

**STATISTIQUES D'ORDRE SUPERIEUR POUR
LES SIGNAUX NON GAUSSIENS, NON
LINEAIRES, NON STATIONNAIRES**

Pierre-Olivier Amblard

10 janvier 1994

Table des matières

Introduction	1
1 Préambule	3
1.1 Statistiques des vecteurs aléatoires	3
1.1.1 Vecteurs aléatoires à valeurs réelles	3
1.1.2 Note sur les tenseurs	5
1.1.3 Vecteurs aléatoires à valeurs complexes	6
1.2 Signaux utilisés	6
1.2.1 Signaux déterministes	7
1.2.2 Signaux aléatoires	7
1.3 Estimation adaptative des cumulants	8
1.3.1 Estimation récursive “naturelle” ou “traditionnelle”	8
1.3.2 Un nouvel algorithme	9
1.3.3 Deux commentaires et l’extension au cas multidimensionnel	13
1.4 Conclusion	14
2 Multicorrélations et Multispectres pour les signaux à valeurs complexes	16
2.1 Définition des multicorrélations et des multispectres	16
2.1.1 Multicorrélations	16
2.1.2 Multispectres symétriques	17
2.2 Influence de la propriété de stationnarité	17
2.2.1 Multicorrélations et stationnarité	18
2.2.2 Multispectres symétriques et stationnarité : multispectres	18
2.2.3 Multiplicités gaussiennes. Un autre avantage des cumulants	20
2.3 Multicorrélations et multispectres de signaux particuliers	21
2.3.1 Signaux analytiques	21
2.3.2 Signaux aléatoires circulaires	22
2.3.3 Signaux modulés par une fréquence pure	22
2.3.4 Signaux bande étroite	23
2.3.5 Quelques remarques	24
2.4 Définitions à temps et fréquence discrets	24
2.4.1 Temps discret et fréquence continue	24
2.4.2 Multispectres et multicorrélations des signaux discrets périodiques	25
2.5 Intermulticorrélations et intermultispectres	25
2.6 Multioutils et filtrage linéaire	26
2.7 Conclusion	27
3 Filtres de Volterra	28
3.1 Définitions	28
3.1.1 Définition à temps continu et fréquence continue	29
3.1.2 Validité du développement de Volterra	32
3.1.3 Définition à temps discret	33
3.2 Implantation sous forme vectorielle	34

3.2.1	Implantation dans le domaine temporel	34
3.2.2	Implantation dans le domaine fréquentiel	36
3.3	Identification en Moyenne Quadratique des filtres de Volterra	37
3.3.1	Identification dans le domaine temporel	37
3.3.2	Identification dans le domaine fréquentiel	39
3.3.3	Identification orthogonale dans le domaine temporel	39
3.4	Implantation adaptative des algorithmes	40
3.4.1	Identification adaptative dans le domaine temporel	41
3.4.2	Identification adaptative dans le domaine fréquentiel	42
3.4.3	Complexité numérique des algorithmes RLS	43
3.4.4	Identification orthogonale adaptative dans le domaine temporel	44
3.5	Applications	44
3.5.1	Modélisation et détection des non-linéarités	45
3.5.2	Applications des filtres de Volterra à la Soustraction de Bruit	46
3.6	Conclusion	52
4	Temps-fréquence d'ordre supérieur pour les signaux nonstationnaires	58
4.1	Rappels sur l'ordre deux	58
4.2	Pourquoi étendre aux ordres supérieurs?	61
4.3	Classes de Cohen d'ordre supérieur pour les signaux déterministes	61
4.3.1	Problèmes posés et propriétés souhaitées	61
4.3.2	Construction déductive des classes de Cohen d'ordre supérieur	62
4.3.3	Rôle particulier de la représentation de Wigner-Ville et exemples	68
4.3.4	Propriétés souhaitées et contraintes sur le noyau	70
4.4	Signaux aléatoires : multispectres de Wigner-Ville	72
4.4.1	Multispectres de Wigner-Ville fondés sur les moments	72
4.4.2	Multispectres de Wigner-Ville fondés sur les cumulants	73
4.5	Réduction de complexité	73
4.6	Application à la détection de signaux transitoires	74
4.6.1	Méthode de Persson et Sangfelt	75
4.6.2	Lien avec les multispectres temps-fréquence	75
4.6.3	Extensions naturelles	76
4.6.4	Interprétation en termes d'information	77
4.6.5	Simulations	79
4.7	Conclusion	84
	Conclusion	85
	Bibliographie	87

Principales notations

- les vecteurs sont notés en gras, $\mathbf{t}, \boldsymbol{\nu}$
- $\mathbf{1}$ est le vecteur ne contenant que des 1
- $\mathbf{0}$ est le vecteur ne contenant que des 0
- $E[\cdot]$ est l'opérateur d'espérance mathématique
- $Cum[\cdot]$ est l'opérateur cumulatif
- \mathcal{C} est l'ensemble des complexes
- \mathcal{R} est l'ensemble des réels
- \mathcal{Z} est l'ensemble des entiers relatifs
- $*$ représente la conjugaison complexe
- T dénote le transposé d'un vecteur ou d'une matrice
- \dagger dénote le transposé conjugué d'un vecteur ou d'une matrice
- \Re est la partie réelle d'un nombre complexe
- \Im est la partie imaginaire d'un nombre complexe
- *Arg Min* signifie argument qui minimise
- $[\cdot]$ dénote la partie entière d'un nombre réel

Introduction

Gaussien. Linéaire. Stationnaire.

Trois mots, trois hypothèses qui gouvernent le Traitement du Signal depuis quatre à cinq décennies. La plupart des problèmes sont traités en utilisant des signaux supposés gaussiens, stationnaires et des systèmes linéaires.

Notons que ces hypothèses ont des liens plus ou moins forts entre elles : la gaussianité et la linéarité sont intimement liées, la linéarité s'accommode particulièrement bien avec la stationnarité.

Dans l'estimation de paramètres aléatoires, il peut être montré que la meilleure estimation en moyenne quadratique est l'espérance du paramètre conditionnée par les observations dépendantes du paramètre. Or, lorsque le paramètre suit une loi gaussienne, cette espérance conditionnelle est linéaire en fonction des observations. Autrement dit, la moyenne quadratique qui mesure l'énergie est la norme "adaptée" aux lois gaussiennes. D'autres faits indiquent cette liaison : la filtrée linéaire d'un signal gaussien est encore gaussienne, le mélange linéaire d'une infinité de variables aléatoires indépendantes de même loi est gaussien. Bref, gaussianité et linéarité sont amies.

La troisième hypothèse, la stationnarité, caractérise les signaux. Pour un processus aléatoire, la stationnarité signifie invariance des propriétés statistiques par changement de l'origine des temps. Cette notion s'étend aux signaux déterministes via cette citation de la Thèse de Patrick Flandrin : "un signal certain réel est stationnaire si et seulement si il peut être mis sous la forme d'une somme dénombrable de composantes à amplitude et fréquence instantanées constantes". Si cette définition semble différente de celle donnée pour les signaux aléatoires, elle garde l'idée fondamentale de l'indépendance vis-à-vis du temps de paramètres pertinents de description du signal. Cette indépendance permet d'utiliser un outil puissant, la transformée de Fourier, qui propose d'examiner un signal comme une superposition de sinusoides. Ce changement de représentation conduit à l'analyse spectrale, complément indispensable à toute analyse temporelle. C'est à travers cette analyse fréquentielle que linéarité et stationnarité se rejoignent puisque les opérateurs linéaires et stationnaires (on dit aussi invariants) ont pour fonctions propres des fonctions sinusoidales. Ainsi, une composante fréquentielle d'un signal peut être récupérée par un filtrage linéaire invariant bien choisi, l'idéal étant la sinusoides oscillant à la fréquence que l'on cherche.

Les recherches en Traitement du Signal ces cinquante dernières années ont montré la puissance des hypothèses de gaussianité, linéarité et stationnarité. De nombreux phénomènes naturels ont été correctement décrits en les utilisant ; beaucoup d'appareils aujourd'hui courants utilisent des traitements bâtis autour de ces hypothèses. Par exemple, en communication, il est d'usage de considérer les milieux de transmission linéaires. Les distorsions du signal provoquées par le milieu sont alors compensées par un dispositif (égaliseur) linéaire. Et bien d'autres exemples pourraient être cités pour montrer le bien fondé de ces propriétés.

Mais parfois, une ou toutes ces hypothèses sont inadéquates pour décrire correctement les phénomènes naturels. Nous allons nous intéresser à ces situations, et donc aux contraires de gaussianité, linéarité, stationnarité. Nous parlerons alors de non-gaussianité, non-linéarité et non-stationnarité, ou plus génériquement de non-propriétés.

Si une propriété conduit normalement à une théorie générale, le caractère négatif d'une non-propriété laisse penser qu'il n'existe pas de théorie générale construite autour d'elle. La caractéristique première d'une non-propriété est de ne pas être caractérisée ! Par exemple, la linéarité d'un système stationnaire se traduit par des relations entrée-sortie simples : convolution dans le domaine temporel et produit dans le domaine de Fourier. Par contre, la diversité des systèmes non linéaires "interdit" l'existence de telles relations caractéristiques. Seules des classes d'opérateurs non linéaires peuvent être définies, au sein desquelles tous les membres agissent de la même façon.

Cette constatation est à tempérer pour la non-gaussianité. La qualité de la loi gaussienne est d'être décrite par seulement deux paramètres : ses statistiques d'ordre un et deux, moyenne et variance. Un processus non gaussien est alors décrit par plus de paramètres. Il est d'ailleurs connu qu'une variable aléatoire est entièrement décrite par l'ensemble de ses moments (lorsqu'ils existent). Nous possédons donc une théorie générale pour la description de la non-gaussianité : les Statistiques d'Ordre Supérieur à deux. Connaître l'ensemble des statistiques d'une variable revient à la décrire complètement.

De plus, si gaussianité et linéarité sont intimement liées, le passage à leurs contraires ne casse pas le lien, et non-gaussianité et non-linéarité sont encore fortement liées. En effet, un signal non gaussien peut être vu comme la sortie d'un filtre non linéaire, et filtrer un signal gaussien par un filtre non linéaire conduit nécessairement à un signal non gaussien. De plus, nous pouvons introduire les notions de signal linéaire et de signal non linéaire. Ces notions reposent sur l'interprétation systémique des signaux selon laquelle un signal peut être considéré comme la sortie d'un système excité par un signal blanc. Un signal linéaire sera issu d'un filtre linéaire alors qu'un signal non linéaire sera généré par un système non linéaire. Les statistiques d'ordre supérieur auront donc un rôle important à jouer dans l'étude des systèmes non linéaires.

Par contre, non-stationnarité et non-linéarité ont moins de relations entre elles. Un signal stationnaire peut provenir d'un système non linéaire, et un système linéaire peut produire un signal non stationnaire. L'introduction des statistiques d'ordre supérieur pour la description des signaux non stationnaires est donc moins naturelle que pour les deux autres non-propriétés. Toutefois, les phénomènes naturels sont souvent non gaussiens non stationnaires, et méritent donc une analyse aux ordres supérieurs.

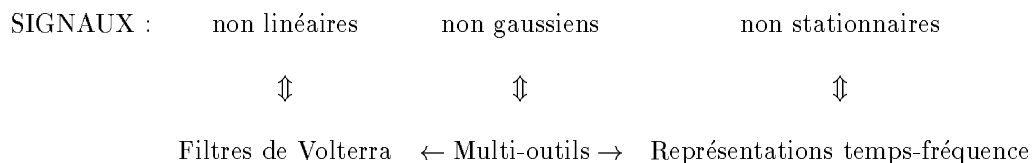
L'objet de ce mémoire est l'étude de la non-gaussianité, de la non-linéarité et de la non-stationnarité à l'aide des statistiques d'ordre supérieur. Cette étude n'est bien sûr pas exhaustive, notre but étant surtout de présenter quelques idées susceptibles d'éclairer ces non-propriétés.

La première partie concerne les signaux non gaussiens stationnaires à valeurs complexes. Nous y définissons des objets adéquats pour la description de tels signaux, objets génériquement appelés multi-outils. Ils seront les multicorrélations lors d'une analyse temporelle et les multispectres lors d'une analyse fréquentielle. Nous étudierons de plus ces objets pour des signaux stationnaires issus de filtres linéaires. La progression se dessine: parmi les trois hypothèses, nous commençons par lever celle de gaussianité.

Dans un deuxième temps, nous oublierons la propriété de linéarité, et les signaux considérés seront alors non gaussiens en général. L'objectif de cette deuxième partie est de présenter la difficulté, mais aussi l'attrait, du non-linéaire à travers une classe particulière de systèmes : les filtres de Volterra. Ces filtres constituent l'extension naturelle des filtres linéaires puisqu'ils sont polynomiaux. Ces systèmes seront présentés dans les domaines temporels et fréquentiels, et appliqués à la méthode de soustraction de bruit.

Enfin, les statistiques d'ordre supérieur pour les signaux non stationnaires feront l'objet du dernier chapitre. Parmi les diverses théories de description des signaux non stationnaires, nous avons choisi d'examiner les représentations temps-fréquence. Ces objets reposent sur l'idée très intuitive de regarder l'évolution au cours du temps du contenu fréquentiel d'un signal. Nous définirons alors à l'aide d'une approche déductive une classe générale de représentations qui constituent l'extension aux ordres supérieurs des représentations temps-fréquence d'ordre deux. Nous proposerons ensuite l'application de ces idées à la détection des signaux transitoires ou brefs.

L'articulation de ce mémoire peut alors être résumée par le schéma suivant.



Mais avant de parcourir ces nouveaux territoires, nous présentons dans un préambule divers outils et techniques dont nous aurons besoin tout au long du chemin.

Chapitre 1

Préambule

Cette partie préliminaire est dédiée à la présentation de tout ce dont nous aurons besoin, statistiques employées (moments et cumulants), types de signaux utilisés Ce chapitre doit donc être considéré comme un bric-à-brac d'outils.

1.1 Statistiques des vecteurs aléatoires

1.1.1 Vecteurs aléatoires à valeurs réelles

Soit \mathbf{X} un vecteur aléatoire à valeurs réelles. Nous supposons que \mathbf{X} est décrit par une densité de probabilité $p_{\mathbf{X}}(\mathbf{x})$ continue et absolument sommable.

Moments

Le moment de \mathbf{X} d'ordre k noté¹ $E[x_1^{i_1} \dots x_n^{i_n}]$ est défini par²

$$E[x_1^{i_1} \dots x_n^{i_n}] = \int x_1^{i_1} \dots x_n^{i_n} p_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \quad (1.1)$$

où $i_1 + \dots + i_n = k$ et où nous supposons que l'intégrale existe pour tout k .

Fonction caractéristique

Soit $\Phi_{\mathbf{X}}(\mathbf{u}) = E[\exp(i\mathbf{u}^T \mathbf{x})]$. En développant l'exponentielle en série entière, il vient

$$\Phi_{\mathbf{X}}(\mathbf{u}) = E\left[\sum \frac{(i\mathbf{u}^T \mathbf{x})^n}{n!}\right] = \sum \frac{i^n}{n!} E[(\mathbf{u}^T \mathbf{x})^n] \quad (1.2)$$

$\Phi_{\mathbf{X}}(\mathbf{u})$ possède la propriété d'être continue et différentiable au voisinage de zéro [59] (de plus, $\Phi_{\mathbf{X}}(\mathbf{0}) = 1$). On obtient alors une propriété caractéristique des moments selon

$$E[x_1^{i_1} \dots x_n^{i_n}] = (-i)^k \left(\frac{\partial^k \Phi_{\mathbf{X}}(\mathbf{u})}{\partial^{i_1} u_1 \dots \partial^{i_n} u_n} \right)_{\mathbf{u}=\mathbf{0}} \quad (1.3)$$

$\Phi_{\mathbf{X}}(\mathbf{u})$ est appelée fonction caractéristique des moments.

¹ $E[\cdot]$ est l'opérateur espérance mathématique.

² Les intégrales multiples sont en général remplacées par un seul signe somme et les éléments différentiels notés sous forme vectorielle, i.e. $\prod_i dx_i = d\mathbf{x}$. De plus, les bornes sont omises lorsque le domaine d'intégration est la droite réelle.

Cumulants

Comme $\Phi_{\mathbf{X}}(\mathbf{0}) = 1$, et comme $\Phi_{\mathbf{X}}(\mathbf{u})$ est continue au voisinage de 0, nous pouvons définir la fonction suivante au voisinage de 0, appelée communément seconde fonction caractéristique, par

$$\Psi_{\mathbf{X}}(\mathbf{u}) = \ln \Phi_{\mathbf{X}}(\mathbf{u}) \quad (1.4)$$

qui est continue et différentiable au voisinage de 0. Tout comme pour les moments, nous différencions $\Psi_{\mathbf{X}}(\mathbf{u})$ en 0, pour définir les cumulants d'ordre k du vecteur \mathbf{X} . Les notations apparaissent sur la formule suivante,

$$Cum[x_1^{i_1}, \dots, x_n^{i_n}] = (-i)^k \left(\frac{\partial^k \Psi_{\mathbf{X}}(\mathbf{u})}{\partial^{i_1} u_1 \dots \partial^{i_n} u_n} \right)_{\mathbf{u}=\mathbf{0}} \quad (1.5)$$

Les propriétés des cumulants seront données après le paragraphe suivant destiné aux relations entre moments et cumulants.

Relations entre moments et cumulants

Au voisinage de zéro

$$\Psi_{\mathbf{X}}(\mathbf{u}) = \ln \Phi_{\mathbf{X}}(\mathbf{u}) \iff \Phi_{\mathbf{X}}(\mathbf{u}) = \exp \Psi_{\mathbf{X}}(\mathbf{u}) \quad (1.6)$$

En utilisant les développements en série de exp et ln, il est possible de montrer par identification terme à terme que les cumulants peuvent s'écrire en fonction des moments d'ordre inférieur ou égal. Cette relation est formellement pour n variables aléatoires a priori distinctes [44, 10, 59, 12]

$$Cum[x_1, \dots, x_n] = \sum_{\forall \nu} (-1)^{q-1} (q-1)! E[\prod_{j \in \nu_1} x_j] \dots E[\prod_{j \in \nu_p} x_j] \quad (1.7)$$

où la sommation s'effectue sur toutes les partitions (ν_1, \dots, ν_p) de $(1, \dots, n)$, et où q est le nombre d'éléments composant la partition. Pratiquement dans cet exposé, ces relations ne seront utilisées que jusqu'à l'ordre 4, soit, pour des vecteurs aléatoires centrés

$$\begin{aligned} Cum[x_1, x_2] &= E[x_1 x_2] \\ Cum[x_1, x_2, x_3] &= E[x_1 x_2 x_3] \\ Cum[x_1, x_2, x_3, x_4] &= E[x_1 x_2 x_3 x_4] - E[x_1 x_2] E[x_3 x_4] - E[x_1 x_3] E[x_2 x_4] - E[x_1 x_4] E[x_2 x_3] \end{aligned} \quad (1.8)$$

Propriétés des moments et des cumulants

Pour les non familiers des cumulants, une question légitime se pose quant à l'intérêt de définir ces statistiques, alors que nous disposons des moments, objets bien connus. Un élément de réponse à cette interrogation est apporté par l'étude des propriétés de ces deux types de statistiques.

- Propriétés des moments.

1. Multilinéarité. Soit λ un vecteur de \mathcal{R}^n . Alors

$$E[(\lambda_1 x_1^{i_1}) \dots (\lambda_n x_n^{i_n})] = \lambda_1 \dots \lambda_n E[x_1^{i_1} \dots x_n^{i_n}] \quad (1.9)$$

Cette propriété permet alors de montrer que les moments sont des tenseurs sur le groupe linéaire (voir la note sur les tenseurs au paragraphe suivant).

2. Si $\{x_1, \dots, x_n\} = \{x_1, \dots, x_j\} \cup \{x_{j+1}, \dots, x_n\}$ où les deux sous-ensembles sont indépendants, alors

$$E[x_1^{i_1} \dots x_n^{i_n}] = E[x_1^{i_1} \dots x_j^{i_j}] E[x_{j+1}^{i_{j+1}} \dots x_n^{i_n}] \quad (1.10)$$

Ces propriétés se déduisent directement de la définition des moments.

- Propriétés des cumulants.

1. Multilinéarité et invariance par translation. Soient λ et α deux vecteurs de \mathcal{R}^n . Alors

$$Cum[\alpha_1 + \lambda_1 x_1^{i_1}, \dots, \alpha_n + \lambda_n x_n^{i_n}] = \lambda_1 \dots \lambda_n Cum[x_1^{i_1}, \dots, x_n^{i_n}] \quad (1.11)$$

Ceci permet de montrer que les cumulants sont des tenseurs sur le groupe affine.

2. Si $\{x_1, \dots, x_n\} = \{x_1, \dots, x_j\} \cup \{x_{j+1}, \dots, x_n\}$ où les deux sous-ensembles sont indépendants, alors

$$Cum[x_1^{i_1}, \dots, x_n^{i_n}] = 0 \quad (1.12)$$

3. Si $\mathbf{X} = \mathbf{Y} + \mathbf{Z}$ où \mathbf{Y} et \mathbf{Z} sont indépendants, alors

$$Cum[x_1^{i_1}, \dots, x_n^{i_n}] = Cum[z_1^{i_1}, \dots, z_n^{i_n}] + Cum[y_1^{i_1}, \dots, y_n^{i_n}] \quad (1.13)$$

4. Les cumulants d'ordre supérieur ou égal à 3 d'un vecteur aléatoire gaussien sont nuls (ceci provient du fait que la fonction génératrice des moments est l'exponentielle d'une forme quadratique, et par suite, la seconde fonction caractéristique génératrice des cumulants se réduit à une forme quadratique).

En examinant les propriétés précédentes, un ensemble d'éléments montre l'intérêt des cumulants par rapport aux moments. Ils permettent de mesurer l'indépendance et la gaussianité. Ils ont la propriété d'être "cumulant" pour la somme de deux vecteurs indépendants.

Ces propriétés sont à la base de l'utilisation maintenant intensive de ces statistiques. Notons que la covariance est le cumulants d'ordre 2 et que cet outil est utilisé depuis bien longtemps pour ses propriétés. La covariance de la somme de variables aléatoires indépendantes est la somme de leurs covariances (cumul). La covariance est nulle si les variables aléatoires sont décorréélées (premier degré d'indépendance).

1.1.2 Note sur les tenseurs

Les tenseurs constituent l'extension multidimensionnelle de la notion de vecteur. Nous donnons ici quelques idées sur leurs définitions dans le cas des tenseurs d'ordre deux, la généralisation à des ordres plus élevés ne posant pas de difficulté. Soient U et V deux espaces vectoriels sur les réels, d'espaces duaux respectifs U^* et V^* . Soient e_i et f_j deux bases de U et V , dont les bases duales seront notées e^i et f^j et vérifient $\langle e^i | e_k \rangle = \delta_k^i$ et $\langle f^j | f_l \rangle = \delta_l^j$, δ étant le symbole de Kronecker.

Le produit tensoriel entre U et V , noté $U \otimes V$, est l'ensemble des applications bilinéaires de $U \times V$ dans \mathcal{R} . Soient alors (α, β) deux éléments de U^* et V^* de coordonnées α_i et β_j , et soit $T \in U \otimes V$. Alors par définition

$$T(\alpha, \beta) = \sum_{i,j} \alpha_i \beta_j T(e^i, f^j) \quad (1.14)$$

On montre que $\alpha_i \beta_j$ forme une base de $U \otimes V$. Ce terme est noté $e_i \otimes f_j$.

Soient alors u et v des éléments de U et V de coordonnées u^i et v^j . Le produit tensoriel entre u et v , qui appartient à $U \otimes V$ est alors défini par

$$u \otimes v(\alpha, \beta) = \langle u | \alpha \rangle \langle v | \beta \rangle = \sum_{i,j} u^i v^j e_i \otimes f_j \quad (1.15)$$

Le tenseur $u \otimes v$ a alors pour coordonnées $u^i v^j$ dans la base $e_i \otimes f_j$.

Une caractérisation des tenseurs est apportée par l'étude des changements de bases. Supposons que $e_i = \sum_k A_i^k e'_k$ et $f_j = \sum_l B_j^l f'_l$, et que dans $e_i \otimes f_j$ les coordonnées de $u \otimes v$ soient T^{ij} , alors

$$u \otimes v = \sum_{i,j} T^{ij} e_i \otimes f_j = \sum_{k,l} \sum_{i,j} T^{ij} A_i^k B_j^l e'_k \otimes f'_l \quad (1.16)$$

Si T'^{kl} sont les coordonnées de $u \otimes v$ dans $e'_k \otimes f'_l$, alors $T'^{kl} = \sum_{i,j} T^{ij} A_i^k B_j^l$. La caractérisation tensorielle suggère alors que tout tableau de nombres se transformant comme précédemment par ces changements de

bases est un tenseur de $U \otimes V$. Il est alors facile de vérifier que les moments se transforment ainsi lorsque le changement de base est linéaire, et que les cumulants réagissent de même si le changement est affine. Ceci justifie les affirmations énoncées précédemment.

Remarque : la construction qui vient d'être faite peut se généraliser à des ordres plus élevés, mais également en envisageant des produits tensoriels mixtes, c'est-à-dire des produits entre un espace vectoriel et son dual.

1.1.3 Vecteurs aléatoires à valeurs complexes

Deux approches possibles peuvent être employées pour la présentation des vecteurs aléatoires à valeurs complexes [18, 43]. La première réside dans une identification de \mathcal{C}^n à \mathcal{R}^{2n} . En effet, un vecteur aléatoire \mathbf{Z} de \mathcal{C}^n pouvant s'écrire $\mathbf{Z} = \mathbf{X} + i\mathbf{Y}$, \mathbf{Z} peut être vu comme un vecteur $(\mathbf{X}^T, \mathbf{Y}^T)^T$ de \mathcal{R}^{2n} . La deuxième approche consiste à écrire toutes les expressions en utilisant la variable complexe. Nous mélangerons les deux approches. En effet, une variable complexe \mathbf{Z} admet une densité $p_{\mathbf{Z}}(\mathbf{z})$ si ses parties réelle et imaginaire possèdent une densité conjointe. Nous supposons qu'il en est ainsi, et les définitions des statistiques d'un vecteur aléatoire complexe se déduisent immédiatement des résultats obtenus à la section précédente. Toutefois, une propriété particulière des variables complexes nécessite la précision de certaines définitions.

Soit donc \mathbf{Z} un vecteur aléatoire de \mathcal{C}^n et $p_{\mathbf{Z}}(\mathbf{z})$ sa densité de probabilité. Sa fonction caractéristique est définie par

$$\Phi_{\mathbf{Z}}(\mathbf{u}) = E[\exp(i \frac{(\mathbf{u}^\dagger \mathbf{z} + \mathbf{z}^\dagger \mathbf{u})}{2})] \quad (1.17)$$

où \mathbf{u} est un vecteur de \mathcal{C}^n . Nous voyons que la description d'une variable aléatoire à valeurs complexes Z fait intervenir z et z^* . En effet, bien que liées, ces deux grandeurs apportent toutes les deux de l'information, et doivent être considérées comme deux variables "indépendantes". Par exemple, la classique variable aléatoire gaussienne complexe circulaire Z est orthogonale à Z^* , puisque $E[z(z^*)^*] = 0$. Cette propriété de circularité est maintenant présentée en détails.

Vecteurs aléatoires complexes circulaires

\mathbf{Z} un vecteur aléatoire de \mathcal{C}^n possédant une densité $p_{\mathbf{Z}}(\mathbf{z})$ est circulaire si $\mathbf{Z}e^{i\theta}$ a même densité de probabilité. Une caractérisation de la circularité se déduit immédiatement de cette définition.

\mathbf{Z} est circulaire si ses moments (ou cumulants) ayant un nombre différent d'arguments conjugués et non-conjugués sont nuls [18, 43], soit

$$E[z_1^{i_1} \dots z_n^{i_n} z_1^{*j_1} \dots z_n^{*j_n}] = 0 \quad (1.18)$$

si $i_1 + \dots + i_n \neq j_1 + \dots + j_n$

La notion de circularité (comme la stationnarité) peut alors être définie à un certain ordre : il suffit de restreindre la caractérisation précédente à n'être vraie que jusqu'à cet ordre.

L'intérêt des variables aléatoires complexes circulaires apparaît par exemple en traitement d'antenne, où les signaux bande étroite vérifient la caractérisation (1.18). D'autre part, la transformée de Fourier d'un signal aléatoire stationnaire³ d'ordre r , est circulaire d'ordre r , sauf pour la fréquence nulle [18].

La partie concernant les définitions nécessaires se termine, et nous entamons une section concernant les signaux manipulés ici.

1.2 Signaux utilisés

Dans tout ce mémoire, différents outils seront considérés, comme les cumulants ou les filtres de Volterra. Ces grandeurs seront utilisées comme descripteurs ou opérateurs sur des signaux aléatoires ou déterministes, et nous devons donc présenter les classes de signaux que nous manipulerons.

³Un signal est stationnaire d'ordre r si ses statistiques d'ordre inférieur ou égal à r sont invariantes par translation temporelle.

1.2.1 Signaux déterministes

Les signaux déterministes interviendront dans la construction des représentations temps-fréquence d'ordre supérieur et dans la partie relative à la définition des filtres de Volterra. Les propriétés dont nous aurons besoin résident essentiellement dans les possibilités d'intégration. Nous utiliserons donc les espaces $L_C^p(\mathcal{R})$ définis par⁴

$$L_C^p(\mathcal{R}) = \{x(t) \ / \ ||x||_p = \left(\int |x(t)|^p dt\right)^{1/p} < +\infty\} \quad (1.19)$$

Nous pourrions par exemple choisir un signal dans un de ces espaces pour que la sortie d'un filtre de Volterra soit d'énergie finie, c'est-à-dire appartienne à $L^2(\mathcal{R})$.

Remarque : les espaces précédents sont des espaces de Banach pour les normes associées [59]. De plus, l'espace $L_C^2(\mathcal{R})$ est Hilbertien car la norme $||\cdot||_2$ "dérive" du produit scalaire $\langle x|y \rangle = \int x(t)y^*(t)dt$.

1.2.2 Signaux aléatoires

Cette partie est essentiellement issue de [9]. Les processus $x(t)$ auxquels nous nous intéressons sont ceux pouvant être mis sous la forme

$$x(t) = \int \exp(2i\pi\nu t) dX(\nu) \quad (1.20)$$

où l'intégrale est une intégrale stochastique de Stieljes-Riemann. Cette représentation est aussi dite de Crâmer. Blanc-Lapierre et Fortet montrent le résultat suivant :

$x(t)$ admet la décomposition harmonique $x(t) = \int \exp(2i\pi\nu t) dX(\nu)$ en moyenne quadratique si et seulement si :

$$E[x(t_1)x^*(t_2)] = \iint \exp(2i\pi(t_1\nu_1 - t_2\nu_2)) d^2\gamma(\nu_1, \nu_2) \quad (1.21)$$

où $d^2\gamma(\nu_1, \nu_2) = E[dX(\nu_1)dX(\nu_2)]$ vérifie

$$\iint |d^2\gamma(\nu_1, \nu_2)| < C < +\infty \quad (1.22)$$

C étant une constante. Le processus $x(t)$ est alors dit harmonisable. Un exemple d'utilisation de ces signaux peut se trouver dans [29, 47] où les auteurs montrent que ces signaux possèdent un spectre de Wigner-Ville. 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 [9].

On dit que $X(\nu) \in \Phi^{(k)}$ si $\forall l < k, \forall \varepsilon_1, \dots, \varepsilon_l$

$$\int \dots \int |E[dX_{\varepsilon_1}(\nu_1) \dots dX_{\varepsilon_l}(\nu_l)]| < C < +\infty \quad (1.23)$$

(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 [9])

Si $x(t) \in f^{(k)}$, si $k = l(l-1)$, alors

$$\forall j \leq l \quad x_{\varepsilon_1}(t_1) \dots x_{\varepsilon_j}(t_j) = \int \dots \int \exp[2i\pi(\sum_i \varepsilon_i t_i \nu_i)] dX_{\varepsilon_1}(\nu_1) \dots dX_{\varepsilon_j}(\nu_j) \quad (1.24)$$

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

$$E[x_{\varepsilon_1}(t_1) \dots x_{\varepsilon_j}(t_j)] = \int \dots \int \exp[2i\pi(\sum_i \varepsilon_i t_i \nu_i)] \cdot E[dX_{\varepsilon_1}(\nu_1) \dots dX_{\varepsilon_j}(\nu_j)] \quad (1.25)$$

Ce résultat nous sera très utile pour la définition de multispectres temps-fréquence.

Remarque 1 : la classe $f^{(2)}$ correspond à l'ensemble des signaux harmonisables.

Remarque 2 : la classe $f^{(\infty)}$ est la classe des signaux appartenant à $f^{(k)}$ pour tout k .

Remarque 3 : les définitions précédentes ont été données initialement sur les moments. Elles pourraient être également présentées à l'aide des cumulants.

⁴Lorsque les signaux sont à valeurs réelles, la notation se simplifie en $L^p(\mathcal{R})$.

1.3 Estimation adaptative des cumulants

Nous aurons besoin pour l'implantation numérique d'une idée que nous proposons d'un estimateur récursif du tenseur cumulant d'ordre 4. L'objet de cette partie est l'introduction d'un nouvel estimateur récursif. Nous présenterons les idées sur le cumulant d'ordre 4 d'un processus aléatoire blanc au sens fort⁵, avant de montrer la forme de l'estimateur des tenseurs cumulants⁶.

1.3.1 Estimation récursive “naturelle” ou “traditionnelle”

Considérons $x(n)$ un processus aléatoire discret à valeurs dans \mathcal{R} , blanc au sens fort, c'est-à-dire dont les échantillons sont indépendants. La relation liant le cumulant d'ordre 4 aux moments d'ordres inférieurs ou égaux s'écrit

$$C_4 = \text{Cum}[x(n)^4] = M_4 - 3M_2^2 = E[x(n)^4] - 3E[x(n)^2]^2 \quad (1.26)$$

La manière naturelle d'obtenir un estimateur de C_4 est de remplacer dans l'expression précédente les moments par des estimateurs. Cette remarque est valide que l'on veuille un estimateur récursif ou non. Pour la forme récursive qui nous intéresse, les estimateurs des moments d'ordre 2 et 4 sont

$$\begin{cases} \widehat{M}_{2,k} &= \widehat{M}_{2,k-1} + \mu(x(k)^2 - \widehat{M}_{2,k-1}) \\ \widehat{M}_{4,k} &= \widehat{M}_{4,k-1} + \gamma(x(k)^4 - \widehat{M}_{4,k-1}) \end{cases} \quad (1.27)$$

Un estimateur récursif du cumulant d'ordre 4 suit selon

$$\widehat{C}_{4,k} = \widehat{M}_{4,k} - 3\widehat{M}_{2,k}^2 \quad (1.28)$$

Dans les algorithmes (1.27), les pas μ et γ sont choisis constants pour pouvoir suivre des non-stationnarités lentes. En cas stationnaire, il suffit de les remplacer par des pas μ_k et γ_k , donc dépendants du temps, qui tendent vers zéro quand k tend vers l'infini.

L'algorithme (1.28) est alors un algorithme asymptotiquement non biaisé, mais dont la variance asymptotique est difficilement calculable. Toutefois un fait évident apparaît sur (1.28) : la variance d'estimation de $\widehat{M}_{2,k}$ participe par son carré à la variance de $\widehat{C}_{4,k}$. Nous dirons que l'estimation du cumulant d'ordre quatre est polluée par la variance d'estimation du moment d'ordre deux.

Or, la variance de l'estimateur récursif de ce moment d'ordre 2 est essentiellement contrôlée ici par le paramètre μ . Une étude expérimentale de la variance asymptotique de (1.28) a été effectuée en fonction de μ . Le résultat est montré figure (1.1). Il apparaît que pour $\gamma = 0.001$, lorsque μ varie de 10^{-3} à 0.5, la variance asymptotique d'estimation du cumulant passe de 0.1 à 100.

La conclusion de cette petite étude est la non robustesse de l'algorithme (1.28) vis-à-vis du paramètre μ .

De plus, un problème différent concerne (1.28) et réside dans l'impossibilité d'écrire $\widehat{C}_{4,k}$ sous la forme traditionnelle des algorithmes adaptatifs, soit par exemple [8]

$$\theta_k = \theta_{k-1} + \mu H(\theta_{k-1}, x(k)) \quad (1.29)$$

où θ est le paramètre à estimer et H une fonction “adaptée”, quantifiant l'erreur instantanée d'estimation. L'impossibilité d'écrire l'algorithme traditionnel d'estimation du cumulant sous cette forme explique les difficultés d'étude de l'algorithme, et surtout les difficultés d'obtention de la variance asymptotique.

Toutefois, en essayant d'écrire (1.28) sous la forme (1.29), c'est-à-dire en insérant (1.27) dans (1.28), et en effectuant quelques manipulations algébriques, un terme simple apparaît, $x(k)^4 - 3x(k)^2\widehat{M}_{2,k-1}$, que nous appellerons cumulant instantané. Ce terme est le pilier du nouvel algorithme que nous présentons et étudions maintenant.

⁵ Pour les notions de blancheur, voir [11].

⁶ Ce travail a été effectué en collaboration avec J.M. Brossier.

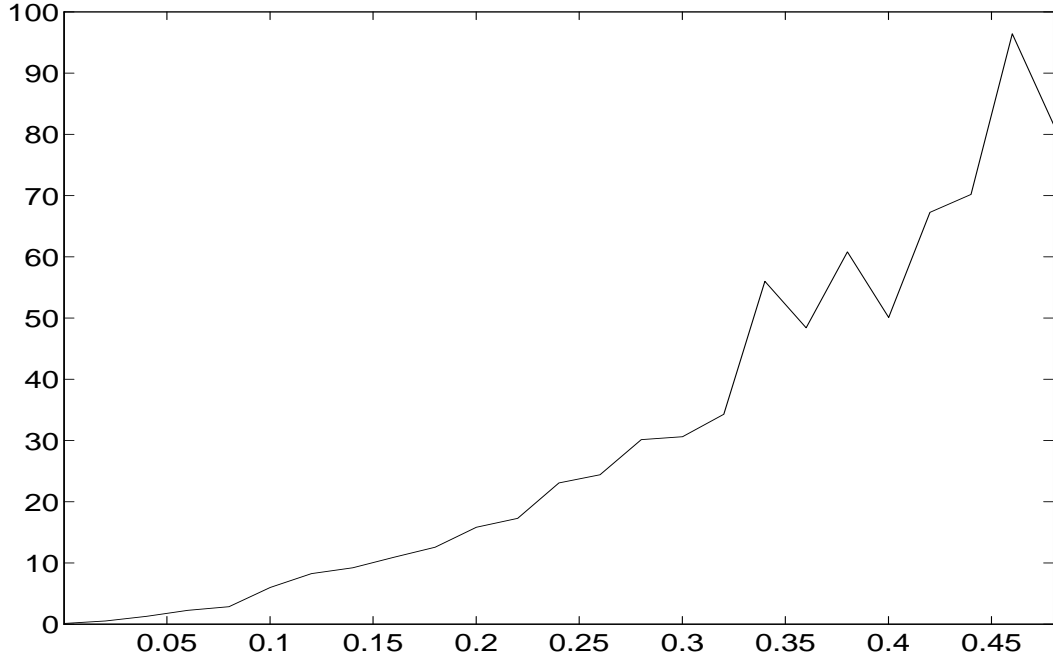


Figure 1.1: Variance expérimentale asymptotique de l'algorithme traditionnel du cumulante en fonction de μ . $\gamma = 0.001$

1.3.2 Un nouvel algorithme

Le terme introduit précédemment nous permet de construire une erreur instantanée d'estimation, soit $x(k)^4 - 3x(k)^2\widehat{M}_{2,k-1} - \widehat{C}_{4,k-1}$, et nous obtenons un algorithme du type (1.29) selon

$$\begin{cases} \widehat{M}_{2,k} &= \widehat{M}_{2,k-1} + \mu(x(k)^2 - \widehat{M}_{2,k-1}) \\ \widehat{C}_{4,k} &= \widehat{C}_{4,k-1} + \gamma(x(k)^4 - 3x(k)^2\widehat{M}_{2,k-1} - \widehat{C}_{4,k-1}) \end{cases} \quad (1.30)$$

Nous n'évitons pas l'estimation du moment d'ordre 2, mais estimons directement le cumulante à l'aide d'un algorithme du "type gradient", où l'estimation est corrigée à chaque instant par un petit saut proportionnel à une quantité chiffrant l'erreur instantanée d'estimation. La suite de ce paragraphe est dédiée à l'étude théorique de cet algorithme en utilisant les outils développés dans [8], soit les techniques O.D.E. (Ordinary Differential Equation) pour la convergence, et les approximations par processus de diffusion pour le comportement asymptotique. Nous ne ferons ici qu'appliquer des théorèmes compliqués justifiés dans [8].

Convergence de l'algorithme : l'O.D.E.

La méthode de l'O.D.E. associée à l'algorithme adaptatif une équation différentielle ordinaire dont la solution reflète le comportement moyen de l'algorithme. Pour l'estimateur (1.30), cette équation s'écrit

$$\begin{cases} \frac{dM_2(s)}{ds} &= E[x(k)^2 - M_2(s)] \\ \frac{dC_4(t)}{dt} &= E[x(k)^4 - 3x(k)^2M_2(\frac{t}{\gamma}) - C_4(t)] \end{cases} \quad (1.31)$$

où les temps continus s et t sont respectivement reliés aux temps discrets de l'algorithme via les relations $s_k = k\mu$ et $t_k = k\gamma$. L'introduction de ces deux temps est nécessaire puisque les deux algorithmes définissant

(1.30) ont des constantes de temps différentes. Toutefois, ces temps ne sont pas indépendants, et l'on a $s_k = \frac{\mu}{\gamma} t_k$.

Avant de résoudre (1.31), précisons les hypothèses de travail. Le processus dont on estime le cumulants est un processus blanc au sens fort, stationnaire pour l'étude, centré, symétrique, dont la variance sera notée σ^2 et le cumulants d'ordre 4 c_4 .

L'évaluation des espérances dans (1.31) conduit au système d'équations du premier ordre

$$\begin{cases} \frac{dM_2(s)}{ds} + M_2(s) = \sigma^2 \\ \frac{dC_4(t)}{dt} + C_4(t) = c_4 + 3\sigma^4 - 3\sigma^2 M_2\left(\frac{\mu}{\gamma}t\right) \end{cases} \quad (1.32)$$

Définissons le paramètre $\eta = \frac{\mu}{\gamma}$. La résolution de ce système conduit, pour $\eta \neq 1$, à

$$\begin{cases} M_2(s) - \sigma^2 = (M_2(0) - \sigma^2)e^{-s} \\ C_4(t) - c_4 = (C_4(0) - c_4)e^{-t} + \frac{3\sigma^2(M_2(0) - \sigma^2)}{1 - \eta}(e^{-t} - e^{-\eta t}) \end{cases} \quad (1.33)$$

Lorsque $\eta = 1$, soit $\mu = \gamma$, la solution pour $C_4(t)$ est

$$C_4(t) - c_4 = (C_4(0) - c_4)e^{-t} - 3\sigma^2(M_2(0) - \sigma^2)te^{-t} \quad (1.34)$$

Remarque : cette solution peut s'obtenir à partir de la solution pour $\eta \neq 1$ en étudiant la limite de $C_4(t)$ quand η tend vers 1.

