\boldsymbol{\nu}) \delta \left[\sum_{i=0}^{p-1} \nu_i \right] d\nu_0 \dots d\nu_{p-1}, \quad (2.7)$$

qui généralise la relation (2.3) pour les statistiques d'ordre supérieur à 2.

2.2.4 Moments ou cumulants ?

Dans la relation (2.7) le multispectre est issu du cumulants des composantes spectrales $dX(\nu)$. Voyons qu'elle serait la définition des multispectres à partir des moments.

Avant de donner une formulation générale limitons-nous aux ordres inférieurs ou égaux à 3. Lorsque le signal étudié est centré, les moments et les cumulants d'ordre 2 et 3 sont identiques. La définition du spectre et du bispectre, pour des signaux centrés, peut être faite en remplaçant les cumulants apparaissant dans (2.7) par des moments : c'est pour cette raison que la distinction entre moments et cumulants n'est pas introduite dans la présentation du spectre ou du bispectre.

Le cas du trispectre est différent. Soit un signal $x(t)$ centré, la tricorrélation des moments est

$$C'_{x(4)}(\tau_1, \tau_2, \tau_3) = E[x(t)x(t+\tau_1)x(t+\tau_2)x(t+\tau_3)],$$

et elle s'écrit en fonction de la tricorrélation et de la corrélation⁵ selon

$$\begin{aligned} C'_{x(4)}(\tau_1, \tau_2, \tau_3) &= C_{x(4)}(\tau_1, \tau_2, \tau_3) + C_{x(2)}(\tau_1)C_{x(2)}(\tau_2 - \tau_3) \\ &+ C_{x(2)}(\tau_2)C_{x(2)}(\tau_3 - \tau_1) \\ &+ C_{x(2)}(\tau_3)C_{x(2)}(\tau_1 - \tau_2). \end{aligned}$$

En appliquant une transformation de Fourier 3D à la tricorrélation des moments, on obtient le trispectre des moments

$$\begin{aligned} S'_{x(4)}(\nu_1, \nu_2, \nu_3) &= S_{x(4)}(\nu_1, \nu_2, \nu_3) + S_{x(2)}(\nu_1)S_{x(2)}(\nu_2) \delta(\nu_2 + \nu_3) \\ &+ S_{x(2)}(\nu_2)S_{x(2)}(\nu_1) \delta(\nu_1 + \nu_3) \\ &+ S_{x(2)}(\nu_3)S_{x(2)}(\nu_1) \delta(\nu_1 + \nu_2). \end{aligned} \quad (2.8)$$

⁵Par corrélation, sans autre précision, nous entendons corrélation des cumulants.

La relation (2.8) entre le trispectre des moments et le trispectre indique que le trispectre des moments introduit des singularités sur les plans

$$\begin{cases} \nu_1 + \nu_2 = 0 \\ \nu_1 + \nu_3 = 0 \\ \nu_2 + \nu_3 = 0. \end{cases}$$

Notons qu'à l'extérieur de ces plans il y a égalité entre les deux trispectres. Pour un signal gaussien, le trispectre est nul. Le trispectre des moments d'un signal gaussien est donc nul en dehors des plans définis ci-dessus. Pour un multispectre d'ordre quelconque le sous-espace contenant le multispectre des moments d'un signal gaussien est dénommé multiplicité gaussienne [169].

Dans le cas du trispectre, si nous n'avions pas considéré un signal centré la présence de la moyenne aurait introduit des singularités sur les axes et à l'origine des fréquences.

De manière générale on peut montrer que le multispectre d'ordre p , défini à partir des moments, est égal au multispectre défini à partir des cumulants presque partout dans l'espace à p dimensions des fréquences. Cela signifie que les ensembles de points sur lesquels les deux multispectres sont différents sont de dimension inférieure à p .

Ces remarques seront utiles quand nous présenterons les méthodes d'estimation des multispectres. Nous verrons alors que la présence de singularités est très gênante : on sait bien qu'avant d'estimer la densité spectrale il faut retirer la valeur moyenne. Ce résultat constitue un nouvel argument en faveur de la définition des multicorrélations et des multispectres à partir des cumulants.

2.2.5 Multispectres symétriques

La multicorrélation $C_{x(p)}(\boldsymbol{\tau})$ met en jeu p valeurs du signal en temps. Il est logique de disposer d'un objet fréquentiel faisant apparaître p valeurs de la fréquence. Cet objet est appelé *multispectre symétrique*, et est la transformée de Fourier de la multicorrélation "générale" (2.1, page 36). Il est noté $\Sigma_{x(p)}(\nu_0, \boldsymbol{\nu})$ avec $\boldsymbol{\nu} = (\nu_1, \dots, \nu_{p-1})$. Il est alors relié aux incréments spectraux par la relation [36]

$$\Sigma_{x(p)}(\nu_0, \boldsymbol{\nu}) d\nu_0 d\boldsymbol{\nu} = \text{Cum}[dX(\nu_0), dX(\nu_1), \dots, dX(\nu_{p-1})]. \quad (2.9)$$

La relation (2.6) nous montre que, par suite de la stationnarité dans l'espace fréquentiel à p dimensions $(\nu_0, \nu_1, \dots, \nu_{p-1})$, le multispectre symétrique est nul en dehors de la *multiplicité stationnaire* [169] définie par

$$\sum_{i=0}^{p-1} \nu_i = 0.$$

Le “confinement” du multispectre à la multiplicité stationnaire vient de la propriété de stationnarité du signal. Ainsi il est bien connu que le spectre (dsp) de signaux stationnaires est défini sur la ligne $\nu_0 + \nu_1 = 0$ du plan ν_0, ν_1 .

On reliera la forme symétrique et la forme non symétrique du multispectre par

$$\Sigma_{x(p)}(\nu_0, \boldsymbol{\nu}) = S_{x(p)}(\boldsymbol{\nu})\delta\left(\nu_0 + \sum_{i=1}^{p-1} \nu_i\right), \quad (2.10)$$

où $\boldsymbol{\nu} = (\nu_1, \nu_2, \dots, \nu_{p-1})$.

2.2.6 Circularité des composantes spectrales

Les relations (2.9) et (2.10) montrent que le cumulante d'ordre p des composantes spectrales est nul en dehors de la multiplicité stationnaire. Cette propriété découle de la stationnarité des signaux. Pour toute fréquence ν non nulle, la condition $\sum_{i=0}^{p-1} \nu_i = 0$ n'est pas satisfaite si le cumulante est calculé avec un nombre différent de composantes conjuguées et non conjuguées. En conséquence, *pour toutes les fréquences non nulles, les composantes spectrales d'un processus stationnaire sont complexes circulaires.*

2.2.7 Symétries

Comme les multicorrélations, dont ils sont la transformée de Fourier multidimensionnelle, les multispectres ont des propriétés de symétrie. On trouvera donc pour le multispectre d'ordre p , $p!$ points, dans l'espace des fréquences, où les valeurs du multispectre sont égales. On sait que le spectre est pair, soit

$$S_{x(2)}(\nu) = S_{x(2)}(-\nu).$$

Il suffit donc de calculer le spectre pour les fréquences positives. Pour le bispectre, on a les relations

$$\begin{aligned} S_{x(3)}(\nu_1, \nu_2) &= S_{x(3)}(\nu_2, \nu_1), \\ S_{x(3)}(\nu_1, \nu_2) &= S_{x(3)}(\nu_1, -\nu_1 - \nu_2), \\ S_{x(3)}(\nu_1, \nu_2) &= S_{x(3)}(-\nu_1 - \nu_2, \nu_1), \\ S_{x(3)}(\nu_1, \nu_2) &= S_{x(3)}(\nu_2, -\nu_1 - \nu_2), \\ S_{x(3)}(\nu_1, \nu_2) &= S_{x(3)}(-\nu_1 - \nu_2, \nu_2). \end{aligned}$$

Il suffit alors de connaître le bispectre sur le domaine $\{\nu_1 \geq 0\} \cap \{\nu_2 \geq 0\} \cap \{\nu_2 \leq \nu_1\}$ pour le connaître partout.

2.2.8 Interprétation des multispectres

L'interprétation "physique" des spectres est classique : ils décrivent la répartition de la puissance du signal en fonction de la fréquence. Les multispectres quant à eux donnent une description des relations statistiques existant entre les composantes du signal étudié à différentes fréquences. Ces relations sont, en particulier, l'indicateur de transformations non linéaires dans la construction des signaux (ceci est développé au chapitre 5). Ainsi, si nous modulons (multiplions) un signal à la fréquence ν_1 par un signal à la fréquence ν_2 il apparaît une composante à la fréquence $\nu_1 + \nu_2$. On a vu plus haut que si le signal est stationnaire les composantes du signal sont décorréées. L'étude, classique, à l'ordre 2, ne peut donc pas mettre en évidence les relations existant entre les fréquences ν_1 , ν_2 et $\nu_1 + \nu_2$. Par contre le bispectre de ces trois signaux aura une valeur non nulle pour la bi-fréquence ν_1, ν_2 . Il permettra ainsi de mettre en évidence la relation entre les trois fréquences ν_1 , ν_2 et $\nu_1 + \nu_2$.

2.2.9 Signaux à valeurs complexes

Comme pour les multicorrélations il existe, dans ce cas, $p + q$ multispectres différents d'ordre $p + q$ qui sont les transformées de Fourier multidimensionnelles des $p + q$ multicorrélations⁶ [16]. On a donc, pour un signal $x(t)$ stationnaire, à valeurs complexes

$$S_{x(p)}^{(q)}(\boldsymbol{\nu}) = \int C_{x(p)}^{(q)}(\boldsymbol{\tau}) \exp\{-2\pi j \boldsymbol{\nu}^T \boldsymbol{\tau}\} d\boldsymbol{\tau},$$

où $\boldsymbol{\nu} = (\nu_1, \dots, \nu_{p+q-1})$ et $\boldsymbol{\tau} = (\tau_1, \dots, \tau_{p+q-1})$. La relation (2.7, page 44) devient

$$\text{Cum}[dX(\nu_0), dX(\nu_1), \dots, dX(\nu_{p-1}), dX^*(\nu_p), \dots, dX^*(\nu_{p+q-1})] =$$

$$S_{x(p)}^{(q)}(\boldsymbol{\nu}) \delta\left(\nu_0 + \sum_{i=1}^{p-1} \nu_i - \sum_{j=p}^{p+q-1} \nu_j\right) d\nu_0 d\nu_1 \dots d\nu_{p+q-1}.$$

La stationnarité impose alors

$$\nu_0 + \sum_{i=1}^{p-1} \nu_i - \sum_{j=p}^{p+q-1} \nu_j = 0,$$

et on obtient donc en notations symétriques

$$\Sigma_{x(p)}^{(q)}(\nu_0, \boldsymbol{\nu}) = S_{x(p)}^{(q)}(\boldsymbol{\nu}) \delta\left(\nu_0 + \sum_{i=1}^{p-1} \nu_i - \sum_{j=p}^{p+q-1} \nu_j\right)$$

⁶ Rappelons que les termes complexes conjugués sont placés après les termes non complexes conjugués, et que le premier terme de la multicorrélation, $x(t)$, n'est par convention jamais conjugué.

où le multispectre symétrique est défini par

$$\Sigma_{x^{(p)}}^{(q)}(\nu_0, \boldsymbol{\nu}) = \int C_{x^{(p)}}^{(q)}(\mathbf{t}) e^{2\pi j(\sum_{k=0}^{p-1} \nu_k t_k - \sum_{k=p}^{p+q-1} \nu_k t_k)} d\mathbf{t},$$

\mathbf{t} étant (t_0, \dots, t_{p+q-1}) .

Il y a, dans ce cas, $p + q$ multiplicités stationnaires. La multiplicité stationnaire M_p^q est définie par

$$\sum_{i=0}^{p-1} \nu_i - \sum_{j=p}^{p+q-1} \nu_j = 0.$$

A titre d'illustration considérons le cas des signaux analytiques et des signaux circulaires.

Signaux analytiques A un signal à valeurs réelles $x(t)$, on associe un signal à valeurs complexes $z(t)$ par les relations

$$\begin{cases} Z(\nu) = 2X(\nu) & \text{pour } \nu \geq 0 \\ Z(\nu) = 0 & \text{pour } \nu < 0. \end{cases}$$

On sait que l'on peut obtenir le signal analytique en utilisant un filtre de quadrature ou une transformation de Hilbert [166]. De tels signaux sont utilisés, en particulier, en télécommunication et dans l'analyse temps-fréquence.

Les signaux analytiques ne contenant que des fréquences positives ont une représentation de Cramér nulle pour les fréquences négatives. Les incréments spectraux $dX(\nu_i)$, $1 \leq i \leq p+q-1$, intervenant dans le multispectre symétrique sont donc non nuls dans le domaine D_A vérifiant

$$\nu_i > 0 \quad 1 \leq i \leq p + q - 1.$$

La multiplicité stationnaire M_p^p du multispectre sans composante complexe conjuguée du signal analytique est définie par

$$\sum_{i=0}^{2p-1} \nu_i = 0.$$

La multiplicité stationnaire M_p^p et le domaine D_A ont une intersection nulle : ceci montre que la multicorrélation $C_{z^{(p)}}^{(p)}(\boldsymbol{\tau})$ et le multispectre $S_{z^{(p)}}^{(p)}(\boldsymbol{\nu})$ sont nuls pour des signaux analytiques stationnaires.

Ceci étend un résultat connu à l'ordre 2. On sait que la corrélation

$$C_{z^{(2)}}(\tau) = E[z(t)z(t + \tau)],$$

et le spectre associé sont nuls pour un signal analytique.

Signaux circulaires Les signaux circulaires sont des signaux à valeurs complexes dont les valeurs prises à un certain nombre d'instants, quelconques pour des signaux circulaires au sens strict, ou jusqu'à p pour des signaux circulaires d'ordre p , sont des variables aléatoires circulaires [171] (chapitre 1).

Il résulte des propriétés des variables aléatoires circulaires que pour ces signaux toutes les multicorrélations $C_{x(p)}^{(q)}(\boldsymbol{\tau})$ et tous les multispectres $S_{x(p)}^{(q)}(\boldsymbol{\nu})$ qui ne vérifient pas la condition $q = p$ sont nuls.

Après avoir défini les multicorrélations et les multispectres, voyons comment ils se transforment par filtrage linéaire.

2.3 Filtrage linéaire

Les statistiques d'ordre supérieur ont une grande importance pour l'étude des systèmes linéaires (chapitre 3). Nous examinons dans ce paragraphe les transformations subies par les multicorrélations et les multispectres lors d'un filtrage linéaire. Commençons dans le domaine fréquentiel.

2.3.1 Relations de filtrage en fréquence.

Plutôt que de donner *in-extenso* les formules de filtrage, ce qui conduit à des expressions arides et assez peu lisibles, nous allons détailler le mécanisme de construction de la relation reliant le multispectre de la sortie au multispectre de l'entrée.

Soit le signal $y(t)$ issu du filtrage du signal $x(t)$ par le filtre de gain complexe $A(\nu)$. La relation entre la représentation de Cramér [36] de l'entrée et de la sortie s'écrit

$$dY(\nu) = A(\nu)dX(\nu). \quad (2.11)$$

Connaissant le multispectre de l'entrée $S_{x(p)}^{(q)}(\boldsymbol{\nu})$, nous cherchons le multispectre de la sortie $S_{y(p)}^{(q)}(\boldsymbol{\nu})$. Toutes les formules de filtrage se déduisent des remarques suivantes.

Dans un multispectre, il apparaît trois types de situations pour les signaux :

- le signal considéré occupe la première place, c'est le signal que nous avons convenu de ne pas remplacer par son complexe conjugué.
- le signal considéré n'occupe pas la première place et n'est pas complexe conjugué.
- le signal considéré n'occupe pas la première place et est complexe conjugué.

Appelons ν_i ($i = 0, \dots, p-1$) les fréquences associées aux signaux non complexes conjugués et ν_j ($j = p, \dots, p+q-1$) les fréquences associées aux signaux complexes conjugués. En utilisant la relation entrée-sortie des représentations de Cramér donnée par (2.11) et le multispectre exprimé comme un cumulants des représentations de Cramér (voir le paragraphe 2.2.9) on voit que le multispectre de la sortie est le multispectre de l'entrée multiplié par :

- $A(-\sum_i \nu_i + \sum_j \nu_j)$: c'est l'effet du filtrage sur le premier terme, la composante fréquentielle de ce terme est $-\sum_i \nu_i + \sum_j \nu_j$ par suite du confinement sur la multiplicité stationnaire,
- $A(\nu_i)$ pour chacun des termes non complexes conjugués ; c'est l'effet de filtrage de chacun de ces termes,
- $A^*(\nu_j)$ pour chacun des termes complexes conjugués ; c'est l'effet de filtrage de ces termes.

L'application de ces règles permet de résoudre tous les problèmes de filtrage. Examinons quelques exemples simples.

Le spectre $S_{y(2)}(\nu)$ du signal filtré est issu du spectre du signal d'entrée multiplié par $A(\nu)$ par suite du filtrage du terme non complexe conjugué et par $A(-\nu)$ par suite du filtrage du premier terme, on obtient donc :

$$S_{y(2)}(\nu) = A(\nu)A(-\nu)S_{x(2)}(\nu) ;$$

pour des signaux à valeurs réelles on a $A(-\nu) = A^*(\nu)$, donnant la formule connue

$$S_{y(2)}(\nu) = A(\nu)A^*(\nu)S_{x(2)}(\nu) = |A(\nu)|^2 S_{x(2)}(\nu). \quad (2.12)$$

Pour le bispectre $S_{y(3)}(\nu_1, \nu_2)$ il faut multiplier par $A(\nu_1)$, $A(\nu_2)$ et $A(-\nu_1 - \nu_2)$ donnant

$$S_{y(3)}(\nu_1, \nu_2) = A(\nu_1)A(\nu_2)A(-\nu_1 - \nu_2)S_{x(3)}(\nu_1, \nu_2),$$

soit, pour des signaux à valeurs réelles

$$S_{y(3)}(\nu_1, \nu_2) = A(\nu_1)A(\nu_2)A^*(\nu_1 + \nu_2)S_{x(3)}(\nu_1, \nu_2). \quad (2.13)$$

La comparaison de (2.12) et (2.13) est intéressante car elle introduit une des propriétés pratiques importantes du bispectre. Dans (2.12), la phase du filtre disparaît : on sait qu'il est impossible de retrouver la phase d'un filtre à partir du spectre de la sortie. Dans (2.13), en appelant $\Phi(\nu)$ la phase du filtre et $\Psi_x(\nu_1, \nu_2)$ et $\Psi_y(\nu_1, \nu_2)$ la phase du bispectre de x et de y , on a la relation

$$\Psi_y(\nu_1, \nu_2) = \Phi(\nu_1) + \Phi(\nu_2) - \Phi(\nu_1 + \nu_2) + \Psi_x(\nu_1, \nu_2),$$

montrant que le bispectre contient des informations sur la phase du filtre. Cette propriété sera utilisée au chapitre 3 pour l'identification aveugle d'un filtre.

2.3.2 Relations de filtrage en temps

En temps la relation sortie-entrée est une convolution qui s'écrit, pour les différentes composantes,

$$y(t) = \int_{-\infty}^{+\infty} a(u)x(t-u)du.$$

On en déduit, par suite de leur propriété de multilinéarité, la formule de filtrage des cumulants [161]. Le cumulants d'ordre 2 de signaux scalaires réels filtrés par un filtre de réponse impulsionnelle $a(t)$ est

$$\begin{aligned} Cum[y(t_0), y(t_0 + \tau)] &= \int \int a(u_0)a(u_1) \\ &\times Cum[x(t_0 - u_0)x(t_0 + \tau - u_1)] du_0 du_1, \end{aligned}$$

donnant pour la corrélation de ces signaux

$$C_{y(2)}(\tau) = \int \int a(u_0)a(u_1)C_{x(2)}(\tau + u_0 - u_1)du_0 du_1.$$

En procédant de même, la bicorrélation de la sortie s'écrit

$$\begin{aligned} C_{y(3)}(\tau_1, \tau_2) &= \int \int \int a(u_0)a(u_1)a(u_2) \\ &\times C_{x(3)}(\tau_1 + u_0 - u_1, \tau_2 + u_0 - u_2)du_0 du_1 du_2. \end{aligned}$$

Un cumulants d'ordre 2 de signaux à valeurs complexes filtrés par un filtre de réponse impulsionnelle $a(t)$ est

$$\begin{aligned} Cum[y(t_0), y^*(t_0 - \tau)] &= \int \int a(u_0)a^*(u_1) \\ &\times Cum[x(t_0 - u_0), x^*(t_0 - \tau - u_1)] du_0 du_1. \end{aligned}$$

donnant pour la corrélation correspondante

$$C_{y(1)}^{(1)}(\tau) = \int \int a(u_0)a^*(u_1)C_{x(1)}^{(1)}(-\tau + u_0 - u_1)du_0 du_1.$$

Pour une bicorrélation, la relation entrée-sortie est

$$\begin{aligned} C_{y(2)}^{(1)}(\tau_1, \tau_2) &= \int \int \int a(u_0)a(u_1)a^*(u_2) \\ &\times C_{x(2)}^{(1)}(\tau_1 + u_0 - u_1, -\tau_2 + u_0 - u_2)du_0 du_1 du_2. \end{aligned}$$

On peut facilement généraliser ces résultats par extrapolation : nous laisserons ce plaisir au lecteur . . .

2.4 Echantillonnage et signaux à temps discret

Ce paragraphe examine l'extension aux signaux à temps discret de la notion de multicorrélation et de multispectre.

2.4.1 L'échantillonnage

Les capteurs nous donnent des signaux $x(t)$ dépendant continûment du temps. On utilise de plus en plus des systèmes de traitement numériques (ordinateurs, microprocesseurs) qui traitent des suites de nombres⁷. Pour rendre un signal à temps continu accessible à un système numérique nous devons prélever un certain nombre de ses valeurs (échantillons). Cette opération d'échantillonnage peut être pratiquée de diverses manières [147]. Généralement on réalise un échantillonnage régulier qui consiste à prélever des échantillons régulièrement répartis dans le temps. L'écart de temps T_E qui sépare deux échantillons successifs est la période d'échantillonnage. La période d'échantillonnage est fixée par une horloge dont la précision est forcément limitée. Les éventuelles erreurs d'horloge constituent le phénomène de gigue qui vient perturber l'échantillonnage. Nous admettrons que ces erreurs sont assez faibles pour pouvoir être négligées. Nous réaliserons ainsi un échantillonnage régulier parfait.

Le théorème d'échantillonnage [147] nous indique que l'on peut reconstituer un signal à temps continu à partir du signal à temps discret, signal obtenu par échantillonnage, si le signal à temps continu a une largeur de bande limitée par B ($X(\nu) = 0$ si $|\nu| > B$) et si la période d'échantillonnage vérifie la condition de Shannon

$$T_E \leq \frac{1}{2B}.$$

Dans ces conditions on peut reconstituer la corrélation à temps continu à partir de la corrélation à temps discret. Par transformation en fréquences réduites de la corrélation à temps discret on obtient exactement la densité spectrale du signal à temps continu.

Ces résultats importants montrent que, à l'ordre 2, on ne perd pas d'information sur le signal étudié si il vérifie les conditions du théorème d'échantillonnage.

Nous nous intéressons ici à des caractéristiques d'ordre supérieur à 2 : multicorrélations et multispectres. Comment ces caractéristiques sont-elles affectées par l'échantillonnage?

L'étude des signaux à largeur de bande limitée nous permettra de poser le problème. Nous montrerons que les conditions d'échantillonnage sont plus

⁷Nous ne considérerons pas ici la quantification de ces nombres qui résulte du nombre fini de décimales.

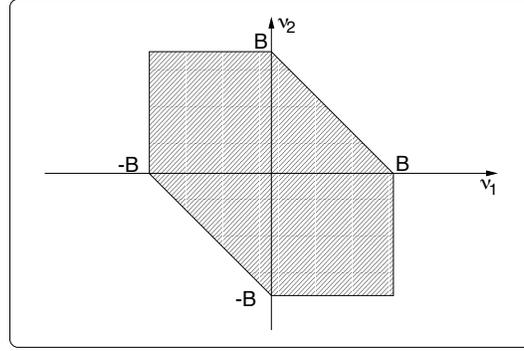


Figure 2.4: Support du bispectre des signaux à bande limitée.

strictes lorsque l'on s'intéresse à des multicorrélations ou à des multispectres d'ordre supérieur à 2 [129, 130]. Nous terminerons par l'étude du filtrage des signaux à temps discret par des filtres numériques (échantillonnés, digitaux).

2.4.2 Signaux à bande limitée et échantillonnage

Soit un signal $x(t)$ dont les multicorrélations, à tous les ordres, sont des impulsions de Dirac à l'origine. Les multispectres de ce signal sont des constantes. On appellera ce signal un signal blanc à tous les ordres ou signal blanc au sens fort. Notons au passage – on en trouvera une présentation plus détaillée dans [34] et nous en reparlerons au chapitre 3 – que la notion de blancheur se complique quand on s'intéresse à des caractéristiques d'ordre supérieur à 2. On peut définir des signaux blancs jusqu'à un certain ordre. Le signal blanc parfait est blanc à tous les ordres. Les signaux gaussiens sont particuliers ; pour ces signaux dont tous les cumulants d'ordre supérieur à 2 sont nuls, la blancheur à l'ordre 2 entraîne la blancheur parfaite. Filtrons le signal blanc $x(t)$ par un filtre passe-bas idéal dont le gain complexe $F(\nu)$ est une fonction porte de largeur⁸ $2B$:

$$F(\nu) = \Pi_{2B}(\nu);$$

nous construisons ainsi un signal $x_B(t)$ à bande limitée à B . Les multispectres de ce signal sont donnés par les relations de filtrage en fréquence (voir 2.3.1). Considérons, pour simplifier la discussion, le bispectre

$$S_{x_B(3)}(\nu_1, \nu_2) = S_{x(3)} F(\nu_1) F(\nu_2) F^*(\nu_1 + \nu_2),$$

$S_{x(3)}$ étant le bispectre de $x(t)$.

Le bispectre de $x_B(t)$ est égal à $S_{x(3)}$ dans l'hexagone représenté sur la figure (2.4) et il est égal à 0 ailleurs. Considérons le pseudo-signal à temps

⁸Ce filtre est irréalisable mais on peut l'approcher aussi près que l'on veut.

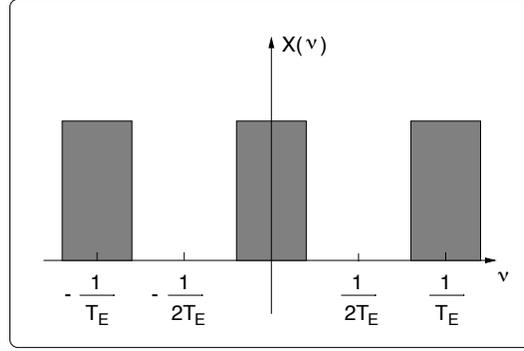


Figure 2.5: Représentation de Cramér des signaux à bande limitée échantillonnés

continu $x_E(t)$ obtenu à partir des échantillons de $x_B(t)$

$$x_E(t) = \sum_n x_B(nT_E)\delta(t - nT_E).$$

Comme à l'ordre 2, le bispectre de $x_E(t)$ est périodisé dans le plan ν_1, ν_2 . On pourrait être tenté de croire que l'on peut retrouver sans erreur le bispectre du signal à temps continu dès que la période d'échantillonnage vérifie la condition de Shannon⁹. Il n'en est pas ainsi et cela est montré dans [129, 130]. La raison est que, contrairement au cas de l'ordre 2, les multispectres d'ordre supérieur font intervenir simultanément plusieurs fréquences. Le bispectre $S_{x(3)}(\nu_1, \nu_2)$, par exemple, fait appel aux fréquences ν_1, ν_2 et $\nu_1 + \nu_2$. Ce sont les représentations de Cramér qui sont en fait périodisées et pour que le bispectre soit non nul il suffit que les trois fréquences ν_1, ν_2 et $\nu_1 + \nu_2$ soient situées dans un des segments de l'axe des fréquences sur lesquels la représentation de Cramér est non nulle. En examinant la figure (2.5) on s'aperçoit que ceci peut se réaliser si $\nu_1 \leq B, \nu_2 \leq B$ et $\nu_E - B \leq \nu_1 + \nu_2 \leq \nu_E + B$ où $\nu_E = 1/T_E$ est la fréquence d'échantillonnage. Le cas le plus défavorable est celui pour lequel $\nu_1 = B, \nu_2 = B$. Pour que, dans ce cas, $\nu_1 + \nu_2$ ne soit pas dans le segment $[\nu_E - B, \nu_E + B]$ il faut que B vérifie

$$2B \leq \nu_E - B,$$

soit

$$B \leq \frac{1}{3}\nu_E. \quad (2.14)$$

Pour retrouver, après échantillonnage le motif du bispectre du signal à temps continu dans le carré

$$|\nu_1| \leq \frac{\nu_E}{2} \quad \text{et} \quad |\nu_2| \leq \frac{\nu_E}{2},$$

⁹Dans ce cas, les différents motifs obtenus par périodisation ne se chevauchent pas.

la largeur de bande du signal à temps continu doit vérifier la condition (2.14) qui est plus forte que la condition de Shannon.

Ce résultat obtenu sur le bispectre peut être généralisé aux multispectres. Le multispectre d'ordre p fait apparaître la représentation de Cramér aux fréquences ν_i ($0 \leq i \leq p-1$) et à la fréquence $\sum_{i=0}^{p-1} \nu_i$. En reprenant les arguments développés plus haut on voit que le multispectre échantillonné sera égal au multispectre du signal à temps continu dans l'hypercube

$$|\nu_i| \leq \frac{\nu_E}{2} \quad 0 \leq i \leq p-1,$$

si la largeur de bande vérifie la condition :

$$B \leq \frac{1}{p} \nu_E.$$

Nous pouvons donc énoncer le théorème d'échantillonnage, pour les multispectres, sous la forme suivante :

Pour que le multispectre d'ordre p du signal échantillonné soit égal au multispectre du signal à temps continu $x(t)$ dans l'hypercube $|\nu_i| \leq \nu_E/2$, ($1 \leq i \leq p-1$), il faut :

- que le signal $x(t)$ ait une largeur de bande limitée par B ,
- que la période d'échantillonnage T_E vérifie la condition

$$T_E \leq \frac{1}{pB}.$$

Cette présentation du théorème d'échantillonnage soulève diverses questions sur lesquelles nous allons revenir en étudiant le support du multispectre.

2.4.3 Sur la géométrie du support des multispectres

Revenons, pour simplifier, au bispectre du signal échantillonné $x_E(t)$ présenté ci-dessus. Considérons (figure (2.6)) dans le plan ν_1, ν_2 le carré défini par

$$|\nu_1| \leq \frac{\nu_E}{2} \quad \text{et} \quad |\nu_2| \leq \frac{\nu_E}{2};$$

par suite des relations de symétrie (paragraphe 2.2.7) il suffit de connaître le bispectre dans les triangles T_1 et T_2 . Par ailleurs, en étudiant le support du bispectre de signaux à bande limitée (voir 2.4.2) nous avons vu que, pour des signaux stationnaires, le bispectre doit être confiné dans le triangle T_1 . Dans la section précédente, nous avons cependant montré que l'on peut voir apparaître des valeurs non nulles dans le triangle T_2 si, tout en respectant la condition de Shannon, on ne vérifie pas la condition plus forte (2.14). Les termes parasites,

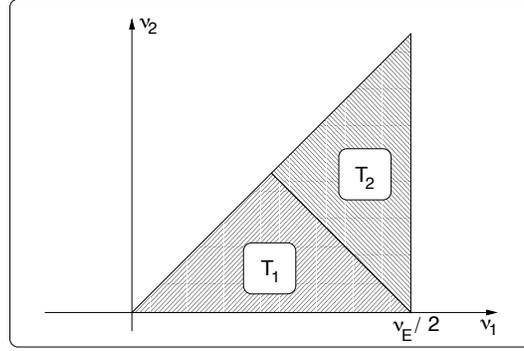


Figure 2.6: Les différentes zones du bispectre. Le bispectre d'un signal correctement échantillonné est nul dans le triangle T_2 .

issus du non-respect de la condition (2.14), ne peuvent pas apparaître dans le triangle T_1 . Ceci nous indique que pour un signal *stationnaire* on peut se contenter de la condition de Shannon et obtenir le bispectre sans erreurs en mettant à 0 les points contenus dans le triangle T_2 .

Cette remarque permet de lever une apparente contradiction dans les résultats donnés en 2.4.2. En effet, on sait que lorsque l'on respecte la condition de Shannon on peut reconstruire le signal sans erreur. L'échantillonnage selon Shannon ne fait donc pas perdre d'information sur le signal. Il peut donc sembler anormal d'être conduit à appliquer une condition plus forte pour l'échantillonnage. Cela n'est pas absolument nécessaire puisque les artefacts engendrés dans le calcul des multispectres peuvent être "effacés".

La situation est cependant différente dans le cas des signaux non stationnaires. Hinich [104] a montré que pour ces signaux il apparaît des valeurs non nulles du bispectre dans le triangle T_2 . Il a proposé sur cette base un test de stationnarité. Si le signal traité peut être non stationnaire, il faudra être prudent dans les opérations d'échantillonnage...

2.4.4 Multispectres des signaux à temps discret

Le signal stationnaire à temps continu $x(t)$ vérifiant une des conditions énoncées ci-dessus est échantillonné pour donner le signal à temps discret (numérique, digital) $x(n)$. Pour simplifier les notations on sous-entend la période d'échantillonnage, et l'instant nT_E est noté n . La multicorrélation de $x(n)$ est définie par¹⁰

$$C_{x(p)}(\mathbf{k}) = \text{Cum}[x(n), x(n + k_1), \dots, x(n + k_{p-1})].$$

où $\mathbf{k} = (k_1, \dots, k_{p-1})$.

¹⁰Nous nous limitons ici aux signaux à valeurs réelles. On peut également considérer des signaux à valeurs complexes.

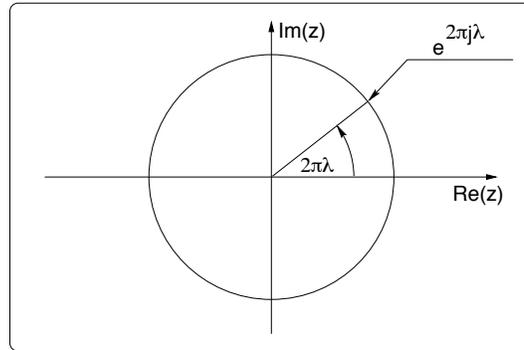


Figure 2.7: *Plan complexe des Z et fréquence réduite.* En restreignant Z au cercle unité, on retrouve la notion de fréquence dans la fréquence réduite λ , qui varie de $-1/2$ à $1/2$.

Pour passer en fréquence on sait qu'il existe, pour les signaux à temps discrets, deux types de transformations : la transformation en Z et la transformation en fréquences réduites [147].

La transformée en Z [126] de la fonction à temps discret $f(n)$ est la fonction de la variable complexe¹¹ z

$$F(z) = \sum_{n=-\infty}^{+\infty} f(n)z^{-n}.$$

Cette transformée est définie dans une couronne du plan complexe. Il existe une transformée inverse permettant de retrouver $f(n)$ à partir de $F(z)$.

La transformation en Z est surtout utile pour la description des systèmes linéaires. Pour la représentation des signaux, on souhaite retrouver une grandeur ayant le sens physique d'une fréquence. Ceci est réalisé par la transformation en fréquences réduites [147] qui n'est, somme toute, que la restriction de la transformée en Z au cercle C_1 de rayon 1 dans le plan complexe des Z .

En se plaçant sur le cercle C_1 dans le plan complexe des Z (figure (2.7)) on peut poser

$$z = e^{2\pi j\lambda}.$$

On obtient alors la transformée en fréquences réduites de $f(n)$, que nous noterons improprement $F(\lambda)$ pour simplifier les notations, définie par

$$F(\lambda) = \sum_{n=-\infty}^{+\infty} f(n)e^{-2\pi jn\lambda}.$$

¹¹Nous considérons ici la transformée en Z bilatérale dont l'usage est nécessaire pour représenter des signaux stationnaires.

La transformation inverse est donnée par

$$f(n) = \int_{-1/2}^{1/2} F(\lambda) e^{2\pi j n \lambda} d\lambda. \quad (2.15)$$

La similitude de cette relation avec l'expression de la transformée de Fourier inverse permet de retrouver dans λ la notion de fréquence. En écrivant

$$2\pi j n \lambda = 2\pi j (nT_E) \left(\frac{\lambda}{T_E} \right),$$

on voit apparaître le temps nT_E et la fréquence λ/T_E . L'intégrale apparaissant dans (2.15) est limitée, en fréquences réduites λ , au segment $[-1/2, +1/2]$, soit en fréquences à $[-1/2T_E, +1/2T_E]$, c'est-à-dire à la bande de Shannon. Les multispectres pourront donc être définis en Z ou en fréquences réduites.

Le multispectre en Z est la transformée en Z multidimensionnelle de la multicorrélation

$$S_{x(p)}(z_1, \dots, z_{p-1}) = \sum_{k_1} \dots \sum_{k_{p-1}} C_{x(p)}(k_1, \dots, k_{p-1}) z_1^{-k_1} \dots z_{p-1}^{-k_{p-1}}. \quad (2.16)$$

Le multispectre en fréquences réduites est la transformée multidimensionnelle en fréquences réduites de la multicorrélation

$$S_{x(p)}(\lambda_1, \dots, \lambda_{p-1}) = \sum_{k_1} \dots \sum_{k_{p-1}} C_{x(p)}(k_1, \dots, k_{p-1}) e^{-2\pi j k_1 \lambda_1} \dots e^{-2\pi j k_{p-1} \lambda_{p-1}}. \quad (2.17)$$

Pratiquement, nous y reviendrons plus loin, les multispectres sont calculés en utilisant la transformation de Fourier discrète, *TFD* [59].

2.4.5 Multispectres et composantes spectrales

Comme nous l'avons vu au paragraphe 2.2.3 pour les signaux à temps continu, on peut également définir la représentation de Cramér des signaux à temps discret par

$$x(n) = \int_{-1/2}^{1/2} e^{2\pi j \lambda n} dX(\lambda).$$

On en déduit le multispectre symétrique en fréquences réduites

$$\Sigma_{x(p)}(\lambda_0, \boldsymbol{\lambda}) d\lambda_0 d\boldsymbol{\lambda} = \text{Cum}[dX(\lambda_0), dX(\lambda_1), \dots, dX(\lambda_{p-1})]. \quad (2.18)$$

où $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_{p-1})$. Dans le cas de signaux stationnaires, ce multispectre symétrique est relié au multispectre par

$$\Sigma_{x(p)}(\lambda_0, \boldsymbol{\lambda}) = S_{x(p)}(\boldsymbol{\lambda}) \delta\left(\lambda_0 + \sum_{i=1}^{p-1} \lambda_i\right). \quad (2.19)$$

Dans cette relation la somme $\sum_{i=0}^{p-1} \lambda_i$ est prise modulo 1. Dans ce cas, la multiplicité stationnaire est formée d'une famille périodique d'hyperplans. Cette relation nous indique que, comme les composantes spectrales des signaux à temps continu, les composantes spectrales d'un signal à temps discret ont des propriétés de circularité. La question est ici un peu plus complexe par suite de la périodicité de la multiplicité stationnaire. Les composantes spectrales, à la fréquence réduite λ , sont circulaires au sens strict si $\forall p, p\lambda \neq 0 \pmod{1}$.

2.4.6 Filtrage linéaire à temps discret

Les relations entre l'entrée et la sortie d'un système linéaire, données en 2.3.1 pour les signaux à temps continu, s'étendent sans difficulté au cas du filtrage linéaire des signaux à temps discret. Nous allons les présenter dans le domaine des fréquences (fréquence réduite ou en Z) et dans le domaine du temps.

2.4.7 Relations de filtrage en fréquence

Relations de filtrage en fréquences réduites Les formules de filtrage se déduisent des formules données au paragraphe 2.3.1 en remplaçant les fréquences ν_i par les fréquences réduites λ_i . Dans le cas d'un filtre à une entrée $x(t)$ et une sortie $y(t)$, de gain complexe $A(\lambda)$, on a en fréquences réduites

$$dY(\lambda) = A(\lambda)dX(\lambda).$$

On obtient pour le spectre

$$S_{y(2)}(\lambda) = A(\lambda)A(-\lambda)S_{x(2)}(\lambda). \quad (2.20)$$

Or, pour les signaux à valeurs réelles le gain complexe du filtre vérifie $A(-\lambda) = A^*(\lambda)$, donnant la formule connue

$$S_{y(2)}(\lambda) = A(\lambda)A^*(\lambda)S_{x(2)}(\lambda) = |A(\lambda)|^2 S_{x(2)}(\lambda).$$

Pour le bispectre et pour des signaux à valeurs réelles la relation de filtrage est

$$S_{y(3)}(\lambda_1, \lambda_2) = A(\lambda_1)A(\lambda_2)A^*(\lambda_1 + \lambda_2)S_{x(3)}(\lambda_1, \lambda_2),$$

On voit, comme au paragraphe 2.3.1, que le bispectre contient des informations sur la phase du filtre.

Relations de filtrage en Z On peut déduire les formules de filtrage en Z des formules de filtrage en fréquences réduites en utilisant les deux remarques suivantes :

- le passage des variables fréquentielles aux variables z est une transformation homomorphique :

$$z = e^{2\pi j\lambda}.$$

Cette transformation applique le groupe additif dans le groupe multiplicatif.

- le fait de remplacer un signal par son complexe conjugué inverse le sens du temps : cette transformation applique l'intérieur du cercle unité sur l'extérieur de ce cercle en réalisant une inversion géométrique. En Z cette transformation revient à remplacer z par $1/z^*$.

Un signal non complexe conjugué, différent du premier (i.e. $x(n)$ dans la multicorrélation), filtré par le filtre de gain complexe $A(z)$ est donc multiplié par $A(z)$. Un signal complexe conjugué, filtré par $A(z)$, est multiplié par $A^*(1/z^*)$. Enfin le premier signal, filtré par $A(z)$ est multiplié par $A(\prod z_j / \prod z_i)$, les z_i étant les variables z des signaux non complexes conjugués et les z_j les variables z des signaux complexes conjugués. Cette dernière règle est issue du remplacement des additions et des soustractions par des produits et des divisions.

En appliquant ces règles, si $y(n)$ est issu du filtrage de $x(n)$ par le filtre de fonction de transfert $A(z)$, on obtient pour les spectres

$$S_{y(1)}^{(1)}(z) = A(z)A^*\left(\frac{1}{z^*}\right)S_{x(1)}^{(1)}(z)$$

et

$$S_{y(2)}^{(1)}(z) = A\left(\frac{1}{z}\right)A(z)S_{x(2)}^{(1)}(z).$$

Pour les bispectres, les relations entrée-sortie sont alors

$$\begin{aligned} S_{y(3)}^{(1)}(z_1, z_2) &= A\left(\frac{1}{z_1 z_2}\right)A(z_1)A(z_2)S_{x(3)}^{(1)}(z_1, z_2) \\ S_{y(2)}^{(1)}(z_1, z_2) &= A\left(\frac{z_2}{z_1}\right)A(z_1)A^*\left(\frac{1}{z_2^*}\right)S_{x(2)}^{(1)}(z_1, z_2) \\ S_{y(1)}^{(2)}(z_1, z_2) &= A(z_1 z_2)A^*\left(\frac{1}{z_1^*}\right)A^*\left(\frac{1}{z_2^*}\right)S_{x(1)}^{(2)}(z_1, z_2) \end{aligned}$$

2.4.8 Relations de filtrage en temps

Les relations de filtrage en temps discret sont identiques à celles que nous avons données pour les signaux à temps continu en remplaçant les convolutions à temps continu par des convolutions à temps discret.

Pour la corrélation de signaux réels filtrés par un filtre de réponse impulsionnelle $a(k)$

$$C_{y(2)}(k) = \sum_{l_0} \sum_{l_1} a(l_0)a(l_1)C_{x(2)}(k + l_0 - l_1),$$

et pour la bicorrélation

$$C_{y(3)}(k_1, k_2) = \sum_{l_0} \sum_{l_1} \sum_{l_2} a(l_0)a(l_1)a(l_2)C_{x(3)}(k_1 + l_0 - l_1, k_2 + l_0 - l_2).$$

Nous avons défini les multicorrélations et les multispectres pour les signaux aléatoires stationnaires à temps continu et à temps discret. Voyons comment on peut étendre ces notions au cas des signaux déterministes.

2.5 Ergodisme, signaux déterministes

Les définitions des multicorrélations et des multispectres pour les signaux aléatoires sont fondées sur les moyennes d'ensemble (moments et cumulants). Ces caractéristiques sont définies à partir des propriétés probabilistes du signal. La notion de moyenne d'ensemble n'a pas de sens pour les signaux certains. Nous pourrions cependant unifier la présentation des propriétés des signaux aléatoires et déterministes en introduisant la notion d'ergodisme.

2.5.1 Ergodisme

Généralités La mesure pratique (estimation) des moyennes d'ensemble est réalisée, en général, en utilisant des moyennes empiriques. Dans l'étude d'un système déterminé, les moyennes empiriques sont obtenues en utilisant un grand nombre de systèmes identiques et en calculant la moyenne sur cet ensemble de systèmes (d'où la dénomination de moyenne d'ensemble). Cette approche est réalisée par les statisticiens pour réaliser, par exemple, des sondages. La loi fait même une obligation aux instituts de sondage de publier la "taille de l'échantillon", nombre d'individus (systèmes) sondés pour estimer la moyenne empirique. La taille de l'échantillon est une mesure de la précision (fiabilité) du sondage.

Les expérimentateurs ne disposent pas toujours d'un nombre suffisant de systèmes identiques leur permettant d'estimer avec précision une moyenne d'ensemble. Par contre, il leur est souvent possible d'observer le système pendant un temps suffisamment long. On a alors pensé à remplacer les moyennes d'ensemble par des moyennes dans le temps : moyennes temporelles [32]. L'ergodisme régit les relations qui existent entre les moyennes d'ensemble et les moyennes temporelles. Pour un système stationnaire et ergodique, les moyennes d'ensemble sont égales aux moyennes temporelles correspondantes [29]. On a donc là un moyen de mesure des moyennes à partir de l'observation d'un seul système pendant un temps suffisamment long (infini à la limite) [147].

La notion d'ergodisme remonte à Boltzmann qui l'a utilisée dans la théorie cinétique des gaz. Un gaz est formé d'un grand nombre de particules identiques. Pour calculer sur ce gaz une moyenne d'ensemble il faudrait, à un

instant donné, mesurer les caractéristiques (vitesse par exemple) de chacune des particules et calculer la moyenne sur l'ensemble des particules. Le théorème ergodique [32] nous indique dans quelles conditions on peut remplacer cette moyenne d'ensemble par la moyenne de la même grandeur calculée en suivant la trajectoire d'une particule pendant un temps suffisamment long.

Des résultats ont été donnés dans [64] sur les moyennes d'ensemble et les moyennes temporelles de signaux. Ces résultats fondent l'ergodisme d'une classe importante de signaux. Ils nous permettront d'étendre les notions de multicorrélation et de multispectre. Notons également, nous y reviendrons au paragraphe 2.6, que ces résultats donnent des moyens pratiques de mesure (estimation) des multicorrélations et des multispectres.

Moyennes d'ensemble et moyennes temporelles Dans [64], on considère L signaux $x_1(t), \dots, x_L(t)$ obtenus par un filtrage de Volterra (chapitre 5) d'ordre éventuellement infini d'une suite $b(t-p)$ de variables aléatoires indépendantes (bruit blanc parfait). La composante l de ce signal s'écrit

$$x_l(t) = \mu_l(t) + \sum_{i=1}^{\infty} f_{i,l}[b(t)],$$

où

$$f_{i,l}[b(t)] = \int h_{i,l}(u_1, \dots, u_i) b(t-u_1) \dots b(t-u_i) du_1 \dots du_i,$$

est un filtre de Volterra homogène (chapitre 5) d'ordre i . La moyenne du signal d'entrée est supposée bornée en module

$$|\mu_l(t)| < \infty,$$

les moments du bruit blanc d'entrée sont également supposés bornés en module

$$E[b(t)^q] < \infty,$$

et les noyaux de Volterra $h_{i,l}(u_1, \dots, u_i)$ sont sommables en module

$$\int \dots \int |h_{i,l}(u_1, \dots, u_i)| du_1 \dots du_i < \infty.$$

Ce modèle est très général et recouvre un grand nombre de situations pratiques. Il généralise au cas des ordres supérieurs le modèle utilisé très souvent consistant à décrire un signal aléatoire comme la sortie d'un filtre linéaire et stationnaire excité par un bruit blanc. Dans ce modèle, on sépare les propriétés probabilistes du signal qui sont fixées par le bruit blanc (il suffit, pour l'analyse à l'ordre 2, d'un bruit blanc à l'ordre 2), de la "forme spectrale" du signal qui est fixée par le filtre.

Dans l'étude des statistiques d'ordre supérieur, les filtres de Volterra fixent les propriétés du signal aux divers ordres. Les conditions imposées au bruit d'entrée et aux filtres de Volterra assurent que tous les moments et les cumulants des signaux aléatoires ainsi construits sont bornés et que les multicorrélations vérifient des conditions d'intégrabilité. Cette intégrabilité des multicorrélations est une condition d'ergodisme [29, 36].

Dans la situation décrite ci-dessus, on a le résultat suivant pour les signaux aléatoires stationnaires :

$$\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x_1(t + \tau_1) \dots x_L(t + \tau_L) dt = E[x_1(t + \tau_1) \dots x_L(t + \tau_L)]. \quad (2.21)$$

Cette relation montre l'identité, asymptotique, entre les moyennes temporelles et les moyennes d'ensemble. Elle va nous permettre de redéfinir les multicorrélations et les multispectres à partir des moyennes temporelles. Cette nouvelle définition donnera un sens aux multicorrélations et aux multispectres des signaux déterministes. Avant de pouvoir donner ces définitions, nous devons définir les cumulants dans cette nouvelle approche.

Les cumulants La relation (2.21) définit les moments à partir des moyennes temporelles. Pour obtenir les cumulants des précautions doivent être prises. Posons le problème dans le cas simple de la corrélation

$$C_{x(2)}(\tau) = E[x(t)x(t + \tau)] - E[x(t)]E[x(t + \tau)] = C'_{x(2)}(\tau) - m_x^2,$$

$C'_{x(2)}(\tau)$ étant la corrélation des moments. Pour estimer $C_{x(2)}(\tau)$ par des moyennes temporelles, nous estimons la fonction de corrélation des moments par une moyenne temporelle en intégrant sur une durée T . En intégrant sur une durée T finie, nous obtenons une estimation qui tend vers la valeur exacte lorsque T tend vers l'infini. Nous représentons les valeurs approchées en coiffant le symbole de la valeur exacte d'un chapeau

$$\widehat{C}'_{x(2)}(\tau) = \frac{1}{T} \int_T x(t)x(t + \tau) dt$$

et

$$\widehat{m}_x = \frac{1}{T} \int_T x(t) dt,$$

cette procédure conduit, comme nous allons le voir à un problème.

Si la fonction de corrélation des moments et la moyenne sont calculées sur la même tranche de signal on aura

$$\int \widehat{C}_{x(2)}(\tau) d\tau = \frac{1}{T} \int_{\Delta} [x(t) - \widehat{m}_x][x(t + \tau) - \widehat{m}_x] dt d\tau,$$

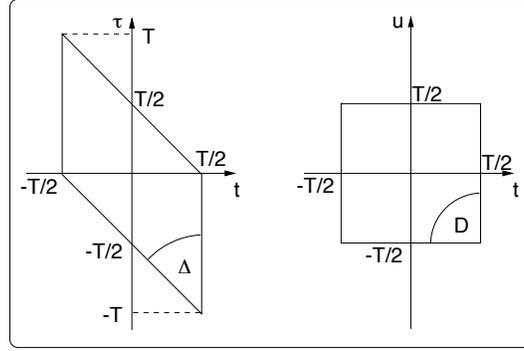


Figure 2.8: Les domaines d'intégration utilisés pour le calcul aboutissant au résultat (2.22).

où Δ est le domaine du plan t, τ précisé sur la figure (2.8); en posant $t + \tau = u$ et en utilisant les variables d'intégration t et u , il vient

$$\int \widehat{C}_{x(2)}(\tau) d\tau = \frac{1}{T} \int_D [x(t) - \widehat{m}_x][x(u) - \widehat{m}_x] dt du = 0,$$

D étant le domaine du plan t, u précisé sur la figure (2.8).

La fonction de corrélation estimée vérifie donc la relation

$$\int \widehat{C}_{x(2)}(\tau) d\tau = 0. \quad (2.22)$$

Ce résultat s'interprète en considérant la densité spectrale. On sait que la densité spectrale des moments d'un signal aléatoire non centré s'écrit

$$S'_{x(2)}(\nu) = S_{x(2)}(\nu) + m_x^2 \delta(\nu),$$

$S_{x(2)}(\nu)$ étant la densité spectrale du signal centré. A la fréquence 0, la densité spectrale du signal non centré (densité spectrale des moments) possède une singularité venant de la moyenne, et en général une composante non singulière $S_{x(2)}(0)$.

Notre objectif, en retranchant la moyenne, est de calculer $S_{x(2)}(\nu)$ et donc d'éliminer la singularité $m_x^2 \delta(\nu)$ en conservant la composante à la fréquence 0, $S_{x(2)}(0)$. La relation (2.22) nous montre que la procédure décrite plus haut conduit à

$$S_{x(2)}(0) = 0$$

et donc, élimine à la fois la singularité et la composante à la fréquence 0. Ce problème se rencontre dans le calcul des multicorrélations en général. Si l'on calcule les cumulants sans précaution on biaise la multicorrélation. On peut montrer que, comme pour la corrélation, les erreurs ainsi introduites

perturbent le multispectre associé sur des sous-ensembles de mesure nulle. Plus précisément, les perturbations se manifestent sur les multiplicités gaussiennes : ensembles de points sur lesquels le multispectre des moments est singulier (voir 2.2.4). Des solutions peuvent être apportées à ce problème : nous les présenterons dans la section 2.6 où seront abordées les techniques d'estimation des multicorrélations et des multispectres.

Dans le cas de la corrélation et de la bicorrélation, les moments et les cumulants sont identiques si le signal est centré. Il suffit donc de disposer d'une "bonne" méthode d'estimation de la moyenne.

2.5.2 Définition étendue des multicorrélations et des multispectres

Pour les signaux aléatoires stationnaires et ergodiques, on peut définir les cumulants à partir de moyennes temporelles. Ceci nous permet de proposer une nouvelle définition des multicorrélations à partir des moyennes temporelles¹². En appliquant la même définition aux signaux déterministes nous définirons les multicorrélations des signaux déterministes. Cette définition des multicorrélations s'applique sans problème aux signaux déterministes à puissance moyenne finie. Nous donnerons également la définition applicable aux signaux à énergie finie (signaux transitoires).

Signaux aléatoires et signaux périodiques Pour ces signaux nous définissons la multicorrélation d'ordre p par le cumulante d'ordre p calculé par une moyenne temporelle. En notant \widetilde{Cum} le cumulante calculé par des moyennes temporelles nous obtenons pour la multicorrélation

$$C_{x(p)}(\boldsymbol{\tau}) = \widetilde{Cum}[x(t), x(t + \tau_1), \dots, x(t + \tau_{p-1})].$$

La définition est donnée ici pour les signaux à valeurs réelles à temps continu. Elle s'étend très simplement aux signaux à valeurs complexes et aux signaux à temps discret.

Comme précédemment les multicorrélations sont les transformées de Fourier, ou les transformées en fréquences réduites, ou les transformées en Z , des multicorrélations.

Signaux transitoires Nous dirons qu'un signal est transitoire à l'ordre p si les intégrales

$$\int_{-\infty}^{+\infty} |x(t)|^s dt,$$

¹²C'est l'approche fonctionnelle introduite par Wiener [199] ou Brillinger [36] en parallèle de l'approche probabiliste, plus classique, utilisée auparavant.

pour $1 \leq s \leq p$ sont bornées. Pour ces signaux nous devons modifier la définition de la moyenne temporelle pour le calcul des moments. En effet, la valeur donnée par (2.21, page 63) est nulle : l'intégrale est bornée et T tend vers l'infini. Pour ces signaux nous définirons les moyennes temporelles par

$$\int_{-\infty}^{+\infty} x_1(t + \tau_1) \dots x_L(t + \tau_L) dt. \quad (2.23)$$

A partir de là on peut, comme précédemment, définir les multicorrélations et, par voie de conséquence, les multispectres. Concluons en considérant le cas des signaux composites.

Signaux composites Nous avons considéré trois types de signaux : aléatoires stationnaires (et ergodiques), déterministes périodiques et déterministes transitoires. Nous appellerons signal composite un signal formé de la somme de signaux de natures différentes. Ces signaux interviennent très souvent pour décrire la somme d'un signal utile et d'un bruit.

Comment calcule-t-on les multicorrélations et les multispectres dans cette situation?

Pour la somme d'un signal aléatoire et d'un signal périodique cela ne pose pas de problème : la définition par des moyennes temporelles est la même pour les deux types de signaux. Il suffit donc, dans ce cas-là, de les calculer par les moyennes temporelles.

Pour la somme d'un signal aléatoire (ou périodique) et d'un signal transitoire on procédera ainsi. Un signal transitoire, c'est pour cela que nous avons retenu cette dénomination, est non nul, ou approximativement non nul¹³, sur une durée finie T . On définira dans ce cas les moyennes temporelles selon la relation (2.23) en calculant les intégrales sur la durée T . On remplace donc, dans les moyennes temporelles l'opérateur $\lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2}$ par l'opérateur $\frac{1}{T} \int_{-T/2}^{T/2}$. En effet, si l'on conservait la définition des moyennes temporelles par l'opérateur $\lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2}$, les termes associés au signal transitoire seraient annulés par la division par $T \rightarrow \infty$.

Notons, en terminant, que cette façon de voir les choses n'a rien de très nouveau. C'est exactement ainsi que l'on procède à l'ordre 2 [147].

¹³Nous faisons référence ici aux signaux à durée approximativement limitée tels qu'ils ont été définis par [189].

2.6 Estimation des multicorrélations et des multispectres

L'estimation des multicorrélations et des multispectres est réalisée en utilisant des moyennes. Les estimateurs ainsi obtenus sont applicables aux différents types de signaux. Pour les signaux aléatoires ces estimateurs sont fondés sur les propriétés d'ergodisme sur lesquelles nous reviendrons.

Les propriétés des estimateurs dépendront du type de signal traité. Pour les signaux déterministes l'estimateur pourra présenter une erreur systématique (biais) mais ses valeurs ne seront pas dispersées (variance). Pour les signaux aléatoires la valeur moyenne de l'estimateur nous donnera le biais (éventuellement nul : estimateur non biaisé). La variance de l'estimateur nous donnera la dispersion des mesures. Nous verrons que, comme dans le cas de la mesure des spectres, l'estimation des multispectres fait apparaître un "antagonisme" entre le biais et la variance.

Les méthodes actuelles d'estimation utilisent des traitements numériques. Nous présenterons les estimateurs des multicorrélations et des multispectres pour des signaux à temps discret.

Revenons sur les propriétés d'ergodisme qui fondent la validité des estimateurs pour les signaux aléatoires stationnaires et ergodiques.

2.6.1 L'ergodisme revisité

Nous avons défini au paragraphe 2.5.1 une large classe de signaux aléatoires stationnaires et ergodiques. Ce sont les signaux obtenus par filtrage de Volterra d'un bruit blanc parfait. Nous avons également indiqué que, pour ces signaux, les cumulants sont bornés et les multicorrélations sont intégrables. Nous allons voir que ceci fonde la consistance des estimateurs obtenus par une moyenne temporelle. Un estimateur est consistant si sa valeur tend presque sûrement vers la vraie valeur lorsque la durée utilisée pour calculer l'estimateur tend vers l'infini.

Rappelons le résultat donné dans [64]. Pour des signaux aléatoires à temps discret de la classe définie en 2.5.1,

$$\lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{n=-N}^N x_1(n+k_1) \dots x_L(n+k_L) = E[x_1(n+k_1) \dots x_L(n+k_L)], \quad (2.24)$$

montre la convergence asymptotique des moyennes temporelles vers les moyennes d'ensemble.

Ce résultat donne la clef des méthodes d'estimation dans lesquelles on estime les moyennes d'ensemble par des moyennes temporelles. Notons que des résultats analogues sont déjà utilisés dans les estimateurs de grandeurs du second ordre. Pratiquement on ne peut pas réaliser le passage à la limite

($N \rightarrow \infty$). Les propriétés statistiques de l'estimateur obtenu pour N fini sont données par le résultat suivant [64].

La variable aléatoire

$$\frac{1}{2N+1} \sum_{n=-N}^N x_1(n+k_1) \dots x_L(n+k_L) - E[x_1(n+k_1) \dots x_L(n+k_L)]$$

tend vers une variable aléatoire gaussienne centrée dont la variance varie en $1/(2N)$ lorsque ($N \rightarrow \infty$).

Ces résultats théoriques fondent la validité des méthodes présentées ci-dessous. Ils nous permettront d'affirmer que les estimateurs donnés plus loin sont consistants. Dorénavant nous supposons que les signaux utilisés vérifient les conditions énoncées en 2.5.1.

2.6.2 Moments ou cumulants ?

Nous avons déjà présenté au paragraphe 2.2.4 les arguments plaidant en faveur des cumulants. L'intérêt de l'usage des cumulants apparaît également lorsque l'on veut mesurer les multicorrélations et les multispectres.

Pour les multicorrélations nous avons vu en 2.2.4 que si l'on calcule la multicorrélation des moments, il apparaît des "branches" s'étendant jusqu'à l'infini et situées sur les sous-ensembles, multiplicités gaussiennes, que nous avons précisés en 2.2.4. L'existence de ces branches infinies pose un problème pratique. En effet, les signaux observés sont de durée finie et, pour des questions de stabilité statistique de l'estimateur (variance), on doit limiter le support de la multicorrélation estimée. Cela conduit à des perturbations importantes des branches s'étendant à l'infini.

Pour les multispectres, si l'on calcule le multispectre des moments il apparaît – comme nous l'avons montré en 2.2.4 – des singularités dans certains sous-ensembles de l'espace des fréquences. Ces sous-ensembles étant de mesure nulle nous pourrions espérer que leur effet soit peu sensible. Ce n'est pas le cas car, comme dans l'analyse spectrale à l'ordre 2, la durée limitée du signal traité entraîne un effet de lissage qui "étale" les singularités. Il s'ensuit l'existence d'un biais important lorsque l'on calcule les multispectres des moments. En conclusion il nous semble raisonnable de préconiser l'utilisation des cumulants dans l'estimation des multicorrélations et des multispectres.

La question qui se pose alors est celle de la "bonne estimation" des cumulants. Nous avons montré en 2.5.1 les déboires qui peuvent accompagner l'estimation de la moyenne. Nous ne connaissons pas de règle générale. Nous allons donner quelques indications pour les multicorrélations et les multispectres jusqu'à l'ordre 4. Commençons par les ordres 2 et 3. Pour ces ordres les cumulants sont les moments du signal centré. Il suffit donc d'estimer la moyenne et de la retrancher (ce problème est bien connu en analyse spectrale classique). Mais il ne faut pas, nous l'avons montré en 2.5.1, calculer la

moyenne sur la même tranche de signal que celle qui est utilisée pour estimer la multicorrélation ou le multispectre. Lorsqu'on est amené à découper le signal en segments, pour diminuer la variance, une solution consiste à retrancher la moyenne estimée sur l'ensemble du signal. Une autre procédure s'appuie sur l'hypothèse suivante, généralement vérifiée. On admet que le signal traité est à mémoire finie. D'une part ceci est conforme à l'intuition¹⁴. D'autre part ceci est une condition nécessaire pour assurer un comportement satisfaisant des estimateurs. En effet, nous avons dit que nous devons, pour des raisons pratiques, limiter le support des multicorrélations estimées. Dans ce cas, la moyenne de l'estimation de la corrélation des moments est

$$C'_{x(2)}(k) = C_{x(2)}(k) + m_x^2;$$

ayant postulé que $C_{x(2)}(k)$ est nul pour $|k| \geq k_M$ on obtiendra la moyenne m pour des retards supérieurs à k_M .

Lorsque l'on veut estimer la tricorrélation il faut non seulement retrancher la moyenne mais également retrancher les termes issus de la corrélation qui apparaissent dans la multicorrélation des moments (voir 2.2.4). On peut alors, comme pour la moyenne, utiliser une estimation de la corrélation (avec précaution) et estimer les termes supplémentaires en se plaçant à l'extérieur du support présumé de la tricorrélation. On a également proposé [90] d'utiliser le trispectre des moments. Comme on connaît la localisation des singularités du trispectre des moments on remplace, en faisant une hypothèse de continuité, les valeurs obtenues au voisinage des zones de singularité par le prolongement des valeurs du trispectre obtenues en dehors des zones de singularité. On peut déterminer l'extension des zones qui doivent être neutralisées en fonction du lissage qui a été appliqué.

Nous avons voulu soulever ce point qui nous paraît important dans l'estimation des multicorrélations et des multispectres. Nous n'avons que très partiellement répondu à la question. Nous pensons qu'il y a là une des tâches importantes qui attendent les chercheurs du domaine.

2.6.3 Estimation des multicorrélations

Définition de l'estimateur Nous allons présenter les expressions pour le cas de la bicorrélation. Nous nous limitons aux signaux à valeurs réelles. La généralisation aux signaux à valeurs complexes ne pose pas de problème. Pour un signal $x(n)$, à valeurs réelles, centré¹⁵ connu à travers une réalisation de

¹⁴Cette analyse ne s'applique pas aux signaux "à longue dépendance" qui posent des problèmes particuliers.

¹⁵Cette hypothèse permet de remplacer le cumulants par le moment. Les problèmes posés par le centrage ont été vus plus haut.

durée N , on estime la bicorrélation par

$$\widehat{C}_{x(3)}(k_1, k_2) = K(N) \sum_{n \in D(N, k_1, k_2)} x(n)x(n+k_1)x(n+k_2).$$

$K(N)$ est un facteur de normalisation qui tient compte de la durée du signal ou du nombre de points. On choisira souvent :

$$K(N) = \frac{1}{N},$$

N étant le nombre de points du signal observé. Le domaine des valeurs parcourues par n , $D(N, k_1, k_2)$, dépend du nombre de points du signal, N et des valeurs des retards, k_1 et k_2 , pour lesquels on calcule la bicorrélation.

Propriétés de l'estimateur Comme nous l'avons indiqué en 2.6.1 et sous les conditions requises, l'estimateur est consistant asymptotiquement.

L'estimateur est asymptotiquement non biaisé si

$$N.K(N) \rightarrow 1 \quad \text{quand} \quad N \rightarrow \infty.$$

Pour obtenir un estimateur non biaisé il faut prendre

$$K(N) = \frac{1}{N - \text{Max}[k_1, k_2]}.$$

Asymptotiquement, la variance de l'estimateur de la bicorrélation à temps discret peut s'écrire¹⁶

$$\text{Var} \left[\widehat{C}_{x(3)}(k_1, k_2) \right] = \frac{1}{N} \sum_k \left[C_{x(2)}(k) \right]^3 \quad (2.25)$$

$$+ \frac{1}{N} \sum_k C_{x(6)}(k+k_1, k+k_2, k, k_1, k_2), \quad (2.26)$$

où $C_{x(2)}(k)$ est la corrélation et $C_{x(6)}(l_1, l_2, l_3, l_4, l_5)$ est la multicorrélation d'ordre 6. Les conditions énoncées en 2.6.1 assurent la convergence des sommes dans (2.26) lorsque N tend vers l'infini. L'estimateur est donc asymptotiquement efficace. Notons enfin que, par suite des propriétés de symétrie de la bicorrélation, on limite le calcul au domaine fondamental.

2.6.4 Estimation des multispectres

Pour estimer les multispectres on utilise le multipériodogramme [36, 37, 177].

¹⁶Ces résultats sont explicités au paragraphe 2.9.6.

Le multipériodogramme Soit une réalisation de longueur N du signal échantillonné $x(n)$, dont la représentation fréquentielle est donnée par la TFD ¹⁷

$$X(m) = \theta \sum_{n=0}^{N-1} x(n) e^{-\frac{2\pi j}{N} nm}$$

Le multipériodogramme d'ordre p est défini par

$$P_{x(p)}(m_1, \dots, m_{p-1}) = \beta X(m_0) X(m_1) \dots X(m_{p-1}).$$

avec

$$m_0 = - \sum_{i=1}^{p-1} m_i.$$

Le multipériodogramme $P_{x(p)}^C$ construit sur les cumulants est présenté au paragraphe 2.9.3, où il est montré que

$$\begin{aligned} P_{x(p)}^C(m_1, \dots, m_{p-1}) &= \beta \text{Cum}[X(m_0), X(m_1), \dots, X(m_{p-1})] \\ &= S_{x(p)}(m_1, \dots, m_{p-1}), \end{aligned}$$

où $S_{x(p)}(m_1, \dots, m_{p-1})$ est le multispectre en fréquences discrètes, défini par

$$\begin{aligned} S_{x(p)}(m_1, \dots, m_{p-1}) &= TFD[C_{x(p)}(k_1, \dots, k_{p-1})] \\ &= \theta^{p-1} \sum_{k_1, \dots, k_{p-1}} C_{x(p)}(k_1, \dots, k_{p-1}) e^{-\frac{2\pi j}{N} \sum_{i=1}^{p-1} k_i m_i}, \end{aligned}$$

Les paramètres θ et β permettent de donner aux grandeurs définies numériquement (multicorrélations, multispectres) leur sens physique, ils sont introduits en 2.9.1. Pour cela on doit interpréter θ comme la période d'échantillonnage et β comme le pas en fréquence¹⁸.

Le bipériodogramme¹⁹ est donné par

$$P_{x(3)}(m_1, m_2) = \beta X(-m_1 - m_2) X(m_1) X(m_2), \quad (2.27)$$

et sa variance est

$$\text{Var}[P_{x(3)}(m_1, m_2)] = E[|P_{x(3)}(m_1, m_2)|^2] - |E[P_{x(3)}(m_1, m_2)]|^2,$$

dont le terme dominant est (voir [105] et 2.9.3)

$$\text{Var}[P_{x(3)}(m_1, m_2)] = \frac{1}{\beta} S_{x(2)}(m_1) S_{x(2)}(m_2) S_{x(2)}(-m_1 - m_2), \quad (2.28)$$

¹⁷Les paramètres θ et β sont précisés dans 2.9.1.

¹⁸On peut ainsi vérifier l'homogénéité des multipériodogrammes et des multispectres.

¹⁹Les composantes spectrales étant centrées, on peut utiliser directement le bipériodogramme.

$S_{x(2)}(m)$ étant le spectre. On a montré également ([105] et 2.9.3) que le terme dominant de la covariance de deux éléments distincts du bipériodogramme est

$$\text{Cov}[P_{x(3)}(m_1, m_2), P_{x(3)}(l_1, l_2)] = \beta S_{x(6)}(m_1, m_2, -m_1 - m_2, l_1, l_2), \quad (2.29)$$

$S_{x(6)}$ étant le multispectre d'ordre 6. Ces résultats permettent de calculer la variance de la bicorrélation que nous avons donnée à la section précédente (2.6.3). Ils nous montrent également que le bipériodogramme n'est pas un estimateur asymptotiquement efficace du bispectre.

Pour obtenir un estimateur asymptotiquement efficace du bispectre nous devons, comme dans l'estimation du spectre, mettre en œuvre une opération de moyenne permettant de faire tendre la variance vers 0 quand le nombre N de points du signal tend vers l'infini. Cette opération de moyenne peut être réalisée de différentes façons. En moyennant en fréquence on obtient le bipériodogramme lissé ou, de manière équivalente, le bicorrélogramme [28]. En moyennant en temps on a le bipériodogramme moyenné [20].

Méthode du bipériodogramme lissé ou du bicorrélogramme Ces deux méthodes d'estimation du bispectre réalisent une moyenne du bipériodogramme

- par lissage dans l'espace des fréquences : c'est la méthode du bipériodogramme lissé,
- de manière équivalente, dans la méthode du bicorrélogramme, on utilise la *TFD* de la bicorrélation estimée apodisée.

L'estimateur par la méthode du *bipériodogramme lissé* est

$$\widehat{S}_{x(3)}(m_1, m_2) = C \sum_{l_1} \sum_{l_2} A(m_1 - l_1, m_2 - l_2) P_{x(3)}(l_1, l_2),$$

C étant une constante de calibration. Le support de la fonction de lissage $A(., .)$ fixe l'étendue de la zone du plan des fréquences sur laquelle la moyenne est faite.

On obtient l'estimateur par la méthode du *bicorrélogramme* en prenant la *TFD* de l'estimation de la fonction de bicorrélation apodisée

$$\widehat{S}_{x(3)}(m_1, m_2) = C.TFD[b(k_1, k_2)\widehat{C}_{x(3)}(k_1, k_2)].$$

La moyenne du bipériodogramme lissé est

$$E[\widehat{S}_{x(3)}(m_1, m_2)] = C \sum_{l_1} \sum_{l_2} A(m_1 - l_1, m_2 - l_2) S_{x(3)}(l_1, l_2).$$

Si l'estimateur de la bicorrélation n'est pas biaisé, la moyenne du bicorrélogramme est

$$\begin{aligned} E[\widehat{S}_{x(3)}(m_1, m_2)] &= C.TFD[b(k_1, k_2)C_{x(3)}(k_1, k_2)] \\ &= C\beta^2 \sum_{l_1} \sum_{l_2} B(m_1 - l_1, m_2 - l_2)S_{x(3)}(l_1, l_2), \end{aligned}$$

avec $B(m_1, m_2) = TFD[b(k_1, k_2)]$. Les estimateurs du bispectre par la méthode du bicorrélogramme et du bipériodogramme lissé sont *biaisés*. Ces deux estimateurs ont des propriétés analogues si

$$A(m_1, m_2) = \beta^2 TFD[b(k_1, k_2)] = \beta^2 B(m_1, m_2)$$

Si le bispectre varie peu sur le support de la fonction de lissage ou, ce qui revient au même, si la bicorrélation a un support petit vis-à-vis du support de la fonction d'apodisation, les deux estimateurs sont approximativement non biaisés si on fixe la constante de calibration à

$$C = \frac{1}{b(0, 0)} = \frac{1}{\sum_{l_1} \sum_{l_2} A(l_1, l_2)} = \frac{1}{\beta^2 \sum_{l_1} \sum_{l_2} B(l_1, l_2)}.$$

La *variance des estimateurs* dépend de l'aire de la fonction de lissage dans le plan des fréquences. En définissant l'aire de la fonction de lissage, normalisée par β^2 , par

$$A_F = \frac{[\sum_{l_1} \sum_{l_2} A(l_1, l_2)]^2}{\sum_{l_1} \sum_{l_2} A^2(l_1, l_2)},$$

et en utilisant la variance et la covariance du bipériodogramme données en (2.28) et (2.29), on obtient pour le terme dominant de la variance des estimateurs [105]

$$\text{Var}[\widehat{S}_{x(3)}(m_1, m_2)] = \frac{1}{\beta A_F} S_{x(2)}(m_1) S_{x(2)}(m_2) S_{x(2)}(-m_1 - m_2). \quad (2.30)$$

Si la fonction de lissage est symétrique en m_1 et m_2 , en introduisant la largeur N_F de la fonction de lissage par

$$A_F = N_F^2,$$

la variance des estimateurs s'écrit

$$\text{Var}[\widehat{S}_{x(3)}(m_1, m_2)] = \frac{1}{\beta N_F^2} S_{x(2)}(m_1) S_{x(2)}(m_2) S_{x(2)}(-m_1 - m_2). \quad (2.31)$$

On peut réécrire les relations donnant la variance en introduisant une variance relative, sans dimension

$$\frac{\beta \text{Var}[\widehat{S}_{x(3)}(m_1, m_2)]}{S_{x(2)}(m_1) S_{x(2)}(m_2) S_{x(2)}(-m_1 - m_2)} = \frac{1}{A_F}.$$

Un choix naturel pour θ et β permettant de retrouver, lorsque le nombre de points N tend vers l'infini, la transformation en fréquences réduites est [36]

$$\theta = 1 \quad \beta = \frac{1}{N},$$

la variance relative est alors

$$\frac{\text{Var}[\widehat{S}_{x(3)}(m_1, m_2)]}{S_{x(2)}(m_1)S_{x(2)}(m_2)S_{x(2)}(-m_1 - m_2)} = \frac{N}{A_F}.$$

Les estimateurs du bipériodogramme ou du bicorrélogramme lissé sont alors asymptotiquement efficaces si lorsque $N \rightarrow \infty$, on a

$$\frac{N}{A_F} \rightarrow 0 \iff \frac{N}{N_F^2} \rightarrow 0.$$

Ces résultats sur la variance mettent en évidence "l'antagonisme" biais-variance. Le biais est issu de la moyenne faite sur le support de la fonction de lissage. Pour le minimiser on doit donner à la fonction de lissage un support le plus étroit possible. Pour diminuer la variance, comme le montrent (2.30) et (2.31), il faut, au contraire, augmenter l'aire du support de la fonction de corrélation.

Notons que les estimateurs présentés ici sont asymptotiquement gaussiens [158]. Dans ces conditions, la connaissance du biais et de la variance donne une description complète des propriétés de ces estimateurs.

Comme pour l'analyse spectrale, un des éléments importants dans l'estimation des multispectres est constitué par la fonction d'apodisation $a(k_1, k_2)$, qui doit vérifier les propriétés suivantes [178]

- $a(k_1, k_2)$ est à support limité,
- éventuellement $a(0,0) = 1$ pour les problèmes de calibration,
- $a(k_1, k_2)$ a les mêmes symétries que la bicorrélation,
- $A(m_1, m_2) = TFD[a(k_1, k_2)] \geq 0 \quad \forall m_1 \text{ et } m_2$

Une famille de fonctions possibles, proposée dans [162], est

$$a(k_1, k_2) = d(k_1)d(k_2)d(k_1 - k_2),$$

où

$$\begin{aligned} d(k_1) &= d(-k_1), \\ d(k_1) &= 0 \text{ pour } k_1 > L, \\ d(0) &= 1, \\ D(m) &= TFD[d(k_1)] \geq 0 \quad \forall m. \end{aligned}$$

On trouve dans la littérature diverses fonctions d'apodisation [178]. La fonction d'apodisation dite "optimale" (minimisation du maximum du biais) est définie par

$$d_0(k_1) = \frac{1}{\pi} \left| \sin \left(\frac{\pi k_1}{L} \right) \right| + \left(1 - \frac{|k_1|}{L} \right) \cos \left(\frac{\pi k_1}{L} \right),$$

avec $|k_1| \leq L$, et $d_0(k_1) = 0$ pour $|k_1| > L$. La fonction d'apodisation de Parzen est donnée par

$$\begin{cases} d_p(k_1) = 1 - 6 \left(\frac{|k_1|}{L} \right)^2 + 6 \left(\frac{|k_1|}{L} \right)^3 & \text{pour } |k_1| \leq \frac{L}{2} \\ d_p(k_1) = 2 \left(1 - \frac{|k_1|}{L} \right)^3 & \text{pour } \frac{L}{2} < |k_1| \leq L \\ d_p(k_1) = 0 & \text{pour } |k_1| > L. \end{cases}$$

Enfin, la fonction d'apodisation uniforme en fréquence est distribuée uniformément dans le domaine hexagonal du bispectre.

L'étude des fonctions d'apodisation applicables à la multicorrélation pour l'estimation du multispectre est encore à l'état naissant. On peut espérer que l'imagination des chercheurs permettra, comme pour l'analyse spectrale, de développer un catalogue de fonctions d'apodisation plus fourni et mieux renseigné que celui que nous pouvons proposer actuellement.

Méthode du bipériodogramme moyenné Dans la méthode du bipériodogramme moyenné, on réalise la moyenne en temps. Considérons une réalisation de longueur M du signal analysé $x(n)$. On découpe le signal observé en L segments de longueur N . Ces segments sont en général disjoints et contigus. On peut utiliser, comme pour l'analyse spectrale [146], des segments se recouvrant partiellement. On choisira souvent la longueur des segments, ou le décalage entre les segments, de manière à ce que les signaux appartenant à deux segments différents soient indépendants. On calcule ensuite, sur chaque segment, la TFD du signal apodisé. On en déduit le bipériodogramme de chaque segment, et l'on moyenne les bipériodogrammes obtenus sur chacun des segments. L'estimateur du bispectre par la méthode du bipériodogramme moyenné est donc

$$\hat{S}_{x(3)}(m_1, m_2) = \frac{C\beta}{L} \sum_{l=0}^{L-1} X_l(m_1)X_l(m_2)X_l(-m_1 - m_2), \quad (2.32)$$

avec

$$X_l(m) = TFD[h(n)x(n - lP)] = \theta \sum_{n=0}^{N-1} h(n)x(n - lP)e^{-\frac{2\pi j}{N}nm}. \quad (2.33)$$

Dans ces relations, C est une constante de calibration, $h(n)$ est la fonction d'apodisation des segments, nulle à l'extérieur de $[0, N[$, et P est le décalage entre les segments $0 < P \leq N$.