Commentaires :

- La figure (1.2) représente la solution $C_4(t)$ pour diverses valeurs de η . Quand η augmente les trajectoires deviennent très similaires. Ainsi, à γ fixé, augmenter μ n'a pas d'influence sur la vitesse de convergence de l'algorithme. Donc, la vitesse de convergence est essentiellement réglée par γ . Sur ce point, l'estimation adaptative du cumulants se comporte comme un algorithme classique.
- Remarquons le comportement "étrange" de $C_4(t)$ lors du transitoire. Une bosse apparaît, et il semble que l'algorithme attende la convergence du moment d'ordre 2 pour faire converger le cumulants. Cette constatation est confirmée par le resserrement de la bosse quand η augmente.
- La figure (1.3) montre la superposition d'une solution de l'O.D.E. et de la trajectoire correspondante de l'algorithme. Cette figure illustre le bon comportement asymptotique moyen de l'algorithme : il est non-biaisé. Ceci se démontre à partir de (1.33) puisqu'en effet

$$\lim_{t \rightarrow \infty} C_4(t) - c_4 = 0 \quad (1.35)$$

Normalité asymptotique

D'après le théorème 2 de la page 103 de [8], $\frac{1}{\sqrt{\gamma}}(\widehat{C}_{4,k} - c_4)$ est gaussien centré asymptotiquement quand γ tend vers zéro. De plus, la variance à l'infini est donnée par $\frac{1}{2}R(c_4)$ où

$$R(c_4) = \sum_{k=-\infty}^{+\infty} Cov[H_k(c_4), H_0(c_4)] \quad (1.36)$$

où $H_k(c_4) = x(k)^4 - 3x(k)^2 \widehat{M}_{2,k-1} - c_4$ et $Cov[\cdot]$ est l'opérateur de covariance. Pratiquement, la variance de $\widehat{C}_{4,k}$ est donnée par $\frac{2}{\gamma}R(c_4)$. Nous présentons maintenant les étapes du calcul qui conduisent à l'expression de $R(c_4)$. Rappelons que le processus étudié est blanc à échantillons identiquement distribués.

1. $H_k(c_4)$ est centré.

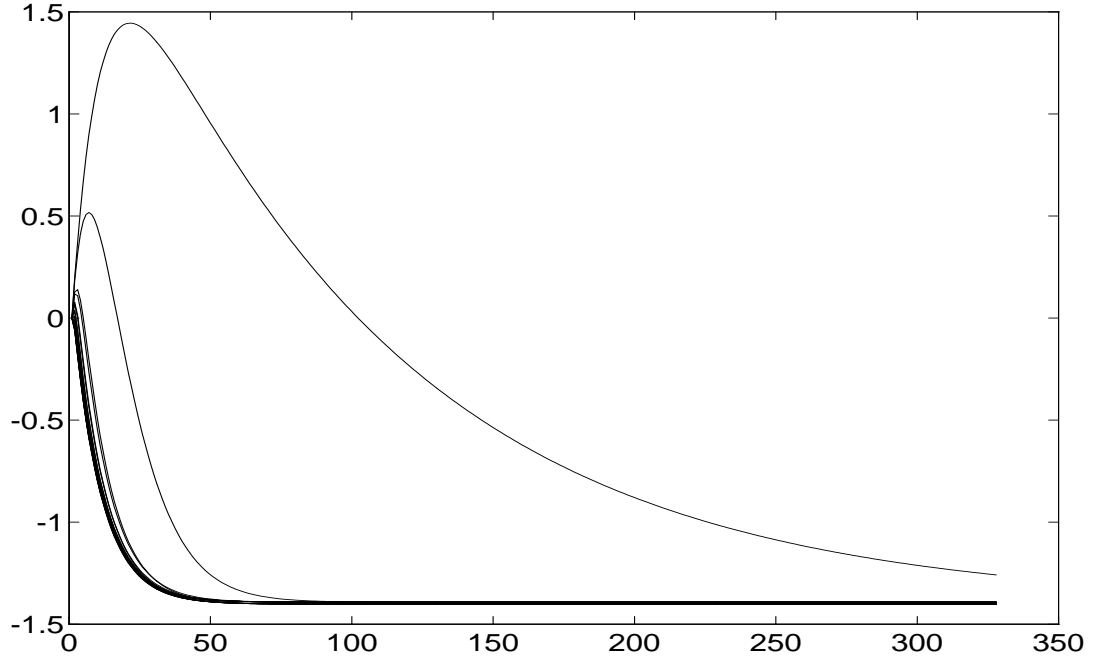


Figure 1.2: Tracé de diverses solutions de l'O.D.E. pour η allant de 10^{-4} à 0.5. Les courbes de droite à gauche correspondent à η croissant.

2. Evaluation de la covariance.

$$\text{Cov}[H_k(c_4), H_0(c_4)] = E[x_k^4 x_0^4] - 3E[x_k^4 x_0^2 M_{2,-1}] - 3E[x_0^4 x_k^2 M_{2,k-1}] + 9E[x_k^2 x_0^2 M_{2,k-1} M_{2,-1}] - c_4^2 \quad (1.37)$$

3. $M_{2,k-1}$ est remplacé par $\mu \sum_{j=-\infty}^{k-1} (1-\mu)^{k-1-j} x_j^2$. Les cas $k > 0$, $k = 0$ et $k < 0$ doivent être distingués. Nous présenterons ici les deux premiers uniquement, le cas $k < 0$ se traitant de la même façon que le cas $k > 0$.

4. Cas $k > 0$.

$$E[x_k^4 x_0^2 M_{2,-1}] = E[x^4] E[x^2]^2 \quad (1.38)$$

car $k > 0$ et car le processus est blanc et identiquement distribué.

$$E[x_0^4 x_k^2 M_{2,k-1}] = \mu E[x^2] \sum_{j=-\infty}^{k-1} (1-\mu)^{k-1-j} E[x_j^2 x_0^4] \quad (1.39)$$

Dans cette expression, $E[x_j^2 x_0^4]$ vaut $E[x^6]$ si $j = 0$ et $E[x^2] E[x^4]$ sinon. Par suite

$$E[x_0^4 x_k^2 M_{2,k-1}] = E[x^2] (\mu (1-\mu)^{k-1} E[x^6] + E[x^2] E[x^4] (1 - \mu (1-\mu)^{k-1})) \quad (1.40)$$

Le terme suivant se calcule selon

$$E[x_k^2 x_0^2 M_{2,k-1} M_{2,-1}] = \mu^2 E[x^2] \sum_{l=-\infty}^{k-1} \sum_{j=-\infty}^{k-1} (1-\mu)^{k-1-l} (1-\mu)^{-1-j} E[x_0^2 M_{2,k-1} M_{2,-1}] \quad (1.41)$$

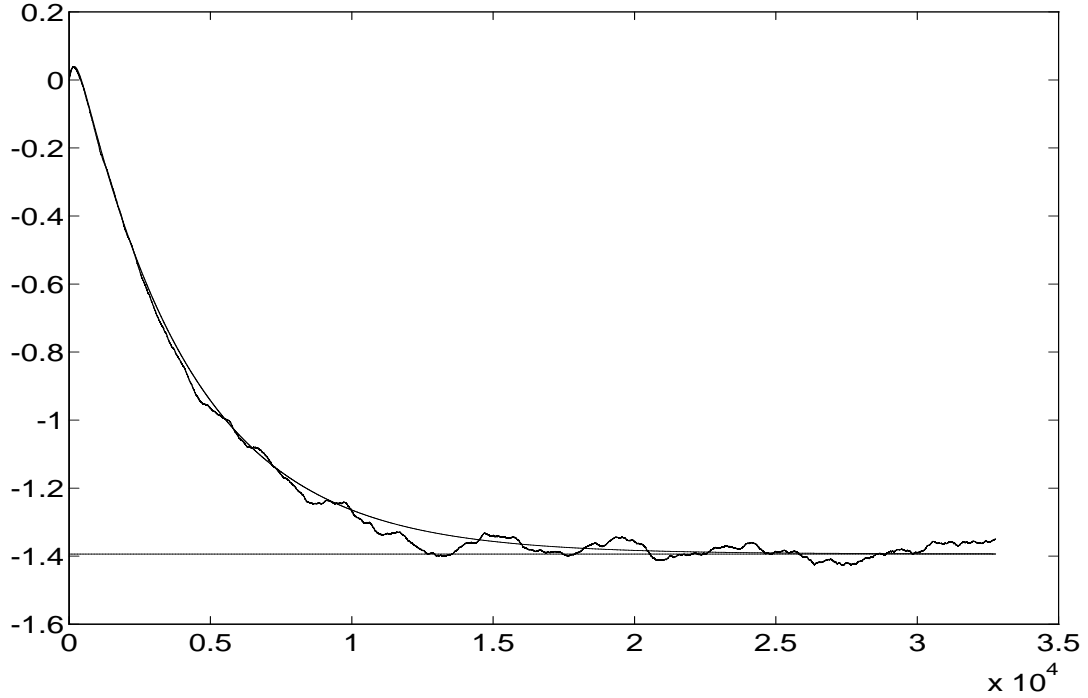


Figure 1.3: Superposition de la trajectoire de l’algorithme et de la solution de l’O.D.E. pour $\mu = 0.005$ et $\gamma = 0.00025$.

La sommation en l est découpée en trois : $l = -\infty, \dots, -1$, $l = 0$ et $l = 1, \dots, k-1$. Après avoir distingué le cas $l = j$ du cas $l \neq j$, on obtient

$$E[x_k^2 x_0^2 M_{2,k-1} M_{2,-1}] = \mu^2 E[x^2].$$

$$(E[x^2]^3 \left(\frac{(1-\mu)^k}{\mu^2} - \frac{(1-\mu)^k}{\mu(2-\mu)} - \frac{(1-\mu)^{k-1} - 1}{\mu^2} \right) + E[x^2] E[x^4] \left(\frac{(1-\mu)^k}{\mu(2-\mu)} + \frac{(1-\mu)^{k-1}}{\mu} \right)) \quad (1.42)$$

5. Cas $k = 0$. La difficulté dans ce cas réside dans l’évaluation de $E[M_{2,-1}^2]$. La double somme qui intervient est alors coupée en deux : les “auto”-termes et les termes croisés. Le calcul s’effectue alors assez simplement.
6. Regroupement des termes. Le cas $k < 0$ conduit pour sa contribution à la covariance au même résultat que la partie due à $k > 0$. En sommant les différents termes puis en laissant k tendre vers l’infini, le résultat final est obtenu.

L’expression de $R(c_4)$ est alors

$$R(c_4) = E[x(n)^8] - 12\sigma^2 E[x(n)^6] + 6\sigma^4 E[x(n)^4] \frac{16 - 11\mu}{2 - \mu} + E[x(n)^4]^2 \frac{-2 - 10\mu}{2 - \mu} - 9\sigma^8 \frac{8 - 5\mu}{2 - \mu} \quad (1.43)$$

Dans le cas gaussien⁷, nous représentons sur la figure (1.4) l’évolution de $R(c_4)$ en fonction de μ . Cette figure est à comparer avec la figure (1.1) précédemment décrite. Ici, lorsque μ passe de 0 à 1, la variance asymptotique est multipliée par un facteur 2, alors qu’il était de 1000 pour l’algorithme traditionnel. Ainsi, le pas utilisé pour l’estimation de l’ordre 2 n’a plus d’influence dans notre algorithme sur la variance d’estimation du cumulatif. C’est en ce sens que nous dirons que l’algorithme (1.30) est robuste au choix de μ .

⁷Présenter l’évolution de $R(c_4)$ en fonction de μ dans le cas gaussien peut paraître curieux, puisque la grandeur à estimer est dans ce cas 0. Toutefois, les cumulants sont aujourd’hui souvent utilisés pour éliminer des bruits gaussiens, et cette étude peut alors être vue comme une étude de performances d’élimination de bruit gaussien.

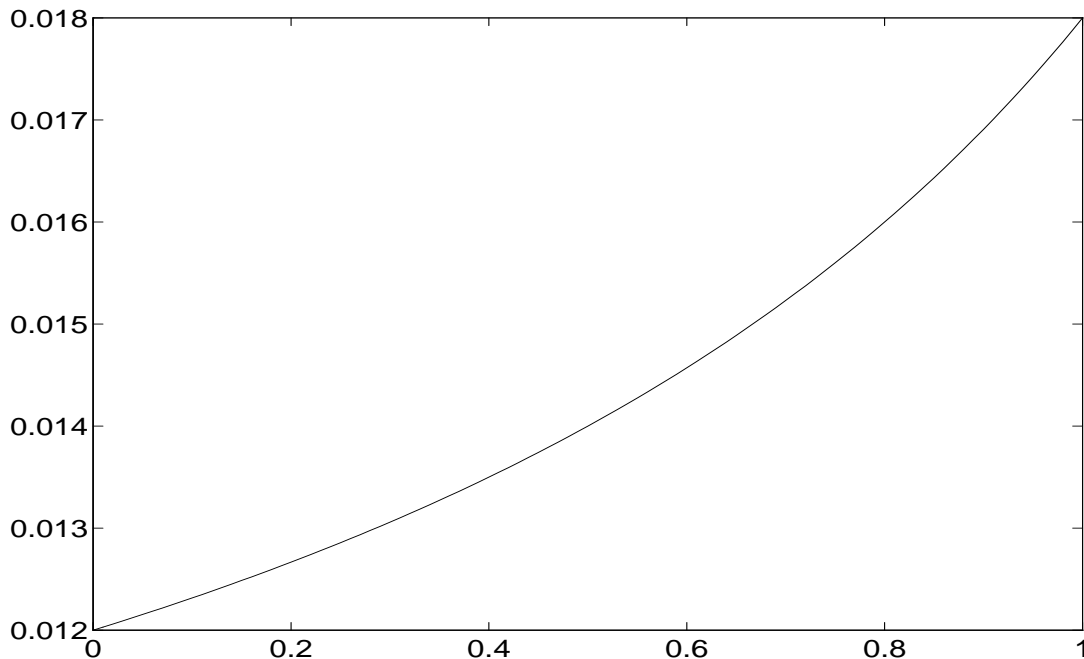


Figure 1.4: Variance asymptotique de l’algorithme en fonction de μ pour $\gamma = 0.001$.

1.3.3 Deux commentaires et l’extension au cas multidimensionnel

- L’algorithme présenté ici est non biaisé si et seulement si le processus étudié est blanc. En effet, le caractère non biaisé provient de ce que $E[x(k)^2 \widehat{M}_{2,k-1}] = E[x(k)^2]E[\widehat{M}_{2,k-1}]$ dans le cas blanc. Dans le cas non blanc, nous n’avons pour l’instant pas de modification susceptible de rendre l’algorithme non biaisé. Toutefois, nous pouvons avancer l’argument suivant. Si le “temps d’indépendance” du processus $x(k)^2$ est T_c , c’est-à-dire $E[x(k)^2 x(k-i)^2] = E[x(k)^2]E[x(k-i)^2]$ dès que $|i| > T_c$, l’algorithme sera non biaisé en remplaçant $\widehat{M}_{2,k-1}$ par $\widehat{M}_{2,k-i}$ avec $i > T_c$. Pratiquement, ce temps d’indépendance est inconnu. Le i invoqué peut être choisi arbitrairement grand, en espérant que le choix satisfait notre argument.

Remarque : cet argument se rapproche de la théorie de la M -indépendance développée par O. Macchi [45].

- Nous avons conclu à la robustesse de l’estimateur quant au choix du paramètre μ . Pour illustrer cela, prenons $\mu = 1$! L’algorithme d’estimation du cumulante est alors basé sur la seule équation

$$\widehat{C}_{4,k} = \widehat{C}_{4,k-1} + \gamma(x(k)^4 - 3x(k)^2 x(k-1)^2 - \widehat{C}_{4,k-1}) \quad (1.44)$$

Le bien-fondé de cet algorithme est montré par l’illustration suivante. Le processus étudié subit un brusque changement de loi, en passant d’uniforme (sur $[-1.8, 1.8]$) à gaussien. La figure (1.5) parle d’elle-même, et à titre de comparaison, l’algorithme (1.30) pour $\mu = 0.01$ et $\gamma = 0.001$ est montré figure (1.6).

- Nous aurons besoin dans ce mémoire de l’extension multidimensionnelle de l’algorithme (1.30). Soit alors \mathbf{X}_n un vecteur aléatoire centré, de loi symétrique, dont nous noterons les composantes x_n^i . Introduisons⁸ $\kappa^{i,j,k,l} = \text{Cum}[x^i, x^j, x^k, x^l]$ le tenseur cumulante d’ordre 4 du vecteur et $\kappa^{i,j} = \text{Cum}[x^i, x^j]$

⁸Les notations introduites ici sont dues à P. MacCullagh [46].

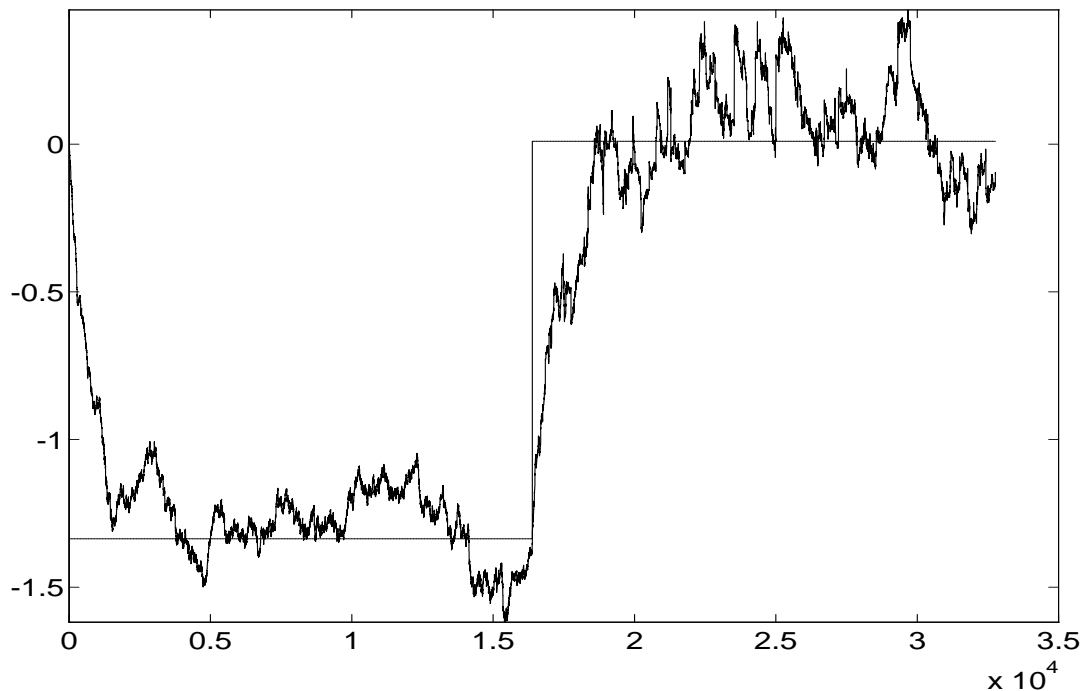


Figure 1.5: Comportement de l'algorithme pour $\mu = 1$ en présence d'une rupture de loi. $\gamma = 0.001$.

son tenseur cumulant d'ordre 2 ou tenseur moment d'ordre 2 (puisque \mathbf{X} est centré). Convenons de plus de noter la somme de trois grandeurs $a^{i,j}b^{k,l} + a^{i,k}b^{j,l} + a^{i,l}b^{j,k}$ par $a^{i,j}b^{k,l}[3]$. L'extension multidimensionnelle de (1.30) est alors

$$\begin{cases} \widehat{\kappa}_n^{i,j} &= \widehat{\kappa}_{n-1}^{i,j} + \mu(x_n^i x_n^j - \widehat{\kappa}_{n-1}^{i,j}) \\ \widehat{\kappa}_n^{i,j,k,l} &= \widehat{\kappa}_{n-1}^{i,j,k,l} + \gamma(x_n^i x_n^j x_n^k x_n^l - x_n^i x_n^j \widehat{\kappa}_{n-1}^{k,l}[3] - \widehat{\kappa}_{n-1}^{i,j,k,l}) \end{cases} \quad (1.45)$$

Evidemment, dans le cas blanc le tenseur cumulant est diagonal, c'est-à-dire que seuls les termes du type $\kappa^{i,i,i,i}$ sont non nuls, et dans le cas stationnaire, il suffit alors d'utiliser l'algorithme pour le cumulante scalaire. De plus, le cas non blanc conduit encore à un estimateur biaisé. Toutefois, ce caractère biaisé n'aura pas d'importance pour l'application dans laquelle nous utiliserons cette extension.

1.4 Conclusion

Ce premier chapitre avait pour but de présenter les outils nécessaires aux chapitres suivants. Plus précisément, nous pouvons mettre en correspondance les outils présentés ici et les chapitres dans lesquels ils apparaîtront.

Les statistiques d'ordre supérieur, moments ou cumulants, et leurs propriétés seront évidemment constamment sur la scène, et agiront essentiellement sur des signaux à valeurs complexes (chapitres 2 et 4) ou réelles (chapitre 3). Les espaces $L^p(\mathcal{R})$ des signaux déterministes à valeurs réelles seront essentiellement utilisés dans le chapitre 3 pour les filtres de Volterra, et les espaces $L_C^p(\mathcal{R})$ dans le chapitre 4. Les classes $f^{(k)}$ de processus stochastiques interviendront tout au long de la présentation. Enfin, l'estimation récursive des cumulants verra son application dans le chapitre 4 relatif aux représentations temps-fréquence d'ordre supérieur.

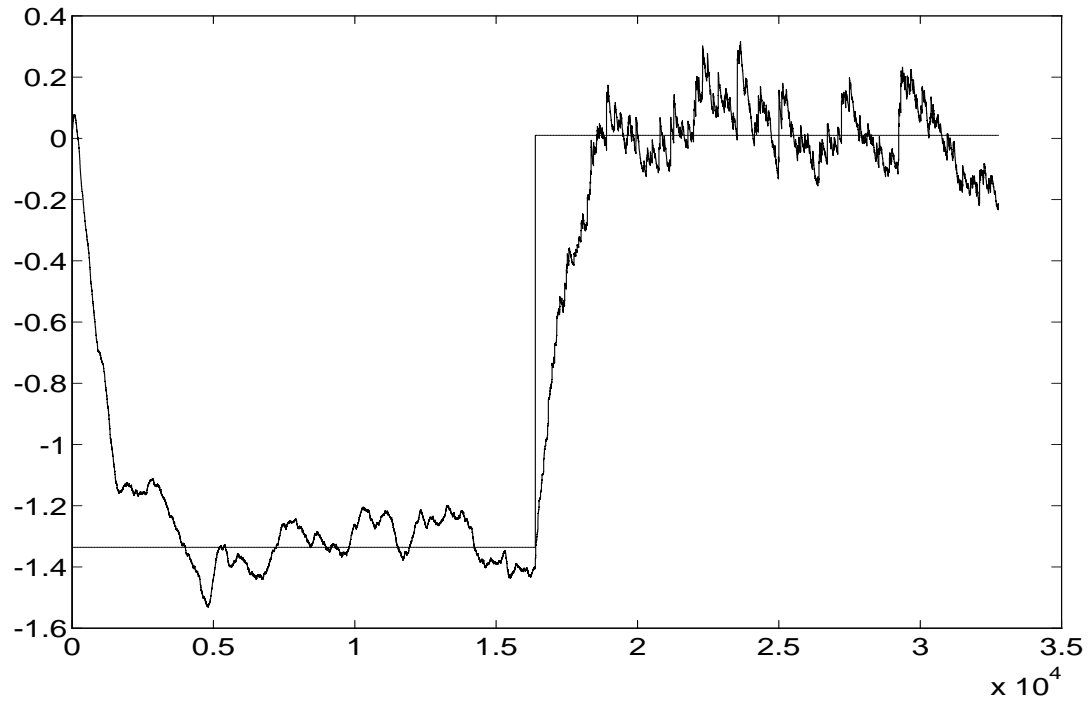


Figure 1.6: Comportement de l'algorithme pour $\mu = 0.01$ en présence d'une rupture de loi. $\gamma = 0.001$.

Il est un sujet que nous n'avons pas mentionné dans ce préambule et qui pourtant sera utilisé au chapitre 3 : la théorie des distributions [6, 54]. Elle sera utilisée pour donner des définitions précises des filtres de Volterra en fréquence, ce qui à notre connaissance n'a jamais été proposé⁹.

Et maintenant ...

⁹L'utilisation de la théorie des distributions dans ce travail a été initiée et réalisée avec l'acérbe E. Moisan.

Chapitre 2

Multicorrélations et Multispectres pour les signaux à valeurs complexes

Commençons notre présentation par l'introduction des statistiques d'ordre supérieur pour les signaux aléatoires à valeurs complexes. Ces statistiques prennent tout leur sens si les signaux traités sont non gaussiens.

Un signal est non gaussien s'il . . . n'est pas gaussien !

La propriété de gaussianité est très forte puisqu'un processus la possédant est entièrement caractérisé par ses statistiques d'ordre un (moyenne) et deux (corrélation).

Les descriptions d'ordre deux reposent sur une analyse statistique de tous les couples de temps (ou fréquences) que l'on peut former. Cette analyse qui semble restrictive et bien faible donne d'excellents résultats dans un grand nombre d'applications. Toutefois, elle a ses limitations, et ne peut traiter certains problèmes. Dans ces cas, des analyses à trois (ou plus) temps (ou fréquences) doivent être envisagées. Ces analyses sont l'extension naturelle des analyses d'ordre deux, et conduisent aux statistiques d'ordre supérieur.

Les études des statistiques d'ordre supérieur pour les signaux à valeurs réelles ont commencé dès 1953 [9]. L'introduction des cumulants pour les signaux aléatoires ne s'est produite que dans les années soixante, avec de nombreux travaux de notamment A. Shirayev [58], D. Brillinger et M. Rosenblatt [13, 14, 15] et bien d'autres. L'essentiel des travaux théoriques sur le sujet a été réalisé durant cette période, mais pour des signaux à valeurs réelles. Pourtant, certaines applications, notamment en communication ou encore en traitement d'antenne, nécessitent l'extension de ces théories au cas de signaux à valeurs complexes.

L'objet de ce chapitre est la définition des objets d'ordre supérieur pour la description des signaux à valeurs complexes. Nous les appellerons multi-outils. Cette appellation dégenère en multicorrélation lorsque l'analyse est temporelle et en multispectre pour des études fréquentielles. Nous présenterons donc des définitions générales, puis regarderons leur évolution sous l'hypothèse de stationnarité, et leur comportement en présence de signaux stationnaires particuliers. Les définitions à temps et fréquence discrets seront alors présentées, puis les multi-outils pour l'étude des interactions entre deux signaux seront introduits. Ce chapitre ne comporte pas d'application. Toutefois, le chapitre suivant utilisera certains outils définis ici dans une application reliée directement à la non-gaussianité.

2.1 Définition des multicorrélations et des multispectres

Cette première section donne des définitions précises et générales des outils d'ordre supérieur pour les signaux complexes. Nous commencerons par une approche dans le domaine temporel, puis examinerons des définitions équivalentes dans le domaine de Fourier.

2.1.1 Multicorrélations

Soit $x(t)$ un signal aléatoire à valeurs complexes appartenant à $f^{(\infty)}$. Sa multicorrélation d'ordre $p + q$ est définie par

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

où $Cum[\cdot]$ représente l'opérateur cumulatif (voir le chapitre 1) et $\mathbf{t} = (t_0, \dots, t_{p+q-1})$. Remarquons tout de suite que dans cette définition apparaissent q composantes conjuguées et p non conjuguées¹. L'expérience montre en effet qu'il faut considérer une variable aléatoire complexe comme "indépendante" de sa conjuguée. Pour être plus précis, l'information contenue dans la multicorrélation dépend du nombre de composantes conjuguées que l'on place dans la définition. Cette remarque est connue depuis longtemps dans la communauté des traiteurs d'antenne où des signaux à bande étroite sont considérés. Pour de tels signaux la pratique a montré que les statistiques envisagées apportent de l'information si le nombre de composantes conjuguées est égal au nombre de composantes non conjuguées, car dans le cas contraire, les statistiques sont nulles. Un autre exemple réside dans la notion de variable aléatoire gaussienne complexe circulaire, pour laquelle les moments $E[x]$, $E[x^*]$, $E[x^2]$ et $E[x^{*2}]$ sont nuls. Seul le moment d'ordre 2 $E[xx^*]$ est non nul. Par suite, seuls les moments d'ordre supérieur ayant autant de conjugués que de non conjugués sont non nuls. Ainsi, la variable gaussienne complexe circulaire est orthogonale à sa conjuguée. Ces quelques arguments justifient la définition (2.1).

Les propriétés de la multicorrélation découlent immédiatement des propriétés des cumulants. Elle est multilinéaire, nulle dans le cas gaussien sous réserve que $p + q > 2$. Si le signal $x(t)$ est blanc au sens fort, la multicorrélation est nulle, sauf lorsque les instants t_i sont égaux.

2.1.2 Multispectres symétriques

Comme le signal appartient à la classe $f^{(\infty)}$ de Blanc-Lapierre et Fortet, le signal admet la représentation de Crámer

$$x(t) = \int \exp(2i\pi\nu t) dX(\nu) \quad (2.2)$$

En remplaçant dans (2.1) les $x(t_i)$ par leur représentation de Crámer, et en utilisant la multilinéarité de l'opérateur cumulatif, il vient

$$C_{x,p+q,p}(\mathbf{t}) = \int Cum[dX(\nu_0), \dots, dX(\nu_{p-1}), dX^*(\nu_p), \dots, dX^*(\nu_{p+q-1})] \exp(2i\pi(\sum_{j=0}^{p-1} \nu_j t_j - \sum_{j=p}^{p+q-1} \nu_j t_j)) \quad (2.3)$$

Supposant que $C_{x,p+q,p}(\mathbf{t})$ admet une transformée de Fourier, que l'on peut définir par

$$C_{x,p+q,p}(\mathbf{t}) = \int \Sigma_{x,p+q,p}(\boldsymbol{\nu}) \exp(2i\pi(\sum_{j=0}^{p-1} \nu_j t_j - \sum_{j=p}^{p+q-1} \nu_j t_j)) d\boldsymbol{\nu} \quad (2.4)$$

où $\boldsymbol{\nu} = (\nu_0, \dots, \nu_{p+q-1})$, on obtient la définition du multispectre symétrique d'ordre $p + q$

$$Cum[dX(\nu_0), \dots, dX(\nu_{p-1}), dX^*(\nu_p), \dots, dX^*(\nu_{p+q-1})] = \Sigma_{x,p+q,p}(\boldsymbol{\nu}) d\boldsymbol{\nu} \quad (2.5)$$

Ces définitions sont très générales et valides pour tout signal de $f^{(\infty)}$. Toutefois, elles sont difficilement utilisables, et nous allons maintenant nous intéresser à leur devenir en présence de signaux stationnaires.

2.2 Influence de la propriété de stationnarité

Un signal est stationnaire si les statistiques de tous les vecteurs aléatoires qu'il induit sont invariantes par translation temporelle. Cette propriété forte peut être affaiblie. Un signal est dit stationnaire d'ordre n si les statistiques des vecteurs aléatoires de dimension inférieure ou égale à n qu'il induit sont invariantes par translation temporelle.

¹ Voir les définitions des variables aléatoires à valeurs complexes dans le Préambule.

2.2.1 Multicorrélations et stationnarité

Pour une multicorrélation d'ordre $p + q$, les propriétés de la stationnarité se traduisent par

$$C_{x,p+q,p}(\mathbf{t} + \tau\mathbf{1}) = C_{x,p+q,p}(\mathbf{t}) \quad (2.6)$$

pour tout couple (p, q) tel que $p+q \leq n$ et tout retard τ . Cette dernière relation montre que la multicorrélation n'est plus $p + q$ dimensionnelle mais seulement $p + q - 1$ dimensionnelle. En effet, elle exprime que la différentielle de $C_{x,p+q,p}(\mathbf{t})$ est nulle selon une certaine direction. Nous allons alors effectuer le changement de variables suivant

$$\begin{cases} \forall i = 1, \dots, p-1 & \tau_i = t_i - t_0 \\ \forall i = p, \dots, p+q-1 & \tau_i = -t_i + t_0 \end{cases} \quad (2.7)$$

Ce changement fait jouer un rôle particulier à t_0 que nous noterons maintenant t . La multicorrélation devient alors fonction uniquement de $\boldsymbol{\tau} = (\tau_1, \dots, \tau_{p+q-1})$ et s'écrit

$$C_{x,p+q,p}(\boldsymbol{\tau}) = \text{Cum}[x(t), x(t + \tau_1), \dots, x(t + \tau_{p-1}), x^*(t - \tau_p), \dots, x^*(t - \tau_{p+q-1})] \quad (2.8)$$

On remarque que pour un terme non conjugué, le retard est ajouté à l'origine alors qu'il est retranché dans le cas d'un terme conjugué. Ce choix est fait pour conserver le théorème de Wiener-Khinchine comme nous le verrons ultérieurement.

Remarque : le rôle particulier donné à t montre que seul le terme $x(t)$ ne peut être conjugué dans cette définition. On peut inclure cette conjugaison en définissant

$$C'_{x,p+q,p}(\boldsymbol{\tau}) = \text{Cum}[x^*(t), x(t + \tau_1), x(t + \tau_{p-1}), x^*(t - \tau_p), \dots, x^*(t - \tau_{p+q-1})] \quad (2.9)$$

Cette forme peut alors être reliée à la définition (2.8), puisqu'en effet

$$C'^*_{x,p+q,p}(\boldsymbol{\tau}) = C_{x,p+q,p}(-\boldsymbol{\tau}) \quad (2.10)$$

qui montre la redondance de cette deuxième définition.

2.2.2 Multispectres symétriques et stationnarité : multispectres

Nous avons vu précédemment que l'on pouvait relier la multicorrélation au multispectre symétrique via une transformation de Fourier multidimensionnelle. Plus précisément

$$\Sigma_{x,p+q,p}(\boldsymbol{\nu}) = \int C_{x,p+q,p}(\mathbf{t}) \exp(-2i\pi(\sum_{j=0}^{p-1} \nu_j t_j - \sum_{j=p}^{p+q-1} \nu_j t_j)) d\mathbf{t} \quad (2.11)$$

En effectuant alors dans l'intégrale le changement de variables (2.7) il vient

$$\Sigma_{x,p+q,p}(\boldsymbol{\nu}) = \int \int C_{x,p+q,p}(\boldsymbol{\tau}) \exp(-2i\pi\boldsymbol{\tau}^T \boldsymbol{\nu}') \exp(2i\pi(\nu_0 + \sum_{j=1}^{p-1} \nu_j - \sum_{j=p}^{p+q-1} \nu_j)t) dt d\boldsymbol{\tau} \quad (2.12)$$

où $\boldsymbol{\nu}' = (\nu_1, \dots, \nu_{p+q-1})$. En posant

$$S_{x,p+q,p}(\boldsymbol{\nu}') = \int C_{x,p+q,p}(\boldsymbol{\tau}) \exp(-2i\pi\boldsymbol{\tau}^T \boldsymbol{\nu}') d\boldsymbol{\tau} \quad (2.13)$$

il vient

$$\Sigma_{x,p+q,p}(\boldsymbol{\nu}) = S_{x,p+q,p}(\boldsymbol{\nu}') \delta(\nu_0 + \sum_{j=1}^{p-1} \nu_j - \sum_{j=p}^{p+q-1} \nu_j) \quad (2.14)$$

Cette équation montre que dans le cas où le signal analysé est stationnaire, le multispectre symétrique ne prend des valeurs non nulles que sur un hyperplan de dimension $p + q - 1$, défini par

$$\nu_0 + \sum_{j=1}^{p-1} \nu_j - \sum_{j=p}^{p+q-1} \nu_j = 0 \quad (2.15)$$

et appelé multiplicité stationnaire [9]. La valeur du multispectre sur cette multiplicité est alors $S_{x,p+q,p}(\boldsymbol{\nu}')$. Nous appellerons cette fonction multispectre d'ordre $p+q$ du signal $x(t)$. Notons que, comme pour la multicorrélation, le multispectre dépend de $p+q-1$ variables fréquentielles.

Utilisons maintenant (2.5) pour obtenir une expression explicite du multispectre. Rappelons que

$$Cum[dX(\nu_0), \dots, dX(\nu_{p-1}), dX^*(\nu_p), \dots, dX^*(\nu_{p+q-1})] = \Sigma_{x,p+q,p}(\boldsymbol{\nu}) d\boldsymbol{\nu} \quad (2.16)$$

En utilisant alors (2.14), il vient

$$\begin{aligned} Cum[dX(\nu_0), \dots, dX(\nu_{p-1}), dX^*(\nu_p), \dots, dX^*(\nu_{p+q-1})] = \\ S_{x,p+q,p}(\boldsymbol{\nu}') \delta(\nu_0 + \sum_{j=1}^{p-1} \nu_j - \sum_{j=p}^{p+q-1} \nu_j) d\boldsymbol{\nu} \end{aligned} \quad (2.17)$$

Supposons que $x(t)$ admette une transformée de Fourier au sens usuel, cette équation se réduit alors à²

$$\begin{aligned} Cum[X(\nu_0), \dots, X(\nu_{p-1}), X^*(\nu_p), \dots, X^*(\nu_{p+q-1})] d\boldsymbol{\nu} = \\ S_{x,p+q,p}(\boldsymbol{\nu}') \delta(\nu_0 + \sum_{j=1}^{p-1} \nu_j - \sum_{j=p}^{p+q-1} \nu_j) d\boldsymbol{\nu} \end{aligned} \quad (2.18)$$

et par conséquent

$$S_{x,p+q,p}(\boldsymbol{\nu}') = Cum[X(-\sum_{j=1}^{p-1} \nu_j + \sum_{j=p}^{p+q-1} \nu_j), X(\nu_1), \dots, X(\nu_{p-1}), X^*(\nu_p), \dots, X^*(\nu_{p+q-1})] \quad (2.19)$$

L'interprétation physique des multispectres est alors claire. Le multispectre examine les interactions statistiques entre $p+q-1$ fréquences $(\nu_1, \dots, \nu_{p+q-1})$ et la fréquence $-\sum_{j=1}^{p-1} \nu_j + \sum_{j=p}^{p+q-1} \nu_j$. Ainsi, un processus aléatoire dont cette fréquence est issue de l'interaction entre les $p+q-1$ qui la composent sera parfaitement analysé par le multispectre³ $S_{x,p+q,p}(\boldsymbol{\nu})$.

Pour terminer cette partie relative aux définitions, effectuons quelques remarques.

- L'approche adoptée ici permet la conservation (ou l'extension) du théorème de Wiener-Khinchine. En effet, tout comme la densité spectrale ou spectre est la transformée de Fourier de la fonction de corrélation, le multispectre et la multicorrélation correspondante forment un couple de transformées de Fourier.
- Examinons le comportement des définitions adoptées dans le cas où le signal étudié est à valeurs réelles. Dans ce cas $x(t) = x^*(t)$, et le changement de variable (2.7) n'a plus lieu d'être réalisé de la sorte. Plus précisément, le nouveau changement de variable s'effectuera selon $t_i = t + \tau_i$ pour tous les i non nuls. En fait, ceci revient à considérer les définitions précédentes sans terme conjugué. Nous aurons alors

$$C_{x,p}(\boldsymbol{\tau}) = Cum[x(t), x(t + \tau_1), \dots, x(t + \tau_{p-1})] \quad (2.20)$$

dont la transformée de Fourier est le multispectre correspondant. Explicitement, ce multispectre suit les relations

$$S_{x,p}(\boldsymbol{\nu}) = Cum[X(-\sum_{j=1}^{p-1} \nu_j), X(\nu_1), \dots, X(\nu_{p-1})] \quad (2.21)$$

$$= Cum[X^*(\sum_{j=1}^{p-1} \nu_j), X(\nu_1), \dots, X(\nu_{p-1})] \quad (2.22)$$

Les définitions obtenues pour le cas réel sont bien les définitions classiques [13, 14].

²Nous commettons ici l'abus d'écriture $dX(\nu) = X(\nu)d\nu$ pour signifier que les $dX(\nu)$ sont les accroissements harmoniques de $X(\nu)$.

³Dorénavant, le vecteur $\boldsymbol{\nu}'$ sera noté simplement $\boldsymbol{\nu}$.

- Lorsque le signal $x(t)$ est non stationnaire, ce qui vient d'être démontré n'est évidemment plus valide. Des tests de non-stationnarité peuvent alors être envisagés. En effet, il suffit de tester si $S_{x,p+q,p}(\boldsymbol{\nu})$ est nul ou non au dehors de la multiplicité stationnaire, puisque la nullité hors la multiplicité stationnaire est caractéristique de la stationnarité. Ce type de tests qui existe dans le cas de valeurs réelles [39] est toutefois très lourd à mettre en oeuvre.

2.2.3 Multiplicités gaussiennes. Un autre avantage des cumulants

Nous revenons dans ce paragraphe au problème soulevé dans le préambule quant à l'avantage des cumulants sur les moments. Nous avons mentionné diverses propriétés intéressantes des cumulants, non vérifiées par les moments. Toutefois, les définitions précédentes des multicorrélations et multispectres, que les signaux soient stationnaires ou non, sont indépendantes de ces propriétés, et peuvent alors être données pour les moments. Nous garderons les mêmes notations, en ajoutant l'indice M pour les grandeurs fondées sur les moments. Les définitions pour les moments dans le cas de signaux stationnaires seront donc

$$\left\{ \begin{array}{l} C_{M,x,p+q,p}(\boldsymbol{\tau}) = E[x(t)x(t+\tau_1)\dots x(t+\tau_{p-1})x^*(t-\tau_p)\dots x^*(t-\tau_{p+q-1})] \\ \Sigma_{M,x,p+q,p}(\nu_0, \boldsymbol{\nu}) = S_{M,x,p+q,p}(\boldsymbol{\nu})\delta(\nu_0 + \sum_{j=1}^{p-1} \nu_j - \sum_{j=p}^{p+q-1} \nu_j) \\ S_{M,x,p+q,p}(\boldsymbol{\nu}) = E[X(-\sum_{j=1}^{p-1} \nu_j + \sum_{j=p}^{p+q-1} \nu_j)X(\nu_1)\dots X(\nu_{p-1})X^*(\nu_p)\dots X^*(\nu_{p+q-1})] \end{array} \right. \quad (2.23)$$

Nous allons maintenant utiliser les relations existantes entre moments et cumulants. De plus, pour développer les idées, nous nous restreindrons aux objets d'ordre 4 à deux complexes conjugués. Les multicorrélations et multispectres sont alors

$$\left\{ \begin{array}{l} C_{x,4,2}(\boldsymbol{\tau}) = Cum[x(t), x(t+\tau_1), x^*(t-\tau_2), x^*(t-\tau_3)] \\ C_{M,x,4,2}(\boldsymbol{\tau}) = E[x(t)x(t+\tau_1)x^*(t-\tau_2)x^*(t-\tau_3)] \\ S_{x,4,2}(\boldsymbol{\nu}) = Cum[X(-\nu_1 + \nu_2 + \nu_3), X(\nu_1), X^*(\nu_2), X^*(\nu_3)] \\ S_{M,x,4,2}(\boldsymbol{\nu}) = E[X(-\nu_1 + \nu_2 + \nu_3)X(\nu_1)X^*(\nu_2)X^*(\nu_3)] \end{array} \right. \quad (2.24)$$

Supposons dans un premier temps que $x(t)$ soit un signal gaussien centré. Il vient alors

$$\begin{aligned} S_{M,x,4,2}(\boldsymbol{\nu}) &= E[X(-\nu_1 + \nu_2 + \nu_3)X(\nu_1)]E[X^*(\nu_2)X^*(\nu_3)] \\ &+ E[X(-\nu_1 + \nu_2 + \nu_3)X^*(\nu_2)]E[X(\nu_1)X^*(\nu_3)] \\ &+ E[X(-\nu_1 + \nu_2 + \nu_3)X^*(\nu_3)]E[X(\nu_1)X^*(\nu_2)] \end{aligned} \quad (2.25)$$

Or, les moments d'ordre deux intervenant dans cette expression sont du type⁴ $S_{x,2,2}(\nu)$ et $S_{x,2,1}(\nu)$ et sont respectivement égaux à $E[X(-\nu)X(\nu)]$ et $E[X(\nu)X^*(\nu)]$. Ainsi, les termes de la première ligne de l'expression précédente sont non nuls si $\nu_1 = -(-\nu_1 + \nu_2 + \nu_3)$ et $\nu_3 = -\nu_2$, puisque les fréquences invoquées doivent appartenir à la multiplicité stationnaire. Ces deux conditions sont identiques et restreintes à $\nu_2 + \nu_3 = 0$. Pour la deuxième ligne, les termes sont non nuls si $\nu_3 - \nu_1 = 0$, alors que cette condition pour la dernière ligne devient $\nu_2 - \nu_1 = 0$. Finalement, le multispectre étudié devient

$$\begin{aligned} S_{M,x,4,2}(\boldsymbol{\nu}) &= S_{x,2,2}(\nu_1)S_{x,2,2}^*(\nu_2)\delta(\nu_2 + \nu_3) \\ &+ S_{x,2,1}(\nu_2)S_{x,2,1}(\nu_3)\delta(\nu_3 - \nu_1) \\ &+ S_{x,2,1}(\nu_3)S_{x,2,1}(\nu_1)\delta(\nu_2 - \nu_1) \end{aligned} \quad (2.26)$$

⁴Comme le signal est supposé centré, les spectres construits sur les cumulants sont égaux aux spectres construits sur les moments.

Les hyperplans

$$\begin{cases} \nu_2 - \nu_1 = 0 \\ \nu_3 - \nu_1 = 0 \\ \nu_2 + \nu_3 = 0 \end{cases} \quad (2.27)$$

sont appelés multiplicités gaussiennes [53]. Ils appartiennent à la multiplicité stationnaire, et représentent les domaines d'existence du multispectre $S_{M,x,4,2}(\boldsymbol{\nu})$ dans le cas gaussien centré. Toutefois, montrons que ces multiplicités n'interviennent pas uniquement dans le cas gaussien. Pour cela, envisageons un processus centré et de loi symétrique, blanc au sens fort, c'est-à-dire, à multicorrélations nulles à tout ordre, excepté pour les retards tous nuls. Pour fixer les notations, les statistiques d'ordre 2 et 4 seront

$$\begin{cases} C_{x,2,1}(\boldsymbol{\tau}) = \sigma^2 \delta(\boldsymbol{\tau}) & = C_{x,2,2}(\boldsymbol{\tau}) \\ S_{x,2,1}(\boldsymbol{\nu}) = \sigma^2 & = S_{x,2,2}(\boldsymbol{\nu}) \\ C_{x,4,2}(\boldsymbol{\tau}) = \kappa_4 \delta(\tau_1, \tau_2, \tau_3) \\ S_{x,4,2}(\boldsymbol{\nu}) = \kappa_4 \end{cases} \quad (2.28)$$

Or, utilisant la relation liant le cumulants d'ordre 4 aux moments d'ordre 2 et 4 et le résultat (2.26), il vient

$$\begin{aligned} S_{x,4,2}(\boldsymbol{\nu}) &= S_{M,x,4,2}(\boldsymbol{\nu}) \\ &- S_{x,2,2}(\nu_1) S_{x,2,2}^*(\nu_2) \delta(\nu_2 + \nu_3) \\ &- S_{x,2,1}(\nu_2) S_{x,2,1}(\nu_3) \delta(\nu_3 - \nu_1) \\ &- S_{x,2,1}(\nu_3) S_{x,2,1}(\nu_1) \delta(\nu_2 - \nu_1) \end{aligned} \quad (2.29)$$

En remplaçant dans cette expression les grandeurs par leurs valeurs, nous obtenons

$$\kappa_4 = S_{M,x,4,2}(\boldsymbol{\nu}) - \sigma^4 (\delta(\nu_2 + \nu_3) + \delta(\nu_3 - \nu_1) + \delta(\nu_2 - \nu_1)) \quad (2.30)$$

Ce résultat montre que le multispectre fondé sur les moments, $S_{M,x,4,2}(\boldsymbol{\nu})$ vaut κ_4 sur toute la multiplicité stationnaire, excepté sur les multiplicités gaussiennes où apparaissent des pics. Par contre, ces pics sont absents du multispectre construit sur les cumulants. Ce résultat provient du fait que les cumulants d'ordre 4 sont des "mesures" d'interactions d'ordre 4 lorsque l'on s'est affranchi des interactions d'ordre 2. D'une manière générale, les cumulants d'ordre n mesurent les interactions d'ordre n lorsque les interactions d'ordre inférieur ont été éliminées (dans le principe, nous pouvons mettre en parallèle ces remarques avec les méthodes d'orthogonalisation de polynômes). A l'ordre 2, ce phénomène est représenté par le lien entre la fréquence nulle et la moyenne du processus. En effet, pour un processus non centré, le spectre construit sur la covariance (cumulant d'ordre 2) présente une amplitude zéro à la fréquence nulle, alors que le spectre construit sur le moment d'ordre 2 présente un pic.

Ces résultats montrent donc un autre avantage des cumulants, puisqu'ils ne présentent pas de pic sur les multiplicités gaussiennes. Cette remarque prend toute son importance lors de procédures d'estimation de multispectres.

2.3 Multicorrélations et multispectres de signaux particuliers

Rappelons la démarche adoptée. Nous avons tout d'abord présenté des définitions très générales valables pour tout élément de $f^{(\infty)}$. Nous avons alors restreint notre étude aux signaux stationnaires, et dans cette section, nous continuons cette réduction en envisageant des signaux particuliers, dont les multicorrélations et multispectres ont des structures intéressantes.

2.3.1 Signaux analytiques

Un signal $x(t)$ est dit analytique si sa transformée de Fourier vérifie $X(\nu) = 0$ pour $\nu < 0$. Ainsi, examinant (2.19), le multispectre $S_{x,p+q,p}(\boldsymbol{\nu})$ est non nul si l'intersection

$$\begin{cases} \forall i = 1, \dots, p+q-1 & \nu_i > 0 \\ -\sum_{i=1}^{p-1} \nu_i + \sum_{i=p}^{p+q-1} \nu_i > 0 \end{cases} \quad (2.31)$$

est non vide. Or le cas $q = 0$, qui correspond à aucun complexe conjugué, conduit le système précédent à

$$\left\{ \begin{array}{l} \forall i = 1, \dots, p-1 \quad \nu_i > 0 \\ -\sum_{i=1}^{p-1} \nu_i > 0 \end{array} \right. \quad (2.32)$$

qui n'a pas de solution. Donc, dans le cas d'un signal analytique, les multispectres du type $S_{x,p,p}(\boldsymbol{\nu})$ sont nuls.

Ainsi, pour des signaux analytiques, les multispectres ne comprenant que des termes conjugués (ou que des termes non conjugués) sont nuls.

2.3.2 Signaux aléatoires circulaires

Nous avons présenté dans le préambule la définition des variables aléatoires multidimensionnelles circulaires. Des définitions de la circularité peuvent également être données pour des signaux. Ces définitions sont dues à B. Picinbono [52].

Une fonction aléatoire est circulaire strictement si toutes les variables aléatoires multidimensionnelles qu'elle induit sont circulaires. Précisément, soit $\mathbf{X}_n = (x(t_1), \dots, x(t_n))$. Alors $x(t)$ est circulaire strictement si \mathbf{X}_n est circulaire pour tout n . Comme pour la stationnarité stricte, cette définition peut être affaiblie. Une fonction aléatoire est circulaire d'ordre n si les variables aléatoires multidimensionnelles de dimension inférieure ou égale à n qu'elle induit sont circulaires.

Or, pour des vecteurs aléatoires, nous avons vu qu'une caractérisation de la circularité est la nullité de ses tenseurs cumulants (ou moments) comportant un nombre différent de termes conjugués et de termes non conjugués. Cette caractérisation nous permet alors d'affirmer que, pour un signal $x(t)$ circulaire

$$C_{x,p+q,p}(\boldsymbol{\tau}) = 0 \quad \forall q \neq p \quad (2.33)$$

Finalement, pour un signal circulaire, seules les multicorrélations du type $C_{x,2p,p}(\boldsymbol{\tau})$ sont non nulles.

2.3.3 Signaux modulés par une fréquence pure

Soit un signal $x(t)$ défini par $x(t) = z(t) \exp(2i\pi\nu_0 t)$, $z(t)$ étant stationnaire. Notons tout de suite que $x(t)$ n'est pas stationnaire. En effet, supposons que $E[z(t)] = m \neq 0$. Alors $E[x(t)] = m \exp(2i\pi\nu_0 t)$ n'est pas indépendant du temps. Donc, $x(t)$ n'est pas stationnaire, mais ici au moins cyclostationnaire à l'ordre 1 (sa moyenne est périodique). Nous ne pouvons donc a priori envisager les multispectres de ce signal. Passons outre, et effectuons le calcul.

Dans le domaine fréquentiel, la définition de $x(t)$ devient $X(\nu) = Z(\nu - \nu_0)$, et par suite, un multispectre de $x(t)$ s'exprime en fonction d'un multispectre de $z(t)$ selon

$$S_{x,p+q,p}(\boldsymbol{\nu}) = Cum[Z((-\sum_{i=1}^{p-1} \nu_i + \sum_{i=p}^{p+q-1} \nu_i) - \nu_0), \\ Z(\nu_1 - \nu_0), \dots, Z(\nu_{p-1} - \nu_0), Z^*(\nu_p - \nu_0), \dots, Z^*(\nu_{p+q-1} - \nu_0)] \quad (2.34)$$

Or, $S_{x,p+q,p}(\boldsymbol{\nu})$ est nul si $\boldsymbol{\nu}$ sort de la multiplicité stationnaire. Ainsi, pour que ce multispectre de $x(t)$ soit non nul il faut

$$(-\sum_{i=1}^{p-1} \nu_i + \sum_{i=p}^{p+q-1} \nu_i) - \nu_0 = -\sum_{i=1}^{p-1} (\nu_i - \nu_0) + \sum_{i=p}^{p+q-1} (\nu_i - \nu_0) \quad (2.35)$$

soit encore

$$-\nu_0 = (p-1)\nu_0 - q\nu_0 \quad (2.36)$$

impliquant $q = p$. Donc, $S_{x,p+q,p}(\boldsymbol{\nu})$ est non nul s'il comporte autant de termes conjugués que non conjugués dans sa définition. De plus,

$$S_{x,2p,p}(\boldsymbol{\nu}) = S_{z,2p,p}(\boldsymbol{\nu} - \nu_0 \mathbf{1}) \quad (2.37)$$

Donc pour les signaux modulés par des fréquences pures, la notion de non-stationnarité est à tempérer, puisque les multispectres d'ordre pair comportant autant de conjugués que de non conjugués sont stationnaires.

Remarque : pour que ce résultat reste valide pour les signaux à valeurs réelles, une phase aléatoire doit être ajoutée dans l'argument du terme oscillant.

2.3.4 Signaux bande étroite

Soit $x(t)$ un signal bande étroite, c'est-à-dire un signal tel que $X(\nu) = 0$ si $|\nu \pm \nu_0| > \Delta\nu$ avec la condition $\nu_0 \gg \Delta\nu$. Ces signaux sont d'importance en traitement d'antenne, puisque l'hypothèse quasi-monochromatique est souvent faite pour traiter les ondes. Nous montrons le résultat suivant

Pour tout ordre inférieur à⁵ $N = \lceil \frac{\nu_0}{\Delta\nu} \rceil$, les seuls multispectres non nuls sont du type $S_{x,2p,p}(\boldsymbol{\nu})$.

Pour démontrer cette affirmation, nous supposons tout d'abord l'ordre du multispectre impair pour arriver à une contradiction. Nous supposons ensuite l'ordre pair mais $p \neq q$, hypothèse conduisant également à une contradiction.

Envisageons les multispectres d'ordre $p + q \leq \frac{\nu_0}{\Delta\nu}$. Un tel multispectre est non nul si $\nu_0 - \Delta\nu \leq |\nu_i| \leq \nu_0 + \Delta\nu \quad \forall i = 1, \dots, p + q - 1$ est compatible avec $\nu_0 - \Delta\nu \leq |-\sum_{i=1}^{p-1} \nu_i + \sum_{i=p}^{p+q-1} \nu_i| \leq \nu_0 + \Delta\nu$. Soient quatre entiers l_1, l_2, k_1 et k_2 tels que $l_1 + l_2 = p - 1$ et $k_1 + k_2 = q$. Considérons alors les $p + q - 1$ fréquences

$$\begin{cases} \nu_0 - \Delta\nu \leq \nu_i \leq \nu_0 + \Delta\nu & \forall i = 1, \dots, l_1 \\ -\nu_0 - \Delta\nu \leq \nu_i \leq -\nu_0 + \Delta\nu & \forall i = l_1 + 1, \dots, p - 1 \\ \nu_0 - \Delta\nu \leq \nu_i \leq \nu_0 + \Delta\nu & \forall i = p, \dots, p + k_1 \\ -\nu_0 - \Delta\nu \leq \nu_i \leq -\nu_0 + \Delta\nu & \forall i = p + k_1 + 1, \dots, p + q - 1 \end{cases} \quad (2.38)$$

En sommant les deux dernières lignes puis en retranchant les deux premières, il vient

$$(q + p - 1 - 2(l_1 + k_2))\nu_0 - (p + q - 1)\Delta\nu \leq -\sum_{i=1}^{p-1} \nu_i + \sum_{i=p}^{p+q-1} \nu_i \leq (q + p - 1 - 2(l_1 + k_2))\nu_0 + (p + q - 1)\Delta\nu \quad (2.39)$$

- Supposons l'ordre impair, soit $q + p = 2n + 1$. Alors, (2.39) devient

$$2(n - (l_1 + k_2))\nu_0 - 2n\Delta\nu \leq -\sum_{i=1}^{p-1} \nu_i + \sum_{i=p}^{p+q-1} \nu_i \leq 2(n - (l_1 + k_2))\nu_0 + 2n\Delta\nu \quad (2.40)$$

Supposons de plus $n - (l_1 + k_2) \geq 1$. Il vient donc $2(n - (l_1 + k_2))\nu_0 \geq 2\nu_0$. Or, l'ordre est choisi tel que $2n + 1 \leq \frac{\nu_0}{\Delta\nu}$, d'où $\nu_0 \geq (2n + 1)\Delta\nu$. Par suite

$$2(n - (l_1 + k_2))\nu_0 + \nu_0 \geq 2\nu_0 + (2n + 1)\Delta\nu \quad (2.41)$$

c'est-à-dire, $2(n - (l_1 + k_2))\nu_0 - 2n\Delta\nu \geq \nu_0 + \Delta\nu$. Cette expression est, d'après (2.40), en contradiction avec $\nu_0 - \Delta\nu \leq -\sum_{i=1}^{p-1} \nu_i + \sum_{i=p}^{p+q-1} \nu_i \leq \nu_0 + \Delta\nu$. Ainsi, l'ordre ne peut être impair.

Remarque : lorsque $n - (l_1 + k_2) < 0$, la contradiction vient de $2(n - (l_1 + k_2))\nu_0 + 2n\Delta\nu \leq \nu_0 - \Delta\nu$.

- Supposons l'ordre $p + q$ pair, mais $p \neq q$, soit par exemple, $q = p + n$ avec $n \geq 1$. Alors (2.39) se transforme en

$$(2p + n - 1 - 2(l_1 + k_2))\nu_0 - (2p + n - 1)\Delta\nu \leq -\sum_{i=1}^{p-1} \nu_i + \sum_{i=p}^{p+q-1} \nu_i \leq (2p + n - 1 - 2(l_1 + k_2))\nu_0 + (2p + n - 1)\Delta\nu \quad (2.42)$$

Or, l'ordre suit $\nu_0 \geq (2p + n)\Delta\nu$, et en supposant encore $p - (l_1 + k_2) \geq 1$, il vient

$$2(p - (l_1 + k_2))\nu_0 - (2p + n - 1)\Delta\nu \geq \nu_0 + \Delta\nu \quad (2.43)$$

⁵[.] désigne ici la partie entière.

qui entraîne évidemment, puisque $n \geq 1$

$$(2p + n - 1 - 2(l_1 + k_2))\nu_0 - (2p + n - 1)\Delta \geq \nu_0 + \Delta\nu \quad (2.44)$$

Ceci est encore en contradiction avec $\nu_0 - \Delta\nu \leq -\sum_{i=1}^{p-1} \nu_i + \sum_{i=p}^{p+q-1} \nu_i \leq \nu_0 + \Delta\nu$, et donc $p = q$.

Remarque : si $n \leq 1$, on échange les rôles de p et q .

- Les démonstrations se font de même pour examiner les contradictions des hypothèses avec $-\nu_0 - \Delta\nu \leq -\sum_{i=1}^{p-1} \nu_i + \sum_{i=p}^{p+q-1} \nu_i \leq -\nu_0 + \Delta\nu$

Remarque 1 : ce résultat est l'extension au cas complexe du théorème 1 page 274 de [58] démontré par A. Shirayev. Notre démonstration en est largement inspirée. D'autre part, un résultat analogue est proposé dans [17] où les auteurs montrent que les multispectres fondés sur les moments et ayant zéro ou un complexe conjugué sont nuls.

Remarque 2 : en traitement d'antenne, les signaux considérés sont souvent analytiques. Dans ce cas, la démonstration précédente se simplifie, puisqu'il suffit de montrer $p = q$.

2.3.5 Quelques remarques

- Un signal analytique gaussien est circulaire strictement. Les seuls multispectres non nuls d'un signal gaussien sont les spectres $S_{x,2,0}(\boldsymbol{\nu})$, $S_{x,2,1}(\boldsymbol{\nu})$ et $S_{x,2,2}(\boldsymbol{\nu})$. Or, si le signal est aussi analytique, seul $S_{x,2,1}(\boldsymbol{\nu})$ est non nul.
- Un signal bande étroite est circulaire d'ordre $N = \lfloor \frac{\nu_0}{\Delta\nu} \rfloor$. Ceci découle immédiatement du paragraphe précédent.

Cette étude générale est terminée, et nous nous tournons vers l'extension aux cas discrets des diverses définitions.

2.4 Définitions à temps et fréquence discrets

Ce paragraphe est destiné à la présentation des définitions des multicorrélations et multispectres pour des signaux discrets. Dans un premier temps, nous nous intéresserons aux signaux définis sur \mathcal{Z} , avant de considérer les signaux discrets périodiques.

2.4.1 Temps discret et fréquence continue

Soit $x(n)$ un processus stochastique stationnaire strictement, défini sur \mathcal{Z} à valeurs dans \mathcal{C} . Nous définirons son multispectre d'ordre $p + q$ par

$$C_{x,p+q,p}(\boldsymbol{\tau}) = Cum[x(t), x(t + \tau_1), \dots, x(t + \tau_{p-1}), x^*(t - \tau_p), \dots, x^*(t - \tau_{p+q-1})] \quad (2.45)$$

où les retards τ_i varient dans \mathcal{Z} . Travaillant à temps discret, nous utiliserons la transformée de Fourier en fréquence réduite (transformée en \mathbb{Z} restreinte au cercle unité). Nous définissons alors le multispectre correspondant par

$$S_{x,p+q,p}(\boldsymbol{\lambda}) = \sum_{\boldsymbol{\tau}=-\infty}^{\infty} C_{x,p+q,p}(\boldsymbol{\tau}) \exp(-2i\pi\boldsymbol{\tau}^T \boldsymbol{\lambda}) \quad (2.46)$$

où les λ_i varient dans $[-\frac{1}{2}, \frac{1}{2}]$. Le multispectre est donc 1-périodique en ses variables. On montre d'une façon analogue au cas continu que le multispectre suit

$$S_{x,p+q,p}(\boldsymbol{\lambda}) = Cum[X(-\sum_{j=1}^{p-1} \lambda_j + \sum_{j=p}^{p+q-1} \lambda_j), X(\lambda_1), \dots, X(\lambda_{p-1}), X^*(\lambda_p), \dots, X^*(\lambda_{p+q-1})] \quad (2.47)$$

Remarque 1 : dans le cas discret, la multiplicité stationnaire est définie par

$$\nu_0 + \sum_{j=1}^{p-1} \lambda_j - \sum_{j=p}^{p+q-1} \lambda_j \equiv 0[1] \quad (2.48)$$

Notons que l'hyperplan est défini modulo 1.

Remarque 2 : toutes les études faites dans le cas temps et fréquence continus pourraient être refaites ici, pour conduire aux mêmes conclusions.

Intéressons nous maintenant au cas pratique des signaux discrets périodiques.

2.4.2 Multispectres et multicorrélations des signaux discrets périodiques

Cette partie est nécessaire puisqu'elle traite les multicorrélations et multispectres des signaux digitalisés. La définition des multicorrélations est la directe transposition de (2.45). Soit $x(n)$ un signal discret à support fini ($n = 0, \dots, N-1$), stationnaire⁶, rendu à support infini en lui rajoutant N zéros et en le périodisant⁷. Alors

$$C_{x,p+q,p}(\mathbf{k}) = Cum[x(n), x(n+k_1), \dots, x(n+k_{p-1}), x^*(n-k_p), \dots, x^*(n-k_{p+q-1})] \quad (2.49)$$

où les k_i varient dans $(-N+1, \dots, -1, 0, 1, \dots, N-1)$. La multicorrélation est donc périodique, de période $2N$.

Le multispectre correspondant est obtenu par transformée de Fourier discrète de la multicorrélation. Par définition

$$S_{x,p+q,p}(\mathbf{m}) = \sum_{\mathbf{k}=-N+1}^{N-1} C_{x,p+q,p}(\mathbf{k}) \exp(-2i\pi \frac{\mathbf{k}^T \mathbf{m}}{2N-1}) \quad (2.50)$$

On montre alors que

$$S_{x,p+q,p}(\mathbf{m}) = Cum[X(-\sum_{j=1}^{p-1} m_j + \sum_{j=p}^{p+q-1} m_j), X(m_1), \dots, X(m_{p-1}), X^*(m_p), \dots, X^*(m_{p+q-1})] \quad (2.51)$$

où les m_i , fréquences discrètes, varient dans $(-N+1, \dots, -1, 0, 1, \dots, N-1)$.

Remarque : ces définitions deviennent les suivantes dans le cas réel :

$$C_{x,p+q,p}(\mathbf{k}) = Cum[x(n), x(n+k_1), \dots, x(n+k_{p-1})] \quad (2.52)$$

$$S_{x,p+q,p}(\mathbf{m}) = Cum[X^*(\sum_{j=1}^{p-1} m_j), X(m_1), \dots, X(m_{p-1})] \quad (2.53)$$

Jusqu'à présent, nous n'avons envisagé que les "auto"-statistiques d'ordre supérieur pour les signaux aléatoires. Nous devons maintenant présenter quelques définitions relatives aux statistiques d'interaction entre deux signaux.

2.5 Intermulticorrélations et intermultispectres

Dans cette section, nous entrons dans l'étude des statistiques d'interaction entre signaux aléatoires. Nous pourrions définir des grandeurs très générales, valables pour n signaux, mais nous préférons rester dans le formalisme développé jusqu'à présent. Nous nous concentrerons sur des définitions concernant les intermultitoutils entre deux signaux aléatoires $x(t)$ et $y(t)$. De plus nous ne referons pas la construction pour le cas stationnaire, mais donnerons les définitions.

⁶Le terme stationnaire est évidemment impropre ici, puisque la finitude du support interdit la stationnarité. Toutefois, puisque nous envisageons dans ce paragraphe des signaux digitalisés, nous supposons le signal analogique initial stationnaire, et garderons par abus de langage le terme stationnaire pour la version échantillonnée périodisée.

⁷Cet ajout de zéros est effectué pour éviter les problèmes dus à la circularité introduite par l'utilisation de la transformée de Fourier discrète.

Soient donc $x(t)$ et $y(t)$ deux signaux de $f^{(\infty)}$. Nous définissons l'intermulticorrélation entre $x(t)$ et $y(t)$ d'ordre $p + q$ par

$$C_{x,y,p+q,\varepsilon}(\boldsymbol{\tau}) = Cum[x(t), x_{\varepsilon_1}(t + \varepsilon_1 \tau_1), \dots, x_{\varepsilon_{p-1}}(t + \varepsilon_{p-1} \tau_{p-1}), y_{\varepsilon_p}(t + \varepsilon_p \tau_p), \dots, y_{\varepsilon_{p+q-1}}(t + \varepsilon_{p+q-1} \tau_{p+q-1})] \quad (2.54)$$

où les ε_i valent $+1$ pour un complexe non conjugué et -1 pour un complexe conjugué, et où $x_{+1}(t) = x(t)$ et $x_{-1}(t) = x^*(t)$. Dans les notations concernant les multicorrélations, l'information relative à la position des complexes conjugués était contenue dans l'indice p , alors qu'ici, elle apparaît dans le vecteur ε . Notons encore une fois que $x(t)$ est pris comme référence. Par extension, nous définirons l'intermultispectre d'ordre $p + q$ comme étant la transformée de Fourier de l'intermulticorrélation (en supposant qu'elle existe), soit

$$S_{x,y,p+q,\varepsilon}(\boldsymbol{\nu}) = \int C_{x,y,p+q,\varepsilon}(\boldsymbol{\tau}) \exp(-2i\pi\boldsymbol{\tau}^T \boldsymbol{\nu}) d\boldsymbol{\tau} \quad (2.55)$$

De la même façon que pour les multispectres, on montre

$$S_{x,y,p+q,\varepsilon}(\boldsymbol{\nu}) = Cum[X(-\sum_{j=1}^{p+q-1} \varepsilon_j \nu_j), X_{\varepsilon_1}(\nu_1), \dots, X_{\varepsilon_{p-1}}(\nu_{p-1}), Y_{\varepsilon_p}(\nu_p), \dots, Y_{\varepsilon_{p+q-1}}(\nu_{p+q-1})] \quad (2.56)$$

L'interprétation de cet objet est donc claire. Il examine les liens statistiques entre q fréquences de $y(t)$, $p - 1$ de $x(t)$ et la fréquence particulière de $x(t)$ définissant la multiplicité stationnaire.

Remarque 1 : dans le cas réel, aucun complexe conjugué n'est pris en compte, et les définitions deviennent

$$\begin{cases} C_{x,y,p+q}(\boldsymbol{\tau}) &= Cum[x(t), x(t + \tau_1), \dots, x(t + \tau_{p-1}), y(t + \tau_p), \dots, y(t + \tau_{p+q-1})] \\ S_{x,y,p+q}(\boldsymbol{\nu}) &= Cum[X^*(\sum_{j=1}^{p+q-1} \nu_j), X(\nu_1), \dots, X(\nu_{p-1}), Y(\nu_p), \dots, Y(\nu_{p+q-1})] \end{cases} \quad (2.57)$$

Remarque 2 : en posant $\boldsymbol{\varepsilon} = (\underbrace{1, \dots, 1}_{p-1}, \underbrace{-1, \dots, -1}_q)$, on a alors les relations

$$\begin{cases} C_{x,x,p+q,\varepsilon}(\boldsymbol{\tau}) &= C_{x,p+q,p}(\boldsymbol{\tau}) \\ S_{x,x,p+q,\varepsilon}(\boldsymbol{\nu}) &= S_{x,p+q,p}(\boldsymbol{\nu}) \end{cases} \quad (2.58)$$

Nous aurions donc pu présenter une théorie générale des intermultioutils pour n signaux, ayant comme cas particulier la théorie développée au début de ce chapitre.

2.6 Multioutils et filtrage linéaire

Cette partie est plus applicative et concerne les relations entrée-sortie d'un système linéaire invariant pour les multioutils.

Soit un système linéaire invariant de réponse impulsionnelle $h(t)$ et de gain complexe $H(\nu)$, attaqué par un signal $x(t)$ de $f^{(\infty)}$. La sortie sera notée $y(t)$, et nous avons

$$\begin{cases} y(t) &= \int h(t - \tau)x(\tau) d\tau \\ Y(\nu) &= H(\nu)X(\nu) \end{cases} \quad (2.59)$$

Examinons alors⁸ $S_{y,x,p+q,\varepsilon}(\boldsymbol{\nu})$. Utilisant (2.56) et (2.59), il vient

$$\begin{aligned} S_{y,x,p+q,\varepsilon}(\boldsymbol{\nu}) &= Cum[H(-\sum_{j=1}^{p+q-1} \varepsilon_j \nu_j)X(-\sum_{j=1}^{p+q-1} \varepsilon_j \nu_j), \\ &\quad H_{\varepsilon_1}(\nu_1)X_{\varepsilon_1}(\nu_1), \dots, H_{\varepsilon_{p-1}}(\nu_{p-1})X_{\varepsilon_{p-1}}(\nu_{p-1}), X_{\varepsilon_p}(\nu_p), \dots, X_{\varepsilon_{p+q-1}}(\nu_{p+q-1})] \end{aligned} \quad (2.60)$$

⁸Notons l'inversion entre les indices x et y qui a toute son importance, puisque que la référence est ici $y(t)$.

L'opérateur cumulant étant multilinéaire, l'équation précédente se transforme en

$$S_{y,x,p+q,\varepsilon}(\boldsymbol{\nu}) = H\left(-\sum_{j=1}^{p+q-1} \varepsilon_j \nu_j\right) H_{\varepsilon_1}(\nu_1) \dots H_{\varepsilon_{p-1}}(\nu_{p-1}) S_{x,x,p+q,\varepsilon}(\boldsymbol{\nu}) \quad (2.61)$$

Par un raisonnement similaire, et en utilisant la dernière remarque du paragraphe précédent, nous avons également la relation

$$S_{y,p+q,p}(\boldsymbol{\nu}) = H\left(-\sum_{j=1}^{p-1} \nu_j + \sum_{j=p}^{p+q-1} \nu_j\right) H(\nu_1) \dots H(\nu_{p-1}) H^*(\nu_p) \dots H^*(\nu_{p+q-1}) S_{x,p+q,p}(\boldsymbol{\nu}) \quad (2.62)$$

Remarque : Les équivalents dans un contexte réel de ces relations sont

$$\begin{cases} S_{y,x,p+q}(\boldsymbol{\nu}) &= H^*\left(\sum_{j=1}^{p+q-1} \nu_j\right) H(\nu_1) \dots H(\nu_{p-1}) S_{x,p+q}(\boldsymbol{\nu}) \\ S_{y,p+q}(\boldsymbol{\nu}) &= H^*\left(\sum_{j=1}^{p+q-1} \nu_j\right) H(\nu_1) \dots H(\nu_{p+q-1}) S_{x,p+q}(\boldsymbol{\nu}) \end{cases} \quad (2.63)$$

2.7 Conclusion

Nous avons dans ce chapitre défini les multicorrélations et multispectres pour les signaux aléatoires à valeurs complexes. Ces outils ont été étudiés en présence de signaux stationnaires, et de signaux particuliers (analytiques, circulaires, bande étroite). Nous avons pu apporter une justification claire au fait que les traiteurs d'antenne considèrent les signaux complexes bande étroite comme étant circulaires.

Ce chapitre théorique ne comporte pas d'application. Toutefois, la littérature relative aux statistiques d'ordre supérieur des signaux à valeurs réelles regorge d'applications, notamment pour l'identification aveugle de modèles linéaires. En effet, il est connu que les statistiques d'ordre deux ne permettent pas l'identification aveugle de la phase d'un filtre, identification possible en utilisant l'ordre trois par exemple. Nous pensons que la théorie développée ici pourra s'appliquer à ce genre de problème dans des contextes où la modélisation doit se faire en utilisant les complexes (e.g. communication).

Nous venons de présenter des outils généraux capables d'appréhender le caractère non gaussien d'un signal. Il nous faut maintenant réfléchir sur l'existence de tels signaux pour justifier l'intérêt des objets introduits.

Si beaucoup de phénomènes génèrent des signaux gaussiens (bruits électroniques, bruits de fond sous-marins, ...), la Nature n'est pas moins fondamentalement non linéaire. Il suffit pour s'en convaincre de regarder l'écoulement turbulent d'une rivière ou encore d'examiner l'évolution à court terme de la météo. Ainsi, la grande partie des signaux que la Nature crée sont non gaussiens.

Du non-linéaire naît le non-gaussien ...

La sortie d'un système non linéaire est non gaussienne. Ce caractère négatif impose une multitude de non-linéarités, et nous ne pouvons prétendre examiner la non-linéarité dans son ensemble. Le chapitre suivant propose l'étude d'une classe particulière de systèmes : les filtres de Volterra, générateurs de non-gaussien.

Chapitre 3

Filtres de Volterra

Nous abordons ici l'étude d'outils particuliers pour la compréhension des phénomènes non linéaires.

Un phénomène est non linéaire s'il ... n'est pas linéaire !

Si l'on fait remonter l'avènement du Traitement du Signal aux années 1930–1940, on s'aperçoit que l'hypothèse de linéarité des systèmes a en grande partie occupé le devant de la scène. A notre connaissance, les premiers travaux concernant les systèmes non linéaires remontent à 1958, avec la parution de la monographie “Nonlinear problems in random theory” de et encore de Norbert Wiener [63]. Les années 60 ont alors vu deux mouvements principaux. A la suite de Wiener, ses élèves et collègues du M.I.T. ont défriché la voie qu'il avait amorcée. Parmi eux, Lee et Schetzen ont développé des méthodes d'identification de filtres non linéaires appelés filtres de Volterra, qui font l'objet principal de ce chapitre [57]. Une autre communauté, plus automatique, s'est chargée de développer ce qu'on appelle aujourd'hui la théorie du filtrage non linéaire optimal, dont une présentation complète peut être trouvée dans [40]. Alors que les filtres de Volterra emploient une représentation dite externe des signaux, la théorie du filtrage non linéaire emploie une représentation interne, ou représentation d'état. La limite linéaire de cette théorie conduit d'ailleurs au filtre de Kalman, basé aussi sur cette représentation. Lorsque la représentation interne du système est connue, il est montré que les états suivent des processus de diffusion, caractérisés par des densités de probabilités suivant l'équation aux dérivées partielles dite de Fokker-Planck ou Kolmogoroff rétrograde. Cette équation est également satisfaite par la densité de probabilité de l'état à estimer conditionnellement aux observations passées. Or, la moyenne de cette densité est le filtre optimal en moyenne quadratique, et nous voyons donc que résoudre le problème du filtrage optimal avec cette approche requiert la résolution d'une équation aux dérivées partielles fortement non linéaire. De plus, cette théorie requiert de connaître le système à identifier ou à contrôler, connaissance généralement obtenue en Automatique mais rarement en Traitement du Signal. Mais la connaissance interne du système est équivalente à la connaissance des statistiques du signal qu'il représente. Nous opterons alors pour une description externe, plus commune en Traitement du Signal, et à notre sens mieux adaptée aux interprétations physiques. Parmi ces représentations externes, nous allons nous intéresser aux filtres de Volterra.

3.1 Définitions

Deux visions équivalentes des filtres de Volterra peuvent être présentées, selon les points de vue adoptés : filtrage ou identification.

Dans le premier contexte, une structure paramétrique particulière est donnée a priori : un filtre de Volterra est un filtre polynomial (polynomial par rapport à l'entrée) avec mémoire et anticipation. Pour l'identification, le lien supposé non linéaire entre deux grandeurs est approché par un développement de type Taylor avec mémoire, ou développement de Volterra. Dans ce contexte, le filtre linéaire apparaît comme la première approximation du lien. Etant données les applications proposées par la suite, nous nous tiendrons à ce deuxième point de vue, mais utiliserons souvent par abus le terme de filtre.

Nous allons maintenant présenter les définitions du développement de Volterra, pour des temps continus et discrets, ainsi que les définitions dans le domaine fréquentiel. Ces définitions étant données, nous commenterons des problèmes divers, tels stabilité, existence ou validité.

3.1.1 Définition à temps continu et fréquence continue

Commençons par donner les définitions à temps continu et examiner les problèmes de stabilité et d'existence de la transformée de Fourier de la sortie du filtre [57].

Temps continu

Soient deux signaux $x(t)$ et $y(t)$ reliés par une relation F , c'est-à-dire

$$y(t) = F(x(t)) \quad (3.1)$$

La notation peut paraître ambiguë. F est une fonctionnelle, et son application à x fait intervenir les valeurs passées et futures de ce signal. Nous supposons que cette fonctionnelle est continue et qu'elle possède le développement suivant, dit de Volterra¹

$$y(t) = h_0 + \sum_{i=1}^{+\infty} f_i(x(t)) \quad (3.2)$$

$$f_i(x(t)) = \int h_i(\tau_1, \dots, \tau_i) x(t - \tau_1) \dots x(t - \tau_i) d\tau$$

Terminologie :

- La fonction $h_i(\tau_1, \dots, \tau_i)$ est appelée noyau d'ordre i du développement.
- La fonction $f_i(x(t))$ 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

Un développement de Volterra apparaît donc comme une généralisation des développements de Taylor, en y intégrant de la mémoire et de l'anticipation. Nous parlerons ici de filtres de Volterra, par extension du cas linéaire qui n'est autre qu'un développement de Volterra du premier ordre. Pour continuer ce parallèle, il est intéressant d'examiner la version fréquentielle de ces filtres et de voir dans quelle mesure la notion de fonction de transfert peut se généraliser au non-linéaire.

Toutefois, le passage en fréquence de la définition pose tout de suite des questions d'existence.

Problèmes d'existence pour un filtre homogène

Divers problèmes se posent quant à la stabilité du filtre et l'existence de la transformée de Fourier de la sortie.

En ce qui concerne la stabilité, nous employons le critère "BIBO" [57], ou "Bounded Input-Bounded Output", qui signifie qu'à une entrée bornée correspond une sortie bornée.

Soit donc $x(t)$ un signal borné, soit $|x(t)| \leq C < +\infty$. Alors, la sortie $y(t)$ d'un filtre de Volterra homogène d'ordre i est bornée si

$$|y(t)| = \left| \int h_i(\tau_1, \dots, \tau_i) x(t - \tau_1) \dots x(t - \tau_i) d\tau \right| < +\infty \quad (3.3)$$

Or,

$$|y(t)| \leq \int |h_i(\tau_1, \dots, \tau_i) x(t - \tau_1) \dots x(t - \tau_i)| d\tau \quad (3.4)$$

$$\leq C^i \int |h_i(\tau_1, \dots, \tau_i)| d\tau \quad (3.5)$$

Donc, une condition suffisante de stabilité BIBO est

$$\int |h_i(\tau_1, \dots, \tau_i)| d\tau < +\infty \quad (3.6)$$

¹ Les 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.

Pour assurer la stabilité du système global, il faut assurer (3.6) pour tout i .

Venons-en au second problème. Si nous souhaitons représenter des signaux physiques comme sorties de filtres de Volterra, il est légitime de les considérer comme signaux d'énergie finie, soit appartenant à $L^2(\mathcal{R})$. Mais quelles sont les contraintes à donner aux entrées et aux noyaux pour qu'une sortie soit dans cet espace? Considérons $y(t)$ la sortie d'un filtre homogène d'ordre i . Nous souhaitons $\int |y(t)|^2 dt < +\infty$. Or²

$$\int |y(t)|^2 dt = \int \left| \int h_i(\boldsymbol{\tau}) \prod_{j=1}^i x(t - \tau_j) d\boldsymbol{\tau} \right|^2 dt \quad (3.7)$$

$$\leq \int \int \int |h_i(\boldsymbol{\tau}) h_i(\boldsymbol{\tau}')| \left| \prod_{j=1}^i x(t - \tau_j) \prod_{j=1}^i x(t - \tau'_j) \right| d\boldsymbol{\tau} d\boldsymbol{\tau}' dt \quad (3.8)$$

$$\leq \int \int |h_i(\boldsymbol{\tau}) h_i(\boldsymbol{\tau}')| \int |x(t)|^{2i} dt d\boldsymbol{\tau} d\boldsymbol{\tau}' \quad (3.9)$$

la dernière inégalité provenant de l'inégalité de Schwartz. Donc, si $\int |x(t)|^{2i} dt < +\infty$, i.e. $x(t) \in L^{2i}(\mathcal{R})$, et $\int |h(\boldsymbol{\tau})| d\boldsymbol{\tau} < +\infty$, alors $\int |y(t)|^2 dt < +\infty$. Ainsi, pour que $y(t)$ 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 $L^{2i}(\mathcal{R})$. Pour étendre ces résultats au filtre global (3.2), il faut que tous les filtres homogènes soient stables et que $x(t)$ soit dans $L^\infty(\mathcal{R})$. Malheureusement, cette condition n'est que nécessaire. Nous pouvons maintenant nous tourner vers la version fréquentielle des filtres de Volterra.