La moyenne du bipériodogramme moyenné est, par suite de la stationnarité

$$E \left[\widehat{S}_{x(3)}(m_1, m_2) \right] = C\beta E [X_0(m_1)X_0(m_2)X_0(-m_1 - m_2)].$$

En remplaçant les termes $X(m)$ par leur expression en fonction des termes $x(n)$ et en introduisant la bicorrélation de la fonction d'apodisation $h(n)$

$$C_{h(3)}(k_1, k_2) = \frac{1}{N} \sum_n h(n)h(n+k_1)h(n+k_2),$$

on obtient

$$E \left[\widehat{S}_{x(3)}(m_1, m_2) \right] = C.TFD[C_{h(3)}(k_1, k_2)C_{x(3)}(k_1, k_2)],$$

où $C_{x(3)}(k_1, k_2)$ est la bicorrélation de $x(n)$. Cette expression s'écrit aussi

$$E \left[\widehat{S}_{x(3)}(m_1, m_2) \right] = C\beta^2 \sum_{l_1} \sum_{l_2} S_{h(3)}(l_1, l_2)S_{x(3)}(m_1 - l_1, m_2 - l_2),$$

$S_{h(3)}(m_1, m_2)$ étant le bispectre de $h(n)$ et $S_{x(3)}(m_1, m_2)$ le bispectre de $x(n)$.

L'estimateur du bispectre par la méthode du bipériodogramme moyenné est *biaisé*. La bicorrélation de $h(n)$ réalise une apodisation dans le domaine des retards, le bispectre de $h(n)$ effectue un lissage dans le domaine des fréquences. Le comportement de la moyenne du bipériodogramme moyenné est analogue à celui de la moyenne du bipériodogramme lissé.

L'estimateur est *approximativement non biaisé* si le support de la bicorrélation de $h(n)$ est grand, vis-à-vis du support de la bicorrélation de $x(n)$ ou, ce qui revient au même, si le support du bispectre de $h(n)$ est petit par rapport au support du bispectre de $x(n)$. Dans ces conditions on obtient un estimateur approximativement non biaisé en fixant la constante de normalisation C par

$$C = \frac{1}{C_{h(3)}(0, 0)} = \frac{1}{\beta^2 \sum_{l_1, l_2} S_{h(3)}(l_1, l_2)}.$$

En supposant que les signaux appartenant à des segments différents sont indépendants et en utilisant la variance du bipériodogramme donnée par (2.28), la variance relative du bipériodogramme moyenné est

$$\frac{\beta \text{Var} \left[\widehat{S}_{x(3)}(m_1, m_2) \right]}{S_{x(2)}(m_1)S_{x(2)}(m_2)S_{x(2)}(-m_1 - m_2)} = \frac{\left[C_{h(2)}(0) \right]^3}{L \left[C_{h(3)}(0, 0) \right]^2},$$

où $C_{h(2)}(k) = \frac{1}{N} \sum_n h(n)h(n+k)$ est la corrélation de la fonction d'apodisation $h(n)$, et $C_{h(3)}(k_1, k_2) = \frac{1}{N} \sum_n h(n)h(n+k_1)h(n+k_2)$ est la bicorrélation de $h(n)$.

Pour $\beta = \frac{1}{N}$ et pour une apodisation constante, $h(n) = 1$, la variance relative de l'estimateur de la bicorrélation est

$$\frac{\text{Var} \left[\widehat{S}_{x(3)}(m_1, m_2) \right]}{S_{x(2)}(m_1) S_{x(2)}(m_2) S_{x(2)}(-m_1 - m_2)} = \frac{N}{L},$$

N étant le nombre de points sur chaque segment et L le nombre de segments utilisés pour réaliser la moyenne.

Lorsque l'on utilise des segments adjacents, si le nombre total de points du signal traité est $M = LN$, la variance relative est M/L^2 . L'estimateur est alors asymptotiquement *efficace* si, lorsque $M \rightarrow \infty$ on a

$$\frac{M}{L^2} \rightarrow 0.$$

Ces estimateurs, comme les estimateurs obtenus par la méthode du bicorrélogramme, sont asymptotiquement gaussiens [158].

Comparaison des deux estimateurs du bispectre La comparaison des deux estimateurs du bispectre (bipériodogramme lissé ou bipériodogramme moyenné) peut se faire en comparant les problèmes pratiques posés par leur mise en œuvre et leurs propriétés : moyenne et variance.

Considérons un signal formé de M échantillons dont nous voulons estimer le bispectre. La mise en œuvre du bipériodogramme lissé nécessite le calcul du bipériodogramme sur les M points. Cette opération est coûteuse en mémoire et en temps de calcul puisqu'elle fait appel à une *TFD* calculée sur un grand nombre de points. La mise en œuvre du bipériodogramme moyenné est moins exigeante puisque l'on se limite au calcul de la *TFD* sur des segments de longueur nettement plus courte.

Les propriétés statistiques des deux estimateurs par la méthode du bipériodogramme lissé, ou bicorrélogramme, et par la méthode du bipériodogramme moyenné sont équivalentes. Pour rendre le biais des deux méthodes équivalent il faut que l'aire, en bifréquence, des fonctions d'apodisation soit équivalente, soit

$$N_F = L.$$

Dans ces conditions les variances des deux estimateurs sont équivalentes.

2.7 Illustration

Illustrons les techniques d'estimation de la bicorrélation et du bispectre en les appliquant à des signaux certains et à des signaux aléatoires. Nous donnons sur les figures suivantes la bicorrélation et le module du bispectre en perspective, la bicorrélation, le module du bispectre et la phase du bispectre vus de dessus, les amplitudes étant codées en niveaux de gris.

Pour les signaux déterministes nous avons appliqué la technique du bipériodogramme. Dans cette situation il n'est pas nécessaire de moyenner.

La figure (2.9) présente la bicorrélation et le bispectre de sinus cardinaux, l'un de bande $B < \nu_E/3$ (figure supérieure) et l'autre de bande $B > \nu_E/3$ (figure inférieure). Sur la figure supérieure, le signal est dans la bande $[0 - 0, 25]$ en fréquences réduites. Ce signal vérifie la condition "forte" du théorème d'échantillonnage appliqué au bispectre (voir 2.4.2). On n'a donc pas d'artefact. La forme hexagonale du bispectre matérialise le domaine couvert par le bispectre de signaux à bande limitée. Sur la figure inférieure, le signal est dans la bande $[0 - 0, 4]$ en fréquences réduites. On voit dans cette situation les artefacts dus à l'échantillonnage.

La figure (2.10, page 80) montre la bicorrélation et le bispectre de la réponse impulsionnelle d'un filtre autorégressif d'ordre 2 (AR-2). Les pôles de ce filtre, qui fixent les résonances, ont un module égal à 0,85 et une phase de $\pm 60^\circ$. La fréquence réduite de résonance de ce filtre est 0,165. On voit sur le module du bispectre les maximums associés aux résonances du filtre.

La figure (2.11, page 81) présente la bicorrélation et le bispectre de la réponse impulsionnelle d'un filtre à moyenne mobile d'ordre 2 (MA-2). Le zéro du premier filtre (figure supérieure : filtre à phase minimale) est l'inverse par rapport au cercle unité du zéro du second filtre (figure inférieure : filtre à phase non minimale). Les phases de ces zéros sont $\pm 120^\circ$ et les modules, respectivement 0,85 et 1,176. On voit que les modules des deux bispectres sont identiques alors que leurs phases présentent une inversion. Ce phénomène est important car il permet d'identifier le filtre MA à partir de sa seule sortie (chapitre 3).

Nous présentons ensuite le traitement de signaux aléatoires. Nous avons utilisé la méthode du corrélogramme. Les signaux traités contiennent 2048 échantillons. La fonction d'apodisation de la bicorrélation est la bicorrélation d'une porte. Sur la figure (2.12, page 82), nous donnons la bicorrélation et le bispectre estimés d'un signal aléatoire obtenu en filtrant un bruit blanc par le filtre AR-2 introduit précédemment (pôles 0,85, phase $\pm 60^\circ$). Le bruit blanc d'entrée utilisé est construit en prenant le module centré d'un bruit blanc gaussien. Sur la figure du haut, l'apodisation est forte (porte de longueur 8 échantillons). On retrouve certains maximums associés aux résonances mais on constate un fort lissage accompagné d'une réduction des fluctuations statistiques. Pour obtenir le panneau du bas de la figure (2.12), on a utilisé une

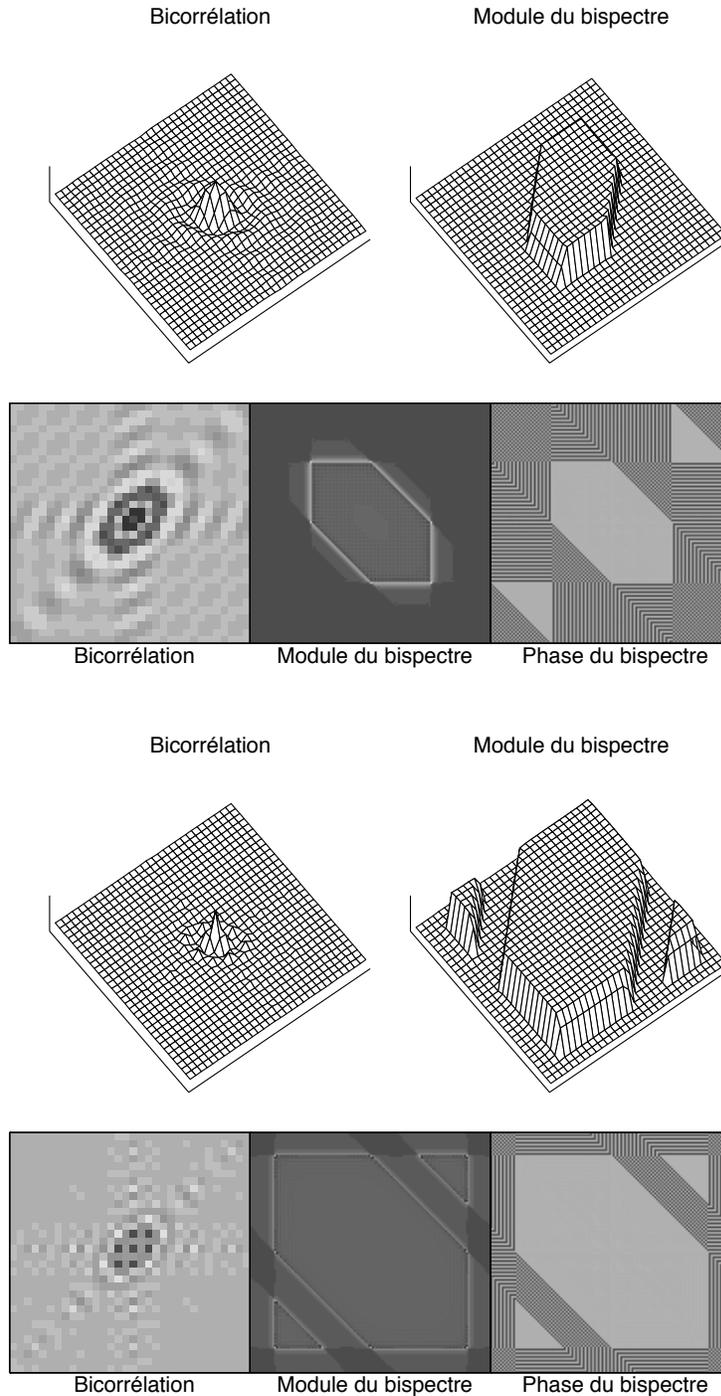


Figure 2.9: *Signal à bande limitée. La figure du haut montre la bicorrélation et le bispectre d'un sinus cardinal correctement échantillonné. La figure du bas correspond à l'analyse d'un sinus cardinal de bande $B > \nu_E/3$. Les repliements apparaissent dans le triangle T_2 , mais ne perturbent pas le calcul du bispectre dans l'hexagone principal.*

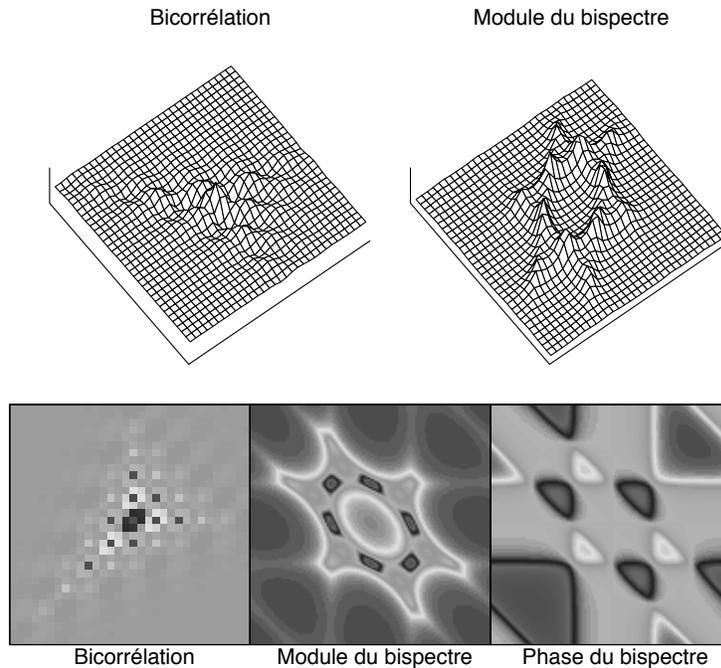


Figure 2.10: *Bispectre et bicorrélation d'un filtre autorégressif d'ordre 2.*

porte de longueur 32 échantillons. Le tracé est plus précis mais, en particulier sur la phase, on voit apparaître des fluctuations statistiques.

Nous présentons enfin, figure (2.13, page 83), la bicorrélation et le bispectre estimés du même bruit blanc filtré par les filtres MA à phase minimale et à phase non minimale utilisés précédemment. Ici, comme le bispectre est lisse nous avons utilisé l'apodisation forte (porte de 8 échantillons). On note que l'inversion de phase du bispectre, significative de la phase du filtre, est bien visible.

Comme cela est généralement fait, nous nous sommes limités à l'estimation du bispectre. Il est important de garder présent à l'esprit le fait que dans de nombreuses situations, la densité de probabilité du signal sera paire. Dans ce cas les cumulants d'ordre 3 sont nuls, la bicorrélation et le bispectre sont alors nuls. Dans ces conditions, les premières caractéristiques d'ordre supérieur à 2 sont la tricorrélation et le trispectre. Il sera alors nécessaire d'estimer la tricorrélation et le trispectre.

Les techniques que nous avons présentées pour la bicorrélation sont applicables à la tricorrélation. Il apparaît cependant deux difficultés d'ordre pratique et théorique. La difficulté pratique vient de la représentation d'une fonction à trois variables. On se contente souvent de présenter des coupes à deux dimensions ou des lignes à une dimension de la tricorrélation. Nous reviendrons sur ce point au chapitre 3. La difficulté principale est entraînée par le calcul des

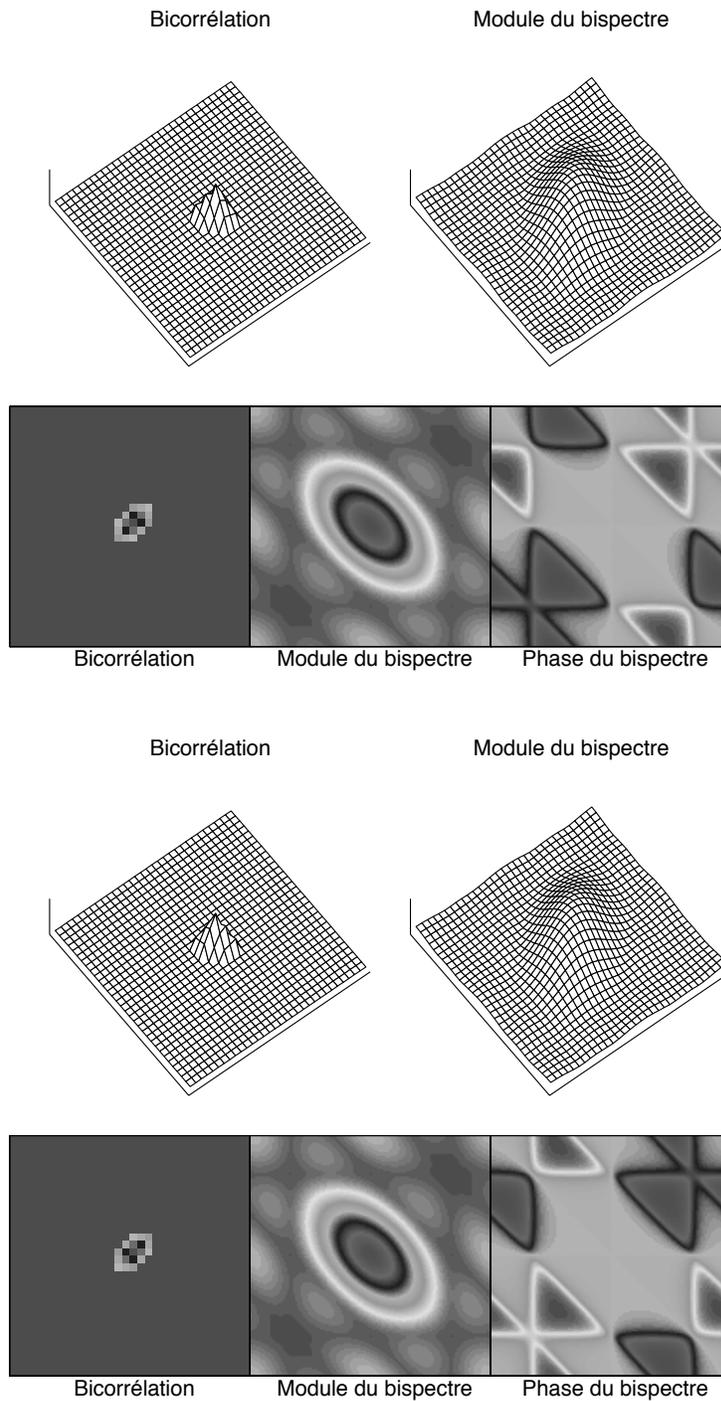


Figure 2.11: *Bispéctre et bicorrélation d'un filtre MA à phase minimale (figure du haut) et d'un filtre MA à phase non minimale (figure du bas). Les phases des bispectres sont inversées : le bispectre est sensible à la phase d'un système.*

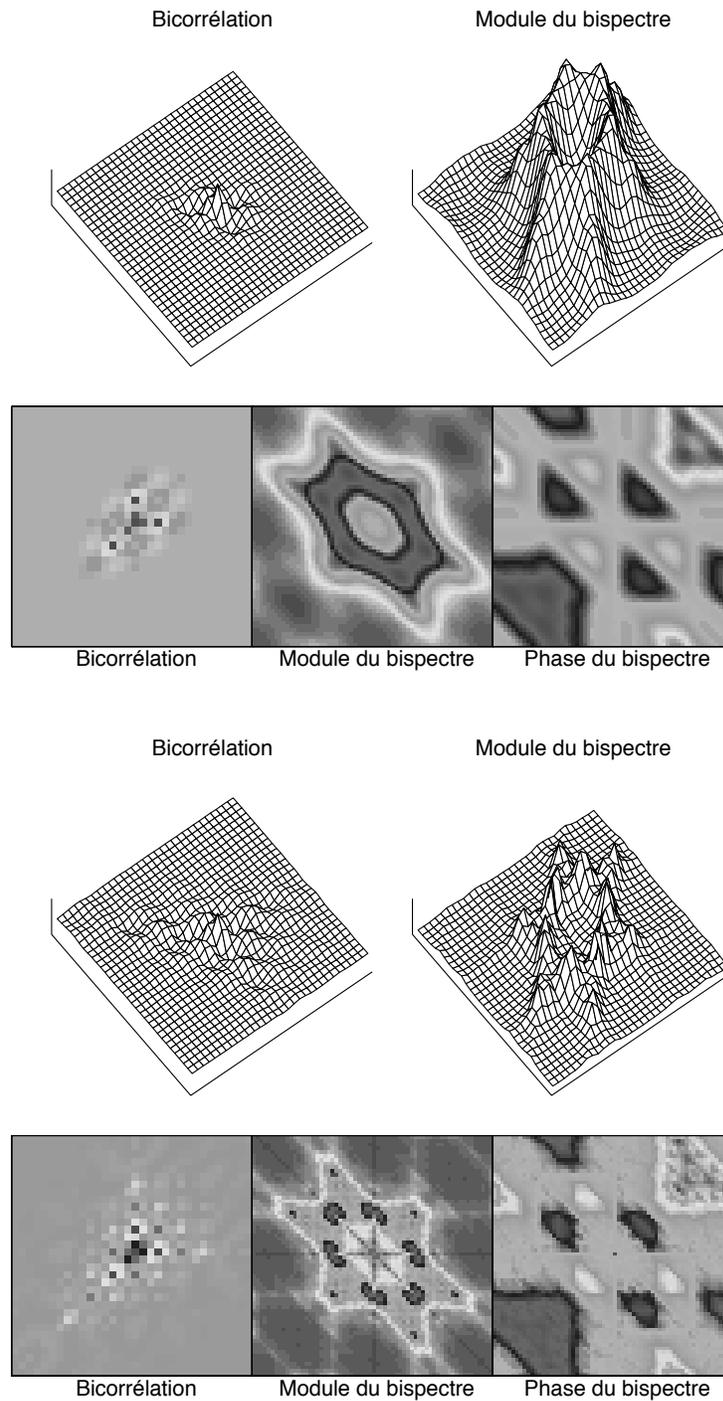


Figure 2.12: Estimation de la bicorrélation et du bispectre d'un signal aléatoire autorégressif. Sur la figure du haut, l'apodisation est de durée courte, alors que la figure du bas est obtenue avec une apodisation de durée plus longue.

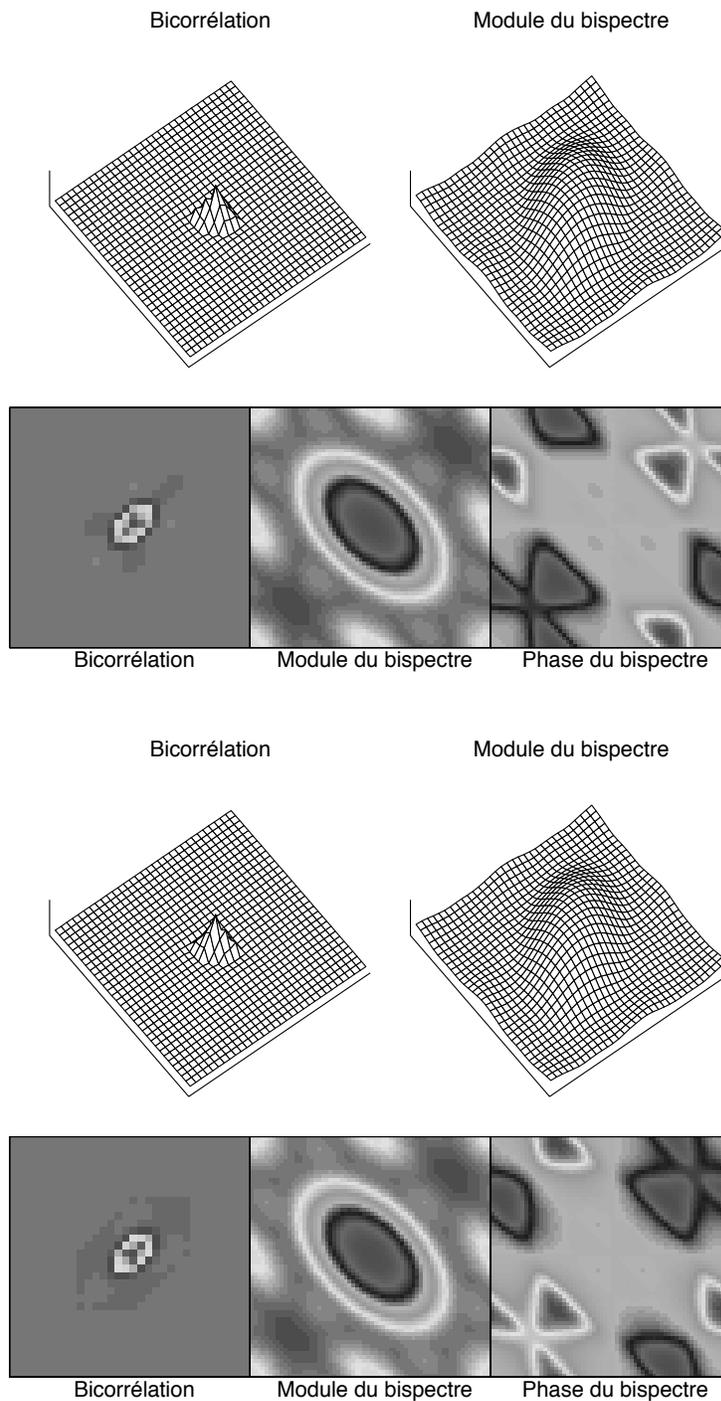


Figure 2.13: Estimation de la bicorrélation et du bispectre d'un signal aléatoire MA, à phase minimale sur la figure supérieure, à phase non minimale sur la figure inférieure. L'apodisation est de durée courte. L'inversion de phase est visible.

cumulants : les cumulants d'ordre 4 sont généralement obtenus en retranchant aux moments d'ordre 4 les produits de moments d'ordres inférieurs. Comme nous l'avons montré, il faut être vigilant si on utilise des moyennes temporelles.

Pour estimer le trispectre il existe, comme à l'ordre 3, deux techniques principales : le tripériodogramme et le tricorrélogramme. Le tricorrélogramme est obtenu par une transformation de Fourier (pratiquement une *TFD*) de la tricorrélation apodisée pour diminuer la variance de l'estimateur. Il reste du travail à faire sur les fonctions d'apodisation tridimensionnelles... Dans la technique du tripériodogramme on retrouve le problème de l'extraction des cumulants d'ordre 2. Là aussi il est encore trop tôt pour présenter un bilan. Enfin, dans tous les cas on retrouve les problèmes de représentation. On peut faire appel à des coupes ou à des lignes.

2.8 Conclusion

Les multicorrélations et les multispectres apparaissent comme une généralisation très naturelle de la corrélation et du spectre. L'analogie entre les propriétés des caractéristiques d'ordre 2 et celles d'ordre supérieur à 2 est très grande dans la définition des grandeurs fondamentales, dans les propriétés de filtrage et dans les méthodes d'estimation.

Un point sensible, que fait apparaître l'étude des grandeurs d'ordre supérieur, est le choix ouvert entre les moments et les cumulants. Nous avons apporté en 2.2.4 et 2.6.2 des éléments qui tranchent en faveur des cumulants.

Les multicorrélations et les multispectres, fonctions de plusieurs variables, apparaissent comme des grandeurs difficiles à calculer et à représenter. Cette complexité explique le fait que les études actuelles se soient cantonnées à la bicorrélation et au bispectre qui ne dépendent "que" de deux variables... Nous pensons que sur ce point les grandeurs vraiment intéressantes seront souvent d'ordre supérieur à 3. Dans ces conditions, comme nous le verrons dans les chapitres ultérieurs, on concentrera l'information en utilisant des projections ou des lignes de multicorrélation et de multispectre.

Les multicorrélations et les multispectres apportent de nouvelles informations sur les signaux non gaussiens. Ils donnent ainsi de nouveaux outils d'analyse et de traitement. Nous verrons dans les chapitres suivants que ces descriptions ouvrent la voie à de nouvelles méthodologies et permettent de réaliser des traitements dans des situations où les descriptions classiques à l'ordre 2 ne sont pas suffisamment précises.

2.9 Compléments

2.9.1 Note sur la TFD

Les TFD utilisées dans les statistiques d'ordre supérieur sont en général à plusieurs dimensions. Elles constituent une généralisation évidente de la TFD à une dimension. Nous avons introduit dans la TFD et dans la TFD^{-1} des facteurs de calibration qui, à notre avis, facilitent les calculs. Présentons rapidement cette normalisation de la TFD .

La TFD fait passer de la suite $x(n)$ de longueur N ($0 \leq n \leq N-1$) des échantillons temporels à la suite de longueur N ($0 \leq m \leq N-1$) des échantillons fréquentiels $X(m)$ par

$$X(m) = \theta \sum_{n=0}^{N-1} x(n) e^{-\frac{2\pi jnm}{N}},$$

la TFD inverse est donnée par

$$x(n) = \beta \sum_{m=0}^{N-1} X(m) e^{\frac{2\pi jnm}{N}},$$

avec

$$N\theta\beta = 1.$$

L'introduction des facteurs θ et β permet d'englober dans une même notation les différents choix qui peuvent être faits. En général on choisit

$$\theta = 1 \quad \text{donnant} \quad \beta = \frac{1}{N},$$

ou

$$\theta = \frac{1}{N} \quad \text{donnant} \quad \beta = 1.$$

Cette notation permet également de redonner à la TFD et à la TFD^{-1} la même dimension que celle de la TF ou de la TF^{-1} . Pour cela, si T_E est la période d'échantillonnage, il faut choisir,

$$\theta = T_E \quad \text{donnant} \quad \beta = \frac{1}{NT_E},$$

avec ce choix θ joue le rôle de dt apparaissant dans la TF . β est l'inverse de la durée du signal ; c'est donc l'écart de fréquence entre deux canaux fréquentiels : β joue donc le rôle de $d\nu$ apparaissant dans la TF^{-1} . Cette analogie avec le cas des signaux à temps continu donne un moyen mnémotechnique utile pour retrouver les relations entre le domaine temps et le domaine fréquence.

Considérons la relation entre la convolution en temps et le produit en fréquence. La convolution en temps fait apparaître l'élément différentiel dt . En TFD , avec nos conventions, cette relation s'écrira

$$z = \theta x * y \implies Z(m) = X(m)Y(m),$$

de même pour la relation duale :

$$z(n) = x(n).y(n) \implies Z = \beta X * Y.$$

Quant à la relation de Parseval

$$\int_{-\infty}^{+\infty} x(t)y^*(t)dt = \int_{-\infty}^{+\infty} X(\nu)Y^*(\nu)d\nu,$$

elle devient en TFD

$$\theta \sum_0^{N-1} x(n)y^*(n) = \beta \sum_0^{N-1} X(m)Y^*(m).$$

Pour le cas qui nous intéresse ici, nous devons utiliser des TFD à p dimensions. La généralisation de ce qui précède est très simple. La TFD p dimensionnelle du signal à p dimensions $x(n_1, \dots, n_p)$ est

$$X(m_1, \dots, m_p) = \left[\prod_{i=1}^p \theta_i \right] \sum_{n_1=0}^{N-1} \dots \sum_{n_p=0}^{N-1} x(n_1, \dots, n_p) e^{-\frac{2\pi j}{N} \sum n_i m_i},$$

et la TFD inverse

$$x(n_1, \dots, n_p) = \left[\prod_{i=1}^p \beta_i \right] \sum_{m_1=0}^{N-1} \dots \sum_{m_p=0}^{N-1} X(m_1, \dots, m_p) e^{\frac{2\pi j}{N} \sum n_i m_i},$$

avec

$$N\theta_i\beta_i = 1.$$

Les relations entre le produit et la convolution se généralisent de même. x et y étant des signaux à p dimensions et $*$ représentant une convolution à p dimensions on aura

$$z = \left[\prod_{i=1}^p \theta_i \right] x * y \implies Z(m_1, \dots, m_p) = X(m_1, \dots, m_p)Y(m_1, \dots, m_p),$$

et

$$z(n_1, \dots, n_p) = x(n_1, \dots, n_p).y(n_1, \dots, n_p) \implies Z = \left[\prod_{i=1}^p \beta_i \right] X * Y.$$

Le théorème de Parseval s'écrit

$$\begin{aligned} & \left[\prod_{i=1}^p \theta_i \right] \sum_{n_1=0}^{N-1} \dots \sum_{n_p=0}^{N-1} x(n_1, \dots, n_p) y^*(n_1, \dots, n_p) \\ = & \left[\prod_{i=1}^p \beta_i \right] \sum_{m_1=0}^{N-1} \dots \sum_{m_p=0}^{N-1} X(m_1, \dots, m_p) Y^*(m_1, \dots, m_p) \end{aligned}$$

Lorsque les différentes composantes des signaux sont homogènes on prendra la même valeur, quel que soit i , pour les constantes de calibration θ_i et β_i . On aura alors

$$\prod_{i=1}^p \theta_i = \theta^p \quad \text{et} \quad \prod_{i=1}^p \beta_i = \beta^p.$$

2.9.2 Sur les classes de signaux

Cette partie est essentiellement issue de [29]. Les processus $x(t)$ auxquels nous nous intéressons sont ceux pouvant être mis sous la forme

$$x(t) = \int \exp(2\pi j\nu t) dX(\nu)$$

où l'intégrale est une intégrale stochastique de Stieljes-Riemann. Cette représentation est aussi dite de Cramér. Blanc-Lapierre et Fortet montrent le résultat suivant :

$x(t)$ admet la décomposition harmonique $x(t) = \int \exp(2\pi j\nu t) dX(\nu)$ en moyenne quadratique si et seulement si

$$E[x(t_1)x^*(t_2)] = \int \int \exp[2\pi j(t_1\nu_1 - t_2\nu_2)] d^2\gamma(\nu_1, \nu_2),$$

où $d^2\gamma(\nu_1, \nu_2) = E[dX(\nu_1)dX(\nu_2)]$ vérifie

$$\int \int |d^2\gamma(\nu_1, \nu_2)| < C < +\infty$$

Le processus $x(t)$ est alors dit harmonisable. Mais pour traiter de statistiques d'ordre supérieur, l'harmonisabilité ne suffit plus. Les classes des signaux $f^{(k)}$ et $\Phi^{(k)}$ sont alors introduites dans [29].

On dit que $X(\nu) \in \Phi^{(k)}$ si $\forall l < k, \forall \varepsilon_1, \dots, \varepsilon_l$

$$\int \dots \int |E[dX_{\varepsilon_1}(\nu_1) \dots dX_{\varepsilon_l}(\nu_l)]| < C < +\infty$$

(où les ε_l correspondent à un complexe conjugué s'ils valent -1 et à un complexe non conjugué s'ils valent +1). $x(t) \in f^{(k)}$ si sa transformée de Fourier est dans $\Phi^{(k)}$. Le principal résultat est alors le suivant (page 427 de [29])

Si $x(t) \in f^{(k)}$, si $k = l(l-1)$, alors $\forall j \leq l$

$$x_{\varepsilon_1}(t_1) \dots x_{\varepsilon_j}(t_j) = \int \dots \int \exp[2\pi j (\sum_i \varepsilon_i t_i \nu_i)] dX_{\varepsilon_1}(\nu_1) \dots dX_{\varepsilon_j}(\nu_j)$$

existe en moyenne quadratique. En conséquence de ce type de convergence, $E[x_{\varepsilon_1}(t_1) \dots x_{\varepsilon_j}(t_j)]$ existe et vérifie

$$E[x_{\varepsilon_1}(t_1) \dots x_{\varepsilon_j}(t_j)] = \int \dots \int \exp[2\pi j (\sum_i \varepsilon_i t_i \nu_i)] \cdot E[dX_{\varepsilon_1}(\nu_1) \dots dX_{\varepsilon_j}(\nu_j)]$$

Remarquons que la classe $f^{(2)}$ correspond à l'ensemble des signaux harmonisables. Précisons enfin que la classe $f^{(\infty)}$ est la classe des signaux appartenant à $f^{(k)}$ pour tout k .

2.9.3 Le multipériodogramme construit sur les cumulants

Le multipériodogramme est défini au paragraphe 2.6.4 (page 70) par

$$P_{x(p)}(m_1, \dots, m_{p-1}) = \beta X(m_0)X(m_1) \dots X(m_{p-1}) \quad \text{avec} \quad m_0 = -\sum_{i=1}^{p-1} m_i.$$

Comme

- $X(m_0) = \theta \sum_0^{N-1} x(n) e^{2\pi j \frac{nm_0}{N}}$,
- $X(m_i) = \theta \sum_0^{N-1} x(n+k_i) e^{2\pi j \frac{(n+k_i)m_i}{N}}$ pour $i \neq 0$, la sommation est faite de 0 à $N-1$ car les échantillons sont indicés modulo N ,

le multipériodogramme construit à partir des cumulants est

$$\begin{aligned} P_{x(p)}^C(m_1, \dots, m_{p-1}) &= \beta \text{Cum}[X(m_0), X(m_1), \dots, X(m_{p-1})] \\ &= \beta \theta^p \sum_n e^{2\pi j \frac{n \sum_0^{p-1} m_i}{N}} \sum_{k_1} \dots \sum_{k_{p-1}} C_{x(p)}(k_1, \dots, k_{p-1}) e^{2\pi j \frac{\sum_1^{p-1} k_i m_i}{N}}, \end{aligned}$$

soit encore

$$P_{x(p)}^C(m_1, \dots, m_{p-1}) = \beta \theta N d\left(\sum_0^{p-1} m_i\right) TFD[C_{x(p)}(k_1, \dots, k_{p-1})],$$

où

$$\begin{cases} d(\sum_0^{p-1} m_i) = 1 & \text{si } \sum_0^{p-1} m_i = 0 \\ d(\sum_0^{p-1} m_i) = 0 & \text{si } \sum_0^{p-1} m_i \neq 0. \end{cases}$$

Finalement, on obtient

$$P_{x(p)}^C(m_1, \dots, m_{p-1}) = S_{x(p)}(m_1, \dots, m_{p-1}).$$

2.9.4 Comportement asymptotique

Le multispectre d'ordre p est la TFD à $p - 1$ dimensions de la multicorrélation

$$S_{x(p)}(m_1, \dots, m_{p-1}) = \theta^{p-1} \sum_{k_1=0}^{N-1} \dots \sum_{k_{p-1}=0}^{N-1} C_{x(p)}(k_1, \dots, k_{p-1}) e^{2\pi j \frac{\sum_{i=1}^{p-1} k_i m_i}{N}}.$$

Nous voulons étudier le comportement du multispectre lorsque N tend vers l'infini. Pour cette étude le choix naturel pour θ est $1/N$. En supposant la multicorrélation sommable²⁰, la TFD à $p - 1$ dimensions de la multicorrélation tend vers une limite finie lorsque N tend vers l'infini.

Le comportement du multispectre d'ordre p , lorsque N tend vers l'infini, est donné par $\theta^{p-1} = \frac{1}{N^{p-1}}$. Il s'ensuit que le mutipériodogramme d'ordre p est également homogène à $\frac{1}{N^{p-1}}$ lorsque N tend vers l'infini. Ces résultats nous permettront dans le paragraphe suivant de calculer la variance de l'estimateur du bispectre.

Notons pour terminer que la condition de sommabilité des multicorrélations est remplie pour les signaux appartenant à la classe que nous avons introduite en 2.5.1. Cette condition de sommabilité des multicorrélations est une condition d'ergodisme. On s'intéresse actuellement à des signaux dits "à longue dépendance" dont les multicorrélations ne sont pas sommables : pour ces signaux, l'estimation des multicorrélations pose des problèmes particuliers et difficiles que nous n'abordons pas ici.

2.9.5 Variance du bipériodogramme

Le bipériodogramme a été défini (2.27, page 71) par

$$P_{x(3)}(m_1, m_2) = \beta X(-m_1 - m_2)X(m_1)X(m_2).$$

La variance du bipériodogramme, calculée dans [105], est

$$\begin{aligned} \text{Var} \left[P_{x(3)}(m_1, m_2) \right] = & \\ & \beta^2 E \left[X(-m_1 - m_2)X(m_1)X(m_2)X(m_1 + m_2)X(-m_1)X(-m_2) \right] \\ & - \left| E \left[P_{x(3)}(m_1, m_2) \right] \right|^2. \end{aligned}$$

Cette variance est un moment d'ordre 6 que l'on peut schématiser par

$$\text{Var} \left[P_{x(3)}(m_1, m_2) \right] = \beta^2 E[1.2.3.4.5.6].$$

²⁰Le fait d'avoir défini les multicorrélations comme des cumulants est ici fondamental car, si les multicorrélations ainsi définies peuvent être sommables, les multicorrélations des moments ne sont généralement pas sommables par suite de la présence de branches s'étendant à l'infini.

Appelons i_6 l'ensemble $(1, 2, 3, 4, 5, 6)$ et i_p , ($p < 6$), les sous-ensembles contenant p éléments de i_6 . $E[1.2.3.4.5.6]$ est la somme des cumulants obtenus en considérant toutes les partitions possibles de i_6 (voir le chapitre 1).

- 1 partition : $Cum[i_6]$ contenant 1 terme,
- 2 partitions
 - $Cum[i_5]Cum[i_1]$ contenant 6 termes,
 - $Cum[i_4]Cum[i_2]$ contenant 15 termes,
 - $Cum[i_3]Cum[i_3]$ contenant 10 termes,
- 3 partitions
 - $Cum[i_4]Cum[i_1]Cum[i_1]$ contenant 15 termes,
 - $Cum[i_3]Cum[i_2]Cum[i_1]$ contenant 60 termes,
 - $Cum[i_2]Cum[i_2]Cum[i_2]$ contenant 15 termes,
- 4 partitions
 - $Cum[i_3]Cum[i_1]Cum[i_1]Cum[i_1]$ contenant 20 termes,
 - $Cum[i_2]Cum[i_2]Cum[i_1]Cum[i_1]$ contenant 45 termes,
- 5 partitions $Cum[i_2]Cum[i_1]Cum[i_1]Cum[i_1]Cum[i_1]$ contenant 15 termes,
- 6 partitions $Cum[i_1]Cum[i_1]Cum[i_1]Cum[i_1]Cum[i_1]Cum[i_1]$ contenant 1 terme.

Hinich [105] montre que dans cet ensemble de termes, il existe un terme dominant. Donnons le principe de cette sélection. Tout d'abord, comme les signaux sont supposés centrés, tous les termes faisant apparaître la moyenne sont nuls. Il ne reste alors que les termes associés à une, deux ou trois partitions. Les termes associés à une partition sont homogènes à un multispectre d'ordre 6 : leur comportement asymptotique est donc en $1/N^5$. Les termes associés à deux partitions contiennent le terme $\left| \left[P_{x(3)}(m_1, m_2) \right] \right|^2$ qui se simplifie dans l'expression de la variance et des termes produits d'un bispectre par un quadrispectre dont le comportement asymptotique est en $1/N^4$. Enfin, les termes non nuls associés à trois partitions sont homogènes au cube d'un spectre. Leur comportement asymptotique est en $1/N^3$. Asymptotiquement, les termes associés à trois partitions seront donc dominants. Parmi les 15 termes concernés, en tenant compte du fait que deux fréquences différentes sont décorréélées puisque le signal est stationnaire, il ne reste finalement qu'un seul terme, donnant au premier ordre pour la variance du bipériodogramme

$$\text{Var} \left[P_{x(3)}(m_1, m_2) \right] = \beta^2 E [X(m_1)X(-m_1)] E [X(m_2)X(-m_2)] E [X(m_1 + m_2)X(-m_1 - m_2)].$$

Comme

$$S_{x(2)}(m) = \beta E[X(m)X(-m)],$$

on obtient finalement

$$\text{Var} \left[P_{x(3)}(m_1, m_2) \right] = \frac{1}{\beta} S_{x(2)}(m_1) S_{x(2)}(m_2) S_{x(2)}(-m_1 - m_2). \quad (2.34)$$

On montre de même dans [105] que la covariance de deux termes différents du bipériodogramme est :

$$\text{Cov} \left[P_{x(3)}(m_1, m_2), P_{x(3)}(l_1, l_2) \right] = \beta S_{x(6)}(m_1, m_2, -m_1 - m_2, l_1, l_2). \quad (2.35)$$

2.9.6 Variance de la bicorrélation

La bicorrélation estimée est la TFD^{-1} du bipériodogramme

$$\widehat{C}_{x(3)}(k_1, k_2) = \beta^2 \sum_{m_1} \sum_{m_2} P_{x(3)}(m_1, m_2) e^{\frac{2\pi j}{N}(m_1 k_1 + m_2 k_2)}.$$

La variance de la bicorrélation estimée est donc

$$\begin{aligned} \text{Var} \left[\widehat{C}_{x(3)}(k_1, k_2) \right] &= \beta^4 \sum_{m_1, m_2} \sum_{l_1, l_2} \text{Cov} \left[P_{x(3)}(m_1, m_2), P_{x(3)}(-l_1, -l_2) \right] \\ &\quad \times e^{\frac{2\pi j}{N}(m_1 k_1 - l_1 k_1 + m_2 k_2 - l_2 k_2)}. \end{aligned}$$

En utilisant (2.34) et (2.35) on obtient

$$\text{Var} \left[\widehat{C}_{x(3)}(k_1, k_2) \right] = I + II,$$

avec

$$I = \beta^3 \sum_{m_1} \sum_{m_2} S_{x(2)}(m_1) S_{x(2)}(m_2) S_{x(2)}(-m_1 - m_2),$$

et

$$\begin{aligned} II &= \beta^3 \sum_{m_1, m_2} \sum_{l_1, l_2} S_{x(6)}(m_1, m_2, -m_1 - m_2, -l_1, -l_2) \\ &\quad \times e^{\frac{2\pi j}{N}(m_1 k_1 - l_1 k_1 + m_2 k_2 - l_2 k_2)}. \end{aligned}$$

Calcul de I Avec

$$S_{x(2)}(m_1) = \theta \sum_{k_1} C_{x(2)}(k_1) e^{-\frac{2\pi j}{N} k_1 m_1},$$

et

$$S_{x(2)}(-m_1 - m_2) = \theta \sum_{k_2} C_{x(2)}(k_2) e^{\frac{2\pi j}{N}(k_2 m_1 + k_2 m_2)},$$

on obtient

$$I = \theta^2 \beta^3 \sum_{m_1, m_2} \sum_{k_1, k_2} C_{x(2)}(k_1) C_{x(2)}(k_2) S_{x(2)}(m_2) e^{\frac{2\pi j}{N}(m_1(k_2 - k_1) + m_2 k_2)}.$$

Comme

$$\sum_{m_1} e^{\frac{2\pi j}{N} m_1 (k_2 - k_1)} = Nd(k_2 - k_1),$$

avec

$$d(k_2 - k_1) = \begin{cases} 1 & \text{si } k_2 = k_1 \\ 0 & \text{sinon} \end{cases}$$

et $\beta^2 \theta^2 = 1/N^2$, il vient

$$I = \frac{\beta}{N} \sum_{k_1} \sum_{m_2} \left[C_{x(2)}(k_1) \right]^2 S_{x(2)}(m_2) e^{\frac{2\pi j}{N} m_2 k_1},$$

donnant finalement

$$I = \frac{1}{N} \sum_{k_1} \left[C_{x(2)}(k_1) \right]^3.$$

Calcul de II On peut écrire le terme II sous la forme

$$II = (\beta\theta)^5 \sum_{i_1} \sum_{i_2} \sum_{i_3} \sum_{i_4} \sum_{i_5} C_{x(6)}(i_1, i_2, i_3, i_4, i_5) ABCD,$$

avec

- $A = \sum_{m_1} e^{\frac{2\pi j}{N} m_1 (-i_1 + i_3 + k_1)} = Nd(-i_1 + i_3 + k_1),$
- $B = \sum_{m_2} e^{\frac{2\pi j}{N} m_2 (-i_2 + i_3 + k_2)} = Nd(-i_2 + i_3 + k_2),$
- $C = \sum_{l_1} e^{\frac{2\pi j}{N} l_1 (i_4 - k_1)} = Nd(i_4 - k_1),$
- $D = \sum_{l_2} e^{\frac{2\pi j}{N} l_2 (i_5 - k_2)} = Nd(i_5 - k_2).$

En posant

$$i_3 = i,$$

on obtient finalement

$$II = \frac{1}{N} \sum_i C_{x(6)}(i + k_1, i + k_2, i, k_1, k_2).$$

La variance de la bicorrélation estimée est donc

$$\text{Var} \left[\widehat{C}_{x(3)}(k_1, k_2) \right] = \frac{1}{N} \sum_i \left[C_{x(2)}(i) \right]^3 + C_{x(6)}(i + k_1, i + k_2, i, k_1, k_2).$$

CHAPITRE 3

Signaux et systèmes

Il existe des relations étroites entre les signaux et les systèmes : les signaux permettent d'identifier les systèmes et un système peut caractériser un signal. Ces relations sont largement exploitées dans l'approche des signaux à l'ordre 2. On connaît à cet ordre de nombreuses méthodes d'identification [101, 142]. Parallèlement la modélisation d'un signal par un système (*AR*, *MA*, *ARMA*) est à la base de nombreuses méthodes de caractérisation des signaux à l'ordre 2 : en analyse spectrale, dans l'égalisation...

La caractérisation à l'ordre 2 est cependant partielle et elle ne permet pas de résoudre les problèmes d'identification d'un système linéaire dont on ne connaît que la sortie (identification aveugle) [65, 150]. Nous allons voir que les relations entre les signaux et les systèmes peuvent être approfondies et élargies lorsque l'on pousse la caractérisation des signaux à un ordre supérieur à 2. On peut alors, en particulier, résoudre le problème de l'identification aveugle.

3.1 Les systèmes échantillonnés

Ce paragraphe donne les rappels nécessaires à ce chapitre en ce qui concerne les filtres à temps discret.

3.1.1 Présentation générale

Présentons les systèmes linéaires et stationnaires qui sont à la base des représentations “systémiques” des signaux [142, 164]. En temps, la relation entrée x -sortie y , pour un système linéaire et stationnaire, est une convolution $y = h * x$, soit

$$y(n) = \sum_p h(n-p) x(p) = \sum_p h(p) x(n-p).$$

La fonction $h(n)$ qui caractérise le système linéaire et stationnaire est la réponse impulsionnelle. En fréquences réduites, ou en transformée en Z , la relation

entrée-sortie est le produit simple $Y(\lambda) = H(\lambda)X(\lambda)$, ou $Y(z) = H(z)X(z)$ [147]. $H(\lambda)$ est la fonction de transfert ou gain complexe en fréquences réduites, et $H(z)$ est la fonction de transfert en Z .

On peut également décrire un système linéaire et stationnaire par une équation aux différences [147, 164]

$$y(n) + \sum_{k=1}^P a(k) y(n-k) = \sum_{l=0}^Q b(l) x(n-l).$$

Nous avons choisi les bornes de sommation de manière à ce que le système soit causal : nous construisons la sortie à l'instant n à partir de l'entrée à l'instant n , des entrées passées et des sorties passées. Cette contrainte n'est pas absolue. En pratique les sommes portent sur un nombre fini de termes. En théorie, on peut considérer des systèmes linéaires et stationnaires pour lesquels on a une sommation infinie. La fonction de transfert en Z est

$$H(z) = \frac{\sum_0^Q b(l)z^{-l}}{1 + \sum_1^P a(k)z^{-k}}.$$

En factorisant le numérateur et le dénominateur de la fonction de transfert on peut l'écrire

$$H(z) = b(0)z^{P-Q} \frac{\prod_{i=1}^Q (z - z_{0i})}{\prod_{j=1}^P (z - z_{pj})}.$$

On fait ainsi apparaître les P pôles z_{pj} et les Q zéros z_{0i} du système linéaire et stationnaire. Les pôles et les zéros constituent une description très parlante des propriétés du système linéaire et stationnaire donnant la fonction de transfert à un facteur près $b(0)$ et à un retard $Q - P$ près [147].

On distingue :

- les systèmes linéaires et stationnaires autorégressifs (*AR*) pour lesquels $b(0) = 1$ et $b(l) = 0$ pour l non nul. La relation entrée-sortie est pour ces systèmes

$$y(n) = - \sum_{k=1}^P a(k)y(n-k) + x(n),$$

montrant que la sortie à l'instant n est construite à partir des P sorties précédentes et de la seule entrée à l'instant n . Ces systèmes ont P pôles et un zéro d'ordre P à l'origine.

- les systèmes linéaires et stationnaires à moyenne mobile ou à moyenne ajustée (*MA*) pour lesquels $a(k) = 0$ pour tout k . La relation entrée-sortie est

$$y(n) = \sum_{l=0}^Q b(l)x(n-l),$$

et la sortie à l'instant n est une moyenne des entrées concomittante et passées. Ces systèmes ont Q zéros et un pôle d'ordre Q à l'origine.

- enfin le cas général est celui des systèmes linéaires et stationnaires autorégressifs à moyenne ajustée (*ARMA*).

On classe également les systèmes linéaires et stationnaires selon la structure de leur réponse impulsionnelle. Les filtres *MA* ont une réponse impulsionnelle de longueur finie : filtres à réponse impulsionnelle finie (RIF). Les filtres *AR* et *ARMA* ont une réponse impulsionnelle de longueur infinie : filtres à réponse impulsionnelle infinie (RII) .

3.1.2 Stabilité et causalité

Pour être “physiquement réalisable” un système linéaire et stationnaire doit être stable et causal [99, 164].

La stabilité BIBO (bounded input-bounded output) exige qu'à toute entrée bornée le système associe une sortie bornée. Autrement dit, un système est stable, si pour toute entrée $x(n)$ vérifiant

$$|x(n)| < +\infty,$$

la sortie $y(n)$ vérifie

$$|y(n)| < +\infty.$$

La causalité stipule que l'effet (la sortie) ne doit pas précéder la cause (l'entrée). La condition de causalité peut se formuler par l'implication suivante :

$$x(n) = 0 \text{ pour } n < n_0 \quad \Rightarrow \quad y(n) = 0 \text{ pour } n < n_0.$$

La stabilité et la causalité peuvent s'exprimer dans les différentes représentations des systèmes linéaires et stationnaires.

Une condition nécessaire et suffisante de stabilité est que la réponse impulsionnelle soit sommable en module, c'est-à-dire

$$\sum_{-\infty}^{+\infty} |h(n)| < +\infty,$$

et la condition de causalité est simplement

$$h(n) = 0 \text{ pour } n < 0.$$

En fréquences réduites le critère de Paley-Wiener [166] nous indique qu'il existe au moins un système linéaire et stationnaire, stable et causal, dont le module de la fonction de transfert est $|H(\lambda)|$ si et seulement si

$$\int_{-1/2}^{1/2} \log |H(\lambda)| d\lambda < +\infty. \quad (3.1)$$

Enfin, en Z , un système sera stable et causal si et seulement si :

1. le nombre de pôles est supérieur ou égal au nombre de zéros,
2. tous les pôles sont à l'intérieur du cercle de rayon 1, dans le plan complexe des Z .

Système linéaire et stationnaire, stable et causal Pour un système stable et causal la réponse impulsionnelle vérifie :

$$h(n) = 0 \quad \text{pour} \quad n < 0.$$

C'est la condition duale de la condition imposée au signal analytique. La partie réelle et la partie imaginaire du gain complexe en fréquences réduites sont donc reliées par une transformation de Hilbert en fréquences réduites (THFr). Ceci est exprimé par les *relations de Kramers-Kronig*

$$\begin{aligned} \text{Im}[H(\lambda)] &= THFr^{-1}[\text{Re}(H(\lambda))] \\ \text{Re}[H(\lambda)] &= THFr[\text{Im}(H(\lambda))]. \end{aligned}$$

La $THFr$ et la $THFr^{-1}$ sont les transformées de Hilbert directe et inverse en fréquences réduites.

Système linéaire et stationnaire, stable, causal et causalement inversible (ou à phase minimale) Un système linéaire et stationnaire sera dit stable, causal et causalement inversible si :

1. le système linéaire et stationnaire est stable et causal ;
2. son inverse (permettant de retrouver l'entrée à partir de la sortie) est stable et causal.

Soit un système linéaire et stationnaire de fonction de transfert en Z , $H(z)$, la fonction de transfert du système inverse est

$$H_i(z) = \frac{1}{H(z)}.$$

Les pôles de $H(z)$ sont les zéros de $H_i(z)$ et les zéros de $H(z)$ sont les pôles de $H_i(z)$. Il s'ensuit qu'un système linéaire et stationnaire sera stable, causal et causalement inversible si et seulement si :

1. le nombre de pôles est égal au nombre de zéros,
2. les pôles et les zéros sont dans le disque de rayon 1.

Pour un système linéaire et stationnaire, stable, causal et causalement inversible les pôles du logarithme de la fonction de transfert, qui sont les pôles et les zéros de la fonction de transfert, sont dans le cercle de rayon 1. La transformée en Z inverse de ce logarithme, que l'on appelle le *cepstre*, est donc une fonction causale (nulle pour les temps négatifs), soit

$$TZ^{-1} [\log H(z)] = C(n) = 0 \quad \text{pour } n < 0.$$

La partie réelle et la partie imaginaire de ce logarithme sont donc reliées par une transformation de Hilbert. En fréquences réduites on a

$$\log [H(\lambda)] = \log |H(\lambda)| + j\phi_H(\lambda),$$

d'où l'on déduit les *relations de Bode-Bayard*

$$\begin{aligned} \phi_H(\lambda) &= THFr^{-1} [\log |H(\lambda)|] \\ \log |H(\lambda)| &= THFr [\phi_H(\lambda)], \end{aligned} \quad (3.2)$$

qui relie le module et la phase de la fonction de transfert d'un système linéaire et stationnaire, stable, causal et causalement inversible. La relation (3.2) nous montre que si l'on connaît le module de la fonction de transfert d'un système linéaire et stationnaire, stable, causal et causalement inversible, alors sa phase est déterminée par (3.2). En rapprochant ce résultat du critère de Paley-Wiener (3.1) on voit que si le module d'un système linéaire et stationnaire, stable, causal et causalement inversible vérifie ce critère, qui porte uniquement sur le module de la fonction de transfert, on peut calculer sa phase.

Les systèmes linéaires et stationnaires, stables, causaux et causalement inversibles sont dénommés également filtres à phase minimale. Cette dénomination vient de l'étude de la famille de filtres ayant une valeur fixée pour le module de la fonction de transfert en fréquences réduites (gabarit). On sait que cette condition définit une famille de filtres. Dans cette famille le filtre stable, causal et causalement inversible est celui pour lequel la rotation de phase totale, lorsque la fréquence réduite varie de $-1/2$ à $+1/2$, est minimale. Ce filtre est intéressant car, dans la famille de filtres considérée, il est celui qui répond le plus rapidement. En effet, en appelant $h(n)$ la réponse impulsionnelle d'un filtre de la famille, et $h_{min}(n)$ celle du filtre à phase minimale, pour tout $n_0 > 0$ on a

$$\sum_0^{n_0} |h(n)|^2 \leq \sum_0^{n_0} |h_{min}(n)|^2.$$

Lorsque l'on réalise l'identification d'un filtre *ARMA*, les pôles doivent être dans le cercle de rayon 1 pour que le filtre soit stable¹. Les zéros peuvent

¹Dans certaines situations, on pourra considérer des filtres ayant des pôles à l'extérieur du cercle unité en inversant le sens du temps : filtres anticausaux.

occuper indifféremment deux positions inverses par rapport au cercle de rayon 1 sans modifier le module du gain complexe du système linéaire et stationnaire. Cette indétermination sur la position des zéros engendre la famille de filtres ayant le même gabarit, et elle ne peut pas être levée par les techniques d'identification à l'ordre 2 pour des signaux et des filtres stationnaires.

Conclusion Le but essentiel de cette brève introduction sur les systèmes linéaires et stationnaires était de préciser les différents modèles (*AR*, *MA* et *ARMA*) et de situer les problèmes importants de causalité, de stabilité et de phase minimale (ou d'inversibilité causale). En général, quand on identifie un système linéaire et stationnaire, on est amené à supposer qu'il est stable et causal, bien que ces hypothèses puissent être remises en cause. Par contre, rien ne permet de supposer qu'il est à phase minimale. Nous verrons que ce problème de phase est au centre des techniques d'identification aveugle.

Le lecteur peut approfondir ses connaissances sur les systèmes linéaires et stationnaires en consultant de nombreux ouvrages, comme par exemple [101, 132, 142, 164].

3.2 Modélisation et filtrage

Avant d'aborder les techniques d'identification aveugle, il est bon de donner quelques considérations générales qui nous permettront de situer les divers problèmes de modélisation et d'identification. Nous préciserons et comparerons également les notions importantes de filtre blanchisseur et filtre inverse. Ces différentes notions seront illustrées par un exemple faisant ressortir les différences entre le blanchiment et l'inversion d'un signal. Cet exemple nous permettra également de présenter la position spéciale occupée par les signaux gaussiens.

3.2.1 Modélisation systémique

Un signal aléatoire stationnaire $x(n)$ peut être modélisé par la sortie $x(n)$ d'un filtre non linéaire de Volterra d'ordre éventuellement infini [64]. Le terme d'ordre i est

$$x_i(n) = \sum_{k_1=-\infty}^{+\infty} \dots \sum_{k_i=-\infty}^{+\infty} h_i(k_1, \dots, k_i) b(n - k_1) \dots b(n - k_i).$$

donnant la sortie du filtre de Volterra

$$x(n) = \sum_{i=1}^{\infty} x_i(n) \quad (3.3)$$

Les fonctions $h_i(k_1, \dots, k_i)$ sont les noyaux de Volterra et le signal d'entrée $b(n)$ est une suite de variables aléatoires indépendantes de même loi². Le système, ou filtre, en général non linéaire, permettant d'engendrer le signal $x(n)$ à partir d'une suite de variables aléatoires indépendantes est appelé *système générateur*. Ce modèle fait apparaître les relations entre les signaux et les systèmes, linéaires ou non linéaires.

Dans certaines situations, en particulier dans les canaux de transmission, où l'on peut imposer au système d'être linéaire, on obtient alors le modèle simplifié suivant dit *modèle linéaire* :

$$x(n) = \sum_{k_1=-\infty}^{+\infty} h_1(k_1) b(n - k_1), \quad (3.4)$$

$h_1(k_1)$ étant la réponse impulsionnelle du système linéaire.

Les signaux dont le système générateur est linéaire sont appelés *signaux linéaires*.

3.2.2 Filtre blanchisseur. Les différents bruits blancs

Soit un signal $x(n)$, on appelle filtre blanchisseur le filtre linéaire qui, ayant à son entrée $x(n)$, donnera en sortie un signal (bruit) blanc à l'ordre 2. La notion très classique de signal blanc (on dit plutôt bruit blanc, et nous suivrons la tradition bien qu'elle affecte ce signal d'une connotation péjorative qu'il ne mérite pas) doit être repensée quand on s'intéresse aux propriétés d'ordre supérieur.

En fait, il n'existe pas un bruit blanc, mais plusieurs bruits blancs qui doivent être introduits [34]. Selon le degré de finesse de l'analyse on aura, si l'on peut dire, plusieurs degrés de blancheur.

En analyse spectrale un bruit blanc est un signal dont le spectre est constant. La corrélation de deux échantillons différents du bruit blanc à l'ordre 2 est donc nulle : un *bruit blanc à l'ordre 2* est alors formé par une suite de variables aléatoires décorréliées.

A un ordre supérieur, on est concerné par des cumulants d'ordre supérieur à 2. Un *bruit blanc à l'ordre q* est formé d'une suite de variables aléatoires dont les cumulants d'ordre inférieur ou égal à q sont nuls dès que l'on s'intéresse à un cumulant mettant en jeu plus d'un échantillon.

Enfin un bruit blanc est dénommé *bruit blanc parfait* si les échantillons de ce signal sont des variables aléatoires indépendantes.

Il existe des relations d'ordre partiel entre ces diverses blancheurs [34]. La blancheur parfaite entraîne la blancheur à tous les ordres. A l'inverse, la blancheur à l'ordre 2 n'entraîne en général ni la blancheur à un ordre supérieur : des variables aléatoires peuvent être décorréliées et avoir des cumulants croisés

²Les problèmes d'existence sont discutés au chapitre 5.

d'ordre supérieur à 2 non nuls, ni la blancheur parfaite : la décorrélation n'entraîne pas l'indépendance. Toutes les blancheurs sont équivalentes dans le cas d'un bruit gaussien : on sait que dans le cas gaussien la décorrélation entraîne l'indépendance. Nous avons vu également au chapitre 2 que les ordres supérieurs de signaux gaussiens n'apportent pas d'information : les cumulants d'ordre supérieur à 2 sont nuls. La notion de filtre blanchisseur n'est donc pas unique.

Classiquement, quand on parle de filtre blanchisseur, on sous-entend un filtre blanchisseur à l'ordre 2, c'est-à-dire un filtre dont la sortie est un bruit blanc d'ordre 2. Pour les signaux aléatoires décrits par (3.3), il existe toujours un filtre linéaire blanchisseur à l'ordre 2. La connaissance de ce filtre est suffisante quand on fait de l'analyse spectrale. Enfin ce filtre blanchisseur est un filtre blanchisseur parfait si le signal blanchi est gaussien.

À un ordre supérieur à 2 et pour les signaux non gaussiens, la situation est plus complexe. . .

3.2.3 Filtre inverse

Le filtre inverse d'un signal $x(n)$ modélisé par (3.3) est le système permettant de retrouver le signal d'entrée $b(n)$ à partir de la sortie $x(n)$. Pour des signaux linéaires (engendrés selon la relation (3.4)) le filtre inverse est linéaire. Dans le cas général, le filtre inverse s'il existe n'est pas linéaire : ce filtre inverse est donc différent du filtre blanchisseur à l'ordre 2 qui est linéaire.

Quelle est, pour les signaux linéaires, la relation entre le filtre blanchisseur et le filtre inverse?

3.2.4 Relation entre le filtre blanchisseur et le filtre inverse pour des signaux linéaires. Identification aveugle

Soit un signal $x(n)$ aléatoire, stationnaire, linéaire et de puissance finie. Connaissant le signal $x(n)$ et ses caractéristiques à l'ordre 2, corrélation ou spectre, on veut trouver le filtre blanchisseur et le filtre inverse de $x(n)$. Le signal étudié vérifie le modèle (3.4), son spectre est donc

$$S_{x(2)}(\lambda) = |H(\lambda)|^2 P_b,$$

où $H(\lambda)$ est le gain complexe du filtre engendrant $x(n)$, et P_b est la puissance du bruit générateur $b(n)$.

La connaissance de $S_{x(2)}(\lambda)$ nous permet d'accéder au module $1/|H(\lambda)|$ du gain complexe du filtre blanchisseur, mais ne nous donne pas d'information sur la phase de ce filtre. On pourra construire le filtre blanchisseur de $x(n)$ en utilisant les diverses méthodes développées dans les approches à l'ordre 2 [142].

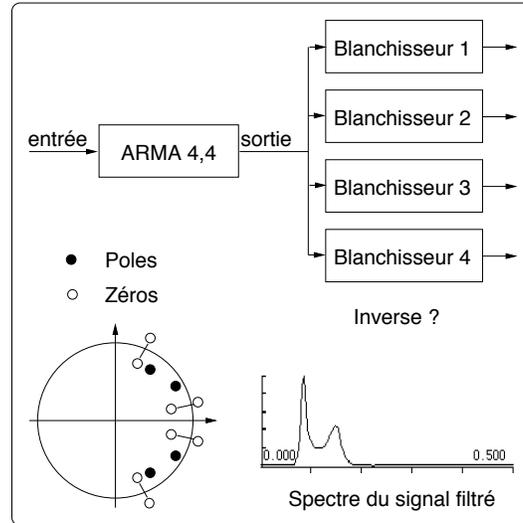


Figure 3.1: *Le filtre direct et les filtres blanchisseurs. A chaque zéro, on associe son inverse géométrique par rapport au cercle de rayon 1. Les zéros peuvent être choisis à l'intérieur ou à l'extérieur du cercle de rayon 1. La réalité du filtre restreint à 4 le nombre de filtres blanchisseurs. Un seul de ces filtres est le filtre inverse.*

- en temps, par la corrélation et les équations de Yule-Walker, en donnant au filtre blanchisseur une structure MA ,
- en temps, par l'algorithme de Levinson et les coefficients de corrélation partielle, en donnant au filtre blanchisseur une structure en treillis,
- en fréquence, en prenant une détermination particulière pour la phase du filtre blanchisseur : phase nulle, phase minimale...

Pour obtenir le filtre inverse nous devons connaître la phase qui n'est pas donnée par les caractéristiques d'ordre 2 du signal de sortie. Nous allons voir que l'on peut réaliser cette identification aveugle en utilisant les caractéristiques d'ordre supérieur du signal analysé.

3.2.5 Illustration

Les analogies et les différences entre les notions de filtre inverse et de filtre blanchisseur, ainsi que le cas singulier présenté par les signaux gaussiens sont illustrés par la simulation suivante inspirée par [65] (figure 3.1).

Un signal $x(n)$ est engendré par filtrage d'un signal blanc parfait $b(n)$ par un filtre $ARMA(4, 4)$ présentant quatre pôles (stables) et quatre zéros (2 dans le disque de rayon 1 et 2 extérieurs). On construit ensuite les quatre filtres blanchisseurs en réalisant un filtre $ARMA(4, 4)$ dont les zéros sont les pôles

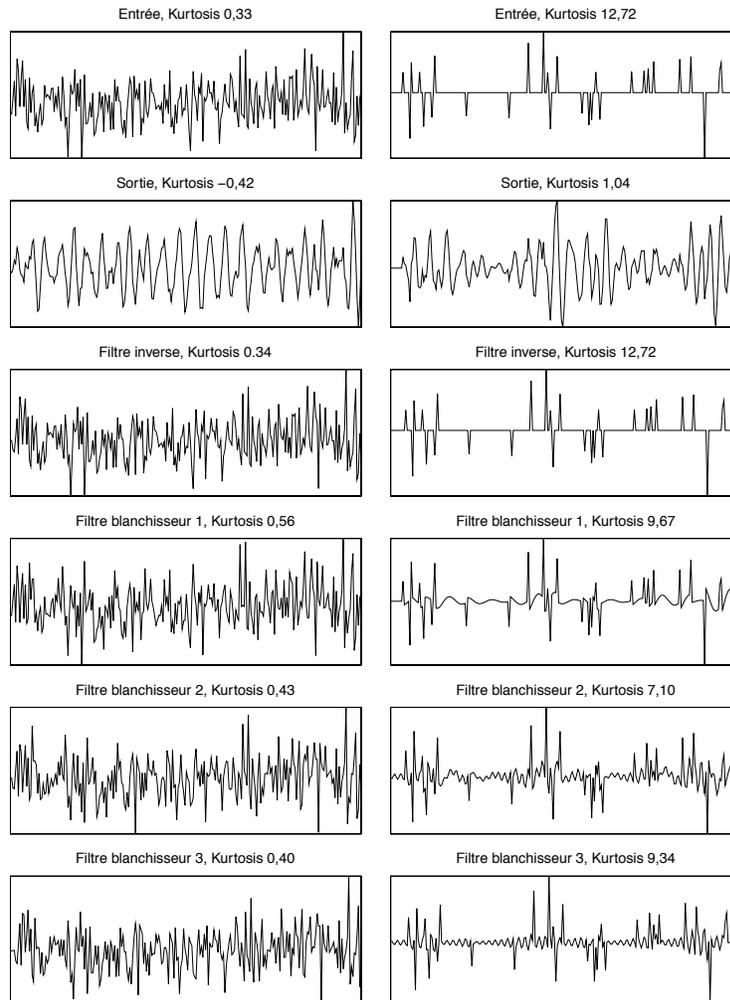


Figure 3.2: *Entrées gaussiennes et non gaussiennes, sortie du filtre direct. Sorties du filtre inverse et des filtres blanchisseurs.*

du filtre direct, et dont les pôles sont les zéros du filtre direct ou les inverses par rapport au cercle unité des zéros du filtre direct. Un seul de ces filtres blanchisseurs est l'inverse. Les positions des pôles et zéros des filtres direct et blanchisseurs, ainsi que la densité spectrale de puissance du signal $x(n)$ sont illustrés sur la figure (3.1).

Deux types de signaux blancs d'entrée ont été utilisés. Le premier est un signal gaussien. Les propriétés de ce signal gaussien sont contenues complètement dans ses caractéristiques d'ordre 2 : on peut prévoir que les sorties des filtres 1, 2, 3 et 4 ne pourront pas être distinguées. Ce fait est illustré, non démontré !, sur les panneaux de gauche de la figure (3.2) qui montrent le signal d'entrée, le signal filtré, et les sorties des quatre filtres. L'observation des seules sorties ne permet pas de distinguer les quatre filtres. On trouve,

évidemment, le filtre inverse quand on compare point à point les sorties et l'entrée...

Le second signal est un signal blanc, non gaussien. Nous avons choisi une séquence d'impulsions qui rappellent les coefficients de réflexion des interfaces entre couches géologiques dans une expérimentation de prospection sismique [174].

On voit sur les panneaux de droite de la figure (3.2) que les signaux de sortie sont différents ; on peut donc penser qu'il est possible de trouver le filtre inverse en observant seulement la sortie : c'est ce que l'on appelle la déconvolution aveugle dont nous reparlerons en détails plus loin. Donnons, en anticipant, l'idée d'une solution.

Puisque le problème n'a pas de solution à l'ordre 2 nous devons utiliser des caractéristiques d'ordre supérieur des signaux de sortie. Nous avons choisi le kurtosis qui est le cumulants d'ordre 4 normé (chapitre 1)

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

Les kurtosis des différents signaux de sortie sont donnés sur les figures. Dans le cas gaussien ils sont théoriquement nuls : leur valeur proche de 0 est issue des erreurs d'estimation du kurtosis. Dans le cas non gaussien, ils sont significativement différents. Le filtre inverse engendre un signal dont le kurtosis est maximum : ce critère permet donc de choisir entre les diverses solutions.

3.2.6 Conclusion

En prenant du recul vis à vis des approches habituelles de la modélisation des signaux à l'ordre 2, nous pouvons mieux discerner les relations existant entre les différents types de filtres appliqués aux signaux : filtre blanchisseur, filtre inverse linéaire ou non linéaire...

En bref, le filtre linéaire blanchisseur à l'ordre 2 est donné par les propriétés à l'ordre 2 du signal : spectre ou corrélation. Le filtre inverse dont le but est de retrouver le bruit blanc parfait générateur est linéaire dans le cas de signaux linéaires, non linéaire (si il existe) dans les autres cas. Dans le cas linéaire, le filtre inverse n'est pas donné par les propriétés du signal à l'ordre 2. Nous allons voir comment les caractéristiques d'ordre supérieur à 2 permettent de trouver le filtre inverse en réalisant une identification aveugle.

Le cas des signaux gaussiens est singulier. Cette situation est très défavorable : lorsque les signaux sont gaussiens et stationnaires, il est impossible de réaliser une déconvolution aveugle...

3.3 Identification aveugle (extralucide?)

La situation considérée est décrite figure (3.3, page 106). Le signal d'entrée $e(n)$ est un bruit blanc parfait, suite de variables aléatoires indépendantes de même loi, et le signal observé $o(n)$ est une filtrée linéaire et stationnaire du signal d'entrée. On ne connaît que ce signal de sortie et l'on veut retrouver le signal d'entrée.

On pourra seulement retrouver un signal *proportionnel* au signal d'entrée à un retard inconnu près (n_r). Dans le cas de signaux à valeurs complexes la constante de proportionnalité est complexe : le signal sera donc également reconstitué à une phase (ϕ) globale près :

$$s(n) = Ae^{j\phi}e(n - n_r),$$

les paramètres A , n_r et ϕ ne sont pas identifiables.

Cette situation se rencontre dans plusieurs domaines d'application pour des signaux à une ou plusieurs dimensions [101, 132, 135]. Dans les systèmes de communication, le filtre inverse, dénommé égaliseur, permet de reconstituer le message émis. Dans de nombreux problèmes de mesure, le signal observé est convolué par la fonction de transfert du système de mesure ; l'identification de ce filtre permet de réaliser une déconvolution. Citons dans ce domaine la prospection sismique [35, 174], le sondage radar ou sonar, l'échographie, le contrôle non destructif...

Pour retrouver $e(n)$ nous devons construire le filtre inverse. Le point fondamental que nous avons introduit en 3.2.5 est l'impossibilité de réaliser cette identification à partir de la sortie seule par des techniques faisant appel aux moments d'ordre 2. Nous allons voir comment ce problème peut être résolu en utilisant, de manière directe ou indirecte, les statistiques d'ordre supérieur. Ces nouvelles techniques se sont développées dans les vingt dernières années en étendant les algorithmes connus à l'ordre 2. Nous allons les présenter en distinguant les systèmes d'identification en boucle qui consistent à construire un système linéaire et stationnaire adaptatif convergeant vers le filtre inverse, et les systèmes en ligne qui donnent une estimation directe du filtre inverse. Comme toute classification cette distinction n'est pas absolue et l'on trouvera des systèmes relevant des deux approches.

Le tableau (3.1) synthétise le contenu des trois paragraphes qui suivent.

3.4 Identification aveugle “en boucle”

Les systèmes d'identification “en boucle” ont été développés dans le cadre des études sur les égaliseurs. Les références de base sont [25, 39, 65, 92, 179, 185]. Nous nous appuyons également sur des articles de synthèse [23, 150].

IDENTIFICATION AVEUGLE	
Non paramétrique	
bouclée, §3.4	Critères non quadratiques, p. 108 kurtosis p. 111 Godard p. 113 Sato p. 113
en fréquence, §3.5	Méthode récursive linéaire p. 116 Moindres carrés, p. 117 Méthode de projection, p. 118 Retard de groupe, p. 119
Paramétrique	
filtre MA, §3.6.2	Formule $C(q, k)$, p. 125 Méthodes algébriques, p. 126 Optimisation, p. 128 Multicepstre, p. 129
filtre AR, §3.6.3	Yule-Walker ordre supérieur, p. 130
filtre ARMA, §3.6.4	p. 131

Tableau 3.1: *Synthèse des méthodes d'identification aveugles présentées dans ce chapitre*

3.4.1 L'algorithme du gradient stochastique

La structure générale de ces systèmes est donnée sur la figure (3.3). Le signal observé $o(n)$ est issu du filtrage d'un bruit blanc parfait (suite de variables aléatoires indépendantes) $e(n)$ par le filtre direct de réponse impulsionnelle $d(k)$.

$$o(n) = \sum_p d(p)e(n-p).$$

On cherche un filtre inverse transversal décrit par sa réponse impulsionnelle de longueur finie $i(k)$ pour $-p_1 \leq k \leq p_2$, donnant en sortie $s(n)$ via

$$s(n) = \sum_p i(p)o(n-p) = \mathbf{i}^T \mathbf{o}(n),$$

avec $\mathbf{i}^T = (i(-p_1) \dots i(p_2))$ et $\mathbf{o}^T(n) = (o(n+p_1) \dots o(n-p_2))$. Dans le cas général, le filtre cherché peut contenir une partie causale $0 \leq k \leq p_2$ et une partie non-causale $-p_1 \leq k < 0$.

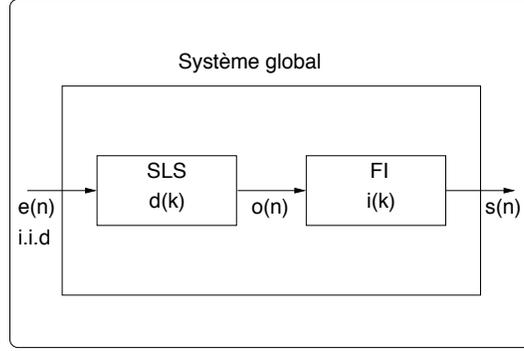


Figure 3.3: *Système global de déconvolution aveugle. Les seules données disponibles sont les échantillons du signal $o(n)$ issu du passage de l'entrée $e(n)$ dans le canal $d(k)$. L'hypothèse fondamentale de la déconvolution aveugle est le caractère *i.i.d.* de l'entrée. Le déconvolveur est le filtre inverse $i(k)$. La sortie $s(n)$ est égale à l'entrée, à une constante multiplicative près et à un retard près.*

Le bruit blanc d'entrée $e(n)$ et la sortie $s(n)$ sont reliés par le filtre global (figure (3.3)) de réponse impulsionnelle $t(k)$ par

$$s(n) = \sum_p t(p)e(n-p). \quad (3.5)$$

Pour obtenir le filtre inverse, on minimise un critère $V(s)$ dépendant de la sortie³. Les signaux traités étant aléatoires, le critère est généralement un moment du type

$$V(s) = E[v(s)].$$

La convergence vers le filtre inverse est obtenue par une méthode du gradient [101, 132, 135, 198]. En pratique, on ne peut pas calculer l'espérance mathématique donnant $V(s)$. On la remplace par sa valeur instantanée $v(s)$. On obtient alors une méthode d'adaptation dite du gradient stochastique [132, 172].

Pendant la période d'adaptation du filtre, sa réponse impulsionnelle est variable et nous la noterons $\mathbf{i}(n)$. En théorie, $\mathbf{i}(n)$ doit tendre vers une limite. A la convergence, le filtre devient donc stationnaire en moyenne. En réalité, par suite des erreurs d'estimation, le filtre ne se stabilise jamais complètement...

L'adaptation du filtre est obtenue par

$$\mathbf{i}(n+1) = \mathbf{i}(n) - \mu \nabla v[s(n)],$$

où $\nabla v(s)$ est le gradient de v par rapport aux coefficients du filtre, et μ est une constante dont la valeur fixe la vitesse de convergence et l'amplitude des fluctuations résiduelles de la réponse impulsionnelle estimée.

³En physique, on rencontre le même problème de minimisation de l'énergie des systèmes décrits par un potentiel.

Quand on utilise un critère quadratique (ordre 2), μ doit être inférieur à une valeur déterminée pour assurer la convergence [198]. On peut diminuer la valeur de μ lorsque l’on approche de la convergence [135].

Cette technique est très étudiée dans les méthodes utilisant des critères quadratiques. Elle s’appelle alors la méthode des moindres carrés moyens et une très vaste littérature a été consacrée à ce sujet [101]. Les principales variantes portent sur la structure du filtre estimé : transversal comme ici, en treillis [97, 137]. . . On peut également définir le filtre en temps, en fréquence [163], on peut optimiser la vitesse de convergence : moindres carrés. . .

En résumé l’algorithme du gradient stochastique s’écrit

$$\begin{cases} s(n) & = \mathbf{i}^T(n)\mathbf{o}(n), \\ \mathbf{i}(n+1) & = \mathbf{i}(n) - \mu\nabla v[s(n)]. \end{cases}$$

Parfois $\mathbf{i}(n) = \mathbf{0}$ est une solution des algorithmes de minimisation. Pour éviter de converger vers cette solution triviale on impose des contraintes à la réponse impulsionnelle. Les contraintes imposées sont généralement soit⁴ $\mathbf{i}_0 = 1$ soit $\|\mathbf{i}\|^2 = \mathbf{i}^T\mathbf{i} = 1$. On obtient alors un algorithme de minimisation sous contrainte.

On peut voir l’algorithme du gradient stochastique de la façon imagée suivante. Le filtre est un randonneur qui se déplace dans un terrain boisé (ce qui ne lui permet pas de voir à distance) dont la topographie est fixée par le critère. Le randonneur ne possède pas de carte mais dispose d’un niveau à bulles lui permettant de mesurer la pente du sol exactement sous ses pieds. Le randonneur veut atteindre le point le plus haut (ou le plus bas, mais pour les randonneurs nous opterions plutôt pour le point le plus haut. . .) du territoire. Il va donc chercher à monter constamment en mesurant, grâce à son niveau à bulles, la déclivité du sol. Le sol, comme les signaux, est aléatoire (mottes de terre, touffes d’herbe, fourmilières. . .) la déclivité locale mesurée est donc le gradient stochastique

$$\nabla v[s(n)].$$

Le randonneur peut “optimiser” sa route : moindres carrés. Le randonneur peut également adapter l’amplitude de ses pas par μ . Lorsque le randonneur est près du but, on imagine que si ses pas sont d’amplitude constante il va osciller autour du sommet, voire redescendre. Pour se rapprocher du sommet il doit donc faire des pas de plus en plus petits (diminution de μ). Le cas des algorithmes sous contraintes est ici celui du randonneur qui veut atteindre le sommet en suivant un chemin. On lui suggère alors deux solutions : soit se contenter de mesurer la pente le long du chemin : c’est le plus simple pour un randonneur, soit mesurer la pente sans tenir compte du chemin, faire un pas dans la direction trouvée et gagner le point du chemin le plus proche

⁴ \mathbf{i}_0 est la réponse impulsionnelle au temps 0, et donc la première composante du vecteur filtre dans le cas causal.

de sa nouvelle position : c'est le plus simple pour les systèmes adaptatifs. Quand le randonneur peut enfin se reposer sur le sommet tant désiré, il a le temps de se poser la grave question suivante : ai-je atteint le véritable sommet ? Seule la métaphysique du critère, c'est-à-dire les propriétés générales du critère (concavité par exemple), lui permettra de s'assurer de ce résultat. S'il dispose d'un altimètre, on lui suggère une solution (fatigante) qui consiste à repartir d'un autre point et à recommencer de multiples fois son ascension. Il verra bien à la fin si il rencontre, ou non, un sommet plus élevé : c'est la méthode du recuit simulé dont le nom semble bien adapté à cette histoire de randonneur...

Les méthodes du gradient stochastique ont été développées avec des critères quadratiques qui ont l'avantage d'être concaves. Dans le contexte considéré ici (identification aveugle), ces méthodes permettent de trouver un filtre blanchisseur mais ne donnent pas forcément le filtre inverse. Pour obtenir le filtre inverse il faut utiliser un critère plus complexe, ce qui fait apparaître dans le traitement des moments ou des cumulants d'ordre supérieur à 2. Après avoir présenté une vue d'ensemble, nous donnerons une solution simple et élégante de ce problème due à [185], nous discuterons les propriétés des algorithmes que l'on appelle les algorithmes de Bussgang [39] et nous les illustrerons sur des exemples.

3.4.2 Sur les critères

Les méthodes classiques, à l'ordre 2, utilisent la puissance de la sortie comme critère à minimiser

$$V(s) = E[|s|^2].$$

Le filtre ainsi identifié est stable, causal et causalement inversible ; il ne peut être le filtre inverse que si celui-ci est stable, causal et causalement inversible. En général le filtre inverse est stable et causal mais il n'est pas toujours causalement inversible (à phase minimale) comme nous l'avons montré dans l'illustration du paragraphe 3.2.5. Pour identifier le filtre inverse, nous ne pouvons pas nous contenter d'un critère quadratique.

Une propriété fondamentale pour l'identification du filtre inverse a été mise en évidence dans [25]. La sortie $s(n)$ du filtre inverse est obtenue par filtrage linéaire de l'entrée $e(n)$ formée d'une suite de variables aléatoires indépendantes de même loi. Lorsque l'entrée $e(n)$ n'est pas gaussienne, le seul filtre pour lequel la loi de probabilité (à un instant) de la sortie $s(n)$ du filtre inverse est identique à celle de l'entrée est le filtre dont tous les coefficients sont nuls sauf un. Ceci donne la solution du problème de la recherche du filtre inverse : il suffit de rendre la loi de la sortie du filtre inverse identique à celle de l'entrée. Les auteurs de [25] étudient alors les critères permettant de converger vers le filtre inverse. Pour obtenir le filtre inverse, nous devons prendre en compte la forme de la loi de probabilité du signal d'entrée. Nous verrons

au paragraphe 3.4.3 que la condition d’identité des lois peut être simplifiée en imposant seulement l’identité de la valeur quadratique moyenne et du kurtosis.

Le cas des signaux gaussiens est singulier. Comme le caractère gaussien se conserve par transformation linéaire, il n’est pas possible de réaliser l’identification aveugle du filtre inverse lorsque les signaux sont gaussiens.

Une analyse de ce problème d’inversion, permettant de caractériser le comportement des signaux sous l’effet d’un filtrage, est également donnée dans [65]. D. Donoho considère un ensemble de signaux aléatoires fermé vis-à-vis des transformations linéaires : toute combinaison linéaire (filtrage) de signaux de l’ensemble appartient à l’ensemble. Il montre que dans cet ensemble, il existe une relation d’ordre partiel, notée \geq^\bullet . Selon cette relation d’ordre, si $x(n)$ est un signal blanc parfait, $z(n)$ un signal gaussien et $y(n) = \sum_i a_i x(n-i)$ une filtrée de $x(n)$, on a alors

- $x(n) \geq^\bullet y(n) \geq^\bullet z(n)$,
- cet ordre est strict : si $y(n) = \sum_i a_i x(n-i)$ est équivalent à $x(n)$, tous les a_i sont nuls sauf un.

Ces deux résultats donnent la règle permettant de déterminer le filtre inverse : il faut trouver dans la famille des filtrées le signal qui est le plus éloigné d’un signal gaussien. Ces résultats établissent l’unicité du filtre inverse à un retard et à un facteur de proportionnalité près.

Une règle permettant de choisir les critères acceptables se dégage naturellement. Les critères $V(\cdot)$ acceptables doivent respecter la relation d’ordre introduite par D. Donoho :

$$x \geq^\bullet y \iff V(x) \geq V(y).$$

Les critères acceptables présentés par D. Donoho sont

- les cumulants standardisés (chapitre 1), en particulier le kurtosis [35] que l’on retrouvera en 3.4.3 et que nous avons utilisé dans l’illustration du paragraphe 3.2.5,
- l’information de Fisher

$$I(x) = - \int \frac{d^2 \log[p_x(u)]}{du^2} p_x(u) du,$$

- l’entropie de Shannon,

$$S(x) = - \int \log[p_x(u)] p_x(u) du,$$

qui avait été introduite auparavant par [201] dans la méthode du maximum d’entropie.

Cadzow [41], en généralisant les résultats de Shalvi et Weinstein [185], montre que la recherche des extremums des cumulants standardisés constitue un critère.

Rappelons l'équation (3.5) qui établit le lien entre la sortie $s(n)$ du filtre inverse, en fonction du bruit blanc parfait d'entrée $e(n)$:

$$s(n) = \sum_p t(p)e(n-p),$$

$e(n)$ étant une suite de variables aléatoires indépendantes. Les cumulants d'ordre k de $s(n)$ sont

$$\kappa_{s(k)} = \kappa_{e(k)} \sum_p t^k(p), \quad (3.6)$$

$\kappa_{e(k)}$ étant le cumulants d'ordre k de $e(n)$. Pour les cumulants standardisés (chapitre 1), cette relation devient

$$\mathcal{K}_{s(k)} = \mathcal{K}_{e(k)} \frac{\sum_p t^k(p)}{\left[\sum_p t^2(p)\right]^{k/2}}.$$

En normalisant la réponse impulsionnelle du filtre total par $\sum_p t^2(p) = 1$, on obtient

$$\mathcal{K}_{s(k)} = \mathcal{K}_{e(k)} \sum_p t^k(p).$$

On sait alors que sous la contrainte $\sum_p t^2(p) = 1$, la somme $\sum_p t^k(p)$ vérifie la condition

$$\left| \sum_p t^k(p) \right| \leq 1,$$

l'égalité ayant lieu si et seulement si tous les $t(p)$ sont nuls sauf un, qui est alors égal à ± 1 .

Ce résultat illustre la relation d'ordre introduite par Donoho [65]. Les cumulants d'ordre supérieur à 2 de toute filtrée linéaire du bruit blanc d'entrée $e(n)$ sont plus proches de 0 que les cumulants de ce bruit blanc d'entrée : une filtrée linéaire d'un bruit blanc est donc plus proche d'un signal gaussien que le bruit blanc initial. Cette propriété nous montre aussi que les cumulants standardisés respectent la relation d'ordre introduite par Donoho ; ils constituent donc un critère acceptable. Enfin ce résultat corrobore l'approche de Benveniste [25]. En effet pour que les signaux d'entrée $e(n)$ et de sortie $s(n)$ aient la même loi, il faut qu'ils aient les mêmes cumulants standardisés et donc que tous les $t(p)$ soient nuls sauf 1.

3.4.3 Identification aveugle par le kurtosis

En se limitant au kurtosis, cumulants d'ordre 4 standardisé, Shalvi et Weinstein ont développé une méthodologie très simple de déconvolution aveugle que nous allons présenter⁵.

Les bases de l'algorithme La situation est la suivante (figure (3.3)) :

$$o(n) = \sum_l d(n-l)e(l),$$

où $e(n)$ est une suite de variables aléatoires indépendantes (bruit blanc parfait) de même loi.

$$s(n) = \sum_l i(n-l)o(l)$$

est la sortie du filtre inverse. En combinant les deux relations précédentes on obtient

$$s(n) = \sum_l t(n-l)e(l),$$

où $t(n)$ est la réponse impulsionnelle du filtre faisant passer de l'entrée $e(n)$ à la sortie $s(n)$.

On veut que le vecteur $\mathbf{t}(n)$, en pratique de longueur finie, ait tous ses coefficients nuls sauf un : on aura alors réalisé l'inversion à un retard près et à un facteur multiplicatif près.

La relation (3.6) appliquée à la puissance (cumulant d'ordre 2) et aux kurtosis $\mathcal{K}_{s(4)}$ et $\mathcal{K}_{e(4)}$ de $s(n)$ et de $e(n)$ donne

$$\kappa_{s(2)} = \kappa_{e(2)} \sum_l t^2(l), \quad (3.7)$$

$$\mathcal{K}_{s(4)} = \mathcal{K}_{e(4)} \sum_l t^4(l). \quad (3.8)$$

Sous la contrainte $\sum_l t^2(l) = 1$ on obtient $|\mathcal{K}_{s(4)}| \leq |\mathcal{K}_{e(4)}|$, l'égalité ayant lieu si et seulement si tous les $t(l)$ sont nuls sauf 1. On en déduit la procédure permettant de trouver le filtre inverse :

maximiser $|\mathcal{K}_{s(4)}|$, en fonction des coefficients du filtre inverse, sous la contrainte :

$$E[|s(n)|^2] = E[|e(n)|^2].$$

L'algorithme proposé donne le filtre inverse cherché.

Revenons sur la contrainte $E[|s(n)|^2] = E[|e(n)|^2]$. On sait trouver un signal donnant la puissance de $e(n)$: c'est le signal blanchi. On sait par les

⁵Nous présentons l'algorithme de Shalvi et Weinstein pour des signaux à valeurs réelles, on peut l'étendre facilement aux signaux à valeurs complexes.

études à l'ordre 2 que le filtre blanchisseur peut être obtenu en minimisant $E[|s(n)|^2]$. La procédure proposée consiste donc à obtenir un filtre qui minimise $E[|s(n)|^2]$ et qui maximise $|\mathcal{K}_{s(4)}|$. Ces deux opérations peuvent être faites successivement ou simultanément. On en déduit deux techniques de mise en œuvre.

Mise en œuvre de l'algorithme Deux solutions sont proposées dans [185]. Un algorithme avec contrainte qui nécessite le blanchiment préalable des signaux observés, et un algorithme sans contrainte.

- Dans le premier cas on doit d'abord blanchir les données $o(n)$ pour obtenir un signal blanchi $o_b(n)$. On adapte ensuite les coefficients du filtre inverse en deux étapes :
 1. poursuite du maximum selon le gradient

$$\mathbf{i}(n+1) = \mathbf{i}(n) - \mu \nabla V(s),$$

2. vérification de la contrainte par

$$\mathbf{i}(n+1) \leftarrow \frac{\mathbf{i}(n+1)}{\|\mathbf{i}(n+1)\|} \quad \text{avec : } \|\mathbf{i}(n+1)\| = \sqrt{\mathbf{i}^T(n+1)\mathbf{i}(n+1)}.$$

Le critère retenu est $V(s) = |\mathcal{K}_{s(4)}|$. Tous calculs faits on obtient la récursion

$$\mathbf{i}(n+1) = \mathbf{i}(n) + \mu \text{sign}(\mathcal{K}_{e(4)}) E[|s(n)|^2 s(n) \mathbf{o}_b^*(n)],$$

avec $\mathbf{o}_b^T(n) = (o_b(n), o_b(n-1), \dots, o_b(n-p+1))$ – rappelons que $o_b(n)$ est le signal blanchi –.

- Le second algorithme part directement des signaux observés. La fonction objectif est dans ce cas du type suivant

$$F(\mathbf{t}) = \sum_l |t(l)|^4 + f\left(\sum_l |t(l)|^2\right),$$

telle que la fonction $h(x) = x^2 + f(x)$ croisse pour $x < 1$, décroisse pour $x > 1$ et ait un maximum unique pour $x = 1$. On montre alors que $F(\mathbf{t})$ est maximale pour un vecteur \mathbf{t} ayant une seule valeur non nulle⁶. On obtient la fonction objectif en remplaçant $\sum_l |t(l)|^2$ et $\sum_l |t(l)|^4$ par leurs expressions en fonction des cumulants de $e(n)$ et $s(n)$ donnés par les équations (3.7) et (3.8). Les auteurs montrent que l'on peut retrouver ainsi des algorithmes connus comme l'algorithme de Godard.

⁶La solution est unique pour un filtre inverse de longueur finie ; pour un filtre de longueur infinie, il peut apparaître des minimums parasites.

3.4.4 Les algorithmes de Godard et Sato

Godard [92] propose de minimiser un critère, appelé dispersion, qui s’écrit

$$D_p = E [(|s(n)|^p - R_p)^2] \quad \text{avec} \quad R_p = \frac{E [|e(n)|^{2p}]}{E [|e(n)|^p]}.$$

Pour $p = 1$ l’algorithme de gradient stochastique s’écrit

$$\mathbf{i}(n+1) = \mathbf{i}(n) - \mu \mathbf{o}^*(n) s(n) \left[1 - \frac{R_1}{|s(n)|} \right],$$

et pour $p = 2$ on obtient

$$\mathbf{i}(n+1) = \mathbf{i}(n) - \mu \mathbf{o}^*(n) s(n) [|s(n)|^2 - R_2]. \quad (3.9)$$

Dans l’algorithme de Sato [179] :

$$\mathbf{i}(n+1) = \mathbf{i}(n) - \mu \mathbf{o}(n) [s(n) - g \text{sign}\{s(n)\}],$$

avec

$$g = \frac{E [|e(n)|^2]}{E [|e(n)|]} = R_1.$$

Notons que la connaissance des statistiques du signal $e(n)$ est nécessaire pour le fonctionnement de ces algorithmes alors que la méthode fondée sur le kurtosis ne requiert que la connaissance du signe du kurtosis de $e(n)$.

3.4.5 Illustration

Pour illustrer les propriétés de l’identification aveugle par des systèmes en boucle nous présentons les résultats obtenus avec un algorithme de Godard en les comparant à un algorithme d’ordre 2.

Les signaux traités sont des séquences binaires pseudo-aléatoires (BPSK). Nous avons choisi cet exemple car il utilise des signaux très fréquents dans les systèmes de communication, et parce que l’algorithme de Godard est bien adapté à ces signaux.

La séquence émise $e(n)$ est présentée sur la figure (3.4-A). Le canal de transmission est un filtre passe-bas à phase non minimale. Le signal reçu après filtrage $o(n)$ est présenté figure (3.4-B). Nous mettons ensuite en œuvre les deux algorithmes.

L’algorithme “d’ordre 2” minimise la puissance du signal reçu. Écrivons le filtre MA , de longueur p , identifié par cet algorithme selon

$$s(n) = \mathbf{i}_1(n)^T \mathbf{o}_1(n),$$

avec $\mathbf{o}_1(n) = (o(n), \dots, o(n-p+1))$ et $\mathbf{i}_1(n) = (i_1(n), \dots, i_1(n-p+1))$. Cette écriture fait apparaître directement la contrainte $i_0 = 1$ qui évite la convergence

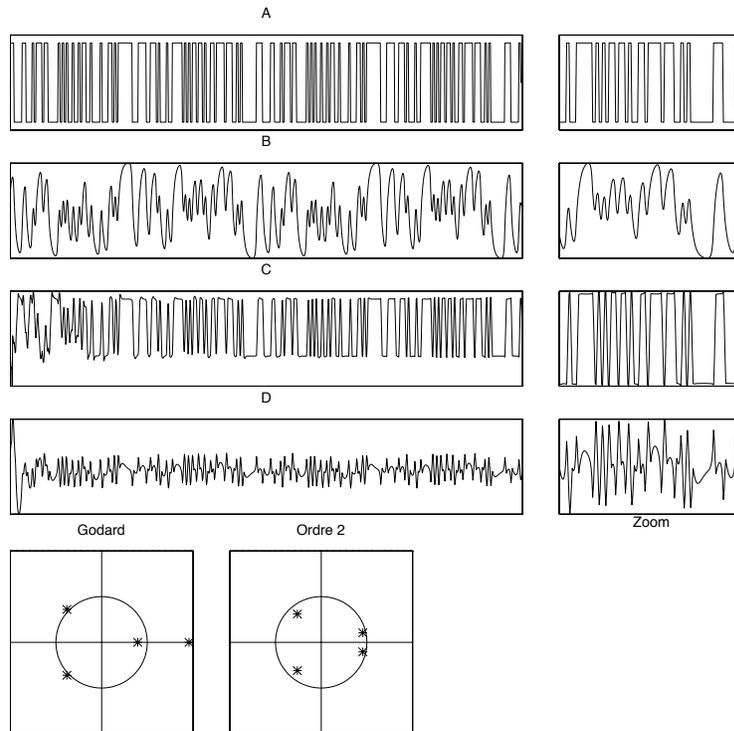


Figure 3.4: Comparaison de la déconvolution par l'algorithme de Godard et par un algorithme d'ordre 2. A : signal émis, B : signal reçu, C : signal estimé par l'algorithme de Godard, D : signal estimé à l'ordre 2. Les cadres du bas de la figures représentent la configuration des zéros estimés, à gauche par l'algorithme de Godard, à droite par l'algorithme d'ordre 2.

du filtre MA vers $\mathbf{0}$. La relation de récurrence donnant les coefficients du filtre est

$$\mathbf{i}_1(n+1) = \mathbf{i}_1(n) - \mu s(n)\mathbf{o}_1(n).$$

L'algorithme de Godard est implanté en utilisant la relation (3.9). Les résultats sont présentés sur la figure (3.4). Nous donnons en (C) le signal reconstruit par l'algorithme de Godard, en (D) le signal reconstruit par l'algorithme à l'ordre 2. Nous donnons également un zoom de ces différents signaux, zoom effectué sur la fin des signaux, après convergence. Enfin nous présentons les zéros du filtre inverse identifié par l'algorithme de Godard et les zéros du filtre inverse identifié par l'algorithme d'ordre 2.

On voit que l'algorithme de Godard, après une période de convergence, retrouve le signal émis⁷. L'algorithme à l'ordre 2 blanchit le signal observé mais ne l'inverse pas comme le montre la figure D. L'observation des zéros montre que l'algorithme de Godard trouve un zéro à l'extérieur du cercle unité. Par

⁷Dans cet exemple, le signal retrouvé est multiplié par -1 pour faciliter l'observation.

contre, l'algorithme d'ordre 2 place, comme prévu, tous les zéros dans le cercle unité.

On peut s'inquiéter du fait que les zéros obtenus par l'algorithme d'ordre 2 ne soient pas identiques, ou inverses par rapport au cercle unité, de ceux du filtre de Godard. Dans ce cas d'espèce les deux zéros sont "presque" inverses par rapport au cercle unité. L'algorithme d'ordre 2 "voit" donc un zéro "double" qu'il interprète, par suite des erreurs d'estimation, comme deux zéros complexes conjugués (on passe continûment d'un zéro double à deux zéros complexes conjugués).

3.5 Identification aveugle en fréquence

Les méthodes de déconvolution aveugle en fréquence ont été développées à l'origine par les astronomes pour réduire le brouillage atmosphérique des images fournies par des télescopes de grandes dimensions [133]. Elles sont également dues aux théoriciens du traitement du signal [131] et aux géophysiciens de la prospection sismique [145].

3.5.1 Principe

Le filtre F à identifier est excité par un bruit $\varepsilon(n)$ blanc à un ordre supérieur à 2. On veut réaliser l'identification du filtre en ne connaissant que sa sortie $o(n)$. A l'ordre 2, le spectre de la sortie est

$$S_{o(2)}(\lambda) = |H(\lambda)|^2 S_{\varepsilon(2)}(\lambda)$$

où $H(\lambda)$ est la fonction de transfert du filtre en fréquences réduites et $S_{\varepsilon(2)}(\lambda)$ est la densité spectrale de puissance du bruit blanc d'entrée. Cette relation nous montre qu'à l'ordre 2 on ne peut retrouver, à un facteur d'amplitude près, que le module du filtre F . Pour identifier le filtre il faut disposer d'informations sur sa phase : informations dites *a priori*. On est ainsi conduit à faire des hypothèses sur la phase, par exemple que la phase est nulle, minimale, ...

Pour un signal d'entrée blanc à l'ordre 3, le bispectre du signal de sortie s'écrit

$$S_{o(3)}(\lambda_1, \lambda_2) = H(\lambda_1)H(\lambda_2)H(-\lambda_1 - \lambda_2)S_{\varepsilon(3)}(\lambda_1, \lambda_2),$$

$S_{\varepsilon(3)}(\lambda_1, \lambda_2)$ étant le bispectre de l'entrée, constant puisque ce signal d'entrée est blanc d'ordre 3. La phase $\Phi(\lambda)$ du filtre F vérifie la relation

$$\Psi(\lambda_1, \lambda_2) = \Phi(\lambda_1) + \Phi(\lambda_2) + \Phi(-\lambda_1 - \lambda_2), \quad (3.10)$$

$\Psi(\lambda_1, \lambda_2)$ étant la phase de $S_{o(3)}(\lambda_1, \lambda_2)$. Lorsque les signaux sont à valeurs réelles, la phase du bispectre est impaire et l'on obtient

$$\Psi(\lambda_1, \lambda_2) = \Phi(\lambda_1) + \Phi(\lambda_2) - \Phi(\lambda_1 + \lambda_2). \quad (3.11)$$

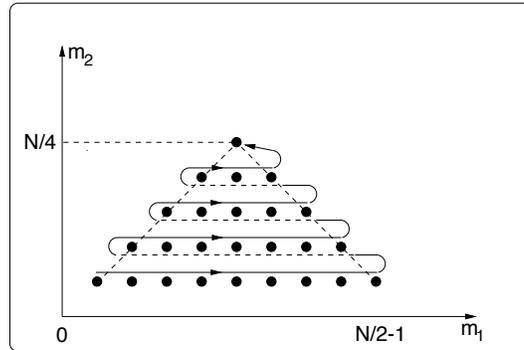


Figure 3.5: Les valeurs du bispectre utilisées pour l'inversion en fréquence. La trajectoire indiquée correspond à l'ordre d'empilement des données dans un vecteur pour la méthode des moindres carrés.

Ces relations sont à la base des techniques mises au point pour retrouver la phase du filtre F en utilisant le bispectre.

Ces techniques nécessitent que le bispectre de $x(n)$ soit non nul : on retrouve là le fait que l'on ne peut identifier ainsi un filtre excité par un bruit blanc gaussien. Si le signal d'entrée n'est pas gaussien il se peut que son bispectre soit nul par suite de la symétrie de sa densité de probabilité. On peut alors faire appel à des multispectres d'ordre supérieur : trispectre...

3.5.2 Mise en œuvre

Pour retrouver la phase du filtre à partir de la phase du bispectre, il faut inverser les relations (3.10) ou (3.11). Nous allons présenter les différentes techniques qui ont été développées.

Auparavant arrêtons nous sur la question du déroulement de la phase. Dans les relations (3.10) et (3.11) on additionne trois phases qui sont comprises entre $-\pi$ et $+\pi$. Le résultat est donc compris entre -3π et $+3\pi$. La mesure de la phase du bispectre donne la phase à 2π près. Certaines méthodes de reconstruction nécessitent que l'on "déroule" la phase du bispectre entre -3π et $+3\pi$. Des techniques de déroulement de la phase du bispectre ont été proposées par [143] et par [173].

Pratiquement l'estimation du bispectre fait appel à la TFD et l'on obtient la phase du bispectre en fréquences discrètes. Les valeurs du bispectre occupent une grille régulière dans le plan des bifréquences (figure (3.5)). On a vu au chapitre 2 que par suite des symétries du bispectre, il suffit de le calculer dans le triangle représenté sur la figure (3.5).

Méthode récursive linéaire La méthode de Lohman, Weigelt et Wirnitzer [133] fait appel aux valeurs de la phase du bispectre sur la droite $m_2 = 1$. Sur

cette droite, la relation (3.11) donne

$$\Phi(m+1) = \Phi(m) + \Phi(1) - \Psi(m, 1).$$

Cette relation de récurrence permet de calculer toutes les valeurs de $\Phi(m)$ à partir de $\Phi(1)$. La valeur de $\Phi(0)$ est nulle par suite de l'imparité de la phase du filtre. La valeur de $\Phi(1)$ correspond à une variation linéaire de la phase en fonction de la fréquence qui se traduit par un retard en temps. La phase $\Phi(1)$ est indéterminée car on obtient le filtre à un retard près.

Cette technique ne nécessite pas de déroulement de la phase. De plus, cette méthode n'utilise qu'une partie des valeurs mesurées de la phase du bispectre, ceci expliquant sa sensibilité au bruit.

Pour obtenir des mesures plus robustes on a développé des techniques utilisant l'ensemble des valeurs de la phase du bispectre : ce sont la méthode des moindres carrés [145], la méthode de projection [129, 130] et la méthode du retard de groupe [156].

Méthode des moindres carrés Dans cette méthode [145], on dispose les données (phase du bispectre) et les inconnues (phase du filtre) dans deux vecteurs. Le vecteur des inconnues est donné simplement par la suite des valeurs de la phase du filtre en fonction de la fréquence. Si le signal a N points en fréquence⁸, on construit le vecteur

$$\Phi = (\Phi(1) \dots \Phi(\frac{N}{2} - 1)).$$

Le vecteur des données est obtenu en balayant la suite des valeurs de la phase du bispectre contenues dans le triangle principal selon la trajectoire indiquée sur la figure (3.5). On obtient alors

$$\Psi = \left(\Psi(1, 1) \dots \Psi(1, N/2 - 1) \Psi(2, 2) \dots \Psi\left(\frac{N}{4}, \frac{N}{4}\right) \right).$$

Les relations entre les données et les inconnues sont obtenues en réécrivant (3.11) en fréquences discrètes, soit

$$\Psi(k, l) = \Phi(k) + \Phi(l) - \Phi(k + l).$$

La relation entre Ψ et Φ , qui s'en déduit est

$$\Psi = A \cdot \Phi,$$

la matrice A étant formée de 0, 1 et 2.

⁸En général N est une puissance de deux par suite de l'utilisation de la *FFT*.

Le vecteur des inconnues a $N/2 - 1$ composantes et le vecteur des données $(N/4)^2$ composantes. On résoud ce système linéaire surdéterminé par les moindres carrés. Le vecteur solution s'écrit alors

$$\Phi = (A^T A)^{-1} A^T \Psi.$$

Cette solution nécessite le déroulement de la phase du bispectre. Le résultat est nettement plus robuste vis-à-vis des incertitudes dans l'estimation de la phase du bispectre.

Méthode de la projection Elle est issue de la propriété de projection. Elle a, en particulier l'intérêt de s'appliquer aux multispectres d'ordre quelconque⁹. Elle présente également une bonne robustesse vis-à-vis des incertitudes d'estimation.

La propriété de projection a été établie par [130, 165]. Pour un multispectre d'ordre p considérons la phase du bispectre $\Psi_p(m, \mathbf{l})$, m représentant une multifréquence, et \mathbf{l} étant le vecteur à $p - 1$ composantes des $p - 1$ autres multifréquences. La relation de filtrage s'écrit

$$\Psi_p(m, \mathbf{l}) = \Phi(m) + \sum_i \Phi(l_i) - \Phi(m + l_1 + \dots + l_p).$$

En sommant cette relation sur toutes les composantes de \mathbf{l} on obtient, par suite des périodicités de Φ ,

$$\Phi(m) = \frac{1}{N^{p-1}} \sum_{\mathbf{l}} \Psi_p(m, \mathbf{l}). \quad (3.12)$$

Cette relation permet de calculer la phase du filtre à partir de la phase du multispectre.

Cette méthode de calcul donne l'estimation de la phase en moyenne quadratique [125, 129]. Cette propriété s'obtient en considérant le sous-ensemble des fonctions discrètes à plusieurs variables $\Psi_p(m, \mathbf{l})$ engendré par les fonctions de la forme $\Phi(m) + \sum_i \Phi(l_i) - \Phi(m + l_1 + \dots + l_p)$. Ce sous-ensemble est convexe et la relation (3.12) est une projection sur ce sous-ensemble. On sait, d'après le théorème de projection, que cette projection minimise l'erreur quadratique moyenne.

On a montré que la méthode de la projection est équivalente, du point de vue de la robustesse aux erreurs d'estimation, à la méthode des moindres carrés présentée plus haut. Ces deux méthodes utilisent toutes les mesures issues du multispectre. L'intérêt de la méthode de la projection est son application aux multispectres d'ordre supérieur. En effet, on rencontre souvent des signaux dont le bispectre est nul, par suite de la parité de la densité de probabilité du

⁹On pourrait aussi étendre à des ordres supérieurs à 3 la méthode des moindres carrés.

bruit blanc d'entrée. Dans cette situation on peut estimer la phase du filtre à partir du trispectre.

Enfin, comme la méthode des moindres carrés, la méthode de la projection nécessite le déroulement de la phase.

Méthode du retard de groupe Dans cette méthode, proposée par [156], on reconstruit la phase du signal à partir du retard de groupe, dérivée de la phase du signal par rapport à la fréquence.

En appelant $\Phi(\nu)$ la phase du signal et $\Psi(\nu_1, \nu)$ la phase du bispectre, on a

$$\Psi(\nu_1, \nu) = \Phi(\nu_1) + \Phi(\nu) - \Phi(\nu_1 + \nu);$$

en dérivant par rapport à ν_1 on obtient

$$\frac{\partial \Psi(\nu_1, \nu)}{\partial \nu_1} = \frac{\partial \Phi(\nu_1)}{\partial \nu_1} - \frac{\partial \Phi(\nu_1 + \nu)}{\partial \nu_1},$$

donnant pour $\nu_1 = 0$

$$\tau_{g,21}(0, \nu) = \tau_g(0) - \tau_g(\nu).$$

Dans cette relation,

$$\tau_g(\nu) = -\frac{1}{2\pi} \frac{d\Phi(\nu)}{d\nu}$$

est le retard de groupe du signal, et

$$\tau_{g,21}(\nu_1, \nu_2) = -\frac{1}{2\pi} \frac{\partial \Psi(\nu_1, \nu_2)}{\partial \nu_1}$$

est issu de la phase du bispectre.

On calcule $\tau_{g,21}(\nu_1, \nu_2)$ à partir de la phase du bispectre $e^{j\Psi(\nu_1, \nu_2)}$. Posons $F_1(\nu_1, \nu_2) = e^{j\Psi(\nu_1, \nu_2)}$, et soient $f_1(\tau_1, \tau_2)$ la transformée de Fourier inverse de $F_1(\nu_1, \nu_2)$ et $F_2(\nu_1, \nu_2)$ la transformée de Fourier de $\tau_1 f_1(\tau_1, \tau_2)$. Un calcul simple donne

$$-\frac{1}{2\pi} \frac{\partial \Psi}{\partial \nu_1} = \frac{j}{2\pi} F_1^* \frac{\partial F_1}{\partial \nu_1},$$

c'est-à-dire

$$\tau_{g,21}(\nu_1, \nu_2) = F_2(\nu_1, \nu_2) F_1^*(\nu_1, \nu_2),$$

conduisant au retard de groupe $\tau_g(\nu)$, à une constante non significative près. De $\theta(\tau) = TF^{-1}[\tau_g(\nu)]$, et de $\tau_g(\nu) = -\frac{1}{2\pi} \frac{d\Phi}{d\nu}$ on déduit

$$\Phi(\nu) = TF \left[j \frac{\theta(\tau)}{\tau} \right].$$

La procédure de calcul des dérivées, en passant dans le domaine dual, peut sembler lourde. On a montré qu'elle est optimale dans le calcul de

dérivées de signaux à bande limitée. Pratiquement les traitements sont faits en fréquences discrètes en utilisant la *TFD*.

Cette méthode de construction de la phase du signal a l'avantage d'utiliser toutes les valeurs de la phase du bispectre et de ne pas nécessiter de déroulement de sa phase.

3.5.3 Domaines d'application

Ces techniques d'identification aveugle ont été mises en œuvre dans des situations réelles, particulièrement en prospection sismique [131, 145] et en astronomie [175].

En propection sismique, un signal bref émis dans le sous-sol est réfléchi par les interfaces entre les différentes couches du sous-sol. La mesure des retards associés à chacune des réflexions permet de déterminer la profondeur des différentes couches. Cette mesure donne des informations utiles pour l'interprétation géologique de la structure du sous-sol. Elle est très utilisée, en particulier dans la recherche pétrolière. Le signal émis n'est généralement pas suffisamment bref. On cherche donc à l'inverser pour affiner la localisation temporelle des échos. Des méthodes d'ordre 2 ont été développées. Elles ne donnent pas de résultats satisfaisants lorsque l'on ne connaît pas la phase du signal émis. Les techniques d'identification aveugle présentées plus haut permettent de déterminer la phase du signal et de réaliser l'inversion du signal émis.

En astronomie, les images données par un télescope sont perturbées par les fluctuations aléatoires de l'atmosphère. Si on n'élimine pas ces perturbations, le pouvoir de résolution des grands télescopes, au delà de 1 mètre, est limité. On a montré que l'on peut s'affranchir de cette limitation en utilisant le bispectre. La technique employée consiste à moyenner des images avec un temps de pose court (1 ms). Cette durée de pose courte "gèle" les perturbations dues à l'atmosphère. On obtient ainsi une suite de réalisations de l'image "aléatorisée" par les fluctuations de l'atmosphère. A l'ordre 2, on calcule la fonction de corrélation de l'image qui permet de reconstituer son module [120]. A cet ordre la phase de l'objet est perdue. Comme nous l'avons montré plus haut la phase peut être reconstituée en calculant le bispectre par une moyenne (moyenne d'ensemble) sur les images en pose courte. Diverses solutions à ce problème de reconstitution de la phase ont été obtenues en utilisant les techniques présentées ci-dessus [134]. Les objets à reconstituer sont des fonctions dépendant de deux variables d'espace. Le bispectre est calculé dans le domaine des fréquences spatiales obtenues par transformation de Fourier sur les variables d'espace.

3.6 Identification aveugle paramétrique

Dans ces techniques, on modélise le filtre F par un des modèles classiques : AR , MA , $ARMA$ et l'on identifie les coefficients de ce filtre. Au premier abord, on peut penser que seule la partie MA , qui peut être à phase non minimale en restant stable, pose un problème nécessitant l'intervention des moments d'ordre supérieur. Il apparaît cependant intéressant d'utiliser ces techniques pour les filtres AR pour deux raisons :

- le signal peut être entaché d'un bruit gaussien inconnu qui perturbe l'estimation par les moments d'ordre 2 mais n'influence pas les estimateurs d'ordre supérieur¹⁰,
- on peut parfois rencontrer des filtres AR ayant des pôles extérieurs au cercle unité. Ces filtres sont instables s'ils sont causaux. Ils sont stables s'ils sont anticausaux. On obtient simplement ces filtres en "remontant" le temps.

Nous proposons de dénommer les filtres combinant deux filtres $AR-2$ causaux et anticausaux des filtres $BIAR-2$. Les filtres $BIAR-2$ ont deux pôles à l'intérieur du cercle de rayon 1 et deux pôles à l'extérieur du cercle de rayon 1. Ils ne peuvent pas être identifiés par des statistiques d'ordre 2. Les statistiques d'ordre supérieur permettent d'identifier ces filtres $BIAR$.

De nombreuses techniques d'identification paramétrique ont été développées à l'ordre 2 [101, 142]. En introduisant les statistiques d'ordre supérieur, on augmente considérablement la collection de méthodes possibles. Nous allons présenter un panorama de ces méthodes qui ne peut prétendre être exhaustif par suite du nombre important de possibilités, et du fait que ce domaine reste un secteur actif de recherche qui n'a sans doute pas encore livré tous ses secrets. Nous essayerons de nous concentrer sur les principes présidant à l'élaboration des principales méthodes. Nous distinguerons le cas des filtres AR , des filtres MA et des filtres $ARMA$. Auparavant, nous donnons des définitions et des relations utiles pour la suite.

3.6.1 Introduction

Lignes de multicorrélation-Spectres associés Quand on utilise les multicorrélations on obtient des fonctions dépendant de plusieurs variables : deux pour la bicorrélation $C_{x(3)}(n, m)$, trois pour la tricorrélation $C_{x(4)}(n, m, l)$. En général, l'ensemble de ces valeurs donne trop d'informations. On se limitera à certaines parties de l'espace des retards, ne dépendant que d'un seul

¹⁰Il faut être cependant conscient du fait que la présence d'un bruit gaussien disperse les estimateurs des statistiques d'ordre supérieur à 2.

paramètre : on appelle ces parties de la multicorrélation des lignes de multicorrélation. Giannakis et Mendel ont montré l'intérêt particulier des diagonales [89, 150]. Pour la multicorrélation d'ordre r du signal $x(n)$, la diagonale est

$$C_{x(r)}^d(k) = C_{x(r)}(k, \dots, k).$$

On peut associer aux diagonales des multicorrélations des spectres en Z ou en fréquences réduites qui ne dépendent que d'une variable fréquentielle. Nous les appellerons spectres diagonaux. Le spectre diagonal en Z d'ordre r de $x(n)$ est

$$S_{x(r)}^d(z) = \sum_k C_{x(r)}^d(k) z^{-k}.$$

En transformée en fréquences réduites, les spectres diagonaux sont obtenus en sommant les valeurs du multispectre sur un hyperplan dans l'espace des multifréquences. La relation

$$C_{x(r)}^d(k) = \int \dots \int S_{x(r)}(\boldsymbol{\lambda}) e^{2\pi j \sum_{i=1}^{r-1} \lambda_i k} d\boldsymbol{\lambda},$$

peut se réécrire, après le changement de variables $\lambda = \sum_{i=1}^{r-1} \lambda_i$

$$C_{x(r)}^d(k) = \int \left[\int \dots \int S_{x(r)}(\lambda - \sum_{i=2}^{r-1} \lambda_i, \lambda_2, \dots, \lambda_{r-1}) d\lambda_2 \dots d\lambda_{r-1} \right] e^{2\pi j \lambda k} d\lambda,$$

qui conduit à

$$S_{x(r)}^d(\lambda) = \int \dots \int S_{x(r)}(\lambda - \sum_{i=2}^{r-1} \lambda_i, \lambda_2, \dots, \lambda_{r-1}) d\lambda_2 \dots d\lambda_{r-1}.$$

Cette relation nous montre que le spectre diagonal d'ordre r est obtenu en sommant le multispectre d'ordre r sur l'hyperplan $\sum \lambda_i = \lambda$. Cet hyperplan contient les fréquences qui interagissent dans une transformation non linéaire d'ordre r . On peut interpréter le spectre diagonal d'ordre r comme un résumé des propriétés d'ordre r du signal...

On utilise également des lignes de multicorrélation obtenues en fixant toutes les valeurs des retards sauf une. La ligne de $C_{x(r)}(k_1, k_2, \dots, k_{r-1})$ obtenue en fixant

$$\begin{cases} k_2 & = & l_2 \\ \vdots & & \vdots \\ k_{r-1} & = & l_{r-1} \end{cases},$$

sera notée

$$C_{x(r)}^l(k) = C_{x(r)}(k, l_2, \dots, l_{r-1}),$$

et on lui associera un spectre de ligne d'ordre r défini par

$$S_{x(r)}^{\mathbf{l}}(z) = TZ[C_{x(r)}^{\mathbf{l}}(k)].$$

Le vecteur \mathbf{l} contient les positions de la ligne de multicorrélation. On pourrait définir r corrélations de lignes à l'aide d'un seul vecteur \mathbf{l} , en coupant suivant différentes variables. Mais étant données les symétries des multicorrélations de signaux à valeurs réelles, ces r corrélations de ligne sont égales. Nous convenons donc de toujours couper selon la première variable de la multicorrélation.

Des relations utiles Soit $h(k)$ la réponse impulsionnelle d'un filtre que nous cherchons à identifier, l'entrée $x(n)$ est un bruit blanc, au moins jusqu'à un certain ordre, la sortie est $y(n)$ et s'écrit

$$y(n) = \sum_i h(n-i)x(i).$$

La diagonale de la multicorrélation d'ordre r de $y(n)$ est

$$C_{y(r)}^d(k) = \underbrace{Cum[y(n), y(n+k), \dots, y(n+k)]}_{r-1 \text{ fois}} = C_{x(r)} \sum_n h(n)h^{r-1}(n+k),$$

$C_{x(r)}$ étant la valeur de la multicorrélation d'ordre r de $x(n)$ à l'origine, qui est la seule valeur non nulle de cette multicorrélation, en supposant $x(n)$ blanc d'ordre supérieur ou égal à r . Le spectre diagonal d'ordre r de $y(n)$ est alors

$$S_{y(r)}^d(z) = C_{x(r)} \sum_k \sum_n h(n)h^{r-1}(n+k)z^{-k}.$$

Soit $H(z)$ la transformée en Z de la réponse impulsionnelle $h(n)$ du filtre. Introduisons

$$H_{r-1}^d(z) = TZ[h^{r-1}(p)].$$

Le spectre diagonal d'ordre r de $y(n)$ devient

$$S_{y(r)}^d(z) = C_{x(r)} H^*\left(\frac{1}{z^*}\right) H_{r-1}^d(z). \quad (3.13)$$

En particulier, le spectre de $y(n)$ est :

$$S_{y(2)}(z) = C_{x(2)} H^*\left(\frac{1}{z^*}\right) H(z),$$

En isolant dans ces deux dernières relations le terme $H^*(1/z^*)$, on déduit la relation, utile pour l'identification des filtres *MA*

$$H_{r-1}^d(z) S_{y(2)}(z) = \frac{C_{x(2)}}{C_{x(r)}} H(z) S_{y(r)}^d(z). \quad (3.14)$$

Les spectres de lignes vérifient une relation analogue

$$H_{r-1}^1(z^{-1})S_{y(2)}^1(z) = \frac{C_{x(2)}}{C_{x(r)}} H^*\left(\frac{1}{z^*}\right) S_{y(r)}^1(z), \quad (3.15)$$

avec $H_{r-1}^1(z) = TZ[h(n)h(n+l_2)\dots h(n+l_{r-1})]$.

Lorsque le filtre étudié est un filtre *AR* causal, appelons $A(z)$ la transformée en Z de la suite de ses coefficients $a(k)$

$$A(z) = \sum_{k=0}^p a(k)z^{-k}.$$

La fonction de transfert en Z , $H(z)$, et le spectre sont reliés à $A(z)$ par

$$H(z) = \frac{1}{A(z)},$$

et

$$S_{y(2)}^1(z) = \frac{C_{x(2)}}{A(z)A^*\left(\frac{1}{z^*}\right)},$$

d'où l'on déduit en reportant dans la relation (3.15) que

$$A(z)S_{y(r)}^1(z) = C_{x(r)}H_{r-1}^1(z^{-1}).$$

Par transformée en Z inverse, cette relation devient

$$\sum_{n=0}^p a(n)C_{y(r)}^1(k-n) = C_{x(r)}h(-k)h(-k+l_2)\dots h(-k+l_{r-1}).$$

Pour un filtre causal la réponse impulsionnelle est nulle pour $k > 0$. On en déduit

$$\sum_{n=0}^p a(n)C_{y(r)}^1(k-n) = 0 \quad \text{pour } k > 0. \quad (3.16)$$

Ces relations généralisent les relations de Yule-Walker aux ordres supérieurs.

3.6.2 Identification d'un filtre *MA*

Seuls les filtres *MA* ayant tous leurs zéros dans le cercle de rayon 1 (filtres à phase minimale) sont identifiables à l'ordre 2. Dans le cas général, on doit faire appel à des statistiques d'ordre supérieur. La relation entre l'entrée $x(n)$ et la sortie $y(n)$ d'un filtre *MA* d'ordre q est

$$y(n) = \sum_{i=0}^q b(i)x(n-i). \quad (3.17)$$

On normalise le filtre par $b(0) = 1$.

La première technique prouvant l'identifiabilité des filtres MA par les statistiques d'ordre supérieur est l'algorithme $C(q, k)$. On a également proposé des méthodes algébriques, des méthodes cepstrales, et des méthodes d'optimisation.

La formule $C(q, k)$ Giannakis [87] a donné des formules explicites montrant la faisabilité de l'identification aveugle d'un filtre MA . Prenons le cas de l'identification par la bicorrélation $C_{y(3)}(n, m)$. On a

$$C_{y(3)}(n, m) = C_{x(3)} \sum_{i=0}^q b(i)b(i+n)b(i+m),$$

avec $b(k) = 0$ si $k < 0$ ou $k > q$. En faisant $n = q$, soit $i = 0$ pour le seul terme restant, et en tenant compte de la normalisation de $b(0)$, on obtient

$$C_{y(3)}(q, k) = C_{x(3)} b(q)b(k) \quad \text{pour } k \leq q,$$

avec en particulier $C_{y(3)}(q, 0) = C_{x(3)} b(q)$. On déduit alors la formule dite de $C(q, k)$

$$b(k) = \frac{C_{y(3)}(q, k)}{C_{y(3)}(q, 0)},$$

donnant une expression explicite des coefficients du filtre MA à partir de la bicorrélation.

Ce résultat est intéressant sur le plan théorique car il établit l'identifiabilité "aveugle" d'un filtre MA .

On peut étendre cette relation à des ordres supérieurs à 3. En particulier, on obtient à l'ordre 4

$$b(k) = \frac{C_{y(4)}(q, q, k)}{C_{y(4)}(q, q, 0)},$$

$C_{y(4)}(i, j, k)$ étant la tricorrélation.

En pratique ces expressions donnant les coefficients du filtre MA sont peu stables par suite des erreurs statistiques intervenant dans l'estimation des multicorrélations. Pour améliorer la stabilité statistique, P. Comon [56] propose d'utiliser un ensemble de valeurs de la multicorrélation. Il détermine les coefficients du filtre MA en résolvant le système surdéterminé :

$$b(k)C_{y(4)}(l, m, q) - b(l)C_{y(4)}(k, m, q) = 0 \quad \text{pour } 0 \leq k \leq l \leq q \text{ et } 0 \leq m < q.$$

On a mis également au point des algorithmes faisant appel à des méthodes d'algèbre linéaire qui présentent une meilleure stabilité statistique.

Les méthodes algébriques Elles sont présentées dans [89, 150, 194, 206]. La relation (3.14)

$$H_{r-1}^d(z)S_{y(2)}(z) = \frac{C_{x(2)}}{C_{x(r)}}H(z)S_{y(r)}^d(z),$$

appliquée à la bicorrélation, donne

$$H_2^d(z)S_{y(2)}(z) = \frac{C_{x(2)}}{C_{x(3)}}H(z)S_{y(3)}^d(z), \quad (3.18)$$

avec $H_2^d(z) = H(z) * H(z) = TZ[h^2(n)]$. La réponse impulsionnelle $h(n)$ est donnée par les coefficients du filtre, soit

$$\begin{aligned} h(n) &= b(n) \quad \text{pour } n \leq q \\ h(n) &= 0 \quad \text{pour } n > q. \end{aligned}$$

Comme

$$S_{y(3)}^d(z) = TZ \left[C_{y(3)}^d(k) \right] = TZ \left[C_{y(3)}(k, k) \right],$$

par transformée en Z inverse de (3.18) on obtient, avec $b(0) = 1$

$$C_{y(2)}(k) + \sum_{n=1}^q b^2(n)C_{y(2)}(k-n) = \frac{C_{x(2)}}{C_{x(3)}} \left[C_{y(3)}^d(k) + \sum_{n=1}^q b(n)C_{y(3)}^d(k-n) \right]. \quad (3.19)$$

En concaténant ces relations pour $k = -q, \dots, 0, \dots, 2q$ on peut construire le système linéaire

$$C \cdot \mathbf{b} = \mathbf{r}.$$

Dans cette relation,

- C est une matrice construite avec les valeurs de $C_{y(2)}(n)$ et $C_{y(3)}^d(n) = C_{y(3)}(n, n)$,
- $\mathbf{b}^T = (\alpha b(1), \dots, \alpha b(q), b^2(1), \dots, b^2(q))$ avec $\alpha = C_{x(2)}/C_{x(3)}$,
- $\mathbf{r}^T = (\alpha C_{y(3)}^d(-q) - C_{y(2)}(-q), \dots, \alpha C_{y(3)}^d(2q) - C_{y(2)}(2q))$.

On déduit le vecteur \mathbf{b} par inversion selon la méthode des moindres carrés. Cette méthode est loin d'être optimale. Elle est même assez surprenante puisqu'elle calcule indépendamment les $b(k)$ et les $b^2(k)$.

Zheng et ses co-auteurs [206] proposent des méthodes algébriques, plus orthodoxes, fondées sur un jeu de relations linéaires. Le filtre direct est

$$y(n) = \sum_{i=0}^q b(i)x(n-i). \quad (3.20)$$

Le filtre inverse cherché est un filtre *MA* représenté par :

$$x(n) = \sum_{i=r_1}^{r_2} c(i)y(n-i), \quad (3.21)$$

avec $r_1 < 0$ et $r_2 > 0$. On normalise $c(i)$ en posant $c(0) = 1$. La partie à phase minimale est représentée par la partie causale du filtre ($i > 0$) et la partie à phase maximale par la partie anti-causale ($i < 0$). Les modules des coefficients aux retards non nuls sont inférieurs à 1.

Les méthodes algébriques sont fondées sur les relations suivantes entre les coefficients du filtre inverse et les éléments du bispectre ou du trispectre du signal observé :

$$\begin{cases} \frac{1}{C_{x(2)}} \sum_{j=r_1}^{r_2} c(j)C_{y(2)}(i+j) = b(i) & \text{si } 0 \leq i \leq q \\ = 0 & \text{sinon} \end{cases} \quad (3.22)$$

et

$$\begin{cases} \frac{1}{C_{x(3)}} \sum_{j=r_1}^{r_2} c(j)C_{y(3)}^d(i+j) = b^2(i) & \text{si } 0 \leq i \leq q \\ = 0 & \text{sinon} \end{cases} \quad (3.23)$$

Ces relations s'obtiennent en calculant de deux façons, à partir des relations (3.20) et (3.21) les cumulants $Cum[x(k), y(k+i)]$ et $Cum[x(k), y^2(k+i)]$. Deux schémas sont alors proposés :

- le premier utilise seulement les équations issues des relations (3.22) et (3.23) dont les seconds membres sont nuls. Parmi les $2(r_2 - r_1 + q)$ équations non triviales possibles, on retient les $2(r_2 - r_1 - q)$ relations issues de (3.22) et (3.23) avec

$$i = -q - r_2, \dots, -1, q + 1, \dots, q - r_1.$$

En tenant compte de la normalisation de $c(0)$ on obtient un système linéaire surdéterminé

$$A \cdot \mathbf{c} = \mathbf{b},$$

la matrice A et les vecteurs \mathbf{c} et \mathbf{b} sont construits à partir des relations (3.22) et (3.23).

- le second schéma utilise l'ensemble des équations (3.22) et (3.23) en éliminant les termes quadratiques de l'équation (3.23) grâce à la relation (3.19) qui s'écrit, avec $b(0) = 1$,

$$C_{x(3)} \sum_{i=0}^q C_{y(2)}(n-i)b^2(i) = C_{x(2)} \sum_{i=0}^q C_{y(3)}^d(n-i)b(i). \quad (3.24)$$

En reportant (3.22) et (3.23) dans (3.24) on obtient¹¹

$$\sum_{j=r_1, j \neq 0}^{r_2} f(k, j)c(j) = g(k), \quad k \in [-q, 2q],$$

avec

$$f(k, j) = \sum_{i=0}^q [C_{y(2)}(k-i)C_{y(3)}^d(i+j) - C_{y(3)}^d(k-i)C_{y(2)}(i+j)]$$

et

$$g(k) = \sum_{i=0}^q [C_{y(2)}(i)C_{y(3)}^d(k-i) - C_{y(3)}^d(i)C_{y(2)}(k-i)].$$

En retenant les relations

$$\sum_{j=r_1, j \neq 0}^{r_2} c(j)C_{y(2)}(k+j) = -C_{x(2)}(k) \quad k \in [-r_2, -1] \cup [q-r_1, q+1]$$

et

$$\sum_{j=r_1, j \neq 0}^{r_2} f(k, j)c(j) = g(k), \quad k \in [0, q],$$

on obtient un système linéaire surdéterminé qui donne les coefficients du filtre inverse.

Lorsque le signal étudié a une bicorrélation nulle on ne peut pas appliquer ces méthodes d'identification. On peut alors utiliser la tricorrélation. Les techniques décrites ci-dessus s'étendent facilement au quatrième ordre à partir de la relation :

$$\begin{cases} \frac{1}{C_{x(4)}} \sum_{j=r_1}^{r_2} c(j)C_{y(4)}^d(i+j) = b^3(i) & \text{si } 0 \leq i \leq q \\ = 0 & \text{sinon} \end{cases} \quad (3.25)$$

et de la relation (3.19) à l'ordre 4.

Les méthodes d'optimisation Le principe de ces méthodes est le suivant. Lorsque l'on cherche un filtre, on peut obtenir une solution par les méthodes d'ordre 2; cette solution n'a pas dans tous les cas la bonne loi de phase. L'indétermination vient du fait que les zéros du filtre *MA* identifié peuvent occuper deux positions inverses par rapport au cercle de rayon 1. Les méthodes à l'ordre 2 sont insensibles au remplacement d'un zéro par son inverse par rapport au cercle de rayon 1. Ayant obtenu, par une méthode d'ordre 2, *M*

¹¹Rappelons que $c(0) = 1$.

zéros à l'intérieur du cercle unité on choisit, parmi les 2^M solutions possibles, le filtre qui maximise une fonction objectif définie à partir des statistiques d'ordre supérieur. Cette méthode de recherche exhaustive peut être lourde si le nombre des zéros est important. Les différentes solutions proposées dépendent du choix de la fonction objectif.

Lii et Rosenblatt [131] minimisent la norme quadratique de l'écart entre les trispectres des sorties des différents filtres et le trispectre estimé du signal observé

$$\sum_l \sum_m \sum_n \left[C_{y(4)}(l, m, n) \widehat{b}(i) - \widehat{C}_{y(4)}(l, m, n) \right]^2.$$

Le critère utilisé par Lii et Rosenblatt, qui fait appel à l'ensemble des valeurs du trispectre est fortement redondant. On a vu au paragraphe 3.2.5 qu'il suffit de maximiser le kurtosis pour sélectionner le filtre optimal. Chi et Kung [52] ont proposé de maximiser le module du trispectre à l'origine. Dans la même ligne, en s'appuyant sur les résultats de Shalvi et Weinstein [185], M. Boumahdi a proposé de maximiser le kurtosis de la sortie [35].

Ces méthodes d'optimisation ont le grand intérêt d'utiliser au maximum les possibilités des statistiques d'ordre 2. On peut penser qu'elles conduisent ainsi à une meilleure stabilité statistique. La présence d'un bruit gaussien important donne, par contre, un solide argument pour utiliser exclusivement les statistiques d'ordre supérieur à 2.

Kammeyer et Jelonnek [108, 114] proposent de lever l'ambiguïté sur la position des zéros en utilisant simultanément le spectre en Z , issue d'une technique d'ordre 2, et une transformée en Z de la diagonale de la tricorrélation. Appelons $H(z)$ la fonction de transfert en Z cherchée. Le spectre est

$$S_{y(2)}(z) = C_{x(2)} H(z) H^*\left(\frac{1}{z^*}\right), \quad (3.26)$$

Soit $C_{y(4)}^d(k) = C_{y(4)}(k, k, k)$ la diagonale de la tricorrélation et $S_{y(4)}^d(z) = TZ \left[C_{y(4)}^d(k) \right]$ le spectre diagonal. La relation (3.13, page 123) s'écrit ici

$$S_{y(4)}^d(z) = C_{x(4)} H^*\left(\frac{1}{z^*}\right) H_3^d(z), \quad (3.27)$$

avec $H_3^d(z) = TZ[h^3(n)]$. Les zéros non communs à (3.26) et (3.27) donnent les zéros de $H(z)$...

Les méthodes cepstrales. Les multicepstres généralisent le cepstre défini à l'ordre 2 [164, 165]. Pour simplifier les notations nous allons présenter le cas de l'ordre 3 mais il est facile de généraliser ces résultats. Soit $S_{y(3)}(z_1, z_2)$ le bispectre en Z de $y(n)$

$$S_{y(3)}(z_1, z_2) = TZ \left[C_{y(3)}(k_1, k_2) \right] = \sum_{k_1=-\infty}^{+\infty} \sum_{k_2=-\infty}^{+\infty} C_{y(3)}(k_1, k_2) z_1^{-k_1} z_2^{-k_2}.$$

Le bicepstre $D_{y(3)}(l_1, l_2)$ de $y(n)$ est la transformée en Z inverse du logarithme du bispectre

$$D_{y(3)}(l_1, l_2) = TZ^{-1} \left[\log \left(S_{y(3)}(z_1, z_2) \right) \right].$$

Soit $h(k)$ la réponse impulsionnelle du filtre que nous cherchons à identifier. Le logarithme du bispectre du signal de sortie est

$$\log \left(S_{y(3)}(z_1, z_2) \right) = \log \left[C_{x(3)} \right] + \log [H(z_1)] + \log [H(z_2)] + \log [H(z_1^{-1}z_2^{-1})],$$

donnant, par transformée en Z inverse, le bicepstre de $y(n)$

$$\begin{aligned} D_{y(3)}(l_1, l_2) &= \log[C_{x(3)}] \delta(l_1) \delta(l_2) + D_h(l_1) \delta(l_2) \\ &+ D_h(l_2) \delta(l_1) + D_h(-l_2) \delta(l_1 - l_2), \end{aligned}$$

où $D_h(l)$ est le cepstre de la réponse impulsionnelle cherchée. On voit que le bicepstre est localisé sur les axes et sur la bissectrice, et qu'il donne une mesure du bicepstre $D_h(l) = TZ^{-1}[\log H(z)]$ de la réponse impulsionnelle. On déduit de là le module et la phase de la fonction de transfert $H(z)$, ce qui permet d'identifier le filtre. Dans la détermination de la phase de la fonction de transfert, on doit veiller à rétablir la continuité de cette phase qui est connue à 2π près (déroulement de la phase).

3.6.3 Identification d'un filtre AR

Pour un filtre AR stable et causal, tous les pôles sont situés à l'intérieur du cercle de rayon 1, le filtre est à phase minimale et son identification peut être faite par les formules de Yule-Walker [142]. Cette procédure peut être étendue aux ordres supérieurs. La relation entre la sortie $y(n)$ et l'entrée $x(n)$ du filtre AR est

$$\sum_{k=0}^p a(k)y(n-k) = x(n) \quad a(0) = 1.$$

La relation (3.16) appliquée à la corrélation donne les équations de Yule-Walker

$$\sum_{k=0}^p a(n)C_{y(2)}(k-n) = 0 \quad \text{pour } k > 0.$$

On obtient des relations analogues avec les lignes de multicorrélation. Pour les lignes de la bicorrélation, définies en 3.6.1

$$\sum_{k=0}^p a(n)C_{y(3)}^1(k-n) = 0 \quad \text{pour } k > 0 \quad \forall l_2.$$

En concaténant $p + M$ équations pour différentes valeurs de k on obtient une relation matricielle, exactement comme dans les méthodes à l'ordre 2

$$C \cdot \mathbf{a} = 0,$$

où \mathbf{a} est le vecteur des coefficients du filtre AR , et C la matrice $(p + M) \times (p + M)$ formée des lignes de la multicorrélation pour les différentes valeurs de k . Cette procédure se généralise facilement à un ordre quelconque. Une question importante, résolue dans [194], est de savoir combien il faut prendre d'équations pour que la matrice C soit de rang supérieur ou égal à p , et s'il est préférable de choisir certaines lignes particulières (en fixant l_2).

Pour un filtre AR causal on peut généraliser les méthodes développées au second ordre. Dans ce cas, l'intérêt des procédures d'ordre supérieur est qu'elles permettent théoriquement d'éliminer un bruit gaussien blanc ou non blanc. Ceci résulte de la propriété d'additivité des cumulants et de la nullité des cumulants d'ordre supérieur pour un signal gaussien. Il faut quand même souligner que les valeurs des lignes de la multicorrélation utilisées pour construire la matrice sont estimées. Les lignes estimées seront affectées par le bruit gaussien et également par des erreurs d'estimation des cumulants du signal. La précision des estimateurs de \mathbf{a} , que l'on appelle souvent stabilité statistique, dépend de ces erreurs d'estimation. Il existe encore peu de résultats théoriques sur ce sujet. Des simulations ont cependant montré que, dans le cas bruité en particulier, l'estimateur utilisant les corrélations d'ordre supérieur à 2 peut être plus précis que l'estimateur classique d'ordre 2.

Les pôles des filtres AR peuvent être extérieurs au cercle unité. Ces filtres conservent la propriété de stabilité s'ils sont anti-causaux : filtres $BIAR$. Le problème du choix des pôles se pose donc également pour les filtres AR . Une méthode d'optimisation a été proposée par M. Boumahdi [35]. La méthode de Yule-Walker à l'ordre 2 place tous les pôles à l'intérieur du cercle de rayon 1 (En effet la fonction de corrélation est insensible au remplacement d'un pôle par son inverse par rapport au cercle de rayon 1). On choisit ensuite la localisation des pôles par optimisation.

3.6.4 Identification d'un filtre $ARMA$

Deux types de filtres $ARMA$ doivent être considérés :

- les filtres $ARMA$ sans filtre de phase,
- les filtres $ARMA$ avec filtre de phase.

Nous appelons filtre de phase un filtre déphaseur pur. Le gain complexe en Z d'un filtre de phase a un pôle et un zéro inverses par rapport au cercle de rayon 1. Les méthodes d'ordre 2 sont totalement insensibles à la présence de ces termes dans le gain complexe d'un filtre $ARMA$. Deux familles de techniques d'identification des filtres $ARMA$ ont été développées [71]:

- l'identification directe dans laquelle on considère le filtre *ARMA* comme un filtre *AR* en série avec un filtre *MA* et l'on identifie successivement la partie *AR* et la partie *MA* : méthode *AR – MA*
- la représentation du filtre *ARMA* par un filtre à phase minimale en série avec un filtre de phase; on identifie successivement le filtre à phase minimale et le filtre de phase : méthode *MP – FP*.

Enfin nous distinguerons deux situations :

- dans le premier cas nous supposons la partie *AR* causale. Tous les pôles sont alors à l'intérieur du cercle de rayon 1, la partie *MA* pouvant être, ou non, à phase minimale.
- dans le second cas la partie *AR* pourra être non causale : il pourra donc y avoir des pôles extérieurs au cercle de rayon 1.

Lorsque le filtre *ARMA* est stable et causal les pôles du filtre *AR* sont à l'intérieur du cercle de rayon 1.

Dans la méthode *AR – MA* on identifie la partie *AR* par les équations de Yule-Walker modifiées¹². S'il n'y a pas de filtre de phase dans le filtre *ARMA* identifié, on peut utiliser les équations de Yule-Walker d'ordre 2. S'il y a des filtres de phase dans le filtre *ARMA* identifié, on doit utiliser les équations de Yule-Walker modifiées d'ordre supérieur à 2. Le signal est ensuite filtré par le filtre *MA*, inverse du filtre *AR* identifié. Le signal obtenu a un modèle *MA* auquel on applique une des techniques d'identification des filtres *MA* décrites plus haut. Le filtre *MA* ainsi identifié peut-être, ou non, à phase minimale.

Dans la méthode *MP – FP* le filtre *ARMA* spectralement équivalent, à phase minimale, est identifié par une des méthodes classiques à l'ordre 2. On doit ensuite identifier le filtre de phase dont les pôles sont à l'intérieur du cercle unité et dont les zéros sont inverses des pôles par rapport au cercle de rayon 1. Giannakis et Mendel [89] identifient la partie *AR* du filtre de phase par les équations de Yule-Walker modifiées d'ordre supérieur à 2. Les zéros se déduisent des pôles par inversion par rapport au cercle de rayon 1. On a également proposé des méthodes d'optimisation. Ayant obtenu le filtre à phase minimale spectralement équivalent on applique au signal les différents filtres possibles obtenus en remplaçant les zéros par leur inverse par rapport au cercle de rayon 1. On maximise ensuite une fonction objectif : valeur absolue de la bicorrélation à l'origine [52], valeur absolue du kurtosis [35].

La partie *AR* peut être anti-causale (pôles extérieurs au cercle de rayon 1). Cette situation n'a pas été très étudiée. La méthode *MP-FP* peut-être appliquée. On identifie tout d'abord le filtre stable, causal et à phase minimale

¹²Dans cette méthode on utilise les équations de Yule-Walker à partir d'un retard supérieur à l'ordre de la partie *MA*.

spectralement équivalent par une des méthodes connues à l'ordre 2. On cherche ensuite la localisation, à l'intérieur ou à l'extérieur du cercle de rayon 1, de chaque zéro et de chaque pôle, en maximisant une fonction objectif.

3.6.5 Conclusion

En étendant l'arsenal des méthodes développées à l'ordre 2, de nouveaux algorithmes d'identification ont été proposés. Le progrès essentiel concerne les filtres qui ne sont pas à phase minimale ou qui sont anti-causaux. Dans ce cas, il n'y a pas de solution à l'ordre 2 et les ordres supérieurs sont absolument nécessaires pour identifier le filtre sans faire d'hypothèse sur sa phase. L'art de l'identification des systèmes linéaires et stationnaires par les moments d'ordre supérieur est dans l'enfance, et de nombreuses questions se posent encore, comme par exemple le choix des lignes de multicorrélation, l'identifiabilité des filtres *AR*, la stabilité statistique des estimations...

La recherche active et productive sur les méthodes d'ordre 2 a duré environ 20 ans : nous avons donc encore du travail à faire sur les ordres supérieurs.

Avant de conclure ce chapitre revenons sur les méthodes d'estimation des multispectres qui se déduisent des techniques d'identification présentées ci-dessus.

3.7 Modélisation et estimation

A l'ordre 2 de nombreuses méthodes d'estimation du spectre ont été construites à partir de la modélisation systémique des signaux. Nous allons étendre ici ces techniques aux multispectres. Nous verrons que, comme à l'ordre 2, il existe des techniques particulières pour les signaux harmoniques.

3.7.1 Modélisation systémique

Les techniques d'identification développées ci-dessus représentent un signal par la sortie d'un filtre linéaire, *AR*, *MA* ou *ARMA* excité par un signal d'entrée "simple" : bruit blanc jusqu'à l'ordre considéré. Comme dans le cas de l'analyse spectrale, à l'ordre 2, cette modélisation permet d'estimer les multispectres en utilisant la relation de filtrage donnant le multispectre de sortie en fonction du multispectre de l'entrée et de la fonction de transfert du filtre. En prenant l'exemple du bispectre nous partons de

$$S_{y(3)}(\lambda_1, \lambda_2) = S_{x(3)} H(\lambda_1) H(\lambda_2) H^*(\lambda_1 + \lambda_2),$$

$S_{x(3)}$ est le bispectre, constant, de l'entrée blanche $x(n)$.

La fonction de transfert est, dans le cas $ARMA(p, q)$:

$$H(\lambda) = \frac{\sum_0^q b(l)e^{-2\pi jl\lambda}}{1 + \sum_1^p a(k)e^{-2\pi jk\lambda}}.$$

Cette méthode d'estimation conduit à des estimateurs dont on peut penser qu'ils permettront, comme cela est le cas dans l'analyse spectrale, de donner des estimateurs plus stables et plus résolvents. Le développement et la caractérisation de ces techniques est un des axes de recherche ouverts aux statistiques d'ordre supérieur.

3.7.2 Signaux harmoniques

On sait qu'à l'ordre 2, les méthodes de modélisation sont mal adaptées aux signaux formés de sommes de fréquences pures : signaux harmoniques. Dans cette situation, des méthodes spécifiques ont été développées [142] : méthode de PISARENKO, MUSIC... Nous allons voir que ces techniques peuvent être étendues aux ordres supérieurs. Ces techniques sont également développées en traitement d'antenne : nous en reparlerons au chapitre 6.

Les signaux harmoniques sont définis par J. Mendel [150] comme les signaux pouvant s'écrire selon le modèle

$$y(n) = \sum_{i=1}^p a(i)s(n; \lambda_i).$$

Les $a(i)$ sont des variables aléatoires et les $s(n; \lambda_i)$ des signaux certains. Ces signaux dépendent d'un nombre fini p de variables aléatoires, contrairement aux signaux engendrés par filtrage d'un bruit blanc. Les signaux considérés jusqu'à maintenant étaient des signaux aléatoires (non totalement prédictibles) alors que les signaux harmoniques sont des signaux déterministes (totalement prédictibles).

Les sommes de fréquences pures, en représentation complexe, sont de ce type

$$y(n) = \sum_{i=1}^p \alpha(i)e^{2\pi j\lambda_i n + j\phi_i},$$

où $\alpha(i)$ est en général une variable certaine. Les variables aléatoires ϕ_i sont indépendantes. Lorsque la densité des ϕ_i est constante sur $[0, 2\pi]$ les variables aléatoires complexes $\alpha(i)e^{j\phi_i}$ sont circulaires. La propriété de circularité entraîne que tous les cumulants ne comportant pas un nombre égal de termes non complexes conjugués et de termes complexes conjugués sont nuls. Dans le cas présent on peut vérifier cette propriété directement. Les cumulants d'ordre 3, et donc aussi la bicorrélation, sont nuls. La tricorrélation non nulle de ces signaux est

$$C_{y(2)}^{(2)}(m_1, m_2, m_3) = \text{Cum}[y(n), y(n + m_1), y^*(n - m_2), y^*(n - m_3)].$$

soit

$$C_{y(2)}^{(2)}(m_1, m_2, m_3) = - \sum_{i=1}^p \alpha^4(i) e^{2\pi j \lambda_i (m_1 + m_2 + m_3)}.$$

Pour des signaux à valeurs réelles, la représentation des signaux est

$$y(n) = \sum_{i=1}^p \alpha(i) \cos(2\pi \lambda_i n + \phi_i),$$

et l'on obtient

$$\begin{aligned} C_{y(4)}(m_1, m_2, m_3) &= -\frac{1}{8} \sum_{i=1}^p \alpha^4(i) \{ \cos [2\pi \lambda_i (m_1 - m_2 - m_3)] \\ &\quad + \cos [2\pi \lambda_i (m_2 - m_1 - m_3)] \\ &\quad + \cos [2\pi \lambda_i (m_3 - m_1 - m_2)] \}. \end{aligned}$$

Dans le cas complexe, les informations sur les fréquences présentes sont contenues dans la ligne du trispectre définie par $m = m_1 = m_2 = -m_3$. On obtient alors

$$C_{y(2)}^{(2)}(m, m, -m) = - \sum_{i=1}^p \alpha^4(i) e^{2\pi j \lambda_i m}.$$

Dans le cas réel, la diagonale du trispectre s'écrit

$$C_{y(4)}(m, m, m) = -\frac{3}{8} \sum_{i=1}^p \alpha^4(i) \cos(2\pi \lambda_i m).$$

Les méthodes de super-résolution appliquées à l'ordre 2 sont fondées sur la corrélation des signaux sommes de fréquences pures qui s'écrivent, dans le cas de signaux à valeurs complexes

$$C_{y(1)}^{(1)}(m) = \sum_{i=1}^p \alpha^2(i) e^{2\pi j \lambda_i m},$$

et dans le cas de signaux à valeurs réelles

$$C_{y(2)}(m) = \frac{1}{2} \sum_{i=1}^p \alpha^2(i) \cos(2\pi \lambda_i m).$$

L'analogie entre la forme de la corrélation et de la diagonale (ou de la ligne dans le cas complexe) de la tricorrélation montre que toutes les méthodes développées à l'ordre 2 s'appliquent à l'ordre 4. Le fait d'utiliser la tricorrélation permet d'éliminer un bruit additif gaussien (au moins en théorie).

3.8 Conclusion

L'utilisation des statistiques d'ordre supérieur permet de conduire une analyse plus fine des signaux. Les techniques développées à l'ordre 2 peuvent, comme nous l'avons vu, se transposer dans une grande mesure à des ordres supérieurs. Le grand succès des techniques utilisant des statistiques d'ordre supérieur est dû au fait qu'elles permettent de résoudre le problème de l'identification aveugle. On peut également penser qu'elles permettront de réaliser une extraction plus efficace des signaux pollués par un bruit gaussien. Enfin de nombreuses questions restent ouvertes. Le choix des lignes de multicorrélation les mieux adaptées et la stabilité statistique des estimateurs nous paraissent les domaines essentiels sur lesquels doit porter l'effort de recherche...

CHAPITRE 4

Signaux non stationnaires

La stationnarité est une propriété forte qui permet de définir précisément les fonctions d'analyse adéquates dans les domaines temporel et fréquentiel : les multicorrélations et les multispectres (voir le chapitre 2).

Lorsque l'hypothèse de stationnarité n'est pas vérifiée, on entre dans la classe très générale des signaux non stationnaires. La non-stationnarité est une "non-propriété" : elle n'est définie que par son contraire. Cette remarque implique d'énormes difficultés pour la définition de méthodes d'analyse.

Il faut toutefois tempérer le propos, puisqu'il est possible de définir des classes de non-stationnarités, pour lesquelles des outils d'analyse précis peuvent être définis. Une classe très importante de signaux non stationnaires fait l'objet de la première moitié de ce chapitre : les signaux *cyclostationnaires*. Ces signaux apparaissent dans de nombreuses applications du traitement du signal : communication, radar, signaux physiologiques, etc . . . Nous verrons que le cadre d'analyse approprié est celui des fonctions périodiques ou presque-périodiques. Les rappels sur les fonctions presque-périodiques nous permettront d'examiner l'influence de l'hypothèse de cyclostationnarité sur les statistiques d'ordre supérieur. L'application présentée dans ce chapitre concernera également les signaux cyclostationnaires dans le contexte des communications.

La deuxième partie de ce chapitre se concentre sur les signaux non stationnaires au sens général. Nous montrons que les représentations temps-fréquence d'ordre supérieur (ou représentations temps-multifréquence) peuvent être des outils utiles pour l'analyse des signaux non stationnaires. Ces représentations seront explicitées pour les signaux déterministes, l'extension aux signaux aléatoires ne venant qu'après.

Si des classes particulières de signaux non stationnaires peuvent être définies, leur étude présente toutefois certaines difficultés. La plus fondamentale est la multiplicité des représentations de l'information que contient le signal. Cette multiplicité vient de la notion de temps de référence. En stationnaire, le temps de référence n'a pas d'importance puisque les statistiques du signal sont invariantes par translation dans le temps. Cette invariance n'existant pas pour les signaux non stationnaires, la représentation de l'information

dépend étroitement de la date à laquelle est effectuée l'analyse. Ce problème est évidemment présent dans les représentations temps-fréquence bilinéaires. Par exemple, la représentation de Wigner-Ville, qui examine les liens existant dans le signal entre les instants $t - \tau/2$ et $t + \tau/2$, est différente de la représentation de Rihaczek, qui étudie les corrélations du signal entre les instants t et $t - \tau$. L'information contenue dans ces deux représentations est la même, mais révélée de façon différente. Ce problème pose alors celui de trouver l'outil adéquat pour l'analyse d'un signal.

Le problème de l'instant de référence (t dans l'exemple précédent) sera présent constamment dans ce chapitre, bien qu'un peu oublié lors du traitement des signaux cyclostationnaires.

Un autre problème difficile dans le cas non stationnaire est l'estimation des grandeurs que nous définirons. Le fait d'observer une seule réalisation d'un signal interdit *a priori* l'obtention d'estimateurs efficaces. Nous verrons que la cyclostationnarité contredit cette affirmation. Par contre, les estimateurs des représentations temps-multifréquence sont non consistants et ne seront pas (ou si peu) discutés ici.

Mentionnons enfin que les thèmes présentés sont très récents et susceptibles de subir d'importantes évolutions. Ce chapitre ne montre que quelques résultats et le lecteur peut approfondir les sujets à l'aide de la littérature récente en traitement du signal.

4.1 Signaux cyclostationnaires

Avant de présenter les statistiques d'ordre supérieur des signaux cyclostationnaires à valeurs complexes, commençons par quelques rappels sur la cyclostationnarité à l'ordre 2.

4.1.1 Cyclostationnarité d'ordre 2

Pour illustrer la cyclostationnarité d'ordre 2 [147], nous nous limitons à un signal $x(t)$ de moyenne nulle, à valeurs réelles. $x(t)$ est supposé cyclostationnaire à l'ordre 2 : sa fonction de corrélation est périodique en temps, c'est-à-dire

$$C_{x(2)}(t, \tau) = E[x(t)x(t + \tau)] = C_{x(2)}(t + T, \tau),$$

où T est la période de cyclostationnarité de $x(t)$.

Un exemple simple est donné par un signal codé en binaire à la cadence T , soit $x(t) = \sum_k a_k \delta(t - kT)$, les a_k formant une suite de variables aléatoires indépendantes et identiquement distribuées (i.i.d), prenant les valeurs ± 1 avec

la probabilité $1/2$. Ce signal est de moyenne nulle, et sa corrélation s'écrit

$$C_{x(2)}(t, \tau) = \sum_k \delta(t - kT)\delta(\tau).$$

Cette fonction est périodique en t de période T .

Comme $C_{x(2)}(t, \tau)$ est périodique suivant t , elle admet une représentation en série de Fourier. De plus, pour conserver la notion d'analyse spectrale du cas stationnaire, une transformée de Fourier suivant τ peut être effectuée. Ainsi, par double transformée de Fourier, on définit la corrélation spectrale

$$S_{x(2)}(\alpha, \nu) = \int \int C_{x(2)}(t, \tau) e^{-2\pi j(\alpha t + \nu \tau)} dt d\tau.$$

Cette corrélation spectrale peut également s'écrire

$$S_{x(2)}(\alpha, \nu) = \sum_k s_{x(2)}(k, \nu) \delta\left(\alpha - \frac{k}{T}\right),$$

où les $s_{x(2)}(k, \nu)$ sont les transformées de Fourier selon τ des coefficients $c_{x(2)}(k, \tau)$ de la décomposition en série de Fourier (suivant t) de $C_{x(2)}(t, \tau)$.

La dénomination corrélation spectrale vient du fait que cette fonction examine les interactions statistiques entre les composantes harmoniques aux fréquences ν et $\nu - k/T$.

Le paragraphe suivant présente les outils pour l'étude de la cyclostationnarité (fonctions presque-périodiques) et une définition plus précise de la cyclostationnarité.

4.1.2 Définitions

Les signaux cyclostationnaires comportent des périodicités cachées. Le cadre mathématique approprié est donc celui des fonctions périodiques et des séries de Fourier. Toutefois, ce cadre peut être limitatif puisqu'il exclut la classe très importante des signaux dits presque-périodiques. Dans ce paragraphe, nous discuterons donc la définition des signaux cyclostationnaires dans le cadre des fonctions presque-périodiques. Notons que les fonctions périodiques forment évidemment une sous-classe des fonctions presque-périodiques.

Les développements donnés dans la suite se fondent sur une représentation aléatoire des signaux. W. Gardner et C. Spooner [84, 85] ont construit une théorie dans un cadre déterministe, en ne travaillant que sur une réalisation d'un signal aléatoire. Cette approche ne sera pas considérée ici.

Fonctions presque-périodiques L'objet de ce paragraphe est d'effectuer quelques rappels sur la théorie des fonctions presque-périodiques ou p-périodiques. Une introduction plus complète peut être trouvée dans le livre de J. Bass [21].

La définition des fonctions p-périodiques généralise celle des fonctions périodiques. Pour une fonction périodique de période de base T , les nombres multiples de T constituent également des périodes. Ainsi, dans tout intervalle de longueur T , on peut trouver une période de la fonction. Cette remarque conduit alors à la définition suivante :

Définition : Une fonction à valeurs complexes $f(t)$ est dite presque périodique si pour tout réel $\varepsilon > 0$, il existe un nombre $l(\varepsilon)$ tel que tout intervalle de longueur $l(\varepsilon)$ contienne un nombre τ vérifiant

$$|f(t + \tau) - f(t)| < \varepsilon \quad \forall t.$$

Les nombres τ s'appellent des presque-périodes de la fonction f et le nombre $l(\varepsilon)$ longueur d'inclusion. Cette définition montre qu'une fonction périodique est p-périodique.