Fréquence continue

Il est bien connu que la quadrature de fréquences pures 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 , h_i et y possèdent des transformées de Fourier. Alors

$$Y(\nu) = TF(y(t)) = \int y(t) \exp(-2i\pi\nu t) dt \quad (3.10)$$

Etant donnée la linéarité de la transformation de Fourier, il vient

$$Y(\nu) = h_0 \delta(\nu) + \sum_{i=1}^{+\infty} TF(f_i(x(t))) \quad (3.11)$$

L'évaluation de $TF(f_i(x(t)))$ s'effectue de la façon suivante. En utilisant à nouveau la linéarité de la transformée de Fourier, il vient

$$TF(f_i(x(t))) = \int h_i(\tau_1, \dots, \tau_i) TF(x(t - \tau_1) \dots x(t - \tau_i)) d\boldsymbol{\tau} \quad (3.12)$$

Or, il est facile de montrer que pour un signal $x(t)$

$$TF\left(\prod_{j=1}^i x(t)\right) = *_{j=1}^i X(\nu) = \int X(\nu_1) \dots X(\nu_{i-1}) X(\nu - \nu_1 - \dots - \nu_{i-1}) d\boldsymbol{\nu} \quad (3.13)$$

où $d\boldsymbol{\nu} = d\nu_1 \dots d\nu_{i-1}$. Par suite, il vient

$$TF(f_i(x(t))) = \int h_i(\tau_1, \dots, \tau_i) \int X(\nu_1) \dots X(\nu_{i-1}) X(\nu - \nu_1 - \dots - \nu_{i-1}) \exp(-2i\pi(\sum_{j=1}^{i-1} \nu_j \tau_j)) \exp(-2i\pi(\nu - \nu_1 - \dots - \nu_{i-1})\tau_i) d\boldsymbol{\nu} d\boldsymbol{\tau} \quad (3.14)$$

²Nous remplacerons parfois $h_i(\tau_1, \dots, \tau_i)$ par $h_i(\boldsymbol{\tau})$.

soit finalement

$$TF(f_i(x(t))) = \int H_i(\nu_1, \dots, \nu_{i-1}, \nu - \nu_1 - \dots - \nu_{i-1}) X(\nu_1) \dots X(\nu_{i-1}) X(\nu - \nu_1 - \dots - \nu_{i-1}) d\nu \quad (3.15)$$

Enfin, l'expression cherchée est la suivante

$$Y(\nu) = h_0 \delta(\nu) + \sum_{i=1}^{+\infty} \int H_i(\nu_1, \dots, \nu_{i-1}, \nu - \nu_1 - \dots - \nu_{i-1}) X(\nu_1) \dots X(\nu_{i-1}) X(\nu - \nu_1 - \dots - \nu_{i-1}) d\nu \quad (3.16)$$

où $H_i(\nu_1, \dots, \nu_i)$ est la transformée de Fourier du noyau $h_i(\tau_1, \dots, \tau_i)$. Ce résultat est valide pour des signaux possédant une transformée de Fourier. Supposons maintenant que nous traitions des signaux d'entrée n'ayant une transformée de Fourier qu'au sens des distributions. Supposons de plus que les noyaux soient infiniment dérivables. La transformée de Fourier de la sortie $y(t)$ d'un filtre homogène doit être traitée au sens des distributions. Calculons la distribution "transformée de Fourier" de $y(t)$, que nous noterons symboliquement $Y(\nu) = TF(y(t))$. Alors par définition³

$$\langle Y(\nu) | \varphi(\nu) \rangle = \langle y(t) | TF(\varphi(\nu)) \rangle \quad (3.17)$$

où $\varphi(\nu)$ est une fonction test infiniment dérivable, tendant vers zéro à l'infini plus vite que tout polynôme en $\frac{1}{\nu}$ (il doit en être de même pour ses dérivées). En remplaçant $y(t)$, sortie d'un filtre homogène d'ordre i , par son expression, il vient

$$\langle y(t) | TF(\varphi(\nu)) \rangle = \langle \int h_i(\tau) \prod_{j=1}^i x(t - \tau_j) d\tau | TF(\varphi(\nu)) \rangle \quad (3.18)$$

$$= \int h_i(\tau) \langle \prod_{j=1}^i x(t - \tau_j) | TF(\varphi(\nu)) \rangle d\tau \quad (3.19)$$

Or,

$$\langle \prod_{j=1}^i x(t - \tau_j) | TF(\varphi(\nu)) \rangle = \langle TF(\prod_{j=1}^i x(t - \tau_j)) | \varphi(\nu) \rangle \quad (3.20)$$

$$= \langle *_{j=1}^i X(\nu) \exp(-2i\pi\nu\tau_j) | \varphi(\nu) \rangle \quad (3.21)$$

où $X(\nu)$ est la transformée de Fourier de $x(t)$ au sens des distributions. De plus, par définition du produit de convolution de deux distributions, soit

$$\langle x(t) * y(t) | \varphi(t) \rangle = \langle x(t_1) | \langle y(t_2) | \varphi(t_1 + t_2) \rangle \rangle \quad (3.22)$$

et en utilisant le fait que l'exponentielle complexe est infiniment dérivable, (3.21) devient

$$\begin{aligned} & \langle \prod_{j=1}^i x(t - \tau_j) | TF(\varphi(\nu)) \rangle \\ &= \langle X(\nu_1) | \langle X(\nu_2) | \dots \langle X(\nu_i) | \exp(-2i\pi(\sum_{j=1}^i \nu_j \tau_j)) \varphi(\sum_{j=1}^i \nu_j) \rangle \dots \rangle \end{aligned} \quad (3.23)$$

$$= \langle \otimes_{j=1}^i X(\nu_j) | \exp(-2i\pi(\sum_{j=1}^i \nu_j \tau_j)) \varphi(\sum_{j=1}^i \nu_j) \rangle \quad (3.24)$$

³ Lorsqu'une distribution est l'associée d'une fonction, $\langle Y(\nu) | \varphi(\nu) \rangle = \int Y(\nu) \varphi(\nu) d\nu$.

où \otimes est le produit tensoriel de deux distributions⁴. Finalement, nous obtenons

$$\langle Y(\nu) | \varphi(\nu) \rangle = \langle \otimes_{j=1}^i X(\nu_j) | H_i(\nu_1, \dots, \nu_i) \varphi\left(\sum_{j=1}^i \nu_j\right) \rangle \quad (3.25)$$

qui est l'application à une fonction test φ de la transformée de Fourier de la sortie d'un filtre de Volterra homogène d'ordre i , dont l'entrée ne possède de transformée de Fourier qu'au sens des distributions. L'obtention de la version fréquentielle pour le filtre total suit alors par linéarité.

Remarque : lorsque l'entrée est une fonction, les résultats (3.15) et (3.25) sont identiques.

L'expression obtenue va nous servir immédiatement dans le cas d'un exemple, mais nous permettra également d'obtenir très facilement les versions échantillonnées des filtres.

La forme (3.25) 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 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, contrairement au cas temporel. 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 d'après (3.25)

$$\langle Y(\nu) | \varphi(\nu) \rangle = \langle X(\nu_1) \otimes X(\nu_2) | H_2(\nu_1, \nu_2) \varphi(\nu_1 + \nu_2) \rangle \quad (3.26)$$

soit encore

$$4 \langle Y(\nu) | \varphi(\nu) \rangle = \langle \delta(\nu_1 - \nu_0) + \delta(\nu_1 + \nu_0) | \langle \delta(\nu_2 - \nu_0) + \delta(\nu_2 + \nu_0) | H_2(\nu_1, \nu_2) \varphi(\nu_1 + \nu_2) \rangle \rangle \quad (3.27)$$

$$= \langle \delta(\nu_1 - \nu_0) + \delta(\nu_1 + \nu_0) | H_2(\nu_1, \nu_0) \varphi(\nu_1 + \nu_0) + H_2(\nu_1, -\nu_0) \varphi(\nu_1 - \nu_0) \rangle \rangle \quad (3.28)$$

$$= H_2(\nu_0, \nu_0) \varphi(2\nu_0) + [H_2(-\nu_0, \nu_0) + H_2(\nu_0, -\nu_0)] \varphi(0) + H_2(-\nu_0, -\nu_0) \varphi(-2\nu_0) \quad (3.29)$$

Reconnaissant que $\varphi(\nu_0) = \langle \delta(\nu - \nu_0) | \varphi(\nu) \rangle$, on obtient alors

$$4Y(\nu) = H_2(\nu_0, \nu_0) \delta(\nu - 2\nu_0) + [H_2(\nu_0, -\nu_0) + H_2(-\nu_0, \nu_0)] \delta(\nu) + H_2(-\nu_0, -\nu_0) \delta(\nu + 2\nu_0) \quad (3.30)$$

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 (3.30), $H_2(\nu_0, \nu_0) \delta(\nu - 2\nu_0)$ signifie que $2\nu_0$ provient de l'interaction quadratique entre ν_0 et elle-même. De plus $H_2(-\nu_0, \nu_0) \delta(\nu)$ montre que la fréquence nulle provient du couplage entre les composantes $-\nu_0$ et ν_0 du signal d'entrée. Ces explications physiques étant données, nous nous intéressons maintenant à la validité du développement de Volterra.

3.1.2 Validité du développement de Volterra

L'étude des développements de fonctionnelles en séries de Volterra remonte au début du siècle avec les travaux de Maurice Fréchet [32]. 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 petite à l'aide de développements de Volterra.

⁴ Le produit tensoriel de deux distributions $x(t)$ et $y(t)$ est par définition $\langle x(a) \otimes y(b) | \varphi(a, b) \rangle = \langle x(a) | \langle y(b) | \varphi(a, b) \rangle \rangle$.

3.1.3 Définition à temps discret

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, l'on a⁵

$$\begin{cases} y(n) &= h_0 + \sum_{i=1}^{+\infty} f_i(x(n)) \\ f_i(x(n)) &= \sum_{\mathbf{k}=-\infty}^{+\infty} h_i(k_1, \dots, k_i) x(n - k_1) \dots x(n - k_i) \end{cases} \quad (3.31)$$

Utilisant la transformée de Fourier en fréquences réduites, nous obtenons l'analogie de cette définition dans le domaine des fréquences

$$Y(\lambda) = h_0 \delta(\lambda) + \sum_{i=1}^{+\infty} \int_{-\frac{1}{2}}^{+\frac{1}{2}} H_i(\lambda_1, \dots, \lambda_{i-1}, \lambda - \lambda_1 - \dots - \lambda_{i-1}) X(\lambda_1) \dots X(\lambda_{i-1}) X(\lambda - \lambda_1 - \dots - \lambda_{i-1}) d\lambda \quad (3.32)$$

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 linéaire ne peuvent être rendus infinis. De plus comme dans le cas des multispectres, les signaux sont à support fini et doivent être supposés périodiques si l'on désire travailler dans le domaine fréquentiel. Nous définissons donc un développement (ou filtre selon le premier point de vue) de Volterra à ordre K et mémoire M finis par

$$\begin{cases} y(n) &= h_0 + \sum_{i=1}^K f_i(x(n)) \\ f_i(x(n)) &= \sum_{\mathbf{k}=0}^{M-1} h_i(k_1, \dots, k_i) x(n - k_1) \dots x(n - k_i) \end{cases} \quad (3.33)$$

Nous supposons de plus que le signal discret ainsi défini est périodique pour pouvoir lui appliquer la transformée de Fourier discrète⁶.

Version fréquentielle pour un signal échantillonné périodisé

Soit $x(t)$ un signal continu observé sur une durée finie. Il est échantillonné, et périodisé pour pouvoir faire des calculs numériques utilisant la transformée de Fourier discrète. Le signal s'écrit alors

$$x_e(t) = (x(t) * \sum_m \delta(t - mP)) \sum_n \delta(t - nT) \quad (3.34)$$

où T et P sont respectivement les périodes d'échantillonnage et de périodisation, et sont relié par $P = NT$, N étant le nombre d'échantillons. Nous souhaitons obtenir la version fréquentielle discrète d'une filtrée de Volterra de ce signal. Il suffit d'obtenir cette forme pour un filtre homogène. De plus, nous pouvons appliquer le résultat de la première section relatif à la version fréquentielle au sens des distributions. Rappelons que la transformée de Fourier au sens des distributions de la sortie d'un filtre homogène satisfait

$$\langle Y(\nu) | \varphi(\nu) \rangle = \langle \otimes_{j=1}^i X(\nu_j) | H_i(\nu_1, \dots, \nu_i) \varphi(\sum_{j=1}^i \nu_j) \rangle \quad (3.35)$$

⁵ La notation $\sum_{\mathbf{k}=a}^b$ est équivalente à $\sum_{k_1=a}^b \dots \sum_{k_i=a}^b$.

⁶ Avant de périodiser, la longueur du support de $x(n)$ est doublée par ajout de zéros. Les problèmes de circularité dus à la transformée de Fourier discrète sont ainsi évités

Nous allons alors remplacer dans cette expression la distribution $X(\nu)$ par la distribution

$$X_e(\nu) = \frac{1}{PT} (X(\nu) \sum_m \delta(\nu - \frac{m}{P})) * \sum_n \delta(\nu - \frac{n}{T}) \quad (3.36)$$

En omettant le facteur $\frac{1}{PT}$, il vient alors

$$\begin{aligned} & \langle Y(\nu) | \varphi(\nu) \rangle \\ &= \langle X_e(\nu_1) | \langle X_e(\nu_2) | \dots \langle X_e(\nu_i) | H_i(\nu_1, \dots, \nu_i) \varphi(\sum_{j=1}^i \nu_j) \rangle \dots \rangle \\ &= \langle X_e(\nu_1) | \langle X_e(\nu_2) | \\ & \quad \dots \sum_{n_i, m_i} X(\frac{m_i}{P}) \langle X_e(\nu_{i-1}) | H_i(\nu_1, \dots, \nu_{i-1}, \frac{n_i}{T} + \frac{m_i}{P}) \varphi(\frac{n_i}{T} + \frac{m_i}{P} + \sum_{j=1}^{i-1} \nu_j) \rangle \dots \rangle \\ & \quad \vdots \\ &= \sum_{n_1, m_1} \dots \sum_{n_i, m_i} X(\frac{m_1}{P}) \dots X(\frac{m_i}{P}) H_i(\frac{n_1}{T} + \frac{m_1}{P}, \dots, \frac{n_i}{T} + \frac{m_i}{P}) \varphi(\frac{n_1}{T} + \frac{m_1}{P} + \dots + \frac{n_i}{T} + \frac{m_i}{P}) \quad (3.37) \end{aligned}$$

Posons alors $m = m_1 + \dots + m_i$ et $n = n_1 + \dots + n_i$. On obtient

$$\begin{aligned} \langle Y(\nu) | \varphi(\nu) \rangle &= \sum_{n, m} \sum_{n_1, m_1} \dots \sum_{n_{i-1}, m_{i-1}} X(\frac{m_1}{P}) \dots X(\frac{m_{i-1}}{P}) X(\frac{m - m_1 - \dots - m_{i-1}}{P}) \\ & \quad H_i(\frac{n_1}{T} + \frac{m_1}{P}, \dots, \frac{n_{i-1}}{T} + \frac{m_{i-1}}{P}, \frac{n - n_1 - \dots - n_{i-1}}{T} + \frac{m - m_1 - \dots - m_{i-1}}{P}) \varphi(\frac{n}{T} + \frac{m}{P}) \quad (3.38) \end{aligned}$$

Or, la transformée de Fourier au sens des distributions de la sortie d'un filtre échantillonné périodisé s'écrit

$$\langle Y(\nu) | \varphi(\nu) \rangle = \sum_{n, m} Y(\frac{m}{P}) \varphi(\frac{n}{T} + \frac{m}{P}) \quad (3.39)$$

Donc, par identification, terme à terme et en posant $Y(\frac{m}{P}) = Y(m)$, $X(\frac{m}{P}) = X(m)$ et $H_i(\frac{\mathbf{m}}{P} + \frac{\mathbf{n}}{T}) = H_i(\mathbf{m} + \mathbf{n})$, il vient

$$\begin{aligned} Y(m) &= \sum_{n_1, m_1} \dots \sum_{n_{i-1}, m_{i-1}} X(m_1) \dots X(m_{i-1}) X(m - m_1 - \dots - m_{i-1}) \\ & \quad H_i(n_1 + m_1, \dots, n_{i-1} + m_{i-1}, n - n_1 - \dots - n_{i-1} + m - m_1 - \dots - m_{i-1}) \quad (3.40) \end{aligned}$$

Enfin, en ne conservant que les termes de la première période (i.e. $\mathbf{n} = \mathbf{0}$) nous obtenons

$$\begin{aligned} Y(m) &= \sum_{m_1, \dots, m_{i-1}} \\ & \quad X(m_1) \dots X(m_{i-1}) X(m - m_1 - \dots - m_{i-1}) H_i(m_1, \dots, m_{i-1}, m - m_1 - \dots - m_{i-1}) \quad (3.41) \end{aligned}$$

3.2 Implantation sous forme vectorielle

Cette partie concerne l'aspect pratique de l'exploitation des filtres de Volterra. Nous nous intéressons donc aux dernières définitions données : les filtres ou développements de Volterra discrets à ordre et mémoire finis. Nous distinguerons encore tout au long de cette partie la version temporelle de la version fréquentielle des filtres.

3.2.1 Implantation dans le domaine temporel

Dans cette partie, nous nous intéresserons à des signaux $y(n)$, sorties de filtres de Volterra d'ordre fini K et de mémoire finie M attaqués par des signaux $x(n)$, soit

$$y(n) = h_0 + \sum_{i=1}^K \sum_{\mathbf{k}=0}^{M-1} h_i(k_1, \dots, k_i) x(n - k_1) \dots x(n - k_i) \quad (3.42)$$

Les filtres de Volterra ont la propriété remarquable, comme il apparaît sur (3.42), d'être linéaires en paramètres (i.e. par rapport aux noyaux). En effet, soient $y_1(n)$ et $y_2(n)$ deux signaux générés selon (3.42) (même excitation $x(n)$), soit

$$y_1(n) = h_{0,1} + \sum_{i=1}^K \sum_{\mathbf{k}=0}^{M-1} h_{i,1}(k_1, \dots, k_i) x(n - k_1) \dots x(n - k_i) \quad (3.43)$$

et

$$y_2(n) = h_{0,2} + \sum_{i=1}^K \sum_{\mathbf{k}=0}^{M-1} h_{i,2}(k_1, \dots, k_i) x(n - k_1) \dots x(n - k_i) \quad (3.44)$$

Alors

$$y_1(n) + y_2(n) = h_0 + \sum_{i=1}^K \sum_{\mathbf{k}=0}^{M-1} h_i(k_1, \dots, k_i) x(n - k_1) \dots x(n - k_i) \quad (3.45)$$

où

$$h_0 = h_{0,1} + h_{0,2} \quad (3.46)$$

et

$$h_i(k_1, \dots, k_i) = h_{i,1}(k_1, \dots, k_i) + h_{i,2}(k_1, \dots, k_i) \quad (3.47)$$

et, pour toute constante réelle a

$$ay_1(n) = h_0 + \sum_{i=1}^K \sum_{\mathbf{k}=0}^{M-1} h_i(k_1, \dots, k_i) x(n - k_1) \dots x(n - k_i) \quad (3.48)$$

où

$$h_0 = ah_{0,1} \quad (3.49)$$

et

$$h_i(k_1, \dots, k_i) = ah_{i,1}(k_1, \dots, k_i) \quad (3.50)$$

Cette propriété n'est pas uniquement attribuable aux filtres discrets à ordre et mémoire finis, mais aussi aux filtres à temps continu à ordre et mémoire infinis. Par contre, la linéarité en paramètres ajoutée à la finitude de l'ordre et de la mémoire suggère une implantation aisée des filtres de Volterra sous forme vectorielle [24, 25, 50, 51].

En effet, si l'on s'intéresse au filtre homogène d'ordre i , les termes $x(n - k_1) \dots x(n - k_i)$ peuvent être rangés dans un certain ordre dans un vecteur que nous noterons $\mathbf{x}_{i,n}$, dont l'élément générique est

$$(x(n - k_1) \dots x(n - k_i)) \quad (3.51)$$

De même, et bien sûr en respectant l'ordre utilisé pour créer $\mathbf{x}_{i,n}$, les paramètres $h_i(k_1, \dots, k_i)$ sont rangés dans un vecteur que nous noterons \mathbf{h}_i , d'élément générique

$$(h_i(k_1, \dots, k_i)) \quad (3.52)$$

Donc le terme d'ordre i du développement de Volterra s'écrit simplement $\mathbf{h}_i^T \mathbf{x}_{i,n}$. Pour obtenir le développement complet, les vecteurs \mathbf{x}_n et \mathbf{h} sont alors créés selon

$$\mathbf{x}_n^T = (\mathbf{x}_{1,n}^T \dots \mathbf{x}_{K,n}^T) \quad (3.53)$$

et

$$\mathbf{h}^T = (\mathbf{h}_1^T \dots \mathbf{h}_K^T) \quad (3.54)$$

et le développement (3.42) devient

$$y(n) = h_0 + \mathbf{h}^T \mathbf{x}_n \quad (3.55)$$

Le terme h_0 constant sert à régler la moyenne de $y(n)$. Par exemple, et cette convention sera adoptée dans la suite, si $y(n)$ est de moyenne nulle, il faudra alors choisir $h_0 = -\mathbf{h}^T \cdot E[\mathbf{x}_n]$ [24, 50].

Remarque : pour alléger l'écriture, le vecteur centré sera noté de la même façon que le vecteur non centré, soit symboliquement, $\mathbf{x}_n = \mathbf{x}_n - E[\mathbf{x}_n]$. Dans la suite, sauf si h_0 intervient explicitement, la sortie du filtre sera supposée centrée, et par suite le vecteur \mathbf{x}_n aussi.

A des fins d'optimisation, étant donnée la structure linéaire présentée avant, il est intéressant de présenter rigoureusement le filtrage de Volterra d'ordre et mémoire finis dans un contexte d'espaces vectoriels. Nous nous intéressons maintenant à des signaux aléatoires, et nous souhaitons pour des raisons physiques que les sorties de filtres de Volterra soient des signaux au pire du deuxième ordre. Définissons l'ensemble suivant

$$VF_{x,K,M} = \{y(n) = \mathbf{h}^T \mathbf{x}_n \quad / x(n) \in f^{(2K)} \quad \text{et} \quad \sum_i |h_i| < +\infty\} \quad (3.56)$$

où h_i est le i ème élément de \mathbf{h} . Alors $VF_{x,K,M}$ est un sous-espace vectoriel de l'espace vectoriel des signaux aléatoires du second ordre. En effet, si $y_1(n)$ et $y_2(n)$ sont deux éléments de $VF_{x,K,M}$, et si a est un réel quelconque, alors d'après la remarque sur la linéarité en paramètres des filtres, $y_1(n) + ay_2(n)$ appartient à $VF_{x,K,M}$. De plus, il est facile de voir que les moments d'ordre 2 de $y(n)$ sont finis car $x(n)$ est dans $f^{(2K)}$.

Nous pouvons alors utiliser sur $VF_{x,K,M}$ le produit scalaire habituel de l'espace des variables ou signaux aléatoires du second ordre, soit $\langle x|y \rangle = E[xy]$. De plus, en utilisant la norme associée $\|x\| = \sqrt{\langle x|x \rangle}$, $VF_{x,K,M}$ est rendu Hilbertien. Cette structure d'espace vectoriel nous sera utile pour l'identification des filtres.

Remarque importante : la taille des vecteurs mis en jeu est considérable puisque égale à $\sum_{i=1}^K M^i$. A titre d'exemple, un filtre linéaire-quadratique ($K = 2$) de mémoire 50 est implanté à l'aide de vecteurs de taille 2550. Cette dimension peut être notablement diminuée en considérant les noyaux $h_i(k_1, \dots, k_i)$ insensibles à toute permutation d'indices. Nous dirons que les noyaux sont symétriques (si la fonction d'un nombre fini de variables dont le filtre de Volterra étudié est un développement est continuellement différentiable jusqu'à l'ordre du filtre, alors le théorème de Schwartz sur l'inversion de l'ordre des différentiations conduit à cette symétrie des noyaux). Nous supposons par la suite cette symétrie. Nous pouvons alors réécrire le développement de Volterra sous la forme

$$y(n) = h_0 + \sum_{i=1}^K \sum_{k_i \geq k_{i-1}} h_i(k_1, \dots, k_i) x(n - k_1) \dots x(n - k_i) \quad (3.57)$$

Les vecteurs \mathbf{h}_i et $\mathbf{x}_{i,n}$ créés avec ces termes ne comportent plus que $\frac{(M+i-1)!}{(M-1)!i!}$ composantes, et les vecteurs \mathbf{h} et \mathbf{x}_n contiennent donc $\sum_{i=1}^K \frac{(M+i-1)!}{(M-1)!i!}$ termes. Evidemment dans les applications présentées par la suite, cette implantation supposant la symétrie des noyaux sera utilisée. De plus, les remarques faites sur les espaces vectoriels restent valides, la dimension de l'espace étant diminuée.

3.2.2 Implantation dans le domaine fréquentiel

Rappelons que la version fréquentielle d'un filtre de Volterra discret à ordre et mémoire finis est

$$Y(m) = h_0 \delta(m) + \sum_{i=1}^K \sum_{m_1, \dots, m_{i-1}} X(m_1) \dots X(m_{i-1}) X(m - m_1 - \dots - m_{i-1}) H_i(m_1, \dots, m_{i-1}, m - m_1 - \dots - m_{i-1}) \quad (3.58)$$

L'important est encore ici de remarquer la linéarité en paramètres de l'expression et par suite, une implantation linéaire est envisageable. En effet, l'expression précédente se réécrit

$$Y(m) = \mathbf{H}_m^T \mathbf{X}_m \quad (3.59)$$

où

$$\mathbf{X}_m^T = (X(m)\mathbf{X}_{2,m}^T \dots \mathbf{X}_{K,m}^T) \quad (3.60)$$

et

$$\mathbf{H}_m^T = (H_{1,m}\mathbf{H}_{2,m}^T \dots \mathbf{H}_{K,m}^T) \quad (3.61)$$

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.

Remarque tout aussi importante : aucune considération de symétrie n'a été prise en compte. Or, la symétrie supposée des noyaux $h_i(k_1, \dots, k_i)$ implique évidemment des symétries sur leur transformée de Fourier $H_i(m_1, \dots, m_i)$. L'étude de ces symétries étant délicate pour un ordre quelconque, nous montrons à titre d'exemple leur influence sur la partie quadratique d'un filtre de Volterra. Soit $h_2(k_1, k_2)$ le noyau quadratique vérifiant $h_2(k_1, k_2) = h_2(k_2, k_1)$. La fonction de transfert associée est la transformée de Fourier discrète de $h_2(k_1, k_2)$ et s'écrit

$$H_2(m_1, m_2) = \sum_k h_2(k_1, k_2) \exp(-2i\pi\mathbf{m}^T \mathbf{k}/M) = H_2(m_2, m_1) \quad (3.62)$$

De plus,

$$H_2(-m_1, -m_2) = \sum_k h_2(k_1, k_2) \exp(+2i\pi\mathbf{m}^T \mathbf{k}/M) = \left\{ \sum_k h_2(k_1, k_2) \exp(-2i\pi\mathbf{m}^T \mathbf{k}/M) \right\}^* \quad (3.63)$$

car $h_2(k_1, k_2)$ est réel, soit finalement

$$H_2(-m_1, -m_2) = H_2(m_1, m_2)^* \quad (3.64)$$

De plus, pour la fréquence m de la sortie Y , seuls les X d'arguments (m_1, m_2) tels que $m_1 + m_2 = m$ se couplent. Donc, si (m_1, m_2) varie dans $\{-N, N\}^2$, il suffira de connaître $H_2(m_1, m_2)$ sur le domaine

$$D = \{(m_1, m_2) \in \{-N, N\}^2\} \cap \{m_2 \leq N - m_1\} \cap \{m_2 \geq -N - m_1\} \quad (3.65)$$

dont la représentation est donnée en hachuré sur la figure (3.1). La ligne en gras représente les composantes du vecteur $\mathbf{X}_{2,m}$ qui participent à la fréquence m de la sortie. Ainsi, dans $\mathbf{X}_{2,m}$ les termes extra-diagonaux seront affectés d'un facteur 2 du à la symétrie, alors que le terme diagonal ne sera compté qu'une fois.

En conclusion à ce paragraphe, rappelons le point essentiel qui sera dorénavant utilisé constamment : les filtres de Volterra sont linéaires en paramètres. Une application importante de l'implantation qui découle de cette propriété est l'identification en Moyenne Quadratique des filtres.

3.3 Identification en Moyenne Quadratique des filtres de Volterra

Nous avons vu précédemment le fait remarquable que dans sa version discrète, ordre et mémoire finis, le filtre de Volterra est linéaire en paramètres (que ce soit dans le domaine temporel ou dans le domaine fréquentiel). Cette constatation a permis de représenter le filtre sous forme de produit scalaire. 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 temporels et fréquentiels avant de présenter deux autres idées importantes: l'implantation adaptative des algorithmes et l'identification orthogonale des filtres de Volterra.

3.3.1 Identification dans le domaine temporel

Soient $x(n)$ et $y(n)$ deux signaux appartenant respectivement aux classes $f^{(2K)}$ et $f^{(2)}$, 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. Etant donnés les paragraphes précédents, trouver le meilleur filtre de Volterra revient à trouver le meilleur vecteur \mathbf{h}_{opt} tel que $\mathbf{h}_{opt}^T \mathbf{x}_n$ approche au mieux le signal de sortie $y(n)$. De plus,

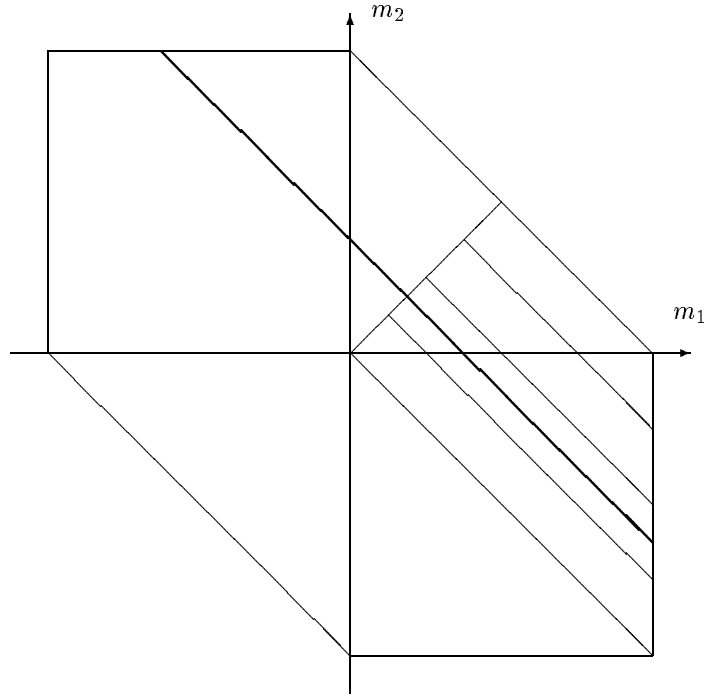


Figure 3.1: Support de la fonction de transfert quadratique.

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

$$\mathbf{h}_{opt} = \text{Arg Min}_{\mathbf{h}} E[(y(n) - \mathbf{h}^T \mathbf{x}_n)^2] = \text{Arg Min}_{\mathbf{h}} E[(y(n) - \hat{y}(n))^2] \quad (3.66)$$

où $\hat{y}(n) = \mathbf{h}^T \mathbf{x}_n$. La solution à cette équation existe puisque la fonction à minimiser est convexe 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 (3.66) 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 ce signal sur cet espace. À l'aide du produit scalaire introduit précédemment, le théorème de la projection orthogonale se traduit par

$$\langle \mathbf{h}^T \mathbf{x}_n | (y(n) - \mathbf{h}_{opt}^T \mathbf{x}_n) \rangle = 0 \quad \forall \quad \mathbf{h}^T \mathbf{x}_n \in VF_{x,K,M} \quad (3.67)$$

Cette dernière équation se résout aisément, la solution étant

$$E[\mathbf{x}_n \mathbf{x}_n^T] \mathbf{h}_{opt} = E[y(n) \mathbf{x}_n] \quad (3.68)$$

$E[\mathbf{x}_n \mathbf{x}_n^T]$ qui sera notée $\mathbf{C}_{\mathbf{xx}}$ est la matrice de covariance du vecteur \mathbf{x}_n , et $E[y(n) \mathbf{x}_n]$ noté $\mathbf{R}_{y\mathbf{x}}$ est le vecteur intercorrélations entre la donnée $y(n)$ et le vecteur \mathbf{x}_n (rappelons que \mathbf{x}_n est centré). La suppression de l'indice n se justifie par la stationnarité de $x(n)$. Ainsi, (3.68) se réécrit

$$\mathbf{C}_{\mathbf{xx}} \mathbf{h}_{opt} = \mathbf{R}_{y\mathbf{x}} \quad (3.69)$$

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 (3.68). En effet, \mathbf{x}_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 cette matrice inversible (l'étude de l'inversibilité de la matrice des moments est faite dans [10]).

Le filtre optimal s'écrit finalement

$$\mathbf{h}_{opt} = \mathbf{C}_{\mathbf{X}\mathbf{X}}^{-1} \mathbf{R}_{y\mathbf{x}} \quad (3.70)$$

L'erreur d'estimation pour le filtre optimal peut donc être calculée selon

$$Eq_{min} = E[(y(n) - \mathbf{h}_{opt}^T \mathbf{x}_n)^2] \quad (3.71)$$

qui devient immédiatement en vertu du théorème de la projection orthogonale

$$Eq_{min} = E[y(n)(y(n) - \mathbf{h}_{opt}^T \mathbf{x}_n)] = E[y(n)^2] - \mathbf{R}_{y\mathbf{x}}^T \mathbf{C}_{\mathbf{X}\mathbf{X}}^{-1} \mathbf{R}_{y\mathbf{x}} \quad (3.72)$$

Il est montré dans [17, 25] que l'augmentation de l'ordre du filtre diminue l'erreur quadratique minimale. Nous remonterons ce résultat au paragraphe 3.3.3 d'une façon très simple.

Dans le paragraphe suivant, l'identification est opérée dans le domaine fréquentiel.

3.3.2 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

$$Y(m) = \mathbf{H}_m^T \mathbf{X}_m \quad (3.73)$$

où les vecteurs \mathbf{H}_m et \mathbf{X}_m sont définis au paragraphe 3.2.2. 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

$$\mathbf{H}_{m,opt} = Arg Min_{\mathbf{H}_m} E[|Y(m) - \mathbf{H}_m^T \mathbf{X}_m|^2] = Arg Min_{\mathbf{H}_m} E[|Y(m) - \hat{Y}(m)|^2] \quad (3.74)$$

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 (3.74) est

$$\mathbf{H}_{m,opt} = \mathbf{C}_{\mathbf{X}_m \mathbf{X}_m}^{-1} \mathbf{R}_{Y(m)\mathbf{X}_m} \quad (3.75)$$

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

$$Eq_{min}(m) = E[|Y(m) - \mathbf{H}_{m,opt}^T \mathbf{X}_m|^2] = E[|Y(m)|^2] - \mathbf{R}_{Y(m)\mathbf{X}_m}^{\dagger} \mathbf{C}_{\mathbf{X}_m \mathbf{X}_m}^{-1} \mathbf{R}_{Y(m)\mathbf{X}_m} \quad (3.76)$$

Dans le paragraphe suivant, nous envisageons une variante de cette procédure, l'identification orthogonale.

3.3.3 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 (3.69). 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

$$\begin{pmatrix} \mathbf{C}_{\mathbf{x}_l \mathbf{x}_l} & \mathbf{C}_{\mathbf{x}_l \mathbf{x}_q} \\ \mathbf{C}_{\mathbf{x}_q \mathbf{x}_l} & \mathbf{C}_{\mathbf{x}_q \mathbf{x}_q} \end{pmatrix} \begin{pmatrix} \mathbf{h}_l \\ \mathbf{h}_q \end{pmatrix} = \begin{pmatrix} \mathbf{R}_{y\mathbf{x}_l} \\ \mathbf{R}_{y\mathbf{x}_q} \end{pmatrix} \quad (3.77)$$

Ce détail montre que les parties linéaire et quadratique du filtre optimal sont couplées via les statistiques d'ordre 3 du signal d'entrée, contenues dans les matrices $\mathbf{C}_{\mathbf{x}_l \mathbf{x}_q}$ [25]. 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 [63] (voir aussi [57]), il réalise une orthogonalisation des fonctionnelles de Volterra sujettes à une entrée gaussienne, et introduit les G-fonctions, système orthogonal de polynômes à mémoire pour une entrée gaussienne. Plus récemment, cette méthode a été reprise par 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 [42]

$$y(n) = \mathbf{h}^T \mathbf{x}_n = \sum_{i=0}^{P-1} a_i p_i(n) \quad (3.78)$$

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 $i = 0 \dots M$, $p_{M+1}(n) = x^2(n)$, $p_{M+2}(n) = x(n)x(n-1)$, etc. . . . Nous souhaitons maintenant réécrire ce développement sur un système de polynômes orthogonaux, soit

$$y(n) = \sum_{i=0}^{P-1} \alpha_i \pi_i(n) \quad (3.79)$$

où $\langle \pi_i(n) | \pi_j(n) \rangle = \text{cte.} \delta_{ij}$, 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

$$\alpha_{k,opt} = \frac{\langle y(n) | \pi_k(n) \rangle}{\langle \pi_k(n) | \pi_k(n) \rangle} \quad \forall k = 0 \dots P-1 \quad (3.80)$$

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, $\langle x|y \rangle = \frac{1}{T} \sum_{i=0}^{T-1} x(i)y(i)$. Si la réalisation provient d'un processus ergodique, il y a équivalence entre les deux approches (si T est effectivement le support du processus; sinon, cela est vrai dans la limite $T \rightarrow \infty$).

L'obtention des polynômes orthogonaux peut s'effectuer tout simplement en utilisant la procédure de Gram-Schmit, en utilisant le produit scalaire adéquat. Lorsque l'identification est effectuée, le calcul de l'erreur quadratique d'estimation de $y(n)$ est

$$Eq_{min} = \langle y(n) | y(n) \rangle - \sum_{i=0}^{P-1} \alpha_{i,opt}^2 \langle \pi_i(n) | \pi_i(n) \rangle \quad (3.81)$$

Cette expression, plus simple que celle obtenue dans le cadre classique ((3.72) du paragraphe 3.3.1) apporte une démonstration claire du fait que l'ajout d'un ordre de non-linéarité résulte en 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.

Remarque 1 : cette procédure ne fait pas que découpler les filtres homogènes entre eux (découplage du linéaire et du quadratique pour l'ordre 2), mais rend diagonale la matrice des moments. Donc, les $\mathbf{x}_{i,n}$ sont non seulement décorrélés entre eux après orthogonalisation, mais de plus ils sont rendus blancs à l'ordre⁷ 2.

Remarque 2 : lorsque l'entrée du filtre est gaussienne, les polynômes orthogonaux naturels sont les polynômes d'Hermite multivariés. Les coefficients optimaux en moyenne quadratique sont alors les inter-cumulants entre l'entrée et la sortie prise à divers instants. On retrouve ici la remarque faite dans le chapitre 2 sur la notion d'orthogonalité qui est naturellement attachée aux cumulants.

⁷Pour la notion de blancheur à un certain ordre, voir [11].

3.4 Implantation adaptative des algorithmes

L'implantation adaptative 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.

Les formes adaptatives 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.

3.4.1 Identification adaptative dans le domaine temporel

Le critère quadratique à minimiser s'écrit

$$Eq(\mathbf{h}) = E[(y(n) - \mathbf{h}^T \mathbf{x}_n)^2] \quad (3.82)$$

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

$$\mathbf{h}_{k+1} = \mathbf{h}_k - \frac{1}{2} \mu_k \nabla_{\mathbf{h}} Eq(\mathbf{h}) \quad (3.83)$$

qui devient en effectuant explicitement la différentiation

$$\mathbf{h}_{k+1} = \mathbf{h}_k + \mu_k E[\mathbf{x}_{k+1}(y(k+1) - \mathbf{h}_k^T \mathbf{x}_{k+1})] \quad (3.84)$$

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

$$\mathbf{h}_{k+1} = \mathbf{h}_k + \mu_k \mathbf{x}_{k+1}(y(k+1) - \mathbf{h}_k^T \mathbf{x}_{k+1}) \quad (3.85)$$

Remarque : Dans les algorithmes précédents, le paramètre μ_k est choisi dépendant du temps pour les cas stationnaires. 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 [8].

Il est possible de montrer (ce résultat est classique) que l'algorithme du gradient converge en moyenne si le pas μ est inférieur à $\frac{2}{\alpha_{max}}$ où α_{max} est la plus grande valeur propre de la matrice covariance du vecteur de régression, soit la matrice $\mathbf{C}_{\mathbf{xx}}$. En effet, en retranchant dans (3.85) le filtre optimal \mathbf{h}_{opt} puis en prenant l'espérance de l'expression obtenue, il vient

$$\epsilon_{k+1} = \epsilon_k + \mu(\mathbf{R}_{y\mathbf{x}} - E[\mathbf{x}_{k+1}\mathbf{x}_{k+1}^T \mathbf{h}_k]) \quad (3.86)$$

où $\epsilon_{k+1} = E[\mathbf{h}_{k+1} - \mathbf{h}_{opt}]$ est l'erreur d'estimation instantanée. En faisant l'hypothèse que $x(n)$ est une suite de variables indépendantes et en utilisant l'expression du filtre optimal, il vient⁸

$$\epsilon_{k+1} = (\mathbf{I} - \mu \mathbf{C}_{\mathbf{xx}}) \epsilon_k \quad (3.87)$$

$$= (\mathbf{I} - \mu \mathbf{C}_{\mathbf{xx}})^{k+1} \epsilon_0 \quad (3.88)$$

⁸L'analyse développée ici est erronée !, ceci est bien connu. La seule indépendance de la suite des $x(n)$ ne suffit pas pour écrire la loi d'évolution qui suit, puisque les vecteurs \mathbf{x}_{k+1} et \mathbf{x}_k ne sont bien sûr pas indépendants. Toutefois, les résultats apportés sont "relativement" corrects. Il faudrait en toute rigueur utiliser les résultats de [8] ou la théorie de la M-indépendance [45] pour montrer la convergence en moyenne de cet algorithme. La M-indépendance est utilisé par P. Duvaut dans [27] dans le cas linéaire-quadratique.

Pour que l'algorithme converge en moyenne, soit $\lim_{k \rightarrow \infty} \epsilon_k = 0$, il faut donc que la matrice $\mathbf{I} - \mu \mathbf{C}_{\mathbf{X}\mathbf{X}}$ ait ses valeurs propres comprises entre -1 et 1 strictement. Or, ces valeurs propres λ_i sont reliées à celles de la matrice covariance α_i par $\alpha_i = \frac{1-\lambda_i}{\mu}$. Imposer que $|\lambda_i| < 1$ pour tout i revient alors à imposer que $0 < \mu\alpha_i < 2$ pour tout i . Cette condition est assurée si $\mu < \frac{2}{\alpha_{max}}$.

Algorithme des moindres carrés récursifs (RLS)

Nous avons vu que le filtre optimal s'écrit sous la forme

$$\mathbf{h}_{opt} = \mathbf{C}_{\mathbf{X}\mathbf{X}}^{-1} \mathbf{R}_{y\mathbf{X}} \quad (3.89)$$

Nous pouvons utiliser cette expression pour le calcul d'une estimation de \mathbf{h}_{opt} , en remplaçant $\mathbf{C}_{\mathbf{X}\mathbf{X}}^{-1}$ et $\mathbf{R}_{y\mathbf{X}}$ par des estimateurs $\widehat{\mathbf{C}}_{\mathbf{X}\mathbf{X}k}^{-1}$ et $\widehat{\mathbf{R}}_{y\mathbf{X}k}$.

Or, $\mathbf{R}_{y\mathbf{X}} = E[y(n)\mathbf{x}_n]$ peut être approché par l'estimateur naturel

$$\widehat{\mathbf{R}}_{y\mathbf{X}k} = \frac{1}{k} \sum_{i=1}^k y(i)\mathbf{x}_i \quad (3.90)$$

qui s'écrit de façon récursive selon

$$\widehat{\mathbf{R}}_{y\mathbf{X}k} = \frac{k-1}{k} \widehat{\mathbf{R}}_{y\mathbf{X}k-1} + \frac{1}{k} y(k)\mathbf{x}_k \quad (3.91)$$

De même, un estimateur naturel de la matrice covariance du vecteur \mathbf{x}_n est

$$\widehat{\mathbf{C}}_{\mathbf{X}\mathbf{X}k} = \frac{1}{k} \sum_{i=1}^k \mathbf{x}_i \mathbf{x}_i^T \quad (3.92)$$

$$\widehat{\mathbf{C}}_{\mathbf{X}\mathbf{X}k} = \frac{k-1}{k} \widehat{\mathbf{C}}_{\mathbf{X}\mathbf{X}k-1} + \frac{1}{k} \mathbf{x}_k \mathbf{x}_k^T \quad (3.93)$$

Or, nous souhaitons avoir un estimateur de l'inverse de la matrice covariance. A partir de la forme récursive précédente, et en utilisant le lemme d'inversion matricielle de Woodbury, il vient

$$\widehat{\mathbf{C}}_{\mathbf{X}\mathbf{X}k}^{-1} = \frac{k}{k-1} \left\{ \widehat{\mathbf{C}}_{\mathbf{X}\mathbf{X}k-1}^{-1} - \frac{\widehat{\mathbf{C}}_{\mathbf{X}\mathbf{X}k-1}^{-1} \mathbf{x}_k \mathbf{x}_k^T \widehat{\mathbf{C}}_{\mathbf{X}\mathbf{X}k-1}^{-1}}{(k-1) + \mathbf{x}_k^T \widehat{\mathbf{C}}_{\mathbf{X}\mathbf{X}k-1}^{-1} \mathbf{x}_k} \right\} \quad (3.94)$$

Ainsi, le filtre estimé à l'instant k se calcule selon (3.89) en utilisant (3.91) et (3.94). L'initialisation de $\widehat{\mathbf{C}}_{\mathbf{X}\mathbf{X}k}^{-1}$ se fait traditionnellement par $\widehat{\mathbf{C}}_{\mathbf{X}\mathbf{X}0}^{-1} = cte \cdot \mathbf{I}$ où $cte \gg 1$ et \mathbf{I} représente la matrice identité.