Les fonctions p-périodiques jouissent de nombreuses propriétés : l'ensemble des fonctions p-périodiques est un espace vectoriel, la fonction conjuguée, le module de la fonction, les décalées de la fonction sont p-périodiques, et le produit de deux fonctions p-périodiques est p-périodique.

Un exemple illustratif simple est donnée par la somme de deux fréquences pures incommensurables. En effet,

$$f(t) = \exp(2\pi j\nu_0 t) + \exp(2\pi j\nu_1 t), \quad (4.1)$$

où ν_0 et ν_1 sont dans un rapport irrationnel, est p-périodique. L'allure de cette fonction est montrée sur la figure (4.1). La p-périodicité est clairement vue sur ce graphe. J. Bass a démontré rigoureusement la p-périodicité pour cet exemple, en exhibant des presque-périodes et des longueurs d'inclusion.

Les fonctions périodiques admettent une représentation fréquentielle à travers la décomposition en série de Fourier. Cette propriété existe également pour les fonctions p-périodiques. En effet, il peut être démontré que la moyenne d'une fonction p-périodique existe toujours [21], c'est-à-dire

$$\lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^T f(t) dt = M < +\infty. \quad (4.2)$$

Comme le produit de deux fonctions p-périodiques est p-périodique, comme une fonction périodique est p-périodique, alors, en vertu de (4.2), les nombres $c(\nu)$ définis par

$$c(\nu) = \lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^T f(t) \exp(-2\pi j\nu t) dt$$

existent. Les nombres $c(\nu)$ non nuls sont appelés coefficients de Fourier-Bohr de la fonction $f(t)$. L'ensemble des coefficients de Fourier-Bohr d'une fonction p-périodique est un ensemble fini ou dénombrable [21]. Nous noterons ces

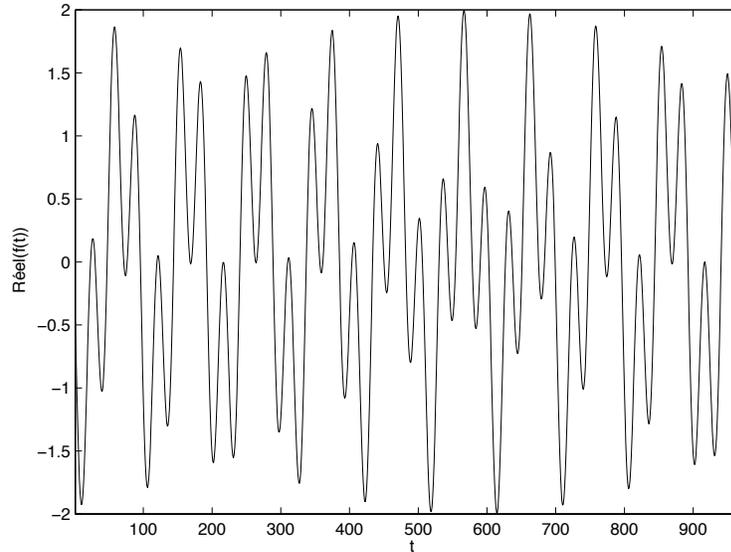


Figure 4.1: *Graphes de la fonction p-périodique $f(t) = \exp(2\pi j\nu_0 t) + \exp(2\pi j\nu_1 t)$, où les fréquences sont dans un rapport irrationnel.*

coefficients par $c_k = c(\nu_k)$. La célèbre formule de Parseval reste valide, et la puissance moyenne de la fonction se répartit sur les coefficients de Fourier-Bohr selon

$$\lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^T |f(t)|^2 dt = \sum_k |c_k|^2.$$

Enfin, la série de Fourier associée aux c_k converge en moyenne quadratique vers la fonction $f(t)$ et nous écrivons

$$f(t) = \sum_k c_k \exp(2\pi j\nu_k t).$$

Pour l'exemple (4.1), il existe deux coefficients de Fourier-Bohr associés aux fréquences ν_0 et ν_1 ; ces deux coefficients sont égaux à 1.

Les rappels sur les fonctions p-périodiques étant faits, nous pouvons maintenant nous tourner vers la définition des signaux cyclostationnaires.

Cyclostationnarité La cyclostationnarité s'inscrit entre la stationnarité et la non-stationnarité pour les signaux aléatoires. Tout comme la stationnarité qui peut être forte ou faible, la discussion du paragraphe précédent montre que la définition de la cyclostationnarité doit comporter divers degrés (périodicité totale ou approximative). De plus, comme la stationnarité, la définition de la cyclostationnarité porte sur les statistiques du signal. Nous pouvons alors donner la définition :

Définition : *Un signal aléatoire est dit cyclostationnaire d'ordre $p+q$ fortement (resp. faiblement) si ses statistiques jusqu'à l'ordre $p+q$ sont périodiques (resp. p -périodiques).*

Si un signal est cyclostationnaire à **tout ordre**, il sera cyclostationnaire **au sens strict**. Mais bien entendu, il pourra l'être faiblement ou fortement.

La périodicité (ou presque-périodicité) s'entend ici sur le temps. Nous voyons donc apparaître la notion fondamentale de temps de référence, absente de toutes les théories stationnaires. En effet, pour un signal stationnaire, les corrélations entre $p+q$ instants du signal ne dépendent que des écarts entre les instants. Par contre, cette propriété disparaît en non-stationnaire où le rôle de l'instant de référence devient primordial.

Cette étape de définition étant présentée, nous nous intéressons dans la suite aux conséquences de l'hypothèse de cyclostationnarité sur les statistiques des signaux.

4.1.3 Influence de la cyclostationnarité sur les SOS

Dans ce paragraphe, nous examinons l'influence de la cyclostationnarité sur les multicorrélations et les multispectres. La construction suit une démarche descendante : partant des définitions les plus générales, nous construisons les objets adéquats pour la description des signaux cyclostationnaires.

Multicorrélation cyclique Considérons la multicorrélation¹ $C_{x^{(p)}}^{(q)}(\mathbf{t})$ d'un signal aléatoire à valeurs complexes ($\mathbf{t} = (t_0, t_1, \dots, t_{p+q-1})$). Nous paramétrons cette multicorrélation comme dans le cas stationnaire², c'est-à-dire

$$C_{x^{(p)}}^{(q)}(t, \boldsymbol{\tau}) = \text{Cum}[x(t), x(t + \tau_1), \dots, x(t + \tau_{p-1}), x^*(t - \tau_p), \dots, x^*(t - \tau_{p+q-1})]. \quad (4.3)$$

où $\boldsymbol{\tau} = (\tau_1, \dots, \tau_{p+q-1})$. Notons que le temps t apparaît explicitement dans les arguments de la multicorrélation puisqu'aucune hypothèse de stationnarité n'est faite.

Supposons que le signal analysé $x(t)$ soit faiblement cyclostationnaire au sens strict. Ses statistiques sont alors p -périodiques quel que soit l'ordre. La multicorrélation précédente est donc p -périodique en t . D'après les résultats sur les fonctions p -périodiques (voir 4.1.2), les multicorrélations admettent une

¹Toute la suite peut être faite également sur les moments. Rappelons que la dénomination multicorrélation sous-entend multicorrélation des cumulants. Lorsqu'il s'agit de moments, nous précisons par multicorrélation des moments.

²Le problème de l'instant de référence sera discuté au paragraphe 4.3.1 .

décomposition en série de Fourier-Bohr, par rapport à la variable t

$$C_{x(p)}^{(q)}(t, \boldsymbol{\tau}) = \sum_{k \in F_{x(p)}^{(q)}} c_{x(p)}^{(q)}(k, \boldsymbol{\tau}) \exp(2\pi j \alpha_k t), \quad (4.4)$$

où $F_{x(p)}^{(q)}$ est l'ensemble des entiers pour lesquels les coefficients de Fourier-Bohr $c_{x(p)}^{(q)}(k, \boldsymbol{\tau})$ sont non nuls. Rappelons que

$$c_{x(p)}^{(q)}(k, \boldsymbol{\tau}) = \lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^T C_{x(p)}^{(q)}(t, \boldsymbol{\tau}) \exp(-2\pi j \alpha_k t) dt.$$

et que $F_{x(p)}^{(q)}$ est un ensemble fini ou dénombrable.

En effectuant la transformée de Fourier de la multicorrélation de la variable t vers la variable α , on définit la *multicorrélation cyclique* qui s'écrit

$$C_{x(p)}^{(q)}(\alpha, \boldsymbol{\tau}) = \sum_{k \in F_{x(p)}^{(q)}} c_{x(p)}^{(q)}(k, \boldsymbol{\tau}) \delta(\alpha - \alpha_k).$$

La fréquence α , variable conjuguée de t , est appelée fréquence cyclique. La multicorrélation cyclique contient uniquement de l'information aux fréquences cycliques α_k . Dans le cas stationnaire, la seule composante p -périodique ou périodique étant le continu, la multicorrélation cyclique est non nulle uniquement pour $\alpha = 0$.

Multicorrélation spectrale Tout comme pour le cas stationnaire, nous allons maintenant examiner l'influence de l'hypothèse de cyclostationnarité sur le multispectre symétrique. Rappelons que le multispectre symétrique est défini à partir de la multicorrélation selon (voir le chapitre 2)

$$\Sigma_{x(p)}^{(q)}(\boldsymbol{\nu}) = \int C_{x(p)}^{(q)}(\mathbf{t}) \exp(-2\pi j (\sum_{i=0}^{p-1} \nu_i t_i - \sum_{i=p}^{p+q-1} \nu_i t_i)) d\mathbf{t}.$$

Si l'on effectue un changement de variables pour obtenir la paramétrisation en $(t, \boldsymbol{\tau})$ de la multicorrélation (4.3), le multispectre symétrique devient

$$\begin{aligned} \Sigma_{x(p)}^{(q)}(\boldsymbol{\nu}) &= \int C_{x(p)}^{(q)}(t, \boldsymbol{\tau}) \exp(-2\pi j (\sum_{i=0}^{p-1} \nu_i - \sum_{i=p}^{p+q-1} \nu_i) t) \\ &\quad \times \exp(-2\pi j \boldsymbol{\nu}'^T \boldsymbol{\tau}) d\boldsymbol{\tau} dt, \end{aligned} \quad (4.5)$$

où le vecteur $\boldsymbol{\nu}'$ est défini par $\boldsymbol{\nu}' = (\nu_1, \dots, \nu_{p+q-1})$. En plongeant dans (4.5) la représentation (4.4), on obtient

$$\Sigma_{x(p)}^{(q)}(\boldsymbol{\nu}) = \sum_{k \in F_{x(p)}^{(q)}} s_{x(p)}^{(q)}(k, \boldsymbol{\nu}') \delta(\alpha_k - (\sum_{i=0}^{p-1} \nu_i - \sum_{i=p}^{p+q-1} \nu_i)),$$

où les coefficients $s_{x^{(p)}}^{(q)}(k, \boldsymbol{\nu}')$ sont les transformées de Fourier des $c_{x^{(p)}}^{(q)}(k, \boldsymbol{\tau})$.

Définissons maintenant la *multicorrélation spectrale* (que l'on pourrait également appeler multispectre cyclique) par

$$S_{x^{(p)}}^{(q)}(\alpha, \boldsymbol{\nu}') = \int C_{x^{(p)}}^{(q)}(t, \boldsymbol{\tau}) \exp(-2\pi j \alpha t) \exp(-2\pi j \boldsymbol{\nu}' \boldsymbol{\tau}^T) d\boldsymbol{\tau} dt. \quad (4.6)$$

On obtient alors la représentation suivante

$$S_{x^{(p)}}^{(q)}(\alpha, \boldsymbol{\nu}') = \sum_{k \in F_{x^{(p)}}^{(q)}} s_{x^{(p)}}^{(q)}(k, \boldsymbol{\nu}') \delta(\alpha - \alpha_k). \quad (4.7)$$

Ces résultats montrent que la multicorrélation spectrale est non nulle uniquement pour les fréquences cycliques α_k du signal analysé.

D'autre part, pour faire le lien avec les théories stationnaires, les résultats sur le multispectre symétrique montrent que la cyclostationnarité crée des hyperplans supplémentaires à la multiplicité stationnaire (voir le chapitre 2). Précisément, le multispectre symétrique est nul partout, sauf sur une famille d'hyperplans parallèles à la multiplicité stationnaire et dont les équations sont

$$\alpha_k = \sum_{i=0}^{p-1} \nu_i - \sum_{i=p}^{p+q-1} \nu_i.$$

Dans le cas de cyclostationnarité forte (périodicité des statistiques), nous parlerons de démultiplication de la multiplicité stationnaire, puisque les hyperplans sont séparés d'elle de $\alpha_k = k/T$, T étant la période fondamentale de la multicorrélation.

Pour terminer cette partie relative aux définitions, il nous reste à examiner la signification physique de la multicorrélation spectrale. Nous avons vu que le multispectre symétrique peut s'écrire en fonction des incréments spectraux du signal selon

$$\Sigma_{x^{(p)}}^{(q)}(\nu_0, \boldsymbol{\nu}') d\nu_0 d\boldsymbol{\nu}' = \text{Cum}[dX(\nu_0), dX(\nu_1), \dots, dX(\nu_{p-1}), dX^*(\nu_p), \dots, dX^*(\nu_{p+q-1})].$$

où nous rappelons que $\boldsymbol{\nu}' = (\nu_1, \dots, \nu_{p+q-1})$. En utilisant la multicorrélation spectrale on obtient

$$\sum_{k \in F_{x^{(p)}}^{(q)}} s_{x^{(p)}}^{(q)}(k, \boldsymbol{\nu}') \delta(\alpha_k - (\sum_{i=0}^{p-1} \nu_i - \sum_{i=p}^{p+q-1} \nu_i)) d\nu_0 d\boldsymbol{\nu}' = \text{Cum}[dX(\nu_0), dX(\nu_1), \dots, dX(\nu_{p-1}), dX^*(\nu_p), \dots, dX^*(\nu_{p+q-1})]. \quad (4.8)$$

La multicorrélation spectrale examine donc les interactions statistiques entre les composantes fréquentielles ν_i , $i = 1, \dots, p+q-1$ et la composante $\alpha - \sum_{i=1}^{p-1} \nu_i + \sum_{i=p}^{p+q-1} \nu_i$. Il faut toutefois se rappeler que ces interactions ne sont non nulles que pour les $\alpha = \alpha_k$.

Propriétés Les multicorrélations cycliques et spectrales jouissent de toutes les propriétés des multicorrélations et multispectres stationnaires. Par exemple, si le signal $x(t)$ est cyclostationnaire et circulaire, seules les multicorrélations cycliques et spectrales non nulles sont du type $\mathcal{C}_{x^{(p)}}^{(p)}(\alpha, \boldsymbol{\tau})$ et $\mathcal{S}_{x^{(p)}}^{(p)}(\alpha, \boldsymbol{\nu}')$. De plus, les propriétés classiques des cumulants leur reviennent (nullité dans le cas gaussien pour un ordre supérieur à deux, ...).

Lorsque le signal est stationnaire, la seule période (ou presque-période) des statistiques du signal est infinie. Par suite, la seule fréquence cyclique ayant des contributions non nulles est la fréquence $\alpha = 0$. On retrouve donc pour cette fréquence les objets statistiques du régime stationnaire.

Examinons maintenant l'influence des transformations linéaires sur les statistiques de signaux cyclostationnaires. Considérons un filtre linéaire invariant, de réponse impulsionnelle $h(t)$ et de fonction de transfert $H(\nu)$. Supposons que ce filtre soit attaqué par un signal $x(t)$ faiblement cyclostationnaire au sens strict. La sortie $y(t)$ est donc la convolution de $x(t)$ par $h(t)$. De plus, dans le domaine des fréquences, la relation entre les incréments spectraux des deux signaux est

$$dY(\nu) = H(\nu)dX(\nu).$$

Par suite, en utilisant le résultat (4.8), le lien entre les multicorrélations spectrales est

$$\begin{aligned} \mathcal{S}_y^{(q)}(\alpha, \boldsymbol{\nu}') &= H\left(\alpha - \sum_{i=1}^{p-1} \nu_i + \sum_{i=p}^{p+q-1} \nu_i\right) H(\nu_1) \dots \\ &\times H(\nu_{p-1}) H^*(\nu_p) H^*(\nu_{p+q-1}) \mathcal{S}_x^{(q)}(\alpha, \boldsymbol{\nu}'). \end{aligned}$$

Cette relation entrée-sortie généralise celle du cas stationnaire que l'on retrouve évidemment pour $\alpha = 0$ (chapitre 2).

4.1.4 Eléments d'estimation

La définition d'objets d'analyse devient intéressante en pratique lorsque des estimateurs de ces objets sont disponibles. Le caractère non stationnaire des signaux cyclostationnaires peut mener à penser que l'estimation des SOS cycliques est impossible. Or il n'en est rien et il existe des estimateurs consistants des multicorrélations cycliques et spectrales.

La théorie de l'estimation a été développée essentiellement par A. Dandawaté et G. Giannakis pour les signaux cyclostationnaires à valeurs réelles [61, 62] (voir aussi [84, 85]). Nous ne reprendrons pas ici toutes les théories développées par ces auteurs mais essaierons plutôt de faire comprendre la technique d'estimation en nous concentrant sur une méthode également applicable en stationnaire [37, 38]. De plus, nous illustrerons l'estimation en nous restreignant à l'ordre 4 symétrique ($p = 2$ et $q = 2$) qui présente toutes les difficultés.

Relation entre cumulants et moments Nous avons vu dans le chapitre 1 la relation liant les moments aux cumulants. Cette relation nous permet d'écrire la multicorrélation en fonction de la multicorrélation des moments d'ordre égaux ou inférieurs. Pour un signal centré, cette relation à l'ordre 4 symétrique s'écrit

$$\begin{aligned} C_{x(2)}^{(2)}(t, \boldsymbol{\tau}) &= C_{x(2)}^{\prime(2)}(t, \boldsymbol{\tau}) \\ &- C_{x(2)}(t, \tau_1) \left[C_{x(2)}(t, \tau_3 - \tau_2) \right]^* \\ &- C_{x(1)}^{(1)}(t, \tau_2) C_{x(1)}^{(1)}(t, \tau_1 - \tau_3) \\ &- C_{x(1)}^{(1)}(t, \tau_3) C_{x(1)}^{(1)}(t, \tau_1 - \tau_2). \end{aligned}$$

Si l'on effectue une double transformée de Fourier en t et $\boldsymbol{\tau}$ nous obtenons la multicorrélation spectrale en fonction des multicorrélations spectrales des moments d'ordre égaux ou inférieurs à 4 selon

$$\begin{aligned} \mathcal{S}_{x(2)}^{(2)}(\alpha, \boldsymbol{\nu}) &= \mathcal{S}_{x(2)}^{\prime(2)}(\alpha, \boldsymbol{\nu}) \\ &- \left\{ \int \mathcal{S}_{x(2)}(u, \nu_1) \left[\mathcal{S}_{x(2)}(\alpha - u, \nu_3) \right]^* du \right\} \delta(\nu_2 + \nu_3) \\ &- \left\{ \int \mathcal{S}_{x(1)}^{(1)}(u, \nu_2) \mathcal{S}_{x(1)}^{(1)}(\alpha - u, \nu_1) du \right\} \delta(\nu_1 - \nu_3) \\ &- \left\{ \int \mathcal{S}_{x(1)}^{(1)}(u, \nu_3) \mathcal{S}_{x(1)}^{(1)}(\alpha - u, \nu_2) du \right\} \delta(\nu_1 - \nu_2). \end{aligned}$$

Remarquons que la notation adoptée ici est un peu ambiguë. En effet, nous avons fait apparaître des produits de convolution qui doivent s'entendre comme suit. La multicorrélation spectrale est non nulle uniquement pour les α appartenant à $F_{x(p)}^{(q)}$: nous n'avons pas un continuum en fréquence cyclique. Les convolutions de l'expression précédente sont donc des convolutions discrètes réalisées sur une grille qui n'est pas forcément à échantillonnage régulier. A un α_k donné, les contributions de produits de corrélations spectrales devraient s'écrire en fonction des coefficient $s_{x(\cdot)}^{(\cdot)}(k, \cdot)$, définis par (4.7), selon des convolutions discrètes. Par exemple, le terme $\int \mathcal{S}_{x(2)}(u, \nu_1) \left[\mathcal{S}_{x(2)}(\alpha - u, \nu_3) \right]^* du$ s'interprète comme

$$\sum_{k_1 \in F_{x(2)}^{(2)}} s_{x(2)}(k, \nu_1) \left[s_{x(2)}(k - k_1, \nu_3) \right]^* .$$

Autrement dit, il faut sommer les produits de termes dont la somme des fréquences cycliques égale α_k .

D'après ce résultat, la multicorrélation spectrale est égale à la multicorrélation spectrale des moments excepté sur les hyperplans d'équation

$$\begin{cases} \nu_1 - \nu_3 = 0 \\ \nu_1 - \nu_2 = 0 \\ \nu_2 + \nu_3 = 0. \end{cases} \quad (4.9)$$

Ce résultat suggère une méthode d'estimation. Si l'on suppose la multicorrélation spectrale continue autour des hyperplans précédents, alors il suffit d'estimer la multicorrélation des moments hors ces hyperplans et d'interpoler l'estimation à travers ces hyperplans. L'interpolation s'effectue en moyennant les valeurs de l'estimée autour des hyperplans.

Estimation du moment d'ordre 4 Considérons un signal $x(t)$ *a priori* cyclostationnaire à l'ordre 4 observé sur une durée T . Nous noterons $X_T(\nu)$ la transformée de Fourier de $x(t)$ sur cette durée, c'est-à-dire

$$X_T(\nu) = \int_{-T/2}^{T/2} x(t) \exp(-2\pi j\nu t) dt.$$

L'estimateur le plus simple du moment est le périodogramme cyclique [61] qui s'écrit

$$\mathcal{P}_{x,T}^{(2)}(\alpha, \boldsymbol{\nu}) = \frac{1}{T} X_T(\alpha - \nu_1 + \nu_2 + \nu_3) X_T(\nu_1) X_T^*(\nu_2) X_T^*(\nu_3).$$

On montre facilement que le périodogramme est asymptotiquement non biaisé, c'est-à-dire

$$\lim_{T \rightarrow +\infty} E[\mathcal{P}_{x,T}^{(2)}(\alpha, \boldsymbol{\nu})] = \mathcal{S}'_{x(2)}(\alpha, \boldsymbol{\nu})$$

en dehors des hyperplans définis par (4.9). Par contre, on montre que la variance du périodogramme cyclique ne tend pas vers zéro lorsque la durée T d'observation croît indéfiniment [61]. Ce fait, bien connu en analyse spectrale, montre qu'il est nécessaire de lisser le périodogramme cyclique pour améliorer l'estimation.

Lisser à l'aide d'une fenêtre bien choisie le périodogramme cyclique conduit à une estimation asymptotiquement consistante (non biais et variance nulle) de la multicorrélation spectrale des moments.

Estimation du cumulants Nous avons vu que la multicorrélation spectrale est égale à la multicorrélation spectrale des moments, sauf sur certains hyperplans. Autrement dit, la multicorrélation spectrale des moments contient des singularités sur ces hyperplans. Le périodogramme, estimateur du moment, présente donc ces singularités. Ainsi, si l'on lisse le périodogramme cyclique

en omettant les singularités, on obtient un estimateur de la multicorrélation spectrale. Cet estimateur peut alors s'écrire

$$\widehat{\mathcal{S}}_{x(2)}^{(2)}(\alpha, \boldsymbol{\nu}) = \int W_T(\boldsymbol{\nu} - \boldsymbol{\lambda}) \mathcal{P}_{x,T(2)}^{(2)}(\alpha, \boldsymbol{\lambda}) \phi(\boldsymbol{\lambda}) d\boldsymbol{\lambda}.$$

La fonction $\phi(\boldsymbol{\nu})$ vaut 1 partout sauf sur les hyperplans (4.9) sur lesquels elle prend la valeur 0. La fonction $W_T(\boldsymbol{\nu})$ est la fenêtre de lissage. L'indice T montre que son comportement dépend étroitement de T . En effet, cette fenêtre doit tendre vers l'impulsion de Dirac lorsque T tend vers l'infini pour assurer le non biais asymptotique à l'estimateur. Ceci signifie que sa bande équivalente doit tendre vers 0 lorsque T va à l'infini. Une façon agréable de jouer sur la bande de la fenêtre est de la définir à partir d'une fenêtre de référence $W(\boldsymbol{\nu})$ via $W_T(\boldsymbol{\nu}) = B_T^{-1} W(B_T^{-1} \boldsymbol{\nu})$, où $W(\boldsymbol{\nu})$ vérifie certaines propriétés assurant la convergence de $W_T(\boldsymbol{\nu})$ vers un Dirac lorsque T tend vers l'infini.

A. Dandawaté et G. Giannakis ont prouvé dans [61] la consistance asymptotique de cet estimateur.

De plus, pour T fini mais assez grand, leur résultat montre que

$$\text{Var}[\widehat{\mathcal{S}}_{x(2)}^{(2)}(\alpha, \boldsymbol{\nu})] \approx \frac{E_W}{B_T^3 T} O_8,$$

où O_8 représente des termes statistiques d'ordre 8 et E_W est l'énergie de la fenêtre W . Encore une fois, augmenter la durée du lissage (i.e. B_T) réduit la variance mais augmente le biais. Le compromis biais-variance existe toujours.

4.1.5 Application en Communication

Dans ce paragraphe nous présentons un exemple de signal cyclostationnaire issu des télécommunications.

Modélisation du signal Nous souhaitons transmettre un signal codé par des modulations de phase à N états (N -PSK pour Phase Shift Keying en anglais) à travers un canal de transmission. Une N -PSK est représentée par les valeurs qu'elle peut prendre. Ces valeurs sont complexes et de module constant. Il suffit donc de les considérer sur le cercle unité. Ces valeurs seront alors prises dans l'ensemble $\{\exp(\pi j(2n-1)/N)\}_{n=1,\dots,N}$. Pour réaliser le codage, un "digit" du code sera répété sur une durée T . Le signal est alors codé par le message

$$m(t) = \sum_k a_k f(t - kT),$$

où la fonction f est appelée fonction de forme et caractérise la forme du signal sur un intervalle de longueur T . La suite $\{a_k\}$ est supposée indépendante et identiquement distribuée (i.i.d.) et stationnaire. A chaque instant k , a_k prend

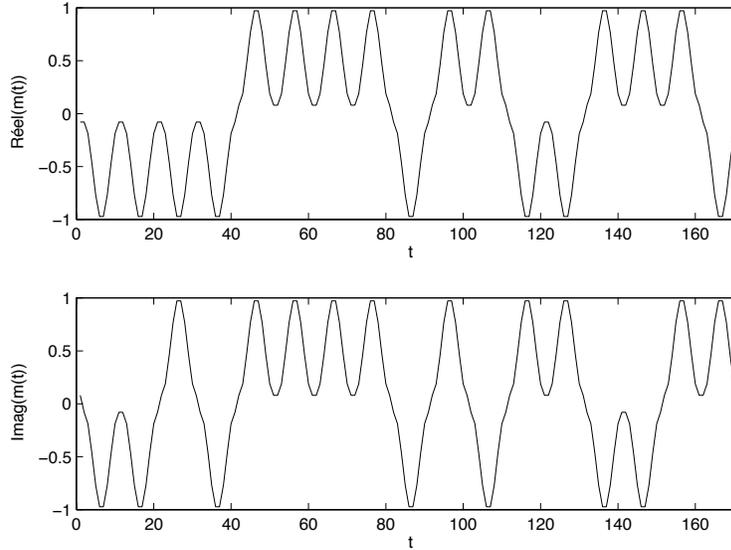


Figure 4.2: *Exemple de Modulation de Phase   4  tats. La fonction de forme est une fen tre de Hamming.*

une des N valeurs pr c dentes, avec une probabilit   gale, ind pendamment des instants pass s. $m(t)$ est un processus al atoire   valeurs complexes. Un exemple illustrant $m(t)$ est pr sent  sur la figure (4.2). Ici, la fonction de forme est une fen tre de Hamming d finie sur l'intervalle $[0, T]$.

Pour s'adapter   la bande passante du canal, le signal doit  tre centr  sur une fr quence porteuse ν_0 . Le signal  mis peut alors s' crire

$$x_e(t) = \text{Re}(m(t) \exp(2\pi j\nu_0 t)),$$

ou de fa on  quivalente

$$x_e(t) = \text{Re}(m(t)) \cos(2\pi\nu_0 t) - \text{Im}(m(t)) \sin(2\pi\nu_0 t).$$

A la r ception, si l'on n glige toute distorsion due au canal,

$$r(t) = x_e(t) + b_r(t),$$

o  $b_r(t)$ repr sente le bruit de mesure. Ce bruit est mod lis  par un processus al atoire ind pendant du signal  mis, et ce bruit pourra  tre gaussien ou non, blanc ou non et stationnaire ou non. Nous reviendrons sur ces hypoth ses.

Une d modulation complexe est effectu e pour s'affranchir de la modulation et ramener le signal en "bande de base". Le signal   traiter s' crit alors

$$y(t) = m(t) + b(t),$$

o  $b(t)$ est un processus al atoire   valeurs complexes.

SOS du signal reçu Le signal $m(t)$ comporte la fréquence caractéristique $1/T$ et est donc fortement cyclostationnaire.

Commençons par évaluer les multicorrélations du signal reçu. Etant donnée la multilinéarité des cumulants et leur propriété quant à la somme de signaux indépendants, la multicorrélation de $y(t)$ s'écrit

$$C_{y(p)}^{(q)}(t, \boldsymbol{\tau}) = C_{m(p)}^{(q)}(t, \boldsymbol{\tau}) + C_{b(p)}^{(q)}(t, \boldsymbol{\tau}).$$

En effectuant une transformée de Fourier par rapport à t , nous obtenons la multicorrélation cyclique

$$\mathcal{C}_{y(p)}^{(q)}(\alpha, \boldsymbol{\tau}) = \mathcal{C}_{m(p)}^{(q)}(\alpha, \boldsymbol{\tau}) + \mathcal{C}_{b(p)}^{(q)}(\alpha, \boldsymbol{\tau}).$$

Enfin, la multicorrélation spectrale est obtenue par transformation de Fourier par rapport aux τ_i

$$\mathcal{S}_{y(p)}^{(q)}(\alpha, \boldsymbol{\nu}) = \mathcal{S}_{m(p)}^{(q)}(\alpha, \boldsymbol{\nu}) + \mathcal{S}_{b(p)}^{(q)}(\alpha, \boldsymbol{\nu}).$$

Il nous faut maintenant évaluer les statistiques de $m(t)$. Sa multicorrélation se calcule en utilisant la multilinéarité des cumulants et s'exprime selon

$$\begin{aligned} C_{m(p)}^{(q)}(t, \boldsymbol{\tau}) &= \sum_{k_0, \dots, k_{p+q-1}} f(t - k_0 T) f(t + \tau_1 - k_1 T) \dots f(t + \tau_{p-1} - k_{p-1} T) \\ &\times f^*(t - \tau_p - k_p T) \dots f^*(t - \tau_{p+q-1} - k_{p+q-1} T) \\ &\times \text{Cum}[a_{k_0}, a_{k_1}, \dots, a_{k_{p-1}}, a_{k_p}^*, \dots, a_{k_{p+q-1}}^*]. \end{aligned}$$

Comme la suite $\{a_k\}$ est une suite de variables aléatoires i.i.d., le cumulants est non nul seulement si les k_i sont tous égaux. Introduisons alors

$$C_{f(p)}^{(q)}(t, \boldsymbol{\tau}) = f(t) f(t + \tau_1) \dots f(t + \tau_{p-1}) f^*(t - \tau_p) \dots f^*(t - \tau_{p+q-1}),$$

pour obtenir

$$C_{m(p)}^{(q)}(t, \boldsymbol{\tau}) = C_{a(p)}^{(q)} \left(C_{f(p)}^{(q)}(t, \boldsymbol{\tau}) * \sum_k \delta(t - kT) \right),$$

où $C_{a(p)}^{(q)}$ est "l'auto-cumulants" de $\{a_k\}$, ou encore sa multicorrélation en $\boldsymbol{\tau} = \mathbf{0}$. La multicorrélation cyclique de $m(t)$ est alors

$$\mathcal{C}_{m(p)}^{(q)}(\alpha, \boldsymbol{\tau}) = \frac{1}{T} C_{a(p)}^{(q)} \mathcal{C}_{f(p)}^{(q)}(\alpha, \boldsymbol{\tau}) \sum_k \delta(\alpha - \frac{k}{T}).$$

Une dernière transformée de Fourier multidimensionnelle permet d'obtenir la multicorrélation spectrale de $m(t)$. Finalement, la multicorrélation spectrale du signal reçu peut s'écrire selon

$$\mathcal{S}_{y(p)}^{(q)}(\alpha, \boldsymbol{\nu}) = \frac{1}{T} C_{a(p)}^{(q)} \sum_k \mathcal{S}_{f(p)}^{(q)}\left(\frac{k}{T}, \boldsymbol{\nu}\right) \delta\left(\alpha - \frac{k}{T}\right) + \mathcal{S}_{b(p)}^{(q)}(\alpha, \boldsymbol{\nu}).$$

Reconnaissance de modulation en bande de base Le but de ce paragraphe est d'utiliser les résultats précédents pour mettre au point une technique de reconnaissance de modulation [139]. Nous savons que le signal émis est une N -PSK, mais N est à déterminer.

Lorsque le signal a été démodulé, nous avons vu que

$$\mathcal{S}_{y(p)}^{(q)}(\alpha, \boldsymbol{\nu}) = \frac{1}{T} C_{a(p)}^{(q)} \sum_k \mathcal{S}_{f(p)}^{(q)}\left(\frac{k}{T}, \boldsymbol{\nu}\right) \delta\left(\alpha - \frac{k}{T}\right) + \mathcal{S}_{b(p)}^{(q)}(\alpha, \boldsymbol{\nu}).$$

Nous devons maintenant faire des hypothèses sur le bruit. Différents cas de figure se présentent. Si le bruit est supposé gaussien, $\mathcal{S}_{b(p)}^{(q)}(\alpha, \boldsymbol{\nu}) = 0$ puisque les multicorrélations sont construites sur les cumulants. Si le bruit est non gaussien mais stationnaire, alors $\mathcal{S}_{b(p)}^{(q)}(\alpha, \boldsymbol{\nu}) = 0$ pour tout α différent de zéro. Il suffit alors de supposer le bruit stationnaire et d'examiner le comportement des multicorrélations spectrales pour des α non nuls.

Nous supposons donc que

$$\mathcal{S}_{y(p)}^{(q)}(\alpha, \boldsymbol{\nu}) = \frac{1}{T} C_{a(p)}^{(q)} \mathcal{S}_{f(p)}^{(q)}(\alpha, \boldsymbol{\nu}) \sum_{k \neq 0} \delta\left(\alpha - \frac{k}{T}\right).$$

A un ordre $p + q$ donné, la forme de la multicorrélation spectrale du signal est indépendante de N . En effet, les multicorrélations spectrales de deux PSK ne diffèrent que par le facteur multiplicatif $C_{a(p)}^{(q)}$. Nous ne pouvons donc *a priori* pas distinguer deux PSK différentes. Toutefois, en conjuguant la cyclostationnarité avec les propriétés de circularité des N -PSK, il est possible de faire la distinction cherchée. En effet, une N -PSK est circulaire à l'ordre $N - 1$ (voir le chapitre 1, paragraphe 1.3). Le principe de la reconnaissance est alors le suivant :

1. $p + q = 2$,
2. Si $\mathcal{S}_{y(p)}^{(q)}(\alpha, \boldsymbol{\nu}) \neq 0$ pour $p \neq q$ alors la suite $\{a_k\}$ est non circulaire à l'ordre $p + q$. La modulation est alors une $(p + q)$ -PSK.
3. Sinon, $p + q \leftarrow p + q + 1$, et recommencer en 2.

Les signaux cyclostationnaires sont très importants dans toutes les applications où des périodicités cachées existent. Citons les signaux issus de machines tournantes ou d'engrenages, les signaux biologiques et physiologiques (EEG, ECG, température, ...), les signaux de télécommunications, ... Si ces signaux sont couplés ou issus de non-linéarités alors les objets définis ici ne peuvent qu'aider à la compréhension des phénomènes mis en jeu.

4.2 Rappel sur le temps-fréquence

Les analyses à l'ordre deux s'intéressent aux propriétés énergétiques des signaux. Dans le cas stationnaire, la densité spectrale représente la distribution de l'énergie sur les fréquences constituant le signal. Cette grandeur est fondamentale et se doit d'être la première quantité étudiée pour la compréhension d'un signal. Toutefois, si le spectre est calculé pour un signal non stationnaire, cette information est incomplète. Par exemple, le spectre d'une modulation linéaire de fréquence montre une partie très large bande, allant des plus basses aux plus hautes fréquences balayées par le signal. Or, dans cet exemple, un tel spectre n'est pas révélateur de la structure précise du signal. En effet, la transformée de Fourier, base du spectre, considère les fréquences à amplitudes non nulles comme présentes à tout instant dans le signal, et ce n'est manifestement pas le cas pour une modulation linéaire de fréquence. Une philosophie s'est alors développée, à partir de 1948 et l'article de J. Ville [195], proposant de regarder l'évolution du contenu fréquentiel (ou énergétique) d'un signal en fonction du temps. Ceci a conduit à diverses approches, montrant encore une fois qu'une non-propriété n'a pas de caractérisation unique. Dans ce paragraphe, nous rappellerons quelques éléments concernant les représentations temps-fréquence bilinéaires, dont les extensions aux ordres supérieurs font l'objet de la suite de ce chapitre.

4.2.1 Représentations temps-fréquence

La construction d'une représentation temps-fréquence d'ordre deux n'est pas du tout triviale. En effet, une telle représentation se doit de vérifier certaines propriétés. Pour les représentations bilinéaires³, le but est d'obtenir une répartition de l'énergie du signal dans un plan temps-fréquence, c'est-à-dire construire une fonction $R_x(t, \nu)$. Nous allons exposer quelques propriétés qu'une "bonne représentation" devrait posséder pour atteindre ce but. Nous ne voulons pas ici faire une liste exhaustive de ces propriétés, mais plutôt dégager la philosophie conduisant aux représentations acceptables (des listes complètes de propriétés souhaitables se trouvent dans [74]).

- Réalité : une densité spectrale étant à valeurs réelles, on souhaite conserver cette propriété pour une densité temps-fréquence.
- Positivité : l'énergie étant physiquement positive, il est souhaitable de conserver cette contrainte dans le cas d'une répartition d'énergie dans le plan temps-fréquence.
- Énergie : en tant que répartition de l'énergie dans le plan temps-fréquence, la représentation doit permettre de retrouver l'énergie du si-

³Précisons que la bilinéarité est relative au signal.

gnal. De même, la puissance instantanée et la densité spectrale peuvent être déduites de la représentation (“marginales”).

- Invariances par translation : deux événements identiques arrivant à des dates différentes doivent avoir la même représentation, à un retard près (la même propriété doit être vérifiée dans le cas d’événement identiques occupant des zones de fréquence différentes – modulation d’un signal par une fréquence pure –). Notons que l’on parle également de covariance.
- Conservation des supports (localisation) : un évènement à support temporel fini aura une représentation ayant le même support temporel (même propriété pour un évènement fréquentiel à support fini).
- Indépendance vis-à-vis du temps dans le cas d’un signal stationnaire.
- etc ...

Diverses autres propriétés souhaitables pourraient être citées. Mais il n’est pas certain qu’il existe une représentation temps-fréquence vérifiant toutes ces propriétés. En fait, toutes ces propriétés ne peuvent être simultanément vérifiées par les représentations bilinéaires ; par exemple, la propriété de positivité interdit les propriétés de marginales [74].

Le terme bilinéaire est apparu plusieurs fois jusqu’à présent. En effet, répétons que nous voulons une transformation non paramétrique capable de répartir aussi bien que possible l’énergie d’un signal dans un plan temps-fréquence. Examiner l’énergie impose de travailler sur le produit des valeurs du signal à deux instants, $x(t_1)x^*(t_2)$, ou à deux fréquences, $X(\nu_1)X^*(\nu_2)$.

Dans le cas stationnaire, le choix des instants n’a pas d’importance, seule la différence entre les deux comptant réellement. Dans le cas non stationnaire, les deux instants doivent être considérés par rapport à un temps de référence, soit l’instant courant t .

La représentation temps-fréquence doit garder un caractère local, et les évènements apparaissant longtemps avant ou après l’instant de référence t ne doivent pas avoir autant d’importance qu’un évènement proche de t . Nous devons alors pondérer $x(t_1)x^*(t_2)$ par une fonction k dépendant de t , mais aussi de t_1 et t_2 qui chiffrent l’éloignement par rapport à t (l’analogue stationnaire de cette pondération est l’apodisation dans l’analyse spectrale). Cette opération peut se répéter sur l’axe dual de l’axe temporel, l’axe fréquentiel. En effet, la non-stationnarité fait perdre l’orthogonalité entre deux composantes fréquentielles $X(\nu_1)$ et $X(\nu_2)$. Ainsi, la pondération k doit aussi dépendre de la fréquence d’analyse ν . Ces considérations conduisent alors à la forme générique d’une représentation temps-fréquence bilinéaire, soit pour un signal $x(t)$ complexe

$$\mathcal{R}_{x(1)}^{(1)}(t, \nu) = \int \int k(t, \nu; t_1, t_2)x(t_1)x^*(t_2)dt_1dt_2. \quad (4.10)$$

La forme (4.10) est la plus générale possible, et est donc trop compliquée. Les propriétés souhaitées indiquées précédemment se transforment en contraintes sur la fonction de pondération $k(t, \nu; t_1, t_2)$ qui prendra donc des formes plus simples. Cette constatation est à la base de la définition de la classe de Cohen : classe des représentations temps-fréquence bilinéaires invariantes par translations temporelles et fréquentielles.

4.2.2 Classe de Cohen

Certaines des propriétés précédentes apparaissent comme très naturelles. Parmi celles-ci, les invariances par translations temporelle et fréquentielle correspondent à des vœux très forts. Il est naturel qu'une représentation se décale dans le temps si l'évènement qu'elle décrit se déplace également. La traduction mathématique de ces propriétés est la suivante:

- translation temporelle : si $y(t) = x(t - \tau)$ alors

$$\mathcal{R}_{y(1)}^{(1)}(t, \nu) = \mathcal{R}_{x(1)}^{(1)}(t - \tau, \nu),$$

- translation fréquentielle : si $y(t) = x(t) \exp(2\pi j \nu_0 t)$ alors

$$\mathcal{R}_{y(1)}^{(1)}(t, \nu) = \mathcal{R}_{x(1)}^{(1)}(t, \nu - \nu_0).$$

On montre alors que $\mathcal{R}_{x(1)}^{(1)}(t, \nu)$ prend la forme [74]

$$\mathcal{R}_{x(1)}^{(1)}(t, \nu) = \int \int F(t - \theta, \tau) x(\theta + \frac{\tau}{2}) x^*(\theta - \frac{\tau}{2}) \exp(-2\pi j \nu \tau) d\tau d\theta, \quad (4.11)$$

où $F(t, \tau)$ est une fonction paramétrant la représentation. Les représentations temps-fréquence bilinéaires vérifiant ces propriétés d'invariance et s'écrivant donc comme (4.11) appartiennent à la classe dite de Cohen. Un élément fameux de cette classe est la représentation ou transformée de Wigner-Ville $\mathcal{WV}_{x(1)}^{(1)}(t, \nu)$ qui est obtenue pour $F(t, \tau) = \delta(t)$, soit

$$\mathcal{WV}_{x(1)}^{(1)}(t, \nu) = \int x(t + \frac{\tau}{2}) x^*(t - \frac{\tau}{2}) \exp(-2\pi j \nu \tau) d\tau.$$

Cette représentation est centrale dans la classe de Cohen, puisqu'elle permet d'engendrer tous les membres de la classe par lissage temps-fréquence selon

$$\mathcal{R}_{x(1)}^{(1)}(t, \nu) = \Pi(t, \nu) *_t *_\nu \mathcal{WV}_{x(1)}^{(1)}(t, \nu),$$

où $\Pi(t, \nu)$ est une fenêtre bidimensionnelle (cette fenêtre est la transformée de Fourier par rapport à τ de la fonction $F(t, \tau)$).

A titre d'exemple, le lissage de la représentation de Wigner-Ville d'un signal $x(t)$ par la transformée de Wigner-Ville d'une fenêtre $h(t)$ conduit au spectrogramme (ou module carré de la transformée de Fourier glissante) lié à la fenêtre $h(t)$, soit plus explicitement

$$\left[\mathcal{WV}_{h(1)}^{(1)}(t, \nu) \right]^* *_t *_\nu \mathcal{WV}_{x(1)}^{(1)}(t, \nu) = \left| \int h^*(t - \tau)x(\tau) \exp(-2\pi j\nu\tau) d\tau \right|^2.$$

Remarquons enfin que la construction de la classe de Cohen est effectuée pour des signaux déterministes. W. Martin et P. Flandrin ont montré que pour des signaux aléatoires harmonisables (classe $f^{(2)}$ de Blanc-Lapierre et Fortet, voir le paragraphe 2.9.2), l'espérance mathématique de $\mathcal{WV}_{x(1)}^{(1)}(t, \nu)$ existe et définit le spectre de Wigner-Ville [74, 144]. Ceci constitue l'extension au cas aléatoire des représentations temps-fréquence bilinéaires, extension qui sera discutée plus en détail au paragraphe 4.3.2.

Nous avons rappelé dans ce paragraphe la philosophie des représentations temps-fréquence, en nous concentrant sur les représentations bilinéaires. Examinons maintenant un exemple de problème pour lequel l'utilisation des représentations ou distributions temps-fréquence s'avère insuffisante.

4.2.3 Non-linéarité et distributions bilinéaires

Considérons le signal $x(t)$ défini par

$$x(t) = e^{-2\pi j(a_1 + b_1 t)t + j\phi_1} + e^{-2\pi j(a_2 + b_2 t)t + j\phi_2} + e^{-2\pi j(a_1 + a_2 + (b_1 + b_2)t)t + j\phi_3},$$

où ϕ_1 et ϕ_2 sont des phases aléatoires indépendantes, uniformément réparties sur $[0, 2\pi]$. Nous envisageons deux cas pour ϕ_3 :

1. ϕ_3 est indépendante de ϕ_1 et ϕ_2 , et est uniformément distribuée sur $[0, 2\pi]$.
2. $\phi_3 = \phi_1 + \phi_2$.

Ces deux situations sont très différentes, puisque la deuxième signifie que la modulation $a_1 + a_2 + (b_1 + b_2)t$ provient d'une interaction quadratique entre les deux premières modulations. Calculons le spectre de Wigner-Ville de ce signal. On doit évaluer au préalable $E[x(t + \frac{\tau}{2})x^*(t - \frac{\tau}{2})]$. Ce calcul se simplifie en remarquant qu'après avoir développé le produit $x(t + \frac{\tau}{2})x^*(t - \frac{\tau}{2})$, six termes conservent leur caractère aléatoire et ont une contribution nulle par moyennage d'ensemble. Les trois termes restants sont du type $\exp(-2\pi j(a + 2bt)\tau)$. Le résultat vient après le calcul de la transformée de Fourier, soit

$$\begin{aligned} \mathcal{WV}_{x(1)}^{(1)}(t, \nu) &= \delta(\nu - (a_1 + 2b_1 t)) + \delta(\nu - (a_2 + 2b_2 t)) \\ &+ \delta(\nu - (a_1 + a_2 + 2(b_1 + b_2)t)). \end{aligned}$$

Le résultat est surprenant : il ne dépend absolument pas du choix fait sur ϕ_3 lors de la modélisation. La conclusion est donc que le spectre de Wigner-Ville ne peut faire la distinction entre un phénomène linéaire et un phénomène non linéaire.

Cet exemple simple montre que les représentations temps-fréquence bilinéaires ne sont sensibles qu'aux propriétés du second ordre, et par suite ne peuvent exhiber que des phénomènes linéaires.

L'exemple précédent démontre la nécessité d'outils différents, adaptés aux phénomènes non linéaires et aux propriétés d'ordre supérieur des signaux non stationnaires. De plus, sans montrer d'exemple particulier, l'idée de caractériser l'évolution au cours du temps de propriétés d'ordre supérieur est naturelle. Les applications de telles idées à des phénomènes physiques ne sont pas encore très développées, mais nul doute qu'elles peuvent être utiles.

Les motivations étant présentées, nous proposons maintenant une extension aux ordres supérieurs des représentations temps-fréquence bilinéaires.

4.3 Représentations temps-multifréquence

Les rappels effectués précédemment sur les représentations bilinéaires montrent l'aspect déductif ou axiomatique de leur construction. Toutefois les interprétations que l'on peut faire de ces objets ont été délibérément cachées. En effet, ces interprétations vont nous servir de fil conducteur pour la présentation des représentations temps-fréquence d'ordre supérieur ou représentations temps-multifréquence.

La première extension aux ordres supérieurs des représentations temps-fréquence est due à N. Gerr en 1988 [86]. Ce premier travail a conduit un certain nombre de chercheurs au développement des représentations temps-multifréquence [17, 75, 88, 191].

4.3.1 Cas des signaux déterministes

Nous commençons la présentation par l'étude des représentations temps-multifréquence des signaux déterministes. Dans la suite, les signaux sont tels que les intégrales envisagées existent.

Multicorrélation locale L'analyse spectrale de signaux stationnaires examine le contenu fréquentiel des signaux, ou la répartition de l'énergie sur les fréquences éternelles contenues dans le signal. D'autre part, une analyse multi-spectrale révèle d'éventuels couplages existant entre les fréquences composant un signal. Par exemple, le bispectre est l'outil idéal pour la recherche de couplages quadratiques.

Le fait important est que ces analyses supposent les fréquences éternelles, puisque le signal est stationnaire. Dans un cas non stationnaire, des fréquences peuvent s'évanouir dans le temps ou apparaître. Effectuer une analyse supposant le signal stationnaire peut alors conduire à des interprétations erronées, et il faut incorporer à l'analyse la dimension temporelle pour espérer interpréter correctement le contenu du signal.

Il est donc très intuitif pour effectuer des analyses d'ordre supérieur sur des signaux non-stationnaires d'essayer de rendre les objets déjà définis dépendant du temps.

Pour les signaux certains, les multicorrélations sont définies par

$$C_{x(p)}^{(q)}(\boldsymbol{\tau}) = \int x(t)x(t + \tau_1) \dots x(t + \tau_{p-1})x^*(t - \tau_p) \dots x^*(t - \tau_{p+q-1})dt.$$

Cette définition inclut un moyennage temporel et suppose donc que les interactions entre les différents instants mis en jeu ne dépendent que des positions relatives de ces instants. Ceci n'est donc valide que si les signaux considérés sont "stationnaires". Cette notion pour des signaux déterministes peut être définie comme suit : un signal déterministe est stationnaire s'il est la somme dénombrable de composantes à fréquence instantanée constante et à amplitude constante [74]. Le moyennage au cours du temps est alors justifié.

Dans le cas contraire, le moyennage n'est plus licite. Nous introduisons alors

$$C_{x(p)}^{\ell(q)}(t, \boldsymbol{\tau}) = x(t)x(t + \tau_1) \dots x(t + \tau_{p-1})x^*(t - \tau_p) \dots x^*(t - \tau_{p+q-1}),$$

appelée *multicorrélation locale*. Elle quantifie les interactions entre $p + q$ dates du signal, ces instants étant repérés par rapport à une référence t .

Un problème important apparaît déjà : le choix du positionnement des dates par rapport à t est arbitraire. Dans la définition précédente, le choix adopté est le même que pour la construction des multicorrélations dans le cas aléatoire stationnaire (dans ce cas, les justifications de ce choix sont claires). Dans le cas non stationnaire des choix différents peuvent être adoptés, comme par exemple remplacer les $p + q$ temps $\{t_0 = t, t_i = t \pm \tau_i, \forall i = 1, \dots, p + q - 1\}$ par $\{t_i = t + f_i(\tau_1, \dots, \tau_{p+q-1}), \forall i = 0, \dots, p + q - 1\}$, f_i étant une fonction linéaire. On obtiendra alors

$$C_{x(p)}^{\ell(q)}(t, \boldsymbol{\tau}) = x(t_0)x(t_1) \dots x(t_{p-1})x^*(t_p) \dots x^*(t_{p+q-1}).$$

Un argument permet toutefois de contraindre un peu le choix de ces instants : l'analyse des signaux non stationnaires devrait se réduire à une analyse stationnaire si les signaux considérés sont stationnaires. Autrement dit, après moyennage temporel, $C_{x(p)}^{\ell(q)}(t, \boldsymbol{\tau})$ ne doit plus dépendre du temps si $x(t)$ est

stationnaire. Ceci est assuré si les $p+q$ instants t_i vérifient les $p+q-1$ équations suivantes

$$\left\{ \begin{array}{l} t_1 - t_0 = \tau_1 \\ \vdots \\ t_{p-1} - t_0 = \tau_{p-1} \\ t_0 - t_p = \tau_p \\ \vdots \\ t_0 - t_{p+q-1} = \tau_{p+q-1}. \end{array} \right. \quad (4.12)$$

Il reste donc un degré de liberté pour figer le choix de ces instants. Un argument alors en général avancé est un argument de symétrie : la date t doit être au barycentre des dates. Mais cette propriété reste arbitraire et n'est pas aussi forte que l'argument de stationnarité précédemment invoqué. Notons que ce problème existe également pour les signaux cyclostationnaires.

Par la suite, les instants t_i seront notés $t + f_i(\boldsymbol{\tau})$ et il sera sous entendu que les deux arguments de compatibilité avec le cas stationnaire et de symétrie sont vérifiés.

Par exemple, à l'ordre 4 avec $p = 2$, la fonction de multicorrélation locale s'écrit

$$\begin{aligned} C_{x(2)}^{\ell(2)}(t, \boldsymbol{\tau}) &= x\left(t + \frac{-\tau_1 + \tau_2 + \tau_3}{4}\right)x\left(t + \frac{3\tau_1 + \tau_2 + \tau_3}{4}\right) \\ &\times x^*\left(t + \frac{-\tau_1 - 3\tau_2 + \tau_3}{4}\right)x^*\left(t + \frac{-\tau_1 + \tau_2 - 3\tau_3}{4}\right). \end{aligned}$$

Les arguments de x apparaissant dans cette expression sont obtenus en résolvant le système (4.12).

Nous disposons maintenant d'un objet qui capture le caractère non stationnaire des signaux. Pour obtenir une représentation temps-multifréquence il faut passer dans un domaine fréquentiel.

Représentation de Wigner-Ville Par analogie avec le cas stationnaire, un spectre peut être obtenu à partir de la fonction de multicorrélation locale en effectuant une transformée de Fourier des variables $\boldsymbol{\tau}$ vers des variables fréquentielles $\boldsymbol{\nu}$ pour obtenir

$$\mathcal{W}\mathcal{V}_{x(p)}^{(q)}(t, \boldsymbol{\nu}) = \int C_{x(p)}^{\ell(q)}(t, \boldsymbol{\tau}) \exp(-2\pi j \boldsymbol{\nu}^T \boldsymbol{\tau}) d\boldsymbol{\tau}.$$

La quantité $\mathcal{W}\mathcal{V}_{x(p)}^{(q)}(t, \boldsymbol{\nu})$ est une représentation temps-multifréquence du signal complexe $x(t)$ et est appelée représentation de Wigner-Ville. A l'ordre 2, cette représentation est maintenant bien connue [74]. Elle a été introduite à l'ordre 3 par N. Gerr en 1987 [86] et étendue à tout ordre par J. Fonollosa [75].

Cette dernière référence et [191] contiennent de nombreuses propriétés de la représentation de Wigner-Ville.

La représentation de Wigner-Ville d'ordre 4 avec $p = 2$ s'écrit alors

$$\begin{aligned} \mathcal{WV}_{x^{(2)}}^{(2)}(t, \boldsymbol{\tau}) &= \int x\left(t + \frac{-\tau_1 + \tau_2 + \tau_3}{4}\right) x\left(t + \frac{3\tau_1 + \tau_2 + \tau_3}{4}\right) \\ &\times x^*\left(t + \frac{-\tau_1 - 3\tau_2 + \tau_3}{4}\right) x^*\left(t + \frac{-\tau_1 + \tau_2 - 3\tau_3}{4}\right) \\ &\times e^{-2\pi j(\nu_1 \tau_1 + \nu_2 \tau_2 + \nu_3 \tau_3)} \mathbf{d}\boldsymbol{\tau}. \end{aligned}$$

L'intérêt des représentations de Wigner-Ville apparaît dans le paragraphe suivant.

Classes Générales Pour les signaux stationnaires, le multispectre est la transformée de Fourier de la multicorrélation qui est obtenue par moyennage de la fonction de multicorrélation locale. Si l'on examine la représentation de Wigner-Ville, on s'aperçoit que cette structure est conservée, au moyennage près. Tous les événements contenus dans un signal, qu'ils soient éloignés ou non de la date de référence t participent avec le même "poids" à la représentation en t . Or, il est légitime d'accorder plus d'importance aux instants proches de l'instant d'analyse.

Une façon de réaliser cela est de pondérer la multicorrélation locale avant d'effectuer la transformée de Fourier. Puisque la fonction de multicorrélation locale dépend à la fois de t et de $\boldsymbol{\tau}$, la pondération la plus générale doit dépendre également de t et $\boldsymbol{\tau}$.

Un autre point de vue pour comprendre cette pondération est de considérer la fonction de multicorrélation comme un estimateur de la multicorrélation des moments d'un signal aléatoire. Pour passer au multispectre, il faut alors "apodiser" en temps et/ou lisser en fréquence.

Une fenêtre de pondération sera alors $F(t, \boldsymbol{\tau})$ où la dépendance en t correspond au lissage et la dépendance en $\boldsymbol{\tau}$ concerne l'apodisation. Prendre la transformée de Fourier de la fonction de multicorrélation locale ainsi pondérée conduit à la définition de représentations temps-multifréquence sous la forme

$$\mathcal{R}_{x^{(p)}}^{(q)}(t, \boldsymbol{\nu}) = \int \int F(t - \theta, \boldsymbol{\tau}) C_{x^{(p)}}^{\ell(q)}(\theta, \boldsymbol{\tau}) \exp(-2\pi j \boldsymbol{\nu}^T \boldsymbol{\tau}) \mathbf{d}\boldsymbol{\tau} d\theta.$$

L'ensemble de ces représentations (paramétrées par F) définit une classe générale des représentations temps-multifréquence d'ordre $p + q$.

Les propriétés d'une représentation de cette classe générale dépendent alors évidemment de la fonction de pondération F . Toutefois, aux ordres pairs symétriques ($p = q$), toutes les représentations de ces classes possèdent la covariance par translation temporelle et par translation fréquentielle. Autrement dit, décaler un signal en temps décale d'autant suivant t la représentation temps-multifréquence.

Il est évident que la représentation de Wigner-Ville appartient à la classe générale. On montre alors facilement que toute représentation de la classe générale d'ordre $p + q$ s'obtient comme convolution multidimensionnelle de la représentation de Wigner-Ville d'ordre $p + q$,

$$\mathcal{R}_{x(p)}^{(q)}(t, \boldsymbol{\nu}) = \Pi(t, \boldsymbol{\nu}) *_t *_\boldsymbol{\nu} \mathcal{WV}_{x(p)}^{(q)}(t, \boldsymbol{\nu}), \quad (4.13)$$

où $\Pi(t, \boldsymbol{\nu})$ est la transformée de Fourier selon $\boldsymbol{\tau}$ de $F(t, \boldsymbol{\tau})$. L'argument de symétrie pris pour la définition de la transformée de Wigner-Ville peut alors se transférer dans la fonction de pondération. Cette remarque permet de comprendre l'importance de la disposition des instants dans la multiscorrélation locale : des choix différents de cette disposition n'altèrent pas l'information mais conduisent à des représentations (de l'information) différentes.

Pour illustrer la structure de la classe de Cohen, reprenons l'exemple récurrent de l'ordre 4, et lisons la représentation de Wigner-Ville du signal $x(t)$ par la représentation de Wigner-Ville d'une fenêtre $h(t)$. On montre alors facilement que $\mathcal{M}_{x(2)}^{(2)}(t, \boldsymbol{\nu}) = \left[\mathcal{WV}_{h(2)}^{(2)}(t, \boldsymbol{\nu}) \right]^* *_t *_\boldsymbol{\nu} \mathcal{WV}_{x(2)}^{(2)}(t, \boldsymbol{\nu})$ s'écrit

$$\mathcal{M}_{x(2)}^{(2)}(t, \boldsymbol{\nu}) = ST_{x,h}(t, -\nu_1 + \nu_2 + \nu_3) ST_{x,h}(t, \nu_1) ST_{x,h}^*(t, \nu_2) ST_{x,h}^*(t, \nu_3),$$

où $ST_{x,h}(t, \nu)$ est la transformée de Fourier à court terme de $x(t)$ à travers la fenêtre $h(t)$ définie par

$$ST_{x,h}(t, \nu) = \int x(\tau) h^*(t - \tau) \exp(2\pi j \nu \tau) d\tau.$$

La représentation $\mathcal{M}_{x(2)}^{(2)}(t, \boldsymbol{\nu})$ rejoint l'intuition : elle apparaît comme une version instantanée d'un multispectre fondé sur les moments. Par analogie avec le spectrogramme, cette représentation est appelée multispectrogramme et est la représentation temps-multifréquence la plus intuitive.

Remarques sur une construction axiomatique Les rappels effectués sur les représentations bilinéaires ont montré l'aspect axiomatique de leur construction. Cette construction est également possible pour les représentations temps-multifréquence.

Le principe de cette construction repose sur les significations souhaitées d'une représentation temps-multifréquence.

Tout d'abord, une représentation d'ordre $p + q$ a pour vocation d'étaler "l'énergie d'ordre $p + q$ " dans un hypercube temps-multifréquence. La forme générale est donc multilinéaire, et comme pour l'ordre deux, nous obtenons une forme générique en pondérant les produits de $p + q$ valeurs du signal pris à des instants différents. La forme multilinéaire d'un signal $x(t)$ la plus générale est

$$\mathcal{R}_{x(p)}^{(q)}(t, \boldsymbol{\nu}) = \int k(t, \boldsymbol{\nu}; \mathbf{t}) x(t_1) \dots x(t_p) x^*(t_{p+1}) \dots x^*(t_{p+q}) d\mathbf{t}.$$

Pour être utilisable, cette forme doit subir quelques contraintes. Or, une représentation temps-multifréquence doit posséder certaines propriétés, plus ou moins naturelles. Imposer ces propriétés à la forme générale permet de réduire sa complexité.

Nous proposons ici quelques propriétés :

1. Dans le cas stationnaire, la représentation devra être indépendante du temps.
2. Les marginales doivent permettre de retrouver les puissances instantanées et spectres d'ordre supérieur.
3. Les représentations seront covariantes par translations temporelle et fréquentielle.
4. Les supports des signaux seront conservés.
5. Le signal pourra être retrouvé à partir des représentations.
6. les représentations seront compatibles avec le filtrage linéaire et les modulations produits.
7. ...

Ces propriétés constituent des vœux, et toutes ne seront pas forcément respectées.

Toutefois, la propriété 3 permet une simplification étonnante puisque l'imposer à la forme générale conduit à la classe générale (voir [17]). Ce fait remarquable montre que la construction fondée sur l'intuition est valide.

4.3.2 Extension aux signaux aléatoires

Afin de traiter des signaux aléatoires non-stationnaires, il est nécessaire d'étendre les résultats précédents au cas aléatoire. Nous nous limiterons à la version aléatoire de la représentation de Wigner-Ville.

Multispectre de Wigner-Ville fondé sur les moments L'idée simple pour définir le multispectre de Wigner-Ville est de prendre l'espérance mathématique de la représentation de Wigner-Ville, c'est-à-dire $E[\mathcal{R}_{x(p)}^{(q)}(t, \boldsymbol{\nu})]$. Sous des hypothèses de mélange (sommabilité des moments), Blanc-Lapierre et Fortet [29] ont montré que l'espérance mathématique commute avec les intégrales. Ainsi, la définition du multispectre de Wigner-Ville fondé sur les moments est

$$\mathcal{WV}_{x(p)}^{(q)}(t, \boldsymbol{\nu}) = \int E[C_{x(p)}^{\ell(q)}(t, \boldsymbol{\tau})] \exp(-2\pi j \boldsymbol{\nu}^T \boldsymbol{\tau}) d\boldsymbol{\tau}. \quad (4.14)$$

Explicitement, à l'ordre 4 pour $p = 2$, cette définition devient

$$\begin{aligned} \mathcal{WV}_{x(2)}^{(2)}(t, \boldsymbol{\nu}) &= \int E[x(t + \frac{-\tau_1 + \tau_2 + \tau_3}{4})x(t + \frac{3\tau_1 + \tau_2 + \tau_3}{4}) \\ &\times x^*(t + \frac{-\tau_1 - 3\tau_2 + \tau_3}{4})x^*(t + \frac{-\tau_1 + \tau_2 - 3\tau_3}{4})] \\ &\times e^{-2\pi j(\nu_1 \tau_1 + \nu_2 \tau_2 + \nu_3 \tau_3)} d\boldsymbol{\tau}. \end{aligned}$$

Importance de la classe générale Un problème crucial levés par les signaux aléatoires non stationnaires est celui de l'estimation des grandeurs statistiques qui les décrivent. En effet, nous ne disposons en général que d'une réalisation du signal. Les méthodes du type périodogramme moyenné appliquées sur les signaux stationnaires ne sont donc plus envisageables. Les seuls moyens disponibles sont du type moyennage par lissage.

Ces lissages sont de deux types : lissage temporel et lissage fréquentiel. Nous avons déjà rencontré cette interprétation lors de la discussion sur la classe générale. En fait, l'estimation du multispectre de Wigner-Ville s'effectue en apodisant la multicorrélation locale (construite à partir de la réalisation du signal aléatoire observée) et en lissant temporellement par une fenêtre adéquate. L'estimateur le plus général s'écrit donc

$$\widehat{\mathcal{WV}}_{x(p)}^{(q)}(t, \boldsymbol{\nu}) = \int \int F(t - \theta, \boldsymbol{\tau}) C_{x(p)}^{\ell(q)}(\theta, \boldsymbol{\tau}) \exp(-2\pi j \boldsymbol{\nu}^T \boldsymbol{\tau}) d\boldsymbol{\tau} d\theta.$$

La classe générale apparaît donc sous un jour nouveau : elle est l'ensemble des estimateurs du multispectre de Wigner-Ville fondé sur les moments. Il est évident que la classe de Cohen conduit à des estimations biaisées à cause de la relation (4.13). Les variances d'estimation (non calculées) sont d'ordre $2(p + q)$ et dépendent des fenêtres de pondération choisies. Ces variances limitent l'utilisation de représentations d'ordres élevés.

Représentation fondée sur les cumulants Les constructions précédentes ont été effectuées sur les moments. Or, comme il apparaît souvent au cours de ce livre, les statistiques d'ordre supérieur les plus utilisées sont les cumulants, pour leurs diverses propriétés ("distance" à la gaussianité, mesure d'indépendance, ...). Pour les signaux aléatoires non stationnaires, il est tout à fait possible de définir les multispectres de Wigner-Ville fondés sur les cumulants. Il suffit pour cela de remplacer dans (4.14) l'opérateur espérance mathématique par l'opérateur cumulants. A l'ordre 4 pour $p = 2$ et $q = 2$, cette définition conduit à

$$\begin{aligned} \mathcal{WV}_{x(2)}^{(2)}(t, \boldsymbol{\nu}) &= \int Cum[x(t + \frac{-\tau_1 + \tau_2 + \tau_3}{4}), x(t + \frac{3\tau_1 + \tau_2 + \tau_3}{4}), \\ &x^*(t + \frac{-\tau_1 - 3\tau_2 + \tau_3}{4}), x^*(t + \frac{-\tau_1 + \tau_2 - 3\tau_3}{4})] \\ &e^{-2\pi j(\nu_1 \tau_1 + \nu_2 \tau_2 + \nu_3 \tau_3)} d\boldsymbol{\tau}. \end{aligned}$$

L'estimation de tels multispectres est évidemment plus difficile que l'estimation des multispectres fondés sur les moments, et leur utilisation est très limitée. Toutefois, le maniement théorique des objets fondés sur les cumulants peut conduire à développer des idées de traitement, dont les implantations numériques utilisent d'autres techniques (voir par exemple [18]).

Pour conclure ce petit aperçu sur les représentations temps-multifréquence, mentionnons la possibilité de réduire leur complexité (tout en perdant de l'information) en effectuant des coupes ou en sommant la représentation selon certaines fréquences pour obtenir des représentations temps-fréquence, mais fondées sur les ordres supérieurs. Ce genre d'idée fait l'objet de la première partie de la section suivante qui présente quelques autres approches.

4.3.3 Quelques autres approches

Temps-fréquence polynômial Une propriété de la représentation de Wigner-Ville bilinéaire est de se concentrer idéalement sur les modulations linéaires de fréquence (ou "chirp") : la représentation est alors une droite dans le plan temps-fréquence et reflète complètement la structure du signal. Pour d'autres types de modulation, il est parfois possible de trouver une représentation bilinéaire qui se concentre idéalement sur la loi de modulation. Toutefois, pour des modulations polynômiales, des représentations bilinéaires respectant cette concentration idéale ne sont pas connues.

Un type particulier de représentation temps-fréquence a alors été introduit [30, 80, 167]. L'idée est de construire une représentation non plus sur un noyau bilinéaire mais sur un noyau multilinéaire et de profiter de l'augmentation du nombre de degrés de liberté pour concentrer idéalement l'objet sur une modulation polynômiale.

Ces représentations peuvent être qualifiées d'adaptées car elle ne révèlent intimement la structure que d'un type de signal. De plus un de leurs problèmes est la multilinéarité qui en cas de mélanges de plusieurs signaux provoque un grand nombre d'interférences. Quoiqu'il en soit, l'utilisation de ces représentations adaptées peut s'avérer intéressante dans des circonstances très particulières.

La transformée bicorspectrale P. Duvaut et ses collaborateurs ont défini une alternative aux représentations temps-multifréquence à l'ordre trois. Cette transformée appelée bicorspectrale [70] s'écrit

$$B_{x,g}(t, \nu) = \frac{1}{2T} \int x_g(u-t)x_g(u+\frac{\tau}{2})x_g^*(u-\frac{\tau}{2})e^{-2\pi j\nu\tau} dud\tau,$$

où $x_g(t)$ représente le signal $x(t)$ pondéré par une fenêtre d'apodisation $g(t)$ de longueur $2T$.

Cette transformée s'interprète comme l'intercorrélation entre le signal et sa représentation de Wigner-Ville d'ordre 2. Cette remarque autorise une implantation numérique aisée.

De plus, dans le cas stationnaire, la transformée de Fourier de $B_{x,g}(t, \nu)$ constitue un estimateur asymptotiquement non biaisé du bispectre (dans le cas de signaux centrés).

Des textures ont été analysées à l'aide de cette transformée qui s'avère utile dans des contextes de classification.

CHAPITRE 5

Filtres de Volterra

La place d'un chapitre sur les filtres non linéaires dans un livre dédié aux statistiques d'ordre supérieur peut paraître étrange. Cette place devient naturelle dès que l'on examine d'un peu plus près le contenu du vocable “non-linéaire”.

Nous avons vu jusqu'ici que les statistiques d'ordre supérieur sont intéressantes lorsque les signaux considérés sont non gaussiens. Or, les systèmes non linéaires engendrent en général des signaux non gaussiens, et l'utilisation des SOS pour la description de ces systèmes est nécessaire. Contrairement au cas des systèmes linéaires, l'analyse du non linéaire requiert d'aller au-delà de l'ordre 2. Pour s'en convaincre, considérons l'exemple simple suivant.

Soit x une variable aléatoire de moyenne m . La moyenne de toute fonction linéaire de x est évidente à calculer puisque l'espérance mathématique est un opérateur linéaire. Soit maintenant la variable aléatoire $y = |x|$, fonction non linéaire de x . Calculer la moyenne de y nécessite de connaître les probabilités qu'a x d'être négative ou positive. Connaître ces probabilités revient à connaître la densité de probabilité de x , et donc de manière équivalente, de connaître l'ensemble de ses moments !

Une autre différence de taille entre le linéaire et le non-linéaire réside dans la présence du “non”. Si la linéarité est parfaitement définie, la non-linéarité est une “non-propriété” : il existe une infinité de types de non-linéarités. Tempérons en précisant que des classes de non-linéarités peuvent être construites, mais insistons sur le fait qu'une unification paraît impossible. Cette remarque conduit les utilisateurs de non-linéaire à restreindre leurs domaines d'étude à certaines classes de non-linéarités.

Dans cette optique, nous nous concentrerons dans ce chapitre sur un type particulier de systèmes non linéaires : les filtres de Volterra. Toutefois, quelques autres types de filtres non linéaires seront mentionnés en conclusion.

Nous commencerons par présenter les filtres de Volterra à temps continu. L'examen de ces filtres permettra d'appréhender leur comportement. En vue d'applications, le cas des filtres à temps discret sera présenté en détail, et le problème de leur identification en moyenne quadratique développé. Enfin, une

application des filtres de Volterra viendra conclure le chapitre.

5.1 Filtres de Volterra à temps continu

Dans ce premier paragraphe, nous donnons la définition des filtres de Volterra à temps continu. Pour comprendre leur fonctionnement, la représentation fréquentielle de ces filtres est également présentée. De plus, le calcul de quelques statistiques entrée-sortie permet d'étudier ces filtres en présence de signaux aléatoires.

5.1.1 Définition en temps et interprétations

Soient deux signaux $x(t)$ et $y(t)$ reliés par la relation fonctionnelle

$$y(t) = F(x(t)).$$

La notation peut paraître ambiguë. $F(\cdot)$ est une fonctionnelle¹, et son application à $x(t)$ fait intervenir les valeurs passées et futures de ce signal. Dans un cas plus général, $F(\cdot)$ peut dépendre explicitement du temps et s'écrire $F(\cdot, t)$. Toutefois, nous n'envisageons pas cette possibilité ici. Nous supposons que cette fonctionnelle est continue et qu'elle possède le développement suivant, dit de Volterra²

$$\begin{cases} y(t) &= h_0 + \sum_{i=1}^{+\infty} f_i(x(t)) \\ f_i(x(t)) &= \int h_i(\tau_1, \dots, \tau_i) x(t - \tau_1) \dots x(t - \tau_i) d\boldsymbol{\tau}. \end{cases} \quad (5.1)$$

La fonction $h_i(\tau_1, \dots, \tau_i)$ est appelée noyau d'ordre i du développement. La fonctionnelle multilinéaire $f_i(x(t))$ est dite homogène car elle ne comporte que des termes du même ordre, et par suite, $f_i(ax(t)) = a^i f_i(x(t))$ pour tout nombre a . On l'appelle filtre homogène d'ordre i .

Les noyaux sont en général considérés comme insensibles à toute permutation de leurs variables. Ils sont alors dits symétriques. Par exemple, nous supposons toujours que le noyau quadratique vérifie $h_2(\tau_1, \tau_2) = h_2(\tau_2, \tau_1)$. Si le noyau n'est pas symétrique, alors il suffit de le remplacer par $(h_2(\tau_1, \tau_2) + h_2(\tau_2, \tau_1))/2$ qui est symétrique et qui conduit à la même sortie $y(t)$ [182]. Cette méthode se généralise à tous les ordres.

¹Une fonctionnelle est une fonction de fonction.

²Les intégrales sont prises de $-\infty$ à $+\infty$, sauf mention explicite. De plus les intégrales multiples sont représentées sous forme d'un seul signe somme – s'il n'y a pas ambiguïté – et les éléments différentiels notés sous forme vectorielle.

Un développement de Volterra apparaît donc comme une généralisation des développements de Taylor, en y intégrant de la mémoire et de l'anticipation, puisqu'il consiste à ajouter des termes polynômiaux au linéaire. En effet, le développement tronqué à l'ordre 1 n'est autre qu'un filtre linéaire liant x à y , et l'on retrouve l'équation de convolution

$$y_1(t) = h_0 + \int h_1(\tau_1)x(t - \tau_1)d\tau_1.$$

Si le terme d'ordre 2 est ajouté, nous obtenons un filtre linéaire-quadratique qui s'écrit

$$y_2(t) = y_1(t) + \int h_2(\tau_1, \tau_2)x(t - \tau_1)x(t - \tau_2)d\tau_1d\tau_2.$$

Nous parlerons ici de filtres de Volterra, par extension du cas linéaire qui n'est autre qu'un développement de Volterra du premier ordre. Toutefois, le terme de développement de Volterra est aussi utilisé. En fait, deux visions équivalentes des filtres de Volterra peuvent être présentées, selon les points de vue adoptés : filtrage ou identification.

Pour le filtrage, une structure particulière est donnée *a priori* : un filtre de Volterra est un filtre polynômial (par rapport à l'entrée) avec mémoire et anticipation. Pour l'identification, la relation fonctionnelle supposée non linéaire entre deux signaux est approchée par un développement de type Taylor avec mémoire. Dans ce contexte, le terme utilisé est alors développement de Volterra. Mais bien évidemment, ces deux points de vue se raccordent, puisqu'en général, l'identification d'une relation est paramétrique : elle s'effectue dans une classe donnée d'objets dépendant de paramètres.

La question importante est celle de la validité du développement de Volterra comme approximation d'un lien supposé non linéaire. L'étude de cette validité est complexe et nous ne donnons que quelques éléments de réponse, ces théories dépassant l'objet de ce livre. L'étude des développements de fonctionnelles en séries de Volterra remonte au début du siècle avec les travaux de Maurice Fréchet [79]. Le résultat principal de ses investigations est le suivant : l'ensemble des fonctionnelles de Volterra forme un système complet dans l'ensemble des fonctionnelles continues. Il en résulte que les systèmes F continus, réalisables, peuvent être approchés avec une précision arbitrairement grande à l'aide de développements de Volterra (au moins sur un horizon de temps fini).

5.1.2 Critères physiques

Les signaux traités en théorie du signal sont soit certains, soit aléatoires. Pour rester proche de la physique, ces signaux doivent posséder des propriétés de type énergie finie ou puissance moyenne finie. De plus, pour approcher des

systèmes physiques par des développements de Volterra, ceux-ci se doivent de respecter certaines conditions de stabilité.

Examinons les conditions assurant cette stabilité. Le critère retenu ici est la stabilité “BIBO” [182], ou “Bounded Input-Bounded Output”, qui signifie qu’à une entrée bornée correspond une sortie bornée (voir le chapitre 3).

Soit donc $x(t)$ un signal borné ($|x(t)| \leq C < +\infty$). Alors, la sortie $y(t)$ d’un filtre de Volterra homogène d’ordre i est bornée si

$$|y(t)| = \left| \int h_i(\tau_1, \dots, \tau_i) x(t - \tau_1) \dots x(t - \tau_i) d\boldsymbol{\tau} \right| < +\infty.$$

Or,

$$\begin{aligned} |y(t)| &\leq \int |h_i(\tau_1, \dots, \tau_i) x(t - \tau_1) \dots x(t - \tau_i)| d\boldsymbol{\tau} \\ &\leq C^i \int |h_i(\tau_1, \dots, \tau_i)| d\boldsymbol{\tau}, \end{aligned}$$

et donc une condition suffisante de stabilité BIBO pour un filtre homogène³ est

$$\int |h_i(\tau_1, \dots, \tau_i)| d\boldsymbol{\tau} < +\infty. \quad (5.2)$$

Pour assurer la stabilité du système global, il faut assurer (5.2) pour tout i . Notons que cette condition n’est que nécessaire.

Venons-en au second problème. Si nous souhaitons représenter des signaux physiques comme sorties de filtres de Volterra, il est légitime de se limiter aux signaux d’énergie finie, c’est-à-dire appartenant à $L^2(\mathbb{R})$ dans le cas de signaux certains, ou à $f^{(2)}$ pour les signaux aléatoires⁴. Les traitements étant ici similaires, nous ne regardons que le cas de signaux certains.

Quelles sont les contraintes à donner aux entrées et aux noyaux pour qu’une sortie soit dans $L^2(\mathbb{R})$? Considérons $y_i(t)$ la sortie d’un filtre homogène d’ordre i . Nous souhaitons obtenir $\int |y_i(t)|^2 dt < +\infty$. Or⁵

$$\begin{aligned} \int |y_i(t)|^2 dt &= \int \left| \int h_i(\boldsymbol{\tau}) \prod_{j=1}^i x(t - \tau_j) d\boldsymbol{\tau} \right|^2 dt \\ &\leq \int \int \int |h_i(\boldsymbol{\tau}) h_i(\boldsymbol{\tau}')| \left| \prod_{j=1}^i x(t - \tau_j) \prod_{j=1}^i x(t - \tau'_j) \right| d\boldsymbol{\tau} d\boldsymbol{\tau}' dt \\ &\leq \int \int |h_i(\boldsymbol{\tau}) h_i(\boldsymbol{\tau}')| d\boldsymbol{\tau} d\boldsymbol{\tau}' \int |x(t)|^{2i} dt, \end{aligned}$$

³Cette condition est nécessaire uniquement dans le cas linéaire, i.e. $i = 1$. Pour $i \geq 1$, il est possible de construire des noyaux non intégrables qui génèrent un filtre stable [182].

⁴Pour la définition des classes $f^{(i)}$ de Blanc-Lapierre et Fortet, se reporter au paragraphe 2.9.2.

⁵Nous remplacerons parfois $h_i(\tau_1, \dots, \tau_i)$ par $h_i(\boldsymbol{\tau})$.

la dernière inégalité provenant de l'inégalité de Schwartz. Donc, si $\int |x(t)|^{2i} dt < +\infty$, i.e. $x(t) \in L^{2i}(\mathbb{R})$, et $\int |h(\boldsymbol{\tau})| d\boldsymbol{\tau} < +\infty$, alors $\int |y_i(t)|^2 dt < +\infty$. Ainsi, pour que $y_i(t)$ soit d'énergie finie, et donc possède une transformée de Fourier, il est suffisant que le filtre homogène soit stable et que l'entrée soit choisie dans $L^{2i}(\mathbb{R})$. Pour étendre ces résultats au filtre global (5.1), il faut que tous les filtres homogènes soient stables et que $x(t)$ soit dans $L^\infty(\mathbb{R})$. Cette condition n'est que nécessaire. Pour les signaux aléatoires, il faut que tous les filtres homogènes soient stables et que $x(t)$ soit dans $f^{(\infty)}$.

Ces critères physiques assurent la bonne adéquation entre théorie et signaux réels. De plus, ils permettent d'aborder la présentation des filtres dans le domaine de Fourier, où d'autres spécificités des filtres de Volterra apparaissent.

5.1.3 Point de vue fréquentiel

Il est bien connu que la quadrature de signaux à fréquence pure provoque un doublement de fréquence (effet rendu spectaculaire en optique non linéaire). Aussi, l'interprétation et la compréhension des phénomènes non linéaires peuvent être rendues aisées dans le domaine de Fourier. Ainsi, il est important dans certaines applications de connaître l'équivalent du développement de Volterra dans le domaine fréquentiel. Nous supposons que les différentes grandeurs $x(t)$, $h_i(\boldsymbol{\tau})$ et $y(t)$ possèdent des transformées de Fourier, ce qui est assuré si les critères du paragraphe précédent sont respectés.

Etant donnée la linéarité de la transformation de Fourier, il vient

$$Y(\nu) = TF[y(t)] = h_0\delta(\nu) + \sum_{i=1}^{+\infty} TF[f_i(x(t))].$$

L'évaluation de la transformée de Fourier de $f_i(x(t))$ s'effectue de la façon suivante. En utilisant à nouveau la linéarité de la transformée de Fourier, il vient

$$TF[f_i(x(t))] = \int h_i(\boldsymbol{\tau}) TF[x(t - \tau_1) \dots x(t - \tau_i)] d\boldsymbol{\tau}.$$

Or, il est facile de montrer que pour un signal $x(t)$

$$TF\left[\prod_{j=1}^i x(t)\right] = \int X(\nu_1) \dots X(\nu_{i-1}) X(\nu - \nu_1 - \dots - \nu_{i-1}) d\boldsymbol{\nu},$$

où $d\boldsymbol{\nu} = d\nu_1 \dots d\nu_{i-1}$. Par suite, on obtient

$$TF[f_i(x(t))] = \int h_i(\boldsymbol{\tau}) \int X(\nu_1) \dots X(\nu_{i-1}) X(\nu - \nu_1 - \dots - \nu_{i-1}) \exp(-2\pi j(\sum_{k=1}^{i-1} \nu_k \tau_k)) \exp(-2\pi j(\nu - \nu_1 - \dots - \nu_{i-1})\tau_i) d\boldsymbol{\nu} d\boldsymbol{\tau},$$

soit enfin

$$TF[f_i(x(t))] = \int H_i(\nu_1, \dots, \nu_{i-1}, \nu - \nu_1 - \dots - \nu_{i-1}) \\ \times X(\nu_1) \dots X(\nu_{i-1}) X(\nu - \nu_1 - \dots - \nu_{i-1}) d\boldsymbol{\nu}.$$

L'expression cherchée est finalement

$$Y(\nu) = h_0 \delta(\nu) + \sum_{i=1}^{+\infty} \int H_i(\nu_1, \dots, \nu_{i-1}, \nu - \nu_1 - \dots - \nu_{i-1}) \\ \times X(\nu_1) \dots X(\nu_{i-1}) X(\nu - \nu_1 - \dots - \nu_{i-1}) d\boldsymbol{\nu}, \quad (5.3)$$

où $H_i(\nu_1, \dots, \nu_i)$ est la transformée de Fourier du noyau $h_i(\tau_1, \dots, \tau_i)$. Notons que les symétries supposées des noyaux impliquent évidemment des symétries sur les $H_i(\boldsymbol{\nu})$. Ces fonctions sont insensibles à toute permutation de leurs variables, et vérifient $H_i(\boldsymbol{\nu}) = H_i^*(-\boldsymbol{\nu})$ puisque les noyaux sont à valeurs réelles. D'autres symétries existent mais sont difficiles à mettre en évidence de manière générale. La symétrie pour $i = 2$ sera examinée au paragraphe 5.2.1.

Le résultat (5.3) est valide pour des signaux possédant une transformée de Fourier (au sens des fonctions). Lorsque les signaux d'entrée ne possèdent pas de transformée de Fourier, il est tout de même possible à l'aide de la théorie des distributions tempérées de définir une version fréquentielle des filtres de Volterra [19]. L'utilisation des distributions est, par exemple, utile lorsque l'on traite des signaux harmoniques.

La forme (5.3) englobe bien évidemment le cas particulier du filtre linéaire pour lequel la sortie en fréquence est égale au produit entre la fonction de transfert $H_1(\nu)$ et la transformée de Fourier de l'entrée. Remarquons de plus la structure "un peu" simplifiée du filtre par le passage en fréquence. Il s'agit toujours d'une sorte de convolution, mais dont l'ordre est abaissé de 1 par rapport à la version temporelle. Enfin, l'interaction entre les diverses composantes fréquentielles de l'entrée apparaît clairement sur la version fréquentielle. A titre d'exemple, examinons le cas d'un filtre quadratique (ordre 2 seul) attaqué par une fréquence pure ν_0 d'amplitude unité. La sortie d'un tel filtre s'écrit

$$Y(\nu) = \frac{1}{4} (H_2(\nu_0, \nu_0) \delta(\nu - 2\nu_0) + [H_2(\nu_0, -\nu_0) + H_2(-\nu_0, \nu_0)] \delta(\nu) \\ + H_2(-\nu_0, -\nu_0) \delta(\nu + 2\nu_0)). \quad (5.4)$$

Cet exemple met en évidence le phénomène de doublement de fréquence avec l'apparition dans la sortie $y(t)$ de la fréquence double de l'entrée, et l'interaction non linéaire entre les fréquences ν_0 et $-\nu_0$ de l'entrée qui crée une composante continue dans la sortie.

Nous conviendrons d'appeler les fonctions $H_i(\nu_1, \dots, \nu_i)$ fonctions de transfert d'ordre i du filtre de Volterra. Ces fonctions pondèrent, comme on

l'a remarqué dans l'exemple précédent, les interactions non linéaires entre les diverses fréquences du signal. Elles indiquent de plus quelles fréquences ont interagi. Dans (5.4), $H_2(\nu_0, \nu_0)\delta(\nu - 2\nu_0)$ indique que $2\nu_0$ provient de l'interaction quadratique entre ν_0 et elle-même. $H_2(-\nu_0, \nu_0)\delta(\nu)$ montre que la fréquence nulle provient du couplage entre les composantes $-\nu_0$ et ν_0 du signal d'entrée.

Ces interactions peuvent également être mises en évidence dans le cas de signaux aléatoires en examinant des statistiques entrée-sortie.

5.1.4 Cas des signaux aléatoires : statistiques entrée-sortie

Pour le cas de signaux aléatoires, les statistiques de la sortie et les statistiques entrée-sortie sont fondamentales pour comprendre quels effets produit un filtre de Volterra. Toutefois, le caractère polynômial des filtres rend le calcul de ces statistiques excessivement lourd. Nous nous limiterons donc à évaluer les statistiques d'ordre 2 de la sortie, et certaines statistiques d'ordre 3, uniquement dans le cas de filtres linéaire-quadratique. Ces quelques calculs mettront en lumière le comportement des filtres de Volterra.

Soit donc $x(t)$ un signal aléatoire stationnaire centré, et soit $y(t)$ la sortie d'un filtre linéaire-quadratique excité par $x(t)$

$$y(t) = h_0 + \int h_1(\tau)x(t - \tau)d\tau + \int h_2(\tau_1, \tau_2)x(t - \tau_1)x(t - \tau_2)d\tau_1d\tau_2.$$

La première statistique est la moyenne de $y(t)$. Elle s'écrit

$$E[y(t)] = h_0 + \int h_2(\tau_1, \tau_2)C_{x(2)}(\tau_1 - \tau_2)d\tau_1d\tau_2.$$

Cette équation montre que la stationnarité d'ordre 2 de l'entrée et la stationnarité du filtre implique la stationnarité d'ordre 1 de la sortie. D'une manière plus générale, pour un filtre linéaire-quadratique stationnaire, la stationnarité d'ordre $2p$ de l'entrée assure la stationnarité d'ordre p de la sortie. Nous supposons cette condition vérifiée.

La corrélation de $y(t)$ est donnée par $C_{y(2)}(\mu) = Cum[y(t), y(t + \mu)]$. Etant donné la multilinéarité des cumulants et leur invariance par translation, le calcul conduit à

$$\begin{aligned} C_{y(2)}(\mu) &= \int \int h_1(\tau)h_1(\tau')C_{x(2)}(\mu - \tau' + \tau)d\tau d\tau' \\ &+ \int \int h_1(\tau)h_2(\tau_1, \tau_2)C_{x(3)}(\tau_1 - \tau_2, \mu - \tau + \tau_1)d\tau d\tau_1d\tau_2 \\ &+ \int \int h_1(\tau)h_2(\tau_1, \tau_2)C_{x(3)}(\tau_1 - \tau_2, -\mu - \tau + \tau_1)d\tau d\tau_1d\tau_2 \end{aligned}$$

$$\begin{aligned}
& + \int \int h_2(\tau_1, \tau_2) h_2(\tau'_1, \tau'_2) \\
& \times [C_{x(4)}(\tau_1 - \tau_2, \mu + \tau_1 - \tau'_1, \mu + \tau_1 - \tau'_2) + \\
& 2C_{x(2)}(\mu - \tau'_1 + \tau_1) C_{x(2)}(\mu - \tau'_2 + \tau_2)] d\tau_1 d\tau_2 d\tau'_1 d\tau'_2,
\end{aligned}$$

où l'on rappelle que $C_{x(3)}(\tau_1, \tau_2) = Cum[x(t), x(t + \tau_1), x(t + \tau_2)]$ et $C_{x(4)}(\tau_1, \tau_2, \tau_3) = Cum[x(t), x(t + \tau_1), x(t + \tau_2), x(t + \tau_3)]$. Le point intéressant à noter est l'interaction statistique qui apparaît entre les termes linéaire et quadratique via les statistiques d'ordre 3. Il est donc très difficile de chiffrer l'influence du terme quadratique sur la sortie, puisqu'il intervient à la fois seul et par le couplage avec le linéaire.

Si l'entrée est gaussienne, les parties linéaire et quadratique se découplent, et l'analyse des contributions à la sortie des deux termes est plus aisée. En effet, sous l'hypothèse $x(t)$ gaussien, par suite de l'annulation des multicorrélations d'ordre supérieur à 2, la corrélation de $y(t)$ se réduit à

$$\begin{aligned}
C_{y(2)}(\mu) & = \int \int h_1(\tau) h_1(\tau') C_{x(2)}(\mu - \tau' + \tau) d\tau d\tau' \\
& + 2 \int \int h_2(\tau_1, \tau_2) h_2(\tau'_1, \tau'_2) \\
& \times C_{x(2)}(\mu - \tau'_1 + \tau_1) C_{x(2)}(\mu - \tau'_2 + \tau_2) d\tau_1 d\tau_2 d\tau'_1 d\tau'_2.
\end{aligned}$$

Par suite de la non-linéarité du filtre, les multicorrélations d'ordre supérieur à 2 de $y(t)$ sont non nulles. Ces multicorrélations pourraient se calculer de la même manière, mais conduisent à des calculs très fastidieux.

D'autres statistiques intéressantes ici sont les statistiques d'interaction entrée-sortie, puisqu'elles permettent de comprendre comment s'effectue le transfert d'énergie entre $x(t)$ et $y(t)$. Commençons par l'intercorrélation $C_{y,x(2)}(\mu) = Cum[y(t), x(t + \mu)]$ pour une entrée non gaussienne

$$\begin{aligned}
C_{y,x(2)}(\mu) & = \int h_1(\tau) C_{x(2)}(\mu + \tau) d\tau \\
& + \int h_2(\tau_1, \tau_2) C_{x(3)}(\tau_1 - \tau_2, \mu + \tau_1) d\tau_1 d\tau_2. \quad (5.5)
\end{aligned}$$

L'interbicorrélation $C_{y,x(3)}(\mu_1, \mu_2) = Cum[y(t), x(t + \mu_1), x(t + \mu_2)]$ s'écrit

$$\begin{aligned}
C_{y,x(3)}(\mu_1, \mu_2) & = \int h_1(\tau) C_{x(3)}(\mu_1 + \tau, \mu_2 + \tau) d\tau \\
& + \int h_2(\tau_1, \tau_2) [C_{x(4)}(\tau_1 - \tau_2, \mu_1 + \tau_1, \mu_2 + \tau_1) + \\
& 2C_{x(2)}(\mu_1 + \tau_1) C_{x(2)}(\mu_2 + \tau_2)] d\tau_1 d\tau_2. \quad (5.6)
\end{aligned}$$

Encore une fois, ces deux statistiques d'interaction font clairement apparaître le couplage entre les termes linéaire et quadratique. De plus, elles suggèrent une méthode pour trouver les noyaux $h_1(\tau)$ et $h_2(\tau_1, \tau_2)$ lorsqu'elles sont connues et que les statistiques de l'entrée sont à disposition : il suffit de résoudre le système d'équations intégrales défini par (5.5) et (5.6) !

Ces différentes relations entrée-sortie ont évidemment leurs équivalents dans le domaine fréquentiel. Nous les donnons dans le cas où $x(t)$ est supposé gaussien. Alors le spectre de $y(t)$ s'écrit pour une fréquence ν non nulle

$$S_{y(2)}(\nu) = |H_1(\nu)|^2 S_{x(2)}(\nu) + 2 \int |H_2(\nu_1, \nu - \nu_1)|^2 S_{x(2)}(\nu_1) S_{x(2)}(\nu - \nu_1) d\nu_1.$$

L'interspectre entre $y(t)$ et $x(t)$ est (puisque'il n'y a pas d'interaction entre le linéaire et le quadratique)

$$S_{y,x(2)}(\nu) = H_1^*(\nu) S_{x(2)}(\nu),$$

alors que l'interbispectre prend la forme

$$S_{y,x(3)}(\nu_1, \nu_2) = 2H_2^*(\nu_1, \nu_2) S_{x(2)}(\nu_1) S_{x(2)}(\nu_2).$$

Ces deux dernières équations permettent, dans le cas gaussien, d'obtenir les noyaux linéaire et quadratique.

5.1.5 Modéliser et détecter une non-linéarité

Dans beaucoup de domaines de la physique, l'approximation linéaire est faite, souvent à bon escient, mais également souvent par souci de simplicité. Toutefois, des problèmes curieux ne peuvent être traités dans cette approximation, et des modélisations non linéaires doivent être adoptées. A titre d'exemple, examinons le doublement de fréquence en optique [78].

Lorsque la lumière traverse un matériau, le champ électrique exerce une force de polarisation sur les électrons (cet effet joue surtout sur les électrons de valence, ou électrons assurant la liaison entre atomes). Lorsque la source lumineuse est peu puissante, la polarisation est en général linéaire en fonction des champs, soit $\mathbf{P} = \chi^1 \mathbf{E}$ où χ^1 est appelée susceptibilité électrique. L'effet est ici une petite perturbation sur le champ liant l'électron à l'atome. Par contre, si le champ électrique est puissant, soit de l'ordre de grandeur du champ liant l'électron à l'atome, les effets du champ électrique ne peuvent plus être considérés comme perturbations et la loi liant la polarisation au champ électrique n'est plus linéaire. Les physiciens ont alors modélisé cette loi selon

$$\mathbf{P} = \chi^1 \mathbf{E} + \chi^2 \mathbf{E}\mathbf{E} + \chi^3 \mathbf{E}\mathbf{E}\mathbf{E} + \dots,$$

où les termes χ^i sont des tenseurs et où \mathbf{EE} représente le produit tensoriel de \mathbf{E} par lui-même. Les études se limitent souvent à l'ordre 2, et le développement précédent se réécrit pour la i -ème composante de \mathbf{P}

$$P_i = \sum_j \chi_{i,j}^1 E_j + \sum_{j,k} \chi_{i,j,k}^2 E_j E_k.$$

Cette dernière relation n'est autre qu'un filtrage linéaire-quadratique!

Si le champ incident est une onde plane de fréquence ν , la dernière équation montre que la polarisation va osciller non seulement à ν mais également à 2ν . Or, dans les matériaux l'influence de la polarisation dans les équations de Maxwell peut être vue comme un terme source. Donc, le terme en 2ν de la polarisation "génère" un terme en 2ν sur le champ électrique, terme qui se propagera à travers le matériau et sera vu en sortie.

Cet exemple montre que les développements de Volterra sont utiles pour modéliser des non-linéarités et par suite pour les détecter. En fait, la détection des non-linéarités utilise deux outils. Les filtres de Volterra pour la modélisation, et les multispectres pour la détection. Pour illustrer ceci, nous nous restreignons encore au cas linéaire-quadratique.

Soit $y(t)$ la sortie d'un filtre linéaire-quadratique attaqué par un signal $x(t)$, centré, gaussien. Le spectre de $y(t)$ s'écrit pour une fréquence ν non nulle

$$S_{y(2)}(\nu) = |H_1(\nu)|^2 S_{x(2)}(\nu) + 2 \int |H_2(\nu_1, \nu - \nu_1)|^2 S_{x(2)}(\nu_1) S_{x(2)}(\nu - \nu_1) d\nu_1, \quad (5.7)$$

et l'interspectre entre $x(t)$ et $y(t)$ est donné par

$$S_{y,x(2)}(\nu) = H_1(\nu)^* S_{x(2)}(\nu). \quad (5.8)$$

Nous disposons des outils pour évaluer la cohérence entre $x(t)$ et $y(t)$, grandeur qui chiffre la linéarité entre ces deux signaux. La cohérence est pour une fréquence non nulle

$$\gamma_{y,x(2)}(\nu) = \frac{|S_{y,x(2)}(\nu)|^2}{S_{x(2)}(\nu) S_{y(2)}(\nu)}.$$

En utilisant les résultats précédents, elle s'écrit

$$\gamma_{y,x(2)}(\nu) = \frac{|H_1(\nu)|^2 S_{x(2)}(\nu)}{|H_1(\nu)|^2 S_{x(2)}(\nu) + 2 \int |H_2(\nu_1, \nu - \nu_1)|^2 S_{x(2)}(\nu_1) S_{x(2)}(\nu - \nu_1) d\nu_1}.$$

Dans le cas d'un filtre linéaire, elle vaut 1, alors que la présence d'un terme quadratique la rend inférieure à l'unité. Toutefois, une cohérence inférieure à

1 ne signifie pas forcément présence d'une non-linéarité, car un bruit polluant la sortie aurait le même effet. La cohérence ne peut donc pas détecter les non-linéarités puisqu'en contexte réel, un bruit pollueur est toujours présent. Tempérons en disant qu'une très faible cohérence doit nous mettre "la puce à l'oreille" quant à l'existence d'une non-linéarité de transfert.

Il est un outil mieux adapté aux non-linéarités quadratiques, la bicohérence. Sa définition est un prolongement de celle de la cohérence, puisqu'elle s'écrit

$$\gamma_{y,x(3)}(\nu_1, \nu_2) = \frac{|S_{y,x(3)}(\nu_1, \nu_2)|^2}{S_{x(2)}(\nu_1)S_{x(2)}(\nu_2)S_{y(2)}(\nu_1 + \nu_2)}. \quad (5.9)$$

Dans le cas que nous traitons, $S_{y,x(3)}(\nu_1, \nu_2) = 2S_{x(2)}(\nu_1)S_{x(2)}(\nu_2)H_2(\nu_1, \nu_2)^*$ permet d'obtenir la fonction de transfert d'ordre 2. La bicohérence n'a alors pas en général de forme exploitable, mais une expression agréable peut être obtenue à partir du spectre de $y(t)$ [116]. En effet, en remplaçant dans (5.7) $H_2(\nu_1, \nu_2)$ par sa valeur, puis en divisant par $S_{y(2)}(\nu)$ on obtient

$$\gamma_{y,x(2)}(\nu) + \frac{1}{2} \int \gamma_{y,x(3)}(\nu_1, \nu - \nu_1) d\nu_1 = 1.$$

Cette relation montre comment l'énergie se transfère de l'entrée vers la sortie. La contribution du linéaire est quantifiée par la cohérence, alors que la contribution du quadratique est quantifiée par le second terme. Pour une fréquence ν de la sortie, il faut sommer toutes les contributions des fréquences ν_1 et ν_2 de l'entrée qui interagissent quadratiquement pour participer à la fréquence $\nu = \nu_1 + \nu_2$ de la sortie. Ainsi, dans le cas linéaire-quadratique, la bicohérence permet de quantifier l'importance du quadratique dans le transfert. De plus, son comportement est remarquable en présence de fréquences pures⁶. Reprenons l'exemple présenté au paragraphe 5.1.3 concernant le passage d'une fréquence pure ν_0 dans un filtre quadratique. Si l'on ne s'intéresse qu'aux fréquences positives, une application directe de (5.9) montre que

$$\gamma_{y,x(3)}(\nu_1, \nu_2) = \delta(\nu_1 - \nu_0)\delta(\nu_2 - \nu_0).$$

Ce résultat signifie que les fréquences de $x(t)$ $\nu_1 = \nu_0$ et $\nu_2 = \nu_0$ ont interagi quadratiquement pour former la fréquence somme, ici $2\nu_0$. La bicohérence est donc un outil adapté pour détecter des non-linéarités quadratiques de transfert, et elle sera utilisée dans le paragraphe (5.4.2) pour démontrer la présence d'une non-linéarité.

⁶ Lorsque les signaux sont des fréquences pures, les définitions données ne sont plus valides. Il faut en toute rigueur donner des définitions en terme de distributions ou d'accroissements harmoniques si une modélisation stochastique est adoptée. Toutefois, les résultats peuvent se déduire des définitions présentées, les démonstrations étant faites "avec les mains". Pour des définitions précises, voir [66].

Symétries de la bicohérence La bicohérence entrée-sortie (5.9) possède les symétries de l'interbispectre $S_{y,x(3)}(\nu_1, \nu_2)$ qui sont différentes des symétries du bispectre $S_{x,3}$. En effet, les symétries sont en nombre plus restreint puisque les seules relations sont maintenant $S_{y,x(3)}(\nu_1, \nu_2) = S_{y,x(3)}(\nu_2, \nu_1) = \left[S_{y,x(3)}(-\nu_1, -\nu_2) \right]^*$. La première égalité induit une symétrie par rapport à la première bissectrice alors que la deuxième est une symétrie centrale par rapport à l'origine. Ainsi, la connaissance de la bicohérence entrée-sortie est totale sur le domaine $\{\nu_1 - \nu_2 \geq 0\} \cap \{\nu_1 + \nu_2 \geq 0\}$. Pour le bispectre (et donc pour "l'auto"-bicohérence) il faut ajouter la condition $\{\nu_2 \geq 0\}$.

5.2 Filtres de Volterra discrets

Dans ce paragraphe, nous examinons les filtres de Volterra discrets à ordre et mémoire finis. Les définitions données, nous présentons la mise sous forme vectorielle des filtres en vue de leur utilisation pratique.

5.2.1 Définitions

Nous pouvons, comme pour les multispectres, donner les définitions pour des signaux discrets à support infini. Ce sont les transpositions directes du cas temps continu. En temps discret, on a⁷

$$\begin{cases} y(n) &= h_0 + \sum_{i=1}^{+\infty} f_i(x(n)) \\ f_i(x(n)) &= \sum_{\mathbf{k}=-\infty}^{+\infty} h_i(k_1, \dots, k_i) x(n - k_1) \dots x(n - k_i). \end{cases}$$

En utilisant la transformée de Fourier en fréquences réduites, nous obtenons l'analogie de cette définition dans le domaine des fréquences

$$\begin{aligned} Y(\lambda) &= h_0 \delta(\lambda) + \sum_{i=1}^{+\infty} \int_{-\frac{1}{2}}^{+\frac{1}{2}} H_i(\lambda_1, \dots, \lambda_{i-1}, \lambda - \lambda_1 - \dots - \lambda_{i-1}) \\ &\times X(\lambda_1) \dots X(\lambda_{i-1}) X(\lambda - \lambda_1 - \dots - \lambda_{i-1}) d\boldsymbol{\lambda}. \end{aligned}$$

Filtres à ordre et mémoire finis Ces définitions étant présentées, nous pouvons remarquer que pratiquement, elles sont inapplicables ! En effet, l'ordre et la mémoire du développement en tant qu'approximation d'un lien non

⁷La notation $\sum_{\mathbf{k}=a}^b$ est équivalente à $\sum_{k_1=a}^b \dots \sum_{k_i=a}^b$.

linéaire ne peuvent être rendus infinis. Nous définissons donc un développement de Volterra à ordre K et mémoire M finis par

$$\left\{ \begin{array}{l} y(n) = h_0 + \sum_{i=1}^K f_i(x(n)) \\ f_i(x(n)) = \sum_{\mathbf{k}=0}^{M-1} h_i(k_1, \dots, k_i) x(n - k_1) \dots x(n - k_i). \end{array} \right. \quad (5.10)$$

Si $X(m)$ est la transformée de Fourier discrète du signal échantillonné $x(n)$, alors la définition généralement admise dans la littérature des filtres de Volterra en fréquence est

$$\begin{aligned} Y(m) &= h_0 \delta(m) + \sum_{i=1}^K \sum_{m_1, \dots, m_{i-1}} H_i(m_1, \dots, m_{i-1}, m - m_1 - \dots - m_{i-1}) \\ &\quad \times X(m_1) \dots X(m_{i-1}) X(m - m_1 - \dots - m_{i-1}). \end{aligned}$$

Toutefois, l'obtention de cette équation n'est pas simple et cette définition est sujette à caution. En effet, il est possible de montrer à l'aide de la théorie des distributions qu'elle ne constitue qu'une approximation de la discrétisation de la définition continue [19] (contrairement au cas linéaire). De plus, elle sous-entend des contraintes sévères sur l'échantillonnage des signaux, beaucoup plus restrictives que la condition de Shannon. Illustrons ceci en prenant l'exemple des filtres quadratiques. Supposons que l'entrée soit de bande passante $[-B, B]$. Alors la sortie sera de bande double, soit $[-2B, 2B]$. Si l'on cherche le lien entre l'entrée et la sortie, il est alors clair que la fréquence d'échantillonnage doit vérifier $\nu_e \geq 4B$ pour respecter le théorème de Shannon sur la sortie!

Symétries et réduction du nombre de paramètres Nous avons vu au paragraphe 5.1.1 que les noyaux sont en général supposés symétriques (insensibles à toute permutation de leurs variables). Il est évident que cette propriété est conservée dans le cas discret. Mais dans ce contexte, ces symétries deviennent intéressantes en vue de la réduction du nombre de paramètres définissant les filtres.

Le nombre de paramètres définissant un filtre de Volterra d'ordre K et de mémoire M est $\sum_{i=1}^K M^i$. Etant donné la symétrie supposée des noyaux, il est intéressant de réécrire (5.10) sous la forme

$$y(n) = h_0 + \sum_{i=1}^K \sum_{k_i \geq k_{i-1}} h_i(k_1, \dots, k_i) x(n - k_1) \dots x(n - k_i).$$

Ce filtre est alors dit non redondant. Dans ce cas, le nombre de paramètres se réduit à $\sum_{i=1}^K \frac{(M+i-1)!}{(M-1)!i!}$. L'ordre de grandeur reste toujours le même mais la réduction est tout de même significative.

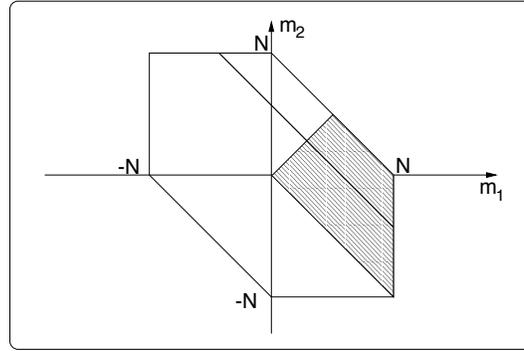


Figure 5.1: *Support de la fonction de transfert quadratique. La zone hachurée permet de connaître cette fonction sur tout son support. La ligne grasse représente la droite d'équation $m = m_1 + m_2$ des fréquences m_1 et m_2 de l'entrée qui participent à la fréquence m de la sortie.*

En ce qui concerne les fonctions de transfert, l'étude est plus délicate. Examinons toutefois le cas des noyaux quadratiques. Puisque $h_2(k_1, k_2) = h_2(k_2, k_1)$ et est réel, $H_2(m_1, m_2) = H_2(m_2, m_1) = H_2^*(-m_1, -m_2)$. De plus, pour la fréquence m de la sortie, seules les fréquences m_1 et m_2 interviennent via $m_1 + m_2 = m$. Donc, si (m_1, m_2) varie dans $\{-N, N\}^2$, il suffira de connaître $H_2(m_1, m_2)$ sur le domaine

$$D = \{(m_1, m_2) \in \{-N, N\}^2\} \cap \{m_2 \leq N - m_1\} \\ \cap \{m_2 - m_1 \geq 0\} \cap \{m_2 + m_1 \geq 0\},$$

dont la représentation est donnée en hachuré⁸ sur la figure (5.1). Les études de symétries pour $K = 3$ peuvent être obtenue dans [155].

5.2.2 Mise sous forme vectorielle

Une propriété importante des filtres de Volterra n'a pas encore été mise en avant : *ils sont linéaires en paramètres*. Ceci signifie que la somme pondérée de deux filtres de Volterra attaqués par une même entrée $x(t)$ (ou $x(n)$ en discret) est encore un filtre de Volterra dont les noyaux d'ordre i sont les sommes pondérées des noyaux d'ordre i des deux filtres. Autrement dit, l'ensemble des filtres de Volterra d'entrée $x(t)$ est un espace vectoriel.

Dans le cas discret à ordre et mémoire finis, la dimension de cet espace est finie. Ainsi, la sortie $y(n)$ peut s'écrire sous la forme d'un produit scalaire entre un vecteur paramètre et un pseudo-vecteur de données⁹.

⁸Remarquons que les symétries du noyau quadratique sont les mêmes que celles de la bicohérence entrée-sortie.

⁹La dénomination pseudo-vecteur de données est explicitée dans les pages suivantes.

Cas temporel L'idée consiste à ranger dans un pseudo-vecteur la suite des combinaisons non linéaires des entrées participant au filtre. Pour la partie linéaire, les termes $x(n), x(n-1), \dots, x(n-M+1)$ se rangent dans le pseudo-vecteur $\mathbf{x}_{n,1} = (x(n)x(n-1)\dots x(n-M+1))^T$. Le vecteur paramètre correspondant sera écrit $\mathbf{h}_1 = (h(0)h(1)\dots h(M-1))^T$ de sorte que

$$\mathbf{h}_1^T \mathbf{x}_{n,1} = \sum_{k_1=0}^{M-1} h(k_1)x(n-k_1).$$

Les termes quadratiques sont ensuite rangés dans un pseudo-vecteur $\mathbf{x}_{n,2}$ et le filtre quadratique dans un vecteur \mathbf{h}_2 . $\mathbf{x}_{n,2}$ et \mathbf{h}_2 s'écrivent

$$\mathbf{x}_{n,2} = \begin{pmatrix} x(n)^2 \\ x(n)x(n-1) \\ \vdots \\ x(n)x(n-M+1) \\ x(n-1)^2 \\ x(n-1)x(n-2) \\ \vdots \\ x(n-1)x(n-M+1) \\ \vdots \\ x(n-M+2)^2 \\ x(n-M+2)x(n-M+1) \\ x(n-M+1)^2 \end{pmatrix} \quad \text{et} \quad \mathbf{h}_2 = \begin{pmatrix} h(0,0) \\ h(0,1) \\ \vdots \\ h(0,M-1) \\ h(1,1) \\ h(1,2) \\ \vdots \\ h(1,M-1) \\ \vdots \\ h(M-2,M-2) \\ h(M-2,M-1) \\ h(M-1,M-1) \end{pmatrix}.$$

Le terme quadratique participant à $y(n)$ s'écrit alors simplement $\mathbf{h}_2^T \mathbf{x}_{n,2}$. D'une façon générale, un pseudo-vecteur de données à l'ordre i est noté $\mathbf{x}_{n,i}$ et contient les termes du type $x(n-k_1)\dots x(n-k_i)$ pour $k_i \leq k_{i+1}$. Le vecteur paramètre \mathbf{h}_i correspondant contient les $h_i(k_1, \dots, k_i)$ rangés de la même façon. Le terme d'ordre i dans $y(n)$ est donc $\mathbf{h}_i^T \mathbf{x}_{n,i}$.

La dernière étape consiste à concaténer les différents vecteurs et pseudo-vecteurs dans deux vecteurs

$$\begin{cases} \mathbf{x}_n^T = \mathbf{x}_{n,1}^T \dots \mathbf{x}_{n,i}^T \\ \mathbf{h}^T = \mathbf{h}_1^T \dots \mathbf{h}_i^T. \end{cases}$$

Le filtre de Volterra non redondant s'écrit finalement

$$y(n) = h_0 + \mathbf{h}^T \mathbf{x}_n.$$

Structure d'espace vectoriel Définissons l'ensemble suivant

$$VF_{x,K,M} = \{y(n) = \mathbf{h}^T \mathbf{x}_n \mid x(n) \in f^{(2K)} \text{ et } \sum_k |h_k| < +\infty\},$$

où h_k est le k ième élément du vecteur \mathbf{h}^T . Notons que cet ensemble dépend explicitement de l'entrée $x(n)$ choisie. Il est clair que cet espace est un espace vectoriel (sur les réels) de dimension finie égale à $\sum_{i=1}^K \frac{(M+i-1)!}{(M-1)!i!}$. Précisons que $V_{F_{x,K,M}}$ est un sous-espace vectoriel de l'espace vectoriel des signaux discrets de puissance moyenne finie. Ainsi, nous pouvons le doter du produit scalaire "naturel" $\langle y_1(n)|y_2(n) \rangle = E[y_1(n)y_2(n)]$. En utilisant la norme associée $\|y(n)\| = \sqrt{\langle y(n)|y(n) \rangle}$, $V_{F_{x,K,M}}$ est rendu hilbertien.

Remarquons que le terme constant h_0 n'a pas été pris en compte. En fait, pour prendre en compte h_0 , il est possible d'étendre le pseudo-vecteur \mathbf{x}_n en lui ajoutant le terme 1 et en incluant le terme h_0 au vecteur \mathbf{h} .

Une autre possibilité consiste à remarquer que h_0 ne joue que sur la moyenne de $y(n)$. Il peut alors être convenu que la sortie est centrée. Dans ce cas, $h_0 = -\mathbf{h}^T E[\mathbf{x}_n]$, et considérer le vecteur $\mathbf{x}_{c,n} = (\mathbf{x}_n - E[\mathbf{x}_n])$ permet de réécrire le filtre de Volterra sous la forme

$$y(n) = \mathbf{h}^T \mathbf{x}_{c,n}$$

Ceci est un commodité d'écriture et n'a pas d'influence sur l'utilisation des filtres de Volterra, notamment dans les procédures d'identification (voir le paragraphe 5.3.1).

Ainsi, dans toute la suite, la sortie d'un filtre de Volterra sera considérée comme centrée, sauf mention explicite.

Remarque sur la terminologie pseudo-vecteur Nous employons le terme pseudo-vecteur pour les $\mathbf{x}_{n,i}$ car se ne sont pas des vecteurs à strictement parler. En effet la somme de deux de ces tableaux de nombres ne peut pas s'écrire sous la même forme. Par exemple, $(x_1^2, x_1x_2, x_2^2) + (y_1^2, y_1y_2, y_2^2)$ ne peut pas de façon générale se mettre sous la forme (z_1^2, z_1z_2, z_2^2) .

Ceci explique aussi que l'espace vectoriel précédemment défini dépend explicitement des données. Le caractère vectoriel est contenu dans les filtres \mathbf{h} , ceci provenant de la linéarité en paramètres des filtres de Volterra.

Cas fréquentiel Les filtres fréquentiels de Volterra peuvent également se mettre sous forme vectorielle. Le principe est le même que dans le cas temporel : ranger les données linéaires et non linéaires dans un vecteur et construire en accord un vecteur filtre.

Ainsi, le filtre de Volterra en fréquence se met sous la forme

$$Y(m) = \mathbf{H}_m^T \mathbf{X}_m,$$

où

$$\begin{cases} \mathbf{X}_m^T &= (X(m) \quad \mathbf{X}_{2,m}^T \quad \dots \quad \mathbf{X}_{K,m}^T) \\ \mathbf{H}_m^T &= (H_{1,m} \quad \mathbf{H}_{2,m}^T \quad \dots \quad \mathbf{H}_{K,m}^T), \end{cases}$$

où $\mathbf{X}_{i,m}$ contient les données du type $X(m_1) \dots X(m_{i-1})X(m - m_1 - \dots - m_{i-1})$ et $\mathbf{H}_{i,m}$ les paramètres correspondants $H_i(m_1, \dots, m_{i-1}, m - m_1 - \dots - m_{i-1})$. Notons que contrairement au cas temporel, le vecteur filtre dépend explicitement de la fréquence et comporte m en indice.

Examinons l'effet des symétries dans le cas quadratique. Pour la fonction de transfert quadratique, les termes $H_2(m_1, m_2)$ participant à la fréquence m de la sortie sont ceux correspondant à $m = m_1 + m_2$ (trait gras sur la figure (5.1)). Compte tenu de la symétrie par rapport à la diagonale principale, il suffit de ne considérer que les termes de la partie hachurée de la figure (5.1). Toutefois, les termes extradiagonaux sont alors affectés d'un facteur 2 alors que les termes de la diagonale ne sont comptés qu'une fois [116]. Par exemple, pour les fréquences paires, les pseudo-vecteurs et vecteurs s'écrivent

$$\mathbf{X}_{2,m} = \begin{pmatrix} X(m/2)^2 \\ 2X(m/2 + 1)X(m/2 - 1) \\ \vdots \\ 2X(N)X(m - N) \end{pmatrix},$$

et

$$\mathbf{H}_{2,m} = \begin{pmatrix} H_2(m/2, m/2) \\ H_2(m/2 + 1, m/2 - 1) \\ \vdots \\ H_2(N, m - N) \end{pmatrix}.$$

Une subtilité intervient lorsque m est impair. En effet, le terme diagonal n'intervient pas dans ce cas et les pseudo-vecteurs et vecteurs deviennent

$$\mathbf{X}_{2,m} = \begin{pmatrix} 2X((m+1)/2)X((m-1)/2) \\ 2X((m+3)/2)X((m-3)/2) \\ \vdots \\ 2X(N)X(m-N) \end{pmatrix},$$

et

$$\mathbf{H}_{2,m} = \begin{pmatrix} H_2((m+1)/2, (m-1)/2) \\ H_2((m+3)/2, (m-3)/2) \\ \vdots \\ H_2(N, m-N) \end{pmatrix}.$$

Une dernière remarque doit être faite concernant l'échantillonnage. Dans les expressions précédentes, N est la fréquence maximale de la transformée de Fourier de $x(n)$. D'après la remarque faite au paragraphe 5.2.1, N correspond au quart de la fréquence d'échantillonnage. Donc, si m est dans $[0, N]$, les

vecteurs contiennent les parties linéaires et quadratiques. Mais rien n'interdit de chercher des liens linéaires dans la partie $[N + 1, 2N]$. Donc, sur cet "intervalle", le filtre contient uniquement la partie linéaire. En synthétisant, les pseudo-vecteurs et vecteurs peuvent alors s'écrire

$$\begin{cases} \mathbf{X}_m^T &= (X(m) \quad \mathbf{X}_{2,m}^T \chi_{[0,N]}(m)) \\ \mathbf{H}_m^T &= (H_{1,m} \quad \mathbf{H}_{2,m}^T \chi_{[0,N]}(m)), \end{cases}$$

où $\chi_{[0,N]}(m)$ est la fonction indicatrice de l'ensemble $\{0, 1, \dots, N\}$.

5.3 Identification en moyenne quadratique

Nous avons vu précédemment le fait remarquable que dans sa version discrète à ordre et mémoire finis le filtre de Volterra peut se représenter sous forme d'un produit scalaire entre un vecteur filtre et un pseudo-vecteur de données. L'objet de cette section est de tirer partie de cette représentation pour l'optimisation en moyenne quadratique des filtres de Volterra. Nous présentons donc la solution à ce problème pour les domaines temporel et fréquentiel avant de présenter deux autres idées importantes : l'implantation récursive des algorithmes et l'identification orthogonale des filtres de Volterra.

5.3.1 Identification dans le domaine temporel

Soient $x(n)$ et $y(n)$ deux signaux stationnaires appartenant respectivement aux classes $f^{(2K)}$ et $f^{(2)}$ (voir 2.9.2), reliés *a priori* non linéairement. On souhaite approcher le lien entre ces signaux par un développement en série de Volterra d'ordre et mémoire finis. Trouver le meilleur filtre de Volterra revient à trouver le meilleur vecteur \mathbf{h}_{opt} tel que $\mathbf{h}_{opt}^T \mathbf{x}_{c,n}$ approche au mieux le signal de sortie $y(n)$. De plus, l'identification en moyenne quadratique consiste à minimiser l'écart quadratique moyen entre la vraie sortie et son estimée. Résoudre le problème posé consiste donc à résoudre

$$\begin{aligned} \mathbf{h}_{opt} &= \text{Arg Min}_{\mathbf{h}} E[(y(n) - \mathbf{h}^T \mathbf{x}_{c,n})^2] \\ &= \text{Arg Min}_{\mathbf{h}} E[(y(n) - \hat{y}(n))^2], \end{aligned} \quad (5.11)$$

où $\hat{y}(n) = \mathbf{h}^T \mathbf{x}_{c,n}$. La solution à cette équation existe puisque la fonction à minimiser est concave en \mathbf{h} . Pour obtenir \mathbf{h}_{opt} , on peut par exemple évaluer à zéro la dérivée de l'erreur quadratique moyenne par rapport à \mathbf{h} . Toutefois, la présentation de ces filtres dans un contexte d'espace vectoriel permet de présenter la solution de (5.11) sous une forme très élégante en utilisant le théorème de la projection orthogonale : *l'erreur produite par la meilleure estimation est orthogonale à l'espace d'estimation*. Dans le cas qui nous intéresse,

l'espace d'estimation est l'espace $VF_{x,K,M}$ introduit précédemment. Trouver le meilleur élément de cet espace pour approcher un signal revient à projeter le signal sur cet espace. A l'aide du produit scalaire introduit précédemment, le théorème de la projection orthogonale se traduit par

$$\langle \mathbf{h}^T \mathbf{x}_{c,n} | (y(n) - \mathbf{h}_{opt}^T \mathbf{x}_{c,n}) \rangle = 0 \quad \forall \quad \mathbf{h}^T \mathbf{x}_{c,n} \in VF_{x,K,M}.$$

Cette dernière équation se résout aisément, la solution étant

$$E[\mathbf{x}_{c,n} \mathbf{x}_{c,n}^T] \mathbf{h}_{opt} = E[y(n) \mathbf{x}_{c,n}]. \quad (5.12)$$

$E[\mathbf{x}_{c,n} \mathbf{x}_{c,n}^T]$ qui est notée $\mathbf{C}_{\mathbf{xx}}$ est la matrice de corrélation du vecteur $\mathbf{x}_{c,n}$, et $E[y(n) \mathbf{x}_{c,n}]$ notée $\mathbf{R}_{y\mathbf{x}}$ est le vecteur intercorrélations entre la donnée $y(n)$ et le vecteur $\mathbf{x}_{c,n}$. La suppression de l'indice n se justifie par la stationnarité de $x(n)$. Ainsi, (5.12) se réécrit

$$\mathbf{C}_{\mathbf{xx}} \mathbf{h}_{opt} = \mathbf{R}_{y\mathbf{x}}. \quad (5.13)$$

Cette équation ressemble à s'y méprendre à l'équation donnant le filtre de Wiener linéaire. En fait la simplicité introduite par la représentation linéaire cache une complexité importante de (5.12). En effet, $\mathbf{x}_{c,n}$ contient les observations linéaires et non linéaires construites à partir de $x(n)$. Donc $\mathbf{C}_{\mathbf{xx}}$ contient les moments d'ordre 2 à $2K$ du signal $x(n)$, rangés d'une façon particulière. Cette matrice est aussi appelée matrice des moments. La même remarque est faite pour le vecteur $\mathbf{R}_{y\mathbf{x}}$ qui contient les intercorrélations d'ordre 2 à $2K$ entre $y(n)$ et les versions linéaires et non linéaires de $x(n)$. Nous supposons la matrice des moments inversible (l'étude de l'inversibilité de cette matrice est faite dans [33]).

Le filtre optimal s'écrit finalement

$$\mathbf{h}_{opt} = \mathbf{C}_{\mathbf{xx}}^{-1} \mathbf{R}_{y\mathbf{x}}.$$

L'erreur d'estimation pour le filtre optimal peut donc être calculée selon

$$Eq_{min} = E[(y(n) - \mathbf{h}_{opt}^T \mathbf{x}_{c,n})^2],$$

qui devient immédiatement en vertu du théorème de la projection orthogonale

$$Eq_{min} = E[y(n)(y(n) - \mathbf{h}_{opt}^T \mathbf{x}_{c,n})] = E[y(n)^2] - \mathbf{R}_{y\mathbf{x}}^T \mathbf{C}_{\mathbf{xx}}^{-1} \mathbf{R}_{y\mathbf{x}}. \quad (5.14)$$

Il est montré dans [51, 67] que l'augmentation de l'ordre du filtre diminue l'erreur quadratique minimale. Nous remontrons ce résultat au paragraphe 5.3.3 d'une façon très simple.

Revenons sur le problème du terme constant h_0 . Si $y(n)$ n'est pas centré, alors nous travaillons avec les vecteurs $(1 \ \mathbf{x}_n^T)^T$ et $(h_0 \ \mathbf{h}^T)^T$. L'optimisation s'effectue de la même manière et conduit au système

$$\begin{pmatrix} 1 & E[\mathbf{x}_n^T] \\ E[\mathbf{x}_n] & E[\mathbf{x}_n \mathbf{x}_n^T] \end{pmatrix} \begin{pmatrix} h_0 \\ \mathbf{h} \end{pmatrix} = \begin{pmatrix} E[y(n)] \\ E[y(n) \mathbf{x}_n] \end{pmatrix},$$

qui conduit à résoudre $(E[\mathbf{x}_n \mathbf{x}_n^T] - E[\mathbf{x}_n]E[\mathbf{x}_n^T])\mathbf{h} = E[y(n)\mathbf{x}_n] - E[y(n)]E[\mathbf{x}_n]$ (qui n'est autre que (5.13)), c'est-à-dire travailler avec la matrice covariance de \mathbf{x}_n et le vecteur covariance entre $y(n)$ et \mathbf{x}_n . Autrement dit, ceci revient à centrer la sortie du filtre et le pseudo-vecteur de données, et confirme la remarque faite au paragraphe 5.2.2.

Dans le paragraphe suivant, l'identification est opérée dans le domaine fréquentiel.

5.3.2 Identification dans le domaine fréquentiel

Dans le domaine de Fourier, une relation de filtrage de Volterra s'écrit encore une fois sous forme linéaire selon

$$Y(m) = \mathbf{H}_m^T \mathbf{X}_m, \quad (5.15)$$

où les vecteurs \mathbf{H}_m et \mathbf{X}_m sont définis au paragraphe 5.2.2. Nous souhaitons maintenant trouver le meilleur vecteur \mathbf{H}_m de sorte à minimiser l'écart quadratique moyen entre $Y(m)$ et son estimée, c'est-à-dire résoudre

$$\begin{aligned} \mathbf{H}_{m,opt} &= \text{Arg Min}_{\mathbf{H}_m} E[|Y(m) - \mathbf{H}_m^T \mathbf{X}_m|^2] \\ &= \text{Arg Min}_{\mathbf{H}_m} E[|Y(m) - \hat{Y}(m)|^2], \end{aligned} \quad (5.16)$$

où $\hat{Y}(m) = \mathbf{H}_m^T \mathbf{X}_m$. La résolution de cette équation peut s'effectuer de la même façon que dans le domaine temporel. Encore une fois, en supposant l'inversibilité de la matrice covariance de \mathbf{X}_m , la solution de (5.16) est

$$\mathbf{H}_{m,opt} = \mathbf{C}_{\mathbf{X}_m \mathbf{X}_m}^{-1} \mathbf{R}_{Y(m)\mathbf{X}_m}. \quad (5.17)$$

Cette équation est encore simple mais voile sa réelle complexité. Il faut se rappeler que \mathbf{X}_m contient également des composantes fréquentielles autres que m , et par suite, $\mathbf{C}_{\mathbf{X}_m \mathbf{X}_m}$ contient des multispectres de $x(n)$ construits sur les moments. L'erreur d'estimation obtenue est alors

$$\begin{aligned} Eq_{min}(m) &= E[|Y(m) - \mathbf{H}_{m,opt}^T \mathbf{X}_m|^2] \\ &= E[|Y(m)|^2] - \mathbf{R}_{Y(m)\mathbf{X}_m}^\dagger \mathbf{C}_{\mathbf{X}_m \mathbf{X}_m}^{-1} \mathbf{R}_{Y(m)\mathbf{X}_m}. \end{aligned}$$

Dans le paragraphe suivant, nous envisageons une variante de cette procédure, l'identification orthogonale.

5.3.3 Identification orthogonale dans le domaine temporel

La simplicité de l'implantation vectorielle des filtres de Volterra, comme déjà mentionné, cache sa réelle complexité. De même, les procédures d'identification utilisées apparaissent simples, mais cachent certaines subtilités. A titre d'exemple, détaillons le filtre optimal linéaire-quadratique dans le domaine temporel. Il est obtenu par résolution de l'équation (5.13). En détaillant les parties linéaire et quadratique des vecteurs (les lettres l et q apparaissant en indice pour respectivement les vecteurs linéaire et quadratique), cette équation devient

$$\begin{pmatrix} \mathbf{C}_{\mathbf{x}_l \mathbf{x}_l} & \mathbf{C}_{\mathbf{x}_l \mathbf{x}_q} \\ \mathbf{C}_{\mathbf{x}_q \mathbf{x}_l} & \mathbf{C}_{\mathbf{x}_q \mathbf{x}_q} \end{pmatrix} \begin{pmatrix} \mathbf{h}_l \\ \mathbf{h}_q \end{pmatrix} = \begin{pmatrix} \mathbf{R}_{y \mathbf{x}_l} \\ \mathbf{R}_{y \mathbf{x}_q} \end{pmatrix}.$$

Ce détail confirme les remarques du paragraphe 5.1.4 : les parties linéaire et quadratique du filtre optimal sont couplées via les statistiques d'ordre 3 du signal d'entrée, contenues dans la matrice $\mathbf{C}_{\mathbf{x}_l \mathbf{x}_q}$ [67, 69]. Ceci rend difficile l'interprétation physique du filtre : comment quantifier l'importance du linéaire par rapport au quadratique? Ce genre de problème de couplage apparaît aussi par exemple dans les systèmes MISO (Multi-Inputs Single Output) où les couplages (ici des cohérences) entre entrées ne permettent pas de bien chiffrer l'influence de chaque entrée sur la sortie. Le problème dans le cas linéaire-quadratique vient du fait que les observations linéaires et les observations quadratiques sont évidemment corrélées. Ainsi, pour obtenir une bonne interprétation physique, il faut provoquer une décorrélation entre ces termes. La procédure obtenue sera appelée identification orthogonale. Le pionnier en la matière a été Norbert Wiener, en 1955. En effet, dans [200] (voir aussi [182]), il réalise une orthogonalisation des fonctionnelles de Volterra dont l'entrée est gaussienne, et introduit les G-fonctions, système orthogonal de polynômes à mémoire pour une entrée gaussienne. Plus récemment, cette méthode a été reprise par M. Korenberg, mais pour des systèmes à ordre et mémoire finis. De plus la mesure de l'orthogonalité ne s'effectue plus sur un horizon infini mais sur un horizon fini. Le principe de ces méthodes est le suivant.

Réécrivons la sortie d'un filtre de Volterra de la façon suivante [117]

$$y(n) = h_0 + \mathbf{h}^T \mathbf{x}_n = \sum_{i=0}^{P-1} a_i p_i(n),$$

où P est la taille des vecteurs \mathbf{h} et \mathbf{x}_n , et où a_i et $p_i(n)$ sont respectivement les i èmes composantes des vecteurs \mathbf{h} et \mathbf{x}_n . Par exemple, $p_0(n) = 1$, $p_i(n) = x(n-i)$ pour $i = 0 \dots M$, $p_{M+1}(n) = x^2(n)$, $p_{M+2}(n) = x(n)x(n-1)$, \dots . Nous souhaitons maintenant réécrire ce développement sur un système de polynômes

orthogonaux, soit

$$y(n) = \sum_{i=0}^{P-1} \alpha_i \pi_i(n),$$

où $\langle \pi_i(n) | \pi_j(n) \rangle = cte. \delta_{ij}$, le produit scalaire étant à définir. Supposons que le système $\{\pi_i(n)\}_{i=0 \dots P-1}$ soit connu. La procédure d'identification au sens de l'erreur quadratique moyenne minimale se réduit à

$$\alpha_{k,opt} = \frac{\langle y(n) | \pi_k(n) \rangle}{\langle \pi_k(n) | \pi_k(n) \rangle} \quad \forall k = 0 \dots P - 1$$

Dans ce système orthogonal, α_k représente la composante de $y(n)$ sur l'axe $\pi_k(n)$ (ce résultat est encore une application du théorème de la projection orthogonale). Le choix du produit scalaire dépend du contexte : dans le cas théorique de processus stochastiques, $\langle x|y \rangle = E[xy]$; dans le cas réel d'applications où une réalisation sur T points est disponible, $\langle x|y \rangle = \frac{1}{T} \sum_{i=0}^{T-1} x(i)y(i)$.

L'obtention des polynômes orthogonaux peut s'effectuer tout simplement avec la procédure de Gram-Schmidt, en utilisant le produit scalaire adéquat. Lorsque l'identification est effectuée, le calcul de l'erreur quadratique d'estimation de $y(n)$ est

$$Eq_{min} = \langle y(n) | y(n) \rangle - \sum_{i=0}^{P-1} \alpha_{i,opt}^2 \langle \pi_i(n) | \pi_i(n) \rangle .$$

Cette expression, plus simple que celle obtenue dans le cadre classique ((5.14) du paragraphe 5.3.1) apporte une démonstration claire du fait que l'ajout d'un ordre de non-linéarité provoque une baisse de l'erreur quadratique moyenne d'estimation. De plus, cette approche permet de chiffrer précisément la diminution en erreur quadratique moyenne. Cette remarque peut alors permettre la définition d'algorithmes de sélection de termes prépondérants. En effet, si l'ajout d'un terme provoque une baisse non négligeable de l'erreur quadratique, il est conservé, alors qu'une faible baisse entraîne son éviction.

5.3.4 Implantation récursive

L'implantation récursive des algorithmes d'identification est rendue nécessaire par deux principales raisons :

- traitement en temps réel de données,
- éventuelles non-stationnarités lentes du système à identifier. Dans ce cas, les algorithmes seront également dits adaptatifs.

Les formes récursives des algorithmes précédents s'obtiennent en réalisant une minimisation récursive des critères quadratiques. Deux approches peuvent être envisagées : méthode type gradient et méthode type Newton-Raphson. Nous allons encore ici distinguer les domaines temporel et fréquentiel.

Identification récursive dans le domaine temporel Le critère quadratique à minimiser s'écrit

$$Eq(\mathbf{h}) = E[(y(n) - \mathbf{h}^T \mathbf{x}_{c,n})^2].$$

- Méthode du gradient (LMS)

La solution de la minimisation par la méthode du gradient consiste à s'approcher de la solution par sauts successifs proportionnels à la variation locale de la fonction que l'on minimise. Plus précisément, l'estimée à l'instant $k + 1$ se déduit de l'estimée à l'instant k selon

$$\mathbf{h}_{k+1} = \mathbf{h}_k - \frac{1}{2} \mu_k \nabla_{\mathbf{h}} Eq(\mathbf{h})|_{\mathbf{h}_k},$$

qui devient en effectuant explicitement la différentiation

$$\mathbf{h}_{k+1} = \mathbf{h}_k + \mu_k E[\mathbf{x}_{c,k+1}(y(k+1) - \mathbf{h}_k^T \mathbf{x}_{c,k+1})].$$

Cet algorithme est appelé gradient déterministe car il ne fait intervenir qu'une fonction déterministe des données. Bien évidemment il est inutilisable en pratique (puisque les statistiques requises pour son fonctionnement sont inaccessibles). Pour les applications, il faut donc introduire sa version dite stochastique, où l'espérance mathématique est remplacée par un estimateur. L'estimateur choisi est le plus simple : l'espérance mathématique est remplacée par sa valeur "instantanée" pour conduire au gradient stochastique

$$\mathbf{h}_{k+1} = \mathbf{h}_k + \mu_k \mathbf{x}_{c,k+1}(y(k+1) - \mathbf{h}_k^T \mathbf{x}_{c,k+1}). \quad (5.18)$$

Remarquons que dans les algorithmes précédents, le paramètre μ_k est choisi dépendant du temps pour le cas stationnaire. Dans un cas où le filtre à identifier possède des non-stationnarités lentes, μ_k sera choisi constant pour pouvoir suivre les variations du vecteur paramètre [26]. L'algorithme est alors adaptatif.

L'analyse de la convergence de l'algorithme LMS est délicate, même dans le cas d'un signal de régression blanc. Cette analyse peut être effectuée en utilisant les résultats de [26] ou la théorie de la M-indépendance [136] pour montrer la convergence en moyenne de cet algorithme. La M-indépendance est utilisée par P. Duvaut dans [68] dans le cas linéaire-quadratique.

- Algorithme des moindres carrés récurrents (RLS)

Nous avons vu que le filtre optimal s'écrit sous la forme

$$\mathbf{h}_{opt} = \mathbf{C}_{\mathbf{xx}}^{-1} \mathbf{R}_{y\mathbf{x}}. \quad (5.19)$$

Nous pouvons utiliser cette expression pour le calcul d'une estimation de \mathbf{h}_{opt} , en remplaçant $\mathbf{C}_{\mathbf{xx}}^{-1}$ et $\mathbf{R}_{y\mathbf{x}}$ par des estimateurs $\widehat{\mathbf{C}}_{\mathbf{xx}k}^{-1}$ et $\widehat{\mathbf{R}}_{y\mathbf{x}k}$.

Or, $\mathbf{R}_{y\mathbf{x}} = E[y(n)\mathbf{x}_{c,n}]$ peut être approché par l'estimateur naturel

$$\widehat{\mathbf{R}}_{y\mathbf{x}k} = \frac{1}{k} \sum_{i=1}^k y(i)\mathbf{x}_{c,i},$$

qui s'écrit de façon récursive selon

$$\widehat{\mathbf{R}}_{y\mathbf{x}k} = \frac{k-1}{k} \widehat{\mathbf{R}}_{y\mathbf{x}k-1} + \frac{1}{k} y(k)\mathbf{x}_{c,k}.$$

De même, un estimateur naturel de la matrice de corrélation du vecteur $\mathbf{x}_{c,n}$ est

$$\widehat{\mathbf{C}}_{\mathbf{xx}k} = \frac{1}{k} \sum_{i=1}^k \mathbf{x}_{c,i}\mathbf{x}_{c,i}^T,$$

$$\widehat{\mathbf{C}}_{\mathbf{xx}k} = \frac{k-1}{k} \widehat{\mathbf{C}}_{\mathbf{xx}k-1} + \frac{1}{k} \mathbf{x}_{c,k}\mathbf{x}_{c,k}^T.$$

Nous souhaitons avoir un estimateur de l'inverse de la matrice de covariance^{10,11}. Remarquons que nous allons effectuer le produit $\widehat{\mathbf{C}}_{\mathbf{xx}k}^{-1} \widehat{\mathbf{R}}_{y\mathbf{x}k}$. Par suite, les termes en $1/k$ des estimateurs précédents vont se simplifier, et nous travaillerons donc sur

$$\begin{cases} \widehat{\mathbf{C}}_{\mathbf{xx}k} &= \widehat{\mathbf{C}}_{\mathbf{xx}k-1} + \mathbf{x}_{c,k}\mathbf{x}_{c,k}^T \\ \widehat{\mathbf{R}}_{y\mathbf{x}k} &= \widehat{\mathbf{R}}_{y\mathbf{x}k-1} + y(k)\mathbf{x}_{c,k}. \end{cases}$$

Soit \mathbf{K}_k la matrice inverse de $\widehat{\mathbf{C}}_{\mathbf{xx}k}$. A partir de cette forme récursive, et en utilisant le lemme d'inversion matricielle de Woodbury $(A + BC)^{-1} = A^{-1} - A^{-1}B(1 + CA^{-1}B)^{-1}CA^{-1}$, il vient

$$\mathbf{K}_k = \mathbf{K}_{k-1} - \frac{\mathbf{K}_{k-1}\mathbf{x}_{c,k}\mathbf{x}_{c,k}^T\mathbf{K}_{k-1}}{1 + \mathbf{x}_{c,k}^T\mathbf{K}_{k-1}\mathbf{x}_{c,k}}. \quad (5.20)$$

¹⁰On n'utilise pas une estimation de l'inverse de la matrice de covariance mais plutôt l'inverse de l'estimée de cette matrice. Ceci n'est bien sûr pas équivalent mais l'algorithme ainsi défini converge vers le filtre optimal.

¹¹Dans un contexte où le récursif n'est pas nécessaire, l'inversion de la matrice de covariance peut être effectuée à la fin du calcul.

Evaluons maintenant $\mathbf{h}_{opt,k} = \mathbf{K}_k \mathbf{R}_{y\mathbf{x}}$. En utilisant (5.20) on obtient après quelques manipulations

$$\mathbf{h}_{opt,k} = \mathbf{h}_{opt,k-1} + \mathbf{K}_k \mathbf{x}_{c,k} (y(k) - \mathbf{h}_{opt,k-1}^T \mathbf{x}_{c,k}). \quad (5.21)$$

Cette dernière équation couplée à l'équation (5.20) constitue l'algorithme RLS. La matrice \mathbf{K}_k est appelée gain de Kalman. Elle pondère à chaque instant la correction à effectuer sur le filtre pour s'approcher de la solution optimale. Imposer à cette matrice d'être proportionnelle à la matrice identité permet de retrouver l'algorithme LMS (5.18). L'initialisation de \mathbf{K}_k se fait traditionnellement par $\mathbf{K}_0 = cte \cdot \mathbf{I}$ où $cte \gg 1$ et \mathbf{I} représente la matrice identité.

5.3.5 Identification récursive dans le domaine fréquentiel

Rappelons qu'en fréquence, le critère quadratique à minimiser est $E[|Y(m) - \mathbf{H}_m^T \mathbf{X}_m|^2]$. L'estimation récursive est ici plus subtile. En effet, dans le domaine temporel, nous disposons directement des échantillons participant au vecteur de régression puisque ce sont les échantillons même du signal. Dans le cas fréquentiel, le filtre est calculé fréquence par fréquence, et pour réussir l'adaptation du filtre nous devons disposer de nombreuses réalisations *pour chaque fréquence*. Les algorithmes pour l'identification dans le domaine fréquentiel devront donc commencer par un tronçonnage du signal temporel en blocs, dont les transformées de Fourier fourniront les réalisations requises. Dans ce qui suit, l'indice k qui apparaîtra sera alors relatif à la réalisation numéro k (issue du k ème bloc). L'adaptation ne se fait plus ici directement au cours du temps, mais "au cours des blocs"! Nous supposons donc posséder un continuum de blocs pour simplifier la présentation. L'indice k apparaîtra alors explicitement.

- Algorithme du gradient

De même que pour l'identification dans le domaine temporel, nous obtenons un algorithme du gradient selon

$$\mathbf{H}_{m,k+1} = \mathbf{H}_{m,k} - \frac{1}{2} \mu(m) \nabla_{\mathbf{H}_m} (E[|Y(m) - \mathbf{H}_m^T \mathbf{X}_m|^2])|_{\mathbf{H}_{m,k}}.$$

Le pas d'adaptation dépend explicitement de la fréquence puisque l'identification se fait canal par canal. Effectuer la différentiation conduit à¹²

$$\mathbf{H}_{m,k+1} = \mathbf{H}_{m,k} + \mu(m) E[(Y(m)_{k+1} - \mathbf{X}_{m,k+1}^T \mathbf{H}_{m,k}) \mathbf{X}_{m,k+1}^*].$$

La version stochastique de cet algorithme est alors la suivante

$$\mathbf{H}_{m,k+1} = \mathbf{H}_{m,k} + \mu(m) (Y(m)_{k+1} - \mathbf{X}_{m,k+1}^T \mathbf{H}_{m,k}) \mathbf{X}_{m,k+1}^*.$$

¹²On utilise la dérivation formelle $\frac{dz z^*}{dz^*} = 2z$.

- Algorithme RLS

L'approche adoptée pour le domaine temporel peut être reprise ici [155]. Nous partons de la forme du filtre optimal, et, à l'aide de la formule d'inversion de Woodbury, nous obtenons une estimée récursive de l'inverse de la matrice de covariance du vecteur de régression. L'algorithme RLS est alors

$$\begin{cases} \mathbf{C}_{\mathbf{X}_m \mathbf{X}_m, k+1}^{-1} &= \mathbf{C}_{\mathbf{X}_m \mathbf{X}_m, k}^{-1} - \frac{\mathbf{C}_{\mathbf{X}_m \mathbf{X}_m, k}^{-1} \mathbf{X}_{m, k+1}^* \mathbf{X}_{m, k+1}^T \mathbf{C}_{\mathbf{X}_m \mathbf{X}_m, k}^{-1}}{1 + \mathbf{X}_{m, k+1}^T \mathbf{C}_{\mathbf{X}_m \mathbf{X}_m, k}^{-1} \mathbf{X}_{m, k+1}^*} \\ \mathbf{R}_{Y(m) \mathbf{X}_m, k+1} &= \mathbf{R}_{Y(m) \mathbf{X}_m, k} + Y(m)_{k+1} \mathbf{X}_{m, k+1}^* \\ \mathbf{H}_{m, k+1} &= \mathbf{C}_{\mathbf{X}_m \mathbf{X}_m, k+1}^{-1} \mathbf{R}_{Y(m) \mathbf{X}_m, k+1}. \end{cases}$$

5.3.6 Complexité numérique des algorithmes RLS

Nous effectuons ici une rapide analyse de la complexité numérique des algorithmes RLS dans les domaines temporel et fréquentiel. La complexité est ici étudiée pour un filtre quadratique pur.

- Domaine temporel.

Il est connu que la résolution d'un système linéaire d'ordre P par un algorithme RLS requiert P^2 opérations par échantillon. Si le signal comporte N points, ce nombre s'élève à NP^2 .

- Domaine fréquentiel.

Supposons que le signal comporte N points. Il est découpé en tranches de M points, M étant une puissance de 2. La transformée de Fourier rapide sur un bloc requiert $M \log_2 M$ opérations. L'algorithme RLS effectue son adaptation au cours des blocs. De plus, la taille du vecteur à adapter dépend de la fréquence (voir la figure (5.1)). La taille de ces vecteurs est majorée par M . Par fréquence et par bloc, l'algorithme requiert alors un nombre d'opérations inférieur ou égal à M^2 . Or, il y a M fréquences à traiter et $\frac{N}{M}$ blocs. Les opérations requises après les transformations de Fourier sont donc au nombre de NM^2 . Finalement, l'algorithme RLS fréquentiel requiert environ $N(M^2 + \log_2 M)$ opérations.

- Comparaison.

Si la mémoire du filtre temporel est choisie égale à M , alors $P = \frac{M(M-1)}{2}$ et nous obtenons

algorithme temporel : NM^4 opérations.

algorithme fréquentiel : $N(M^2 + \log_2 M)$ opérations.

Le temps de calcul est donc beaucoup plus court dans le cas fréquentiel. Par contre, la dualité temps-fréquence se retrouve au niveau numérique dans une dualité temps de calcul-place mémoire! En effet, la place requise pour l'algorithme temporel est de l'ordre de M^4 cases mémoire, alors qu'il est de M^5 cases pour l'algorithme fréquentiel.

Notons que des versions rapides de l'algorithme RLS temporel ont été proposées par V.J. Matthews et ses collaborateurs, ramenant la complexité à NM^3 opérations [128, 192].

5.4 Application : soustraction de bruit

La méthode de Soustraction de Bruit (SdB) (Noise Cancellation en Anglais) apparaît pour la première fois en 1975 dans un célèbre article de B. Widrow et de ses collègues [197]. L'application de cette méthode à l'écoute du cœur d'un fœtus est remarquable. En effet, les battements du cœur d'un fœtus sont faibles, et sont en grande partie cachés par ceux du cœur de la mère. Widrow propose alors de placer un capteur de référence au niveau du cœur de la mère, et montre comment estimer l'influence de cette référence sur un capteur d'écoute placé près du cœur du fœtus. Une fois l'influence quantifiée, il suffit de la retrancher du capteur d'écoute pour obtenir le rythme cardiaque propre du fœtus. La soustraction de bruit est donc une méthode d'élimination de bruit d'un signal lorsque l'on a accès à une référence liée au bruit pollueur. Elle est fondée sur des hypothèses de type indépendance et son principe est d'identifier le lien entre la référence (souvent appelée référence bruit seul) et le bruit. L'identification étant faite, une estimation du bruit est obtenue et est retranchée du signal bruité. Cette démarche est résumée par la figure (5.2).

Les travaux de pionnier de B. Widrow et de ses collaborateurs ont conduit de nombreuses équipes à travailler dans le domaine. Toutefois, les études ont été principalement menées dans le cas où le lien entre le bruit et sa référence est supposé linéaire. Plus précisément, nous pouvons dire qu'aucune hypothèse n'était faite sur la nature du lien, mais que ce lien était identifié dans la classe très restreinte des systèmes linéaires. De plus, de nombreux travaux ont concerné la soustraction de bruit en linéaire pour des signaux non stationnaires. Une synthèse détaillée des divers algorithmes utilisables et une bibliographie étendue peuvent être trouvées dans [22]. Durant ces deux dernières décennies, peu d'intérêt a été porté au cas où l'identification du lien s'effectue par des filtres non linéaires. L'objet de ce paragraphe est l'application des filtres de Volterra à la soustraction de bruit. Dans un premier temps, nous donnons la solution au problème, dans les domaines temporel et fréquentiel, puis nous présentons un cas réel issu d'expériences de sonar.

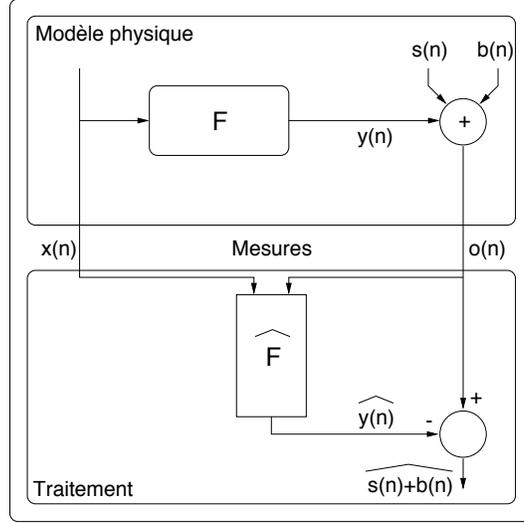


Figure 5.2: *Principe de la soustraction de bruit. La partie modèle physique représente la propagation de la référence bruit vers le capteur du signal utile. Les signaux $o(n)$ et $x(n)$ sont mesurés. La partie traitement réalise l'identification du milieu de propagation F et la soustraction du signal $y(n)$ estimé.*

5.4.1 Filtres de Volterra et Soustraction de Bruit

Considérons à nouveau le schéma de la figure (5.2) en supposant que le filtre liant la référence au bruit est non linéaire. La modélisation du système physique est la suivante

$$\begin{cases} o(n) &= s(n) + y(n) + b(n) \\ y(n) &= F(x(n)). \end{cases} \quad (5.22)$$

où $x(n)$ et $o(n)$ sont observés et $b(n)$ est le bruit de fond. De plus, $b(n)$, $x(n)$ et $s(n)$ sont supposés indépendants et centrés. Nous allons identifier le lien F par un filtre de Volterra discret d'ordre K et mémoire M finis. Le problème est de trouver le meilleur élément de $VF_{x,K,M}$ au sens de l'erreur quadratique moyenne minimale Eq . Ainsi, nous cherchons $\mathbf{h}_{opt} \mathbf{x}_{c,n}$ de $VF_{x,K,M}$ tel que

$$\mathbf{h}_{opt} = \text{Arg Min}_{\mathbf{h}} E[(y(n) - \mathbf{h}^T \mathbf{x}_{c,n})^2].$$

Mais $y(n)$ est inaccessible!

Toutefois, l'hypothèse d'indépendance entre $x(n)$ et $s(n) + b(n)$ et leur caractère centré permettent de conclure que

$$Eq' = E[(o(n) - \mathbf{h}^T \mathbf{x}_{c,n})^2] = Eq + E[(s(n) + b(n))^2].$$

Donc, minimiser Eq est équivalent à minimiser Eq' . Le filtre optimal est donc obtenu en utilisant les méthodes présentées au paragraphe 5.3. Ceci est bien sûr valable dans les domaines temporel et fréquentiel :

- domaine temporel :

$$\mathbf{h}_{opt} = \mathbf{C}_{\mathbf{xx}}^{-1} \mathbf{R}_{y\mathbf{x}}.$$

L'erreur d'estimation est alors

$$\begin{aligned} Eq'_{min} &= E[(s(n) + b(n))^2] + E[y(n)^2] \\ &- \mathbf{R}_{o\mathbf{x}}^T \mathbf{C}_{\mathbf{xx}}^{-1} \mathbf{R}_{o\mathbf{x}}. \end{aligned}$$

- domaine fréquentiel :

$$\mathbf{H}_{m,opt} = \mathbf{C}_{\mathbf{X}_m \mathbf{X}_m}^{-1} \mathbf{R}_{Y(m)\mathbf{X}_m}.$$

L'erreur d'estimation est alors

$$\begin{aligned} Eq'_{min}(m) &= E[|S(m) + B(m)|^2] + E[|Y(m)|^2] \\ &- \mathbf{R}_{O(m)\mathbf{X}_m}^\dagger \mathbf{C}_{\mathbf{X}_m \mathbf{X}_m}^{-1} \mathbf{R}_{O(m)\mathbf{X}_m}. \end{aligned}$$

Avant d'illustrer cette méthode sur un cas réel (paragraphe 5.4.2), faisons une remarque sur l'hypothèse d'indépendance entre le signal utile et la référence bruit-seul.

Non indépendance entre signal utile et la référence bruit-seul

Comme présenté dans les paragraphes précédents, la méthode de soustraction de bruit ne fonctionne que si des hypothèses d'indépendance entre le signal utile et la référence bruit-seul sont vérifiées. Comme l'a remarqué H. Mermoz [151], si ces hypothèses ne sont pas vérifiées, effectuer le traitement peut conduire à une dégradation importante des performances, et il est préférable en général de ne rien faire!

L'invalidité de l'hypothèse d'indépendance entre signal utile et référence bruit-seul peut venir du fait que le signal utile se mélange également à la référence bruit-seul. La modélisation (5.22) doit alors être modifiée et peut se réécrire selon

$$\begin{cases} o_1(n) &= s(n) + F(x(n)) + b_1(n) \\ o_2(n) &= G(s(n)) + x(n) + b_2(n). \end{cases}$$

Dans cette modélisation, on suppose l'indépendance entre les signaux $s(n)$, $x(n)$, $b_1(n)$ et $b_2(n)$. Ce modèle représente le mélange de deux signaux indépendants, et reconstituer $s(n)$ et $x(n)$ à partir des observations $o_1(n)$ et $o_2(n)$ relève de la séparation de sources. Lorsque les transferts $F(\cdot)$ et $G(\cdot)$ sont linéaires, la séparation de $s(n)$ et $x(n)$ est possible en utilisant les techniques développées au chapitre 7. Cette remarque est illustrée par quelques exemples au paragraphe 7.3.5. Si les transferts sont non linéaires, la séparation doit attendre les progrès des nombreux chercheurs travaillant encore sur ce domaine! Notons que les travaux de M. Krob et M. Benidir [119] concernant l'identification aveugle de filtres de Volterra multidimensionnels offrent une piste intéressante pour cette recherche.

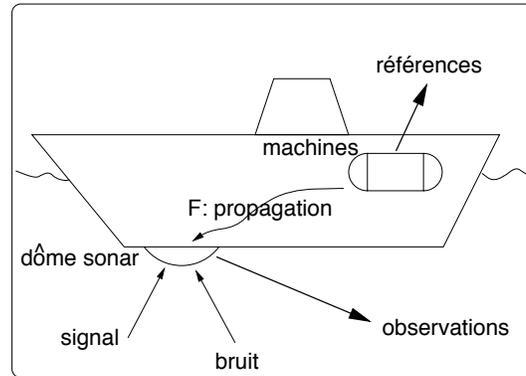


Figure 5.3: *Expérience réelle étudiée. Le sonar est fixé sous la coque du bateau et reçoit le signal utile, le bruit ambiant sous-marin et les bruits propres des machines. La sortie du sonar et les références bruit sont enregistrées.*

5.4.2 Etude d'un cas réel

Nous proposons maintenant une application réelle de la soustraction de bruit utilisant les filtres de Volterra. Notre but est uniquement de montrer que dans ce cas la modélisation non linéaire apporte un gain non négligeable dans les performances.

La situation réelle à laquelle nous nous intéressons ici est issue de mesures sonar réalisées sur un bateau. Le sonar se trouve sous la coque du bateau dans le dôme sonar. Il écoute les bruits sous-marins, espérant y détecter des signaux issus d'autres bateaux. Lors de la présence d'un tel signal, le bruit sous-marin ambiant est inévitablement présent. Mais la mesure est également entachée d'une autre source de bruit : le bruit propre du bateau, provoqué par ses moteurs et autres auxiliaires. Ces bruits étant provoqués par les machines du bateau, il est bien sûr possible de disposer près de ces sources de bruit des capteurs qui donnent des signaux représentatifs de ces bruits : ce sont les références bruit seul. La description faite est résumée par la figure (5.3).

Rappelons la modélisation mathématique du problème :

$$\begin{cases} o(n) &= s(n) + y(n) + b(n) \\ y(n) &= F(x(n)). \end{cases}$$

Effectuons maintenant la description des signaux observés $o(n)$ et $x(n)$. Leurs spectres respectifs apparaissent sur les figures (5.4) et (5.5). La figure (5.6) représente la cohérence entre ces deux signaux. Ces figures appellent quelques commentaires :

- Le signal $o(n)$ issu du sonar semble composé de trois parties distinctes. La première est le plateau très large bande représentant certainement le bruit de fond. La seconde apparaît sous forme d'une bosse dans la partie

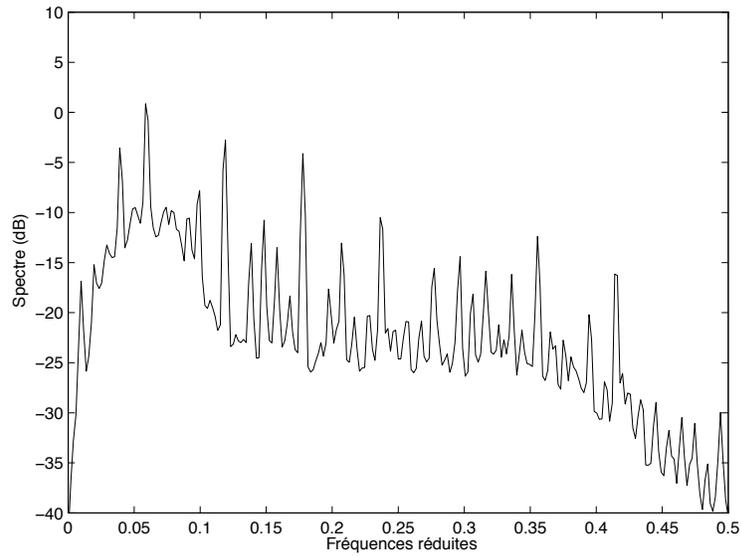


Figure 5.4: *Spectre en dB de la sortie du sonar. 256 canaux de fréquence sont représentés (195 moyennes).*

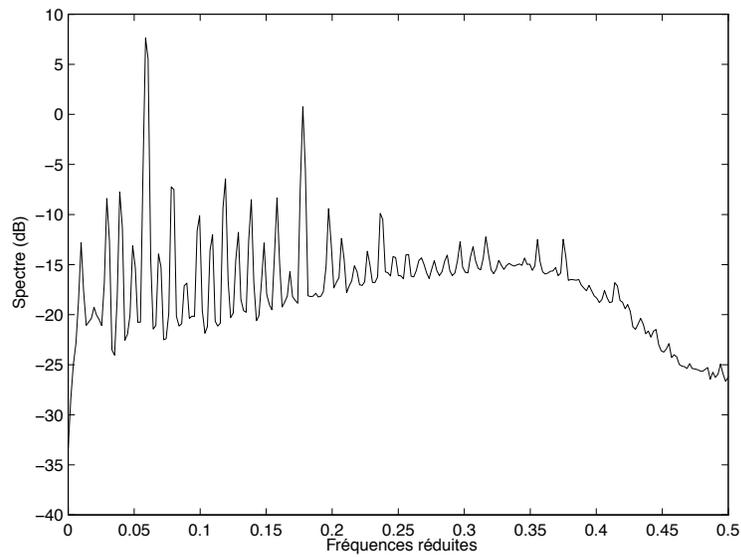


Figure 5.5: *Spectre en dB de la référence bruit seul. 256 canaux de fréquence sont représentés (195 moyennes).*

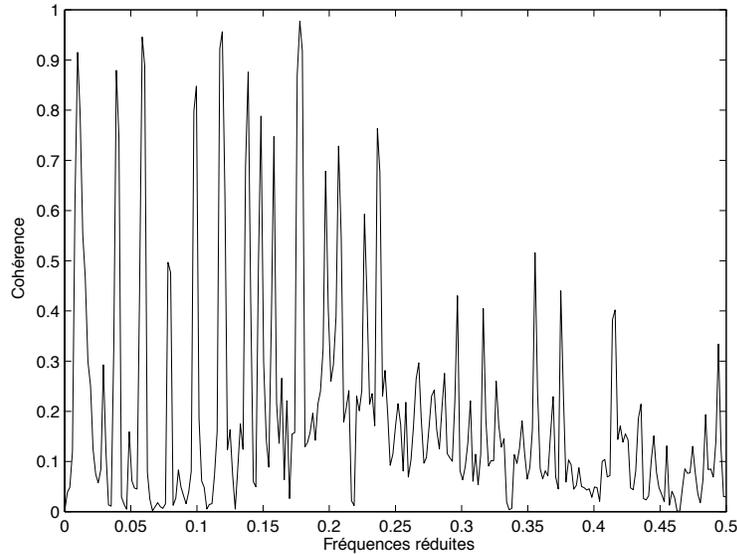


Figure 5.6: *Cohérence entre la référence et la sortie sonar. 256 canaux de fréquence sont représentés (195 moyennes).*

$[0, 0.25]$ du spectre qui doit être due au signal utile. Enfin, se superpose un spectre de raies provenant probablement du bruit provoqué par les machines du bateau. Notons que les raies composant ce spectre discret sont puissantes sur toute la gamme de fréquence.