3.4.2 Identification adaptative 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 adaptative 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 kème 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 [49]

$$\mathbf{H}_{m,k+1} = \mathbf{H}_{m,k} - \frac{1}{2}\mu(m)\nabla_{\mathbf{H}_m}(E[|Y(m) - \mathbf{H}_m^T \mathbf{X}_m|^2]) \quad (3.95)$$

Le pas d'adaptation dépend explicitement de la fréquence puisque l'identification se fait canal par canal. Effectuer la différentiation conduit à⁹

$$\mathbf{H}_{m,k+1} = \mathbf{H}_{m,k} + \mu(m)E[(Y(m)_{k+1} - \mathbf{X}_{m,k+1}^T \mathbf{H}_{m,k})\mathbf{X}_{m,k+1}^*] \quad (3.96)$$

La version stochastique de cet algorithme est alors la suivante

$$\mathbf{H}_{m,k+1} = \mathbf{H}_{m,k} + \mu(m)(Y(m)_{k+1} - \mathbf{X}_{m,k+1}^T \mathbf{H}_{m,k})\mathbf{X}_{m,k+1}^* \quad (3.97)$$

Algorithme RLS

L'approche adoptée pour le domaine temporel peut être reprise ici [41, 49]. 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 covariance du vecteur de régression. L'algorithme RLS est alors

$$\begin{cases} \mathbf{C}_{\mathbf{X}_m \mathbf{X}_m, k+1}^{-1} &= \frac{k}{k-1} \left\{ \mathbf{C}_{\mathbf{X}_m \mathbf{X}_m, k}^{-1} - \frac{\mathbf{C}_{\mathbf{X}_m \mathbf{X}_m, k}^{-1} \mathbf{X}_{m, k+1}^* \mathbf{X}_{m, k+1}^T \mathbf{C}_{\mathbf{X}_m \mathbf{X}_m, k}^{-1}}{k-1 + \mathbf{X}_{m, k+1}^T \mathbf{C}_{\mathbf{X}_m \mathbf{X}_m, k}^{-1} \mathbf{X}_{m, k+1}^*} \right\} \\ \mathbf{R}_{Y(m) \mathbf{X}_m, k+1} &= \frac{k-1}{k} \mathbf{R}_{Y(m) \mathbf{X}_m, k} + \frac{1}{k} Y(m)_{k+1} \mathbf{X}_{m, k+1}^* \\ \mathbf{H}_{m, k+1} &= \mathbf{C}_{\mathbf{X}_m \mathbf{X}_m, k+1}^{-1} \mathbf{R}_{Y(m) \mathbf{X}_m, k+1} \end{cases} \quad (3.98)$$

3.4.3 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 échantillons. 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 (3.1)). La taille de ces vecteurs est majoré par M . Par fréquence et par bloc, l'algorithme requiert alors un nombre d'opération 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.

⁹On utilise la dérivation formelle $\frac{dzz^*}{dz^*} = 2z$.

3.4.4 Identification orthogonale adaptative dans le domaine temporel

Dans la théorie des espaces vectoriels, une méthode d'orthogonalisation consiste à utiliser la décomposition de Choleski : si \mathbf{A} est carrée, symétrique, définie positive, alors il existe \mathbf{L} triangulaire inférieure telle que $\mathbf{A} = \mathbf{L}\mathbf{L}^T$. Comme $\mathbf{C}_{\mathbf{X}\mathbf{X}}$ suit les hypothèses précédentes, on peut écrire $\mathbf{C}_{\mathbf{X}\mathbf{X}} = \mathbf{L}\mathbf{L}^T$. Alors le vecteur $\mathbf{x}'_n = \mathbf{L}^{-1}\mathbf{x}_n$ est de covariance identité, et par suite a ses composantes orthogonales. En effet

$$\mathbf{C}_{\mathbf{X}'\mathbf{X}'} = E[\mathbf{x}'_n \mathbf{x}'_n{}^T] = E[\mathbf{L}^{-1} \mathbf{x}'_n \mathbf{x}'_n{}^T \mathbf{L}^{-1T}] = \mathbf{I} \quad (3.99)$$

Le filtre optimal calculé dans la nouvelle base est alors

$$\mathbf{h}_{opt} = \mathbf{C}_{\mathbf{X}'\mathbf{X}'}^{-1} \mathbf{R}_{y\mathbf{X}'} = E[y(n) \mathbf{x}'_n] \quad (3.100)$$

Dans le but de traiter des signaux dans des cas non stationnaires ou en temps réel, il est nécessaire de rendre adaptative la procédure d'orthogonalisation, qui sera suivie par l'estimation adaptative du filtre. La procédure utilisée est décrite dans [37] et déjà employée par P. Comon [21] dans le cadre d'une orthogonalisation de signaux à l'ordre 2. Cette procédure est liée à l'estimation de la matrice covariance. En effet, on se propose, lorsque l'estimation est

$$\widehat{\mathbf{C}}_{\mathbf{X}\mathbf{X}k} = a_k \widehat{\mathbf{C}}_{\mathbf{X}\mathbf{X}k-1} + b_k \mathbf{x}_k \mathbf{x}_k^T \quad (3.101)$$

de modifier récursivement la matrice \mathbf{L}_k , où \mathbf{L}_k résulte de la décomposition de Choleski de $\widehat{\mathbf{C}}_{\mathbf{X}\mathbf{X}k}$, en fonction de \mathbf{L}_k . Soit alors \mathbf{A} la matrice définie par

$$\mathbf{A} = \begin{pmatrix} \sqrt{a_k} \mathbf{L}_{k-1} & \sqrt{b_k} \mathbf{x}_{k-1} \\ \mathbf{0}^T & \frac{1}{\sqrt{b_k}} \end{pmatrix} \quad (3.102)$$

En appliquant une séquence de rotations de Givens sur \mathbf{A} annulant la partie supérieure droite de \mathbf{A} [37], on obtient :

$$\mathbf{B} = \begin{pmatrix} \mathbf{L}_k & \mathbf{0} \\ \mathbf{x}'_k{}^T & \gamma \end{pmatrix} \quad (3.103)$$

et l'on vérifie que

$$\mathbf{L}_k \mathbf{L}_k^T = a_k \mathbf{L}_{k-1} \mathbf{L}_{k-1}^T + b_k \mathbf{x}_k \mathbf{x}_k^T \quad (3.104)$$

et

$$\mathbf{L}_k \mathbf{x}'_k = \mathbf{x}_k \quad (3.105)$$

On récupère donc par cette opération le vecteur à composantes orthogonales sans avoir à inverser la matrice \mathbf{L} . L'estimée du signal utile s'écrira alors simplement

$$\hat{y}(n) = \widehat{\mathbf{R}}_{y\mathbf{X}'}^T \mathbf{x}'_n \quad (3.106)$$

où $\widehat{\mathbf{R}}_{y\mathbf{X}}$ est l'estimateur récursif défini précédemment.

Remarque : dans le cas d'une entrée gaussienne, les coefficients dans une base de polynômes d'Hermite sont des intercumulants, et l'algorithme adaptatif présenté dans le Préambule pourrait s'appliquer.

3.5 Applications

Nous proposons maintenant deux applications des filtres de Volterra. La première est liée aux multispectres et concerne la détection de non-linéarités. Cette application est plus une idée puisqu'appliquée à aucune donnée. La deuxième utilise les développements de Volterra en tant que filtres et emploie les techniques présentées précédemment.

3.5.1 Modélisation et détection des non-linéarités

Dans beaucoup de domaines de la Physique, l'approximation linéaire est faite, souvent à bon escient, mais également souvent par simplicité. Toutefois, des problèmes curieux ne peuvent être traités dans cette approximation, et des modélisations non linéaires doivent être adoptées. A titre d'exemple, nous reciterons le cas du doublement de fréquence en optique [31].

Lorsque la lumière traverse un matériau dit optique, 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 le champ, 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 ne peut plus être linéaire. Les physiciens ont alors modélisé cette loi selon

$$\mathbf{P} = \chi^1 \mathbf{E} + \chi^2 \mathbf{E}\mathbf{E} + \chi^3 \mathbf{E}\mathbf{E}\mathbf{E} + \dots \quad (3.107)$$

où les termes χ^i sont des tenseurs. Les études se limitent souvent à l'ordre 2, et le développement précédent se réécrit pour la i ème composante de \mathbf{P}

$$P_i = \sum_j \chi_{i,j}^1 E_j + \sum_{j,k} \chi_{i,j,k}^2 E_j E_k \quad (3.108)$$

Cette dernière relation n'est autre qu'un filtre 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 va "générer" 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 à besoin de deux outils. Les filtres de Volterra pour la modélisation, et les multispectres pour la détection. Pour illustrer ceci, nous nous restreindrons encore au cas linéaire-quadratique.

Soit $y(t)$ la sortie d'un filtre linéaire-quadratique attaqué par un signal $x(t)$, centré, gaussien, ayant une transformée de Fourier. La relation fréquentielle entrée-sortie s'écrit

$$Y(\nu) = H_1(\nu) + \int H_2(\nu_1, \nu - \nu_1) X(\nu_1) X(\nu - \nu_1) d\nu_1 \quad (3.109)$$

Le spectre de $y(t)$ pour une fréquence $\nu \neq 0$ est alors, sachant que les moments d'ordre impair de $x(t)$ sont nuls

$$S_{y,2}(\nu) = |H_1(\nu)|^2 S_{x,2}(\nu) + \int H_2(\nu_1, \nu - \nu_1) H_2^*(\nu_2, \nu - \nu_2) E[X(\nu_1) X(\nu - \nu_1) X^*(\nu_2) X^*(\nu - \nu_2)] d\nu_1 d\nu_2 \quad (3.110)$$

Or, le moment d'ordre 4 s'exprime en fonction de trois produits de moments d'ordre 2. Notons que le moment $E[X(\nu_1) X(\nu - \nu_1)]$ est nul¹⁰, puisque nous avons supposé ν non nul, et qu'il est donc impossible d'avoir $-\nu_1 = \nu - \nu_1$. Par suite, il reste les produits

$$E[X(\nu_1) X^*(\nu_2)] E[X(\nu - \nu_1) X^*(\nu - \nu_2)] + E[X(\nu_1) X^*(\nu - \nu_2)] E[X(\nu - \nu_1) X^*(\nu_2)] = S_{x,2}(\nu_1) S_{x,2}(\nu - \nu_1) \delta(\nu_2 - \nu_1) + S_{x,2}(\nu_1) S_{x,2}(\nu - \nu_1) \delta(\nu_2 - \nu + \nu_1) \quad (3.111)$$

Finalement, en supposant la fonction de transfert d'ordre 2 symétrique

$$S_{y,2}(\nu) = |H_1(\nu)|^2 S_{x,2}(\nu) + 2 \int |H_2(\nu_1, \nu - \nu_1)|^2 S_{x,2}(\nu_1) S_{x,2}(\nu - \nu_1) d\nu_1 \quad (3.112)$$

¹⁰ Voir le chapitre 2 sur les multispectres et les multiplicités stationnaires.

L'interspectre entre $x(t)$ et $y(t)$ est donné par

$$S_{x,y,2}(\nu) = H_1(\nu)^* S_{x,2}(\nu) \quad (3.113)$$

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

$$\gamma_{x,y,2}(\nu) = \frac{|S_{x,y,2}(\nu)|^2}{S_{x,2}(\nu)S_{y,2}(\nu)} \quad (3.114)$$

En utilisant les résultats précédents, elle devient

$$\gamma_{x,y,2}(\nu) = \frac{|H(\nu)|^2 S_{x,2}(\nu)^2}{S_{x,2}(\nu)(|H_1(\nu)|^2 S_{x,2}(\nu) + 2 \int |H_2(\nu_1, \nu - \nu_1)|^2 S_{x,2}(\nu_1) S_{x,2}(\nu - \nu_1) d\nu_1)} \quad (3.115)$$

Dans le cas d'un filtre linéaire seul, elle vaut donc 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 déceler 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, puisque elle s'écrit

$$\gamma_{x,y,3}(\nu_1, \nu_2) = \frac{|S_{x,y,3}(\nu_1, \nu_2)|^2}{S_{x,2}(\nu_1)S_{x,2}(\nu_2)S_{y,2}(\nu_1 + \nu_2)} \quad (3.116)$$

Dans le cas que nous traitons, $S_{x,y,3}(\nu_1, \nu_2) = E[X(\nu_1)X(\nu_1)Y^*(\nu_1 + \nu_2)]$ se réduit à $S_{x,2}(\nu_1)S_{x,2}(\nu_2)H_2(\nu_1, \nu_2)^*$ qui 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 $y(t)$ [41]. En effet, en remplaçant dans (3.116) $H_2(\nu_1, \nu_2)$ par sa valeur, puis en divisant par $S_{y,2}(\nu)$ on obtient

$$\gamma_{x,y,2}(\nu) + \int \gamma_{x,y,3}(\nu_1, \nu - \nu_1) d\nu_1 = 1 \quad (3.117)$$

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 pures¹¹. Reprenons l'exemple présenté au paragraphe 3.1.1 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 (3.116) montre que

$$\gamma_{x,y,3}(\nu_1, \nu_2) = \delta(\nu_1 - \nu_0)\delta(\nu_2 - \nu_0) \quad (3.118)$$

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 quadratique de transfert, et elle sera utilisée dans le prochain paragraphe pour démontrer la présence d'une non-linéarité.

3.5.2 Applications des filtres de Volterra à la Soustraction de Bruit

La soustraction de bruit : c'est quoi?

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 [62]. L'application de cette méthode à l'écoute du coeur d'un foetus est remarquable. En effet, les battements du coeur d'un foetus sont faibles, et sont en grande partie cachés par ceux du coeur de la mère. Widrow propose alors de placer un capteur

¹¹ 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 [22]

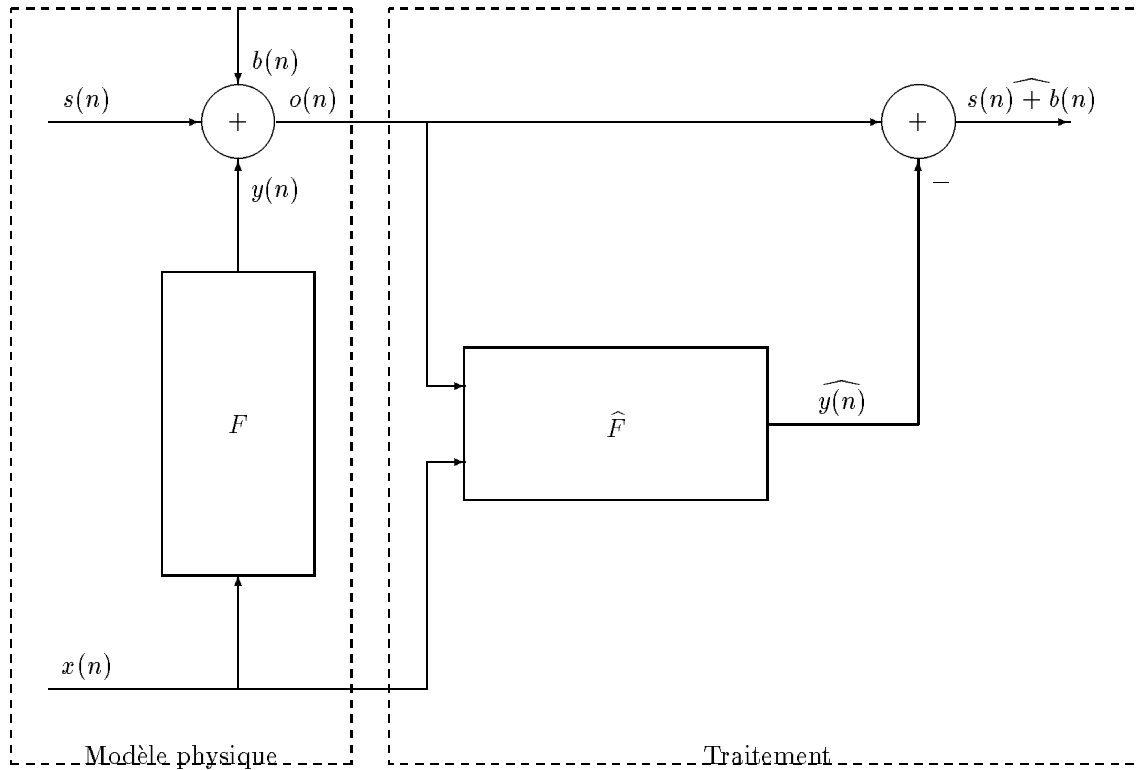


Figure 3.2: Principe de la Soustraction de Bruit.

de référence au niveau du coeur de la mère, et montre comment estimer l'influence de cette référence sur un capteur d'écoute placé près du coeur 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 (3.2).

Les travaux de pionniers 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 trouvés dans [7]. Durant ces deux 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 justement 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 présentons un cas réel issu d'expériences de SONAR.

Filtres de Volterra et Soustraction de Bruit

Considérons à nouveau le schéma de la figure (3.2) en supposant que le filtre liant la référence au bruit est non linéaire. La modélisation du système physique est la suivante

$$\begin{cases} o(n) &= s(n) + y(n) + b(n) \\ y(n) &= F(x(n)) \end{cases} \quad (3.119)$$

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, hypothèse qui sera légèrement affaiblie plus tard. 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 minimum Eq . Ainsi, nous cherchons $\mathbf{h}_{opt}\mathbf{x}_n$ de $VF_{x,K,M}$ tel que

$$\mathbf{h}_{opt} = Arg Min_{\mathbf{h}} E[(y(n) - \mathbf{h}^T \mathbf{x}_n)^2] \quad (3.120)$$

Mais $y(n)$ est inaccessible !

Toutefois, l'hypothèse d'indépendance entre $x(n)$ et $s(n)$ et leur caractère centré permettent de conclure que

$$Eq' = E[(o(n) - \mathbf{h}^T \mathbf{x}_n)^2] = Eq + E[(s(n) + b(n))^2] \quad (3.121)$$

Donc, minimiser Eq est équivalent à minimiser Eq' . Le filtre optimal est donc obtenu en utilisant "l'artillerie" développée précédemment au paragraphe 3.3. Ceci est bien sûr valable dans les domaines temporel et fréquentiel.

Remarque : comme mentionné précédemment, l'hypothèse d'indépendance peut être affaiblie. En effet, les développements précédents restent valides si les intermulticorrélations du type $C_{s,x,1+K}(\tau)$ sont nulles, et nous dirons que $x(n)$ et $s(n)$ sont décorrélés jusqu'à l'ordre K .

Nous donnons maintenant les solutions dans les deux domaines :

- domaine temporel :

$$\mathbf{h}_{opt} = \mathbf{C}_{\mathbf{xx}}^{-1} \mathbf{R}_{y\mathbf{x}} \quad (3.122)$$

L'erreur d'estimation est alors

$$Eq'_{min} = E[(s(n) + b(n))^2] + E[y(n)^2] - \mathbf{R}_{o\mathbf{x}}^T \mathbf{C}_{\mathbf{xx}}^{-1} \mathbf{R}_{o\mathbf{x}} \quad (3.123)$$

- domaine fréquentiel :

$$\mathbf{H}_{m,opt} = \mathbf{C}_{\mathbf{X}_m \mathbf{X}_m}^{-1} \mathbf{R}_{Y(m)\mathbf{X}_m} \quad (3.124)$$

L'erreur d'estimation est alors

$$Eq'_{min}(m) = E[|S(m) + B(m)|^2] + E[|Y(m)|^2] - \mathbf{R}_{O(m)\mathbf{X}_m}^\dagger \mathbf{C}_{\mathbf{X}_m \mathbf{X}_m}^{-1} \mathbf{R}_{O(m)\mathbf{X}_m} \quad (3.125)$$

Pour réaliser pratiquement la soustraction de bruit, les algorithmes adaptatifs présentés avant sont employés, dans lesquels l'observation est $o(n)$ ou $O(m)$. La section suivante est une illustration de cette méthode sur une expérience réelle.

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. Mais, ces bruits étant provoqués par les machines du bateau, il est bien sûr possible de disposer près de ces sources de bruits des capteurs qui donneront des signaux représentatifs de ces bruits : ce seront les références bruit seul. La description faite est résumée par la figure (3.3). Rappelons la modélisation mathématique du problème :

$$\begin{cases} o(n) & = & s(n) + y(n) + b(n) \\ y(n) & = & F(x(n)) \end{cases} \quad (3.126)$$

Effectuons maintenant la description des signaux observés $o(n)$ et $x(n)$. Leurs spectres respectifs apparaissent sur les figures (3.4) et (3.5). La figure (3.6) représente la cohérence entre ces deux signaux. Ces figures appellent quelques commentaires :

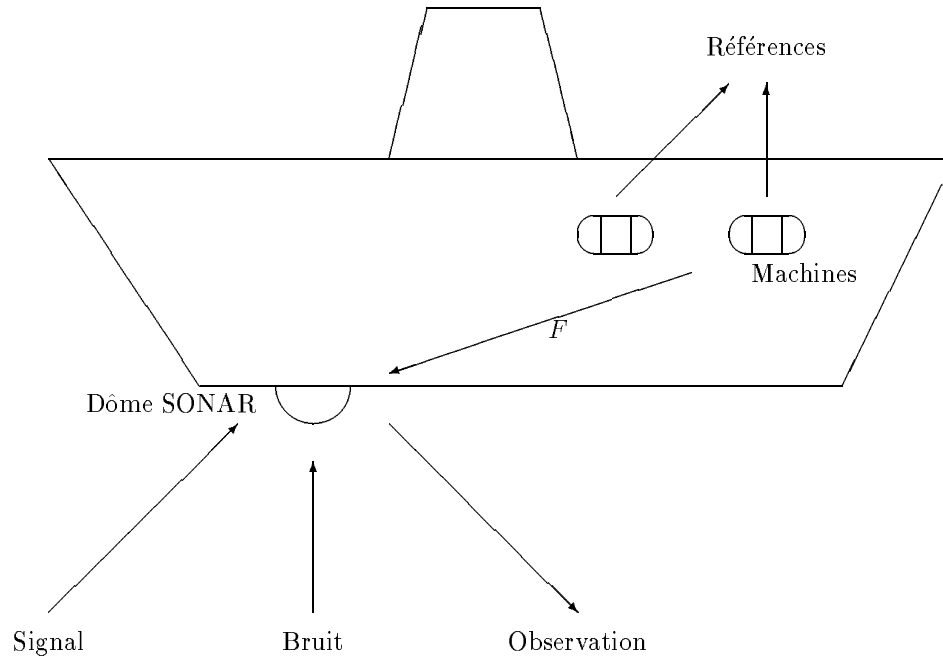


Figure 3.3: Cas réel étudié : écoute du milieu marin par un SONAR.

- Le signal $o(n)$ issu du SONAR semble composé de trois parties distinctes. La première est ce plateau très large bande représentant certainement le bruit de fond. La seconde apparaît sous forme d'une bosse dans la partie $[0, 0.25]$ 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 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. Un outil adapté à cette recherche est la bicohérence, comme nous l'avons montré au paragraphe précédent. Rappelons sa définition

$$\gamma_{x,o,3}(m_1, m_2) = \frac{E[X(m_1)X(m_2)O^*(m_1 + m_2)]}{(E[|X(m_1)|^2]E[|X(m_2)|^2]E[|O(m_1 + m_2)|^2])^{1/2}} \quad (3.127)$$

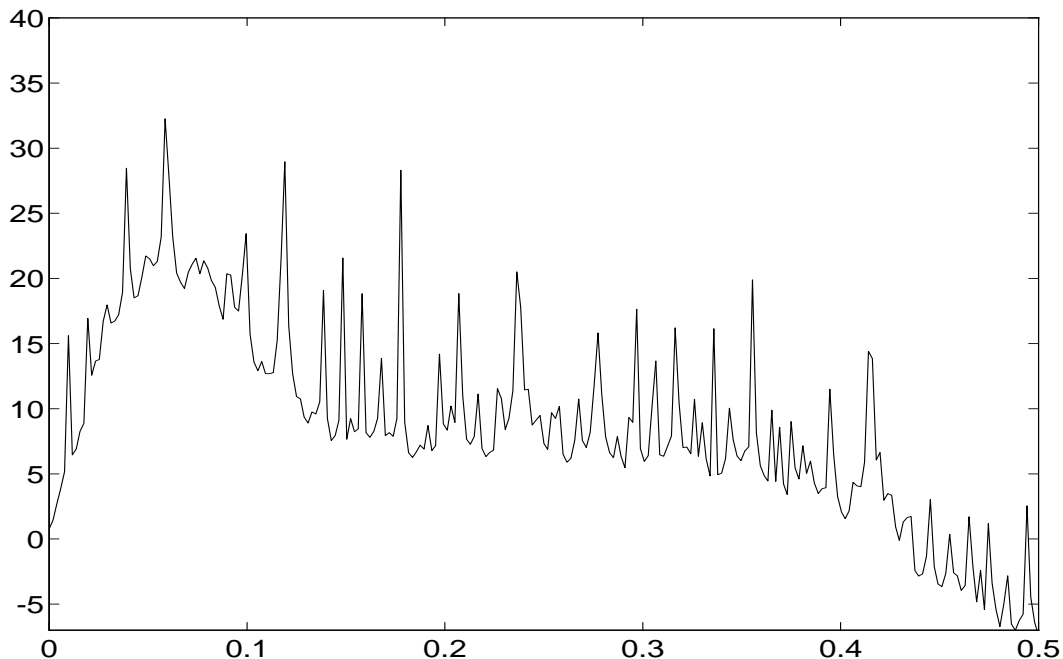


Figure 3.4: Spectre en dB de la sortie du SONAR. 512 canaux de fréquence sont représentés (128 moyennes).

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 (3.7).

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 provient (peut être en partie) de l'interaction non linéaire entre les deux composantes du couple. 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 techniques linéaires par C. Servière [55], J.F Guerre-Chaley [38] et R. Zaknoute [64] dans leurs travaux de Thèse. La conclusion obtenue dans [55] 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 [38], où une approche multiréférences est développée, il est montré que la référence choisie ici donne de très bons résultats lorsqu'utilisée seule. Cette conclusion est confirmée dans [64] 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

Nous avons employé les trois approches décrites précédemment, à savoir, identification d'un filtre linéaire en fréquences (puisque d'après les remarques précédentes elle fournit les meilleurs résultats sur ces signaux), identification d'un filtre linéaire-quadratique dans le domaine temporel et identification dans le domaine fréquentiel. Détaillons maintenant les conditions expérimentales utilisées dans chacune des approches.

- Filtre linéaire identifié en fréquence.

Nous utilisons 128 canaux fréquentiels, et l'adaptation s'est faite sur 1024 blocs.

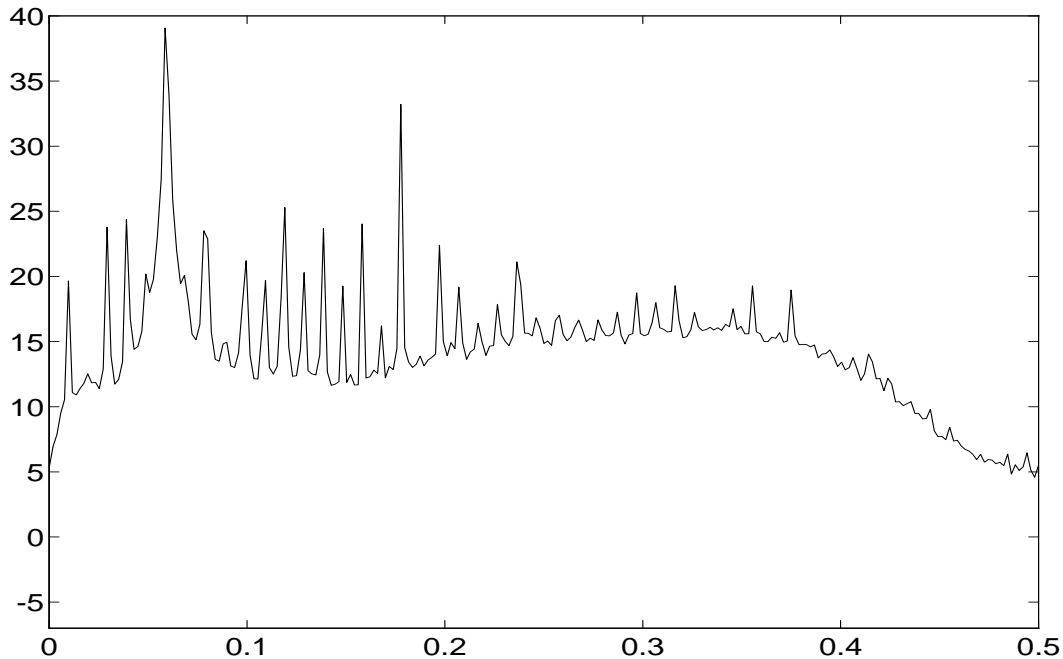


Figure 3.5: Spectre en dB de la référence bruit seul. 512 canaux de fréquence sont représentés (128 moyennes).

- Filtre linéaire-quadratique identifié en temps.

La paramétrisation est ici explicitement transversale. La mémoire choisie est de 15 échantillons (taille des vecteurs de 135). 20000 échantillons sont utilisés pour l'adaptation. Les spectres sont alors calculés sur le signal filtré par le filtre final, et évalués pour 128 canaux sur 128 blocs.

- Filtre linéaire-quadratique identifié en fréquence.

Comme pour le cas linéaire, 1024 blocs nous ont permis l'adaptation du filtre calculé sur 128 canaux de fréquence.

Pour ces trois cas, l'algorithme utilisé est un moindres carrés récursif (RLS).

Remarque : dans le cas non linéaire identifié en fréquence, les signaux ont dû être suréchantillonnés par un facteur 2 pour pouvoir atteindre sans problème les fréquences comprises entre 0.25 et 0.5. En effet, la fréquence f_{max} maximale de la sortie peut provenir de l'interaction quadratique de la fréquence maximale de l'entrée. Celle-ci doit donc être $\frac{f_{max}}{2}$.

La performance de soustraction est chiffrée par la différence entre le spectre en dB de sortie avant et après traitement. Elle est arbitraire, et nous aurions tout aussi bien pu choisir le spectre de la différence entre les signaux 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 (3.8) à (3.13).

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 (3.8) et (3.9)). Le traitement linéaire-quadratique temporel (figures (3.10) et (3.11)) 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). Toutefois, la mémoire limitée choisie n'autorise pas d'amélioration exceptionnelle. Nous pouvons dire que cette franche amélioration est obtenue par l'algorithme fréquentiel, malgré un artefact aux alentours de 0.12 (figures (3.12) et (3.13)).

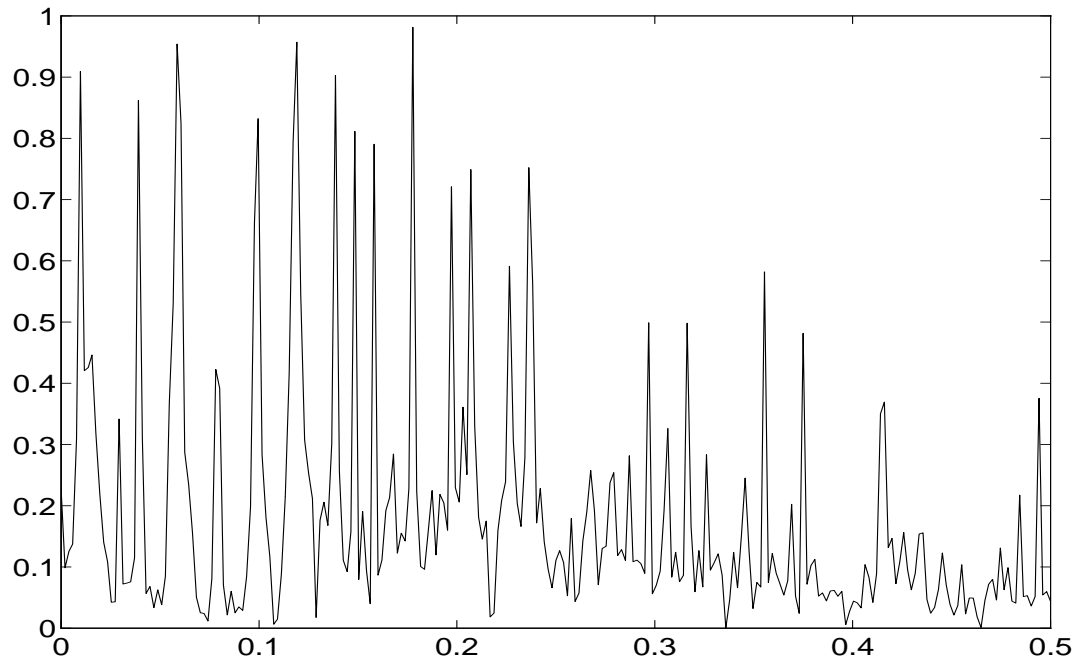


Figure 3.6: Cohérence entre la référence et la sortie SONAR. 512 canaux de fréquence sont représentés (128 moyennes).

Ces résultats appellent quelques commentaires quant à la supériorité de l'algorithme fréquentiel sur les approches temporelles¹². Le choix arbitraire de la mémoire est pour partie responsable du semi-échec de l'approche temporelle. En effet, il n'y a pas équivalence entre le filtre linéaire-quadratique temporel et le filtre fréquentiel, puisqu'il n'y a pas le même nombre de coefficients. Ce choix a simplement été fait pour des problèmes de temps de calcul. Pour l'anecdote, nous avons estimé le temps de calcul pour une mémoire de 128 en temps à . . . une semaine sur les calculateurs dont nous disposons. En effet, ceci requiert la propagation de l'inverse d'une matrice 8378 par 8378! Combien alors faut-il d'échantillons pour arriver à convergence?

Mais nous pensons que cette raison n'est pas la plus profonde. L'argument principal à notre sens réside dans la non-équivalence entre les définitions des filtres temporels à ordre et mémoire finis et les filtres fréquentiels discrets. Ces derniers sont en effet obtenus par discrétisation de la forme fréquentielle continue, et non comme transformée de Fourier de la version temporelle. Ainsi, ce problème repose sur la paramétrisation. Alors qu'en temps les filtres sont explicitement transversaux, les versions fréquentielles peuvent correspondre à des réponses impulsionnelles infinies, d'où leur supériorité quant à la description de raies. De plus, il n'y a pas équivalence entre les critères minimisés en temps et en fréquence. En effet, le critère fréquentiel minimise l'erreur quadratique moyenne fréquence par fréquence, et est donc adapté aux fréquences pures.

3.6 Conclusion

Nous venons de présenter un outil susceptible de prendre en compte les non-linéarités des systèmes, et par suite l'éventuelle non-gaussianité des signaux. Mais répétons que cette non-propriété peut être abordée de diverses façons. Les filtres de Volterra ne supposent rien de connu sur les types de non-linéarités mises en jeu. Par contre, toutes connaissances a priori disponibles doivent être exploitées, et elles peuvent conduire à préférer un modèle particulier aux filtres de Volterra. Ainsi, ces dernières années ont vu l'émergence

¹²La supériorité du non-linéaire sur le linéaire n'en appelle plus!, cette supériorité étant clairement démontrée ici –sur cet exemple particulier bien entendu–.

Figure 3.7: Bicohérence entre la référence et la sortie SONAR. 512 canaux fréquentiels utilisés (128 moyennes).

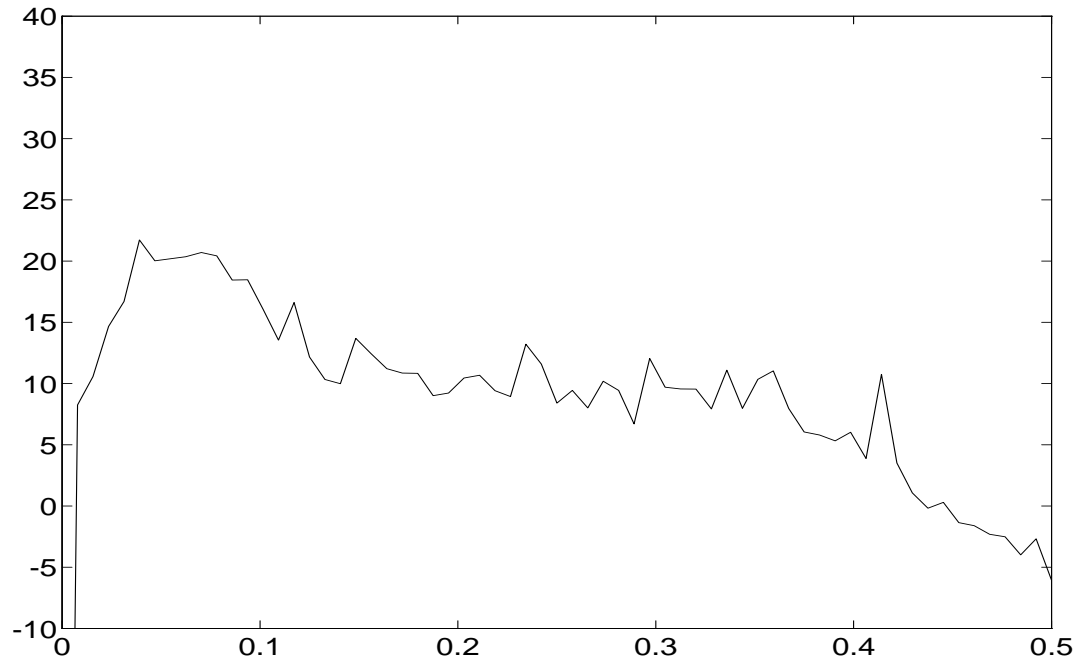


Figure 3.8: Spectre en dB de l'estimée du signal utile par traitement linéaire en fréquence.

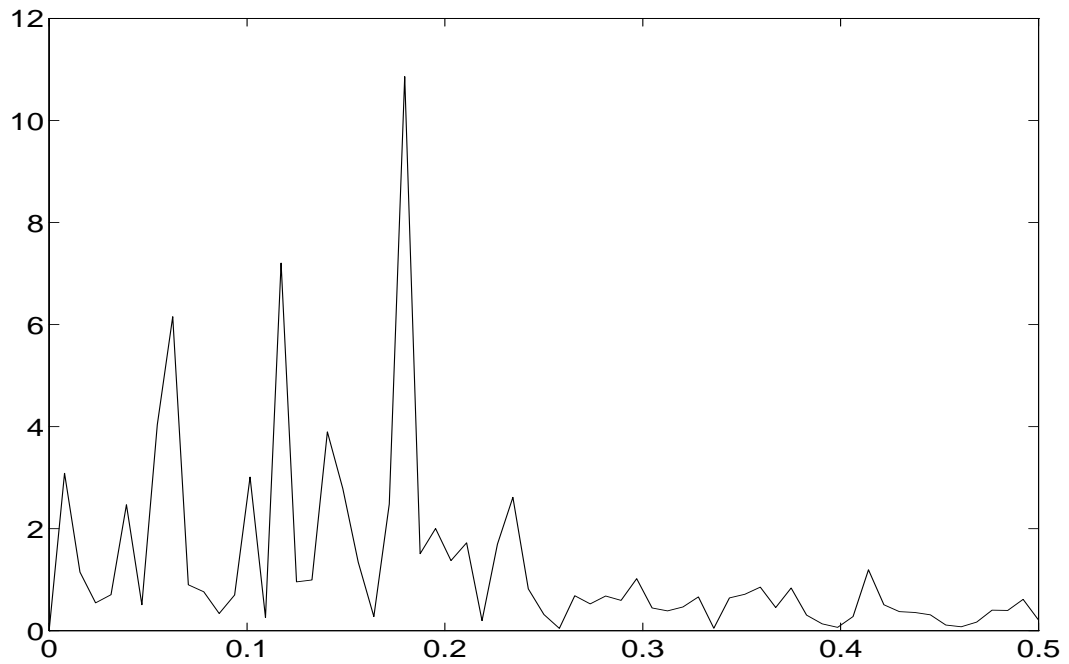


Figure 3.9: Gain en dB pour le traitement linéaire.

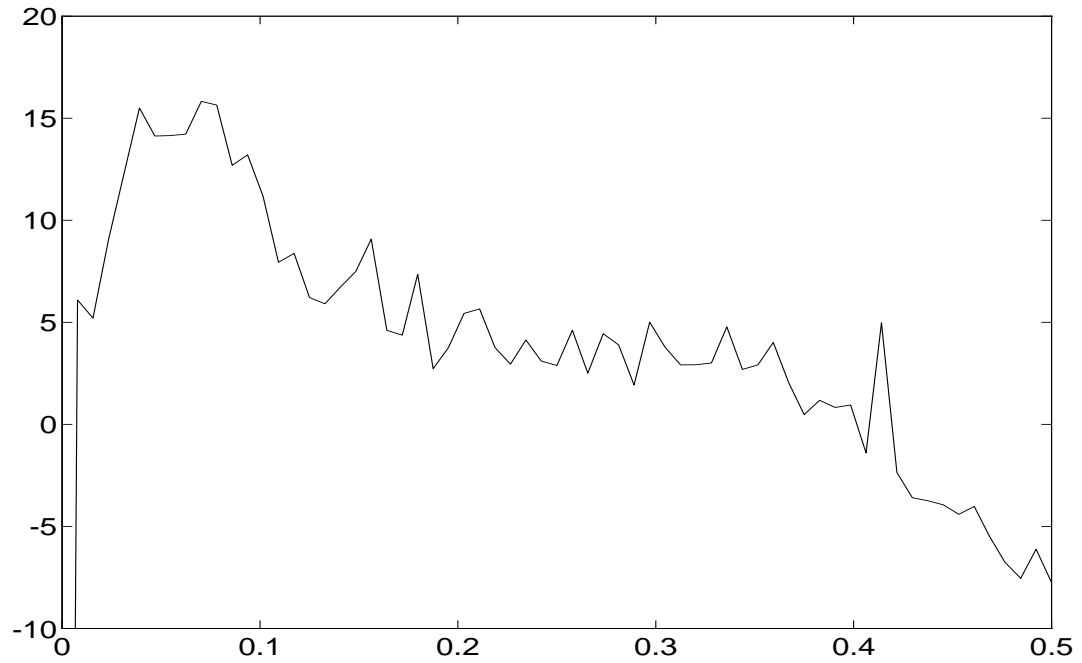


Figure 3.10: Spectre en dB de l'estimée du signal utile par traitement linéaire-quadratique en temps.

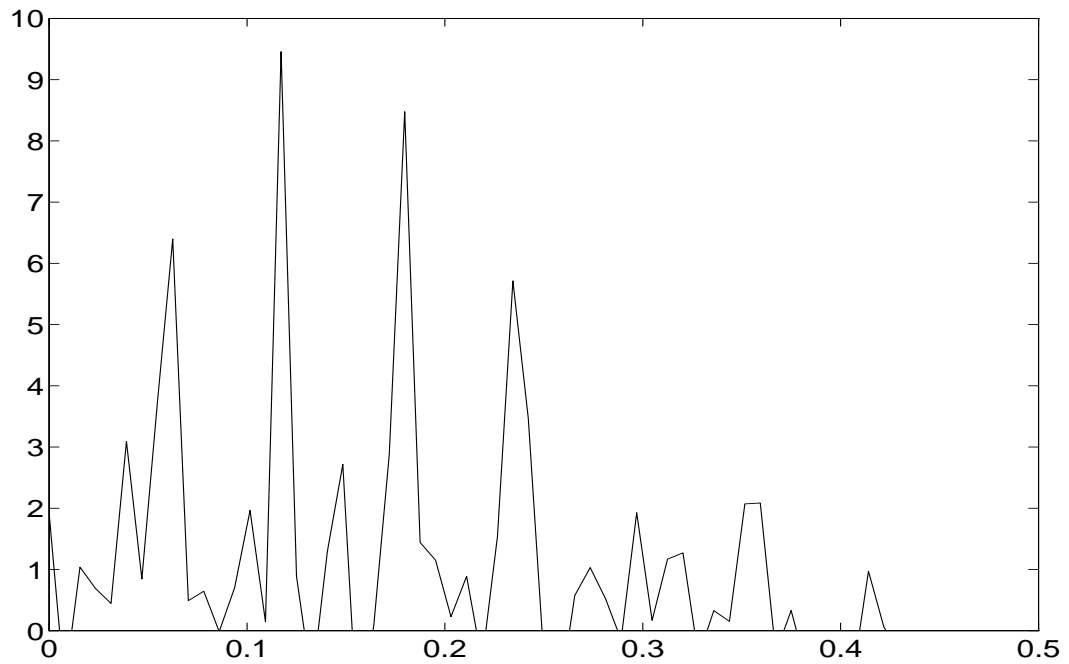


Figure 3.11: Gain en dB pour le traitement linéaire-quadratique en temps.

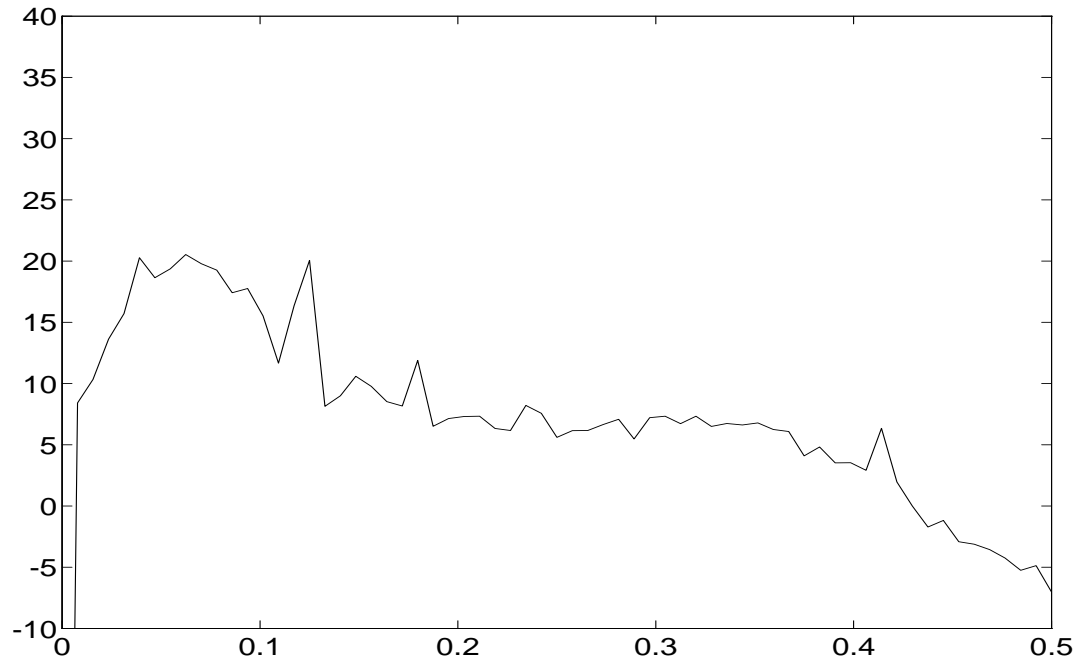


Figure 3.12: Spectre en dB de l'estimée du signal utile par traitement linéaire-quadratique en fréquence.

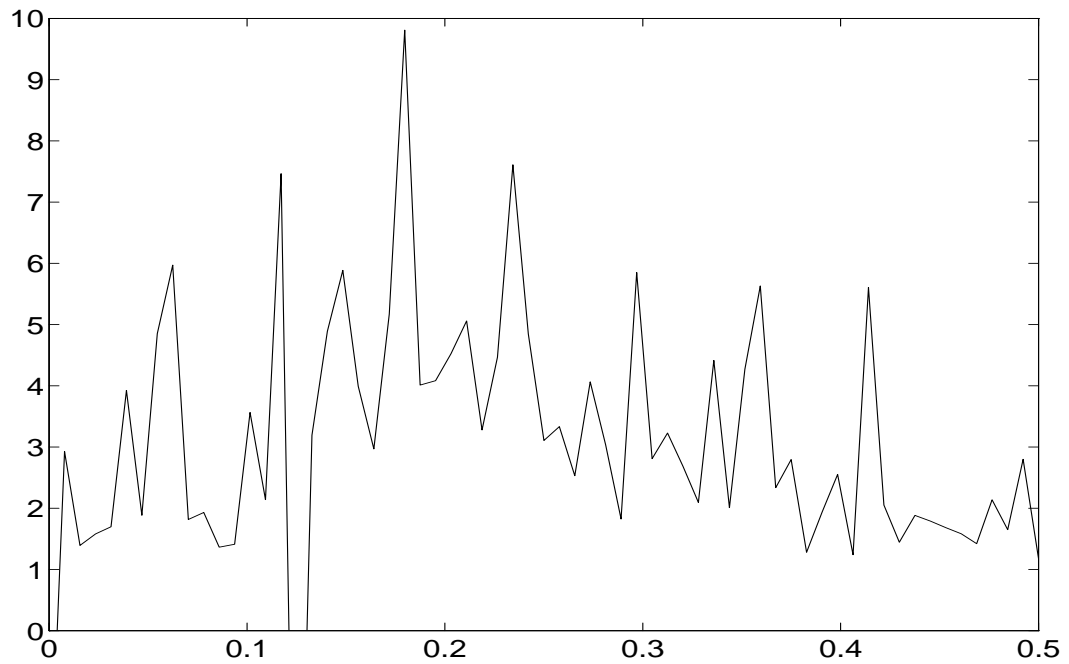


Figure 3.13: Gain en dB pour le traitement linéaire-quadratique en fréquence.

d'un certain nombre de modèles non linéaires, chacun étant adapté à un type de problème. Notons que ces modèles sont souvent de nature très différente, et que leur unification en une théorie générale semble difficile. Toutefois, un effort doit être fait pour unifier certaines classes de modèles. Par exemple, il existe un grand nombre de filtres non linéaires basés sur les statistiques d'ordre, et il est sensé de penser que tous ces filtres entrent dans une classe générale.

Soulignons à nouveau que les filtres de Volterra, en tant que systèmes non linéaires ici invariants, peuvent engendrer des signaux non gaussiens stationnaires, dont les outils de description ont été présentés au chapitre précédent. Ces signaux ne forment qu'une petite partie des signaux non gaussiens, les autres, plus généraux, étant les signaux non gaussiens non stationnaires. Ces derniers, au moins en partie, peuvent être générés à l'aide de filtres de Volterra rendus non invariants (les noyaux des filtres deviennent dépendant du temps).

L'appréhension des signaux non gaussiens non stationnaires ne peut plus se faire avec les objets développés précédemment. De nouveaux outils sont à introduire. Mais une fois encore, la non-stationnarité est une nonpropriété, et d'innombrables approches pour la décrire peuvent être envisagées. Nous nous concentrerons dans le chapitre suivant sur une méthode non paramétrique de description des signaux non gaussiens non stationnaires : les représentations temps-fréquence d'ordre supérieur.

Chapitre 4

Temps-fréquence d'ordre supérieur pour les signaux non stationnaires

Ce quatrième chapitre concerne la définition d'outils d'ordre supérieur pour l'étude de la troisième non-propriété, la non-stationnarité.

Un signal est non stationnaire s'il ... n'est pas stationnaire !

La description de tels signaux a provoqué un engouement à la fin des années 70 et surtout dans les années 80. De multiples approches pour définir des outils adéquats ont été développées ; dans un premier temps, nous ferons un rappel bref des représentations temps-fréquence bilinéaires. Nous ne nous concentrerons pas sur leurs nombreuses qualités, mais montrerons plutôt leurs limitations à travers un exemple.

Cet exemple illustre l'intérêt de définir des représentations d'ordre supérieur, dont l'idée initiale revient à N. Gerr en 1988 [35]. A partir de cette date, d'autres auteurs se sont intéressés à ces représentations, citons J. Fonollosa, A. Swami ou encore G. Giannakis [30, 60, 36]. Mais leurs définitions ont été données abruptement, et le but de ce chapitre est de montrer qu'une construction complètement déductive de ces objets est possible.

Les sections 4.2, 4.3 et 4.4 sont destinées à cette construction, pour des signaux déterministes aussi bien qu'aléatoires. Diverses questions sont ensuite abordées pour présenter finalement une application de ces représentations temps-fréquence d'ordre supérieur à la détection de signaux transitoires ou brefs. Ceci étant une autre histoire, commençons la promenade.

4.1 Rappels sur l'ordre deux

L'ordre deux s'intéresse aux propriétés énergétiques des signaux. Dans le cas stationnaire, la densité spectrale d'énergie 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 [61], 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 cette section, nous rappellerons quelques éléments concernant les représentations temps-fréquence bilinéaires, dont les extensions aux ordres supérieurs font l'objet de ce chapitre.

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 bilinéaires¹, le but est d'obtenir une répartition de l'énergie du signal dans un plan temps-fréquence. 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 [28, 29]).

- **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.
- **Energie** : en tant que répartition de l'énergie dans le plan temps-fréquence, la représentation doit permettre de retrouver l'énergie du signal. 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–).
- **Conservation des supports** : un événement à support temporel fini aura une représentation ayant le même support temporel (même propriété pour un événement 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 [28].

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 événements apparaissant longtemps avant ou après l'instant de référence t ne doivent pas avoir autant d'importance qu'un événement proche de t . Nous devons alors pondérer $x(t_1)x^*(t_2)$ par une fonction 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 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

$$R_{x,2,1}(t, \nu) = \int \int k(t, \nu; t_1, t_2) x(t_1) x^*(t_2) dt_1 dt_2 \quad (4.1)$$

Remarque 1 : nous pourrions envisager deux autres définitions pour lesquelles les deux termes $x(t_i)$ seraient conjugués ou non. Nous ne débattons pas ici de cette question, puisqu'elle sera abordée dans la section relative aux représentations d'ordre supérieur.

¹Précisons que la bilinéarité est relative au signal.

Remarque 2 : la notation $R_{x,2,1}$ signifie ordre 2 avec un complexe non-conjugué. Elle est l'analogue de la notation dans le cas stationnaire et sera reprise de manière générale pour les ordres supérieurs.

La forme (4.1) est la plus générale possible, et est donc trop compliquée. Les propriétés souhaitées indiquées précédemment se transforment en contraintes sur (4.1) 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 temporelle et fréquentielle.

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 :

$$R_{x(t-\tau),2,1}(t, \nu) = R_{x,2,1}(t - \tau, \nu) \quad (4.2)$$

- translation fréquentielle :

$$R_{x(t) \exp(2i\pi\nu_0 t),2,1}(t, \nu) = R_{x,2,1}(t, \nu - \nu_0) \quad (4.3)$$

Il peut alors être montré que $R_{x,2,1}(t, \nu)$ prend la forme [28, 29]

$$R_{x,2,1}(t, \nu) = \int \int F(t - \theta, \tau) x(\theta + \frac{\tau}{2}) x^*(\theta - \frac{\tau}{2}) \exp(-2i\pi\nu\tau) d\tau d\theta \quad (4.4)$$

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 (4.4) appartiennent à la classe dite de Cohen d'ordre 2. Un élément fameux de cette classe est la représentation ou transformée de Wigner-Ville $WV_{x,2,1}(t, \nu)$ qui est obtenue pour $F(t, \tau) = \delta(t)$, soit

$$WV_{x,2,1}(t, \nu) = \int x(t + \frac{\tau}{2}) x^*(t - \frac{\tau}{2}) \exp(-2i\pi\nu\tau) d\tau \quad (4.5)$$

Cette représentation est centrale dans la classe de Cohen, puisqu'elle permet de générer tous les membres de la classe par lissage temps-fréquence selon

$$R_{x,2,1}(t, \nu) = \Pi(t, \nu) *_t *_\nu WV_{x,2,1}(t, \nu) \quad (4.6)$$

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 lié à la fenêtre $h(t)$, soit plus explicitement

$$WV_{h,2,1}^*(t, \nu) *_t *_\nu WV_{x,2,1}(t, \nu) = \left| \int h^*(t - \tau) x(\tau) \exp(-2i\pi\nu\tau) d\tau \right|^2 \quad (4.7)$$

Remarque : 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), l'espérance mathématique de $WV_{x,2,1}(t, \nu)$ existe et définit le spectre de Wigner-Ville [47]. Ceci constitue l'extension au cas aléatoire des représentations temps-fréquence bilinéaires, extension qui sera discuté plus en détail à la section 4.4.

Nous avons rappelé dans cette section la philosophie des représentations temps-fréquence, en nous concentrant sur les représentations bilinéaires ; la section suivante avance quelques motivations pour généraliser cette approche aux ordres supérieurs.

4.2 Pourquoi étendre aux ordres supérieurs?

Les représentations temps-fréquence bilinéaires (donc d'ordre deux) sont les extensions au non-stationnaire des corrélations et spectres du stationnaire. Elles ne concernent donc que les propriétés d'ordre deux d'un signal. Examinons l'exemple suivant. Soit

$$x(t) = e^{-2i\pi(a_1+b_1t)t+i\phi_1} + e^{-2i\pi(a_2+b_2t)t+i\phi_2} + e^{-2i\pi(a_1+a_2+(b_1+b_2)t)t+i\phi_3} \quad (4.8)$$

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. Remarquons de suite que l'on doit calculer 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 $\exp(-2i\pi(a + 2bt)\tau)$. Le résultat vient après le calcul de la transformée de Fourier, soit

$$WV_{x,2,1}(t, \nu) = \delta(\nu - (a_1 + 2b_1t)) + \delta(\nu - (a_2 + 2b_2t)) + \delta(\nu - (a_1 + a_2 + 2(b_1 + b_2)t)) \quad (4.9)$$

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 Classes de Cohen d'ordre supérieur pour les signaux déterministes

L'objet de cette section est l'extension aux ordres supérieurs des représentations temps-fréquence bilinéaires. La construction de ces représentations se fera d'une manière déductive. Dans un premier temps, nous discuterons les propriétés souhaitables et les problèmes qui se posent. La construction sera ensuite effectuée, en distinguant les ordres pairs des ordres impairs, et nous définirons les classes de Cohen d'ordre supérieur. La section suivante sera dédiée au cas aléatoire.

4.3.1 Problèmes posés et propriétés souhaitées

Avant de présenter des propriétés que l'on souhaite voir vérifiées par des représentations d'ordre supérieur, nous abordons une question délicate, toujours débattue actuellement dans le cas stationnaire.

Moments ou cumulants

Une des propriétés listées dans le cas de l'ordre deux stipule l'indépendance vis-à-vis du temps d'une représentation bilinéaire dans le cas d'un signal stationnaire. Cette propriété peut être renforcée en demandant au

spectre de Wigner-Ville de se réduire au spectre traditionnel si le signal aléatoire étudié est stationnaire. Cette condition est remplie puisque $E[x(t + \frac{\tau}{2})x^*(t - \frac{\tau}{2})]$ se réduit à la fonction de corrélation dans le cas stationnaire. Donc, si $x(t)$ est stationnaire,

$$E[WV_{x,2,1}(t, \nu)] = S_{x,2,1}(\nu) \quad (4.10)$$

Remarquons de plus que ceci est valide car à l'ordre deux, moments et cumulants sont identiques dans le cas centré.

Ces remarques sur l'ordre deux nous conduisent naturellement à des réflexions concernant des représentations d'ordre supérieur, réflexion se concentrant en :

*dans le cas aléatoire, l'espérance mathématique d'une représentation temps-fréquence d'ordre supérieur construite pour des signaux déterministes devra se réduire au multispectre défini au chapitre 2 dans le cas d'un signal aléatoire stationnaire.*²

Or, nous allons voir que la forme générique d'une représentation d'ordre supérieur est une forme multilinéaire qui ne peut se réduire à un cumulants dans le cas aléatoire stationnaire. Le premier désir ne sera donc pas comblé par une approche déductive. La construction sera effectuée pour une représentation fondée sur les moments et non sur les cumulants. Toutefois, nous verrons comment pallier cette difficulté.

Quelques propriétés souhaitables

Nous proposons ici quelques propriétés dont certaines concerneront le cas aléatoire. Nous commençons la liste par le souhait discuté précédemment.

1. Dans le cas stationnaire, la représentation devra être indépendante du temps. De plus, pour des signaux aléatoires stationnaires, la représentation devra se réduire au multispectre stationnaire.
2. Tout comme les inter-multispectres sont nuls dans le cas de signaux indépendants, une représentation d'ordre supérieur devra éliminer les interférences entre composantes indépendantes d'un signal aléatoire non stationnaire.
3. Les marginales doivent permettre de retrouver les puissances instantanées et spectres d'ordre supérieur.
4. Les représentations seront invariantes par translations temporelles et fréquentielles.
5. Les supports des signaux seront conservés.
6. Le signal pourra être retrouvé à partir des représentations.
7. les représentations seront compatibles avec le filtrage linéaire et les modulations produits.

Ces propriétés constituent des vœux, et toutes ne seront pas forcément respectées. Nous reviendrons ultérieurement sur les conditions assurant ces propriétés. La partie suivante concerne la construction mathématique proprement dite.

4.3.2 Construction déductive des classes de Cohen d'ordre supérieur

Nous avons déjà mentionné que la forme générique des représentations temps-fréquence d'ordre supérieur est une forme multilinéaire. De plus, tout comme dans le cas stationnaire, le travail sur des signaux complexes impose de considérer des versions conjuguées et non-conjuguées des termes participant à la multilinéarité. Enfin, 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. Soit alors un signal $x(t)$ à valeurs complexes. Nous définissons la forme générale pour une représentation temps-fréquence d'ordre supérieur de $x(t)$ par

$$R_{x,p+q,p}(t, \nu) = \int k(t, \nu; \mathbf{t}) x(t_1) \dots x(t_p) x^*(t_{p+1}) \dots x^*(t_{p+q}) d\mathbf{t} \quad (4.11)$$

² Effectuer un moyennage d'ensemble de représentations pour des signaux aléatoires sera justifié dans la section suivante.

où $\boldsymbol{\nu} = (\nu_1 \dots \nu_{p+q-1})$ et $\mathbf{t} = (t_1 \dots t_{p+q})$. Notons que cette représentation réalise un filtrage de Volterra homogène à noyau non invariant. Cette forme est évidemment inutilisable, et tout comme à l'ordre deux, nous allons montrer maintenant comment la simplifier.

Invariances par translations temporelle et fréquentielle

La propriété souhaitable d'invariance par translation temporelle paraît très naturelle. Par contre, la deuxième, invariance par translation fréquentielle, mérite d'être examinée plus en détails.

Pour simplifier le débat, examinons cette condition pour des signaux aléatoires stationnaires. L'invariance par translation fréquentielle signifie que décaler en fréquence un signal translate également en fréquence ses statistiques. Or, nous avons étudié cette propriété au chapitre 2 pour les multispectres, et nous avons montré que le multispectre $S_{x,p+q,p}(\boldsymbol{\nu})$ d'un signal $x(t) = z(t) \exp(2i\pi\nu_0 t)$ est nul si $p \neq q$. Autrement dit, les seuls multispectres non nuls de $x(t) = z(t) \exp(2i\pi\nu_0 t)$ sont les multispectres d'ordre pair comportant un nombre égal de conjugués et de non-conjugués. Nous dénommerons cette situation par ordre pair symétrique. Dans ce cas, le multispectre de $x(t)$ est égal à celui de $z(t)$ décalé de ν_0 sur toutes les fréquences. Physiquement, ceci signifie que le multispectre se déplace le long de la multiplicité stationnaire lorsque l'ordre est pair symétrique. Par contre, pour un ordre impair ou pair non symétrique, le multispectre sort de la multiplicité stationnaire et s'annule.

Ainsi, la propriété d'invariance par translation fréquentielle apparaît naturelle pour un ordre pair symétrique, mais s'avère sans consistance pour un autre cas. Nous devons donc couper la construction des représentations en deux : ordre pair symétrique et ordre impair ou pair non symétrique.

Ordre pair symétrique

Le problème soumis est donc de simplifier

$$R_{x,2p,p}(t, \boldsymbol{\nu}) = \int k(t, \boldsymbol{\nu}; \mathbf{t}) x(t_1) \dots x(t_p) x^*(t_{p+1}) \dots x^*(t_{2p}) d\mathbf{t} \quad (4.12)$$

où maintenant $\boldsymbol{\nu} = (\nu_1 \dots \nu_{2p-1})$ et $\mathbf{t} = (t_1 \dots t_{2p})$, sachant que l'on souhaite

1. Invariance par translation temporelle :

$$R_{x(t-\tau),2p,p}(t, \boldsymbol{\nu}) = R_{x,2p,p}(t - \tau, \boldsymbol{\nu}) \quad \forall \tau \quad (4.13)$$

2. Invariance par translation fréquentielle :

$$R_{x(t) \exp(2i\pi\nu_0 t),2p,p}(t, \boldsymbol{\nu}) = R_{x,2p,p}(t, \boldsymbol{\nu} - \nu_0 \mathbf{1}) \quad \forall \nu_0 \quad (4.14)$$

Ces deux dernières contraintes vont se traduire par des contraintes sur le noyau $k(t, \boldsymbol{\nu}; \mathbf{t})$. En effet,

$$R_{x(t-\tau),2p,p}(t, \boldsymbol{\nu}) = \int k(t, \boldsymbol{\nu}; \mathbf{t}) x(t_1 - \tau) \dots x(t_p - \tau) x^*(t_{p+1} - \tau) \dots x^*(t_{2p} - \tau) d\mathbf{t} \quad (4.15)$$

soit, après un changement de variable évident

$$R_{x(t-\tau),2p,p}(t, \boldsymbol{\nu}) = \int k(t, \boldsymbol{\nu}; \mathbf{t} + \tau \mathbf{1}) x(t_1) \dots x(t_p) x^*(t_{p+1}) \dots x^*(t_{2p}) d\mathbf{t} \quad (4.16)$$

Or,

$$R_{x,2p,p}(t - \tau, \boldsymbol{\nu}) = \int k(t - \tau, \boldsymbol{\nu}; \mathbf{t}) x(t_1) \dots x(t_p) x^*(t_{p+1}) \dots x^*(t_{2p}) d\mathbf{t} \quad (4.17)$$

donc, en égalant selon la contrainte 1 les équations (4.16) et (4.17), le noyau doit vérifier

$$k(t - \tau, \boldsymbol{\nu}; \mathbf{t}) = k(t, \boldsymbol{\nu}; \mathbf{t} + \tau \mathbf{1}) \quad (4.18)$$

Examinons maintenant l'effet de la contrainte 2 sur le noyau. Il vient

$$R_{x(t) \exp(2i\pi\nu_0 t), 2p, p}(t, \boldsymbol{\nu}) = \int k(t, \boldsymbol{\nu}; \mathbf{t}) x(t_1) e^{2i\pi\nu_0 t_1} \dots x(t_p) e^{2i\pi\nu_0 t_p} x^*(t_{p+1}) e^{-2i\pi\nu_0 t_{p+1}} \dots x^*(t_{2p}) e^{-2i\pi\nu_0 t_{2p}} d\mathbf{t} \quad (4.19)$$

qui doit être égal à

$$R_{x, 2p, p}(t, \boldsymbol{\nu} - \nu_0 \mathbf{1}) = \int k(t, \boldsymbol{\nu} - \nu_0 \mathbf{1}; \mathbf{t}) x(t_1) \dots x(t_p) x^*(t_{p+1}) \dots x^*(t_{2p}) d\mathbf{t} \quad (4.20)$$

Par suite, la contrainte 2 est suivie si le noyau vérifie

$$k(t, \boldsymbol{\nu} - \nu_0 \mathbf{1}; \mathbf{t}) = k(t, \boldsymbol{\nu}; \mathbf{t}) \exp(2i\pi\nu_0 (\sum_{i=1}^p t_i - \sum_{i=p+1}^{2p} t_i)) \quad (4.21)$$

Le problème est donc de simplifier (4.12) sachant (4.18) et (4.21) :

1. (4.18) qui se réécrit

$$k(t - \tau, \boldsymbol{\nu}; \mathbf{t} - \tau \mathbf{1}) = k(t, \boldsymbol{\nu}; \mathbf{t}) \quad (4.22)$$

signifie que la dimension temporelle de la fonction k n'est pas $2p + 1$ mais seulement $2p$. Nous allons donc chercher un changement de variables qui permet de réaliser explicitement cette diminution de dimension.

2. (4.21) qui devient

$$k(t, \boldsymbol{\nu} - \nu_0 \mathbf{1}; \mathbf{t}) \exp(-2i\pi\nu_0 (\sum_{i=1}^p t_i - \sum_{i=p+1}^{2p} t_i)) = k(t, \boldsymbol{\nu}; \mathbf{t}) \quad (4.23)$$

fait naturellement intervenir des exponentielles imaginaires. En utilisant la diminution de dimension temporelle, nous allons montrer que l'invariance par translation fréquentielle rend le noyau dépendant des fréquences uniquement à travers les exponentielles.

Les propriétés d'invariance sont donc équivalentes à des contraintes sur le noyau, exprimées par (4.22) et (4.23). Nous allons maintenant détailler ces contraintes pour obtenir une expression générale du noyau.

- Contrainte 1 ou invariance par translation temporelle :

L'équation fonctionnelle (4.22) montre que le noyau $k(t, \boldsymbol{\nu}; \mathbf{t})$ qui dépend a priori de $2p + 1$ variables temporelles n'est en fait que $2p$ dimensionnel en temps. En effet, considérons le changement de variable suivant

$$\left\{ \begin{array}{l} w = t \\ \alpha_1 = t - \frac{1}{2p} \sum_{i=1}^{2p} t_i \\ \alpha_2 = t_1 - t_{2p} \\ \vdots \\ \alpha_i = t_{i-1} - t_{2p} \\ \vdots \\ \alpha_{2p} = t_{2p-1} - t_{2p} \end{array} \right. \quad (4.24)$$

et évaluons l'opérateur de dérivée partielle $\frac{\partial}{\partial w}$. Il vient

$$\frac{\partial}{\partial w} = \frac{\partial}{\partial t} \frac{\partial t}{\partial w} + \frac{\partial}{\partial t_1} \frac{\partial t_1}{\partial w} + \dots + \frac{\partial}{\partial t_{2p}} \frac{\partial t_{2p}}{\partial w} \quad (4.25)$$

Or, l'inversion du système (4.24) se fait relativement facilement. En effet, en sommant les $2p - 1$ dernières équations de (4.24) et $2p$ fois la deuxième, nous obtenons

$$2p\alpha_1 + \sum_{i=2}^{2p} \alpha_i = 2pw - \sum_{i=1}^{2p} t_i + \sum_{i=1}^{2p-1} (t_i - t_{2p}) \quad (4.26)$$

qui implique

$$t_{2p} = w - \alpha_1 - \frac{1}{2p} \sum_{i=2}^{2p} \alpha_i \quad (4.27)$$

et

$$t_i = w - \alpha_1 + \alpha_{i+1} - \frac{1}{2p} \sum_{i=2}^{2p} \alpha_i \quad \forall i = 1, \dots, 2p - 1 \quad (4.28)$$

Ces deux dernières équations montrent alors que

$$\frac{\partial t_i}{\partial w} = 1 \quad \forall i = 1, \dots, 2p \quad (4.29)$$

Finalement, l'opérateur $\frac{\partial}{\partial w}$ s'écrit

$$\frac{\partial}{\partial w} = \frac{\partial}{\partial t} + \sum_i \frac{\partial}{\partial t_i} \quad (4.30)$$

Réexaminons la contrainte 1. Elle stipule que pour tout décalage τ , $k(t - \tau, \boldsymbol{\nu}; \mathbf{t} - \tau \mathbf{1}) = k(t, \boldsymbol{\nu}; \mathbf{t})$. Or, si cela est valable pour tout τ , c'est aussi vrai pour des valeurs infinitésimales de τ . Donc, écrivant la différentielle du noyau selon les variables temporelles des deux façons

$$dk = k(t - d\tau, \boldsymbol{\nu}; \mathbf{t} - d\boldsymbol{\tau}) - k(t, \boldsymbol{\nu}; \mathbf{t}) \quad (4.31)$$

$$= \frac{\partial k}{\partial t} d\tau + \sum_i \frac{\partial k}{\partial t_i} d\tau_i \quad (4.32)$$

nous voyons que la différentielle est nulle le long de la diagonale (i.e. pour $d\tau = d\tau_i \quad \forall i$), et nous obtenons finalement

$$\frac{\partial k}{\partial w} = 0 \quad (4.33)$$

Ainsi, k ne dépend plus que de $2p$ variables temporelles, et il existe une fonction l telle que

$$k(t, \boldsymbol{\nu}; \mathbf{t}) = l\left(t - \frac{1}{2p} \sum_{i=1}^{2p} t_i, \boldsymbol{\nu}; \mathbf{t} - t_{2p} \mathbf{1}\right) \quad (4.34)$$

où maintenant $\mathbf{t} = (t_1, \dots, t_{2p-1})$.

Le choix du changement de variables n'est pas arbitraire. Les instants t_i sont disposés autour de la date t pour respecter la propriété 1 de stationnarité. En effet, les instants interviennent par leur différence avec t_{2p} pris comme référence. De plus, ils sont choisis pour que t soit leur barycentre, ce qui correspond à une idée intuitive de localisation.

- Contrainte 2 ou invariance par translation fréquentielle :

Nous avons vu que cette contrainte se traduit par

$$k(t, \boldsymbol{\nu} - \nu_0 \mathbf{1}; \mathbf{t}) \exp(-2i\pi\nu_0(\sum_{i=1}^p t_i - \sum_{i=p+1}^{2p} t_i)) = k(t, \boldsymbol{\nu}; \mathbf{t}) \quad (4.35)$$

En reprenant les notations développées précédemment, on montre que

$$\sum_{i=1}^p t_i - \sum_{i=p+1}^{2p} t_i = \sum_{i=1}^p (\alpha_{i+1} + t_{2p}) - \sum_{i=p+1}^{2p-1} (\alpha_{i+1} + t_{2p}) - t_{2p} \quad (4.36)$$

soit

$$\sum_{i=1}^p t_i - \sum_{i=p+1}^{2p} t_i = \sum_{i=1}^p \alpha_{i+1} - \sum_{i=p+1}^{2p-1} \alpha_{i+1} \quad (4.37)$$

Donc, utilisant les résultats concernant la première contrainte, (4.35) se réécrit

$$l(\alpha_1, \boldsymbol{\nu} - \nu_0 \mathbf{1}; \boldsymbol{\alpha}) \exp(-2i\pi\nu_0(\sum_{i=1}^p \alpha_{i+1} - \sum_{i=p+1}^{2p-1} \alpha_{i+1})) = l(\alpha_1, \boldsymbol{\nu}; \boldsymbol{\alpha}) \quad (4.38)$$

où le vecteur $\boldsymbol{\alpha}$ contient les α_i pour $i = 2, \dots, 2p$. Définissons maintenant les vecteurs $\boldsymbol{\alpha}' = (\alpha_3, \dots, \alpha_{2p})$ et $\boldsymbol{\nu}' = (\nu_2, \dots, \alpha_{2p-1})$. Effectuons une transformée de Fourier (à noyau conjugué) par rapport à α_2 vers la variable f . On a alors

$$L(\alpha_1, \nu_1, \boldsymbol{\nu}'; f, \boldsymbol{\alpha}') = L(\alpha_1, \nu_1 - \nu_0, \boldsymbol{\nu}' - \nu_0 \mathbf{1}; f - \nu_0, \boldsymbol{\alpha}') \exp(-2i\pi\nu_0(\sum_{i=2}^p \alpha_{i+1} - \sum_{i=p+1}^{2p-1} \alpha_{i+1})) \quad (4.39)$$

Tout comme pour la première contrainte, ceci montre que les deux variables f et ν_1 sont liées, et que l'on peut réduire la dimension de L de 1 par un changement de variables évident. Par suite, il existe une fonction M telle que (4.39) soit équivalente à

$$M(\alpha_1, \boldsymbol{\nu}'; f - \nu_1, \boldsymbol{\alpha}') \exp(-2i\pi\nu_0(\sum_{i=2}^p \alpha_{i+1} - \sum_{i=p+1}^{2p-1} \alpha_{i+1})) = L(\alpha_1, \nu_1, \boldsymbol{\nu}'; f, \boldsymbol{\alpha}') \quad (4.40)$$

Par transformée de Fourier de f vers α_2 , il vient

$$l(\alpha_1, \boldsymbol{\nu}; \boldsymbol{\alpha}) = m(\alpha_1, \boldsymbol{\nu}'; \boldsymbol{\alpha}) \exp(-2i\pi\nu_1\alpha_2) \exp(-2i\pi\nu_0(\sum_{i=2}^p \alpha_{i+1} - \sum_{i=p+1}^{2p-1} \alpha_{i+1})) \quad (4.41)$$

On réalise cette opération pour les $2p - 2$ fréquences restantes pour conclure qu'il existe une fonction n telle que

$$l(\alpha_1, \boldsymbol{\nu}; \boldsymbol{\alpha}) = n(\alpha_1; \boldsymbol{\alpha}) \exp(-2i\pi(\sum_{i=1}^p \nu_i \alpha_{i+1} - \sum_{i=p+1}^{2p-1} \nu_i \alpha_{i+1})) \quad (4.42)$$

Ainsi, le noyau ne dépend des variables fréquentielles qu'à travers les exponentielles imaginaires. Ceci est en accord avec l'idée intuitive de la forme d'une représentation temps-fréquence : analyse spectrale au cours du temps. Il est donc logique qu'une transformation de Fourier intervienne.

Finalement, les seules propriétés d'invariance par translation temporelle et fréquentielle contraignent le noyau à s'écrire

$$k(t, \boldsymbol{\nu}; \mathbf{t}) = F(t - \frac{1}{2p} \sum_{i=1}^{2p} t_i, \mathbf{t} - t_{2p} \mathbf{1}) \exp(-2i\pi(\sum_{i=1}^p \nu_i(t_i - t_{2p}) - \sum_{i=p+1}^{2p-1} \nu_i(t_i - t_{2p}))) \quad (4.43)$$

Par suite, de sorte à disposer les temps de façon à satisfaire la propriété souhaitable 1 (stationnarité, voir page 62), nous effectuons le changement de variables suivant

$$\left\{ \begin{array}{l} \tau_1 = t_1 - t_{2p} \\ \tau_2 = t_2 - t_{2p} \\ \vdots \\ \tau_p = t_p - t_{2p} \\ -\tau_{p+1} = t_{p+1} - t_{2p} \\ \vdots \\ -\tau_{2p-1} = t_{2p-1} - t_{2p} \\ \theta = \frac{1}{2p} \sum_{i=1}^{2p} t_i \end{array} \right. \quad (4.44)$$

Le Jacobien de cette transformation est unité. De plus, l'inversion de ce système linéaire conduit à

$$\left\{ \begin{array}{l} t_{2p} = \theta - \frac{1}{2p} \left(\sum_{i=1}^p \tau_i - \sum_{i=p+1}^{2p-1} \tau_i \right) \\ t_j = \tau_j + t_{2p} = \theta + \frac{2p-1}{2p} \tau_j - \frac{1}{2p} \left(\sum_{\substack{i=1 \\ i \neq j}}^p \tau_i - \sum_{i=p+1}^{2p-1} \tau_i \right) \quad \forall j = 1, \dots, p \\ t_j = -\tau_j + t_{2p} = \theta - \frac{2p-1}{2p} \tau_j - \frac{1}{2p} \left(\sum_{i=1}^p \tau_i - \sum_{\substack{i=p+1 \\ i \neq j}}^{2p-1} \tau_i \right) \quad \forall j = p+1, \dots, 2p-1 \end{array} \right. \quad (4.45)$$

et l'on peut écrire la version finale de la représentation temps-fréquence d'ordre $2p$

$$R_{x,2p,p}(t, \boldsymbol{\nu}) = \int F(t - \theta, \boldsymbol{\tau}) x(\theta + f_1(\boldsymbol{\tau})) \dots x(\theta + f_p(\boldsymbol{\tau})) x^*(\theta + f_{p+1}(\boldsymbol{\tau})) \dots x^*(\theta + f_{2p}(\boldsymbol{\tau})) \exp(-2i\pi \left(\sum_{i=1}^{2p-1} \nu_i \tau_i \right)) d\boldsymbol{\tau} d\theta \quad (4.46)$$

où les $f_i(\boldsymbol{\tau})$ apparaissent dans le système (4.45). Examinons de plus près cette expression. La partie multilinéaire $x(t_1) \dots x^*(t_{2p})$, où les t_i vérifient (4.45), apparaît comme une multicorrélation relative à l'instant t . Nous l'appellerons multicorrélation locale. Par extension du cas stationnaire, nous obtenons alors une représentation spectrale en prenant la transformée de Fourier de cette multilinéarité.

Mais les retards τ_i varient a priori de $-\infty$ à $+\infty$. Or, il est sensé de penser que les événements très lointains de t n'influent pas sur le comportement du signal autour de t . Ainsi, avant d'effectuer la transformation de Fourier, une apodisation ($F(\cdot, \boldsymbol{\tau})$) est appliquée pour "oublier" plus ou moins les valeurs lointaines de t . Enfin, le lissage ($F(t - \theta, \cdot)$) apparaît comme l'opération duale : il s'agit d'une apodisation en fréquence.

Remarque 1 : le résultat englobe évidemment le cas de l'ordre deux.

Remarque 2 : dans la suite, les arguments $\theta + f_i(\boldsymbol{\tau})$ seront remplacés par les t_i qui vérifient (4.45).

Passons maintenant à une discussion sur les ordres impairs.

Ordre pair non symétrique ou impair

Nous avons mentionné précédemment que la propriété d'invariance par translation fréquentielle n'a pas lieu d'être pour les ordres pairs non symétriques ou impairs. Nous allons présenter ce fait un peu différemment en nous appuyant sur l'exemple de l'ordre 3 et du bispectre. Soit $x(t)$ un signal à valeurs réelles ; considérons $S_{x,3}(\nu_1, \nu_2) = \text{Cum}[X(\nu_1), X(\nu_2), X^*(\nu_1 + \nu_2)]$. Le bispectre constitue donc un quantificateur des interactions statistiques entre les trois fréquences ν_1 , ν_2 et $\nu_1 + \nu_2$. Supposons que le signal $x(t)$ étudié soit non nul pour trois de ces fréquences. Examinons maintenant le bispectre de $x(t) \exp(2i\pi\nu_0 t)$. Les trois fréquences précédentes deviennent donc $\nu_1 - \nu_0$, $\nu_2 - \nu_0$ et $\nu_1 + \nu_2 - \nu_0$. Alors $\text{Cum}[X(\nu_1 - \nu_0), X(\nu_2 - \nu_0), X^*(\nu_1 + \nu_2 - \nu_0)] \neq S_{x \exp(2i\pi\nu_0 t), 3}(\nu_1 - \nu_0, \nu_2 - \nu_0)$ car la troisième fréquence n'est pas sur la multiplicité stationnaire. En fait le signal étudié n'est plus stationnaire, il est devenu cyclostationnaire (ses statistiques sont périodiques). Par contre, envisageons l'inter-bispectre $\text{Cum}[X(\nu_1), Y(\nu_2), Z^*(\nu_1 + \nu_2)]$. Par les mêmes arguments, si $x(t)$ et $z(t)$ (ou $y(t)$ et $z(t)$) sont modulés par une fréquence pure, alors la troisième fréquence reste sur la multiplicité stationnaire, et l'inter-bispectre des signaux modulés se décale selon une des deux fréquences de la quantité ν_0 . Ainsi, dans le cas impair, la propriété d'invariance par translation fréquentielle est valide dans une certaine mesure, et nous pouvons construire des inter-représentations temps-fréquence d'ordre supérieur.

Toutefois, nous n'effectuerons pas à nouveau la construction, ceci n'apportant rien. Il faut pourtant mentionner les difficultés supplémentaires liées essentiellement au fait que la construction doit s'effectuer sur des inter-représentations. Le principal problème provient des complexes conjugués à disposer dans la forme générale. Toute configuration est a priori envisageable³, bien que quelques simplifications apparaissent par

³Ceci est également valable dans le cas cas de construction d'inter-représentations d'ordre pair non symétrique.

raisons de symétrie ou en se restreignant aux signaux analytiques. Quoiqu'il en soit, pour un ordre impair, la forme générale d'une inter-représentation est

$$R_{\mathbf{x}, p+q, p}(t, \boldsymbol{\nu}) = \int k(t, \boldsymbol{\nu}; \mathbf{t}) x_1(t_1) \dots x_p(t_p) x_{p+1}^*(t_{p+1}) \dots x_{p+q}^*(t_{p+q}) d\mathbf{t} \quad (4.47)$$

où $p + q$ est impair. Pour des détails sur l'ordre 3 (qui sera utilisé par la suite) nous renvoyons à [2].

4.3.3 Rôle particulier de la représentation de Wigner-Ville et exemples

Nous avons vu jusqu'à présent la définition des classes de Cohen d'ordre supérieur. Bien que la forme générale d'un membre de ces classes soit considérablement plus simple que la forme générale présentée au début, elle n'en reste pas moins compliquée. L'objet de cette section est de montrer deux exemples simples qui expliciteront, nous l'espérons, l'approche adoptée ici.

Transformée de Wigner-Ville d'ordre $2p$: la mère !