- La référence $x(n)$ présente une structure essentiellement à spectre de raies. Ceci n'est pas étonnant puisque ce signal est représentatif d'une machine tournante. Remarquons que les raies sont très puissantes dans la première moitié du spectre, et beaucoup plus faibles, voire inexistantes, dans la seconde moitié.
- La cohérence entre $o(n)$ et $x(n)$ montre bien que le plateau et la bosse sont "extérieurs" au bateau, et que les raies présentes en sortie du sonar proviennent linéairement de la référence bruit. Tempérons toutefois cette affirmation en remarquant que les raies dans la partie $[0, 0.25]$ du spectre sont très cohérentes, alors qu'elles le sont bien moins dans la seconde moitié du spectre $[0.25, 0.5]$. Or, dans cette partie, les raies sont présentes sur la sortie mais absentes (très faibles) sur la référence. D'où viennent ces raies, puisqu'elles ne parviennent linéairement que très faiblement (cohérence basse)?

Nous pouvons répondre à cette question en supposant que les bruits provoqués par les machines se transfèrent non linéairement entre les machines et le dôme sonar (il y a bien entendu également une partie linéaire dans ce transfert). Cette supposition se fonde sur une sorte de doublement de fréquence

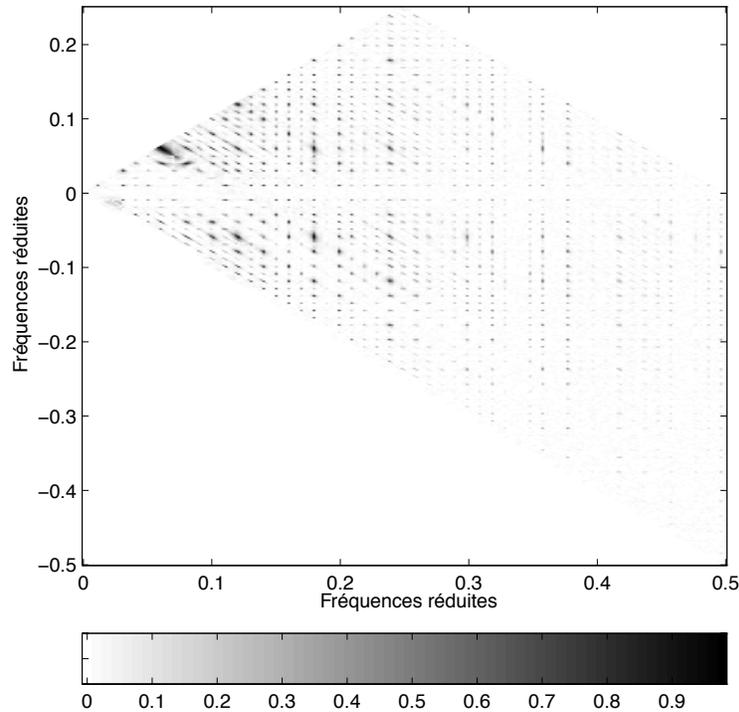


Figure 5.7: *Bicohérence entre la référence et la sortie sonar. 512 canaux fréquentiels utilisés (128 moyennes).*

qui apparaît entre la référence et la sortie du sonar. La non-linéarité est peut-être “pure” mais est peut-être également due à des effets de saturation. La méconnaissance de la physique du bateau nous empêche d’avancer plus dans les suppositions. Toutefois, nous pouvons essayer de vérifier quantitativement ces arguments qualitatifs en utilisant des statistiques d’ordre supérieur sensibles aux non-linéarités. L’outil adapté à cette recherche est la bicohérence, comme nous l’avons montré au paragraphe 5.1.5. Rappelons sa définition

$$\gamma_{o,x(3)}(m_1, m_2) = \frac{|E[X(m_1)X(m_2)O^*(m_1 + m_2)]|^2}{E[|X(m_1)|^2]E[|X(m_2)|^2]E[|O(m_1 + m_2)|^2]},$$

et répétons qu’elle présente des valeurs proches de 1 en (m_1, m_2) si les fréquences m_1 et m_2 de $x(n)$ ont interagi quadratiquement pour créer la fréquence $m_1 + m_2$ de $o(n)$. Nous avons calculé cette bicohérence pour les signaux utilisés ici, et le résultat apparaît sur la figure (5.7).

Commentons ce résultat en prenant un couple de fréquences à titre d’exemple. Sur la bicohérence apparaît un pic prononcé au couple $(0.18, 0.06)$. Ceci signifie donc que la raie située à la fréquence 0.24 de la sortie provient (peut être en partie) de l’interaction non linéaire entre les fréquences 0.18 et 0.06 de la référence. Nous attendons donc sur cette raie une meilleure soustraction en utilisant l’approche non linéaire qu’en utilisant les techniques linéaires.

Venons-en maintenant aux résultats après soustraction. Mentionnons de suite que ces signaux ont été déjà traités d'une façon approfondie par les techniques linéaires dans [98, 184, 205]. La conclusion obtenue dans [184] montre que ces signaux sont mieux adaptés à des traitements fréquentiels, puisque le travail d'identification porte essentiellement sur des raies. La modélisation transversale en temps est donc peu adéquate car elle nécessite un grand nombre de coefficients. Dans [98], où une approche multiréférences (nous disposons de plusieurs références bruit issues de capteurs disposés près de diverses machines) est développée, il est montré que la référence choisie ici donne de très bons résultats lorsqu'utilisée seule. Cette conclusion est confirmée dans [205] où le filtre identifié est à réponse impulsionnelle infinie (identification par prédiction). Donc, dans ce cas réel, choisir comme référence la plus cohérente avec la sortie semble être une décision correcte.

Conditions expérimentales et résultats La soustraction de bruit est effectuée dans le domaine temporel après identification du filtre en temps. Dans les cas linéaire et linéaire-quadratique, la mémoire retenue est $M = 30$. Le calcul des statistiques se réalise sur un horizon de 100 000 échantillons temporels.

La performance de soustraction est chiffrée par la différence entre le spectre en dB de sortie avant et après traitement. Cet indice de performance sera appelé gain. Ses valeurs positives dénotent une bonne soustraction de bruit. Les résultats sont montrés sur les figures (5.8) à (5.11).

Comme prévu, le traitement linéaire apporte de bons résultats dans la première moitié de la bande. Par contre son insuffisance, déjà prédite par la cohérence, apparaît nettement dans la partie $[0.25, 0.5]$ du spectre (figures (5.8) et (5.9)). Le traitement linéaire-quadratique (figures (5.10) et (5.11)) améliore la soustraction dans cette partie du spectre, comme la différence des spectres le montre. Ce gain est surtout visible sur la raie 0.24, ce qui confirme la non-linéarité détectée par la bicohérence pour le couple (0.06, 0.18). Notons également que des gains non négligeables sont obtenus en basse fréquence. Toutefois, la mémoire limitée choisie n'autorise pas d'amélioration exceptionnelle.

5.5 En guise de conclusion

La présentation des filtres de Volterra s'achève. Pour conclure, tentons de relever leurs faiblesses. Ces développements généralisent aux fonctionnelles les développements de Taylor. Ainsi, les systèmes "développables" doivent être suffisamment "doux", et les zones de convergence peuvent être limitées. L'identification de non-linéarités "dures", telles des saturations, sont donc difficilement approchables par des filtres de Volterra.

Un autre point délicat concerne l'aspect pratique des filtres discrets :

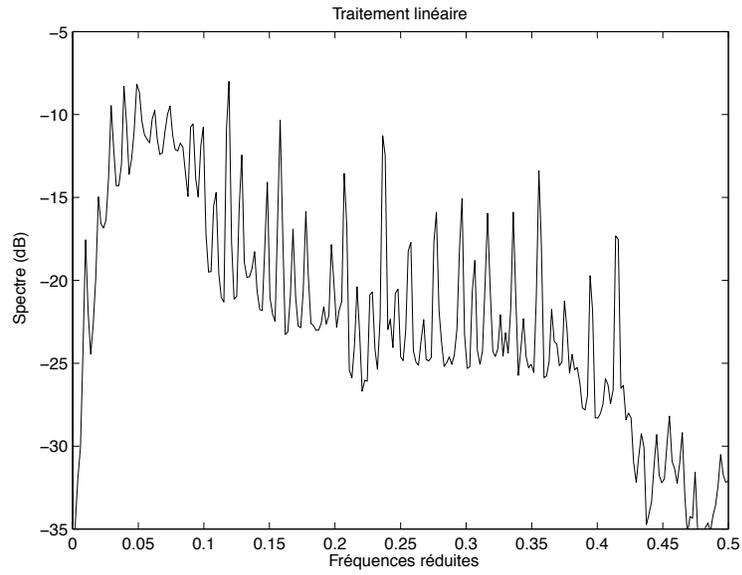


Figure 5.8: *Spectre en dB de l'estimée du signal utile par traitement linéaire.*

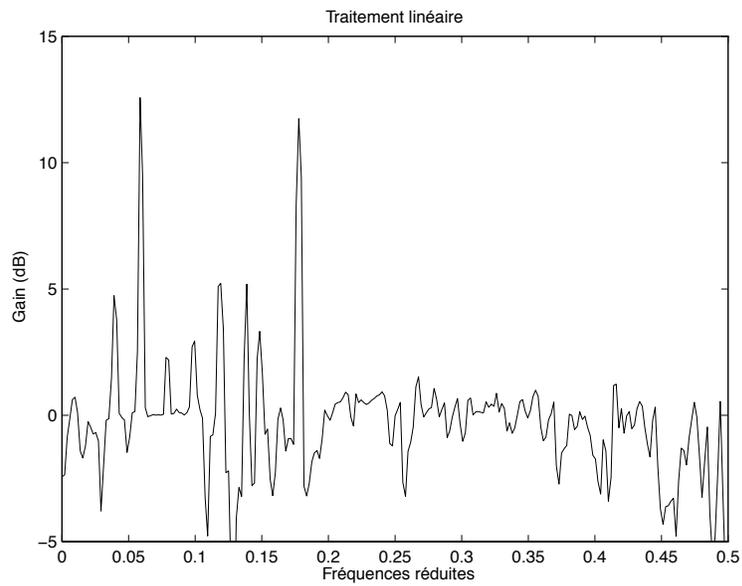


Figure 5.9: *Gain en dB pour le traitement linéaire.*

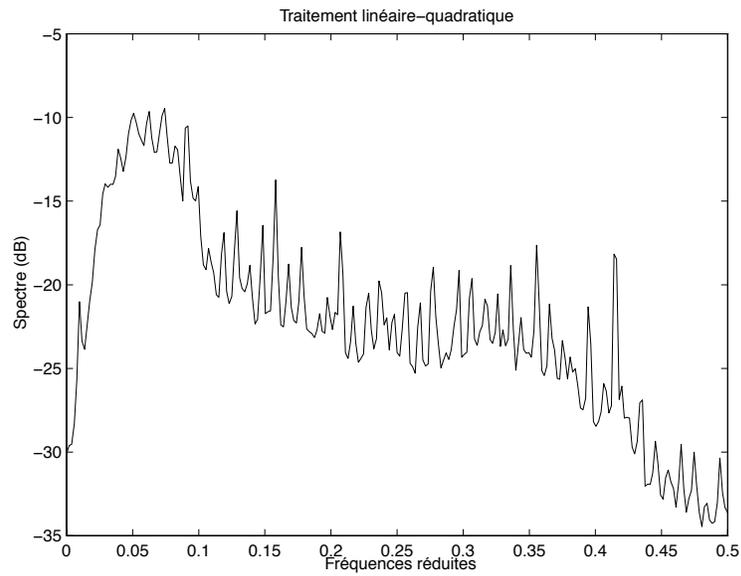


Figure 5.10: *Spectre en dB de l'estimée du signal utile par traitement linéaire-quadratique.*

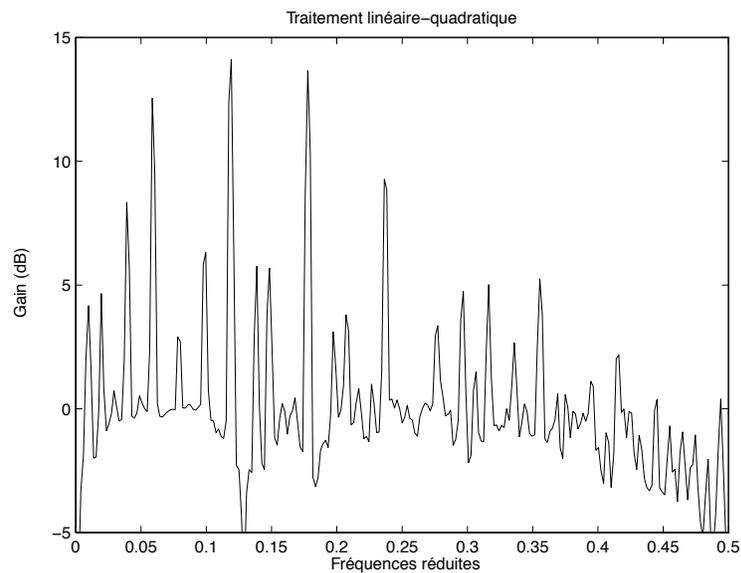


Figure 5.11: *Gain en dB pour le traitement linéaire-quadratique.*

leur identification requiert de chercher un nombre de paramètres croissant exponentiellement avec l'ordre du filtre. Dans les applications, les limitations informatiques empêchent d'identifier des ordres de non-linéarité élevés. Dans ce contexte, les approches orthogonales permettent une recherche non exhaustive de termes polynômiaux qui peuvent être d'ordre élevé. Une approche pour résoudre ce problème repose sur les filtres bilinéaires, qui incluent une récursivité non linéaire assurant un ordre infini au filtre polynômial. A temps discret, leur définition est

$$y(n) = \sum_{k=0}^q h(k)x(n-k) + \sum_{k=1}^p g(k)y(n-k) + \sum_{k=0}^r \sum_{l=1}^s c(k,l)x(n-k)y(n-l)$$

Un filtre bilinéaire est donc un filtre ARMA(p, q) auquel est adjoint un terme bilinéaire assurant un ordre de non-linéarité infini. Les filtres bilinéaires peuvent également se mettre sous forme matricielle. Ils sont appliqués à la soustraction de bruit pour le même exemple traité précédemment dans [152, 153]. L'ordre infini donné par le terme bilinéaire assure un "nettoyage" plus performant que les filtres de Volterra.

Pour terminer, mentionnons que les filtres de Volterra sont appliqués dans bon nombre de domaines, du traitement d'image à l'annulation d'échos, de l'identification de systèmes à l'égalisation de canal, ...

Partie II

Signaux multidimensionnels

Dans les chapitres précédents, nous avons présenté des techniques d'analyse s'appliquant à un signal dépendant d'une seule variable. Ces techniques de traitement sont adaptées aux situations dans lesquelles on ne dispose que d'un seul capteur, qui délivre un signal fonction du temps. Dans de nombreux cas, on dispose de plusieurs capteurs répartis dans l'espace, qui délivrent chacun un signal dépendant du temps. L'ensemble de ces signaux constitue un signal qualifié de spatio-temporel, multidimensionnel, ou multivariable : la terminologie varie selon les ouvrages.

Les signaux multidimensionnels issus d'un réseau de capteurs se rencontrent dans de nombreux domaines d'application. En astronomie ou en astrophysique, on utilise des réseaux d'antennes. Les systèmes radar utilisent de plus en plus les signaux issus d'un réseau de capteurs. Le sonar a toujours utilisé des réseaux d'hydrophones soit portés soit tractés. La prospection sismique met en œuvre des réseaux de géophones répartis sur le sol. L'électrocardiographie, l'électroencéphalographie produisent des signaux spatio-temporels captés en divers points du patient. Le contrôle non destructif réalise également un balayage de la surface de la pièce contrôlée. En télécommunication, ou pour l'interception en guerre électronique, on recourt de plus en plus souvent à plusieurs antennes. On pourrait multiplier les exemples de situations mettant en jeu des signaux multidimensionnels.

Le traitement des signaux multidimensionnels peut avoir plusieurs objectifs. Nous retiendrons ici les deux principaux qui sont la *localisation* des sources et la *séparation* des sources. Nous reviendrons sur la définition de ce que l'on appelle une source.

La localisation des sources est l'objectif traditionnel du *traitement d'antenne*. Le réseau de capteurs constitue alors une antenne. On considère souvent que les sources sont ponctuelles et très éloignées de l'antenne et l'on cherche à déterminer leur direction.

La *séparation de sources* a pour objectif de restituer les signaux émis par chacune des sources, qui sont mélangés lors de la réception sur l'antenne. Ce traitement est nécessaire quand on veut accéder à une estimation de la forme d'onde d'une ou de plusieurs sources reçues. Citons le contrôle de machines fonctionnant ensemble ou la séparation d'échos en radar.

Après avoir précisé le modèle sur lequel sont basées les diverses approches, nous présenterons, pour la localisation et pour la séparation de sources, les techniques classiques fondées sur l'utilisation de grandeurs du second ordre. Nous montrerons ensuite les limitations de ces techniques à l'ordre 2 et nous décrirons les nouveaux moyens de traitement faisant appel à des statistiques d'ordre supérieur à 2.

CHAPITRE 6

Traitement d'antenne

Les techniques classiques utilisées en traitement d'antenne à l'ordre 2 sont fondées sur la formation de voies et sur le goniomètre. Nous allons montrer que ces techniques peuvent être étendues aux ordres supérieurs à 2.

6.1 Modélisation

Le modèle des signaux reçus sur l'antenne est fixé par les signaux émis par les sources, modifiés par la propagation entre les sources et les capteurs.

6.1.1 Les signaux reçus

Nous nous plaçons ici dans le cas de signaux spatio-temporels. D'autres problèmes pourraient être modélisés de la même manière, notamment le traitement de signaux de natures physiques différentes reçus sur un réseau de capteurs. Le signal spatio-temporel est déterministe ou aléatoire, il dépend du temps et de l'espace. De manière générale un tel signal peut se représenter par une fonction de plusieurs variables $y(t, \mathbf{r})$, t étant l'instant de mesure et \mathbf{r} le vecteur définissant la position du capteur. Il est également possible de passer dans le domaine spectral et de traiter le signal spatio-fréquentiel¹ $y(\nu, \mathbf{r})$ résultant d'une transformée de Fourier de $y(t, \mathbf{r})$.

Nous nous limitons ici au traitement des signaux dépendant de la fréquence. Il existe à l'ordre 2 des méthodes de traitement travaillant directement sur les signaux temporels. Le fait de travailler en fréquence nous conduit à traiter des signaux à valeurs complexes alors que les signaux temporels sont en général à valeurs réelles.

Comme nous l'avons indiqué, on utilise les signaux reçus par un réseau de capteurs : ceci revient à discrétiser l'espace. Les signaux reçus sont alors de la forme

$$y(\nu, \mathbf{r}_i), 1 \leq i \leq K,$$

¹Pour simplifier les notations, nous représentons par le même symbole les signaux temporels ou spectraux. Le contexte permet de retrouver le domaine considéré.

K désignant le nombre de capteurs.

Il est bon que la discrétisation spatiale ne perturbe pas les mesures. Pour cela il faut vérifier des conditions de Shannon spatiales. On se limite souvent à la localisation de sources dans un plan, réalisée par un réseau rectiligne de capteurs équidistants – figure (6.1, page 211) –. Dans ce cas, la condition de Shannon spatiale stipule que l'écart entre deux capteurs doit être inférieur à la demi-longueur d'onde. Pour traiter un cas véritablement tridimensionnel on doit mettre en œuvre un réseau plan de capteurs. On les dispose souvent aux nœuds d'une grille rectangulaire ou carrée. La condition de Shannon exige que l'écart entre deux capteurs successifs sur une ligne ou une colonne soit toujours inférieur à la demi-longueur d'onde.

Dans le cas plan et dans le cas général (on parle de traitements 2D ou 3D), la localisation des capteurs est un premier problème important. En particulier dans la prospection sismique, comme l'on déplace le réseau de capteurs, il existe une grande variété de possibilités de combinaisons des signaux captés [141].

La matière première du traitement d'antenne ou de la séparation de sources est donc le signal spatio-fréquentiel $y(\nu, \mathbf{r}_i)$. On simplifie les écritures en adoptant des notations vectorielles, plus compactes. Les K composantes du signal sont regroupées dans un vecteur colonne

$$\mathbf{y}(\nu) = \begin{pmatrix} y(\nu, \mathbf{r}_1) \\ \vdots \\ y(\nu, \mathbf{r}_K) \end{pmatrix},$$

auquel on associe le vecteur ligne transposé et conjugué, $\mathbf{y}^\dagger(\nu)$. En pratique, les signaux sont échantillonnés en temps et la transformation de Fourier est réalisée de façon discrète induisant un échantillonnage en fréquence. Si la durée du signal traité est T , on obtient les valeurs suivantes de la fréquence : $\nu_k = k/T$. Nous reviendrons à cette forme pratique quand cela sera utile mais en général nous utiliserons les signaux continus en fréquence.

Pour compléter le modèle d'observation, nous devons préciser la forme des sources et les conditions de propagation entre les sources et les capteurs.

6.1.2 Les sources

Pour être localisables et séparables, les sources doivent influencer les signaux reçus (observabilité). On distingue deux situations, l'une dite passive et l'autre active. Ces qualificatifs sont relatifs à l'expérimentateur. Dans la situation passive l'expérimentateur reçoit des signaux issus de sources qu'il ne contrôle pas. C'est le cas en astronomie où les astres sont les émetteurs, et ce peut être le cas aussi en radar ou en sonar (écoute passive électromagnétique ou acoustique). C'est le cas également en électrocardiographie ou en électroencéphalographie.

Dans le cas actif, le signal est émis par un système contrôlé par l'expérimentateur. Ce signal est renvoyé (après déformation) par la source qui l'a marqué de son empreinte. C'est le cas en radar et en sonar actifs où les sources, que l'on appelle alors des cibles, réfléchissent un code émis par l'expérimentateur. C'est également le cas en prospection sismique où une onde élastique est envoyée dans le sous-sol. Les sources sont ici les interfaces entre les couches géologiques du sous-sol. Ces interfaces renvoient vers l'antenne un signal réfléchi, réfracté ou diffracté selon les conditions expérimentales.

Nous considérons les sources comme des émetteurs de signaux (même s'il ne s'agit que d'objets réfléchissant ou réfractant d'autres émissions). Nous postulons ainsi une connaissance très faible quant aux signaux émis par les sources, notre objectif étant, dans la séparation de sources, de retrouver la forme de ces signaux. Dans le cas actif, cette approche est clairement sous-optimale, puisque l'on connaît le signal qui a été émis, avant propagation et déformation par la cible². Lorsque la forme du signal est connue, le seul objectif est la localisation des sources.

Si nous ne postulons strictement rien sur les signaux émis par les sources, on imagine aisément que les problèmes de séparation ou de localisation sont insolubles. On ajoute donc une hypothèse fondamentale pour la suite de notre étude : on suppose que les signaux émis par les sources sont *statistiquement indépendants*.

Ce point capital doit être discuté. Pour des signaux aléatoires notre hypothèse a un sens. Qu'en est-il des signaux déterministes? Expliquons-le sur un exemple simple. Soient deux signaux aléatoires $x(t)$ et $y(t)$, centrés, stationnaires et indépendants. En travaillant à l'ordre 2, l'indépendance se réduit au fait que le moment croisé d'ordre 2 est nul, soit

$$C_{xy(2)}(\tau) = E[x(t)y(t+\tau)] = 0, \forall \tau$$

Pratiquement nous ne disposons que d'une seule réalisation des signaux $x(t)$ et $y(t)$, connue sur la durée T , et nous estimons le moment croisé en postulant des propriétés d'ergodisme (voir chapitre 2). La fonction d'intercorrélacion est alors la moyenne temporelle

$$C_{xy(2)}(\tau) = \frac{1}{T} \int_{-T/2}^{T/2} x(t)y(t+\tau)dt.$$

Nous considérons les signaux $x(t)$ et $y(t)$ comme indépendants si cette moyenne temporelle est "approximativement" nulle. Ainsi définie, la décorrélation s'applique aussi bien aux signaux déterministes qu'aux signaux aléatoires. On généralise alors cette approche aux ordres supérieurs : nous dirons que deux si-

²Des techniques prenant en compte la connaissance de la forme du signal ont été développées [96].

gnaux déterministes sont indépendants si tous les cumulants croisés empiriques des deux signaux sont nuls³.

L'hypothèse d'indépendance est assez naturelle dans certains des exemples cités plus haut : sonar passif, astronomie... Elle est plus difficile à réaliser dans d'autres situations comme la prospection sismique, mais on peut se placer dans des conditions simulant l'indépendance en utilisant de "bons" estimateurs des grandeurs caractéristiques [123].

6.1.3 La propagation

Les travaux classiques supposent que la propagation est linéaire, stationnaire et déterministe. Ceci est conforme aux lois de propagation des ondes tant que la perturbation qu'elles engendrent reste petite. On peut parler du "principe de superposition des petits mouvements". La propagation entre chaque source et chaque capteur est alors un filtrage linéaire et invariant dans le temps.

En supposant la présence de P sources émettant des signaux $s_p(t)$, le signal $y_i(t)$ reçu sur le capteur i s'écrit

$$y_i(t) = \sum_{p=1}^P \int a_{ip}(t - \tau) s_p(\tau) d\tau + b_i(t), \quad \forall i = 1, \dots, K.$$

Dans cette relation, $a_{ip}(\cdot)$ représente la réponse impulsionnelle du filtre décrivant la propagation entre la source p et le capteur i ; c'est la fonction de Green des physiciens. Le terme $b_i(t)$ désigne un bruit additif, nécessaire dans toute modélisation réaliste.

Dans le domaine spectral, l'équation précédente devient

$$y_i(\nu) = \sum_{p=1}^P a_{ip}(\nu) s_p(\nu) + b_i(\nu), \quad \forall i = 1, \dots, K. \quad (6.1)$$

On introduit alors les vecteurs

$$\mathbf{a}_p(\nu) = \begin{pmatrix} a_{1p}(\nu) \\ \vdots \\ a_{Kp}(\nu) \end{pmatrix},$$

qui sont dénommés *vecteurs sources*. Ils décrivent la propagation entre la source p et les capteurs du réseau.

Les composantes à valeurs complexes $s_p(\nu)$ peuvent s'interpréter de deux façons. Lorsque le signal spectral est obtenu par transformation de Fourier du signal temporel, elles décrivent la composante spectrale à la fréquence ν du

³Voir la discussion sur les multicorrélations de signaux déterministes page 65.

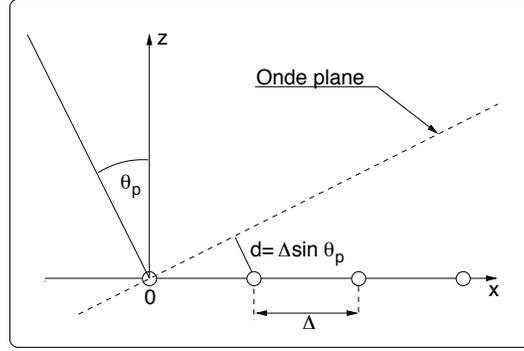


Figure 6.1: Réseau rectiligne de capteurs équidistants. Les capteurs sont séparés par une distance Δ , l'angle entre l'axe Oz et le vecteur d'onde est θ_p , de sorte que la différence de marche entre deux capteurs est $\Delta \sin \theta_p$.

signal émis par la source p . Lorsque, comme cela est souvent le cas, les signaux sont à bande étroite, ce sont les amplitudes complexes des signaux émis par les sources. Enfin, $b_i(\nu)$ est le bruit sur le capteur i .

La relation (6.1) peut s'écrire de manière plus compacte selon

$$\mathbf{y}(\nu) = \sum_{p=1}^P \mathbf{a}_p(\nu) s_p(\nu) + \mathbf{b}(\nu) = A(\nu) \mathbf{s}(\nu) + \mathbf{b}(\nu).$$

La matrice $A(\nu)$, de dimension $K \times P$, rassemble les fonctions de transfert entre les sources et les capteurs, le vecteur $\mathbf{s}(\nu)$ est formé des signaux émis par les P sources et le vecteur $\mathbf{b}(\nu)$ des K composantes du bruit.

Pour conduire l'opération de localisation, on doit préciser les conditions de propagation décrites par les vecteurs $\mathbf{a}_p(\nu)$, colonnes de la matrice $A(\nu)$. Donnons l'exemple de sources à l'infini.

Sources à l'infini On peut souvent admettre que les sources sont à l'infini et que la propagation se fait dans un milieu homogène. Dans ces conditions la localisation revient à déterminer la direction de chaque source.

On peut alors préciser les vecteurs sources. Le signal émis par chaque source est porté par une onde plane. La propagation entre deux capteurs successifs introduit simplement un déphasage dépendant de la célérité des ondes, de la direction de la source et de la géométrie du réseau de capteurs. Le cas le plus simple est celui d'un réseau rectiligne de capteurs équidistants (représenté

sur la figure (6.1)). Dans ce cas le vecteur source est

$$\mathbf{a}_p(\nu) = \beta_p e^{j\psi_p} \begin{pmatrix} 1 \\ e^{-j\varphi_p} \\ e^{-2j\varphi_p} \\ \vdots \\ e^{-(K-1)j\varphi_p} \end{pmatrix}.$$

Dans cette expression

- $\beta_p e^{j\psi_p}$ représente la propagation entre la source et le capteur 1. Cette partie du modèle n'est pas identifiable⁴. Les paramètres β_p et ψ_p sont donc arbitraires. On pourra normaliser les vecteurs sources en choisissant $\beta_p = 1/\sqrt{K}$. Pour ψ_p le choix fait ici fixe à 0 la phase de la composante du vecteur source sur le premier capteur. On choisit parfois de repérer les phases par rapport au "centre de phase" de l'antenne en fixant $\psi_p = -(K-1)\varphi_p/2$.
- φ_p est le déphasage entre deux capteurs successifs. En appelant (figure (6.1)) θ_p l'angle entre la normale à l'onde et la normale à l'antenne, x_p la distance entre le capteur p et le capteur de référence, c la célérité des ondes au niveau des capteurs et ν la fréquence, on obtient

$$\varphi_p = \frac{2\pi\nu x_p \sin \theta_p}{c}$$

, pour un réseau rectiligne de capteurs équidistants, Δ étant la distance entre deux capteurs successifs et $x_p = p\Delta$.

Modèle général d'observation Le vecteur observé sur l'antenne s'écrit

$$\mathbf{y} = \sum_{p=1}^P s_p \mathbf{a}_p + \mathbf{b}, \quad (6.2)$$

en omettant la fréquence ν pour simplifier les écritures.

Les vecteurs sources appartiennent à un ensemble de vecteurs $\mathbf{d}(\boldsymbol{\alpha})$, fixé par le modèle de propagation, et dénommé la *variété d'antenne*. $\boldsymbol{\alpha}$ est un vecteur contenant les paramètres décrivant les ondes portant les sources. Par exemple, pour des ondes sphériques, $\boldsymbol{\alpha}$ contient les angles définissant la direction de la source et la distance entre la source et le centre de l'antenne.

L'objectif du traitement d'antenne est de localiser les sources en déterminant les valeurs des paramètres de localisation $\boldsymbol{\alpha}_p$ correspondant aux

⁴Nous reviendrons au chapitre 7 sur l'identifiabilité.

sources présentes. Lorsque le modèle est adapté à la situation physique, les vecteurs sources sont liés aux vecteurs directionnels $\mathbf{d}(\boldsymbol{\alpha})$ par

$$\mathbf{a}_p = \mathbf{d}(\boldsymbol{\alpha}_p),$$

les paramètres $\boldsymbol{\alpha}_p$ prenant les valeurs correspondant aux sources présentes.

Le modèle que nous venons de présenter sert de base aux nombreuses méthodes de traitement d'antenne, utilisant des grandeurs du second ordre, qui ont été développées durant les vingt dernières années.

Nous allons présenter les idées force sur lesquelles reposent ces méthodes, et les principaux outils de traitement. Nous introduirons ainsi les développements récents du traitement d'antenne utilisant des statistiques d'ordre supérieur.

6.2 Formalisme et outils à l'ordre 2

Ce paragraphe a pour objet le rappel des fondements du traitement d'antenne à l'ordre 2.

6.2.1 Sous-espaces signal et bruit

Le modèle linéaire de propagation présenté dans la section précédente permet de séparer l'espace vectoriel des observations en deux sous-espaces : le sous-espace signal et le sous-espace bruit. Le vecteur d'observation \mathbf{y} appartient à un espace de Hilbert \mathcal{H}_y de dimension K : *l'espace des observations*.

Lorsque le nombre K de capteurs est supérieur au nombre P de sources, et lorsque l'ensemble des vecteurs sources \mathbf{a}_p forme une famille libre de vecteurs⁵, les vecteurs sources \mathbf{a}_p engendrent le sous-espace hilbertien \mathcal{H}_s de l'espace des observations \mathcal{H}_y , de dimension P , que l'on appelle *le sous-espace signal*. Le sous-espace orthogonal à \mathcal{H}_s dans \mathcal{H}_y , de dimension $K - P$, que l'on notera \mathcal{H}_b , ne contient que du bruit. Il est appelé *le sous-espace bruit*. La construction des sous-espaces signal et bruit permet

- de déterminer le nombre de sources, qui est donné par la dimension du sous-espace signal.
- de réduire le bruit. En effet, en projetant les observations sur le sous-espace source on élimine une partie du bruit sans altérer le signal.

⁵Cette hypothèse est nécessaire pour que les sources soient discernables.

6.2.2 La matrice spectrale

Dans la suite de ce chapitre nous pourrons noter les composantes des matrices M_i^j , l'indice inférieur i étant l'indice de ligne et l'indice supérieur j , l'indice de colonne. Les indices inférieurs sont associés à des termes non-complexes conjugués et les indices supérieurs à des termes complexes conjugués.

La matrice spectrale $S_{\mathbf{y}(1)}^{(1)}(\nu) = \text{Cum}[\mathbf{y}(\nu), \mathbf{y}^\dagger(\nu)]$ (qui peut également être notée $\text{Cum}[y_i(\nu), y_j^*(\nu)] = S_{\mathbf{y}i}^j(\nu)$) est la matrice de covariance des observations spectrales $\mathbf{y}(\nu)$; elle est de dimension $K \times K$, K étant le nombre de capteurs. Elle résume toutes les propriétés statistiques d'ordre 2 lorsque les mesures sont stationnaires. La matrice spectrale est formée des spectres et des interspectres des signaux reçus sur les capteurs. Elle est hermitienne car $S_{\mathbf{y}i}^j(\nu)^* = S_{\mathbf{y}j}^i(\nu)$.

De nombreuses méthodes d'estimation de la matrice spectrale à partir des observations ont été développées. Pour estimer la matrice spectrale, on peut utiliser des moyennes temporelles ou fréquentielles [123]. On a aussi imaginé d'autres types de moyennes, basées sur des invariances de la matrice spectrale. Pour un réseau rectiligne de capteurs équidistants recevant des ondes planes, la matrice spectrale est Töplitz, c'est-à-dire que $S_{\mathbf{y}i+\ell}^{j+\ell}(\nu) = S_{\mathbf{y}i}^j(\nu)$, pour toutes les valeurs de (i, j, ℓ) admissibles. C'est cette dernière propriété qui est exploitée dans le lissage spatial [40, 186].

Dans le modèle donné au paragraphe 6.1.3, les observations sont $\mathbf{y} = \mathbf{A}\mathbf{s} + \mathbf{b} = \mathbf{m} + \mathbf{b}$, en appelant \mathbf{m} la partie des observations issue des sources (signal). Le bruit étant supposé statistiquement indépendant du signal, la matrice spectrale des observations est la somme des matrices spectrales du signal et du bruit

$$S_{\mathbf{y}(1)}^{(1)} = S_{\mathbf{m}(1)}^{(1)} + S_{\mathbf{b}(1)}^{(1)}.$$

Dans cette expression la matrice spectrale des observations non-bruitées, $S_{\mathbf{m}(1)}^{(1)} = \text{Cum}[\mathbf{m}, \mathbf{m}^\dagger]$, est

$$S_{\mathbf{m}(1)}^{(1)} = \sum_{p=1}^P \text{Cum}[s_p, s_p^*] \mathbf{a}_p \mathbf{a}_p^\dagger = \sum_p S_{s_p}^p \mathbf{a}_p \mathbf{a}_p^\dagger = \mathbf{A} S_{\mathbf{s}(1)}^{(1)} \mathbf{A}^\dagger, \quad (6.3)$$

$\text{Cum}[s_p, s_p^*]$ étant la puissance de la source p , \mathbf{a}_p les vecteurs sources et $S_{\mathbf{s}(1)}^{(1)} = \text{Cum}[\mathbf{s}, \mathbf{s}^\dagger]$ la matrice spectrale des sources.

Par suite de l'indépendance statistique des P sources la matrice spectrale $S_{\mathbf{s}(1)}^{(1)}$ est une matrice diagonale de rang P . La matrice $S_{\mathbf{m}(1)}^{(1)}$ est donc de rang P et son noyau est le sous-espace bruit. La matrice spectrale $S_{\mathbf{m}(1)}^{(1)}$ permet donc de déterminer le nombre de sources et les sous-espaces signal et bruit.

A partir des observations on n'accède pas à $S_{\mathbf{m}(1)}^{(1)}$ mais à $S_{\mathbf{y}(1)}^{(1)}$. On peut cependant déduire le nombre de sources et les sous-espaces signal et bruit à partir de la matrice spectrale des observations $S_{\mathbf{y}(1)}^{(1)}$:

- lorsque le bruit est nul ou négligeable car alors $S_{\mathbf{y}(1)}^{(1)} = S_{\mathbf{m}(1)}^{(1)}$;
- lorsque la matrice spectrale du bruit est connue ou lorsque le bruit est blanc. Dans ce dernier cas la matrice spectrale du bruit est proportionnelle à l'identité $S_{\mathbf{b}(1)}^{(1)} = P_b I$. Les valeurs propres de la matrice spectrale des observations font apparaître un ensemble de valeurs propres minimales égales à la puissance du bruit blanc. Les vecteurs propres associés à ces valeurs propres minimales engendrent le noyau de la matrice spectrale des observations non-bruitées qui est le sous-espace bruit. Le sous-espace complémentaire du sous-espace bruit, engendré par les vecteurs propres associés aux valeurs propres non minimales, est le sous-espace signal. On montre dans [27] que les éléments propres de la matrice spectrale estimée des observations donnent des estimateurs du maximum de vraisemblance du nombre de sources (nombre de valeurs propres non minimales), de la puissance du bruit (valeurs propres minimales) et des sous-espaces signaux et bruits.

Pour terminer, faisons deux remarques.

Lorsque le bruit n'est pas spatialement blanc des solutions existent si l'on peut paramétrer par un "petit" nombre de valeurs la matrice spectrale du bruit. Ce cas est cependant difficile et les statistiques d'ordre supérieur seront alors utiles si le bruit est gaussien.

Les résultats donnés ci-dessus nous indiquent que l'étude à l'ordre 2 permet de déterminer le sous-espace signal mais ne détermine pas, dans ce sous-espace, les vecteurs sources ; c'est le problème de la séparation aveugle de sources qui, comme nous le montrerons au chapitre 7, est résolu en utilisant des statistiques d'ordre supérieur à 2.

6.3 Formalisme et outils aux ordres supérieurs

Pour étendre les techniques de traitement d'antenne aux ordres supérieurs, on se place souvent à l'ordre 4 qui contient, en général, les premiers cumulants d'ordre supérieur à 2 non nuls . Dans cette approche à l'ordre 4, la matrice spectrale est remplacée par la quadricovariance [43]. Un formalisme plus général faisant intervenir des moments hybrides non linéaires a été développé [107].

6.3.1 La quadricovariance

Aux ordres supérieurs à 2, il convient d'utiliser les cumulants d'ordre 4. En effet, pour diverses raisons que nous avons déjà évoquées dans les chapitres précédents, ces statistiques sont préférées aux moments. De plus, les cumulants d'ordre 3 sont souvent mal conditionnés (*e.g.* quasi nuls) dans la grande majorité des cas (circularité des données, symétrie de la distribution). Pour les signaux à valeurs complexes considérés ici il existe trois cumulants d'ordre 4 différents. Les signaux spectraux, obtenus par transformation de Fourier, ou les signaux bande étroite, utilisés en traitement d'antenne, sont circulaires comme nous l'avons montré au chapitre 2. Le seul cumulant d'ordre 4 non nul est le trispectre contenant deux termes non conjugués et deux termes conjugués. Ce trispectre, cumulant d'ordre 4 des composantes spectrales, est dénommé *quadricovariance*. Dans le traitement d'antenne aux ordres supérieur, la quadricovariance joue un rôle central analogue à celui joué par la matrice spectrale dans le traitement d'antenne à l'ordre 2.

La quadricovariance des observations est⁶

$$S_{\mathbf{y}}^{kl} = Cum[y_i(\nu), y_j(\nu), y_k^*(\nu), y_l^*(\nu)],$$

et peut également être notée sous la forme

$$S_{\mathbf{y}}^{(2)} = Cum[\mathbf{y}, \mathbf{y}, \mathbf{y}^\dagger, \mathbf{y}^\dagger]$$

où l'on omet la variable ν . En reprenant le modèle des observations donné en 6.1.3, $\mathbf{y} = \mathbf{A}\mathbf{s} + \mathbf{b} = \mathbf{m} + \mathbf{b}$, la quadricovariance des observations non-bruitées est

$$S_{\mathbf{m}}^{kl} = \sum_p a_{ip} a_{jp} a_{kp}^* a_{lp}^* S_{\mathbf{s}}^{pp}. \quad (6.4)$$

Dans cette relation les coefficients a_{ip} sont les termes de la matrice de mélange \mathbf{A} , composantes des vecteurs sources. Les termes $S_{\mathbf{s}}^{pp}$ sont les trispectres des sources, quand on travaille après transformation de Fourier, ou les cumulants d'ordre 4 des sources, quand on travaille en bande étroite. Seuls les termes diagonaux des trispectres ou des cumulants des sources sont non nuls puisque les sources sont statistiquement indépendantes.

Dans la relation (6.4), les éléments $a_{jp} a_{kp}^*$ sont les coordonnées des produits de Kronecker⁷ des vecteurs sources $\mathbf{a}_p \otimes \mathbf{a}_p^* = \mathbf{a}_p \mathbf{a}_p^\dagger$. Ces produits de Kronecker sont des grandeurs à K^2 composantes.

Avec ces notations la quadricovariance des observations non-bruitées s'écrit

⁶ Rappelons que les indices inférieurs sont associés aux termes non complexes conjugués et les indices supérieurs aux termes complexes conjugués.

⁷ Pour certains lecteurs, nous éclairerons ce sujet en indiquant que le produit de Kronecker est identique au produit tensoriel.

$$S_{\mathbf{m}}^{kl} = \sum_p S_{\mathbf{s}}^{pp} [\mathbf{a}_p \otimes \mathbf{a}^p] [\mathbf{a}_p \otimes \mathbf{a}^p]^\dagger. \quad (6.5)$$

En comparant cette relation à (6.3), on voit apparaître une analogie forte entre la forme de la matrice spectrale et celle de la quadricovariance. La matrice spectrale est une forme quadratique diagonale construite à partir des vecteurs sources et la quadricovariance est également une forme quadratique diagonale construite non sur les vecteurs sources mais sur les produits de Kronecker des vecteurs sources. Cette analogie est à la base de l'extension des techniques de traitement d'antenne développées à l'ordre 2 à des ordres supérieurs, ici l'ordre 4. Ces résultats permettent d'introduire les sous-espaces signal et bruit à l'ordre 4.

6.3.2 Sous-espaces signal et bruit

Comme nous l'avons vu au paragraphe 6.2.1, les observations engendrent un espace de Hilbert à K dimensions, l'espace des observations et les vecteurs source engendrent un sous-espace de l'espace des observations dénommé l'espace signal. A l'ordre 4, on doit remplacer, comme le montre la forme (6.5) de la quadricovariance, les vecteurs observés et les vecteurs sources par leurs produits de Kronecker. L'espace des observations construit avec les produits de Kronecker des observations est un espace à K^2 dimensions. Le sous-espace signal d'ordre 4, introduit par [43] et dénommé FOSS⁸, est l'espace à P dimensions construit avec toutes les combinaisons linéaires des produits de Kronecker $\mathbf{a}_p \otimes \mathbf{a}^p$ des vecteurs sources. Le sous-espace bruit d'ordre 4 est le complémentaire du sous-espace signal dans l'espace des observations.

6.3.3 Nombre de sources séparables

Le fait de remplacer les vecteurs observés par leur produit de Kronecker augmente la dimension de l'espace des observations et permet ainsi d'augmenter le nombre de sources séparables. Une démonstration très concrète et élégante en est donnée dans [50].

Revenons aux vecteurs sources donnés en fonction des propriétés géométriques du réseau de capteurs pour des sources portées par des ondes planes. La source p est portée par une onde plane repérée par les angles θ_p et χ_p (figure (6.2)). Le déphasage entre un capteur situé à l'origine des coordonnées et un capteur situé au point de coordonnées x_k et y_k est

$$\varphi_k = 2\pi \frac{\sin \chi_p (x_k \cos \theta_p + y_k \sin \theta_p)}{\lambda}.$$

⁸Fourth-Order Signal Subspace.

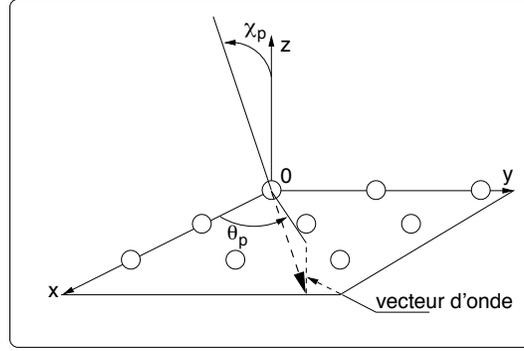


Figure 6.2: Réseau plan de capteurs. Le vecteur d'onde \mathbf{k} fait un angle χ_p avec l'axe Oz , et sa projection orthogonale sur le plan Oxy fait un angle θ_p avec l'axe Ox . La différence de marche entre les capteurs \mathbf{r}_l et \mathbf{r}_m est donnée par le produit scalaire $\mathbf{k}^T(\mathbf{r}_l - \mathbf{r}_m)$.

La phase ϕ_{lm} de la composante $a_{lp}a_{mp}^*$ du produit de Kronecker de ce vecteur source par lui-même est

$$\phi_{lm} = 2\pi \frac{\sin \chi_p [(x_l - x_m) \cos \theta_p + (y_l - y_m) \sin \theta_p]}{\lambda}.$$

Le produit de Kronecker engendre ainsi un *réseau virtuel* de capteurs dont les positions sont données par $x_l - x_m$ et $y_l - y_m$. Les K couples correspondant aux mêmes coordonnées engendrent le même capteur virtuel. Finalement, comme le montrent [50] et [149], le nombre de capteurs virtuels est au maximum $K_v = K^2 - K + 1$. Ce nombre maximum est atteint si la disposition des capteurs n'entraîne pas la confusion de certains capteurs virtuels. Pour un réseau rectiligne de capteurs équidistants il y a de multiples superpositions et, comme les lecteurs pourront le vérifier, le nombre de capteurs virtuels est de $2N - 1$. Les antennes virtuelles associées à une antenne rectiligne à capteurs équidistants et à une antenne circulaire sont présentées sur la figure (6.3).

Cette approche géométrique nous montre qu'en utilisant la quadricovariance on pourra, comme l'avait montré [43], localiser plus de sources qu'il n'y a de capteurs. Un raisonnement similaire montre que le même principe s'applique à une autre méthode haute-résolution, dénommée Esprit, qu'il est possible d'utiliser aux ordres supérieurs, avec éventuellement des capteurs virtuels [149]

On peut aussi utiliser le réseau virtuel pour améliorer la résolution de l'antenne. Mais il faut bien entendu s'attendre à ce que les performances obtenues avec K capteurs virtuels soient en général moins bonnes que celles de K capteurs réels. Ceci a été démontré pour la méthode Esprit dans [204].

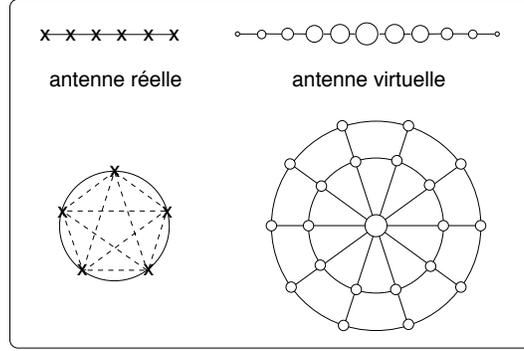


Figure 6.3: Antennes réelles et virtuelles pour un réseau rectiligne et un réseau circulaire.

6.3.4 Les moments hybrides non linéaires

La quadricovariance est associée au cumulants d'ordre 2 du module carré des observations. Elle peut être vue comme un moment d'ordre 2 de transformées non linéaires des observations. Cette approche a été formalisée dans [107] par l'introduction des moments hybrides non linéaires. L'introduction de non-linéarités dans le traitement est déjà apparue au chapitre 3 pour la déconvolution aveugle par les techniques de Bussgang, nous la retrouverons au chapitre 7 pour la séparation de sources. Avant de présenter les techniques de traitement d'antenne utilisant les moments hybrides non linéaires, donnons quelques relations utiles.

Généralisation des relations de Bussgang Soient deux variables aléatoires z_1 et z_2 à valeurs complexes. Soit $g(z_2, z_2^*)$ une fonction non linéaire de z_2 et z_2^* . Le moment hybride non linéaire construit avec la fonction g est $E[z_1 g^*(z_2, z_2^*)]$. En développant g en série de Taylor on obtient

$$E[z_1 g^*(z_2, z_2^*)] = \sum_{r=0}^{\infty} \sum_{t=0}^{\infty} \frac{1}{t!r!} \left. \frac{\partial^{t+r} g^*(z_2, z_2^*)}{\partial z_2^r \partial z_2^{*t}} \right|_{(0,0)} E[z_1 z_2^r z_2^{*t}]. \quad (6.6)$$

Dans cette relation on considère les variables z_2 et z_2^* comme algébriquement indépendantes. Il a été montré dans [15] et dans [180] que cela est licite.

En remplaçant les moments apparaissant dans (6.6) par les cumulants on obtient après quelques manipulations algébriques fastidieuses

$$E[z_1 g^*(z_2, z_2^*)] = \sum_{l=0}^{\infty} \sum_{n=0}^{\infty} \frac{1}{l!n!} E^* \left[\frac{\partial^{n+l} g^*(z_2, z_2^*)}{\partial z_2^n \partial z_2^{*l}} \right] \text{Cum}[z_1, \underbrace{z_2, \dots, z_2}_{n \text{ fois}}, \underbrace{z_2^*, \dots, z_2^*}_{l \text{ fois}}].$$

Lorsque les variables aléatoires z_1 et z_2 sont circulaires la formule se

simplifie par suite de l'annulation des cumulants non symétriques

$$E[z_1 g^*(z_2, z_2^*)] = \sum_{l=0}^{\infty} \frac{1}{l!(l+1)!} E^* \left[\frac{\partial^{2l+1} g^*(z_2, z_2^*)}{\partial z_2^{l+1} \partial z_2^{*l}} \right] \\ \times \text{Cum}[z_1, \underbrace{z_2, \dots, z_2}_{l \text{ fois}}, \underbrace{z_2^*, \dots, z_2^*}_{l+1 \text{ fois}}].$$

Ces relations, données pour des variables aléatoires quelconques, se simplifient lorsque les variables aléatoires sont gaussiennes. Dans ce cas, le seul cumulant non nul est le cumulant d'ordre 2.

Pour des variables aléatoires gaussiennes à valeurs réelles, on obtient la formule de Bussgang [39]

$$E[x_1 g(x_2)] = E \left[\frac{\partial g(x_2)}{\partial x_2} \right] \text{Cum}[x_1, x_2],$$

qui se généralise aux variables aléatoires à valeurs complexes gaussiennes circulaires

$$E[x_1 g(x_2, x_2^*)] = E^* \left[\frac{\partial g(x_2, x_2^*)}{\partial x_2} \right] \text{Cum}[x_1, x_2^*].$$

Moments hybrides non linéaires en traitement d'antenne Les observations sont des variables aléatoires à valeurs complexes, circulaires. La covariance non linéaire du vecteur d'observation \mathbf{y} est définie par

$$S_{\mathbf{y},g}^{(1)} = E[\mathbf{y}\mathbf{z}^\dagger].$$

Dans cette relation, les composantes du vecteur \mathbf{z} sont des transformées, par la fonction non linéaire g , des composantes des observations y_j et de leurs complexes conjuguées y_j^*

$$z_j = g(y_j, y_j^*).$$

Pour le modèle d'observation donné au paragraphe 6.1.3, $\mathbf{y} = \mathbf{A}\mathbf{s} + \mathbf{b}$, le bruit étant supposé gaussien, la matrice de covariance non linéaire s'écrit, en utilisant les relations données ci-dessus,

$$S_{\mathbf{y},g}^{(1)} = \mathbf{A} S_{\mathbf{s},g}^{(1)} \mathbf{A}^\dagger + K_1^0 S_{\mathbf{b}}^{(1)}. \quad (6.7)$$

Dans cette relation,

- $S_{\mathbf{s},g}^{(1)}$ est la matrice de covariance non linéaire des sources. Par suite de l'indépendance des sources cette matrice est diagonale. Les termes diagonaux sont

$$S_{\mathbf{s},g}^{(1)} = \sum_{n=0}^{\infty} \frac{1}{(n+1)!n!} K_{n+1}^n(y, y^*) \text{Cum} \left[\underbrace{s_l, \dots, s_l}_{n+1 \text{ fois}}, \underbrace{s_l^*, \dots, s_l^*}_{n+1 \text{ fois}} \right],$$

où

$$K_{n+1}^n(y, y^*) = E^* \left[\frac{\partial^{2n+1} g(y, y^*)}{\partial y^{n+1} \partial y^{*n}} \right].$$

Dans cette relation, y représente une des composantes des observations. Cette composante n'a pas à être précisée car on suppose que toutes les observations ont les mêmes propriétés statistiques.

- $K_1^0 = E^* \left[\frac{\partial g(y, y^*)}{\partial y} \right]$ est le terme de “Bussgang”,
- $S_{\mathbf{b}(1)}^{(1)}$ est la matrice de covariance du bruit.

La relation (6.7) généralise la relation (6.3) donnant la matrice spectrale des observations. Nous verrons au paragraphe 6.5.3 que cette relation permet de retrouver les sous-espaces signal et bruit en présence de bruit non blanc.

Après cette présentation générale des outils du traitement d'antenne à un ordre supérieur à 2, voyons plus précisément les techniques qui peuvent être mises en œuvre. Toutes les méthodes de traitement d'antenne peuvent se formuler à partir de la matrice spectrale. L'idée directrice pour l'extension à l'ordre 4 est de faire jouer à la quadricovariance, ou aux moments hybrides non linéaires, le rôle joué à l'ordre 2 par la matrice spectrale.

6.4 Formation de voies

La formation de voies est la technique la plus utilisée en traitement d'antenne. Elle est fondée sur des statistiques d'ordre 2, nous allons en donner l'extension à des ordres supérieurs.

6.4.1 Formation de voies à l'ordre 2

La formation de voies à l'ordre 2 utilise le modèle de propagation contenu dans les vecteurs directionnels $\mathbf{d}(\boldsymbol{\alpha})$ constituant la variété d'antenne définie en 6.1.3. Dans cette méthode, on détermine les valeurs des paramètres de localisation $\boldsymbol{\alpha}$ en filtrant “spatialement” le signal. Les filtres spatiaux réalisent la projection, par un produit scalaire, de l'observation $\mathbf{y}(\nu)$ sur le vecteur directionnel $\mathbf{d}(\boldsymbol{\alpha})$. Chacun de ces produits scalaires constitue une voie. Dans la pratique, le nombre de voies est déterminé par le nombre de valeurs discrètes des paramètres de localisation. On mesure ensuite la puissance de sortie de chacune des voies.

Pour des ondes planes, reçues sur un réseau de capteurs équidistants, le paramètre de localisation est le déphasage entre capteurs que nous avons introduit en 6.1.3. Le filtre spatial compense ces déphasages avant de sommer les signaux recalés en phase. Cette approche revient à tourner artificiellement l'antenne, pour la rendre perpendiculaire au vecteur d'onde associé à

une source : on parle de rotation électronique de l'antenne. La sommation des signaux émis par une source située dans la direction de la voie amplifie la contribution de cette source. Les signaux issus des sources situées dans d'autres directions sont atténués par suite de la sommation de signaux déphasés.

La sortie de la formation de voies est la puissance de la sortie du filtre spatial. Les signaux étant centrés cette moyenne est le cumulatif d'ordre 2

$$f_{2,FV}(\boldsymbol{\alpha}) = \text{Cum}[\mathbf{d}^\dagger(\boldsymbol{\alpha}) \mathbf{y}, (\mathbf{d}^\dagger(\boldsymbol{\alpha}) \mathbf{y})^*]. \quad (6.8)$$

En développant cette relation on obtient

$$f_{2,FV}(\boldsymbol{\alpha}) = \mathbf{d}^\dagger(\boldsymbol{\alpha}) S_{\mathbf{y}(1)}^{(1)} \mathbf{d}(\boldsymbol{\alpha}). \quad (6.9)$$

La fonction utilisée pour la formation de voies est une forme quadratique construite sur la matrice spectrale des observations $S_{\mathbf{y}(1)}^{(1)}$. Les observations sont en général formées des signaux provenant des sources (\mathbf{m}) et de bruit additif (\mathbf{b}). La relation (6.9) est alors

$$f_{2,FV}(\boldsymbol{\alpha}) = \mathbf{d}^\dagger(\boldsymbol{\alpha}) S_{\mathbf{m}(1)}^{(1)} \mathbf{d}(\boldsymbol{\alpha}) + \mathbf{d}^\dagger(\boldsymbol{\alpha}) S_{\mathbf{b}(1)}^{(1)} \mathbf{d}(\boldsymbol{\alpha}),$$

soit encore

$$f_{2,FV}(\boldsymbol{\alpha}) = \sum_p |\mathbf{d}^\dagger(\boldsymbol{\alpha}) \mathbf{a}_p|^2 \text{Cum}[s_p, s_p^*] + \mathbf{d}^\dagger(\boldsymbol{\alpha}) S_{\mathbf{b}(1)}^{(1)} \mathbf{d}(\boldsymbol{\alpha}).$$

Les vecteurs directionnels étant normalisés selon $\mathbf{d}(\boldsymbol{\alpha})^\dagger \mathbf{d}(\boldsymbol{\alpha}) = 1$, on obtient, pour un bruit blanc de covariance $S_{\mathbf{b}(1)}^{(1)} = P_b I$,

$$f_{2,FV}(\boldsymbol{\alpha}) = \mathbf{d}(\boldsymbol{\alpha})^\dagger S_{\mathbf{m}(1)}^{(1)} \mathbf{d}(\boldsymbol{\alpha}) + P_b.$$

De nombreux ouvrages (par exemple [110]) ont été consacrés à la formation de voies à l'ordre 2 particulièrement dans le cas d'ondes planes reçues sur un réseau rectiligne de capteurs équidistants. Nous ne détaillerons pas toutes les propriétés de cette méthode très largement utilisée dans de nombreux domaines d'application. Donnons quelques indications sur le pouvoir de résolution.

Le pouvoir de résolution est l'aptitude à séparer des sources. Il se caractérise par la largeur du lobe à 3dB. Pour une source isolée, de puissance unité, située dans la "direction" $\boldsymbol{\alpha}_0$ la matrice spectrale des observations non bruitées est $S_{\mathbf{m}(1)}^{(1)} = \mathbf{d}(\boldsymbol{\alpha}_0) \mathbf{d}(\boldsymbol{\alpha}_0)^\dagger$. La sortie de la formation de voies est alors

$$f_{2,FV}(\boldsymbol{\alpha}) = |\mathbf{d}(\boldsymbol{\alpha}_0) \mathbf{d}(\boldsymbol{\alpha})^\dagger|^2.$$

Le lobe de l'antenne à 3dB est le lieu des points vérifiant

$$f_{2,FV}(\boldsymbol{\alpha}_{3dB}) = f_{2,FV}(\boldsymbol{\alpha}_0)/2.$$

Pour une source à l'infini, arrivant de la direction $\theta = \alpha_0 = 0$ (voir le paragraphe 6.1.3), reçue sur un réseau rectiligne de K capteurs équidistants

$$f_{2,FV}(\boldsymbol{\alpha}) = \left| \frac{\sin K\varphi/2}{K \sin \varphi/2} \right|^2,$$

φ étant le déphasage entre capteurs. L'ouverture angulaire, en θ , du lobe à 3dB est $\Delta\theta_{3dB} = 0,88\lambda/D$ en radians et $\Delta\theta_{3dB} = 51\lambda/D$ en degrés, λ étant la longueur d'onde et D la longueur totale de l'antenne.

Pour obtenir un bon pouvoir de résolution par formation de voies il faut utiliser une antenne longue vis-à-vis de la longueur d'onde. Cette exigence est difficile, sinon impossible, à réaliser pour les ondes à grande longueur d'onde que l'on rencontre, en particulier, en acoustique et en prospection sismique. Nous verrons que la méthode du goniomètre permet d'accroître fortement le pouvoir de résolution.

6.4.2 Formation de voies à l'ordre 4

Dans la formation de voies à l'ordre 4, donnée dans [43] et [50], en sortie du filtre spatial, on remplace le cumulante d'ordre 2 estimant la puissance, par la quadricovariance, cumulante d'ordre 4 de la sortie. Comme nous l'avons indiqué on se place à l'ordre 4 car les cumulants d'ordre 3 sont en général nuls. On obtient

$$f_{4,FV}(\boldsymbol{\alpha}) = \text{Cum}[\mathbf{d}(\boldsymbol{\alpha})^\dagger \mathbf{y}, (\mathbf{d}(\boldsymbol{\alpha})^\dagger \mathbf{y})^*, \mathbf{d}(\boldsymbol{\alpha})^\dagger \mathbf{y}, (\mathbf{d}(\boldsymbol{\alpha})^\dagger \mathbf{y})^*], \quad (6.10)$$

soit

$$f_{4,FV}(\boldsymbol{\alpha}) = \sum_{i,j,k,l} d_i d_j^* d_k d_l^* \text{Cum}[y_i^* y_j y_k^* y_l] = \sum_{i,j,k,l} d_i d_j^* d_k d_l^* S_{\mathbf{y} \, jl}^{ik}.$$

Simplifions les notations en posant⁹ $\mathbf{d} = \mathbf{d}_{(1)}$ et $\mathbf{d}^\dagger = \mathbf{d}^{(1)}$. Introduisons la matrice D issue du produit de Kronecker (ou produit tensoriel) de ces deux vecteurs $\mathbf{d}_{(1)} \otimes \mathbf{d}^{(1)}$

$$D_i^j = \mathbf{d}_{(1)i} \mathbf{d}^{(1)j} = d_i d_j^*.$$

La relation (6.10) se réécrit

$$f_{4,FV}(\boldsymbol{\alpha}) = \sum_{ijkl} D_i^j D_k^l S_{\mathbf{y} \, jl}^{ik}.$$

La fonctionnelle de la formation de voies à l'ordre 4 est une forme quadratique construite sur la quadricovariance des observations dans l'espace engendré par le produit de Kronecker des vecteurs directionnels. On retrouve l'interprétation de l'extension des techniques d'ordre 2 à l'ordre 4 donnée en 6.3.1.

⁹Les lecteurs familiers des notations tensorielles reconnaîtront des tenseurs d'ordre 1 covariants et contravariants.

Le signal issu des sources et le bruit étant statistiquement indépendants, la quadricovariance des observations s'écrit

$$S_{\mathbf{y}^{(2)}} = S_{\mathbf{m}^{(2)}} + S_{\mathbf{b}^{(2)}}.$$

Tous calculs faits, en tenant compte de l'indépendance statistique des sources, la fonctionnelle de la formation de voies à l'ordre 4 s'écrit

$$f_{4,FV}(\boldsymbol{\alpha}) = \sum_p |\mathbf{d}(\boldsymbol{\alpha})^\dagger \mathbf{a}_p(\alpha_p)|^4 \text{Cum}[s_p, s_p, s_p^*, s_p^*] + \sum_{ijkl} D_i^j D_k^l S_{\mathbf{b}^{(2)}}^{ik}.$$

La formation de voies à l'ordre 4 présente deux avantages. Lorsque le bruit est gaussien, le terme de bruit qui dépend de la quadricovariance du bruit est nul. Cette annulation des termes de bruit est particulièrement utile lorsque le bruit n'est pas blanc. Dans ce cas, si l'on ne connaît pas sa covariance, le bruit peut affecter le traitement de manière imprévisible. Par ailleurs, comme nous l'avons montré en 6.3.3, le réseau virtuel engendré par la formation de voies à l'ordre 4 est plus étendu que le réseau réel. Cette technique possède donc un meilleur pouvoir de résolution que la formation de voies à l'ordre 2. L'accroissement du pouvoir de résolution sera illustré au paragraphe 6.6.

6.5 Le goniomètre

Le goniomètre, ou méthode MUSIC, introduit dans les techniques d'ordre 2 par [27] et [183], est fondé sur la décomposition de l'espace des observations en deux sous-espaces, le sous-espace signal et le sous-espace bruit. Nous avons montré en 6.3.2 que cette décomposition de l'espace des observations s'étend à l'ordre 4. Cette constatation est à la base de l'extension du goniomètre à l'ordre 4.

6.5.1 Le goniomètre à l'ordre 2

On utilise, comme dans la formation de voies, le modèle de propagation décrit par les vecteurs directionnels $\mathbf{d}(\boldsymbol{\alpha})$. Les vecteurs sources engendrent le sous-espace signal qui est orthogonal au sous-espace bruit. Lorsque le paramètre de localisation $\boldsymbol{\alpha}$ parcourt l'ensemble de ses valeurs possibles, les vecteurs directionnels engendrent la variété d'antenne, sous-ensemble \mathcal{V} de l'espace des observations. Les vecteurs sources appartiennent à l'intersection de \mathcal{V} et du sous-espace signal. Les valeurs des paramètres de localisation des sources sont donc contenues dans l'ensemble de valeurs de $\boldsymbol{\alpha}$ qui rendent le vecteur directionnel orthogonal au sous-espace bruit. Dans le cas général, les conditions d'orthogonalité peuvent conduire à des solutions parasites. Dans le cas d'ondes planes reçues sur un réseau rectiligne de capteurs équidistants, quand on connaît le nombre de sources, il n'y a pas de solutions parasites.

La mise en œuvre du goniomètre à l'ordre 2 est faite en deux étapes. Elle s'appuie sur une hypothèse fondamentale qui est la blancheur du bruit. La première étape passe par l'estimation de la matrice spectrale des observations définie au paragraphe 6.2.2. En diagonalisant cette matrice on détermine l'ensemble $\{\lambda_i, K - P < i < K\}$ des valeurs propres minimales. Les vecteurs propres $\{\mathbf{v}_i, K - P < i < K\}$ associés à ces valeurs propres minimales engendrent le sous-espace bruit. Les autres vecteurs propres $\{\mathbf{v}_j, 1 \leq j \leq P\}$ engendrent le sous-espace signal. Par suite des erreurs d'estimation les valeurs propres minimales n'ont pas des valeurs strictement égales. Il existe des tests statistiques, donnés par [193] et [196], permettant de déterminer l'ensemble des valeurs propres minimales.

Les vecteurs directionnels recherchés sont orthogonaux au sous-espace bruit. Les valeurs recherchées du paramètre de localisation $\boldsymbol{\alpha}$ sont celles qui rendent minimale la norme¹⁰

$$N(\boldsymbol{\alpha}) = \sum_{k=P+1}^K |\mathbf{d}^\dagger(\boldsymbol{\alpha})\mathbf{v}_k|^2 = 1 - \sum_{k=1}^P |\mathbf{d}^\dagger(\boldsymbol{\alpha})\mathbf{v}_k|^2,$$

de la projection des vecteurs directionnels sur le sous-espace bruit.

Dans le cas d'ondes planes reçues sur un réseau de capteurs, le paramètre de localisation est l'angle φ . Le goniomètre et l'algorithme MUSIC représentent la fonction

$$f_{2,G}(\varphi) = \frac{1}{N(\varphi)}.$$

Les directions des sources sont données par les valeurs de φ associées aux maximums de $f_{2,G}(\varphi)$.

Le goniomètre à l'ordre 2 appliqué à un réseau de K capteurs peut localiser au maximum $K - 1$ sources. En effet, pour ce nombre maximal de sources, le sous-espace bruit est réduit à la dimension 1 qui est sa dimension minimale.

L'intérêt principal du goniomètre est l'accroissement du pouvoir de résolution par rapport à la formation de voies. Pour un réseau rectiligne de capteurs équidistants recevant des ondes planes le goniomètre donne les directions des sources sans biais lorsque le bruit additif est blanc. Les directions estimées sont dispersées par suite des erreurs d'estimation de la matrice spectrale des observations. De nombreux travaux ont présenté la variance de cette estimation [110].

6.5.2 Le goniomètre à l'ordre 4

Le goniomètre à l'ordre 4, comme le goniomètre à l'ordre 2, est fondé sur la décomposition de l'espace des observations en deux sous-espaces

¹⁰Les vecteurs propres \mathbf{v}_k de la matrice spectrale des observations sont normés.

complémentaires, le sous-espace signal d'ordre 4 et le sous-espace bruit d'ordre 4. A l'ordre 4, on se place dans les espaces engendrés par les produits de Kronecker des vecteurs observations et des vecteurs sources. Comme nous l'avons indiqué en 6.3.2, l'espace d'ordre 4 des observations a K^2 dimensions et le sous-espace signal d'ordre 4 a P dimensions. On peut donc, à cet ordre, localiser un nombre de sources supérieur au nombre de capteurs. Ce nombre de sources localisables dépend de la géométrie du réseau comme nous l'avons montré en 6.3.3.

Comme pour le cas de l'ordre 2 la mise en œuvre du goniomètre d'ordre 4 se fait en deux étapes : détermination des sous-espaces bruit et signal, et utilisation de l'orthogonalité des sources au sous-espace bruit.

Les matrices cumulantes La détermination des sous-espaces bruit et signal est faite à partir de la quadricovariance des observations

$$Q_{\mathbf{y}} = S_{\mathbf{y}}^{(2)} = \text{Cum}[\mathbf{y}, \mathbf{y}^\dagger, \mathbf{y}, \mathbf{y}^\dagger].$$

La quadricovariance joue à l'ordre 4 le rôle joué par la matrice spectrale à l'ordre 2. Pour obtenir les sous-espaces signal et bruit on doit diagonaliser la quadricovariance. Cette diagonalisation est faite dans l'espace à K^2 dimensions engendré par les produits de Kronecker des vecteurs d'observation. Une solution à ce problème de diagonalisation a été donnée par [45] en utilisant les matrices cumulantes.

L'application de la quadricovariance Q à une matrice M permet de construire ce que [45] dénomme une *matrice cumulante* dont la composante $Q_i^j(M)$, est donnée par

$$Q_i^j(M) = \sum_{lm} M_l^m Q_{mi}^{lj}.$$

On peut interpréter ces différentes opérations en termes de tenseurs. La quadricovariance est un tenseur d'ordre 4, deux fois covariant et deux fois contravariant; les matrices M , ou les produits de Kronecker, sont des tenseurs d'ordre 2, une fois covariant et une fois contravariant. La matrice cumulante est le tenseur d'ordre 2, une fois covariant et une fois contravariant obtenu par produit contracté du tenseur quadricovariance et du tenseur M .

Diagonalisation de la quadricovariance Etant donnée une quadricovariance Q , construite sur un espace d'observation de dimension K , on montre dans [45] qu'il existe K^2 matrices cumulantes qui diagonalisent la quadricovariance selon la relation

$$Q(\mathcal{M}(l)) = \lambda_l \mathcal{M}(l).$$

Pour obtenir les matrices $\mathcal{M}(l)$ diagonalisant la quadricovariance on peut, par exemple, ordonner leurs composantes dans un "long" vecteur en balayant

$\mathcal{M}(l)$ en lignes ou en colonnes. On est alors ramené à un problème classique de diagonalisation d'une matrice $K^2 \times K^2$, hermitienne par suite des symétries de la quadricovariance, et donc à valeurs propres réelles.

La réponse du goniomètre d'ordre 4 Les résultats précédents appliqués à la quadricovariance des observations $Q_{\mathbf{y}}$ permettent de déterminer le sous-espace bruit et le sous-espace signal. Le sous-espace bruit est engendré par les $K^2 - P$ matrices cumulantes $\mathcal{M}(l)$ associées aux $K^2 - P$ valeurs propres minimales. Le sous-espace signal, complémentaire du sous-espace bruit, est engendré par les matrices cumulantes associées aux P valeurs propres maximales.

La norme de la projection des produits de Kronecker des vecteurs directionnels $\mathbf{d}_{(1)}(\boldsymbol{\alpha}) \otimes \mathbf{d}^{(1)}(\boldsymbol{\alpha})$ sur le sous-espace bruit est

$$N_4(\boldsymbol{\alpha}) = 1 - \sum_{l=1}^P \left(\sum_{pq} \mathcal{M}(l)_p^q \mathbf{d}_q(\boldsymbol{\alpha}) \mathbf{d}^p(\boldsymbol{\alpha}) \right) = \sum_{l=P+1}^{K^2} \left(\sum_{pq} \mathcal{M}(l)_p^q \mathbf{d}_q(\boldsymbol{\alpha}) \mathbf{d}^p(\boldsymbol{\alpha}) \right).$$

On en déduit la réponse du goniomètre d'ordre 4

$$f_{4,G}(\boldsymbol{\alpha}) = \frac{1}{N_4(\boldsymbol{\alpha})}.$$

Les maximums de $f_{4,G}(\boldsymbol{\alpha})$ donnent les paramètres de localisation des sources. Comme à l'ordre 2, le goniomètre à l'ordre 4 a un fort pouvoir de résolution. Le goniomètre à l'ordre 4 est utile, sinon indispensable, dans deux situations :

- lorsque le nombre de sources est supérieur au nombre de capteurs,
- en présence d'un bruit gaussien de covariance inconnue il permet, en moyenne, d'éliminer la contribution du bruit. La présence de bruit, même gaussien, influencera cependant la qualité de la localisation en augmentant la variance de l'estimation des paramètres de localisation.

6.5.3 Utilisation des moments hybrides non linéaires

Nous avons introduit au paragraphe 6.3.4 la covariance non linéaire des observations

$$S_{\mathbf{y},g}^{(1)} = AS_{\mathbf{s},g}^{(1)}A^\dagger + K_1^0 S_{\mathbf{b}}^{(1)}.$$

La technique du goniomètre peut être mise en oeuvre à partir de cette covariance estimée comme l'ont montré [107] et [181].

Lorsque le bruit est blanc $S_{\mathbf{b}}^{(1)} = P_b I$, les vecteurs propres associés aux valeurs propres minimales de la covariance non linéaire des observations engendrent le sous-espace bruit. On peut donc utiliser $S_{\mathbf{y},g}^{(1)}$ comme la covariance classique utilisée à l'ordre 2.

Les techniques d'ordre 2 sont peu efficaces lorsque le bruit n'est pas blanc. Dans cette situation la covariance non linéaire permet de localiser les sources. On peut soit éliminer le bruit en choisissant une non-linéarité adaptée ou utiliser conjointement la covariance non linéaire et la covariance linéaire.

L'élimination de l'effet du bruit est obtenue en adaptant la forme de la non-linéarité. Deux possibilités sont proposées. En prenant la non-linéarité $g(y, y^*) - K_1^0 y$, la covariance non linéaire des observations

$$S_{\mathbf{y},g}^{(1)} = A(S_{\mathbf{s},g}^{(1)} - K_1^0 S_{\mathbf{s}}^{(1)})A^\dagger$$

ne fait plus intervenir le bruit et donne directement les sous-espaces signal et bruit. On élimine également le bruit en prenant la non-linéarité $g(y, y^*) = K_{g_2}^0 g_1(y, y^*) - K_{g_1}^0 g_2(y, y^*)$. On obtient

$$S_{\mathbf{y},g}^{(1)} = A \left(K_{g_2}^0 S_{\mathbf{s},g_1}^{(1)} - K_{g_1}^0 S_{\mathbf{s},g_2}^{(1)} \right) A^\dagger.$$

Enfin, on montre dans [107] que la diagonalisation conjointe de la matrice de covariance et d'une matrice de covariance non linéaire des observations donne les sous-espaces signal et bruit.

6.6 Illustration

Dans cette illustration, nous reprenons l'exemple de l'antenne rectiligne à capteurs équidistants. Nous simulons une antenne de $K = 6$ capteurs, espacés de $\lambda/3$, λ étant la longueur d'onde. L'antenne reçoit 12 sources dont les directions d'arrivée (θ en degrés), les puissances (P_s) et les kurtosis ($\mathcal{K}_{s(4)}$), sont donnés dans le tableau (6.1).

Le bruit, indépendant des sources, est spatialement blanc jusqu'à l'ordre 4, et est supposé gaussien. La puissance du bruit est $P_b = 0.1$. Les cumulants d'ordre 2 et 4 des observations sont calculés exactement, de sorte que tout se passe comme si nous avions observé les données sur une durée infinie.

On a représenté sur la figure (6.4) les réponses obtenues pour la formation de voies et le goniomètre aux ordres 2 et 4. L'antenne utilisée a une longueur

θ	-50	-10	-5	0	5	15	20	25	35	40	60	80
P_s	5	1	2	2	1	.1	.01	.01	.01	.01	.01	.01
$\mathcal{K}_{s(4)}$	-2	0.2	-1.2	-1	-1.4	-1.5	-1.6	0.7	-2	-2	-0.5	-2

Tableau 6.1: Description des sources utilisées pour l'illustration de traitement d'antenne. θ est la direction dans laquelle se trouve une source de puissance P_s et de kurtosis $\mathcal{K}_{s(4)}$.

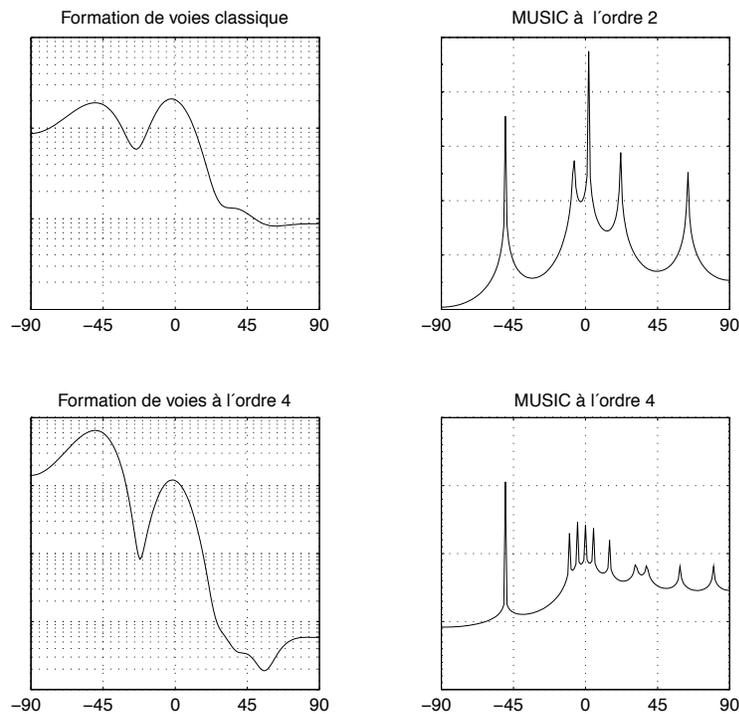


Figure 6.4: Réponses de la formation de voies et du goniomètre. Les panneaux du haut concerne l'ordre 2 et ceux du bas l'ordre 4. La réponse de la formation de voies est en échelle logarithmique.

totale de $5\lambda/3$. Pour la formation de voies à l'ordre 2 l'ouverture du lobe principal est approximativement de 31 degrés. Les sources ne sont pas localisées. On distingue seulement deux zones énergétiques centrées sur la direction 0° et sur la direction -50° . La formation de voies à l'ordre 4, issue d'une "antenne virtuelle" dont la longueur totale est $10\lambda/3$ comme le montre la figure (6.3), a un lobe central plus étroit mais elle ne sépare toujours pas les sources. On voit que le goniomètre à l'ordre 2 a un pouvoir de résolution bien supérieur à celui de la formation de voies. Le goniomètre à l'ordre 2, comme nous l'avons indiqué, ne peut localiser que cinq $(K - 1)$ sources. Il fait apparaître cinq directions centrées sur les zones contenant les sources. La source isolée à l'azimut -50° est bien localisée. Les directions des autres sources sont repérées mais les sources individuelles ne sont pas localisées. Le goniomètre à l'ordre 4, comme nous l'avons indiqué en 6.3.3, peut séparer onze sources. On voit qu'il indique dix directions qui s'approchent des directions des sources.