Rappelons l'expression générale d'un membre de la classe de Cohen d'ordre $2p$.

$$R_{x, 2p, p}(t, \boldsymbol{\nu}) = \int F(t - \theta, \boldsymbol{\tau}) x(t_1) \dots x(t_p) x^*(t_{p+1}) \dots x^*(t_{2p}) \exp(-2i\pi(\sum_{i=1}^{2p-1} \nu_i \tau_i)) d\boldsymbol{\tau} d\theta \quad (4.48)$$

où le système (4.45) est vérifié. Cette expression peut-être interprétée comme la transformée de Fourier d'une version lissée du noyau multilinéaire $x(t_1) \dots x(t_p) x^*(t_{p+1}) \dots x^*(t_{2p})$. Or, ce noyau apparaît comme une multicorrélation locale (fondée sur les moments) du signal $x(t)$. Le lissage a un effet de moyennage, et la façon la plus simple de moyennage est de ... ne pas moyennage ! Ainsi, en remplaçant $F(t, \boldsymbol{\tau})$ par $\delta(t)$, nous obtenons l'expression

$$WV_{x, 2p, p}(t, \boldsymbol{\nu}) = \int x(t_1) \dots x(t_p) x^*(t_{p+1}) \dots x^*(t_{2p}) \exp(-2i\pi(\sum_{i=1}^{2p-1} \nu_i \tau_i)) d\boldsymbol{\tau} \quad (4.49)$$

où le système (4.45) est vérifié en changeant θ en t . Cette expression est la définition de la transformée de Wigner-Ville d'ordre $2p$.⁴

L'interprétation que nous venons de donner et la définition de la transformée de Wigner-Ville indiquent que cette transformée est centrale dans la classe de Cohen correspondante. En effet, considérons $\Pi(t, \boldsymbol{\nu})$ la transformée de Fourier de $F(t, \boldsymbol{\tau})$. Il vient

$$WV_{x, 2p, p}(t, \boldsymbol{\nu}) * \Pi(t, \boldsymbol{\nu}) = \int \int \int \Pi(t - \theta, \boldsymbol{\nu} - \boldsymbol{\nu}') x(t_1) \dots x(t_p) x^*(t_{p+1}) \dots x^*(t_{2p}) \exp(-2i\pi(\sum_{i=1}^{2p-1} \nu'_i \tau_i)) d\boldsymbol{\nu}' d\theta d\boldsymbol{\tau} \quad (4.50)$$

En effectuant le changement de variables $\boldsymbol{\nu}'' = \boldsymbol{\nu} - \boldsymbol{\nu}'$, puis en réalisant la transformation de Fourier $\boldsymbol{\nu}'' \longleftrightarrow \boldsymbol{\tau}$, nous obtenons

$$WV_{x, 2p, p}(t, \boldsymbol{\nu}) * \Pi(t, \boldsymbol{\nu}) = \int F(t - \theta, \boldsymbol{\tau}) x(t_1) \dots x(t_p) x^*(t_{p+1}) \dots x^*(t_{2p}) \exp(-2i\pi(\sum_{i=1}^{2p-1} \nu_i \tau_i)) d\boldsymbol{\tau} d\theta \quad (4.51)$$

où les t_i vérifient (4.45). Ainsi, tout élément de la classe de Cohen d'ordre $2p$ s'obtient par lissage multidimensionnel de la transformée de Wigner-Ville d'ordre $2p$. Dans la section suivante, un exemple clarifie cette idée.

⁴ Pour les ordres impairs, la définition est obtenue identiquement.

Analyse de Fourier à court terme : le multispectrogramme

Les premières idées pour réaliser des analyses temps-fréquences étaient de définir des transformées de Fourier à court-terme. Ce genre de transformée s'écrit

$$ST_{x,h}(t, \nu) = \int h^*(t - \tau)x(\tau)\exp(-2i\pi\nu\tau)d\tau \quad (4.52)$$

où h est une fenêtre quelconque. Pour obtenir un spectre au cours du temps, il suffit de prendre le module carré de cette expression. La quantité obtenue est le très célèbre spectrogramme

$$SP_{x,h}(t, \nu) = \left| \int h^*(t - \tau)x(\tau)\exp(-2i\pi\nu\tau)d\tau \right|^2 \quad (4.53)$$

Nous pouvons généraliser cette approche en définissant un multispectrogramme. Cet outil est la première extension naturelle au non stationnaire des outils développés dans le chapitre 2. Il sera obtenu en multipliant des transformées de Fourier à court-terme prises à la date t , mais évaluées à des fréquences qui devront appartenir à la multiplicité stationnaire, pour rester en accord avec les théories stationnaires. De plus, il peut être généré par lissage de la transformée de Wigner-Ville par la transformée de Wigner-Ville de la fenêtre associée. En effet, évaluons $M_{x,h,2p,p}(t, \nu) = WV_{x,2p,p}(t, \nu) ** WV_{h,2p,p}^*(t, \nu)$. Il vient

$$\begin{aligned} M_{x,h,2p,p}(t, \nu) &= \int x(u_1) \dots x(u_p)x^*(u_{p+1}) \dots x^*(u_{2p})h^*(u'_1) \dots h^*(u'_p)h(u'_{p+1}) \dots h(u'_{2p}) \\ &\quad \exp(-2i\pi(\sum_{i=1}^{2p-1} n_i \tau_i)) \exp(-2i\pi(\sum_{i=1}^{2p-1} (\nu_i - n_i)\tau'_i)) d\tau d\tau' du d\mathbf{n} \end{aligned} \quad (4.54)$$

où les u_i vérifient le système (4.45) en y remplaçant θ par u , et où les u_i vérifient aussi (4.45) en y remplaçant t par $t - u$ et les τ_i par les τ'_i . En effectuant l'intégration sur \mathbf{n} (qui a pour effet de "remplacer" les τ'_i par les $-\tau_i$ et permet de tirer partie du caractère impair des fonctions $f_i(\tau)$ apparaissant en (4.46)), puis en effectuant le changement de variable

$$\begin{cases} u_i &= u_i \quad \forall i \\ u'_i &= t - u_i \quad \forall i \end{cases} \quad (4.55)$$

l'expression (4.54) se réécrit

$$\begin{aligned} M_{x,h,2p,p}(t, \nu) &= \int x(u_1) \dots x(u_p)x^*(u_{p+1}) \dots x^*(u_{2p})h^*(t - u_1) \dots h^*(t - u_p)h(t - u_{p+1}) \dots h(t - u_{2p}) \\ &\quad \exp(-2i\pi(\sum_{i=1}^p \nu_i u_i)) \exp(2i\pi(\sum_{i=p+1}^{2p-1} \nu_i u_i)) \exp(2i\pi(\sum_{i=1}^p \nu_i - \sum_{i=p+1}^{2p-1} \nu_i)u_{2p}) du \end{aligned} \quad (4.56)$$

soit encore⁵

$$M_{x,h,2p,p}(t, \nu) = ST_{x,h}^*(t, \sum_{i=1}^p \nu_i - \sum_{i=p+1}^{2p-1} \nu_i) \prod_{i=1}^p ST_{x,h}(t, \nu_i) \prod_{i=p+1}^{2p-1} ST_{x,h}^*(t, \nu_i) \quad (4.57)$$

Remarque : dans cette définition, la multiplicité stationnaire intervient dans un terme conjugué. Ceci provient du fait que nous avons privilégié dans la construction un instant argument d'un terme conjugué.

Rappelons le résultat en conclusion de ce paragraphe

$$M_{x,h,2p,p}(t, \nu) = WV_{x,2p,p}(t, \nu) ** WV_{h,2p,p}^*(t, \nu) \quad (4.58)$$

qui montre que le multispectrogramme est un élément de la classe de Cohen d'ordre $2p$.

⁵ Une définition analogue est obtenue pour des ordres impairs. Par exemple, pour $x(t)$ réel, le bispectrogramme s'écrit $B_{x,h,3}(t, \nu_1, \nu_2) = ST_{x,h}(t, \nu_1)ST_{x,h}(t, \nu_2)ST_{x,h}^*(t, \nu_1 + \nu_2)$ qui est égal à $WV_{x,3}(t, \nu_1, \nu_2) *** WV_{h,3}^*(t, \nu_1, \nu_2)$.

4.3.4 Propriétés souhaitées et contraintes sur le noyau

Pour illustrer d'autres caractéristiques de la classe de Cohen d'ordre $2p$, nous revenons maintenant aux propriétés souhaitées au paragraphe 4.3.1. Nous avons déjà mentionné le fait que ces propriétés sont équivalentes à des contraintes sur le noyau. Nous montrons comme exemple les contraintes apportées au noyau par les points 3, 5, 6 et 7.

Marginales

- La "puissance instantanée d'ordre $2p$ " est la marginale temporelle d'un élément de la classe de Cohen d'ordre $2p$ si et seulement si $F(t, \mathbf{0}) = \delta(t)$.

En effet

$$\int R_{x,2p,p}(t, \boldsymbol{\nu}) d\boldsymbol{\nu} = \int F(t - \theta, \mathbf{0}) |x(\theta)|^{2p} d\theta \quad (4.59)$$

après avoir intégré en $\boldsymbol{\nu}$ et $\boldsymbol{\tau}$. Or, la propriété voulue impose

$$\int F(t - \theta, \mathbf{0}) |x(\theta)|^{2p} d\theta = |x(t)|^{2p} \quad (4.60)$$

d'où le résultat.

De même, on montre

- La "densité d'énergie d'ordre $2p$ " est la marginale fréquentielle d'un élément de la classe de Cohen d'ordre $2p$ si et seulement si $F(0, \boldsymbol{\tau}) = \delta(\boldsymbol{\tau})$.

Conservation du support temporel

Cette propriété s'énonce comme suit :

$$(x(t) = 0 \forall t > |T|) \implies (R_{x,2p,p}(t, \boldsymbol{\nu}) = 0 \forall t > |T|) \quad (4.61)$$

Cette propriété implique la contrainte

$$F(t, \boldsymbol{\tau}) = 0 \forall |\tau_i| < 2|t| \quad (4.62)$$

En effet, supposons $t < -T$. Alors (4.61) implique que : $\forall \theta$ tel que

$$\left\{ \begin{array}{l} -T < t - \theta - \frac{1}{2p} \left(\sum_{i=1}^p \tau_i - \sum_{i=p+1}^{2p-1} \tau_i \right) < T \\ -T < t - \theta + \frac{2p-1}{2p} \tau_j - \frac{1}{2p} \left(\sum_{\substack{i=1 \\ i \neq j}}^p \tau_i - \sum_{i=p+1}^{2p-1} \tau_i \right) < T \quad \forall j = 1, \dots, p \\ -T < t - \theta - \frac{2p-1}{2p} \tau_j - \frac{1}{2p} \left(\sum_{i=1}^p \tau_i - \sum_{\substack{i=p+1 \\ i \neq j}}^{2p-1} \tau_i \right) < T \quad \forall j = p+1, \dots, 2p-1 \end{array} \right. \quad (4.63)$$

$F(\theta, \boldsymbol{\tau})$ doit être nul (puisque t est tel que $x(t) = 0$).

Or, comme $t < -T$, le système (4.63) entraîne le suivant

$$\left\{ \begin{array}{l} t < t - \theta - \frac{1}{2p} \left(\sum_{i=1}^p \tau_i - \sum_{i=p+1}^{2p-1} \tau_i \right) < -t \\ t < t - \theta + \frac{2p-1}{2p} \tau_j - \frac{1}{2p} \left(\sum_{\substack{i=1 \\ i \neq j}}^p \tau_i - \sum_{i=p+1}^{2p-1} \tau_i \right) < -t \quad \forall j = 1, \dots, p \\ t < t - \theta - \frac{2p-1}{2p} \tau_j - \frac{1}{2p} \left(\sum_{i=1}^p \tau_i - \sum_{\substack{i=p+1 \\ i \neq j}}^{2p-1} \tau_i \right) < -t \quad \forall j = p+1, \dots, 2p-1 \end{array} \right. \quad (4.64)$$

qui implique encore, en effectuant la différence entre la ligne i et la première,

$$2t < \tau_i < -2t \quad \forall i = 1, \dots, 2p-1 \quad (4.65)$$

Donc, pour résumer, si $t < -T$, vérifier la propriété implique $F(\theta, \boldsymbol{\tau}) = 0$ si $2t < \tau_i < -2t \forall i = 1, \dots, 2p-1$. On montre la même chose si $t > T$ (i.e. $-2t < \tau_i < 2t$), et le résultat est acquis.

Remarque : cette condition n'est que suffisante.

Reconstruction du signal

Le signal peut être retrouvé à partir d'un membre de la classe de Cohen d'ordre $2p$ si la transformée de Fourier par rapport à $\boldsymbol{\tau}$ de $F(t, \boldsymbol{\tau})$ ne comporte pas de zéro.

Commençons par retrouver le signal à partir de la transformée de Wigner-Ville. Un calcul immédiat prouve que

$$|x(0)|^{2p-2} x(0) x^*(t) = \int W V_{x, 2p, p}(t, \boldsymbol{\nu}) \exp(-2i\pi \left(\sum_{i=1}^p \nu_i - \sum_{i=p+1}^{2p-1} \nu_i \right) t) d\boldsymbol{\nu} \quad (4.66)$$

Or, tout membre $R_{x, 2p, p}(t, \boldsymbol{\nu})$ de la classe de Cohen s'exprime selon

$$R_{x, 2p, p}(t, \boldsymbol{\nu}) = W V_{x, 2p, p}(t, \boldsymbol{\nu}) * \Pi(t, \boldsymbol{\nu}) \quad (4.67)$$

Ainsi, si nous pouvons obtenir l'opérateur inverse de $\Pi(t, \boldsymbol{\nu})$, soit $\Pi^{-1}(t, \boldsymbol{\nu})$, nous pourrions écrire

$$|x(0)|^{2p-2} x(0) x^*(t) = \int R_{x, 2p, p}(t, \boldsymbol{\nu}) * \Pi^{-1}(t, \boldsymbol{\nu}) \exp(-2i\pi \left(\sum_{i=1}^p \nu_i - \sum_{i=p+1}^{2p-1} \nu_i \right) t) d\boldsymbol{\nu} \quad (4.68)$$

Or, cet opérateur inverse existe clairement si la transformée de Fourier $2p$ dimensionnelle de $\Pi(t, \boldsymbol{\nu})$ ne possède pas de zéro. Mais remarquer que cette transformée de Fourier est $\int F(t, \boldsymbol{\tau}) \exp(-2i\pi \left(\sum_{i=1}^{2p-1} \nu_i \tau_i \right)) d\boldsymbol{\tau}$ donne la conclusion.

Compatibilité avec les convolutions et modulations produits

Soit $f(n, \boldsymbol{\tau}) = \int F(t, \boldsymbol{\tau}) \exp(-2i\pi \left(\sum_{i=1}^{2p-1} \nu_i \tau_i \right)) d\boldsymbol{\tau}$, alors nous donnons les résultats suivants sans démonstration⁶

- Un membre de la classe de Cohen est compatible avec les convolutions, i.e. $(y(t) = x(t) * z(t) \implies R_{y, 2p, p}(t, \boldsymbol{\nu}) = R_{x, 2p, p}(t, \boldsymbol{\nu}) *_t R_{z, 2p, p}(t, \boldsymbol{\nu}))$ si et seulement si

$$f(n, \boldsymbol{\tau} + \boldsymbol{\tau}') = f(t, \boldsymbol{\tau}) f(n, \boldsymbol{\tau}') \quad (4.69)$$

⁶ Les démonstrations ne posent pas de difficulté. Les résultats sont obtenus par identification –voir [28]–.

- Un membre de la classe de Cohen est compatible avec les modulations produits, i.e. $(y(t) = x(t)z(t) \implies R_{y,2p,p}(t, \nu) = R_{x,2p,p}(t, \nu) *_{\nu} R_{z,2p,p}(t, \nu))$ si et seulement si

$$f(n + n', \tau) = f(n, \tau)f(n', \tau) \quad (4.70)$$

Remarque 1 : la transformée de Wigner-Ville d'ordre $2p$ ayant $\delta(t)$ comme noyau, elle vérifie toutes ces propriétés.

Remarque 2 : les résultats obtenus ici sont analogues à ceux de l'ordre deux [28]. Nous pourrions généraliser à d'autres propriétés, mais ceci n'apporterait que peu d'intérêt.

Ce dernier paragraphe termine la partie relative aux signaux déterministes, et nous entrons maintenant dans les extensions pour les signaux aléatoires.

4.4 Signaux aléatoires : multispectres de Wigner-Ville

Les définitions obtenues dans la section précédente concernent les signaux déterministes. Pour pouvoir traiter des signaux aléatoires, ou ne seraient-ce que des signaux déterministes noyés dans du bruit, nous devons généraliser les représentations temps-fréquence d'ordre supérieur au cas stochastique. Rappelons de plus la propriété souhaitable 1 qui stipule que dans le cas de signaux aléatoires stationnaires, une distribution temps-fréquence doit se réduire au multispectre stationnaire. Nous avons déjà débattu sur ce point en disant qu'il serait impossible de réaliser cette propriété en prenant comme définition l'espérance mathématique des outils définis précédemment. Nous allons donc devoir contourner ce problème par un petit artifice, en considérant d'abord des distributions fondées sur les moments, puis en définissant les analogues fondées sur les cumulants. De plus, nous n'envisagerons ici qu'un type de représentation temps-fréquence pour les signaux aléatoires (le terme de représentation devenant d'ailleurs impropre) : les multispectres de Wigner-Ville.

4.4.1 Multispectres de Wigner-Ville fondés sur les moments

Comme l'avaient remarqués W. Martin et P. Flandrin dans le cas de l'ordre 2 [47, 28, 29], prendre comme définition d'un spectre de Wigner-Ville l'espérance mathématique de la transformée de Wigner-Ville nécessite quelques précautions. Dans le cas de l'ordre 2, cette définition est valide si le signal aléatoire considéré est dit harmonisable, c'est-à-dire appartient à la classe $f^{(2)}$ de Blanc-Lapierre et Fortet. Dans le cas des ordres supérieurs, cette condition ne suffit plus, puisque des produits de plus de deux termes interviennent, et nous allons devoir considérer des signaux appartenant aux classes plus restreintes $f^{(k)}$. Pour être plus précis, considérons la représentation d'ordre pair de Wigner-Ville $WV_{x,2p,p}(t, \nu)$. Nous souhaitons adopter comme définition l'espérance mathématique $WV_{x,2p,p}(t, \nu)$, soit la moyenne d'ensemble de

$$WV_{x,2p,p}(t, \nu) = \int x(t_1) \dots x(t_p) x^*(t_{p+1}) \dots x^*(t_{2p}) \exp(-2i\pi(\sum_{i=1}^{2p-1} \nu_i \tau_i)) d\tau \quad (4.71)$$

où

$$\left\{ \begin{array}{l} t_{2p} = t - \frac{1}{2p}(\sum_{i=1}^p \tau_i - \sum_{i=p+1}^{2p-1} \tau_i) \\ t_j = t + \frac{2p-1}{2p}\tau_j - \frac{1}{2p}(\sum_{i=1}^p \tau_i - \sum_{i=p+1}^{2p-1} \tau_i) \quad \forall j = 1, \dots, p \\ t_j = t - \frac{2p-1}{2p}\tau_j - \frac{1}{2p}(\sum_{i=1}^p \tau_i - \sum_{i=p+1}^{2p-1} \tau_i) \quad \forall j = p+1, \dots, 2p-1 \end{array} \right. \quad (4.72)$$

et où le signal étudié $x(t)$ est aléatoire. La question de l'existence de l'écriture précédente se pose. Cette existence est assurée en utilisant le résultat rappelé dans le préambule : si $x(t) \in f^{(2p(2p-1))}$, alors (4.71)

existe en moyenne quadratique, et l'espérance mathématique peut être échangée avec les signes sommes. En conséquence, nous définissons le multispectre de Wigner-Ville d'ordre $2p$ d'un signal de $f^{(2p(2p-1))}$ par

$$WVM_{x,2p,p}(t, \boldsymbol{\nu}) = \int E[x(t_1) \dots x(t_p)x^*(t_{p+1}) \dots x^*(t_{2p})] \exp(-2i\pi(\sum_{i=1}^{2p-1} \nu_i \tau_i)) d\boldsymbol{\tau} \quad (4.73)$$

où (4.72) est vérifié, et où l'on remarque l'ajout de la lettre M dans la notation pour signifier multispectre fondé sur un moment.

Remarque 1 : la même construction s'effectue pour des ordres impairs $2p + 1$. Dans ce cas, les signaux considérés doivent appartenir à la classe $f^{((2p+1)2p)}$.

Remarque 2 : par construction, si le signal $x(t)$ analysé est stationnaire, alors $WVM_{x,2p,p}(t, \boldsymbol{\nu})$ se réduit au multispectre correspondant, construit sur les moments.

4.4.2 Multispectres de Wigner-Ville fondés sur les cumulants

Nous souhaitons maintenant posséder des multispectres de Wigner-Ville construits sur les cumulants. Or, nous venons de définir des multispectres de Wigner-Ville sur des moments et cela pour tout ordre. Remarquons de plus que la définition repose sur l'espérance mathématique du noyau multilinéaire, soit $E[x(t_1) \dots x(t_p)x^*(t_{p+1}) \dots x^*(t_{2p})]$. Nous pouvons alors définir des représentations d'ordre $2p$ fondées sur les cumulants, en remplaçant brutalement dans (4.73) l'opérateur $E[\cdot]$ par l'opérateur $Cum[\cdot]$. L'existence de ce multispectre est assurée par l'existence des multispectres d'ordres inférieurs ou égaux fondés sur les moments. Le multispectre de Wigner-Ville est alors défini par

$$WV_{x,2p,p}(t, \boldsymbol{\nu}) = \int Cum[x(t_1) \dots x(t_p)x^*(t_{p+1}) \dots x^*(t_{2p})] \exp(-2i\pi(\sum_{i=1}^{2p-1} \nu_i \tau_i)) d\boldsymbol{\tau} \quad (4.74)$$

où le système (4.72) est vérifié.

Ce multispectre temps-fréquence de Wigner-Ville répond favorablement aux propriétés souhaitables 1 et 2. En effet, dans le cas stationnaire, il se réduit au multispectre stationnaire fondé sur les cumulants et les composantes indépendantes (au sens statistique) du signal analysé ne provoquent pas d'interférence. Nous préférons cette définition à celle fondée sur les moments, et le terme de multispectre de Wigner-Ville lui fera référence dans la suite.

Remarque : la construction est identique pour des multispectres de Wigner-Ville d'ordre impair.

4.5 Réduction de complexité

Une des propriétés souhaitées des représentations ou spectres temps-fréquence d'ordre 2 est la bonne lisibilité. D'ailleurs, de nombreuses applications utilisent ces objets comme descripteurs et non comme quantificateurs. Il est évident que la lisibilité des représentations d'ordre supérieur est faible !, notre vision se limitant à l'espace. De plus, leur évaluation est gourmande en temps de calcul et en place mémoire. Il faut donc trouver des moyens de réduire cette complexité.

Une possibilité est de regarder certaines tranches de ces représentations (comme il est communément fait pour les multicorrélations). Nous proposons l'approche duale, dont les racines se trouvent dans le principe de contraction des tenseurs cumulants, proposé notamment par J.F. Cardoso en séparation de sources et traitement d'antenne utilisant les statistiques d'ordre supérieur [16]. Cette contraction permet de réduire la complexité en gardant beaucoup d'information. Nous donnons à titre illustratif les définitions de ces versions moins complexes à l'ordre 4.

Définitions à l'ordre 4

Nous allons illustrer cette approche sur l'exemple de l'ordre 4, en commençant par le cas déterministe et en généralisant immédiatement au cas aléatoire. Rappelons que les éléments de la classe de Cohen d'ordre 4

s'écrivent

$$R_{x,4,2}(t, \boldsymbol{\nu}) = \int F(t - \theta, \boldsymbol{\tau}) x\left(\theta + \frac{3\tau_1 - \tau_2 + \tau_3}{4}\right) x\left(\theta + \frac{3\tau_2 - \tau_1 + \tau_3}{4}\right) x^*\left(\theta - \frac{3\tau_3 + \tau_1 + \tau_2}{4}\right) x^*\left(\theta - \frac{\tau_1 + \tau_2 - \tau_3}{4}\right) \exp\left(-2i\pi\left(\sum_{i=1}^3 \nu_i \tau_i\right)\right) d\boldsymbol{\tau} d\theta \quad (4.75)$$

Nous proposons alors de sommer cette représentation le long de deux directions fréquentielles. Il vient alors

$$RC_{x,4,2}(t, \nu) = \int \int R_{x,4,2}(t, \boldsymbol{\nu}) d\nu_1 d\nu_2 \quad (4.76)$$

Les intégrations en ν_1 et ν_2 des exponentielles complexes produisent deux diracs en τ_1 et τ_2 . L'intégration par rapport à ces deux variables conduit alors à l'expression de la distribution contractée

$$RC_{x,4,2}(t, \nu) = \int F(t - \theta, \tau) x\left(\theta + \frac{\tau}{4}\right) x\left(\theta + \frac{\tau}{4}\right) x^*\left(\theta - \frac{3\tau}{4}\right) x^*\left(\theta + \frac{\tau}{4}\right) \exp(-2i\pi\nu\tau) d\tau d\theta \quad (4.77)$$

(où l'on a remplacé $F(t - \theta, \tau, 0, 0)$ par $F(t - \theta, \tau)$)

Dans le cas aléatoire, nous ne referons pas la discussion, le multispectre de Wigner-Ville contracté d'ordre 4 est défini par

$$WVC_{x,4,2}(t, \nu) = \int Cum\left[x\left(t + \frac{\tau}{4}\right), x\left(t + \frac{\tau}{4}\right), x^*\left(t - \frac{3\tau}{4}\right), x^*\left(t + \frac{\tau}{4}\right)\right] \exp(-2i\pi\nu\tau) d\tau \quad (4.78)$$

Remarque 1 : nous obtenons les mêmes définitions pour des multispectres fondés sur les moments en remplaçant dans (4.78) l'opérateur cumulatif par l'opérateur moment. Les notations seront alors obtenues en accolant un M à WVC .

Remarque 2 : dans le cas stationnaire le cumulatif apparaissant dans l'expression précédente peut s'écrire $Cum[x(u), x(u), x^*(u - \tau), x^*(u)]$ qui est indépendant du temps et qui correspond à $C_{x,4,2}(0, 0, \tau)$, et par suite, $WVC_{x,4,2}(t, \nu)$ se réduit au spectre correspondant à cette tranche de la multicorrélation d'ordre 4. La dualité "tranche-contraction" apparaît clairement ici.

4.6 Application à la détection de signaux transitoires

Avant de présenter l'application à la détection de transitoires, faisons quelques remarques quant à l'utilité de telles représentations. Nous avons vu la grande complexité des outils définis et comment les rendre plus simples. Malgré tout, ces représentations ou multispectres temps-fréquence d'ordre supérieur ont la mauvaise idée d'être très difficilement estimables, et par suite, utilisables en pratique. Toutefois, ils reposent sur des arguments solides et peuvent être utilisés pour la compréhension de certains phénomènes et pour guider le développement d'idées de traitement. C'est à cette philosophie que nous adhérons à l'heure actuelle, et l'application qui suit en est le reflet. L'utilisation de ces outils en tant que descripteurs ou quantificateurs est rejetée dans le futur : des travaux sur l'estimation devant être poussés plus loin.

Nous en arrivons à la détection de signaux transitoires noyés dans du bruit gaussien. Un signal transitoire est un signal bref à l'échelle de la durée d'observation du signal qui le contient. Détecter un transitoire représente beaucoup d'intérêt, surtout pour la surveillance, comme on peut les rencontrer en contexte de guerre sous-marine. Nous nous intéressons ici à la détection dite passive, c'est-à-dire que le signal bref n'est pas émis par l'opérateur effectuant la surveillance, mais est issu de l'objet à surprendre. Il s'agira par exemple de bruits brefs provoqués dans des bateaux ennemis, surveillés par un sous-marin.

De nombreuses approches existent. La plus simple consiste à utiliser l'énergie et à suivre son évolution au cours du temps. Les bruits de fond ayant des énergies ne variant pas beaucoup temporellement, l'occurrence d'un signal transitoire provoquera une augmentation de l'énergie. Ce type de détection fonctionne bien lorsque le rapport signal à bruit est suffisamment élevé.

D'autres approches ont récemment été proposées reposant sur des changements d'espaces de représentations (ou de description) du signal. Des méthodes basées sur des transformations linéaires du signal (type Gabor ou ondelettes) ont été étudiées par principalement Friedlander et Porat [33] (voir également [34]). Ces méthodes utilisent le fait que les coefficients obtenus par la transformée sont gaussiens (si le bruit l'est aussi) de moyenne variant lors de la présence d'un transitoire. Ces auteurs ont donc proposé une approche par rapport de vraisemblance qui donne de bons résultats, encore une fois lorsque le rapport signal à bruit est suffisant. Une deuxième voie a été initiée par Hinich [39] en profitant du caractère non stationnaire du signal transitoire. Hinich propose de tester la nullité du bispectre dans une zone hors la multiplicité stationnaire. Enfin, une méthode originale, simple et (ou mais) empirique est apparue en 1992. Elle est due à Persson et Sangfelt [48], et nous allons la décrire maintenant en détails.

4.6.1 Méthode de Persson et Sangfelt

La méthode proposée ici repose sur une mesure d'asymétrie, temporelle et d'amplitude. Une mesure naturelle d'asymétrie est le cumulants d'ordre trois normalisé (skewness en anglais). Persson et Sangfelt ont utilisé une mesure d'asymétrie construite à partir du bispectre [48]. Ils disposent d'une mesure pour l'asymétrie temporelle et une pour l'asymétrie d'amplitude. Ces mesures sont respectivement⁷

$$d_t = \sum_{\nu_1, \nu_2} \Im[S_{x,3}(\nu_1, \nu_2)]^2 \quad (4.79)$$

$$d_a = \sum_{\nu_1, \nu_2} \Re[S_{x,3}(\nu_1, \nu_2)]^2 \quad (4.80)$$

où le domaine de sommation est le triangle principal de définition du bispectre. En effet, on peut vérifier que si $x(-t) = x(t)$ (symétrie temporelle) alors le bispectre est réel, alors que si $x(-t) = -x(t)$ (symétrie d'amplitude) le bispectre est imaginaire pur. Ces mesures sont d'autant plus fortes que l'asymétrie étudiée est prononcée. Nous ne ferons pas la distinction entre ces deux mesures, mais allons envisager leur somme, soit

$$d = \sum_{\nu_1, \nu_2} |S_{x,3}(\nu_1, \nu_2)|^2 \quad (4.81)$$

En tant que somme des deux précédentes, si une des asymétries est forte, cette mesure le sera également. L'astuce de Persson et Sangfelt est de rendre "évolutive" cette mesure. Autrement dit, il font une mesure à court terme de l'asymétrie. Leur algorithme est alors le suivant :

1. Calcul du bispectre dans une fenêtre (cette fenêtre est suffisamment longue pour obtenir une estimation correcte du bispectre).
2. Calcul de la mesure d pour cette position de fenêtre.
3. Avancée de la fenêtre et retour au point 1.

Le fonctionnement de cette méthode est alors clair. Lorsque la fenêtre se trouve dans une zone de bruit seul, ce bruit étant supposé gaussien, la mesure est faible. Par contre, lorsqu'un transitoire est présent dans la fenêtre, ses disymétries sont vues et la mesure est grande. Ces auteurs ont validé leur méthode sur des signaux réels maîtrisés, les résultats étant très satisfaisants. Nous montrons maintenant pourquoi leur méthode est reliée à la théorie objet de ce chapitre.

4.6.2 Lien avec les multispectres temps-fréquence

L'analyse de la méthode de Persson et Sangfelt présentée précédemment doit nous guider intuitivement : la méthode est fondée sur une analyse bispectrale à court terme et utilise le bispectrogramme.

En effet, le détecteur peut se réécrire sous la forme (en utilisant des notations continues)

$$d(t) = \int |\hat{S}_{x,3,t}(\nu_1, \nu_2)|^2 d\nu_1 d\nu_2 \quad (4.82)$$

⁷ \Re et \Im dénotent respectivement les parties réelle et imaginaire d'un nombre complexe.

Nous avons placé t en indice pour montrer que le bispectre est estimé localement autour de la date t . Plus précisément, plaçons nous à la date t fixée, et examinons la grandeur calculée. Dans leur exposé, Persson et Sangfelt prennent une fenêtre rectangulaire de longueur T , la découpent en N petites fenêtres de longueur M . Sur chacune des N fenêtres, ils calculent le bipériodogramme, puis moyennent sur les N fenêtres. Ceci s'écrit plus formellement

$$\frac{1}{N} \sum_{i=0}^{N-1} X_i(\nu_1) X_i(\nu_2) X_i^*(\nu_1 + \nu_2) \quad (4.83)$$

où

$$X_i(\nu) = \int \Pi_T((t + iM) - \tau) x(\tau) \exp(-2i\pi\nu\tau) d\tau = ST_{x,\Pi}((t + iM), \nu) \quad (4.84)$$

est la transformée de Fourier à court terme de $x(t)$ à la date particulière $t + iM$. Ainsi, l'expression (4.82) est clairement une estimée, à la date t , du bispectrogramme pour les signaux aléatoires, soit une estimée de

$$E[B_{x,3,\Pi}(t, \nu_1, \nu_2)] = E[ST_{x,\Pi}(t, \nu_1) ST_{x,\Pi}(t, \nu_2) ST_{x,\Pi}^*(t, \nu_1 + \nu_2)] \quad (4.85)$$

Le détecteur de Persson et Sangfelt est donc une version estimée de

$$d(t) = \int |E[B_{x,3,\Pi}(t, \nu_1, \nu_2)]|^2 d\nu_1 d\nu_2 \quad (4.86)$$

Or, nous avons vu que le bispectrogramme est un membre de la classe de Cohen d'ordre 3, et qu'il est obtenu par triple convolution entre la transformée de Wigner-Ville de sa fenêtre et la transformée de Fourier du signal étudié. En conclusion le détecteur "parfait" de Persson et Sangfelt est

$$d(t) = \int |E[|WV_{x,3,2}(t, \nu_1, \nu_2)| * * * WV_{\Pi,3}^*(t, \nu_1, \nu_2)]|^2 d\nu_1 d\nu_2 \quad (4.87)$$

La méthode entre donc dans le cadre des applications de la théorie des représentations temps-fréquence d'ordre supérieur.

4.6.3 Extensions naturelles

Nous proposons dans ce paragraphe deux extensions à cette approche.

Détecteur fondé sur une autre représentation d'ordre 3

Nous avons vu précédemment que les transformées de Wigner-Ville possèdent la propriété de conservation du support temporel d'un signal. Ainsi, puisque le bispectrogramme résulte d'une convolution de la transformée de Wigner-Ville, il n'a pas cette propriété. Or, à des fins d'estimation de temps d'arrivée et de durée du transitoire, il est intéressant de conserver son support temporel. Ceci n'étant pas possible avec le détecteur de Persson et Sangfelt, nous proposons naturellement d'utiliser une représentation d'ordre 3 possédant la propriété de conservation du support. La transformée de Wigner-Ville est un candidat idéal, et le détecteur s'écrit alors

$$d(t) = \int |E[|WV_{x,3,2}(t, \nu_1, \nu_2)|]^2 d\nu_1 d\nu_2 \quad (4.88)$$

Ce résultat n'est que théorique, puisqu'en pratique il faudra lisser la représentation, et par suite élargir le support du motif correspondant au transitoire. Mais on peut imaginer de lisser fortement en fréquence et faiblement en temps, ce que permet la représentation de Wigner-Ville mais pas le bispectrogramme.

Et l'ordre 4 ?

La seconde extension concerne le passage de l'ordre 3 aux ordres supérieurs. Pratiquement, nous étendrons à l'ordre 4 en proposant (en gardant l'idée de conservation du support présentée avant)

$$d(t) = \int |WV_{x,4,2}(t, \nu_1, \nu_2, \nu_3)|^2 d\nu_1 d\nu_2 d\nu_3 \quad (4.89)$$

Il faut noter que l'espérance mathématique a disparu et qu'il s'agit de la version aléatoire de la représentation de Wigner-Ville, soit le trispectre de Wigner-Ville.

Ce détecteur sera utilisé dans la suite, bien qu'implanté d'une façon différente, évitant l'estimation du trispectre de Wigner-Ville.

4.6.4 Interprétation en termes d'information

Cette section relate complètement l'approche que nous avons eue pour la définition du détecteur de transitoire que nous proposons. Notre première idée réside dans l'utilisation de critères d'information rendus évolutifs, des réflexions ultérieures nous ayant alors permis de relier ces idées à la théorie des représentations temps-fréquence d'ordre supérieur. Nous développons maintenant ces deux points.

Détection de transitoire et critères entropiques

Dans "The mathematical Theory of Communication" [56], C.E. Shannon définit l'entropie d'un ensemble de fonctions. Cet ensemble de fonctions peut être l'ensemble des réalisations d'un processus stochastique. La définition de Shannon s'appuie donc sur les descriptions maintenant classiques des processus aléatoires. Etant donné $x(t)$ un tel processus, il est courant de le caractériser par l'ensemble des variables aléatoires multidimensionnelles induites, soit la collection des vecteurs $\{\mathbf{X}_n\}_{n=1,2,\dots} = (x(t_1), \dots, x(t_n))$. En associant à chaque vecteur une densité de probabilité $p_{\mathbf{X}_n}(\mathbf{x})$, la description du processus est complète. Shannon définit alors l'entropie du processus par degré de liberté selon

$$H(x) = - \lim_{n \rightarrow \infty} \frac{1}{n} \int p_{\mathbf{X}_n}(\mathbf{x}) \ln(p_{\mathbf{X}_n}(\mathbf{x})) d\mathbf{x} \quad (4.90)$$

Cette définition est bien adaptée aux processus strictement stationnaires, puisque la densité $p_{\mathbf{X}_n}(\mathbf{x})$ est alors indépendante d'un quelconque instant de référence. Ceci rend indépendante la limite dans (4.90) d'un tel instant. Par contre, pour un processus non stationnaire, la limite dépend du choix des temps (il suffit de considérer à titre d'exemple un processus dont la puissance change soudainement de valeur).

Ce problème vient du fait que les définitions sont données pour des processus stochastiques, donc relatives à l'ensemble des réalisations possibles. Il faudrait pouvoir obtenir des définitions pour chaque réalisation, la validité de ces définitions étant apportée par des arguments ergodiques lors de l'étude de processus stationnaires.

Nous n'avons pour l'instant pas de réponse à ce souhait. Toutefois, nous parlerons de processus ergodiques au sens de l'entropie, si l'on est capable de calculer l'entropie d'un processus en utilisant une seule de ses réalisations. Ceci peut se faire en utilisant des histogrammes comme estimés de densités, puis en envisageant comme définition de l'entropie la limite pour des résolutions de plus en plus fines de la formule classique de l'entropie calculée avec ces histogrammes.

Calculer l'entropie sur une réalisation permet alors d'envisager des interprétations en terme d'information sur un signal. Une réalisation d'un processus stationnaire aura une entropie constante au cours du temps, c'est-à-dire que l'information nécessaire à la description du signal est constante au cours du temps. Par contre, un processus non stationnaire verra son entropie varier dans le temps. Supposons qu'un signal change brutalement de puissance à un certain instant t . Jusqu'à t , l'entropie du signal est constante, mais le changement de puissance correspond à un apport d'information, que l'entropie à partir de t est capable de chiffrer.

Pour revenir à la détection de transitoire, la discussion précédente rend clair le point de vue que nous allons adopter : *un transitoire en tant que signal bref correspond à une bouffée d'information dans le signal qui le contient. L'information contenue dans ce signal doit donc évoluer dans la zone où existe le transitoire.*

Une variante à ce point de vue peut être avancée. Nous utiliserons la néguentropie, qui chiffre la différence d'entropie entre un signal et le signal gaussien de mêmes première et seconde statistiques. Notre second point de vue est alors le suivant *un signal transitoire est en général déterministe, bref, à nature oscillante, et donc peu gaussien. La néguentropie du signal contenant le transitoire doit donc évoluer dans la zone où vit le transitoire.*

Cette discussion provient d'idées très intuitives qui mériteraient de nombreux approfondissements. Quoiqu'il en soit, le problème posé maintenant est l'obtention d'une estimation évolutive de l'entropie. Or, les travaux

de P. Comon sur l'approximation de la néguentropie en utilisant les développements de densités de probabilités en série de Edgeworth nous apportent la clé. Il est en effet montré dans [19, 20], qu'une approximation de la néguentropie d'un vecteur aléatoire \mathbf{X} normalisé, à distribution symétrique⁸ est

$$N(\mathbf{X}) \approx \frac{1}{48} \sum_{i,j,k,l} Cum[x_i, x_j, x_k, x_l]^2 \quad (4.91)$$

Considérons alors \mathbf{X}_t une famille de vecteurs de \mathcal{R}^n contenant les échantillons successifs du signal $x(t)$ étudié, soit

$$\mathbf{X}_t = (x_1, \dots, x_n)^T = (x_{t-1}, \dots, x_{t-n})^T \quad (4.92)$$

Nous obtenons alors une néguentropie évolutive du signal $x(t)$ selon, après normalisation,

$$N(\mathbf{X})(t) \approx \frac{1}{48} \sum_{i,j,k,l=1}^n Cum[x_i, x_j, x_k, x_l]^2 \quad (4.93)$$

Cette mesure doit être très faible dans les zones vides de transitoire et très forte dans une zone où existe un transitoire.

Dans la section suivante, nous effectuons le lien avec les représentations temps-fréquence d'ordre 4.

Lien avec les représentations temps-fréquence

Revenons à l'interprétation physique du multispectre d'ordre 4 de Wigner-Ville. En examinant sa définition, il est clair que ce multispectre s'obtient par transformée de Fourier de la quantité

$$C_{x,4,2}(t, \boldsymbol{\tau}) = Cum\left[x\left(t + \frac{3\tau_1 - \tau_2 + \tau_3}{4}\right), x\left(t + \frac{3\tau_2 - \tau_1 + \tau_3}{4}\right), x^*\left(t - \frac{3\tau_3 + \tau_1 + \tau_2}{4}\right), x^*\left(t - \frac{\tau_1 + \tau_2 - \tau_3}{4}\right)\right] \quad (4.94)$$

qui apparaît comme une multicorrélation locale de $x(t)$. Comme $C_{x,4,2}(t, \boldsymbol{\tau})$ et $WV_{x,4,2}(t, \boldsymbol{\nu})$ sont reliés par une transformée de Fourier des variables $\boldsymbol{\tau}$ vers $\boldsymbol{\nu}$, la relation de Parseval montre que

$$\int |WV_{x,4,2}(t, \boldsymbol{\nu})|^2 d\boldsymbol{\nu} = \int |C_{x,4,2}(t, \boldsymbol{\tau})|^2 d\boldsymbol{\tau} \quad (4.95)$$

Or, la néguentropie d'un vecteur aléatoire \mathbf{X} centré, à distribution symétrique est approchée par

$$N(\mathbf{X}) = \frac{1}{48} \sum_{i,j,k,l} Cum[x_i, x_j, x_k, x_l]^2 \quad (4.96)$$

Ainsi, si le vecteur \mathbf{X} est une collection d'échantillons successifs d'un signal aléatoire stationnaire, alors en considérant (4.93) et (4.95) nous avons l'approximation suivante

$$d(t) \approx 48N(\mathbf{X}) \quad (4.97)$$

Implantation adaptative

Nous utiliserons comme détecteur la néguentropie. Soit $\{x_n\}$ un processus aléatoire discret. La version discrète du détecteur est

$$d(n) = \frac{1}{48} \sum_{i,j,k,l} Cum[x_{n-i}, x_{n-j}, x_{n-k}, x_{n-l}]^2 \quad (4.98)$$

⁸Les transitoires sont en général dissymétriques, et l'on peut se demander pourquoi nous n'utilisons ici que les cumulants d'ordre quatre, qui ne mesurent pas la dissymétrie. En toute rigueur, nous devrions incorporer dans les développements qui suivent les termes liés aux cumulants d'ordre trois qui prennent en compte les informations de dissymétries. Ceci est omis par soucis de clarté.

où les indices varient entre⁹ 0 et N . Nous pouvons alors obtenir une version récursive ou adaptative de ce détecteur, en remplaçant dans (4.98) le tenseur cumulant par une estimation récursive. Nous utilisons alors la version nouvelle décrite dans le préambule.