Des expériences complémentaires, non présentées ici, ont montré que lorsque le kurtosis du bruit augmente en valeur absolue, de moins en moins de sources sont détectées par le goniomètre à l'ordre 4. Il en est de même pour

le goniomètre à l'ordre 2 lorsque la puissance du bruit augmente. Cependant, la puissance du bruit a moins d'effet sur le goniomètre à l'ordre 4 que son kurtosis ; dans les performances asymptotiques que nous observons ici, cette puissance n'intervient pas.

6.7 Conclusion

Comme nous l'avons montré dans ce chapitre, le traitement d'antenne aux ordres supérieurs étend les techniques classiques de traitement développées à l'ordre 2. Nous avons présenté l'utilisation des statistiques d'ordre 4. Sur les mêmes principes, on pourrait étendre le traitement d'antenne à d'autres ordres.

Les fondements des principales méthodes de traitement d'antenne d'ordre supérieur ont été données. De nombreuses questions concernant la mise en œuvre et les performances restent encore ouvertes.

A l'ordre 2, on a développé des méthodes adaptatives dans lesquelles le filtre spatial, introduit en traitement d'antenne est déterminé en fonction des données observées. Ces techniques trouvent leur origine dans le filtre de Wiener adaptatif et dans le filtre de Capon. Nous pensons que des résultats sont à attendre dans ce domaine¹¹.

¹¹Citons la référence [72] qui étend le filtre de Wiener aux ordres supérieurs.

CHAPITRE 7

Séparation de sources

La séparation de sources concerne les techniques de traitement du signal mises en œuvre pour retrouver plusieurs composantes élémentaires (les sources) à partir de combinaisons linéaires de ces sources reçues sur plusieurs capteurs.

La séparation de sources se rencontre dans de nombreux domaines. En astronomie ou en astrophysique, le télescope ou le radio-télescope observent simultanément plusieurs étoiles dont on veut séparer les contributions. En acoustique sous-marine, l'antenne du sonar reçoit simultanément les signaux acoustiques émis par plusieurs bâtiments et perturbés par le bruit acoustique ambiant. En prospection sismique les échos renvoyés par les surfaces de séparation entre les couches du sous-sol peuvent interférer. Dans un atelier, les vibrations des différentes machines se confondent. Dans les communications par la parole, les situations de mélange sont nombreuses...

Dans ces divers domaines, on s'intéresse à un objet (étoile, bâtiment, machine vibrante...) qui émet des signaux (optiques, électromagnétiques, acoustiques, vibratoires). Après propagation ces signaux sont reçus sur un réseau de capteurs. A partir des signaux reçus, on veut retrouver les signaux émis par chacune des sources.

7.1 Modélisation

L'objet de ce paragraphe est de préciser les modèles de mélanges de sources sur lesquels nous allons travailler. Ces modèles ont déjà été présentés dans le chapitre 6, mais il paraît important de les discuter à nouveau dans le contexte plus général de la séparation de sources. De plus, des remarques fondamentales sur l'identifiabilité des paramètres des modèles sont effectuées.

7.1.1 Modèle

En séparation de sources, la situation physique est semblable à celle qui a été présentée au chapitre 6 pour le traitement d'antenne. Les signaux émis, ou modifiés, par P sources sont reçus sur un réseau de K capteurs. La propagation

entre les sources et les capteurs est représentée par un filtrage à P entrées, les signaux sources, et K sorties, les signaux reçus sur les capteurs. On admet que ce filtrage, introduit par la propagation, est linéaire et stationnaire. Comme nous l'avons indiqué au chapitre 6, les signaux $y_i(t)$ reçus sur les capteurs s'écrivent en fonction des signaux $s_p(t)$ émis par les P sources selon

$$y_i(t) = \sum_{p=1}^P \int a_{ip}(t - \tau) s_p(\tau) d\tau + b_i(t), \quad \forall i = 1, \dots, K. \quad (7.1)$$

Dans cette relation, $a_{ip}(t)$ représente la réponse impulsionnelle du filtre décrivant la propagation entre la source p et le capteur i (fonction de Green). Le terme $b_i(t)$ désigne un bruit additif.

Dans le domaine spectral, cette expression devient

$$Y_i(\nu) = \sum_{p=1}^P A_{ip}(\nu) S_p(\nu) + B(\nu), \quad \forall i = 1, \dots, K,$$

que l'on conviendra d'écrire de manière plus compacte sous la forme

$$\mathbf{y}(\nu) = A(\nu) \mathbf{s}(\nu) + \mathbf{b}(\nu). \quad (7.2)$$

La matrice $A(\nu)$, de dimension $K \times P$, rassemble toutes les fonctions de transfert entre les P sources et les K capteurs.

Nous distinguerons trois types de situations selon les conditions de propagation et selon le domaine, temporel ou spectral, dans lequel sont décrits les signaux.

1. *Le mélange instantané* : dans la modélisation en temps, si on suppose que la propagation n'introduit pas de retard, le signal reçu sur le capteur i s'écrit

$$y_i(t) = \sum_{p=1}^P a_{ip} s_p(t) + b_i(t).$$

En rassemblant les poids affectés aux différents trajets entre les sources et les capteurs dans la matrice A , en introduisant les vecteurs des sources $\mathbf{s}(t)$, du bruit $\mathbf{b}(t)$ et des signaux reçus $\mathbf{y}(t)$, le mélange instantané s'écrit

$$\mathbf{y}(t) = A\mathbf{s}(t) + \mathbf{b}(t). \quad (7.3)$$

2. *Le mélange spectral* est décrit par la relation (7.2).
3. *Le mélange convolutif* correspond au cas général décrit par l'équation (7.1).

Le mélange instantané et le mélange spectral sont formellement identiques puisque décrits par un produit matriciel. La différence essentielle entre ces deux modèles vient du fait que dans le mélange instantané, les signaux sont en général à valeurs réelles, alors que dans le mélange spectral ils sont à valeurs complexes.

Les algorithmes permettant de séparer le mélange spectral peuvent être utilisés pour le mélange instantané. L'inverse n'est pas vrai, et certains algorithmes ne s'appliqueront qu'aux mélanges instantanés. En raison de leur similarité de forme, on convient cependant de rassembler ces deux modèles sous la dénomination de *mélanges simples*.

La séparation de mélanges convolutifs est encore un sujet en cours de développement. Nous présenterons les solutions proposées qui se situent à la frontière des domaines déjà explorés.

Après avoir rappelé les résultats connus sur la séparation de sources à l'ordre 2, nous présenterons les techniques faisant appel aux statistiques d'ordre supérieur pour la séparation de mélanges simples et de mélanges convolutifs. Auparavant, en nous limitant au modèle simple, intéressons nous à l'identifiabilité qui nous conduira à préciser quels sont les paramètres du modèle que l'on peut déterminer et quels sont les paramètres non accessibles à la mesure.

7.1.2 Identifiabilité

Dans le modèle simple, le mélange des sources résulte du produit du vecteur des sources \mathbf{s} par la matrice A . Cette matrice représente les poids affectés aux différents trajets dans le mélange instantané, ou les fonctions de transfert des filtres liant les sources aux capteurs dans le mélange spectral. Le mélange non bruité s'écrit

$$\mathbf{m} = A \mathbf{s}. \quad (7.4)$$

Ce modèle contient des éléments fondamentalement non identifiables et des éléments non identifiables seulement à l'ordre 2.

Non-identifiabilité intrinsèque La forme de la relation (7.4) n'est pas modifiée si on multiplie une quelconque des composantes du vecteur des sources \mathbf{s} par une constante, et si l'on divise la colonne correspondante de A par la même constante.

Dans le mélange instantané, qui met en jeu des variables à valeurs réelles, cette indétermination porte sur l'amplitude, ou sur la puissance des sources, qui est donc non identifiable.

Dans le mélange spectral, la constante multiplicative arbitraire est un nombre complexe qui modifie l'amplitude et la phase des sources. Dans ce cas, l'indétermination porte sur l'amplitude et sur la phase des sources. L'amplitude

(ou la puissance) et la phase des sources ne sont donc pas identifiables. Cette indétermination s'applique pour chaque valeur de la fréquence. Dans le modèle spectral, les sources sont donc définies à un filtrage linéaire stationnaire près.

La connaissance des sources à un filtrage près était *a priori* prévisible. Si l'on remplace chaque source par une filtrée d'elle-même, on conserve l'indépendance des sources qui est l'hypothèse fondamentale permettant de réaliser la séparation.

Les sources sont connues à un filtrage près; on peut cependant déterminer l'effet produit par chacune d'elles sur chacun des capteurs. La matrice séparatrice A^- , pseudo-inverse de la matrice de mélange A , étant fixée à une constante multiplicative près on peut toujours imposer à l'un des coefficients des vecteurs colonnes de la matrice A^- d'être égal à 1. Avec cette normalisation, la séparation donne la composante de chaque source sur le capteur correspondant à la composante unité de la matrice A^- . Après séparation on pourra ainsi reconstituer le signal qu'aurait reçu un capteur si une seule source avait été présente.

Une autre indétermination intrinsèque au problème est celle de l'ordre des sources. En effet, on peut permuter deux composantes du vecteur source \mathbf{s} sans modifier le modèle (7.4), pourvu que l'on permute les colonnes correspondantes de A . Il ne faudra donc pas être surpris de constater, dans les expérimentations, que les sources estimées n'apparaissent pas toujours dans le même ordre que les sources originales.

Non-identifiabilité à l'ordre 2 A l'ordre 2, pour un modèle simple, les techniques classiques utilisent les autocorrélations et les intercorrélations des observations. Ces grandeurs constituent la matrice de covariance, calculée au retard nul, pour les mélanges instantanés, et la matrice spectrale pour les mélanges spectraux.

Normalisons la matrice de mélange A de manière à ce que les sources aient une puissance unité¹. La matrice de covariance des sources est alors la matrice identité $C_{\mathbf{s}(1)}^{(1)} = \text{Cum}[\mathbf{s}, \mathbf{s}^\dagger] = I$. La matrice de mélange A s'écrit (décomposition en valeurs singulières ou SVD)

$$A = U\Lambda^{1/2}V^\dagger. \quad (7.5)$$

Dans cette relation, U et V sont des matrices rectangulaires à P colonnes (le nombre de sources) et à K lignes (le nombre de capteurs), telles que² $UU^\dagger = VV^\dagger = I_K$ et $U^\dagger U = V^\dagger V = I_P$. Λ est une matrice diagonale contenant les P valeurs propres réelles non nulles de AA^\dagger .

La matrice de covariance de la partie non bruitée des observations est

$$C_{\mathbf{m}(1)}^{(1)} = \text{Cum}[\mathbf{m}, \mathbf{m}^\dagger] = AC_{\mathbf{s}(1)}^{(1)}A^\dagger.$$

¹Ceci n'est pas une contrainte limitative puisque la puissance des sources est arbitraire.

² I_Q est la matrice identité de dimension $Q \times Q$.

En tenant compte du fait que la covariance des sources est égale à l'identité, et en remplaçant A par sa SVD (7.5), l'expression de la covariance des observations non bruitées devient

$$C_{\mathbf{m}(1)}^{(1)} = U\Lambda^{1/2}VV^\dagger\Lambda^{*1/2}U^\dagger = U\Lambda^{1/2}\Lambda^{1/2}U^\dagger = U\Lambda U^\dagger. \quad (7.6)$$

La matrice de covariance des observations permet d'estimer les matrices U et Λ . U est la matrice des vecteurs propres normalisés associés aux valeurs propres non nulles de la matrice de covariance des observations $C_{\mathbf{m}(1)}^{(1)}$. Λ est la matrice diagonale des valeurs propres non nulles de $C_{\mathbf{m}(1)}^{(1)}$. Ces valeurs propres sont réelles positives car la matrice de covariance des observations est hermitienne définie non négative. Les éléments diagonaux de $\Lambda^{1/2}$ sont les racines réelles positives des valeurs propres de $C_{\mathbf{m}(1)}^{(1)}$.

La matrice V a disparu de la relation (7.6); par conséquent, cette matrice et donc la matrice de mélange A ne sont pas identifiables à l'ordre 2 en utilisant seulement la matrice de covariance au retard nul pour le mélange instantané, ou la matrice spectrale pour le mélange spectral. Dans certaines conditions, on peut réaliser la séparation à l'ordre 2 de mélanges instantanés, en utilisant les matrices de covariance des observations pour des retards différents de 0.

7.1.3 Blanchiment spatial à l'ordre 2

En projetant le vecteur des observations $\mathbf{y} = \mathbf{m} + \mathbf{b}$ sur les vecteurs propres associés aux valeurs propres non nulles de la matrice de covariance des signaux utiles (les colonnes de U), et en normalisant ces projections, on obtient le vecteur à P composantes

$$\boldsymbol{\sigma} = \Lambda^{-1/2}U^\dagger \mathbf{y} \stackrel{\text{def}}{=} \mathbf{x} + \mathbf{n} = V^\dagger \mathbf{s} + \Lambda^{-1/2}U^\dagger \mathbf{b}. \quad (7.7)$$

Après projection, la matrice de covariance de la partie signal,

$$C_{\mathbf{x}(1)}^{(1)} = E[\mathbf{xx}^\dagger] = \Lambda^{-1/2}U^\dagger C_{\mathbf{m}(1)}^{(1)} U\Lambda^{-1/2} = V^\dagger C_{\mathbf{s}(1)}^{(1)} V = V^\dagger V = I,$$

est la matrice identité. L'étude à l'ordre 2 nous permet donc de blanchir³ le signal en construisant une observation standardisée, formée de composantes décorrélatées et de même puissance.

Le bruit est modifié par la projection. La matrice de covariance du bruit projeté est en effet

$$E[\mathbf{nn}^\dagger] = \Lambda^{-1/2}U^\dagger C_{\mathbf{b}(1)}^{(1)} U\Lambda^{-1/2}.$$

³Nous utilisons ici le terme blanchiment qui est plus courant en traitement du signal. En statistique, comme nous l'avons indiqué au chapitre 1, cette opération est dénommée standardisation.

Lorsque le bruit affectant les données observées est spatialement blanc, $C_{\mathbf{b}(1)}^{(1)} = P_b I$, le bruit projeté a pour matrice de covariance $C_{\mathbf{n}(1)}^{(1)} = P_b \Lambda^{-1}$. Les composantes du bruit projeté conservent la propriété de décorrélation mais elles n'ont plus la même puissance.

7.2 Séparation de sources à l'ordre 2

Dans ce paragraphe, nous examinons les solutions existant pour séparer des sources à l'aide des statistiques d'ordre 2.

7.2.1 Séparation de mélanges instantanés

Plusieurs auteurs (voir [24, 73] par exemple) ont montré que l'on peut séparer un mélange instantané de sources ayant des spectres différents en utilisant les matrices de covariance $C_{\mathbf{y}(1)}^{(1)}(\tau) = \text{Cum}[\mathbf{y}(t), \mathbf{y}^\dagger(t - \tau)]$ des signaux observés prises, pour différentes valeurs du retard.

Comme dans le paragraphe 7.1.2, nous admettons que les sources ont une puissance unité et nous écrivons la matrice de mélange sous la forme $A = U\Lambda V^\dagger$. La matrice de covariance des observations au retard nul $C_{\mathbf{y}(1)}^{(1)}(0) = U\Lambda U^\dagger$ donne les matrices U et Λ . Comme en 7.1.3, on en déduit les observations blanchies

$$\boldsymbol{\sigma}(t) = \Lambda^{-1/2} U^\dagger \mathbf{y}(t).$$

En supposant le bruit temporellement blanc, la matrice de covariance des observations blanchies est, au retard τ ,

$$C_{\boldsymbol{\sigma}(1)}^{(1)}(\tau) = V^\dagger C_{\mathbf{s}(1)}^{(1)}(\tau) V. \quad (7.8)$$

La matrice de covariance des sources $C_{\mathbf{s}(1)}^{(1)}(\tau)$ au retard τ est diagonale. La matrice V est donc la matrice unitaire qui diagonalise la matrice de covariance des observations blanchies. La détermination de V permet de terminer l'identification de la matrice de mélange A et donc de séparer les sources par

$$\mathbf{s}(t) = V \boldsymbol{\sigma}(t).$$

Pour que cette méthode de séparation soit applicable, il faut que la diagonalisation des matrices de covariance pour différents retards apporte des informations nouvelles par rapport à la diagonalisation de la matrice de covariance au retard nul. Cette condition est remplie si les fonctions de corrélation ou, ce qui est équivalent, les densités spectrales des sources, sont différentes. Si les sources ont des spectres de même forme (égaux à un facteur de proportionnalité près) la séparation n'est pas possible. Dans ce cas, la matrice de covariance des sources

est égale à la matrice identité I multipliée par la fonction de corrélation des sources $C_{s(2)}(\tau)$ – qui est la même pour toutes les sources –. La relation (7.8) devient

$$C_{\sigma(1)}^{(1)}(\tau) = C_{s(2)}(\tau)V^{\dagger}V = C_{s(2)}(\tau)I.$$

La matrice V n'intervient pas, elle n'est donc pas identifiable.

Si les sources ont des spectres différents, mais proches, la séparation à l'ordre 2 est théoriquement possible. Le résultat sera cependant très sensible aux erreurs. Une telle méthode de séparation, appliquée à des sources ayant des spectres voisins, sera très peu précise.

7.2.2 Séparation de mélanges spectraux

L'outil de base utilisé pour la séparation de mélanges spectraux à l'ordre 2 est la matrice spectrale $S_{\mathbf{y}(1)}^{(1)}(\nu)$, matrice de covariance des observations spectrales $\mathbf{y}(\nu)$. L'utilisation de la matrice spectrale en traitement d'antenne (chapitre 6) et en séparation de sources est issue des travaux de H. Mermoz [151].

Comme nous l'avons vu au chapitre 6, le nombre de valeurs propres dominantes de la matrice spectrale permet de déterminer le sous-espace signal. La dimension du sous-espace signal donne le nombre de sources. La connaissance du sous-espace signal est évidemment insuffisante pour déterminer les sources qui sont caractérisées par des vecteurs appartenant à ce sous-espace.

Pour réaliser la séparation de sources à l'ordre 2 on doit disposer de relations supplémentaires. Il n'existe pas de cadre général pour déterminer ces relations qui dépendent de chaque cas d'espèce. Comme on l'a vu au chapitre 6, on peut localiser et séparer des sources avec une antenne formée d'un réseau de capteurs lorsque la réponse et la position de ces derniers sont parfaitement connues. Si ce n'est pas le cas, il faut recourir aux statistiques d'ordre supérieur (cf. section 7.3).

7.2.3 Séparation de mélanges convolutifs

La séparation de mélanges convolutifs en utilisant les statistiques d'ordre 2 est possible lorsque l'information est suffisamment redondante. On parle de diversité spatiale, temporelle, ou spectrale selon que les informations sont issues de l'exploration de l'espace, du temps ou du spectre. Des systèmes permettant la séparation à l'ordre 2 de mélanges convolutifs, que nous ne développerons pas ici, ont été décrits dans [95]. Notons que ce thème de recherche est un sujet d'actualité.

7.3 Séparation de mélanges simples

Les techniques utilisant les statistiques d'ordre 2 ne permettent pas de résoudre tous les problèmes de séparation de sources. Dans ces situations, l'utilisation des statistiques d'ordre supérieur à 2 amène des solutions. Ce paragraphe a pour objet de présenter la philosophie de quelques méthodes, et de montrer les techniques de séparation.

7.3.1 L'analyse en composantes indépendantes

La séparation à l'ordre 2 est essentiellement fondée sur la décorrélation des signaux sources. Cette décorrélation est une conséquence de l'indépendance des sources. Sauf dans le cas de signaux sources gaussiens, l'utilisation de la décorrélation n'exploite pas complètement l'indépendance qui est une propriété bien plus forte que la décorrélation. C'est ici qu'apparaissent les statistiques d'ordre supérieur en séparation de sources. En utilisant des statistiques d'ordre supérieur à 2, on exploite plus complètement l'indépendance des sources. On obtient ainsi des critères permettant de réaliser la séparation sans information supplémentaire. Ceci conduit à un nouveau concept : l'Analyse en Composantes Indépendantes (ACI), introduit par C. Jutten et J. Héroult dans [111], et développé dans [57]. Cette nouvelle approche vient compléter l'Analyse en Composantes Principales (ACP), fondée sur la seule décorrélation, qui est une méthode très utilisée en statistique.

7.3.2 Utilisation de l'indépendance statistique

Plusieurs approches utilisant l'indépendance statistique ont été développées. Les réseaux neuromimétiques donnent une solution. Les approches fondées sur les cumulants d'ordre supérieur à 2 utilisent la propriétés d'annulation des cumulants croisés construits sur des composantes indépendantes (chapitre 1 page 19), que l'on peut utiliser pour séparer des sources. On peut également s'appuyer sur le principe du maximum de vraisemblance qui est la méthode standard en statistique pour résoudre un problème d'estimation. Enfin, on peut fonder la séparation sur des distances entre lois, ce qui conduit aux contrastes.

Les réseaux neuromimétiques Une présentation détaillée de cette méthode peut être trouvée dans [102, 112]. Considérons un mélange instantané de deux sources non bruitées, $\mathbf{y}(t) = \mathbf{A}s(t)$, et cherchons à séparer les deux sources par le système bouclé décrit sur la figure (7.1). La sortie du système est $\mathbf{r}(t) = \mathbf{y}(t) - \mathbf{G}\mathbf{r}(t)$, \mathbf{G} étant la matrice séparatrice,

$$\mathbf{G} = \begin{pmatrix} 0 & g_{12} \\ g_{21} & 0 \end{pmatrix}.$$

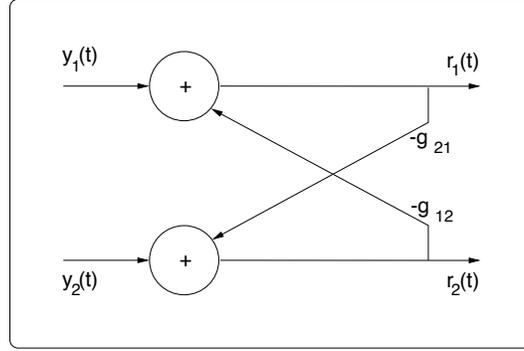


Figure 7.1: *Système bouclé pour la séparation de deux sources. On retranche à l'observation 1 la contribution de la source estimée 2, et à l'observation 2 la contribution de la source 1. Un critère d'adaptation adéquat des coefficients réalise la séparation.*

La sortie du système bouclé est

$$\mathbf{r}(t) = (I + G)^{-1} \mathbf{y}(t) = (I + G)^{-1} A \mathbf{s}(t). \quad (7.9)$$

En faisant $(I + G)^{-1} A = I$ on retrouve les sources. Cet exercice d'algèbre simple montre qu'il existe des solutions. Il ne donne pas de moyen pratique car on ne connaît pas la matrice de mélange A .

Le principe de résolution consiste à appliquer la méthode générale utilisée dans les systèmes bouclés que nous avons présentée au chapitre 3 : faire évoluer les coefficients du système, ici g_{12} et g_{21} , de manière à maximiser ou minimiser un critère.

Dans le cas plus général d'un mélange de P sources, le principe reste le même et est fondé sur le système $\mathbf{r}(t) = \mathbf{y}(t) - G\mathbf{r}(t)$, où la diagonale de la matrice G ne contient que des zéros. La composante n de la sortie s'écrit alors

$$r_n(t) = a_{nn}s_n(t) + \sum_{k \neq n} (a_{nk}s_k(t) - g_{nk}r_k(t)).$$

Ecrivons la solution sous la forme $r_k(t) = a_{kk}[s_k(t) + \delta s_k(t)]$ (on admet ici que l'on retrouve les sources à un facteur d'échelle près a_{kk} non identifiable). Quand $\delta s_k(t) \rightarrow 0$, on obtient

$$r_n(t) \simeq a_{nn}s_n(t) + \sum_{k \neq n} (a_{nk} - g_{nk}a_{kk})s_k(t).$$

Les sources étant indépendantes, la puissance de la sortie est

$$E[r_n^2(t)] = a_{nn}^2 E[s_n^2(t)] + \sum_{k \neq n} (a_{nk} - g_{nk}a_{kk})^2 E[s_k^2(t)].$$

Cette expression nous indique que pour obtenir une sortie $r_n(t)$ proportionnelle à la source n il faut minimiser la puissance de la sortie. Ceci nous fournit un critère d'adaptation : on fera varier les coefficients g_{ij} de manière à minimiser la puissance de la sortie. Après quelques calculs, développés dans [112], on obtient pour la règle d'adaptation des coefficients de la matrice séparatrice

$$\frac{dg_{ik}}{dt} = \mu s_i(t)s_k(t) \quad \forall i \neq k,$$

où μ est le pas d'adaptation, qui fixe la vitesse de convergence.

Avec cette règle d'adaptation, compte tenu de l'effet de mémoire inhérent à ces systèmes bouclés, lorsque l'on atteint l'état stationnaire $dg_{ik}/dt = 0$, on a $E[s_i(t)s_k(t)] = 0$. Le critère d'adaptation permet donc d'obtenir des sorties décorréelées. Nous avons vu que cette condition n'était pas suffisante pour séparer les sources. Il y aura donc de nombreux points d'équilibre parasites : cette solution est donc inacceptable. C'est ici qu'intervient l'apport décisif des systèmes neuromimétiques. Une propriété générale de ces systèmes indique que pour résoudre des problèmes de classification, en utilisant des systèmes bouclés, il est nécessaire d'utiliser une règle d'adaptation non linéaire. Les auteurs de [103] ont proposé comme règle d'adaptation

$$\frac{dg_{ik}}{dt} = \mu f[s_i(t)]h[s_k(t)] \quad \forall i \neq k,$$

les fonctions f et h étant des fonctions non linéaires. Avec ce type de règle, la séparation est possible. Il peut exister selon la forme de f et de h quelques solutions parasites.

On peut comprendre simplement pourquoi une règle d'adaptation non linéaire permet la séparation. Quand le régime stationnaire est atteint, le système conduit à $E[f\{s_i(t)\}h\{s_k(t)\}] = 0$. Les fonctions non linéaires f et h font que cette moyenne est fonction non seulement des moments du second ordre mais aussi de moments d'ordre supérieur. On a donc une règle qui annule simultanément des cumulants de divers ordres et utilise donc plus complètement l'indépendance des sources que ne le fait la décorrélation. Nous verrons plus loin que ce type de règle est justifié par les résultats de l'étude selon le maximum de vraisemblance. Cette procédure établit un parallèle avec les techniques d'ordre 2. La séparation est toujours fondée sur un critère de décorrélation, mais ici il s'applique à des transformées non linéaires des signaux. Enfin, on retrouve des résultats analogues à ceux que l'on avait obtenus en déconvolution aveugle (chapitre 3). L'identification du filtre inverse nécessitait la mise en œuvre de potentiels non quadratiques, et donc de règles d'adaptation non linéaires. L'unification de ces techniques est une question ouverte.

Les cumulants croisés et la quadricovariance A partir d'une analyse heuristique du problème, fondée sur la propriété fondamentale d'annulation

des cumulants croisés, il a été proposé dans [42, 54, 124] des approches faisant explicitement appel à l'annulation des cumulants croisés d'ordre 4. Ces approches ont été mises en forme dans [45, 57], et étendent les procédures de diagonalisation qui s'appliquent habituellement aux statistiques d'ordre 2.

Deux approches principales ont été introduites, fondées sur les matrices cumulantes (introduites au chapitre 6) et sur la diagonalisation conjointe de matrices issues des tenseurs cumulants d'ordre 4.

Les cumulants d'ordre 4 du signal blanchi, $C_{\sigma_{il}}^{jk} = \text{Cum}[\sigma_i, \sigma_j^*, \sigma_k^*, \sigma_l]$ forment un tenseur d'ordre 4 dénommé *quadrivariance*⁴. Comme nous l'avons vu au chapitre 6, il est possible d'associer à toute matrice M la *matrice cumulante* $Q(M)$, dont les composantes sont

$$Q(M)_i^j = \sum_k \sum_l C_{\sigma_{il}}^{jk} M_k^l$$

On peut faire un rapprochement entre cette définition de la matrice cumulante et celle des tranches des multcorrélations et des multispectres qui ont été introduites au chapitre 3.

La relation (7.7, page 235) donne les composantes de l'observation blanchie en fonction du vecteur source

$$\boldsymbol{\sigma} = V^\dagger \mathbf{s} + \mathbf{n}.$$

A partir de cette relation, en tenant compte du fait que les sources sont indépendantes et en supposant le bruit gaussien, on obtient, après quelques calculs développés dans [45],

$$Q(M) = V\Psi V^\dagger,$$

Ψ étant la matrice diagonale

$$\Psi = \text{Diag}[\gamma_1 \mathbf{v}_1^\dagger M \mathbf{v}_1, \dots, \gamma_K \mathbf{v}_K^\dagger M \mathbf{v}_K], \quad (7.10)$$

où les $\gamma_i = \mathcal{K}_{s_i(4)}$ sont les kurtosis des sources, et où les vecteurs \mathbf{v}_i sont les colonnes de la matrice V^\dagger .

Cette relation donne la méthode de séparation : la matrice V est la matrice diagonalisant une matrice cumulante $Q(M)$. La relation (7.10) nous montre également, puisque les cumulants de sources gaussiennes sont nuls, que l'on ne peut séparer les sources que si une au plus est gaussienne. Pour éviter les ambiguïtés, il faut également que les termes diagonaux de la matrice Ψ soient différents. Cette condition d'unicité de la solution a été en partie levée par des amégagements ultérieurs [45].

⁴Nous considérons ici des signaux à valeurs réelles ou des signaux à valeurs complexes circulaires.

Lorsque le vecteur source est formé de variables aléatoires à valeurs complexes circulaires, l'expression de la matrice cumulante devient

$$Q(M) = E[(\boldsymbol{\sigma}^\dagger M \boldsymbol{\sigma}) \boldsymbol{\sigma} \boldsymbol{\sigma}^\dagger] - C_{\boldsymbol{\sigma}}^{(1)} M C_{\boldsymbol{\sigma}}^{(1)} - C_{\boldsymbol{\sigma}}^{(1)} \text{Trace} \left[M C_{\boldsymbol{\sigma}}^{(1)} \right],$$

et si l'on choisit comme matrice M la matrice identité

$$Q(I) = E[|\boldsymbol{\sigma}|^2 \boldsymbol{\sigma} \boldsymbol{\sigma}^\dagger] - (K + 1)I.$$

Dans ce cas, la matrice V est obtenue en diagonalisant la matrice $E[|\boldsymbol{\sigma}|^2 \boldsymbol{\sigma} \boldsymbol{\sigma}^\dagger]$.

Dans les techniques de séparation issues de la matrice cumulante, le choix de la matrice M est arbitraire. Il est bon d'utiliser plusieurs matrices M , la redondance de l'estimation permettant de diminuer les fluctuations statistiques de l'estimation de la matrice V .

Quand on utilise plusieurs matrices cumulantes, on doit les diagonaliser conjointement. Théoriquement, la matrice de diagonalisation conjointe est la matrice V . Cependant, par suite des erreurs d'estimations, les diverses matrices cumulantes conduiront à des résultats différents.

La matrice réalisant la diagonalisation conjointe est définie comme suit. Pour un faisceau de matrices $M_l, 1 \leq l \leq L$, la matrice diagonalisant conjointement les matrices du faisceau est la matrice V , telle que la norme du vecteur construit sur les éléments diagonaux de la matrice $\sum_l V^\dagger M_l V$ soit maximale.

On trouve dans [45] un moyen permettant d'optimiser le choix des matrices cumulantes. Les matrices cumulantes doivent être les K^2 matrices propres P_r de la quadricovariance. Ces matrices propres sont telles que⁵

$$Q(P_r) = \lambda_r P_r. \quad (7.11)$$

Ces résultats conduisent à une méthode pratique de séparation de sources dénommée JADE (pour "Joint Approximate Diagonalisation of Eigenmatrices") que nous présenterons à la section 7.3.3.

Le maximum de vraisemblance La séparation de sources par le maximum de vraisemblance que nous présentons ici a été développée par [168]. Le maximum de vraisemblance est une méthode très couramment utilisée en estimation statistique. Soit $p_{\mathbf{y}|A}(\mathbf{y})$ la densité de probabilité des signaux observés \mathbf{y} , conditionnellement aux paramètres mesurés A . La fonction de vraisemblance est la fonction des paramètres à mesurer $v(A, \mathbf{y}) = p_{\mathbf{y}|A}(\mathbf{y})$. L'estimateur du maximum de vraisemblance est la valeur des paramètres qui rend maximale la fonction de vraisemblance. On peut modifier la fonction de vraisemblance par une application monotone sans changer la valeur des estimateurs (selon le

⁵La définition de la quadricovariance $Q(M)$ montre qu'elle peut être considérée comme l'application d'un opérateur sur une matrice. L'équation (7.11) est alors une équation aux "matrices propres" de cet opérateur.

sens de variation de l'application on retiendra les arguments du maximum ou du minimum). On prend souvent la log-vraisemblance qui est le logarithme de la fonction de vraisemblance.

Dans un mélange simple, le signal observé est

$$\mathbf{y} = \mathbf{m} + \mathbf{b} = A\mathbf{s} + \mathbf{b}.$$

La densité de probabilité de \mathbf{m} conditionnellement à A est

$$p_{\mathbf{m}|A}(\mathbf{m}) = p_{\mathbf{s}}(A^{-1}\mathbf{m})|\text{Det}A|^{-1} = |\text{Det}A|^{-1} \prod_{i=1}^K p_{s_i}(e_i^T A^{-1}\mathbf{m})$$

e_i^T étant un vecteur dont toutes les composantes sont nulles sauf la composante i qui est égale à 1.

La densité de probabilité des observations conditionnées par la matrice A des paramètres à mesurer est, en supposant le bruit indépendant du signal,

$$p_{\mathbf{y}|A}(\mathbf{y}) = \int p_{\mathbf{s}}(A^{-1}\mathbf{u})p_{\mathbf{b}}(\mathbf{y} - \mathbf{u})|\text{Det}A|^{-1}d\mathbf{u}.$$

Cette relation conduit à la fonction de vraisemblance dans le cas général. Ici, pour simplifier, nous supposons le bruit faible, soit $p_{\mathbf{b}}(\mathbf{y} - \mathbf{u}) \simeq \delta(\mathbf{y} - \mathbf{u})$, ce qui donne pour la log-vraisemblance d'une observation

$$\begin{aligned} l_v(A, \mathbf{y}) &= \log \{p_{\mathbf{s}}(A^{-1}\mathbf{y})|\text{Det}A|^{-1}\} \\ &= \sum_i \log \{p_{s_i}(e_i^T A^{-1}\mathbf{y})\} - \log |\text{Det}A|, \end{aligned}$$

La log-vraisemblance de M observations issues de sources blanches (suite de vecteurs aléatoires indépendants⁶) est

$$l_v(A, \mathbf{y}) = \widehat{E}[\log p_{\mathbf{s}}(A^{-1}\mathbf{y})] = \sum_i \widehat{E}[\log \{p_{s_i}(e_i^T A^{-1}\mathbf{y})\}] - \log |\text{Det}A|, \quad (7.12)$$

\widehat{E} représentant une moyenne temporelle sur les échantillons de l'observation. En posant $\Phi_i = [\log p_{s_i}]'$ et en notant que $dA^{-1} = -A^{-1}dAA^{-1}$ et que $d \log |\text{Det}A| = \text{Trace}(A^{-1}dA)$, on obtient pour la différentielle de la log-vraisemblance

$$dl_v(A, \mathbf{y}) = \sum_i \widehat{E}[\Phi_i(e_i^T A^{-1}\mathbf{y})e_i^T(A^{-1}dA)A^{-1}\mathbf{y}] - \text{Trace}(A^{-1}dA),$$

ou, en notant que $e_i^T(A^{-1}dA)A^{-1}\mathbf{y} = \sum_l (A^{-1}dA)_{il}e_l^T A^{-1}\mathbf{y}$,

$$dl_v(A, \mathbf{y}) = \sum_i \widehat{E} \left[\sum_l \Phi_i(e_i^T A^{-1}\mathbf{y})e_l^T A^{-1}\mathbf{y} \Delta_{i,l} \right] - \sum_i \Delta_{i,i},$$

⁶Cette hypothèse simplificatrice peut être levée.

avec $\Delta_{i,l} = (A^{-1}dA)_{i,l}$.

L'estimateur du maximum de vraisemblance de A est donc la solution de

$$\widehat{E} [\Phi_i(e_i^T A^{-1} \mathbf{y}) e_i^T A^{-1} \mathbf{y}] = 0 \quad \forall i \neq l,$$

On montre dans [168], nous ne développerons pas ce point ici, que les termes diagonaux ($i = l$) donnent la puissance des sources.

Les termes $e_i^T A^{-1} \mathbf{y}$ sont les estimateurs des vecteurs sources, on peut les noter $\widehat{s}_i = e_i^T A^{-1} \mathbf{y}$. L'estimateur du maximum de vraisemblance est alors solution de

$$\widehat{E} [\Phi_i(\widehat{s}_i) \widehat{s}_i] = 0 \quad \forall i \neq l. \quad (7.13)$$

Ce résultat donne la forme de la transformation non linéaire qui doit être appliquée pour résoudre le problème de séparation selon le critère du maximum de vraisemblance. Il justifie la règle d'adaptation non linéaire utilisée dans les méthodes de séparation de sources par des systèmes neuromimétiques.

Les conditions (7.13) sont des conditions nécessaires mais non suffisantes. Elles peuvent donc conduire à des solutions parasites. L'existence de ces solutions parasites dépend de la forme des densités de probabilité des sources. Pour des sources gaussiennes les fonctions Φ_i sont linéaires. On aboutit à des conditions de décorrélation qui ne sont pas suffisantes pour réaliser la séparation.

Le point délicat de la méthode du maximum de vraisemblance est la nécessité de connaître la densité de probabilité des sources. En général nous ne disposerons pas d'une connaissance aussi précise. Pour conserver la philosophie de cette approche tout en relaxant la connaissance précise des lois, on a représenté dans [83] les densités de probabilité des sources par un développement de Gram-Charlier limité à l'ordre 4. Dans cette approche, les lois de probabilité sont supposées paires ; les cumulants d'ordre impair sont donc nuls et le terme d'ordre 4, dans le développement de Gram-Charlier, est le premier terme non gaussien, nécessaire pour rendre la séparation possible. Nous n'infligerons pas ce calcul fastidieux à nos lecteurs, il peuvent le trouver dans la référence citée. Il conduit à l'estimateur du maximum de vraisemblance approché de la matrice V . Cet estimateur est la valeur \widehat{V}_{MV} de V qui maximise la somme des carrés des cumulants d'ordre 4 standardisés. Ce résultat est intéressant car nous verrons plus loin qu'il est analogue à celui qui a été obtenu dans les méthodes fondées sur le contraste que nous allons présenter. Ce résultat est également à rapprocher de la méthode utilisant les matrices cumulantes : la diagonalisation de ces matrices est une manière de rendre maximum les termes diagonaux qui sont les autocumulants standardisés.

Les contrastes La séparation de sources est basée sur une hypothèse fondamentale : l'indépendance statistique des signaux émis par les sources. Il paraît naturel d'introduire un critère permettant de mesurer l'indépendance statistique d'un ensemble de variables aléatoires. On pourra alors fonder les

méthodes de séparation sur la maximisation de ce critère. Cette approche conduit aux *contrastes*, tels qu'ils ont été introduits dans [57].

Considérons un vecteur aléatoire \mathbf{x} dont la densité de probabilité est $p_{\mathbf{x}}$. Le but de l'opération est de trouver une transformation qui rende les composantes de ce vecteur indépendantes. Or, on a vu au paragraphe 7.1.2 que certains paramètres sont intrinsèquement non identifiables ; l'amplitude des sources est non identifiable et les sources sont déterminées à une permutation près. Il est donc logique que le contraste soit invariant par changement d'échelle et par permutation des composantes. Pour constituer un critère d'indépendance statistique, le contraste doit être extrême lorsque les composantes du vecteur \mathbf{x} sont indépendantes.

On appelle contraste une application $\Psi(p_{\mathbf{x}})$ de l'ensemble des densités de probabilité sur l'ensemble des nombres réels qui vérifie les propriétés suivantes :

1. $\Psi(p_{P\mathbf{x}}) = \Psi(p_{\mathbf{x}})$, P étant une permutation quelconque des composantes de \mathbf{x} .
2. $\Psi(p_{D\mathbf{x}}) = \Psi(p_{\mathbf{x}})$, D étant une matrice diagonale inversible quelconque.
3. si \mathbf{x} est un vecteur à composantes indépendantes et A une matrice inversible, alors $\Psi(p_{A\mathbf{x}}) \leq \Psi(p_{\mathbf{x}})$, $\forall A$.
4. si $\Psi(p_{A\mathbf{x}}) = \Psi(p_{\mathbf{x}})$, alors $A = DP$, où D est diagonale et P une permutation.

Un certain nombre de contrastes sont proposés dans [57], en se plaçant après blanchiment spatial. L'opposé de l'information mutuelle $\Psi_I(p_{\mathbf{x}}) = -I(p_{\mathbf{x}})$, définie au chapitre 1, est un contraste sur l'ensemble des vecteurs aléatoires ayant au plus une composante gaussienne. On peut interpréter ce contraste sous la forme $\Psi_I(p_{\mathbf{x}}) = -I(p_{x_g}) - J(p_x) + \sum_i J(p_{x_i})$ donnée au chapitre 1. Rappelons que J désigne la néguentropie. Le terme $I(p_{x_g})$ est maximisé par le blanchiment. Le terme $\Psi(p_{\mathbf{x}}) = -J(p_x) + \sum_i J(p_{x_i})$ ne dépend que des statistiques d'ordre supérieur à 2, comme nous l'avons montré au chapitre 1. Sa maximisation correspond donc à la seconde étape de la séparation. En utilisant un développement d'Edgeworth de la densité de probabilité (voir le chapitre 1), on montre que

$$\Psi(p_{\mathbf{x}}) \approx \frac{1}{48} \sum_p \left(\kappa_{\mathbf{x}pp} \right)^2.$$

Le contraste Ψ approche Ψ_I . On retrouve le critère donné par le maximum de vraisemblance approché.

Cette approche de la séparation de sources est analogue à celle que nous avons décrite au chapitre 3 pour la déconvolution aveugle. Dans cette situation

le critère utilisé était la distance à la loi gaussienne, développé par [65]. On applique ici les mêmes principes en introduisant une mesure de la distance à l'indépendance. Ces deux approches se rejoignent dans la définition de contrastes étendus pour la déconvolution aveugle multivariable [58].

En guise de conclusion provisoire En terminant ce panorama des principes sur lesquels reposent les méthodes de séparation de sources on voit apparaître, au-delà de la diversité des approches possibles, une profonde unité. Le principe de séparation est dans tous les cas l'indépendance statistique des sources. Mais les cheminements sont différents : règle d'adaptation non linéaire, annulation des cumulants, diagonalisation de la quadricovariance, maximum de vraisemblance, contraste. Chaque approche apporte ses contributions originales, mais il est agréable de constater qu'elles se retrouvent dans la formulation de la solution par la maximisation de la somme des carrés des "autocumulants" du signal blanchi. Cette maximisation de la somme des carrés des "autocumulants", appliquée aux cumulants standardisés, est équivalente à la minimisation de la somme des carrés des cumulants croisés, qui s'annule pour des variables indépendantes.

Nous avons présenté la trame des différentes approches en citant les articles de référence dans lesquels les lecteurs soucieux d'approfondissement trouveront le développement des résultats annoncés ici. Voyons maintenant les techniques mettant en œuvre ces approches.

7.3.3 Les systèmes de séparation

Les algorithmes développés pour la séparation de sources ont pour objectif de construire le séparateur qui est un système linéaire donnant en sortie des signaux proportionnels aux sources. Deux familles de méthodes ont été utilisées pour construire le séparateur – figure (7.2) –. Dans les méthodes directes, dites aussi *hors ligne*, un système annexe détermine les paramètres du séparateur à partir des signaux observés ; c'est une approche *par blocs*. Les systèmes *en ligne*, ou en boucle, utilisent la rétroaction. Un critère issu des signaux de sortie fournit la règle d'adaptation des paramètres du séparateur. Les systèmes adaptatifs et les réseaux neuromimétiques rentrent dans cette catégorie.

Nous allons décrire les principaux systèmes proposés en commençant par le cas simple de la séparation de deux sources qui permet de bien appréhender la structure des différents systèmes. Nous illustrerons ensuite ces techniques sur des exemples simulés et réels.

Une question importante se pose d'entrée de jeu aux concepteurs de systèmes. Nous avons vu qu'une partie de la séparation peut-être faite à l'ordre 2 par des techniques classiques déjà bien développées, parfois en ayant recours à des informations *a priori*. On peut donc construire des systèmes à deux

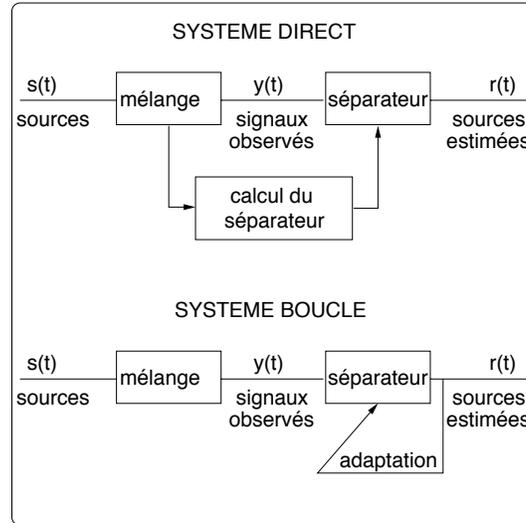


Figure 7.2: Les systèmes directs et les systèmes en boucle pour la séparation de sources.

étages enchaînant le blanchiment, réalisé à l'ordre 2, et la séparation finale, utilisant des statistiques d'ordre supérieur à 2. On peut aussi concevoir un système global. Notre choix se porte plutôt vers les systèmes découplant les deux étapes de traitement, parce qu'ils ont déjà fait leurs preuves. Un argument qui nous semble important est le fait qu'après blanchiment, les signaux sont normalisés (standardisés), ce qui permet d'éviter un certain nombre de problèmes numériques qui pourraient résulter (entre autres) de l'élévation des données à la puissance 4; ceci est surtout sensible lorsque les signaux ont des amplitudes très différentes.

Systèmes directs pour la séparation de deux sources Le blanchisseur a ici pour fonction de réaliser la projection des signaux sur les vecteurs propres de la matrice de covariance des observations et de normaliser la puissance des deux signaux projetés. La matrice de covariance des observations est, dans le cas de deux capteurs, une matrice 2×2 . On peut alors diagonaliser la matrice de covariance en utilisant des expressions analytiques.

Quelques remarques doivent être faites. On utilise souvent plus de deux capteurs, tout en restant dans le cas de deux sources. Dans cette situation, on fait appel pour la diagonalisation de la matrice de covariance des observations aux techniques classiques de diagonalisation (notamment la SVD). Les vecteurs propres sur lesquels on doit projeter les observations correspondent aux valeurs propres les plus grandes. Ici apparaît le rôle du bruit. Il n'est pas toujours très réaliste de supposer le bruit négligeable. Il existe toujours des parasites, ne seraient-ce que les bruits des capteurs. La méthode de blanchiment à l'ordre 2 n'est pas mise en défaut si le bruit est spatialement blanc, c'est-à-dire si ses

composantes sont décorréélées et de même puissance. Si cette condition n'est pas réalisée on doit préalablement blanchir le bruit. Il existe des méthodes prenant en compte un bruit non blanc [31]. Notons enfin que la projection d'un signal reçu (sur un grand nombre de capteurs) sur les deux vecteurs propres dominants réalise une réduction de bruit [91].

Après blanchiment spatial, on dispose d'un signal à deux composantes $\boldsymbol{\sigma} = V^\dagger \mathbf{s} + \mathbf{n}$. Si le bruit résiduel est gaussien il ne pose théoriquement pas de problème car il ne contribue pas aux cumulants d'ordre supérieur à 2. Il limite cependant les performances car il intervient sur les valeurs estimées des cumulants.

La séparation à un ordre supérieur revient à la détermination de la matrice unitaire 2×2 , V^\dagger . La forme de cette matrice est très simple dans le cas de deux sources.

Pour un mélange instantané réel, la matrice V^\dagger est une matrice de rotation de la forme

$$V^\dagger = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}$$

caractérisée par un seul paramètre, l'angle θ .

Pour un mélange spectral, la matrice V^\dagger est unitaire, et peut s'écrire

$$V^\dagger = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta e^{j\chi} & \cos \theta e^{j\chi} \end{pmatrix}$$

Elle est alors caractérisée par deux paramètres réels, l'angle θ et la phase, χ ou φ .

Rappelons que la séparation est fondée sur la minimisation d'un critère. La séparation consiste donc à trouver la valeur de l'angle, et de la phase dans le cas complexe, qui optimisent le critère. Il existe pour cela des méthodes algébriques, qui donnent directement la solution, et des méthodes de recherche exhaustive dans lesquelles on recherche numériquement la valeur du, ou des paramètres, optimisant le critère. Ces deux approches ont été appliquées.

Une question se pose également quant à l'ordre des statistiques que l'on doit considérer. La réponse actuelle est essentiellement heuristique. Comme en général les propriétés des estimateurs des cumulants se dégradent quand l'ordre augmente (voir le chapitre 1), on est conduit à se limiter à l'ordre le plus bas nécessaire. On retient en général l'ordre 4 car, dans la plupart des situations, les signaux ont des densités symétriques et donc des cumulants d'ordre 3 nuls. Cette règle n'est évidemment pas absolue et dans certains cas on pourra se limiter à l'ordre 3.

Pour le mélange instantané réel, et s'il n'y a pas de bruit, on peut obtenir l'angle θ en annulant les cumulants croisés d'ordre 4. On obtient (voir [54])

$$\tan 2\theta = \frac{2 \text{Cum}[\sigma_1, \sigma_1, \sigma_2, \sigma_2]}{\text{Cum}[\sigma_1, \sigma_2, \sigma_2, \sigma_2] - \text{Cum}[\sigma_1, \sigma_1, \sigma_1, \sigma_2]} = \frac{2\kappa_{\boldsymbol{\sigma}} 1122}{\kappa_{\boldsymbol{\sigma}} 1222 - \kappa_{\boldsymbol{\sigma}} 1112}. \quad (7.14)$$

Un certain nombre d'aménagements peuvent être réalisés, si le calcul du rapport ci-dessus est mal conditionné [55], ou s'il y a du bruit [57].

Pour le mélange spectral, il faut tenir compte de la circularité. Ce mélange est issu d'une transformation de Fourier et nous avons montré au chapitre 2 que les composantes spectrales prises à une même fréquence sont des variables aléatoires complexes circulaires⁷. On ne doit donc faire intervenir que les cumulants ayant autant de composantes complexes conjuguées que de composantes non conjuguées. Les autres cumulants, qui sont théoriquement nuls, auraient pour seul effet de détériorer les performances de l'estimation par suite des erreurs d'estimation.

La matrice V est caractérisée par les 2 angles θ et χ . On obtient facilement les relations suivantes entre les cumulants de l'observation blanchie, σ et les angles θ et χ

$$\begin{aligned}\kappa_{\sigma}^{12} &= \cos^2 \theta \sin^2 \theta (\gamma_1 + \gamma_2) \\ \kappa_{\sigma}^{22} &= \cos^2 \theta \sin^2 \theta (\gamma_1 + \gamma_2) e^{-2j\chi} \\ \kappa_{\sigma}^{11} &= \cos \theta \sin \theta (-\cos^2 \theta \gamma_1 + \sin^2 \theta \gamma_2) e^{-j\chi} \\ \kappa_{\sigma}^{22} &= \cos \theta \sin \theta (-\sin^2 \theta \gamma_1 + \cos^2 \theta \gamma_2) e^{-j\chi}\end{aligned}$$

γ_1 et γ_2 étant les cumulants standardisés des sources. Rappelons que $\kappa_{\sigma}^{kl} = \text{Cum}[\sigma_i, \sigma_j, \sigma_k^*, \sigma_l^*]$.

χ est donné par les phases de κ_{σ}^{22} , κ_{σ}^{12} ou κ_{σ}^{11} . On peut moyenner ces différentes estimations de χ . θ est donné par

$$\tan(2\theta) = \frac{2\kappa_{\sigma}^{12} e^{-j\chi}}{\kappa_{\sigma}^{12} - \kappa_{\sigma}^{11}}. \quad (7.15)$$

Dans le cas de signaux à valeurs réelles, $\chi = 0$, on retrouve la relation (7.14).

Dans les méthodes de recherche exhaustive on calcule le critère en fonction du ou des paramètres et on détermine les valeurs des paramètres qui maximisent le critère. On peut utiliser comme critère la somme des carrés des cumulants croisés d'ordre 4 pour des mélanges instantanés ou pour des mélanges spectraux [82, 122]. Il existe souvent des solutions algébriques explicites à ces problèmes d'optimisation [57].

⁷Les variables complexes ne sont pas toujours circulaires. Par exemple, les signaux de communications numériques sont représentés en bande de base par des variables complexes non circulaires pour un certain nombre de modulations très courantes. Le cas des N -PSK est traité dans le chapitre 1, et leur non circularité est utilisée dans une application en communication dans le chapitre 4.

La méthode JADE, reposant sur la diagonalisation de la matrice de quadricovariance, développée par J.F. Cardoso [45], est également applicable dans ce cas.

Systèmes directs pour la séparation de plus de deux sources La séparation de plus de deux sources est une généralisation de la situation précédente. On retrouve les mêmes principes de résolution. L'algorithme proposé dans [57] procède en deux étapes : blanchiment, suivi de la séparation à un ordre supérieur sous contrainte de blancheur. La première étape est classique : elle utilise la décomposition en valeurs singulières de la matrice de covariance des données. Dans la seconde étape on contraint la matrice V à être unitaire en la représentant par des matrices de Givens.

La matrice unitaire V , de dimension K , peut se décomposer (voir par exemple [93]) en un produit de $N = K(K - 1)/2$ matrices de rotation à 2 dimensions. Ces matrices composantes, de la forme

$$R(\theta_n) = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & & & & & & \\ \vdots & & \ddots & \vdots & \vdots & & & \vdots \\ 0 & & & \cos \theta_n & \sin \theta_n & & & 0 \\ 0 & & & -\sin \theta_n & \cos \theta_n & & & 0 \\ \vdots & & & \vdots & \vdots & \ddots & & \vdots \\ & & & & & & 1 & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 1 \end{pmatrix},$$

sont appelées matrices de Givens. On peut donc écrire

$$V = R(\theta_N) \dots R(\theta_1)$$

La détermination de V revient à la détermination des N angles de rotation de Givens, θ_n . Il est montré dans [57] que les tangentes de ces angles de Givens sont les racines d'un polynôme de degré K . Ce résultat généralise le cas de deux sources dans lequel on a donné avec (7.14) une expression explicite de $\tan 2\theta$. Cette procédure donne donc une technique directe de séparation d'un nombre quelconque de sources dont la seule difficulté est la recherche des racines d'un polynôme. On connaît des solutions analytiques jusqu'à l'ordre 5 ; au-delà, il faut faire appel à une résolution numérique.

L'algorithme présenté dans [45], dénommé JADE (Joint Approximate Diagonalization of Eigen-matrices) procède en quatre étapes.

- Dans la première étape on estime la matrice $\widehat{W} = \widehat{\Lambda}^{-1/2} \widehat{U}$ qui diagonalise l'estimée de la matrice de covariance des signaux observés, $C_{\mathbf{y}(1)}^{(1)}$.

- On en déduit la quadricovariance $C_{\hat{\boldsymbol{\sigma}}^{(2)}}^{(2)}$ de l'estimée du signal blanchi $\hat{\boldsymbol{\sigma}} = \widehat{W}\mathbf{y}$ et l'on calcule les matrices propres P_r de la quadricovariance (solutions de $\sum_{kl} C_{\hat{\boldsymbol{\sigma}}^{(2)}}^{il} P_j^i = \lambda_{i,j} P_j^i$) associées aux valeurs propres les plus élevées.
- On détermine la matrice unitaire \widehat{V} diagonalisant conjointement les matrices propres P_r .
- Finalement, on estime les sources par $\widehat{\mathbf{s}} = \widehat{V}\hat{\boldsymbol{\sigma}}$.

Dans l'article [45] les auteurs préconisent, pour la diagonalisation conjointe des matrices propres, la méthode de Jacobi, présentée dans [93], fondée sur les matrices de rotation de Givens.

Systèmes bouclés Le principe de ces systèmes est donné sur la figure (7.2, page 247). Le mélange observé $\mathbf{y}(t)$ est transformé par la matrice séparatrice G donnant la sortie $G\mathbf{y}(t)$ formée de composantes proportionnelles aux sources $\mathbf{s}(t)$. La matrice G est obtenue récursivement en optimisant un critère $v(\mathbf{r})$ calculé à partir de la sortie \mathbf{r} . Les composantes diagonales de G sont égales à 1, et les composantes non diagonales $g(t)$ obéissent à la règle d'adaptation

$$g(t) = g(t-1) + \mu \frac{\partial v\{\mathbf{r}(t-1)\}}{\partial g(t-1)}.$$

Différentes structures ont été proposées. On peut implanter directement le système adaptatif optimisant le critère. On peut aussi tirer parti du fait que le traitement peut se décomposer en deux étapes. La première étape réalise le blanchiment (à l'ordre 2), qui consiste à normaliser la puissance des composantes de l'observation et à les décorrélérer. La seconde étape est la séparation par des critères d'ordre supérieur à 2 sous la contrainte de blancheur. L'étape de blanchiment met en jeu des opérateurs classiques de décorrélation et de normalisation. Dans l'étape de séparation finale on respecte la contrainte de blancheur en imposant à la matrice de séparation d'être unitaire. Les matrices de Givens donnent une solution. Les algorithmes proposés utilisent les différentes approches possibles. Nous pensons que les algorithmes en deux étapes garantissent une meilleure stabilité numérique en opérant sur des données normalisées.

Un élément important, dans la conception de ces systèmes bouclés, est le critère. Comme nous l'avons vu au paragraphe 7.3.2, on peut utiliser différents critères qui sont dans un certain sens équivalents. Les systèmes fondés sur une caractérisation simple de l'indépendance utilisent l'annulation des cumulants croisés ou la minimisation de la somme des autocumulants. En général, comme les sources ont souvent des densités de probabilité paires, on utilise

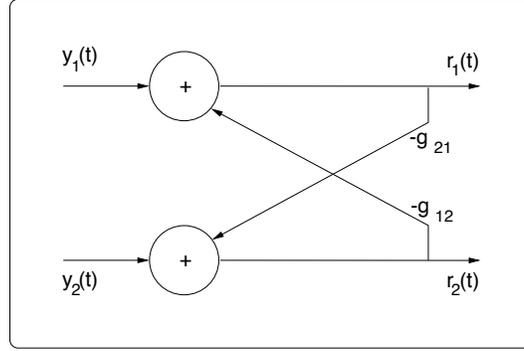


Figure 7.3: *Système bouclé pour la séparation de deux sources.*

des cumulants d'ordre 4. On peut aussi s'appuyer sur le principe du maximum de vraisemblance qui conduit à des règles d'adaptation non linéaires. Les contrastes donnent directement des critères d'adaptation.

Précisons les choses en donnant quelques solutions proposées. Pour simplifier la présentation nous nous limitons à la séparation de deux sources.

Le système proposé par Héroult et Jutten [112], schématisé sur la figure (7.3), s'écrit

$$\mathbf{r}(t) = \mathbf{y}(t - 1) - G(t - 1)\mathbf{r}(t - 1),$$

avec

$$G = \begin{pmatrix} 0 & g_{12} \\ g_{21} & 0 \end{pmatrix}.$$

Pour l'adaptation des coefficients non diagonaux, on doit utiliser des fonctions non linéaires des sorties. La règle d'adaptation utilisée dans les illustrations présentées au paragraphe 7.3.5 est

$$g_{ij}(t) = g_{ij}(t - 1) + \mu \frac{r_i^3(t - 1)r_j(t - 1)}{E[r_i^4(t - 1)]} \quad \text{pour } i, j = 1, 2 \text{ et } i \neq j.$$

Cette règle d'adaptation conduit, à l'équilibre, à l'annulation des cumulants $\kappa_{\mathbf{r}1222}$ et $\kappa_{\mathbf{r}1112}$. On normalise par la puissance 4 pour éviter les problèmes numériques car on n'a pas préalablement blanchi les signaux. Le calcul de cette puissance d'ordre 4 se fait également de façon récursive. Finalement, l'algorithme global dépend alors de deux constantes de temps : une régissant l'estimation du moment d'ordre 4 et l'autre régissant le calcul des g_{kl} .

Notons enfin que cette règle d'adaptation peut conduire à des solutions parasites car l'annulation de $\kappa_{\mathbf{r}1222}$ et $\kappa_{\mathbf{r}1112}$ est une condition nécessaire mais non suffisante. Si on blanchit les observations avant d'appliquer cet algorithme de séparation, la convergence est plus rapide. Ce blanchiment peut être fait de manière adaptative [53].

E. Moreau et O. Macchi dans [154] procèdent en deux étapes. La première étape réalise le blanchiment, la seconde complète la séparation. La règle

d'adaptation est fondée sur des contrastes. Les auteurs montrent que

$$J_{\alpha,\beta} = \alpha \sum_i E[r_i^4] - \frac{\beta}{2} \sum_{i \neq j} E[r_i^2 r_j^2],$$

α et β étant des paramètres, est un contraste si les sources ont des cumulants de même signe. La forme de ce contraste est intéressante car elle conduit à une expression simple de la dérivée du contraste par rapport aux paramètres de la matrice séparatrice. Ils proposent ensuite un système direct de séparation estimant les termes de la matrice de séparation sous la contrainte de blancheur. Ils donnent également un système adaptant les angles des matrices de Givens. Pour deux sources, la matrice de Givens est une simple matrice de rotation dont l'angle θ est adapté selon l'équation

$$\theta(t) = \theta(t-1) - \mu r_1(t-1) r_2(t-1) [r_1^2(t-1) - r_2^2(t-1)].$$

Une solution très élégante, valable pour un nombre quelconque de sources, est donnée par J.F. Cardoso et B. Laheld dans [47]. La sortie $\mathbf{r}(t)$ du système séparateur est reliée à l'entrée $\mathbf{y}(t)$ de ce système par la matrice de séparation $G(t)$, fonction de t dans un système adaptatif. La règle d'adaptation de la matrice séparatrice s'écrit de manière générale,

$$G(t) = G(t-1) - \mu F[\mathbf{r}(t-1)]G(t-1).$$

La définition du système de traitement est contenue dans la forme de la fonction matricielle des sorties $F[\mathbf{r}(t)]$. Les auteurs proposent la forme de $F[\mathbf{r}(t)]$ en s'appuyant sur la notion de contraste, sur la décomposition du traitement en deux étapes et sur la notion générale d'équivariance.

Un critère d'adaptation $v(\mathbf{r})$ est équivariant si, pour toute matrice inversible M , on a la relation

$$v(M\mathbf{r}) = Mv(\mathbf{r}).$$

Cette propriété introduite dans [106] conduit à des systèmes adaptatifs de séparation dont les performances sont indépendantes de la matrice de mélange. Cette robustesse du traitement homogénéise les propriétés des systèmes de séparation.

En développant ces deux principes, on montre dans [47] que la forme générale de la matrice d'adaptation est

$$F[\mathbf{r}] = \mathbf{r}\mathbf{r}^T - I + \mathbf{h}(\mathbf{r})\mathbf{r}^T - \mathbf{r}\mathbf{h}(\mathbf{r})^T.$$

où $\mathbf{h}(\mathbf{r})$ est un vecteur de fonctions non linéaires tel que $h_i(\mathbf{r}) = h_i(r_i)$. Écrivons la matrice d'adaptation $F = F_S + F_{AS}$, $F_S = \mathbf{r}\mathbf{r}^T - I$ étant la partie symétrique et $F_{AS} = \mathbf{h}(\mathbf{r})\mathbf{r}^T - \mathbf{r}\mathbf{h}(\mathbf{r})^T$ la partie antisymétrique. La partie symétrique,

quadratique en fonction des sorties, assure le blanchiment (ordre 2). Dans la partie antisymétrique, \mathbf{h} est une fonction non linéaire. Cette partie antisymétrique assure la séparation à un ordre supérieur à 2. En effet, quand la convergence est atteinte, on a

- $E[F_S] = 0$, conduisant à : $E[\mathbf{r}\mathbf{r}^T] = I$, réalisant ainsi le blanchiment,
- $E[F_{AS}] = 0$, donnant la condition d'annulation des cumulants d'ordre supérieur à 2 : $E[\mathbf{h}(\mathbf{r})\mathbf{r}^T - \mathbf{r}\mathbf{h}(\mathbf{r})^T] = 0$, grâce à la fonction non linéaire h .

7.3.4 Les performances

Les performances des systèmes de séparation dépendent de la méthode utilisée. Des résultats ont été obtenus, mais nous pensons que ce sujet mérite encore de retenir l'attention.

Dans le modèle de mélange présenté au paragraphe 7.1, les signaux observés \mathbf{y} sont obtenus en ajoutant du bruit \mathbf{b} à une transformée linéaire des sources \mathbf{s} par la matrice de mélange A

$$\mathbf{y} = A\mathbf{s} + \mathbf{b}.$$

Le résultat de la séparation est influencé par les erreurs commises sur l'estimation de la matrice de mélange et par les perturbations apportées par le bruit additif. Peu de résultats existent quant à l'effet du bruit. Pour étudier l'effet des erreurs d'estimation de la matrice de mélange on suppose le bruit faible et, dans une première approche, on néglige son effet.

Envisageons successivement les méthodes directes et les méthodes en boucle.

Performances des méthodes directes Les méthodes directes procèdent en deux étapes. Le traitement à l'ordre 2, conduisant au blanchiment, est suivi d'un traitement utilisant des statistiques d'ordre supérieur pour compléter la séparation. Les performances du blanchiment ont été étudiées, notamment par [127] et [190], dans le cadre des techniques classiques de traitement d'antenne. Présentons les résultats actuels sur les performances de la seconde étape qui est la partie du traitement mettant en œuvre des statistiques d'ordre supérieur.

Dans la séparation d'un mélange instantané de deux sources à valeurs réelles, la seconde étape du traitement nécessite l'estimation de l'angle θ qui caractérise la matrice de rotation V (voir le paragraphe 7.3.3). Les travaux de [100] donnent une analyse du biais et de la variance de l'estimateur de θ . L'estimateur n'est pas biaisé. Cette propriété est valide pour tous les critères d'estimation proposés indépendamment de la forme précise de la densité de probabilité des sources, pourvu qu'elles ne soient pas toutes les deux gaussiennes. Une démonstration géométrique est donnée. L'idée de base est que la

valeur de l'angle θ ne dépend pas de la forme précise de la densité de probabilité des sources mais seulement des axes de symétrie de la loi conjointe des deux sources. Tous les critères respectant cette symétrie sont donc, en moyenne, équivalents. Pour la variance de l'estimateur, dans le cas de sources i.i.d. observées sur N échantillons, la borne de Cramer-Rao de l'estimateur de θ donne

$$\text{Var}(\theta) \geq \frac{6}{N(\gamma_1^2 + \gamma_2^2)},$$

γ_1 et γ_2 étant les kurtosis des deux sources. On voit que la séparation n'est pas possible si les deux sources sont gaussiennes.

Dans [44] et [46], une étude générale des performances est donnée. En appelant \hat{A}^- la matrice pseudo-inverse de la matrice de mélange estimée, le résultat de la séparation est

$$\hat{\mathbf{s}} = \hat{A}^- \mathbf{A} \mathbf{x} + \hat{A}^- \mathbf{b}.$$

Les perturbations issues des erreurs d'estimation de la matrice de mélange sont décrites par la distance de la matrice $\hat{A}^- A$ à la matrice identité⁸. Les auteurs établissent une propriété importante. En présence de bruit faible les performances de la séparation sont indépendantes de la forme précise de la matrice de mélange. Elles ne dépendent que de la loi de probabilité des sources. Ce résultat est issu de la propriété d'équivariance⁹ des critères de séparation.

La qualité de la séparation est caractérisée par la diaphonie ou, selon la terminologie utilisée en télécommunication, par les interférences intersymboles. En première approximation, les termes diagonaux de la matrice $\hat{A}^- A$ sont égaux à 1. La diaphonie est alors chiffrée par la valeur des termes non diagonaux de la matrice $\hat{A}^- A$. La composante $P_{p,q}$ de la matrice $\hat{A}^- A$ décrit la pollution de la source estimée p par la source q . Les auteurs donnent des bornes inférieures pour ces termes de diaphonie. Contentons nous du résultat asymptotique, pour des sources i.i.d

$$\lim_{N \rightarrow \infty} P_{p,q} = \frac{1}{4N}, \quad (7.16)$$

N étant le nombre d'échantillons observés.

Pour compléter ces études, on pourra trouver dans [49] une analyse assez exhaustive de certaines performances obtenues à partir d'expérimentations informatiques.

Performances des méthodes en boucle On a vu au paragraphe 7.3.3 que la forme générale de la règle d'adaptation des composantes $g(n)$ de la matrice

⁸On admet ici qu'il n'y a pas eu de permutation des sources : on pourrait en tenir compte ce qui compliquerait inutilement la présentation.

⁹On rappelle qu'un critère $v(\hat{\mathbf{y}})$ est équivariant si $v(M\hat{\mathbf{y}}) = Mv(\hat{\mathbf{y}})$ pour toute matrice M inversible.

séparatrice est

$$g(t) = g(t-1) + \mu \frac{\partial v\{\mathbf{r}(t-1)\}}{\partial g(t-1)},$$

$v\{\mathbf{r}(t-1)\}$ étant la valeur de la règle d'adaptation et μ le pas d'adaptation.

Les propriétés des méthodes en boucle sont données par les points d'équilibre et par la vitesse de convergence. Les points d'équilibre stables sont les solutions vers lesquelles l'algorithme converge asymptotiquement. Ils dépendent de la règle d'adaptation. Une bonne règle d'adaptation admet comme points d'équilibre stables les solutions acceptables. La vitesse de convergence est commandée par le pas d'adaptation. De manière générale un pas d'adaptation faible conduit à de bonnes performances asymptotiques mais à une convergence lente et *vice-versa*. Il est possible, nous ne développerons pas ce point, de faire diminuer le pas d'adaptation quand on approche de l'équilibre de manière à diminuer les fluctuations au voisinage du point d'équilibre¹⁰.

Comme le montrent [47] et [76], pour que les points d'équilibre stables soient des solutions acceptables, il faut qu'une des sources, au plus, soit gaussienne, que les kurtosis des sources soient négatifs et que la règle d'adaptation soit non linéaire. Deux de ces règles étaient déjà apparues. La condition imposant une valeur négative des kurtosis est un critère supplémentaire qui restreint le domaine d'application de ces méthodes.

Quant à la variance asymptotique, on montre dans [44] que le taux de diaphonie asymptotique vérifie

$$\lim_{N \rightarrow \infty} P_{p,q} \geq \frac{\mu}{4}.$$

Cette relation est semblable au résultat donné par (7.16), μ étant, dans un certain sens, l'inverse du temps d'intégration.

7.3.5 Illustrations

Pour illustrer les techniques de séparation de sources nous donnons des exemples de séparation par les méthodes directes et par les méthodes adaptatives.

Méthodes directes Nous nous sommes limités à la séparation d'un mélange instantané de deux sources ayant des densités de probabilité paires. L'étape finale de séparation est réalisée en déterminant la rotation θ selon la relation (7.14, page 248).

Sur les figures suivantes, les planches (s_1) et (s_2) montrent les deux sources, les planches (y_1) et (y_2) les signaux observés, les planches (σ_1) et (σ_2) les signaux blanchis (traitement à l'ordre 2) et les planches (r_1) et (r_2) les sources reconstituées.

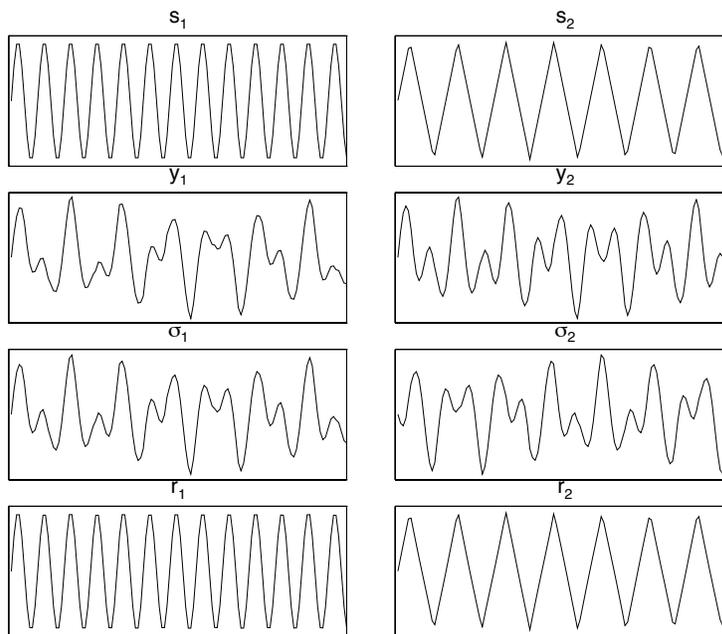


Figure 7.4: *Séparation d'une sinusoïde et d'une dent de scie par une méthode directe de séparation.*

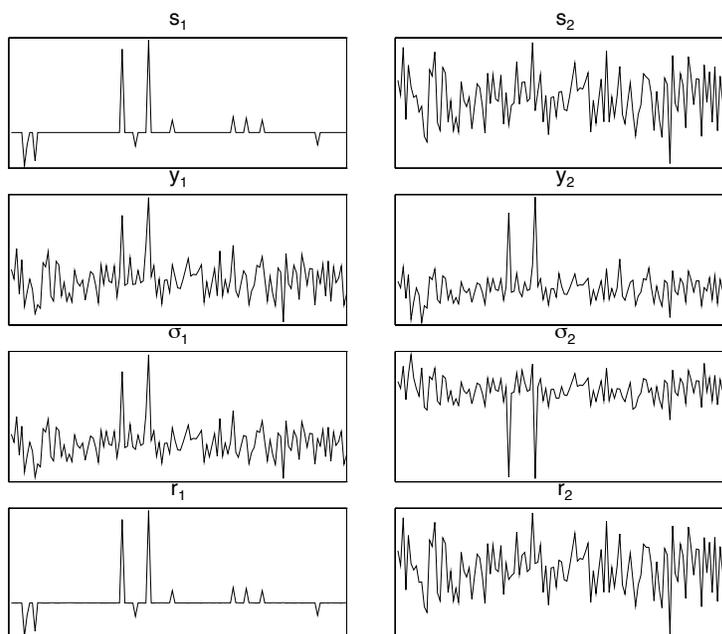


Figure 7.5: *Séparation d'un signal impulsif et d'un bruit gaussien par une méthode directe de séparation.*

Pour les figures (7.4) et (7.5), les sources ont la même puissance, et la matrice de mélange est

$$\begin{pmatrix} 1 & 0,8 \\ 0,8 & 1 \end{pmatrix}.$$

L'exemple de la figure (7.4) est un cas schématique dans lequel nous cherchons à séparer deux signaux simples : une sinusoïde pure et un signal en dent de scie. Bien que ces signaux déterministes soient assez peu respectueux des hypothèses qui s'appuyaient sur la notion d'indépendance au sens statistique, on constate que la séparation se fait bien (voir la discussion sur l'indépendance de signaux déterministes au paragraphe 6.1.2).

Le deuxième exemple, présenté figure (7.5), correspond à une situation rencontrée en prospection sismique. La première source est un signal aléatoire formé d'une suite d'impulsions éparses. Elle représente la réflectivité du sous-sol que l'on cherche à déterminer. La seconde source est un signal gaussien qui représente le bruit. Nous disposons de deux observations du mélange. On constate sur la figure (7.5) la possibilité de séparation de ces types de signaux.

Le troisième exemple décrit les possibilités de la séparation de sources en réduction de bruit (voir la discussion au chapitre 5, paragraphe 5.4). H. Mermoz [151] a montré que la réduction de bruit est possible à l'ordre 2 lorsque l'on possède un signal uniquement lié au bruit, dénommé la référence bruit seul. Dans cette technique, on identifie le filtre permettant de reconstruire le bruit polluant le signal à partir de la référence bruit seul. On retranche ensuite, au signal bruité, le bruit identifié à partir de la référence bruit seul. H. Mermoz a montré que cette technique de réduction de bruit devenait inefficace, et même nocive, lorsque la référence bruit seul contient une contribution, même faible, du signal. En reprenant ce problème sous l'angle de la séparation de sources on peut s'affranchir de la nécessité de disposer d'une référence bruit seul. Dans la figure (7.6), le signal, source 1, est formé de créneaux d'amplitude 0,5 et le bruit, source 2, est gaussien de puissance 0,09. Les signaux observés sont un mélange de signal et de bruit, la matrice de mélange étant

$$\begin{pmatrix} 1 & 0,8 \\ 0,8 & 1 \end{pmatrix}.$$

La mise en œuvre de la technique de séparation permet de retrouver le signal et le bruit et donc de réaliser une réduction de bruit sans disposer d'une référence bruit seul.

Méthodes adaptatives Sur les planches présentées, on fait figurer les sources (s_1 et s_2), les signaux observés (y_1 et y_2) et les sources reconstituées (r_1 et r_2); pour les techniques fonctionnant en deux étapes, les signaux blanchis apparaissent sur les planches (σ_1) et (σ_2).

¹⁰Rappelons l'image du randonneur développée au chapitre 3.

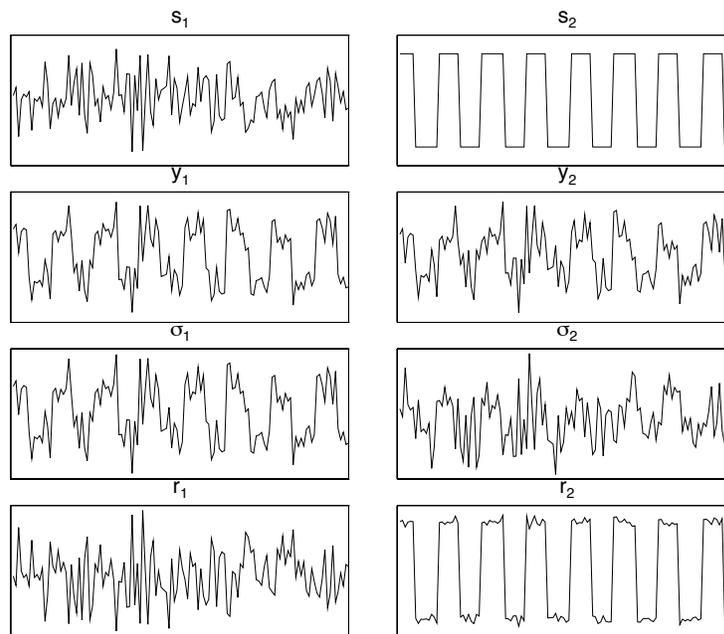


Figure 7.6: Réduction de bruit sur un signal carré par une méthode directe de séparation.

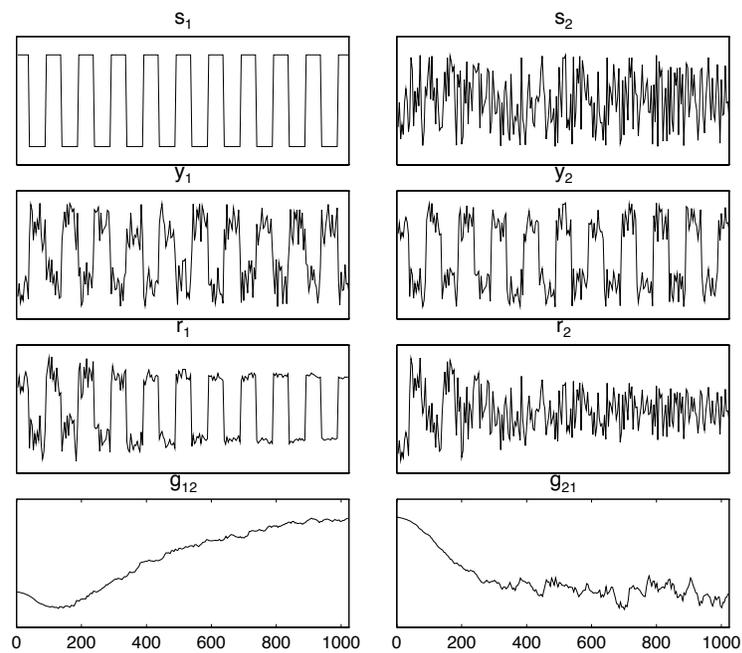


Figure 7.7: Réduction de bruit par un système neuromimétique de Jutten et Héault.

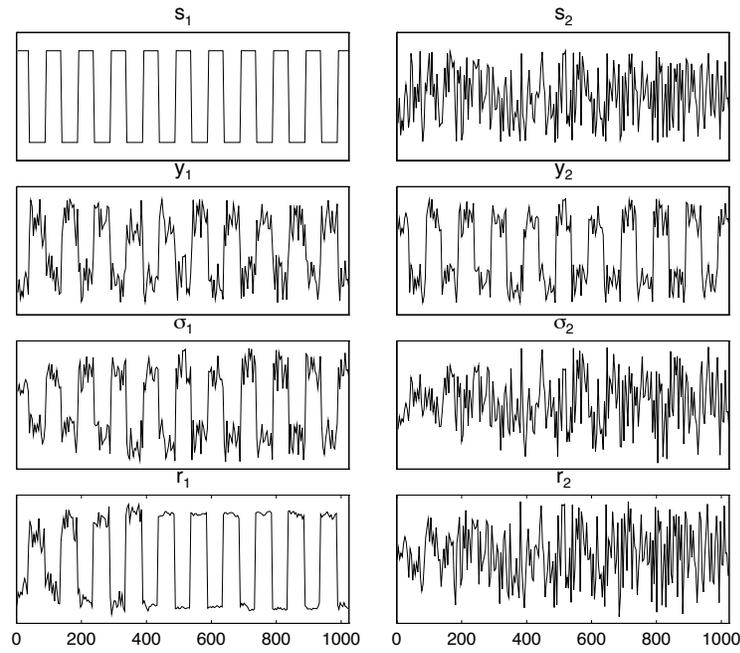


Figure 7.8: Réduction de bruit par un système adaptatif procédant en deux étapes : blanchiment à l'ordre 2 et utilisation des statistiques d'ordre supérieur.

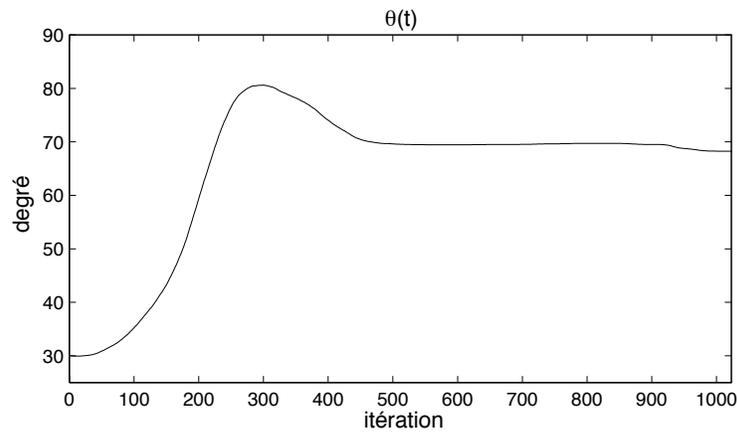


Figure 7.9: Evolution de l'angle obtenu par l'algorithme adaptatif de séparation procédant en deux étapes.

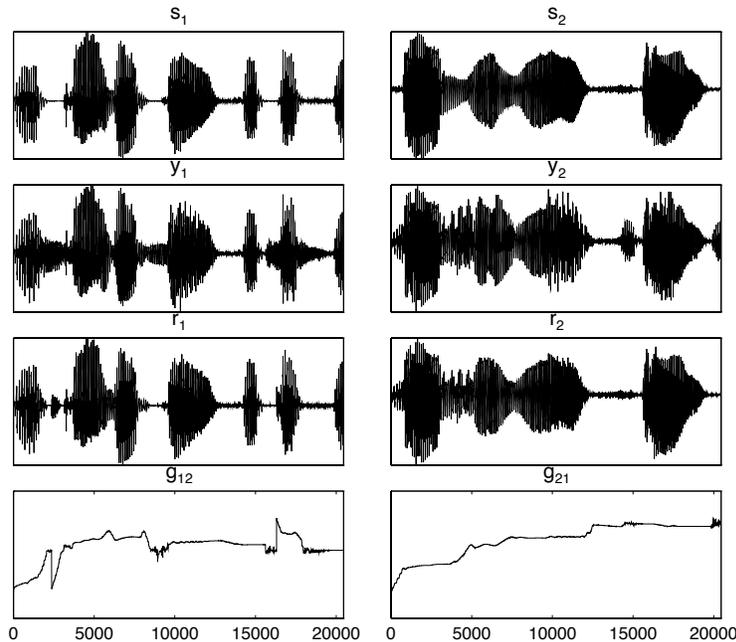


Figure 7.10: *Séparation de signaux de parole par l'algorithme de Héroult et Jutten.*

Nous présentons tout d'abord, figure (7.7), la séparation d'un signal formé de créneaux et d'un bruit par la méthode des réseaux neuromimétiques. Le signal s_1 est formé de créneaux d'amplitude 0,5, et le bruit s_2 est uniformément réparti sur $[-0,5; 0,5]$. Le mélange est obtenu par une rotation de 40° .

Il s'agit d'identifier les coefficients g_{12} et g_{21} du séparateur. On utilise la règle d'adaptation $g_{ij}(t) = g_{ij}(t-1) + \mu r_i^3(t-1)r_j(t-1)/E[r_i^4(t-1)]$, donnée en 7.3.3. Cette méthode conduit directement à l'estimation des sources. Les panneaux g_{12} et g_{21} donnent l'évolution des coefficients du séparateur. La convergence est atteinte dans ce cas en environ 1000 échantillons. On constate qu'après convergence les coefficients du séparateur continuent à fluctuer.

La figure (7.8) donne la séparation du même mélange par une méthode adaptative procédant en deux étapes. La première étape réalise la décorrélation par une orthogonalisation de Gram-Schmidt, suivie d'une normalisation des deux signaux décorrélés. L'étape finale prend l'annulation de la somme des carrés des cumulants croisés d'ordre 4 comme règle d'adaptation pour estimer la rotation θ . La figure (7.9) montre l'évolution de l'angle θ au cours du temps.

Pour terminer nous donnons sur la figure (7.10) la séparation de deux signaux non stationnaires par l'algorithme de Héroult et Jutten présenté au paragraphe 7.3.3. Les signaux séparés sont des signaux de parole¹¹. Ce cas difficile illustre les possibilités des systèmes adaptatifs.

¹¹Ces données sont issues de la thèse de L. Nguyen Thi [159].

7.4 Séparation de mélanges convolutifs

Le cas le plus général de mélange linéaire correspond au mélange convolutif que nous avons présenté au paragraphe 7.1.1. L'étude des systèmes permettant de séparer des mélanges convolutifs est encore un sujet en voie de développement. Nous allons présenter les solutions actuellement proposées en nous limitant à la séparation de deux sources. Les résultats que nous développons ici sont issus des travaux de D. Yellin et E. Weinstein [203] et de H.L. Nguyen Thi et C. Jutten [160].

7.4.1 Le modèle convolutif

Comme cela est fait dans les articles qui nous servent de référence, le bruit est supposé négligeable et nous n'en tenons pas compte. La relation entre les sources \mathbf{s} et les signaux observés \mathbf{y} est,

$$\begin{aligned} y_1(t) &= s_1(t) + \int a_{12}(t - \tau) s_2(\tau) d\tau, \\ y_2(t) &= s_2(t) + \int a_{21}(t - \tau) s_1(\tau) d\tau. \end{aligned}$$

Dans le domaine spectral, cette relation s'écrit

$$\begin{pmatrix} y_1(\nu) \\ y_2(\nu) \end{pmatrix} = \begin{pmatrix} 1 & a_{12}(\nu) \\ a_{21}(\nu) & 1 \end{pmatrix} \begin{pmatrix} s_1(\nu) \\ s_2(\nu) \end{pmatrix}.$$

Le séparateur est un filtre linéaire que nous écrirons dans le domaine spectral

$$\mathbf{r}(\nu) = G_1(\nu) \mathbf{y}(\nu),$$

avec

$$G_1 = \begin{pmatrix} 1 & -g_{12}(\nu) \\ -g_{21}(\nu) & 1 \end{pmatrix}.$$

On obtient alors

$$\begin{pmatrix} r_1(\nu) \\ r_2(\nu) \end{pmatrix} = \begin{pmatrix} 1 - g_{12}(\nu)a_{21}(\nu) & a_{12}(\nu) - g_{12}(\nu) \\ a_{21}(\nu) - g_{21}(\nu) & 1 - g_{21}(\nu)a_{12}(\nu) \end{pmatrix} \begin{pmatrix} s_1(\nu) \\ s_2(\nu) \end{pmatrix}$$

La séparation peut alors être réalisée de deux manières.

1. Avec $g_{12}(\nu) = a_{12}(\nu)$ et $g_{21}(\nu) = a_{21}(\nu)$, on obtient

$$\begin{pmatrix} r_1(\nu) \\ r_2(\nu) \end{pmatrix} = \begin{pmatrix} t_{11}(\nu) & 0 \\ 0 & t_{22}(\nu) \end{pmatrix} \begin{pmatrix} s_1(\nu) \\ s_2(\nu) \end{pmatrix}$$

où

$$t_{11}(\nu) = t_{22}(\nu) = 1 - g_{12}(\nu)g_{21}(\nu).$$

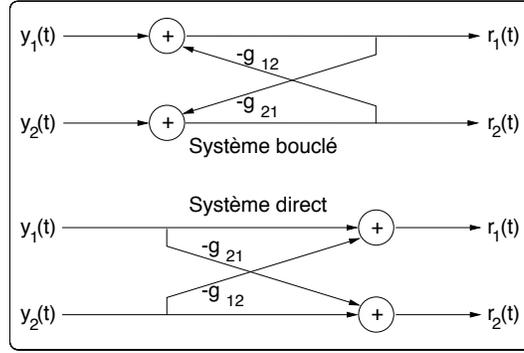


Figure 7.11: Schémas du problème de séparation pour les mélanges convolutifs.

2. Avec $g_{12}(\nu) = 1/a_{21}(\nu)$ et $g_{21}(\nu) = 1/a_{12}(\nu)$, on obtient

$$\begin{pmatrix} r_1(\nu) \\ r_2(\nu) \end{pmatrix} = \begin{pmatrix} 0 & t_{12}(\nu) \\ t_{21}(\nu) & 0 \end{pmatrix} \begin{pmatrix} s_1(\nu) \\ s_2(\nu) \end{pmatrix}$$

où

$$t_{12}(\nu) = \frac{1 - g_{12}(\nu)g_{21}(\nu)}{g_{21}(\nu)} \text{ et } t_{21}(\nu) = \frac{1 - g_{12}(\nu)g_{21}(\nu)}{g_{12}(\nu)}.$$

La séparation ainsi réalisée donne les sources à une filtrage près. La transformation issue du filtrage peut être très gênante dans le cas de signaux large bande. Dans la séparation de signaux de parole par exemple, les locuteurs deviendront incompréhensibles.

Ce filtrage peut être éliminé. Pour cela, remplaçons le séparateur G_1 par

$$G_2 = \frac{1}{1 - g_{12}(\nu)g_{21}(\nu)} \begin{pmatrix} 1 & -g_{12}(\nu) \\ -g_{21}(\nu) & 1 \end{pmatrix}.$$

Cette opération consiste simplement à remplacer le filtre direct associé à G_1 par un filtre en boucle ayant les mêmes gains, comme le montre la figure (7.11). Dans la situation 1 précédente, le séparateur G_2 donne les deux sources par $r_1(\nu) = s_1(\nu)$ et $r_2(\nu) = s_2(\nu)$. Dans la situation 2, on peut obtenir les sources en sortie de G_2 en procédant de deux façons :

- En filtrant r_1 par g_{21} et r_2 par g_{12} , on obtient les sources par $r_1(\nu) = s_2(\nu)$ et $r_2(\nu) = s_1(\nu)$. Notons que les sources sont permutées.
- On retrouve la source 1 sur la sortie 1 et la source 2 sur la sortie 2 en réalisant une combinaison linéaire des entrées et des sorties du séparateur. $y_1(\nu) - r_1(\nu)$ donne $s_1(\nu)$, et $y_2(\nu) - r_2(\nu)$ donne $s_2(\nu)$.

Les systèmes de séparation donnés ici supposent la connaissance du filtre de mélange; dans les situations de séparation aveugle de sources qui

nous intéressent, ici cette information n'est pas disponible. Pour obtenir le séparateur, il suffit de déterminer les gains $g_{ij}(\nu)$. Comme pour la séparation de mélanges simples, l'identification de ces gains est réalisée en utilisant l'indépendance des sources. Deux approches ont été proposées fondées sur les multispectres et les muticorrélations (multi-outils) ou sur les cumulants.

7.4.2 Séparation par les multi-outils

L'indépendance des sources est une condition très forte. Yellin et Weinstein ont donné dans [203] des conditions suffisantes de séparation portant sur les bispectres ou sur les trispectres des signaux observés et des sources estimées. Ces résultats conduisent à des algorithmes de séparation agissant dans le domaine spectral ou dans le domaine temporel.

Séparation par le bispectre

Pour les besoins de ce paragraphe et du suivant, nous introduisons une notation supplémentaire pour les bi- et tricorrélations, et les bi- et trispectres, de façon à manier simplement des inter-outils.

Soient $s_1(t)$ et $s_2(t)$ deux signaux, a priori à valeurs complexes. On définit alors, les indices prenant les valeurs 1 ou 2,

$$C_{s_i s_j}^{s_k}(\tau_1, \tau_2) = \text{Cum}[s_i(t), s_j(t + \tau_1), s_k^*(t - \tau_2)],$$

qui est une bicorrélation si $i = j = k$ et une interbicorrélation sinon. Par analogie avec les notations développées au chapitre 2, l'instant de référence t est l'argument de l'indice en bas à gauche, et par conséquent concerne un terme non conjugué. On peut évidemment définir d'autres bicorrélations en changeant de place les indices j et k , c'est-à-dire en conjugant de façon différente les termes intervenant dans le cumulants. La transformation de Fourier de la bicorrélation précédente définit le bispectre noté $S_{s_i s_j}^{s_k}(\nu_1, \nu_2)$.

De la même manière, on définit la tricorrélation

$$C_{s_i s_j}^{s_k s_l}(\tau_1, \tau_2, \tau_3) = \text{Cum}[s_i(t), s_j(t + \tau_1), s_k^*(t - \tau_2), s_l^*(t - \tau_3)],$$

dont la transformée de Fourier conduit au trispectre $S_{s_i s_j}^{s_k s_l}(\nu_1, \nu_2, \nu_3)$.

Etant donné le mélanges de deux sources $s_1(t)$ et $s_2(t)$, pour pouvoir réaliser la séparation, il suffit que, pour toutes les valeurs des fréquences ν_1 et ν_2 , le bispectre de chacune des sources $S_{s_i s_i}^{s_i}(\nu_1, \nu_2)$ soit différent de 0 et que, pour toutes les valeurs des fréquences ν_1 et ν_2 , le bispectre $S_{s_i s_j}^{s_k}(\nu_1, \nu_2)$ soit nul lorsque l'on n'a pas $i = j = k$.

Le séparateur sera obtenu en annulant les bispectres $S_{r_1 r_2}^{r_1}(\nu_1, \nu_2)$ et $S_{r_2 r_1}^{r_2}(\nu_1, \nu_2)$ des sources estimées. En outre, le système global de séparation doit vérifier la condition $\text{Det}[Q(0)] \neq 0$, pour éviter la solution triviale $Q(\nu) = 0$.

Le système de séparation est donné par

$$\begin{pmatrix} r_1(\nu) \\ r_2(\nu) \end{pmatrix} = \begin{pmatrix} 1 & -g_{12}(\nu) \\ -g_{21}(\nu) & 1 \end{pmatrix} \begin{pmatrix} y_1(\nu) \\ y_2(\nu) \end{pmatrix}, \quad (7.17)$$

dans lequel on a forcé à 1 les termes diagonaux pour obtenir les sources au niveau des capteurs sans filtrage supplémentaire. En remplaçant, dans le bispectre $S_{r_1 r_2}^{r_1}(\nu_1, \nu_2)$, la composante $r_2(\nu_1)$ par son expression donnée par (7.17) on obtient,

$$S_{r_1 r_2}^{r_1}(\nu_1, \nu_2) = S_{r_1 y_2}^{r_1}(\nu_1, \nu_2) - g_{21}(\nu_1) S_{r_1 y_1}^{r_1}(\nu_1, \nu_2).$$

De même

$$S_{r_2 r_1}^{r_2}(\nu_1, \nu_2) = S_{r_2 y_1}^{r_2}(\nu_1, \nu_2) - g_{12}(\nu_1) S_{r_2 y_2}^{r_2}(\nu_1, \nu_2).$$

Les sources étant indépendantes, ces deux bispectres sont nuls. On en déduit

$$g_{12}(\nu_1) = \frac{S_{r_2 y_1}^{r_2}(\nu_1, \nu_2)}{S_{r_2 y_2}^{r_2}(\nu_1, \nu_2)} \quad \forall \nu_2 \quad (7.18)$$

$$g_{21}(\nu_1) = \frac{S_{r_1 y_2}^{r_1}(\nu_1, \nu_2)}{S_{r_1 y_1}^{r_1}(\nu_1, \nu_2)} \quad \forall \nu_2. \quad (7.19)$$

Le membre de droite de l'équation donnant $g_{12}(\nu_1)$ dépend de $g_{21}(\nu_1)$ et *vice-versa*. On peut donc résoudre ce système par itération. Dans le calcul on doit disposer des bispectres qui sont estimés par l'une des méthodes classiques données au chapitre 2. Notons que l'on peut utiliser la redondance des relations (7.18) et (7.19), qui sont vraies pour toute valeur de ν_2 , pour améliorer la précision statistique de l'estimation.

On peut également obtenir la matrice de séparation dans le domaine temporel. Dans cette approche, le séparateur est modélisé par un filtre à réponse impulsionnelle finie (RIF)

$$\begin{aligned} r_1(t) &= y_1(t) - \sum_{k=p_1}^{p_2} a_k y_2(t-k) \\ r_2(t) &= y_2(t) - \sum_{k=q_1}^{q_2} b_k y_1(t-k). \end{aligned}$$

Dans ce modèle on peut introduire des filtres non causaux en prenant des valeurs négatives pour p_1 et q_1 .

La condition suffisante de séparation appliquée aux bicorrélations stipule que la séparation est réalisée si les bicorrélations

$$\begin{aligned} C_{r_1 r_2}^{r_1}(\tau_1, \tau_2) &= \text{Cum}[r_1(t), r_2(t + \tau_1), r_1^*(t - \tau_2)] \\ &= C_{r_1 y_2}^{r_1}(\tau_1, \tau_2) - \sum_{k=q_1}^{q_2} b_k C_{r_1 y_1}^{r_1}(\tau_1 - k, \tau_2) \end{aligned}$$

et

$$\begin{aligned} C_{r_2 r_1}^{r_2}(\tau_1, \tau_2) &= \text{Cum}[r_2(t), r_1(t + \tau_1), r_2^*(t - \tau_2)] \\ &= C_{r_2 y_1}^{r_2}(\tau_1, \tau_2) - \sum_{k=p_1}^{p_2} a_k C_{r_2 y_2}^{r_2}(\tau_1 - k, \tau_2), \end{aligned}$$

sont nulles pour toutes les valeurs de τ_1 et τ_2 . En exprimant la nullité de ces bicorrélations, on obtient pour les coefficients a_k et b_k les équations linéaires

$$\sum_{k=p_1}^{p_2} a_k C_{r_2 y_2}^{r_2}(\tau_1 - k, \tau_2) = C_{r_2 y_1}^{r_2}(\tau_1, \tau_2) \quad \forall \tau_1 \text{ et } \tau_2 \quad (7.20)$$

$$\sum_{k=q_1}^{q_2} b_k C_{r_1 y_1}^{r_1}(\tau_1 - k, \tau_2) = C_{r_1 y_2}^{r_1}(\tau_1, \tau_2) \quad \forall \tau_1 \text{ et } \tau_2. \quad (7.21)$$

Comme précédemment, ces équations peuvent être résolues par un algorithme itératif. Les valeurs des bicorrélations qui interviennent dans ce calcul sont estimées à partir de la réalisation particulière des signaux d'entrée et de sortie par les méthodes données au chapitre 2. La redondance apparaissant dans les relations (7.20) et (7.21) peut être utilisée pour augmenter la précision statistique du résultat. Étant donnée la forme retenue pour le système de séparation (7.17), on obtient, à une permutation près, la source 1 sur le capteur 1 et la source 2 sur le capteur 2.

L'utilisation des bispectres permet donc de réaliser la séparation de sources dont le bispectre est différent de 0. Pour les sources ayant une densité de probabilité paire le bispectre est nul. On peut alors réaliser la séparation en utilisant le trispectre.

Séparation par le trispectre Les résultats donnés pour le bispectre sont étendus dans [203] au trispectre.

La condition suffisante de séparation porte sur les valeurs du trispectre. On pourra séparer deux sources $s_1(t)$ et $s_2(t)$ si leurs trispectres sont non nuls pour toutes les valeurs des fréquences. Comme précédemment, la séparation peut être faite dans le domaine spectral ou dans le domaine temporel.

Dans le domaine spectral, les relations (7.18) et (7.19) deviennent

$$g_{12}(\nu_1) = \frac{S_{r_2 y_1}^{r_2 r_2}(\nu_1, \nu_2, \nu_3)}{S_{r_2 y_2}^{r_2 r_2}(\nu_1, \nu_2, \nu_3)} \quad \forall \nu_2 \text{ et } \nu_3$$

et

$$g_{21}(\nu_1) = \frac{S_{r_1 y_2}^{r_1 r_1}(\nu_1, \nu_2, \nu_3)}{S_{r_1 y_1}^{r_1 r_1}(\nu_1, \nu_2, \nu_3)} \quad \forall \nu_2 \text{ et } \nu_3$$

La redondance de ces relations est deux fois plus importante que précédemment : on peut les moyenner sur ν_2 et ν_3 pour améliorer l'estimation.

En temps, avec les notations précédentes, on obtient pour les coefficients a_k et b_k

$$\sum_{k=p_1}^{p_2} a_k C_{r_2 y_2}^{r_2 r_2}(\tau_1 - k, \tau_2, \tau_3) = C_{r_2 y_1}^{r_2 r_2}(\tau_1, \tau_2, \tau_3) \quad \forall \tau_1, \tau_2, \text{ et } \tau_3$$

$$\sum_{k=q_1}^{q_2} b_k C_{r_1 y_1}^{r_1 r_1}(\tau_1 - k, \tau_2, \tau_3) = C_{r_1 y_2}^{r_1 r_1}(\tau_1, \tau_2, \tau_3) \quad \forall \tau_1, \tau_2, \text{ et } \tau_3.$$

7.4.3 Séparation par les cumulants

Une approche par un algorithme en boucle, généralisant aux mélanges convolutifs la méthode donnée au paragraphe 7.3.3, est décrite dans [160]. Reprenons les notations précédentes et appelons $\mathbf{r}(t)$ les sources estimées. L'annulation des cumulants d'ordre 4 symétriques $Cum[r_i(t), r_i(t), r_j(t), r_j(t)]$ pour $i \neq j$ est une condition suffisante de séparation lorsque les kurtosis des sources sont de même signe. Sur cette base on peut construire un système adaptatif convergeant vers le séparateur en fixant pour chaque composante a_k ou b_k de la matrice de séparation la règle d'évolution

$$a_k(t) = a_k(t-1) - \mu Cum[r_1(t-1), r_1(t-1), r_2(t-1), r_2(t-1)].$$

En parallèle, on doit faire évoluer les cumulants. On peut choisir un estimateur adaptatif des cumulants avec facteur d'oubli.

Il est montré dans [160], sur la base d'une étude expérimentale, que la règle d'adaptation utilisant les cumulants symétriques est mal conditionnée. Les auteurs proposent de les remplacer par les cumulants non symétriques $Cum[r_1(t-1), r_1(t-1), r_1(t-1), r_2(t-1)]$ et $Cum[r_1(t-1), r_2(t-1), r_2(t-1), r_2(t-1)]$. La règle d'adaptation est alors mieux conditionnée mais il existe des solutions parasites que l'on peut identifier, *a posteriori*, en testant la nullité du cumulant symétrique.

Conclusion et commentaires Les voies de recherche conduisant à la solution du problème de la séparation de mélanges convolutifs ont été ouvertes. Notre appréciation est que ce domaine de recherche, très riche d'applications, reste encore largement inexploré. Avant de véritablement maîtriser ces systèmes, de nombreux problèmes de stabilité et de convergence devront être résolus. Enfin les algorithmes proposés laissent encore de nombreux choix ouverts et la question de leur sensibilité au bruit est encore largement inexplorée.

7.5 Domaines d'application

Les applications potentielles de la séparation de sources concernent de nombreux domaines. Si les mélanges simples (sans retards) sont bien maîtrisés, les mélanges complexes (convolutifs) requièrent le développement de nouvelles techniques.

Des algorithmes de séparation de sources par des statistiques d'ordre supérieur ont été proposés dans divers domaines d'application.

- La séparation de sources vibrantes appliquée au contrôle d'un atelier dans lequel fonctionnent simultanément plusieurs machines ou à la discrétion acoustique des navires. Dans ce contexte, des exemples de mise en oeuvre sont présentés dans [81].
- L'augmentation du trafic autour des aéroports conduit à la présence de plusieurs aéronefs dans chacune des voies des radars de contrôle. Dans [48] on décrit un système de séparation d'ordre supérieur permettant de séparer les cibles présentes dans la même voie du radar.
- En traitement des électrocardiogrammes, un exemple célèbre de traitement est, pour une femme portant un enfant, la séparation du signal électrique issu du cœur de la mère de celui émis par le cœur du fœtus [197, 198]. Les auteurs de [63] montrent les améliorations apportées au traitement à l'ordre 2 par la mise en oeuvre de techniques d'analyse en composantes indépendantes.

Ces exemples montrent que les techniques de séparation de sources sont sur le point de déboucher au niveau des applications. On peut, pour terminer, évoquer d'autres domaines où les applications apparaissent :

- la lecture de badges électroniques, lorsque plusieurs badges sont captés simultanément,
- la réduction de bruit sur des signaux de parole et son application au téléphone mains-libres,
- la localisation de sources sonores...

Bibliographie

- [1] *Proc. 1st International workshop on higher-order statistics*, Vail, Colorado, U.S.A., june 1989.
- [2] *Proc. 2nd International workshop on higher-order statistics*, Chamrousse, France, july 1991.
- [3] *Proc. 3rd International workshop on higher-order statistics*, Lake Tahoe, California, U.S.A., july 1993.
- [4] *Proc. IEEE Signal Processing ATHOS Workshop on Higher-Order statistics*, Begur, Girona, Spain, june 1995.
- [5] *Proc. IEEE Signal Processing Workshop on Higher-Order statistics*, Banff, Canada, july 1997.
- [6] “Numéro spécial non gaussien, non linéaire”, *Traitement du Signal*, Vol. 7, 5, 1990.
- [7] “Special issue higher-order statistics”, *IEEE trans. ASSP*, vol. 38, 7, 1990.
- [8] “Special issue higher-order statistics”, *IEEE trans. AC*, vol. 35, 1, 1990.
- [9] “Special issue on applications of high-order statistics”, *IEE Proceedings-Part F*, vol. 140, 6, 1993.
- [10] “Special issue higher-order statistics”, *Signal Processing*, vol. 36, 3, 1994.
- [11] “Special issue higher-order statistics”, *Journ. Franklin Institute*, vol. 333(B), 3, 1996.
- [12] “Special issue higher-order statistics”, *Signal Processing*, vol. 53, 2-3, 1996.
- [13] H. AKAIKE, “Note on higher spectra”, *Ann. of the Inst. of Stat. Math.*, vol. 18, pp. 123–126, 1966.
- [14] P. O. AMBLARD, J.M. BROSSIER, “Adaptive estimation of the fourth-order cumulant of a white stochastic process”, *Signal Processing*, vol. 42, no. 1, pp. 37–43, 1995.
- [15] P. O. AMBLARD, M. GAETA, J. L. LACOUME, “Complex random variables and signals. part 1 : variables”, *Signal Processing*, vol. 42, no. 11, pp. 1–13, Nov. 1996.

- [16] P. O. AMBLARD, M. GAETA, J. L. LACOUME, “Complex random variables and signals. part 2 : signals”, *Signal Processing*, vol. 42, no. 11, pp. 15–25, Nov. 1996.
- [17] P. O. AMBLARD, J. L. LACOUME, “A deductive construction of third-order time-frequency distributions”, *Signal Processing*, vol. 36, no. 3, pp. 277–286, 1994.
- [18] P. O. AMBLARD, J. L. LACOUME, J. M. BROSSIER, “Transient detection, higher-order time-frequency distributions and the entropy”, in *Proc. IEEE SP Workshop on Higher-Order Stat., Lake Tahoe, USA*, 1993, pp. 265–269.
- [19] P. O. AMBLARD, E. MOISAN, “Frequency domain volterra filters in terms of distributions”, *IEEE Sig. Proc. Letters*, vol. 1, no. 11, pp. 169–172, Nov. 1994.
- [20] M. S. BARTLETT, “Smoothing periodograms from time series with continuous spectra”, *Nature*, vol. 161, pp. 686–687, May 1948.
- [21] J. BASS, *Cours de Mathématique, Tome III*, Masson, 1971.
- [22] D. BAUDOIS, C. SERVIERE, A. SILVENT, “Soustraction de bruit : Analyse et synthèse bibliographique”, *Traitement du Signal*, vol. 6, no. 5, pp. 391–497, 1989.
- [23] S. BELLINI et al., *Blind deconvolution : polyspectrum and Bussgang techniques*, pp. 251–263, Elsevier, 1986, Biglieri ed.
- [24] A. BELOUCHRANI, *Séparation autodidacte de sources : algorithmes, performances et applications à des signaux expérimentaux*, Thèse ENST, juillet 1995.
- [25] A. BENVENISTE, M. GOURSAT, G. RUGET, “Robust identification of a non-minimum phase system”, *IEEE Trans. Auto. Control*, vol. 25, no. 3, pp. 385–399, June 1980.
- [26] A. BENVENISTE, M. METIVIER, R. PRIOURET, *Algorithmes Adaptatifs et Approximations Stochastiques*, Masson, 1987.
- [27] G. BIENVENU, L. KOPP, “Optimality of high-resolution array processing using the eigensystem approach”, *IEEE Trans. ASSP*, vol. 31, no. 5, pp. 1235–1248, Oct. 1983.
- [28] R. B. BLACKMAN, J. W. TUCKEY, *The measurement of power spectra from the view of communication engineering*, Dover Publication Inc., New-York, 1958.
- [29] A. BLANC-LAPIERRE, R. FORTET, *Théorie des Fonctions Aléatoires*, Masson, 1953.

- [30] B. BOASHASH, G. J. FRAZER, “Time-varying higher-order spectra, generalized Wigner-Ville distributions and the analysis of underwater acoustic data”, in *Proc. ICASSP, San Francisco*, 1992, pp. 5.193–5.196.
- [31] J.F. BOHME, D. KRAUSS, “On least square methods for direction of arrival estimation in the presence of unknown noise fields”, in *Proc. IEEE ICASSP, New York*, 1988, pp. 2866–2836.
- [32] L. BOLTZMANN, *Vorlesungen über Gastheorie*, J.A. Barth, Liepzig, 2 Vol., 1896.
- [33] P. BONDON, M. BENIDIR, B. PICINBONO, “Sur l’invérsibilité de la matrice de corrélation d’un vecteur aléatoire à composantes polynômiales”, in *Proc. GRETSI, Juans-Les-Pins*, 1991, pp. 13–16.
- [34] P. BONDON, B. PICINBONO, “De la blancheur et de ses transformations”, *Traitement du Signal*, vol. 7, no. 5, pp. 385–395, 1990.
- [35] M. BOUMAHDI, “Blind identification using the kurtosis with applications to field data”, *Signal Processing*, vol. 48, no. 3, pp. 205–216, Feb. 1996.
- [36] D. R. BRILLINGER, *Time Series, Data Analysis and Theory*, Holden-Day, 1981.
- [37] D. R. BRILLINGER, M. ROSENBLATT, *Asymptotic theory of k-th order spectra*, pp. 153–188, Wiley, 1967, in *Spectral Analysis of Time Series*, Harris ed.
- [38] D. R. BRILLINGER, M. ROSENBLATT, *Computation and interpretation of k-th order spectra*, pp. 189–232, Wiley, 1967, in *Spectral Analysis of Time Series*, Harris ed.
- [39] J. J. BUSSGANG, *Cross-correlation function of amplitude-distorted Gaussian signals*, 1952.
- [40] J.P. LE CADRE, Ph. LOPEZ, “Estimation d’une matrice spectrale de structure imposée. Applications”, *Traitement du Signal*, vol. 1, no. 1, pp. 3–19, Oct. 1984.
- [41] A. A. CADZOW, “Blind deconvolution via cumulant extrema”, *IEEE Sig. Proc. Mag.*, vol. 13, pp. 24–42, May 1996.
- [42] J. F. CARDOSO, “Source separation using higher order moments”, in *Proc. ICASSP Glasgow*, 1989, pp. 2109–2112.
- [43] J. F. CARDOSO, “Localisation et identification par la quadricovariance”, *Traitement du Signal*, vol. 7, no. 5, pp. 397–406, Dec. 1990.
- [44] J. F. CARDOSO, “On the performance of source separation algorithms”, in *Proc. EUSIPCO, Edinburgh*, Sept. 1994, pp. 776–779.

- [45] J. F. CARDOSO, A. SOULOUMIAC, “Blind beamforming for non-Gaussian signals”, *IEE Proceedings-Part F*, vol. 140, no. 6, pp. 362–370, Dec. 1993, Special issue on Applications of High-Order Statistics.
- [46] J. F. CARDOSO, A. SOULOUMIAC, “An efficient technique for blind separation of complex sources”, in *Proc. IEEE SP Workshop on Higher-Order Stat., Lake Tahoe, USA*, 1993, pp. 275–279.
- [47] J.F. CARDOSO, B. LAHELD, “Equivariant adaptive source separation”, *IEEE Trans. on SP*, vol. 44, no. 12, pp. 3017–3030, Dec. 1996.
- [48] E. CHAUMETTE, P. COMON, D. MULLER, “An ICA-based technique for radiating sources estimation; application to airport surveillance”, *IEE Proceedings - Part F*, vol. 140, no. 6, pp. 395–401, Dec. 1993, Special issue on Applications of High-Order Statistics.
- [49] P. CHEVALIER, “Méthodes aveugles de filtrage d’antennes”, *Revue d’Electronique et d’Electricité, SEE*, , no. 3, pp. 48–58, Sept 1995.
- [50] P. CHEVALIER, A. FERREOL, J. P. DENIS, “New geometrical results about fourth-order direction finding method performance”, in *EUSIPCO 96*, Trieste, Italy, September 10-13 1996, pp. 923–926.
- [51] P. CHEVALIER, B. PICINBONO, P. DUVAUT, “Le filtrage de volterra transverse réel et complexe”, *Traitement du Signal*, vol. 7, no. 5, pp. 451–476, 1990.
- [52] C. Y. CHI, J. Y. KUNG, “A phase determination method of non minimum phase ARMA system by a single cumulant sample”, *IEEE Trans. on SP*, vol. 41, no. 2, pp. 981–985, Feb. 1993.
- [53] P. COMON, “Séparation de mélanges de signaux”, in *Proc. GretsI Juans-les-Pins*, 12 -16 juin 1989, pp. 137–140.
- [54] P. COMON, “Separation of stochastic processes”, in *Proc. Workshop on Higher-Order Spectral Analysis*, Vail, Colorado, June 28-30 1989, IEEE-ONR-NSF, pp. 174–179.
- [55] P. COMON, “Analyse en Composantes Indépendantes et identification aveugle”, *Traitement du Signal*, vol. 7, no. 3, pp. 435–450, Dec. 1990, Numéro spécial non linéaire et non gaussien.
- [56] P. COMON, “MA identification using fourth order cumulants”, *Signal Processing*, vol. 26, no. 3, pp. 381–388, 1992.
- [57] P. COMON, “Independent Component Analysis, a new concept? ”, *Signal Processing*, vol. 36, no. 3, pp. 287–314, Apr. 1994, Special issue on Higher-Order Statistics.
- [58] P. COMON, “Contrasts for multichannel blind deconvolution”, *Signal Processing Letters*, vol. 3, no. 7, pp. 209–211, July 1996.

- [59] J. W. COOLEY, J. W. TUCKEY, “An algorithm for the machine computation of complex Fourier series”, *Mathematics of Computation*, vol. 19, pp. 297–301, May 1965.
- [60] H. CRAMER, *Mathematical methods of statistics*, Princeton University, 1946.
- [61] A. V. DANDAWATE, G. B. GIANNAKIS, “Nonparametric polyspectral estimators for k th-order (almost) cyclostationary processes”, *IEEE Trans. on IT*, vol. 40, no. 1, pp. 67–84, Jan. 1994.
- [62] A. V. DANDAWATE, G. B. GIANNAKIS, “Asymptotic theory of mixed time averages and k th-order cyclic-moment and cumulant statistics”, *IEEE Trans. on IT*, vol. 41, no. 1, pp. 216–232, Jan. 1995.
- [63] L. DeLATHAUWER, B. DeMOOR, J. VANDEWALLE, “Fetal electrocardiogram extraction by source subspace separation”, in *IEEE-ATHOS Workshop on Higher-Order Statistics*, Begur, Spain, 12–14 June 1995, pp. 134–138.
- [64] A. N. DELOPOULOS, G. B. GIANNAKIS, “Strongly consistent identification algorithms and noise insensitive MSE criteria”, *IEEE Trans. on SP*, vol. 40, no. 8, pp. 1955–1970, Aug. 1992.
- [65] D. DONOHO, “On minimum entropy deconvolution”, in *Applied time-series analysis II*, pp. 565–609. Academic Press, 1981.
- [66] P. DURAND, *Théorie et estimation conventionnelle du bispectre. Application aux processus linéaire-quadratique à spectre de raies*, Doctorat de l’INPG, Grenoble, juillet 1991.
- [67] P. DUVAUT, “Le filtrage de Wiener linéaire-quadratique à horizon fini. Application à la prédiction”, *Traitement du Signal*, vol. 6, no. 3, pp. 152–159, 1989.
- [68] P. DUVAUT, “A unifying and general approach to adaptive linear-quadratic discrete-time volterra filtering”, in *Proc. ICASSP Glasgow*, 1989, pp. 1166–1170.
- [69] P. DUVAUT, “Influence des moments d’ordre trois sur la structure d’un algorithme adaptatif linéaire-quadratique”, *Traitement du Signal*, vol. 7, no. 5, pp. 477–488, 1990.
- [70] P. DUVAUT, T. GOURAUD, C. CORROYER, “Bet : the bicorspectral transform. A new tool for nongaussian processes”, in *Proc. IEEE SP Workshop on Higher-Order Stat., Begur, Spain*, 1995, pp. 424–427.
- [71] G. FAVIER, D. DEMBELE, J. L. PEYRE, “ARMA identification using high-order statistics based linear methods: a unified presentation”, in *Proc. EU-SIPCO*, Edinburgh, Sept. 1994, pp. 203–207.
- [72] C. C. FENG, C. Y. CHI, “Design of Wiener filters using cumulant based mse criterion”, *Signal Processing*, vol. 54, no. 1, pp. 23–48, Oct. 1996.

- [73] L. FETY, *Méthodes de traitement d'antenne adaptées aux radiocommunications*, Doctorat, ENST, 1988.
- [74] P. FLANDRIN, *Temps-Fréquence*, Hermès, 1993.
- [75] J. R. FONOLLOSA, C. M. NIKIAS, “Wigner higher-order moment spectra : Definition, properties, computation and application to transient signal analysis”, *IEEE Trans. on SP*, vol. 41, no. 1, pp. 245–266, Jan. 1993.
- [76] J.C. FORT, “Stabilité de l’algorithme de séparation de sources de Jutten et Héroult”, *Traitement du Signal*, vol. 8, no. 1, pp. 35–42, Jan. 1991.
- [77] R. FORTET, *Eléments de la théorie des probabilités*, CNRS, 1965.
- [78] G. FOWLES, *Introduction to Modern Optics*, Dover, 1975.
- [79] M. FRECHET, “Sur les fonctionnelles continues”, *Ann. Sci. Ecole Normale Sup.*, vol. 27, pp. 193–219, 1910.
- [80] B. FRIEDLANDER, “Parametric signal analysis using the polynomial phase transform”, in *Proc. IEEE SP Workshop on Higher-Order Stat., Lake Tahoe, USA*, 1993, pp. 151–159.
- [81] M. GAETA, F. HARROY, J.L. LACOUME, “Utilisation de statistiques d’ordre supérieur pour la séparation de sources vibrantes”, in *Colloque Bruit et Vibration, SENLIS*, 1992.
- [82] M. GAETA, J. L. LACOUME, “Source separation without a priori knowledge: the maximum likelihood solution”, in *Proc. EUSIPCO*, Barcelona, Spain, 1990, pp. 621–624.
- [83] M. GAETA, J.L. LACOUME, “Estimateurs du maximum de vraisemblance étendus à la séparation de sources non-gaussiennes”, *Traitement du Signal*, vol. 7, no. 5, pp. 419–444, Dec. 1990.
- [84] W. A. GARDNER, C. M. SPOONER, “The cumulant theory of cyclostationary time-series, part 1 : Foundation”, *IEEE Trans. on SP*, vol. 42, no. 12, pp. 3387–3408, Dec. 1994.
- [85] W. A. GARDNER, C. M. SPOONER, “The cumulant theory of cyclostationary time-series, part 2 : Development and applications”, *IEEE Trans. on SP*, vol. 42, no. 12, pp. 3409–3429, Dec. 1994.
- [86] N. GERR, “Introducing a third-order Wigner distribution”, *Proc. IEEE*, vol. 76, no. 3, pp. 290–292, Mar. 1988.
- [87] G. B. GIANNAKIS, “Cumulants: A powerful tool in signal processing”, *Proceedings of The IEEE*, vol. 75, no. 9, pp. 1333–1334, Sept. 1987.
- [88] G. B. GIANNAKIS, A. V. DANDAWATE, “Polyspectral analysis of nonstationary signals : Bases, consistency and hos-wv”, in *Higher-Order Statistics, J.L. Lacoume ed., Elsevier*, 1992, pp. 175–178.

- [89] G. B. GIANNAKIS, J. M. MENDEL, "Identification of non minimum phase systems using higher order statistics", *IEEE Trans. on ASSP*, vol. 37, no. 3, pp. 360–377, Mar. 1989.
- [90] G. B. GIANNAKIS, M. K. TSATSANIS, "A unifying maximum-likelihood view of cumulant and polyspectra measures for non-Gaussian signal classification and estimation", *IEEE Trans. on IT*, vol. 38, no. 2, pp. 386–406, Mar. 1992.
- [91] F. GLANGEAUD, J.L. MARI, *Wave separation*, Editions Technip, 1994.
- [92] D. GODARD, "Self recovering equalization and carrier tracking in two dimensional data communication systems", *IEEE Trans. on COM*, vol. 28, no. 11, pp. 1867–1875, Nov. 1980.
- [93] G. H. GOLUB, C. F. VAN LOAN, *Matrix computations*, The John Hopkins University Press, 1989.
- [94] N. R. GOODMAN, "Statistical analysis based on certain multivariate complex normal distributions", *Annals Math. Stat.*, vol. 34, pp. 152–177, 1963.
- [95] A. GOROKHOV, *Séparation autodidacte de mélanges convolutifs : méthodes du second ordre*, Thèse ENST, mai 1997.
- [96] P. GOUNON, "Analyse spatio-temporelle haute résolution à l'aide d'une antenne active", *Traitement du Signal*, vol. 11, no. 5, pp. 351–361, Nov. 1994.
- [97] L. J. GRIFFITHS, "An adaptive lattice structure for noise-cancelling applications", in *Proc. ICASSP*, Tulsa, 1978, pp. 87–90.
- [98] J. F. GUERRE-CHALEY, *Etude de Différentes Structure de Soustracteurs de Bruit Adaptatifs Multiréférence*, Doctorat de l'INPG, Grenoble, octobre 1990.
- [99] E. A. GUILLEMIN, *Theory of linear physical systems*, John Wiley and Sons, 1963.
- [100] F. HARROY, J.L. LACOUME, "Maximum likelihood estimators and cramer-rao bounds in source separation", *Signal Processing*, vol. 55, no. 2, pp. 167–178, Dec. 1996.
- [101] S. HAYKIN, *Adaptive filter theory*, Prentice Hall, 1991.
- [102] J. HERAULT, C. JUTTEN, *Réseaux neuronaux et traitement du signal*, Hermès, 1994.
- [103] J. HERAULT, C. JUTTEN, B. ANS, "Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé", in *Proc. GRETSI, Nice*, 1985, pp. 1017–1022.
- [104] M. HINICH, "Detecting a transient signal by bispectral analysis", *IEEE Trans. on ASSP*, vol. 38, no. 7, pp. 1277–1283, July 1990.

- [105] M. HINICH, “Higher order cumulants and cumulant spectra”, *Circuits systems and signal processing*, vol. 13, pp. 391–402, Apr. 1994.
- [106] P. J. HUBER, *Robust statistics*, John Wiley and Sons, 1981.
- [107] G. JACOVITTI, G. SCARANO, “Hybrid non linear moments in array processing and spectrum analysis”, *IEEE Trans. on SP*, vol. 42, no. 7, pp. 1708–1718, July 1994.
- [108] J. B. JELONNEK, K. D. KAMMEYER, “A closed-form solution to blind equalization”, *Signal Processing*, vol. 36, pp. 251–259, 1994.
- [109] A. N. JOUNY, G. B. GIANNAKIS, “The bispectrum of complex signals: definitions and properties”, *IEEE Trans. on SP*, vol. 40, no. 11, pp. 2833–2836, Nov. 1992.
- [110] J.H JUSTICE, N.L. OWSLEY, J.L. YEN, A.C. KAK, *Array signal processing*, Prentice Hall, 1985.
- [111] C. JUTTEN, J. HÉRAULT, “Independent component analysis versus PCA”, in *Proc. EUSIPCO*, Grenoble, France, 1988, pp. 643–646.
- [112] C. JUTTEN, J. HERAULT, “Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture”, *Signal Processing*, vol. 24, no. 1, pp. 1–10, 1991.
- [113] A. M. KAGAN, Y. V. LINNIK, C. R. RAO, *Characterization Problems in Mathematical Statistics*, Wiley, 1973.
- [114] K. D. KAMMEYER, B. JELONNEK, “A cumulant zero-matching method for the blind system identification”, in *Proc. Int. Sig. Proc. Workshop on Higher-Order Statistics*, Chamrousse, France, July 10-12 1991, pp. 103–106, Republished in *Higher-Order Statistics*, J.L.Lacoume ed., Elsevier, 1992, pp 153–156.
- [115] M. KENDALL, A. STUART, *The Advanced Theory of Statistics, Distribution Theory*, vol. 1, C. Griffin, 1977.
- [116] K. I. KIM, E. J. POWERS, “A digital method of modeling quadratically nonlinear systems with a general random input”, *IEEE Trans. on ASSP*, vol. 36, no. 11, pp. 1758–1769, Nov. 1988.
- [117] M. KORENBERG, L. PAARMANN, “Orthogonal approaches to time series and system identification”, *IEEE Sig. Proc. Mag.*, pp. 28–43, July 1991.
- [118] S. KOTZ, N. L. JOHNSON, *Encyclopedia of Statistical Sciences*, Wiley, 1982.
- [119] M. KROB, M. BENIDIR, “Blind identification of a linear-quadratic mixture”, in *Proc. IEEE SP Workshop on Higher-Order Stat., Lake Tahoe, USA*, 1993, pp. 351–355.

- [120] A. LABEYRIE, “Attainment of diffraction limited resolution in large telescope by Fourier analysing speckle patterns in star images”, *Astronomy and astrophysics*, vol. 6, pp. 85–87, 1970.
- [121] J. L. LACOUME, M. GAETA, P. O. AMBLARD, “From order 2 to HOS: new tools and applications”, in *Proc. European Signal Processing Conf. EUSIPCO*, Brussels, Aug 24 - 27 1992, pp. 91–98.
- [122] J. L. LACOUME, P. RUIZ, “Separation of independent sources from correlated inputs”, *IEEE Trans. SP*, vol. 40, no. 12, pp. 3074–3078, Dec. 1992.
- [123] J.L. LACOUME, “Modèles et traitements de signaux multidimensionnels”, *Traitement du Signal*, vol. 5, no. 2, pp. 87–106, Apr. 1988.
- [124] J.L. LACOUME, P. RUIZ, “Sources identification: a solution based on the cumulants”, in *Proc. 4th ASSP Workshop Spectral Estimation Modeling, Mineapolis*, 1988, pp. 199–203.
- [125] A. LANNES, “Remarkable algebraic structure of phase closure imaging and their algorithmic implications in apertures synthesis”, *J. Opt. Soc. Amer.*, vol. 7, pp. 500–512, 1990.
- [126] B. P. LATHI, *Signals, systems and communications*, J. Wiley and Sons, New-York, 1965.
- [127] C. LATOMBE, *Détection et caractérisation de signaux à partir de la matrice interspectrale*, Doctorat d’Etat, Grenoble, novembre 1982.
- [128] J. LEE, V. J. MATHEWS, “A fast recursive least squares adaptive second-order volterra filter and its performance analysis”, *IEEE Trans. on SP*, vol. 41, no. 3, pp. 1087–1102, Mar. 1993.
- [129] J. LEROUX, P. SOLE, “Least-square error reconstruction of a sampled signal Fourier transform from its n-th order polyspectrum”, *Signal Processing*, vol. 35, pp. 75–81, 1994.
- [130] J. LEROUX, P. SOLE, A. M. TEKALP, A. T. ERKEM, “Tekalp-Erkem estimator gives the ls estimate for Fourier phase and log-Fourier modulus”, *IEEE Trans. on SP*, vol. 41, no. 4, pp. 1705–1707, Apr. 1993.
- [131] K. S. LIH, M. ROSENBLATT, “Deconvolution and estimation of transfert function phase and coefficients for nongaussian linear processes”, *Ann. Stat.*, vol. 10, pp. 1195–1208, May 1982.
- [132] L. LJUNG, *System identification: Theory for the user*, Prentice Hall, 1987.
- [133] A. W. LOHMAN, G. WEIGELT, B. WIRNITZER, “Speckle masking in astronomy : triple correlation theory and applications”, *Applied Optics.*, vol. 22, pp. 4028–4037, Dec. 1983.

- [134] A. W. LOHMAN, B. WIRNITZER, “Triple correlation”, *Proc. IEEE*, vol. 72, pp. 889–890, July 1984.
- [135] O. MACCHI, *Adaptive processing*, Wiley, 1995.
- [136] O. MACCHI, M. BELLANGER, “Le point sur le filtrage adaptatif transverse”, in *Proc. GRETSI, Nice*, 1987, pp. 1G–14G.
- [137] J. MAKHOUL, “Stable and efficient lattice methods for linear prediction”, *IEEE Trans. on ASSP*, vol. 25, no. 1, pp. 423–428, Jan. 1977.
- [138] A. MANSOUR, *Contribution à la séparation de sources*, Doctorat, INPG, Grenoble, janvier 1997.
- [139] P. MARCHAND, P. O. AMBLARD, J.L. LACOUME, “Statistiques d’ordre supérieur à deux pour des signaux cyclostationnaires à valeurs complexes”, in *Proc. GRETSI, Juans-Les-Pins*, 1995, pp. 69–72.
- [140] K. V. MARDIA, “Tests of univariate and multivariate normality”, in *Handbook of Statistics, Vol.1*, P. R. Krishnaiah, Ed., pp. 279–320. North-Holland, 1980.
- [141] J.L. MARI, F.COPPENS, F. GLANGEAUD, *Traitement du signal pour géologues et géophysiciens*, Technip, 1997.
- [142] S. L. MARPLE, *Digital Spectral Analysis with applications*, Prentice Hall, 1987.
- [143] J. C. MARRON, P. P. SANCHEZ, R. C. SULLIVAN, “Unwrapping algorithm for least-squares phase recovery from the modulo- 2π bispectrum phase”, *J. Opt. Soc. Amer.*, vol. 7, pp. 14–20, 1990.
- [144] W. MARTIN, P. FLANDRIN, “Wigner-Ville spectral analysis of nonstationary processes”, *IEEE Trans. on ASSP*, vol. 33, no. 6, pp. 1461–1470, Dec. 1985.
- [145] T. MATSUOKA, T. J. ULRYCH, “Phase estimation using the bispectrum”, *Proc. IEEE*, vol. 72, pp. 1403–1411, Sept. 1984.
- [146] V. J. MATTHEWS, D. YOUNG, N. AHMED, “A unified approach to non parametric spectrum estimation algorithm”, *IEEE Trans. on ASSP*, vol. 35, no. 3, pp. 338–349, Mar. 1987.
- [147] J. MAX, J. L. LACOUME, *Traitement du Signal, 5e ed.*, Masson, 1996.
- [148] P. McCULLAGH, *Tensor Methods in Statistics*, Monographs on Statistics and Applied Probability. Chapman and Hall, 1987.
- [149] J. MENDEL, M. C. DOGAN, “Higher-order statistics applied to some array signal processing problems”, in *IFAC-SYSID, 10th IFAC Symposium on System Identification*, M. Blanke, T. Soderstrom, Eds., Copenhagen, Denmark, July 4-6 1994, vol. 1, pp. 101–106, invited session.

- [150] J. M. MENDEL, “Tutorial on higher-order statistics (spectra) in signal processing and system theory: Theoretical results and some applications”, *Proc. IEEE*, vol. 79, no. 3, pp. 277–305, Mar. 1991.
- [151] H. MERMOZ, “Imagerie, corrélation et modèles”, *Ann. Telecom.*, vol. 31, no. 1–2, pp. 17–36, 1976.
- [152] K. MIO, *Etudes de filtres et de processus non linéaires discrets*, Doctorat de l’INPG, Grenoble, octobre 1996.
- [153] K. MIO, E. MOISAN, P.O. AMBLARD, “Nonlinear noise cancellation: Bilinear filtering of a sonar experiment”, in *IEEE Workshop on nonlinear signal and image processing*, Halkidiki, Greece, 20–22 June 1995, pp. 416–419.
- [154] E. MOREAU, O. MACCHI, “New self-adaptive algorithms for source separation based on contrast functions”, in *Proc. IEEE Signal Processing Workshop on Higher-Order Statistics, Lake Tahoe*, 1993, pp. 215–219.
- [155] S. W. NAM, E. J. POWERS, “Application of higher-order spectral analysis to cubically nonlinear system identification”, *IEEE Trans. on SP*, vol. 42, no. 7, pp. 1746–1765, July 1994.
- [156] S. V. NARASIMHAN et al., “Phase estimation by bispectrum : A group delay approach”, *Signal Processing*, vol. 27, no. 1, pp. 79–86, Apr. 1992.
- [157] F. D. NESEER, J. L. MASSEY, “Proper complex random processes with applications to information theory”, *IEEE Trans. on IT*, vol. 39, pp. 1293–1302, July 1993.
- [158] J. W. VAN NESS, “Asymptotic normality of bispectral estimate”, *Ann. Math. Stat.*, vol. 37, pp. 1257–1272, 1966.
- [159] H.L. NGUYEN-THI, *Séparation aveugle de sources à large bande dans un mélange convolutif*, Doctorat de l’INPG, Grenoble, Avril 1993.
- [160] H.L. NGUYEN-THI, C. JUTTEN, “Blind source separation for convolutive mixture”, *Signal Processing*, vol. 45, no. 2, pp. 209–229, Aug. 1995.
- [161] C. L. NIKIAS, A. P. PETROPULU, *Higher-Order Spectra Analysis*, Signal Processing Series. Prentice-Hall, Englewood Cliffs, 1993.
- [162] C. L. NIKIAS, M. RAGHUVeer, “Bispectrum estimation: a digital signal processing framework”, *Proc. of IEEE*, vol. 75, pp. 869–891, Jan. 1985.
- [163] “Numéro spécial Algorithmes adaptatifs et soustraction de bruit”, *Traitement du Signal*, Vol. 6, 5, 1989.
- [164] A. V. OPPENHEIM, R. W. SCHAFER, *Digital signal processing*, Prentice Hall, 1975.

- [165] R. PAN, C. L. NIKIAS, “The complex cepstrum of higher order cumulants and non minimum phase systems identification”, *IEEE Trans. on ASSP*, vol. 34, no. 2, pp. 186–205, Feb. 1988.
- [166] A. PAPOULIS, *Signal analysis*, Mac Graw Hill, New York, 1977.
- [167] S. PELEG, B. PORAT, “Estimation and classification of polynomial-phase signals”, *IEEE Trans. on IT*, vol. 37, no. 2, pp. 422–430, Feb. 1991.
- [168] D.T. PHAM, Ph. GARAT, C. JUTTEN, “Separation of a mixture of independent sources through a maximum likelihood approach”, in *Proc. SIGNAL PROCESSING VI, Brussels*, 1992, pp. 771–774.
- [169] B. PICINBONO, “Geometrical concepts in higher-order statistics”, in *Proc. Int. Sig. Proc. Workshop on Higher-Order Statistics*, Chamrousse, France, July 10-12 1991, pp. 1–5, Republished in *Higher-Order Statistics*, J.L.Lacoume ed., Elsevier, 1992, pp 1–5.
- [170] B. PICINBONO, *Random Signals and Systems*, Prentice-Hall, 1993.
- [171] B. PICINBONO, “On circularity”, *IEEE Trans. SP*, vol. 42, no. 12, pp. 3473–3482, Dec. 1994.
- [172] R. L. PLACKETT, “Some theorems in least squares”, *Biometrika*, vol. 37, pp. 149–000, May 1950.
- [173] M. RANGOUSI, G. B. GIANNAKIS, “FIR modeling using log-bispectra: weighted least-squares algorithms and performances analysis”, *IEEE Trans. on CS*, vol. 38, no. 3, pp. 281–296, Mar. 1991.
- [174] E. A. ROBINSON, S. TREITEL, *Geophysical signal analysis*, Prentice Hall, 1990.
- [175] F. RODDIER, “Triple correlation as a phase closure technique”, *Optics communication*, vol. 60, pp. 145–148, Nov. 1986.
- [176] M. ROSENBLATT, *Random Processes*, Oxford press, 1962.
- [177] M. ROSENBLATT, J. W. VAN NESS, “Estimation of the bispectrum”, *Ann. Math. Stat.*, vol. 36, pp. 1120–1136, 1965.
- [178] K. SASAKI, T. SATO, Y. YAMASHITA, “Minimum bias windows for bispectral estimation”, *J. sound and Vib.*, vol. 40, no. 1, pp. 139–148, 1975.
- [179] Y. SATO, “A method of self recovering equalization for multilevel amplitude-modulation systems”, *IEEE Trans. on COM*, vol. 23, pp. 679–682, June 1975.
- [180] G. SCARANO, “Cumulant series expansion of hybrid nonlinear moments of complex random variables”, *IEEE Trans. on SP*, vol. 39, no. 4, pp. 1001–1003, 1991.

- [181] G. SCARANO, G. JACOVITTI, “Applications of generalised cumulants to array processing”, *Signal Processing*, vol. 53, no. 2, pp. 209–229, Sept. 1996.
- [182] M. SCHETZEN, *The Volterra and Wiener Theories of Nonlinear Systems*, Krieger, 1980.
- [183] R. O. SCHMIDT, “Multiple emitter location and signal parameter estimation”, *IEEE Trans. AP*, vol. 34, no. 3, pp. 276–280, Mar. 1986.
- [184] C. SERVIERE, *Eléments de Comparaison entre Différentes Méthodes de Soustraction de Bruit*, Doctorat de l’INPG, Grenoble, mai 1988.
- [185] O. SHALVI, E. WEINSTEIN, “New criteria for blind deconvolution of non-minimum phase systems”, *IEEE Trans. Inf. Theory*, vol. 36, no. 2, pp. 312–321, Mar. 1990.
- [186] T.J. SHAN, M. WAX, T. KAILATH, “On spatial smoothing for direction-of-arrival estimation of coherent signals”, *IEEE Trans. on ASSP*, vol. 33, no. 8, pp. 806–811, Aug. 1985.
- [187] A. SHIRYAYEV, “Some problems in the spectral theory of higher-order moments”, *Th. Prob. and Appl.*, vol. 5, no. 3, pp. 265–284, 1960.
- [188] A. N. SHIRYAYEV, *Probability*, Springer Verlag, 1984.
- [189] D. SLEPIAN, H. T. LANDAU, H. O. POLLACK, “Prolate spheroidal wave functions, Fourier analysis and uncertainty principle”, *Bell Syst. Tech. J.*, vol. 40, no. 5, pp. 43–80, Jan. 1961.
- [190] A. SOULOUMIAC, *Utilisation des statistiques d’ordre supérieur pour la séparation et le filtrage*, Doctorat, ENST, Février 1993.
- [191] A. SWAMI, “Third-order Wigner distributions : Definitions and properties”, in *Proc. ICASSP, Toronto*, 1991, pp. 3081–3084.
- [192] M. A. SYED, V. J. MATHEWS, “QR-decomposition based algorithms for adaptive Volterra filtering”, *IEEE Trans. on Circ. and Syst.*, vol. 40, no. 6, pp. 372–382, June 1993.
- [193] I. TAS, C. LATOMBE, “Détection multiple par les valeurs propres de la matrice spectrale”, *Traitement du Signal*, vol. 3, no. 2, pp. 79–90, Apr. 1986.
- [194] J. K. TUGNAIT, “On identifiability of arma models of non-gaussian processing via cumulant matching”, in *Proc. Int. Sig. Proc. Workshop on Higher-Order Statistics*, Chamrousse, France, July 10-12 1991, pp. 29–31, Republished in *Higher-Order Statistics*, J.L.Lacoume ed., Elsevier, 1992, pp 117–120.
- [195] J. VILLE, “Théorie et applications de la notion de signal analytique”, *Cables et Transmission*, vol. 2, no. 1, pp. 61–74, Jan. 1948.

- [196] M. WAX, T. KAILATH, “Determination of the number of signals by information theoretic criteria”, *IEEE Trans. on Acoust., Speech and Signal Processing*, vol. 32, no. 7, pp. 817–827, July 1985.
- [197] B. WIDROW et al., “Adaptive noise cancelling : Principles and applications”, *Proc. IEEE*, vol. 63, no. 12, pp. 1695–1716, Dec. 1975.
- [198] B. WIDROW, S. D. STEARN, *Adaptive signal processing*, Prentice Hall, 1985.
- [199] N. WIENER, *The extrapolation, interpolation and smoothing of stationary time series*, M.I.T. Press, Cambridge, USA, 1949.
- [200] N. WIENER, *Non-Linear Problems in Random Theory*, M.I.T. Press, 1955.
- [201] R. A. WIGGINS, “On minimum entropy deconvolution”, *Geoexploration*, vol. 16, pp. 21–35, Jan. 1978.
- [202] R. A. WOODING, “The multivariate distribution of complex normal variables”, *Biometrika*, vol. 43, pp. 212–215, 1956.
- [203] D. YELLIN, E. WEINSTEIN, “Criteria for multichannel signal separation”, *IEEE Trans. on SP*, vol. 42, no. 8, pp. 2158–2168, Aug. 1994.
- [204] N. YUEN, B. FRIEDLANDER, “Asymptotic performance analysis of Esprit, higher-order Esprit, and Virtual Esprit algorithms”, *IEEE Trans. on SP*, vol. 44, no. 10, pp. 2537–2550, Oct. 1996.
- [205] R. ZAKNOUNE, *Soustraction Adaptive de Bruit par Prédiction Vectorielle*, Doctorat de l’INPG, Grenoble, octobre 1993.
- [206] F. C. ZHENG, S. MAC LAUGHLIN, B. MULGREW, “Cumulant-based deconvolution and identification: several new families of linear equations”, *Signal Processing*, vol. 30, no. 2, pp. 199–219, Jan. 1993.

Glossaire

SOS	Statistiques d'Ordre Supérieur
TF	Transformée de Fourier
TFD	Transformée de Fourier Discrète
TZ	Transformée en Z
Filtre AR	Filtre AutoRégressif
Filtre MA	Filtre à Moyenne Ajustée
Filtre $ARMA$	Filtre AutoRégressif à Moyenne Ajustée
i.i.d.	indépendant et identiquement distribué
\mathbb{N}	Ensemble des nombres entiers
\mathbb{Z}	Ensemble des nombres entiers relatifs
\mathbb{R}	Ensemble des nombres réels
\mathbb{C}	Ensemble des nombres complexes
\mathbf{x}	Les vecteurs sont distingués en gras
A	Les matrices sont distinguées par des capitales
A^{-1}	Matrice inverse
A^{-}	Matrice pseudo-inverse
z^*, \mathbf{x}^*	Conjugaison complexe
\mathbf{x}^T, A^T	Transposition (sans conjugaison)
$\mathbf{x}^\dagger, A^\dagger$	Transposition et conjugaison
Re	Partie réelle
Im	Partie imaginaire
$\delta(t)$	Distribution de Dirac ou symbole de Kronecker selon le contexte

Variables et vecteurs aléatoires

$E[\cdot]$	Espérance mathématique
$Cum[\cdot]$	Cumulant
$Cum[x, y, z]$	Cumulant des variables x, y, z
$\mu_{\mathbf{x}i}^{jk}$	Moment des variables aléatoires x_i, x_j^*, x_k^*
$\mu'_{x(p)}$	Moment centré d'ordre p de x
$\kappa_{x(2)}^{(2)}$	Cumulant $Cum[x, x, x^*, x^*]$
$\kappa_{\mathbf{x}i\ell}^j$	Cumulant des variables aléatoires x_i, x_j^*, x_ℓ
$\mathcal{K}_{\mathbf{x}i\ell}^{jk}$	Cumulant standardisé de variables aléatoires
$\widehat{k}_{(r)}$	Cumulant estimé par k -statistique

Signaux à valeurs réelles

$C_{x(p)}(\mathbf{t})$ ou $C_{x(p)}(t, \tau)$	Multicorrélation non stationnaire d'ordre p
$C'_{x(p)}(\mathbf{t})$ ou $C'_{x(p)}(t, \tau)$	Multicorrélation non stationnaire des moments d'ordre p
$C_{x(p)}(\tau)$	Multicorrélation d'ordre p
$\Sigma_{x(p)}(\nu)$	Multispectre symétrique
$S_{x(p)}(\nu)$	Multispectre stationnaire
$C'_{x(p)}(\tau)$	Multicorrélation stationnaire des moments
$C_{x(p)}^d(k)$	Diagonale de la multicorrélation
$C_{x(p)}^{\mathbf{l}}(k)$	Coupe de la multicorrélation selon les retards l_2, \dots, l_{p-1}

Signaux à valeurs complexes

$C_{x(p)}^{(q)}(\mathbf{t})$ ou $C_{x(p)}^{(q)}(t, \tau)$	Multicorrélation non stationnaire d'ordre $p + q$
$C'_{x(p)}^{(q)}(\mathbf{t})$ ou $C'_{x(p)}^{(q)}(t, \tau)$	Multicorrélation non stationnaire des moments
$C_{x(p)}^{(q)}(\tau)$	Multicorrélation d'ordre $p + q$
$\mathcal{C}_{x(p)}^{(q)}(\alpha, \tau)$	Multicorrélation cyclique d'ordre $p + q$
$\mathcal{S}_{x(p)}^{(q)}(\alpha, \nu)$	Multicorrélation spectrale d'ordre $p + q$
$\mathcal{S}'_{x(p)}^{(q)}(\alpha, \nu)$	Multicorrélation spectrale des moments
$\widehat{\mathcal{S}}_{x(p)}^{(q)}(\alpha, \nu)$	Multicorrélation spectrale estimée
$P_{x(p)}^{(q)}(\nu)$	Multipériodogramme
$C_{x(p)}^{\ell(q)}(\mathbf{t})$ ou $C_{x(p)}^{\ell(q)}(t, \tau)$	Multicorrélation locale

Index

- Algorithme
 - $C(q, k)$, 125
 - adaptatif, 246
 - de Bussgang, 108
 - de Godard, 113
 - de Levinson, 100
 - de Sato, 113
 - en ligne, 246
 - gradient stochastique, 104
 - LMS, 187, 189
 - par bloc, 246
 - RLS, 188, 190
- Analyse
 - en composantes indépendantes, 238
 - en composantes principales, 238
- Angle
 - de Givens, 250
- Aplatissement, 8
- Asymétrie
 - définition, 8
- Bicepstre, 129
- Bicohérence, 175
 - et filtre linéaire-quadratique, 175
 - symétrie, 176
- Bicorrélation, 36, 264
 - signal à mémoire limitée, 39
 - symétrie, 37
 - variance d'estimation, 91
- Bicorrélogramme, 72
- Bipériodogramme
 - lissé, 72, 77
 - moyenné, 75, 77
 - variance, 89
- Bispectre, 264
- Blancheur
 - à l'ordre 2, 53
 - du bruit, 225
 - parfaite, 53
 - spatiale du bruit, 248
- Blanchiment
 - à l'ordre 2, 248, 251
 - spatial, 235, 248
- Borne
 - de Cramer-Rao, 255
- Bruit blanc, 99
 - à l'ordre q , 99
 - à l'ordre 2, 99
 - gaussien, 99
 - parfait, 99
- Capteurs virtuels, 218
- Causalité, 95
- Centre de phase, 212
- Cepstre, 96
- Circularité, 16, 249
- Coefficients de Fourier-Bohr, 140
- Cohérence, 174
 - et filtre linéaire-quadratique, 174
- Condition de Shannon, 52
 - spatiale, 208
- Contrôle non destructif, 205
- Contraste, 238, 245
- Corrélation, 36
 - signal à mémoire limitée, 39
 - symétrie, 37
- Covariance
 - non linéaire, 220, 227
- Critère
 - d'adaptation, 240, 253
 - de Paley-Wiener, 95
- Crochet
 - notation, 12
- Cumulant, 35
 - additivité, 20

- croisé, 240
- définition, 7
- de variables indépendantes, 19
- estimation, 25
- formule de Leonov-Shiryayev, 13
- invariance par translation, 19
- multilinéarité, 18
- multivariable, 13
- notation, 7
- standardisé, 8, 109
- tenseur, 19
- Cyclostationnarité
 - au sens strict, 142
 - faible, 142
 - forte, 142
- Décomposition en valeurs singulières, 235
- Déconvolution, 103
 - aveugle, 103
- Déphasage
 - entre capteurs, 221
- Développement en série
 - de densités de probabilité, 21
 - de Edgeworth, 21, 22
 - de Gram-Charlier, 21, 244
- Densité de probabilité, 5
- Diagonalisation
 - conjointe, 241
 - de la quadricovariance, 226
- Diaphonie, 255
- Distance
 - à l'indépendance, 246
 - à la loi gaussienne, 246
- Divergence de Kullback, 23
- Diversité
 - spatiale, 237
 - spectrale, 237
 - temporelle, 237
- Echantillonnage, 52
 - condition de Shannon, 52
 - fréquence, 52
 - période, 52
 - spatial, 208
 - théorème, 52
- Egalisation, 103
- Entropie, 23, 109
- Equations
 - aux différences, 94
 - de Yule-Walker, 100
- Equivariance, 253, 255
- Ergodisme, 61, 67
- Espace des observations, 213
- Estimateur
 - de cumulants, 26
 - de cumulants standardisés, 29
 - de l'asymétrie, 29
 - de la bicorrélation, 69
 - de moments, 26
 - du bispectre, 70
 - du kurtosis, 29
 - du maximum de vraisemblance, 244
 - récurif des cumulants, 29
- Filtrage
 - en Z , 59
 - en fréquences réduites, 59
 - linéaire, 49
 - linéaire et invariant, 210
- Filtre
 - à temps discret, 93
 - de mélange, 263
 - de phase, 131
 - de quadrature, 48
 - en treillis, 100
 - linéaire-quadratique, 171, 174
 - non linéaire, 165
 - spatial, 221
- Filtre linéaire
 - AR , 94
 - $ARMA$, 95
 - $BI - AR$, 121
 - MA , 94
 - à phase minimale, 96, 97
 - blanchisseur, 99, 100
 - causalité, 95
 - inverse, 100
 - relations de Bode-Bayard, 97
 - stabilité, 95
- Filtres de Volterra, 62, 98, 165

- à temps continu, 166
- à temps discret, 176
- identification, 182
- modélisation, 173
- soustraction de bruit, 191, 192
- stabilité, 168
- symétrie, 166, 177
- Fonction
 - de vraisemblance, 242
 - caractéristique
 - première, 6
 - seconde, 6
 - de répartition, 5
 - de transfert, 94
 - presque-périodique, 139
- Fonction d'apodisation, 74
 - de Parzen, 75
 - optimale, 75
 - uniforme, 75
- Fonctionnelle du goniomètre
 - d'ordre 2, 225
 - d'ordre 4, 227
- Formation de voies
 - à l'ordre 2, 221
 - à l'ordre 4, 223
- Gain
 - complexe, 94
- Goniomètre, 224
 - à l'ordre 2, 225
 - à l'ordre 4, 226
- Identifiabilité, 104, 125, 212, 233
- Identification aveugle, 100, 103
 - critères, 108
 - en fréquence, 115
 - filtre *AR*, 130
 - filtre *ARMA*, 131
 - filtre *MA*, 124
 - kurtosis, 110
 - multicepstre, 129
 - paramétrique, 121
 - réursive, 104
- Incrément spectral, 42, 46
- Indépendance statistique, 19, 238
- Information de Fisher, 109
- Information mutuelle, 23, 25, 245
- Interférences intersymboles, 255
- Interspectre, 214
- k*-statistiques
 - définition, 27
- Kurtosis, 103, 110
 - définition, 8
 - des sources, 228, 241
- Leonov-Shiryayev
 - formule de, 13
- Limite centrale
 - théorème, 20
- Localisation de sources, 205
- Log-vraisemblance, 243
- Mélange
 - convolutif, 232, 237, 262, 268
 - instantané, 232, 236, 254
 - simple, 233, 268
 - spectral, 232, 237, 249
- Méthode
 - algébrique, 126
 - d'optimisation, 128
 - directe, 254, 256
 - en boucle, 255
- Matrice
 - cumulante, 226, 241
 - de covariance, 11
 - de covariance non linéaire, 220, 228
 - de Givens, 250, 251, 253
 - pseudo-inverse, 234
 - séparatrice, 251, 253
- Matrice spectrale, 237
 - des observations, 214, 225
 - du bruit, 214
 - du signal, 214
- Maximum de vraisemblance, 238, 242
- Modélisation, 98, 133
 - systemique, 133
- Modèle
 - d'observation, 231
 - linéaire, 99
- Moment
 - définition, 6

- estimation, 25
- généralisé, 6
- hybride non linéaire, 220
- multivariable, 13
- Moyenne
 - d'ensemble, 61
 - empirique, 61
- Multicepstre, 129
- Multicorrélation, 35
 - à temps continu, 36
 - à temps discret, 56
 - définition, 36
 - des moments, 37
 - estimation, 67, 69
 - filtrage linéaire, 51
 - filtrage linéaire discret, 60
 - lignes-coupes, 121
 - signal à mémoire limitée, 38
 - signal à valeurs complexes, 39
 - symétrie, 37
- Multicorrélation cyclique, 143
 - propriétés, 145
- Multicorrélation locale, 157
- Multicorrélation spectrale, 144
 - estimation, 145
 - filtrage linéaire, 145
 - propriétés, 145
- Multipériodogramme, 71, 88
- Multiplicité gaussienne, 45
- Multiplicité stationnaire, 44, 46, 48, 59
- Multispectre, 41
 - définition, 43
 - de Wigner-Ville, 161
 - en Z , 58
 - en fréquences réduites, 58
 - estimation, 67, 70
 - filtrage linéaire, 49
 - filtrage linéaire discret, 59
 - interprétation, 47
 - lignes-coupes, 121
 - signal à bande limitée, 53
 - signal à valeurs complexes, 47
 - signal analytique, 48
 - signal circulaire, 49
 - signaux à temps discret, 56
 - spectre diagonal, 121, 123
 - symétrie, 46
- Multispectre symétrique, 45
 - cyclostationnarité, 144
 - signal à valeurs complexes, 48
- MUSIC, 224
- Néguentropie, 24, 245
- Nombre de sources, 213, 215
- Notation
 - crochet, 12
- Onde
 - plane, 211, 217
 - sphérique, 212
- Pôle, 94
- Paramètres
 - de localisation, 221
- Pas
 - d'adaptation, 240, 256
- Phase
 - déroulement, 116
 - minimale, 100
 - nulle, 100
- Point d'équilibre, 240, 256
- Polynômes
 - de Hermite, 22
- Produit
 - de Kronecker, 216, 217, 223, 226
 - tensoriel, 216, 223
- Projection, 235
- Prospection sismique, 115, 120, 205, 209, 258
- Quadricovariance, 216, 223, 226, 240, 241
 - des observations, 216
- Réduction de bruit, 191, 248, 258, 268
- Référence bruit seul, 191, 258
- Réponse impulsionnelle, 93
 - finie, 95
 - infinie, 95
- Réseau
 - de capteurs, 205
 - neuromimétique, 238, 246
 - virtuel, 218

- Règle d'adaptation, 240, 246, 252, 256
- Radar, 205, 209
- Recherche exhaustive, 248
- Relation supplémentaire, 237
- Relations
 - de Bode-Bayard, 97
 - de Kramers-Kronig, 96
- Représentation de Cramér, 41
- Représentation systémique, 93
- Représentation temps-fréquence, 152
 - bilinéaire, 152
 - classe de Cohen, 154
 - de Wigner-Ville, 154
 - spectrogramme, 155
- Représentation temps-multifréquence, 156
 - classe générale, 159
 - de Wigner-Ville, 158
 - estimation, 162
 - multispectrogramme, 160
- Séparateur, 246
- Séparation
 - aveugle, 215
 - de mélanges convolutifs, 237, 262
 - de mélanges instantanés, 236
 - de mélanges simples, 238
 - de mélanges spectraux, 237
- Séparation de sources, 205, 231
 - à l'ordre 2, 236
- Signal
 - à bande limitée, 53
 - à mémoire limitée, 38
 - à temps discret, 52
 - à valeurs complexes, 39
 - aléatoire, 65, 210
 - analytique, 48
 - blanc, 53
 - circulaire, 49
 - composite, 66
 - d'énergie finie, 65
 - déterministe, 61, 209
 - de puissance moyenne finie, 65
 - harmonique, 134
 - linéaire, 100
 - multidimensionnel, 205
 - multivariable, 205
 - périodique, 65
 - sinusoïdal, 10
 - spatio-fréquentiel, 207, 208
 - spatio-temporel, 205, 207
 - transitoire, 65
- Signaux aléatoires
 - cyclostationnaires, 139, 142
 - non stationnaires, 56, 137, 152
 - stationnaires, 35
- Sonar, 194, 205, 209
- Sous-espace bruit, 213, 215, 224
 - d'ordre 4, 217, 226
- Sous-espace signal, 213, 215, 224, 237
 - d'ordre 4, 217, 226
- Soustraction de bruit, 191
- Spectre, 214
 - de ligne, 121
 - diagonal, 121
- Stabilité BIBO, 95, 168
- Standardisation, 16, 247
- Système
 - adaptatif, 246
 - bouclé, 104, 246, 251
 - en ligne, 104
 - linéaire, 49
 - linéaire et stationnaire, 93
 - neuromimétique, 240
 - stable et causal, 96
 - stable, causal, causalement inversible, 96
- Télécommunications, 205
- Tenseur, 19
- Traitement d'antenne, 205
 - aux ordres supérieurs, 230
- Transformée de Fourier
 - discrète, 85
 - filtres de Volterra, 169
- Transformation de Hilbert, 48
- Transformation en Z , 57
- Transformation en fréquences
 - réduites, 57
- Tricorrélation, 36, 266
- Trispectre, 266
- Valeur propre

- dominante, 237
- minimale, 225
- Variété d'antenne, 212
- Variable aléatoire
 - N -PSK, 18, 148
 - à valeurs complexes, 14
 - binaire, 11
 - centrée, 7
 - circulaire, 16
 - circulaire à l'ordre r , 17
 - complexe, 14, 15
 - de Laplace, 9
 - du Chi-deux, 10
 - exponentielle, 9
 - gaussienne, 7, 9
 - gaussienne complexe circulaire, 18
 - gaussienne généralisée, 9
 - indépendance, 19
 - multidimensionnelle, 11
 - Poisson, 10
 - scalaire, 5
 - standardisée, 8, 16
 - uniforme, 8
- Variance
 - d'une variable aléatoire, 8
 - de l'asymétrie estimée, 29
 - du kurtosis estimé, 29
- Vecteur
 - directionnel, 221, 224, 225, 227
 - source, 210, 224
- Vitesse
 - de convergence, 256
- Zéro, 94

Les cinquante dernières années ont vu la naissance du traitement du signal. Cette nouvelle discipline s'inscrit dans la révolution issue de l'émergence de l'information et de la communication. Les systèmes de traitement développés en traitement du signal sont essentiellement fondés sur l'utilisation des statistiques d'ordre 2.

Nous présentons ici le chapitre du traitement du signal qui s'ouvre avec la mise en œuvre de techniques faisant appel aux statistiques d'ordre supérieur à 2. Nous décrivons l'ensemble du nouveau paysage en partant des bases mathématiques conduisant à des outils de description et de caractérisation, et débouchant sur des systèmes de traitement qui viennent compléter les moyens déjà disponibles.

En rassemblant cette matière nous avons voulu permettre aux étudiants, aux chercheurs et aux ingénieurs de s'initier à ce nouveau champ de connaissance qui n'a pas encore livré toutes ses potentialités de recherche et de développement.

Jean-Louis Lacoume, professeur des Universités à l'ENSIEG (Institut National Polytechnique de Grenoble), et **Pierre-Olivier Amblard**, chargé de recherche au CNRS, travaillent au Centre d'Etude des PHénomènes Aléatoires et Géophysique (CEPHAG). **Pierre Comon** est enseignant à EURECOM Sophia-Antipolis et directeur de recherche associé au CNRS.