Remarque : comme l'estimateur récursif est asymptotiquement gaussien, notre test ressemble étrangement à un test du χ^2 . Toutefois, il n'est pas normalisé par la variance de l'estimateur du cumulant, et par suite, n'en est pas un!

La dernière section de ce chapitre est destinée à montrer le bien fondé de cette approche, en l'illustrant sur des signaux synthétiques.

4.6.5 Simulations

Rappelons tout d'abord la forme de l'algorithme utilisé. Soit $x(n) = t(n) + b(n)$ le signal aléatoire contenant le transitoire $t(n)$ noyé dans le bruit gaussien blanc $b(n)$. A chaque instant n , nous créons le vecteur¹⁰ $\mathbf{X}_n = (x_n^1, \dots, x_n^N)^T = (x_{n-1}, \dots, x_{n-N})^T$. Le vecteur est ensuite normalisé, et nous obtenons alors une estimation récursive du tenseur cumulant d'ordre 4 de ce vecteur (sans changer de notation après la normalisation) selon

$$\begin{cases} \kappa_n^{k,l} &= \kappa_{n-1}^{k,l} + \mu(x_n^k x_n^l - \kappa_{n-1}^{k,l}) \\ \kappa_n^{i,j,k,l} &= \kappa_{n-1}^{i,j,k,l} + \gamma(x_n^i x_n^j x_n^k x_n^l - [3]x_n^i x_n^j \kappa_{n-1}^{k,l} - \kappa_{n-1}^{i,j,k,l}) \end{cases} \quad (4.99)$$

La néguentropie est alors estimée selon

$$N(\mathbf{X}_n) = \frac{1}{48} \sum_{i,j,k,l=1}^N (\kappa_n^{i,j,k,l})^2 \quad (4.100)$$

Quelques simulations

Nous présentons dans cette partie quelques simulations illustrant la méthode proposée. Le transitoire envisagé est la somme de deux sinusoides modulée par une fenêtre de Hamming étroite. Les signaux comportent 4096 points, le transitoire ayant un support d'environ 400 points. Une première simulation est effectuée pour un rapport signal à bruit¹¹ de -5.5 dB. La figure (4.1) montre le transitoire bruité. Le signal transitoire est discernable aux environs de l'échantillon 2000. Le détecteur d'énergie, évalué également récursivement à l'aide de la matrice de corrélation, est représenté figure (4.2). La détection du transitoire est réussie. Le détecteur d'ordre 4 apparaît figure (4.3). Pour les deux détecteurs, la fenêtre choisie est de longueur 5. Nous remarquons que le contraste pour l'ordre 4 est plus fort que dans le cas du détecteur d'énergie. Quoiqu'il en soit, l'ordre 2 est ici suffisant.

La deuxième simulation est faite pour un rapport signal à bruit de -10 dB. Sur le signal bruité (figure (4.4)), il est impossible de discerner le transitoire. Les figures (4.5) et (4.6) montrent respectivement les détecteurs d'ordre 2 et 4 pour une fenêtre de 5 points. Le transitoire est encore détectable par l'énergie, mais le contraste est très faible, et on peut raisonnablement penser que la probabilité de fausse alarme est élevée. Par contre, le contraste pour l'ordre 4 est fort, et la détection en sera meilleure.

Ceci est confirmé par les figures (4.7) à (4.9). Le rapport signal à bruit est toujours de -10 B, mais nous utilisons une autre réalisation de bruit. De plus, la longueur de la fenêtre est maintenant de 7 échantillons. L'apport d'énergie du au transitoire est maintenant noyé dans la variance d'estimation de l'énergie (figure (4.8)), et la détection n'est pas obtenue à l'ordre 2. Par contre, le détecteur d'ordre 4 se comporte bien, et permet une bonne détection (contraste fort, figure (4.9)).

Bien entendu, ces simulations sont bien incomplètes. Nous devons maintenant réaliser des études de performances. Il faut donc obtenir les courbes COR et les courbes de probabilité de détection en fonction du rapport signal à bruit. De plus, une comparaison avec le filtre adapté est envisagée, puisque le filtre adapté est le récepteur linéaire optimal pour la détection de signaux certains dans du bruit gaussien.

⁹ N définit la taille de la fenêtre d'observation, et est l'analogie de la fenêtre de Persson et Sangfelt.

¹⁰ Notons que l'indice est placé en exposant !

¹¹ Le rapport signal à bruit est défini dans la zone de support du transitoire par $10 \log \frac{E(t)}{E(b)}$ où $E(\cdot)$ représente l'énergie, t et b le transitoire et le bruit respectivement.

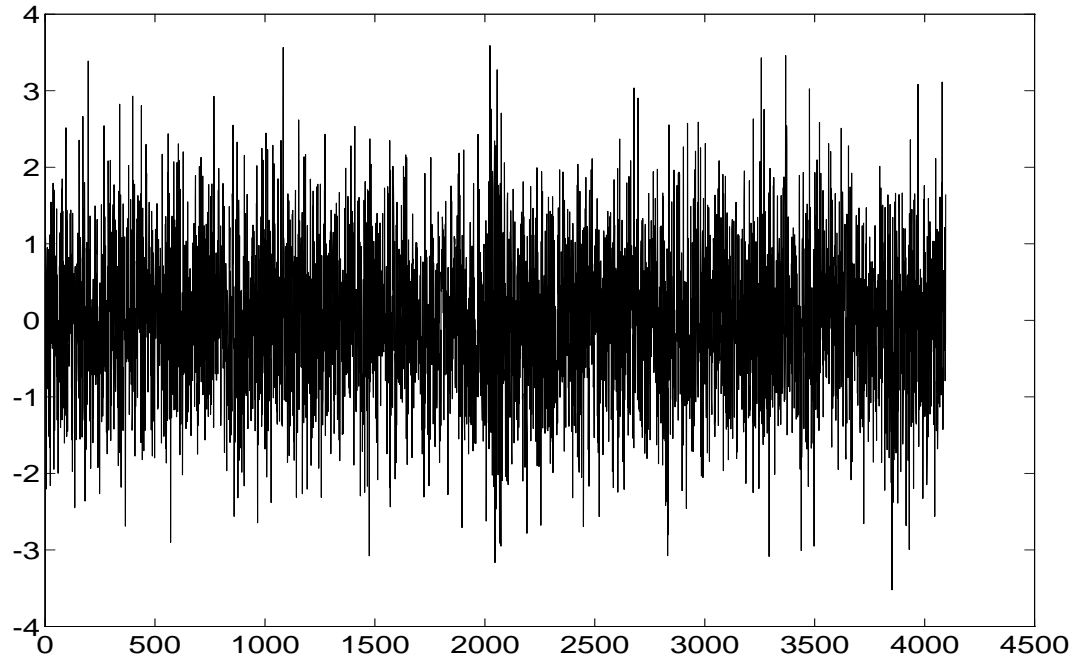


Figure 4.1: Signal transitoire

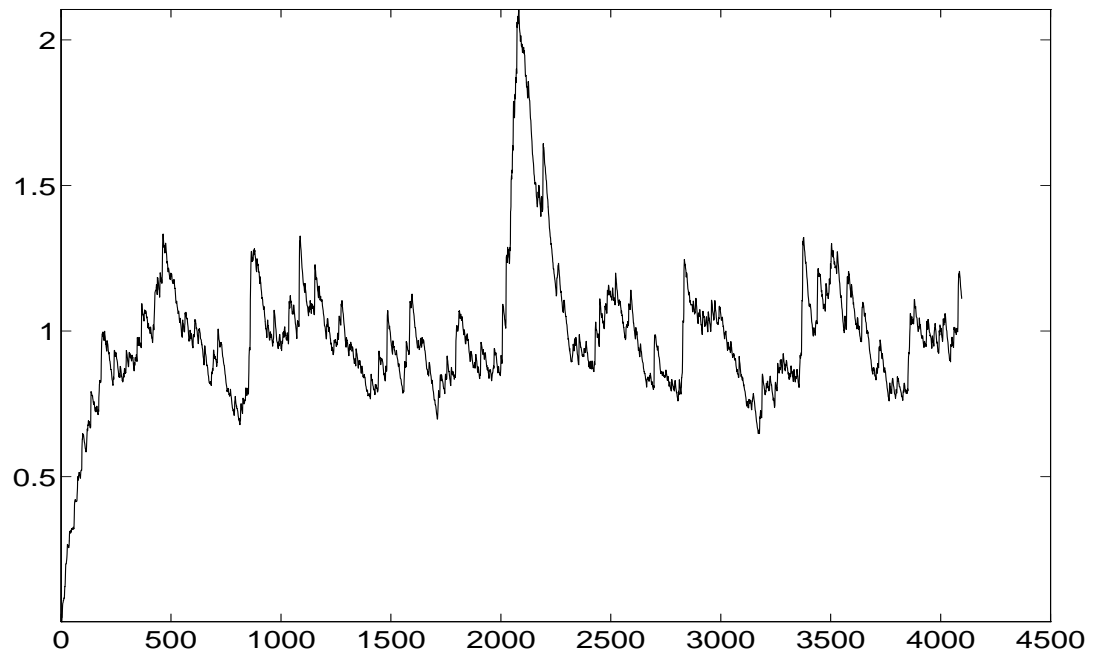


Figure 4.2: Energie

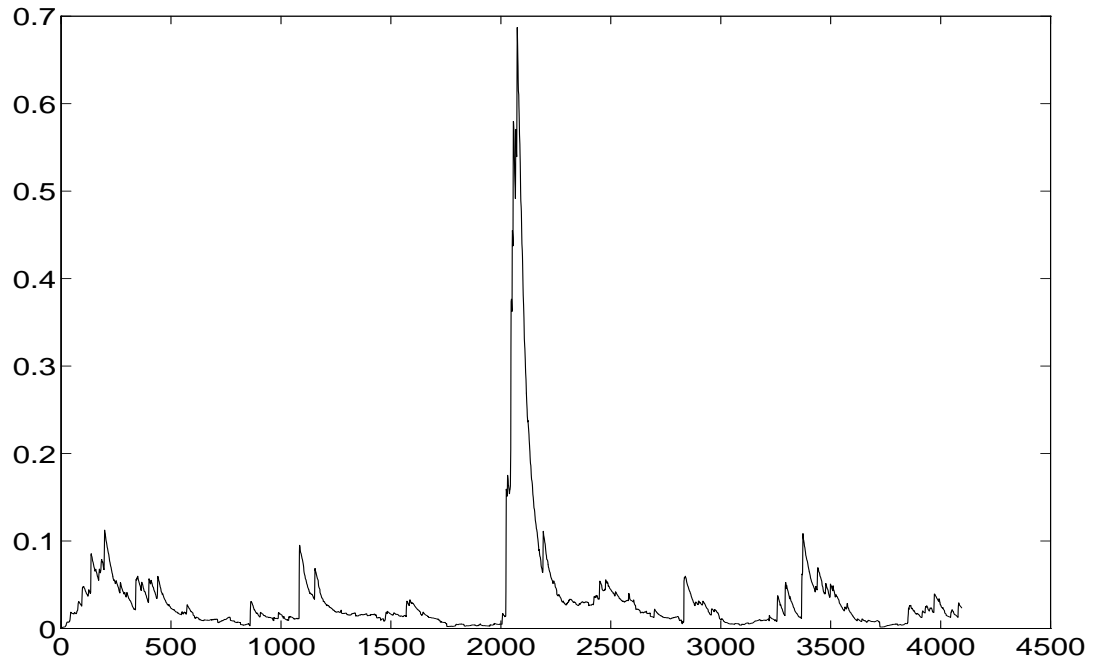


Figure 4.3: Détecteur d'ordre 4

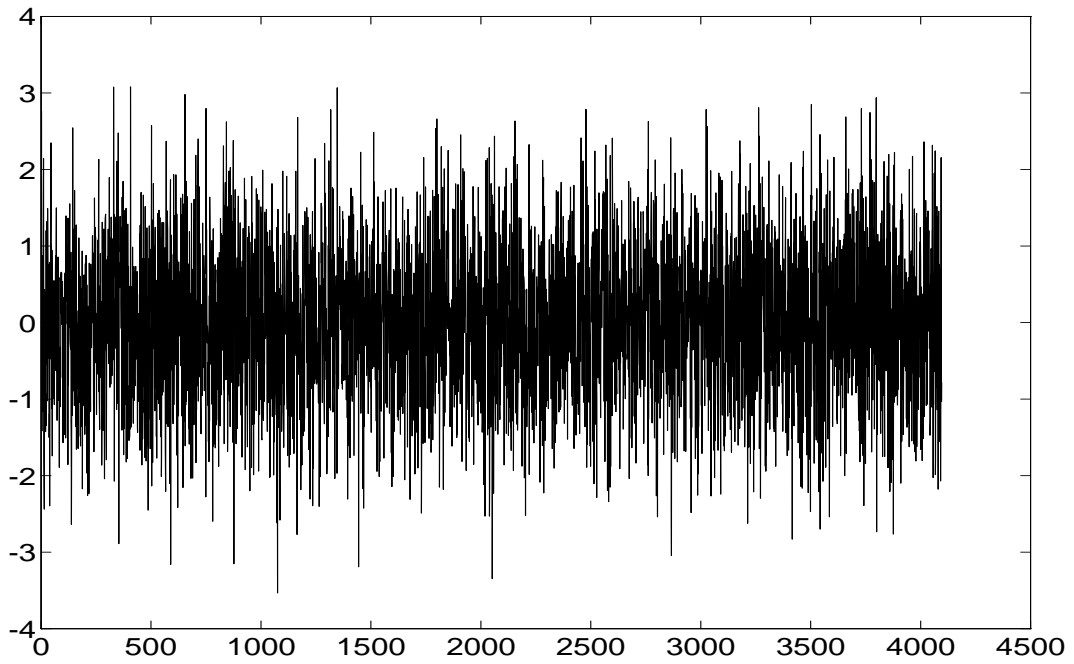


Figure 4.4: Signal transitoire

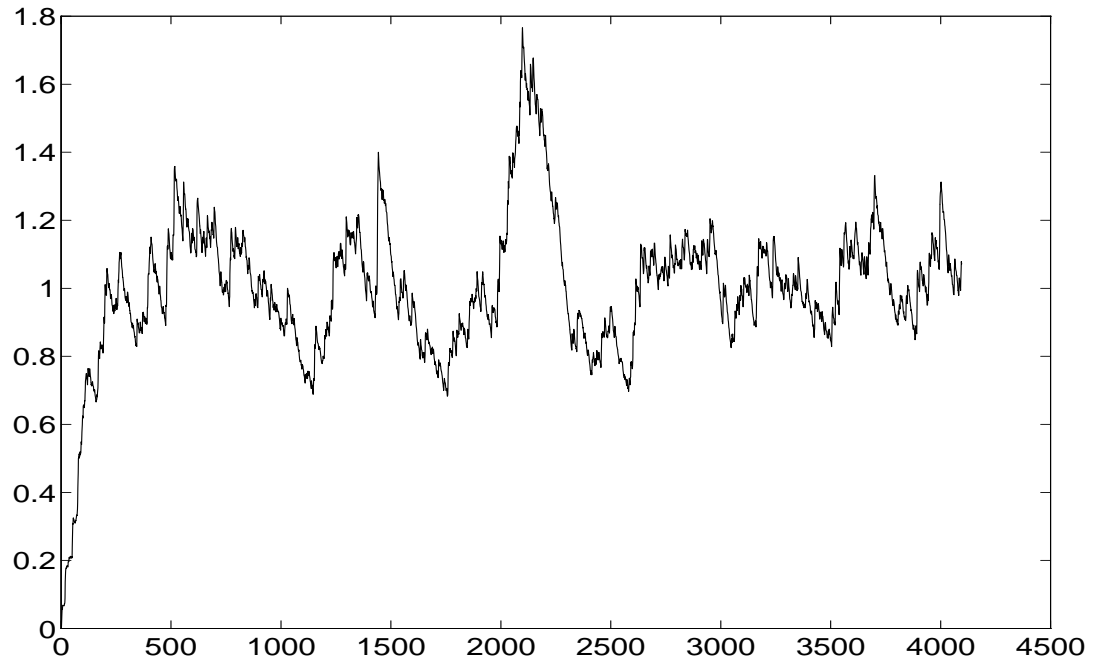


Figure 4.5: Energie

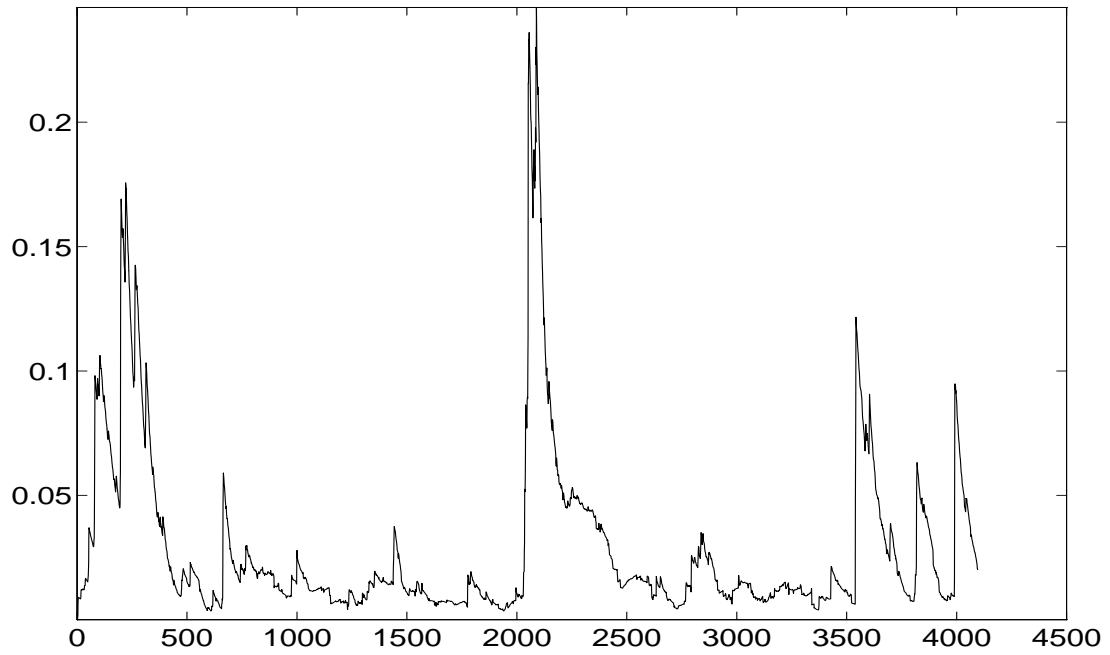


Figure 4.6: Détecteur d'ordre 4

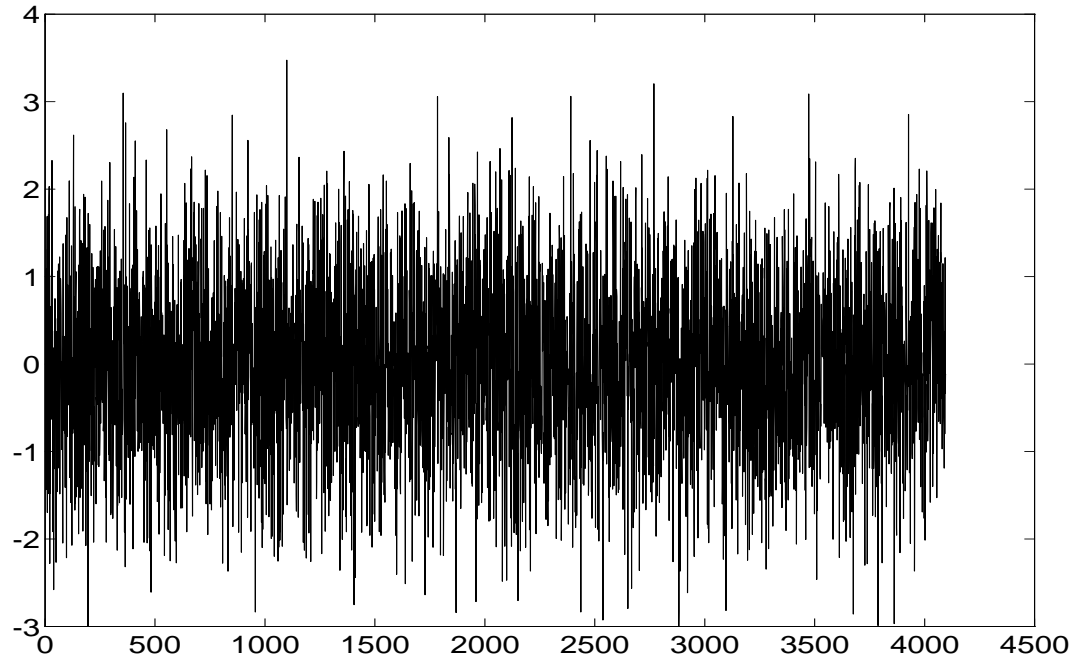


Figure 4.7: Signal transitoire

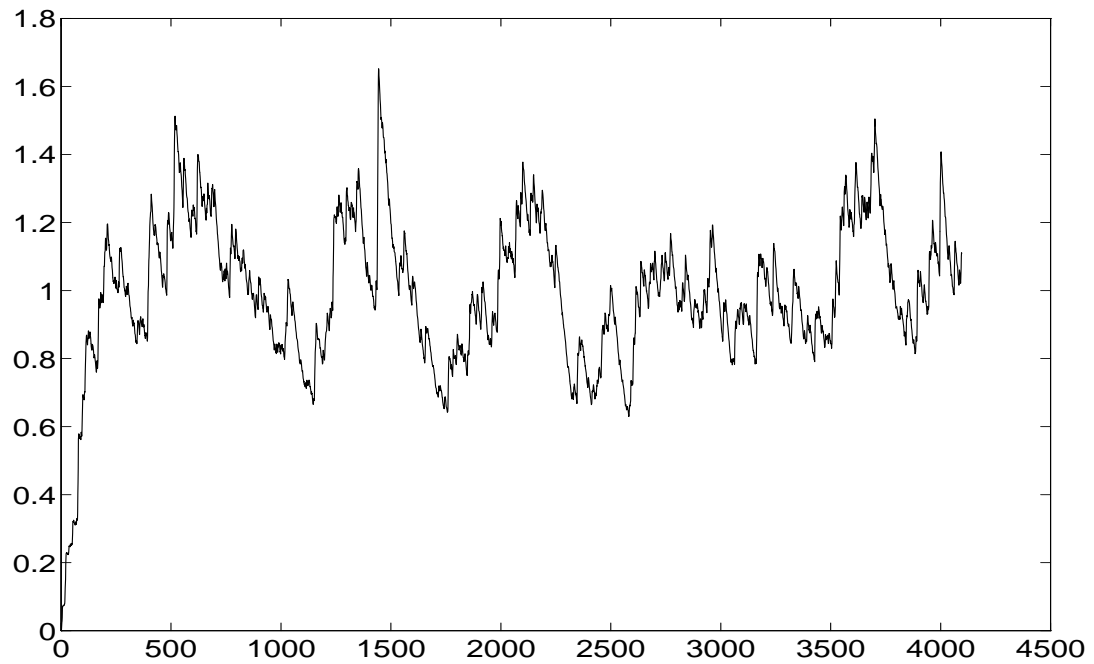


Figure 4.8: Energie

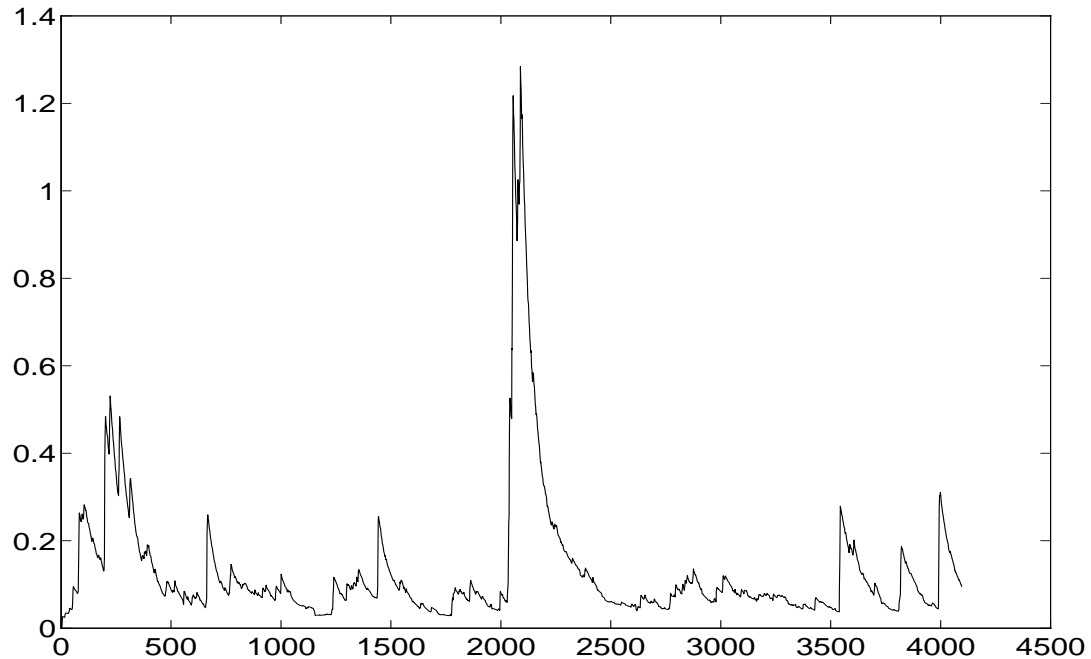


Figure 4.9: Détecteur d'ordre 4

4.7 Conclusion

Nous avons dans ce chapitre proposé une approche déductive de l'extension aux ordres supérieurs des représentations temps-fréquence. Mais comme il apparaît tout au long du développement, ces outils sont très compliqués.

Notre sentiment à l'heure actuelle est de les utiliser comme outils nous permettant de comprendre les propriétés d'ordre supérieur des phénomènes non stationnaires, et par suite de nous permettre de dégager des idées de traitements. Ceci est illustré par l'application à la détection des signaux transitoires que nous avons proposée ici. Bien que l'idée soit apparue en marge des représentations temps-fréquence, nous avons en effet montré comment elle s'inscrit dans la théorie très générale développée dans ce chapitre.

D'autres approches pourraient être développées pour décrire les signaux non gaussiens non stationnaires. Il faut en particulier tenir compte de toute information a priori sur les signaux manipulés, ces informations pouvant conduire à des méthodes simples. De plus, des méthodes paramétriques, adaptées à ces informations peuvent également être envisagées.

Conclusion

La promenade se termine. Partis de la description des signaux non gaussiens stationnaires, nous sommes passés par le monde des filtres de Volterra qui peuvent engendrer ces signaux ; nous avons enfin élargi l'horizon en essayant de comprendre un peu mieux les signaux les plus généraux : non gaussiens (non linéaires) non stationnaires.

Avant de terminer ce mémoire, nous voudrions présenter les points non abordés ici, et avancer quelques idées.

La grande absente de ce travail est l'estimation. Nous ne nous sommes pas du tout souciés des problèmes pratiques liés à l'estimation des multi-outils ou des multispectres de Wigner-Ville. Pourtant leur utilisation nécessite des études approfondies de leurs estimateurs. Si ces études n'ont pas été faites, quelques arguments peuvent être avancés.

En ce qui concerne les multi-outils pour les signaux stationnaires, des études ont été réalisées pour les signaux à valeurs réelles dans les années soixante [13, 14]. Les variances d'estimation obtenues sont évidemment de l'ordre de deux fois l'ordre du multi-outil utilisé ; ceci limite l'utilisation d'ordres élevés, et explique l'emploi maintenant courant des ordres trois et quatre. Les études pour les signaux à valeurs complexes peuvent certainement être menées de la même façon, et conduiront probablement aux mêmes types de résultats.

Le plus gros problème d'estimation est rencontré pour les représentations temps-fréquence. Comme les signaux sont non-stationnaires, les moyennages ne peuvent être que des lissages. Les études à l'ordre deux ont conduit à la définition d'estimateurs plus ou moins performants, telle par exemple la distribution Pseudo-Wigner-Ville Lissée. Cette distribution est un élément de la classe de Cohen pour des signaux déterministes. Le noyau de cette représentation est séparé en les deux variables, et réalise deux moyennages, en temps et en fréquence, de façon indépendante.

La définition d'estimateurs aux ordres supérieurs est délicate, puisque les multispectres de Wigner-Ville sont basés sur les cumulants, combinaisons non-linéaires de moments. Les structures retenues seront alors compliquées, et l'obtention des variances d'estimation très difficile. Nous pensons toutefois que les idées appliquées à l'estimateur récursif du cumulatif d'ordre quatre présenté dans ce travail pourraient conduire à des estimateurs performants et plus simples.

Un dernier point technique concernant l'utilisation des représentations temps-fréquence d'ordre supérieur est leur discrétisation, et les conditions d'échantillonnage des signaux à analyser. Des études utilisant la théorie des distributions peuvent être menée (comme il est fait dans ce mémoire pour les filtres de Volterra) pour obtenir les versions discrètes des représentations (il est à noter que des travaux allant dans ce sens se trouvent dans [30]).

L'utilisation des représentations temps-fréquence d'ordre supérieur ne deviendra valide que lorsque ces études auront été faites. Pour l'instant, elles nous apportent une compréhension un peu meilleure de certains phénomènes, et nous permettent de dégager des idées de traitement.

Une autre voie d'investigation semble intéressante pour les représentations temps-fréquence et concerne la définition de représentations temps-multiéchelle¹².

Il peut être montré que la représentation de Wigner-Ville, qui génère la classe de Cohen, est également centrale dans la classe des représentations affines bilinéaires (temps-échelle), dont l'élément le plus connu est le scalogramme, module carré de la transformée en ondelettes continue. Nous avons essayé de généraliser aux ordres supérieurs l'approche conduisant de la représentation de Wigner-Ville aux représentations affines. Malheureusement, nous n'avons pas réussi. Le problème fondamental à notre avis réside dans la compréh-

¹²Cette idée est initialement due à P. Flandrin.

sion de ce qu'est une échelle, qui n'est pas un concept aussi bien défini que celui de fréquence. Toutefois, les réflexions doivent se poursuivre car nous pensons que de tels outils pourraient être intéressants dans l'étude de la turbulence ou des processus multifractals, ou des interactions non-linéaires entre diverses échelles interviennent.

Les problèmes d'estimation n'ont pas non plus été abordés dans l'étude des filtres de Volterra. Toutefois, nous pensons qu'ils sont d'importance lorsque des procédures orthogonales d'identification sont adoptées. En effet, nous avons vu que les approches orthogonales permettent une sélection automatique simple de termes en examinant la diminution en erreur quadratique moyenne. Mais cette erreur comme critère de sélection n'est peut-être pas le mieux adapté, car les erreurs d'estimations des coefficients ne sont pas pris en compte.

Un dernier point concerne les approches d'identification orthogonale. Toutes les méthodes proposées reposent sur des traitements par blocs de données : l'influence d'un terme est évalué sur tout le signal. Nous dirons que ces méthodes sont adaptatives en ordre, mais pas en temps. Elles sont donc très intéressantes dans un contexte d'identification de système, mais peu en Traitement du Signal. Nous réfléchissons à une procédure adaptative en temps. L'algorithme proposé dans ce travail est une première étape, mais ne permet pas de sélection automatique. Il devra être modifié ou oublié.

Enfin, le lecteur peut se demander pourquoi les filtres de Volterra n'ont pas été présentés pour les signaux à valeurs complexes. La réponse est simple : c'est complexe ! Toutefois, quelques idées peuvent être avancées. Nous avons vu que dans les chapitres 2 et 4 que manipuler des signaux à valeurs complexes impose de jouer à la fois sur les termes conjugués et non conjugués. Des versions complexes des filtres de Volterra devront respecter cette "règle". Mais la complexité du formalisme en sera accrue. Par exemple, un filtre quadratique pur complexe devra considérer les termes du type x^*x^* , xx^* et xx pour rester général. Par contre, le domaine d'application des filtres peut restreindre cette généralité. En traitement d'antenne, les signaux sont circulaires jusqu'à un certain ordre, et l'on peut penser que seul les termes du type xx^* seraient à considérer dans le cas quadratique. Notons que des travaux ont déjà été effectués dans [17] pour des filtres de Volterra discrets à ordre et mémoire finis.

Nous avons essayé de montrer quelques aspects des statistiques d'ordre supérieur pour les signaux non gaussiens, non linéaires et non stationnaires, et d'en dégager les intérêts. Si quelques applications ont été proposées, nous avons occulté la multitude qui existe dans diverses branches du Traitement du Signal.

Les statistiques d'ordre supérieur ont leur place en communications (égalisation aveugle, synchronisation), en sismique et d'une façon générale pour les problèmes inverses (déconvolution aveugle), en traitement d'antenne (séparation aveugle de sources), en traitement d'image, ...

Les statistiques d'ordre supérieur permettent la résolution de problèmes restés incompris avec les techniques d'ordre deux. Leur capacité de description et la quantité d'information qu'elles véhiculent sont plus importantes que pour les statistiques d'ordre deux. Et de là vient leur force et leur légitimité. Qu'elles soient appliquées directement ou à travers des filtres non-linéaires, leur importance pour le Traitement du Signal n'est plus à démontrer. De plus, la puissance informatique toujours croissante n'est plus un obstacle à leur maniement.

Les statistiques d'ordre supérieur ouvrent la voie à de nouveaux traitements et déclenchent de nouveaux virages.

Non-gaussien, non-linéaire, non-stationnaire.

Trois non-propriétés qui partagent maintenant le gouvernement du Traitement du Signal avec leurs contraintes.

Bibliographie

- [1] P.O. Amblard et J.M. Brossier, *A new adaptive estimator of the fourth-order cumulant*, soumis à Signal Processing, avril 1993.
- [2] P.O. Amblard et J.L. Lacoume, *A deductive construction of third-order time-frequency distributions*, à paraître dans Signal Processing, Vol. 36, 3, 1994.
- [3] P.O. Amblard et D. Baudois, *Noise cancellation using Volterra filters : A real case study*, Proc. IEEE Winter Workshop on Non-Linear Digital Signal Processing, Tampere, Finlande, janvier 1993.
- [4] P.O. Amblard, J.L. Lacoume et J.M. Brossier, *Transient detection, Higher-Order time-frequency distributions and the entropy*, Proc. IEEE Workshop on Higher-Order Statistics, pp 265-269, Lake Tahoe, USA, juin 1993.
- [5] P.O. Amblard, D. Baudois, J.L. Lacoume, *Edgeworth series expansion of the conditional mean and the optimality of non-linear Volterra filters*, Proc. IEEE ICASSP, ppV477-V480, San Francisco, USA, mars 1992.
- [6] J. Bass, Cours de Mathématiques, Tome III, Masson, 1971.
- [7] D. Baudois, C. Servièrre, A. Silvent, *Soustraction de bruit : analyse et synthèse bibliographique*, Traitement du Signal, Vol. 6, 5, 1989.
- [8] A. Benveniste, M. Métivier, R. Priouret, Algorithmes adaptatifs et approximations stochastiques, Masson, 1987.
- [9] A. Blanc-Lapierre, R. Fortet, Théorie des fonctions aléatoires, Masson, 1953.
- [10] P. Bondon, Statistiques d'ordre supérieur et modélisation en traitement du signal, Thèse de Doctorat, Paris XI Orsay, 1991.
- [11] P. Bondon, B. Picinbono, *De la blancheur et de ses transformations*, Traitement du Signal, Spécial Non-Linéaire, Non-Gaussien, Vol. 7, 5, 1990.
- [12] D. Brillinger, Time series analysis : data analysis and theory, MacGraw-Hill, 1981.
- [13] D. Brillinger, *An introduction to polyspectra*, Ann. Math. Stat., Vol. 36, pp1351-1374, 1965.
- [14] D. Brillinger, M. Rosenblatt, *Asymptotic theory of k^{th} order spectra*, in Spectral Analysis of Time Series, pp153-188, Wiley, 1967.
- [15] D. Brillinger, M. Rosenblatt, *Computation and interpretation of k^{th} order spectra*, in Spectral Analysis of Time Series, pp189-232, Wiley, 1967.
- [16] J.F. Cardoso, *Higher-order narrow-band array processing*, in Higher-Order Statistics, J.L. Lacoume ed., Elsevier, 1992.
- [17] P. Chevalier, B. Picinbono, P. Duvaut, *Le filtrage de Volterra transverse réel et complexe*, Traitement du Signal Spécial Non-Linéaire, Non-Gaussien, Vol. 7, 5, 1990.

- [18] P. Comon et J.L. Lacoume, Statistiques d'ordre supérieur, Cours de l'Ecole d'été de physique des Houches, Septembre 1993.
- [19] P. Comon, *Independent component analysis*, in Higher-Order Statistics, J.L. Lacoume ed., Elsevier, 1992.
- [20] P. Comon, *Analyse en composantes indépendantes et identification aveugle*, Traitement du Signal Spécial Non-Linéaire, Non-Gaussien, Vol. 7, 5, 1990.
- [21] P. Comon, *Séparation de mélanges de signaux*, Proc. GRETSI 1989, Juan-les-Pins.
- [22] P. Durand, Théorie et estimation conventionnelle du bispectre. Application aux processus linéaire-quadratiques à spectres de raies, Thèse de Doctorat, INP grenoble, 1991.
- [23] P. Duvaut, Traitement du signal : concepts et applications, Hermès, 1991.
- [24] P. Duvaut, *Non-linear filtering in Signal Processing*, in Higher-Order Statistics, J.L. Lacoume ed., Elsevier, 1992.
- [25] P. Duvaut, *Le filtrage de Wiener linéaire-quadratique à horizon fini. Application à la prédiction*, Traitement du Signal, Vol. 6, 3, 1989.
- [26] P. Duvaut, *Influence des moments d'ordre trois sur la structure d'un algorithme adaptatif linéaire-quadratique*, Traitement du Signal Spécial Non-Linéaire, Non-Gaussien, Vol. 7, 5, 1990.
- [27] P. Duvaut, *A unifying and general approach to adaptive linear-quadratic discrete time Volterra filtering*, Proc. ICASSP 1989, Glasgow.
- [28] P. Flandrin, Représentations temps-fréquence des signaux non-stationnaires, Thèse de Doctorat d'Etat, INP Grenoble, 1987.
- [29] P. Flandrin, Temps-fréquence, Hermès, 1993.
- [30] J. Fonollosa, C. Nikias, *Wigner higher-order moment spectra : definition, properties, computation and application to transient signal analysis*, IEEE trans. Signal Processing, SP. 41, 1, 1994.
- [31] G. Fowles, Introduction to modern optics, Dover, 1975.
- [32] M. Fréchet, *Sur les fonctionnelles continues*, Ann. Sci. Ecole Normale Sup., Vol. 27, pp193-219, 1910.
- [33] B. Friedlander, B. Porat, *Performance analysis of transient detectors based on a class of linear data transforms*, IEEE trans. Information Theory, IT 38, 2, 1992.
- [34] M. Frisch, H. Messer, *The use of the wavelet transform in the detection of unknown transient signals* IEEE trans. Information Theory, IT 38, 2, 1992.
- [35] N. Gerr, *Introducing a third-order Wigner distribution*, Proc. IEEE, Vol. 76, 3, 1988.
- [36] G. Giannakis, A. Dandawaté, *Polyspectral analysis of non-stationary signals : bases, consistency and HOS-WV*, in Higher-Order Statistics, J.L. Lacoume ed., Elsevier, 1992.
- [37] P. Gill, G. Gollub, et. al., *Methods for modifying matrix factorization*, Math. Comp., Vol. 28, 126, 1974.
- [38] J.F. Guerre-Chaley, Etude de différentes structures de soustracteurs de bruits adaptatifs multiréférences, Thèse de Doctorat, INP Grenoble, 1990.
- [39] M. Hinich, *Detecting a transient signal by bispectral analysis*, IEEE Trans. on ASSP, ASSP 38, 7, 1990.
- [40] A. Jazwinski, Stochastic processes and filtering theory, Academic Press, 1970.
- [41] K.I. Kim, E.J. Powers, *A digital method of modeling quadratically nonlinear systems with a general random input*, IEEE trans. ASSP, ASSP 36, 11, 1988.

- [42] M. Korenberg, L. Paarmann, *Orthogonal approaches to time-series analysis and system identification*, IEEE Signal Processing Magazine, juillet 1991.
- [43] J.L. Lacoume, M. Gaeta, *Complex random variables : a tensorial approach*, in Higher-Order Statistics, J.L. Lacoume ed., Elsevier, 1992.
- [44] V. Leonov, A. Shirayev, *On a method of calculation of semi-invariants*, Th. Prob. and Appl., Vol. 4, 3, 1959.
- [45] O. Macchi, M. Bellanger, *Le point sur le filtrage adaptatif transverse*, Proc. Colloque GRETSI, Nice, France, juin 1987.
- [46] P. MacCullagh, *Tensor methods in statistics*, Chapman and Hall, 1987.
- [47] W. Martin, P. Flandrin, *Wigner-Ville spectral analysis on non-stationary processes*, IEEE trans. ASSP, ASSP 33, 6, 1985.
- [48] L. Persson, E. Sangfelt, *Detection of hydroacoustic transients by means of higher-order statistics*, Proc. EUSIPCO, Bruxelles, Belgique, 1992.
- [49] S.W. Nam, S.B. Kim, E.J. Powers, *On the identification of a third-order Volterra nonlinear system using a frequency-domain block RLS adaptive algorithm*, proc. ICASSP, pp2407-2410, Albuquerque, USA, 1990.
- [50] B. Picinbono, P. Duvaut, *Optimal Linear-Quadratic Systems for Detection and Estimation*, IEEE trans. ASSP, ASSP 34, 2, 1988.
- [51] B. Picinbono, P. Duvaut, *Geometrical Properties of Optimal Volterra Filters for Signal Detection*, IEEE trans. ASSP, ASSP 36, 5, 1990.
- [52] B. Picinbono, *Variables et signaux aléatoires circulaires*, communication au GDR TSI, groupe 9, réunion du 1er octobre 1990.
- [53] B. Picinbono, *Geometrical concepts in higher-order statistics*, in Higher-Order Statistics, J.L. Lacoume ed., Elsevier, 1992.
- [54] F. Roddier, *Distributions et transformation de Fourier*, MacGraw Hill, 1978.
- [55] C. Servière, *Eléments de comparaison entre différentes méthodes de soustraction de bruit*, Thèse de Doctorat, INP Grenoble, 1988.
- [56] C. Shannon, W. Weaver, *The mathematical theory of communication*, University of Illinois Press, 1963.
- [57] M. Schetzen, *The Volterra and Wiener theories of non-linear systems*, Krieger, 1980.
- [58] A. Shirayev, *Some problems in the spectral theory of higher-order moments*, Th. Prob. and Appl., Vol. 5, 3, 1960.
- [59] C. Soize, *Méthodes mathématiques en analyse du signal*, Masson, 1993.
- [60] A. Swami, *Third-order Wigner distributions : definitions and properties*, Proc. ICASSP, pp3081-3084, Toronto, Canada, 1991.
- [61] J. Ville, *Théorie et applications de la notion de signal analytique*, Cable et Transmission, 2ème année, 1, 1948.
- [62] B. Widrow et.al., *Adaptive noise cancelling : principles and applications*, Proc. IEEE, pp1695-1716, 1975.
- [63] N. Wiener, *Non-Linear problems in random theory*, MIT Press, 1958.
- [64] R. Zaknoune, *Soustraction adaptative de bruit par prédiction vectorielle*, Thèse de Doctorat, INP Grenoble, 1993.